N° ordre : 337 - 2014

THESE de DOCTORAT

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE : Evolution Ecosystèmes Microbiologie Modélisation

Spécialité : Ecologie des communautés, fonctionnement des écosystèmes, écotoxicologie

Soutenue publiquement le 9 décembre 2014 par :

Judith Trapp

Approches protéomiques pour le développement de biomarqueurs chez l'amphipode d'eau douce *Gammarus fossarum* : découverte et caractérisation de protéines impliquées dans la fonction reproductrice

JURY

M. Jean-Pierre THOME, Professeur, Université de Liège	Rapporteur
M. Vianney PICHEREAU, Professeur, Université de Bretagne Occidentale	Rapporteur
M. Jean-Michel DANGER, Professeur, Université du Havre	Examinateur
M. Pierre MARMONIER, Professeur, Université Lyon 1	Examinateur
M. Jean ARMENGAUD, Directeur de Recherche, CEA Marcoule	Directeur de thèse
M. Olivier GEFFARD, Chargé de Recherche, Irstea Villeurbanne	Directeur de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. François-Noël GILLY**
Vice-président du Conseil d'Administration	M. le Professeur Hamda BEN HADID
Vice-président du Conseil des Etudes et de la Vie Universitaire	M. le Professeur Philippe LALLE
Vice-président du Conseil Scientifique	M. le Professeur Germain GILLET
Directeur Général des Services	M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. ETIENNE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : Mme la Professeure C. BURILLON
Faculté d'Odontologie	Directeur : M. le Professeur D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme la Professeure C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. le Professeur Y. MATILLON
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme. la Professeure A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. F. DE MARCHI
Département Biologie	Directeur : M. le Professeur F. FLEURY
Département Chimie Biochimie	Directeur : Mme Caroline FELIX
Département GEP	Directeur : M. Hassan HAMMOURI
Département Informatique	Directeur : M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur : M. le Professeur Georges TOMANOV
Département Mécanique	Directeur : M. le Professeur H. BEN HADID
Département Physique	Directeur : M. Jean-Claude PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	Directeur : M. B. GUIDERDONI
Polytech Lyon	Directeur : M. P. FOURNIER
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur C. VITON
Ecole Supérieure du Professorat et de l'Education	Directeur : M. le Professeur A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	Directeur : M. N. LEBOISNE

REMERCIEMENTS

Je tiens en premier lieu à remercier mes directeurs de thèse : Olivier Geffard et Jean Armengaud. J'ai du mal à trouver les mots pour vous exprimer ma gratitude de m'avoir donné l'opportunité d'effectuer ma thèse au sein de vos deux laboratoires. Ce projet de thèse fut pour moi passionnant, à l'interface de vos deux disciplines, j'ai passé trois années très enrichissantes à votre contact. Vous m'avez donné une grande liberté d'entreprendre mais sans jamais me laisser me disperser. Je vous remercie infiniment pour votre optimisme et votre confiance. Jean, je te remercie d'avoir apporté cette touche d'exotisme inattendu et fascinant dans le monde des bactéries, et de m'avoir fait découvrir cette terre jusque-là inconnue qu'était la MS et toutes les opportunités qu'elle offre. Je te remercie d'avoir eu la volonté de me former non seulement à la recherche mais également au métier de chercheur, ainsi que pour tous tes conseils et tes encouragements. Olivier, je te remercie infiniment de m'avoir inclue dans ton projet de recherche « gammare » pour lequel j'ai la plus grande considération. C'est une grande joie pour moi de savoir que ces travaux seront poursuivis et déboucheront vers une application dans le domaine de la biosurveillance. Je te remercie pour tous les moments de « brain storming » passés dans ton bureau qui m'ont tant appris ainsi que pour ta générosité (même si des fois j'aimerais bien réussir à t'offrir une bière !).

Je tiens à remercier l'ensemble des membres du jury qui ont accepté d'évaluer mon travail. Je remercie Dr. Jean-Pierre Thomé et Dr. Vianney Pichereau pour avoir accepté d'être rapporteurs ainsi que Dr. Jean-Michel Danger et Dr. Pierre Marmonier d'être examinateurs.

J'adresse mes remerciements à mes financeurs, l'Irstea et le CEA par le biais du programme transversal de toxicologie, ainsi que l'ANR par le projet « Protéogam ».

Je souhaite remercier toutes les personnes qui ont activement participé à ces travaux. A Arnaud, mon 3ème encadrant, pour m'avoir initié à la phylogénie et m'avoir expliqué à de multiples reprises la différence entre paralogue et orthologue. J'ai énormément appris à ton contact. A J-C, pour ta passion de la MS que tu sais si bien transmettre. Partager son travail entre deux labos est parfois un casse-tête mais j'ai toujours pu compter sur toi pour faire avancer les choses en mon absence. A Gilles et Olivier P., pour m'avoir accompagné dans le challenge du big-data et pour avoir aidé à donner un sens biologique à toutes ces séquences protéiques ! A Yassine pour son implication dans la dernière manip de tox et à Khedidja pour avoir contribué à imager un peu mes travaux ! Je remercie également Anne-Hélène et Guylaine pour leur aide en biochimie.

Je tiens également à remercier les membres de mon comité de pilotage : Jeanne Garric, Eric Quemeneur, Nathalie Mondy, Arnaud Salvador et Catherine Moali pour avoir contribuer à nos échanges scientifiques.

Je remercie mes prédécesseurs qui ont su répondre à mes diverses interrogations. A Guillaume, mon voisin de bureau, pour avoir toujours répondu avec patience à mes « pourquoi-ci et pourquoi-ça ? ». A Emilie pour m'avoir expliqué les subtilités du comptage spermatique ainsi que Benoit, initiateur des travaux sur la Vtg. Je souhaite bon courage aux prochains !

Ayant eu la chance de réaliser ma thèse entre deux laboratoires, mon environnement scientifique a été particulièrement stimulant et je remercie tous mes collègues pour y avoir contribué. Au personnel du CEA : Emie pour son accueil et son hospitalité, François (mais pas pour m'avoir complexée à jamais sur mes qualités de dactylo), Erica (j'espère venir te visiter prochainement aux US !), Céline, Alain L., Christine, Laetitia, Carine, Ramon et Damien. Au personnel d'Irstea : A mes poules Delphine et Amandine. Delphine, j'espère que cette connexion quasi « mystique » perdurera entre nous, ta présence quotidienne me manquera énormément. Amandine, également transfuge de Metz, merci pour tous les pétages de plomb de 17h15 et pour tous ces échanges scientifiques fascinants, on se retrouve en post-doc ? A Romain, mon voisin du bureau envahissant mais tellement attachant ! A Remy (bon vent à BIOMAE !), Hervé (j'ai hésité … mais bon je sais bien qu'au fond tu aurais aimé être alsacien, je te pardonne), Madame Adeline (merci de nous faire toujours profiter de tes escapades gourmandes), Paul (respire, tu es tranquille à présent !), Alex, Renaud, Patrice, Marc, Marion, Jean-Phi, Manuella, Juliette ainsi que Vanessa pour son aide dans les méandres administratifs. J'espère n'avoir oublié personne !

Une petite pensée pour toutes les personnes qui m'ont soutenu, poussé, challengé au cours de ma scolarité : mes potes de fac, mes profs, mes pauvres binômes qui m'ont subi (sauf Hadrien, là c'est moi qui a subi), mes premiers mentors de stage. Je pense notamment au personnel du LIEBE, à Sandrine, Vincent, Laetitia, Simon et Maël.

Je termine avec une mention à ma famille. A mes beaux-parents, Françoise, pour avoir traqué les fautes de français de tous mes rapports de stage et Philipe, pour avoir grandement contribué à me fournir un logement « habitable » en cette dernière année. A ma mère, pour m'avoir fourni toute l'aide possible afin de mener à bien mes études et m'avoir toujours pousser à saisir les opportunités qui s'offrent à moi. Enfin, à mon meilleur ami, Raphi. Ta présence a ponctué toute ma scolarité, de la rentrée de 6ème, en passant par l'annonce des résultats du bac, jusqu'à la thèse. Je te remercie d'avoir épousé la bonne femme compliquée que je suis et j'espère te voir présent à mon pot de départ en retraite.

TABLE DES MATIERES

LISTE DES ABREVIATIONS VII

LISTE DES FIGURES ET DES TABLEAUX XI

LISTE DES PUBLICATIONS ET DES COMMUNICATIONS XV

INTRODUCTION 1

CHAPITRE I : Démarches et outils pour l'évaluation écotoxicologique des milieux aquatiques – effets sur la reproduction 7

1. Les biomarqueurs pour l'évaluation de la qualité biologique des hydrosystèmes 9

1.1. Approches pour la surveillance de la qualité environnementale 9

1.2. Le rôle des biomarqueurs dans l'évaluation écotoxicologique 10

1.2.1. Classification 10

1.2.2. Exemples de biomarqueurs couramment employés en écotoxicologie 11

1.2.3. Intérêt et pertinence des biomarqueurs 14

1.3. Bilan de l'utilisation des biomarqueurs en biosurveillance environnementale : exemple avec les organisme sentinelles du genre *Gammarus* 15

1.3.1. Utilisation de l'organisme sentinelle *Gammarus fossarum* 17

1.3.2. Contraintes actuelles pour l'utilisation des biomarqueurs en biosurveillance 27

2. Endocrinologie et perturbations de la reproduction chez les crustacés 34

2.1. Reproduction et endocrinologie des crustacés 34

2.1.1. Reproduction 35

2.1.2. Endocrinologie 43

2.2. Perturbations de la fonction reproductrice chez les crustacés 53

2.2.1. Exemples d'altération de la fonction reproductrice chez des populations naturelles 53

2.2.2. Origine de la perturbation de la fonction reproductrice 54

2.3. Outils pour l'évaluation de la reprotoxicité chez les crustacés 58

3. Apports des nouveaux outils omiques en écotoxicologie 62

3.1. La toxicologie environnementale à l'ère des omiques 62

3.1.1. Génomique et transcriptomique 63

3.1.2. Protéomique 65

3.1.3. Vers une approche multi-omiques 67

3.2. Outils et étapes de l'analyse protéomique 69

3.2.1. Les méthodes séparatives 70

3.2.2. L'analyse par spectrométrie de masse 73

3.2.3. Stratégies pour l'interprétation des données protéomiques 78

3.3. Applications et limites des outils protéomiques en écotoxicologie aquatique 84
4. Valorisation sous la forme de revue bibliographique 89
4.1. Les outils protéomiques pour le renouveau de l'approche biomarqueurs en écotoxicologie 91
4.2. La protéogénomique au secours des espèces non-modèles 121
5. Objectifs des travaux 149
CHAPITRE II : Matériels et méthodes 153
1. Biotests chez *Gammarus fossarum* 155
1.1. Matériel biologique 155
1.1.2. Maintien des organismes en laboratoire 156
1.1.3. Identification et acquisition des échantillons biologiques 157
1.2. Expérimentations mises en œuvre pour répondre aux objectifs de la thèse 158
1.3. Détermination des stades de mue et du développement embryonnaire 162
1.4. Dénombrement spermatique 163
2. Analyses moléculaires 164
2.1. Analyse transcriptomique 164
2.2. Analyse protéomique 166
2.2.1. Extraction des protéines 166
2.2.2. Digestion enzymatique des protéines 168
2.2.3. Techniques séparatives du protéome 170
2.2.4. Séquençage du protéome par spectrométrie de masse 172
3. Traitement informatique des données de séquençage 173
3.1. Assemblage *de novo* du transcriptome 173
3.2. Interprétation des spectres MS/MS et identification des séquences peptidiques 174
3.3. Interprétation des données 175
CHAPITRE III : Découverte de protéines impliquées dans la reproduction de *Gammarus fossarum* 177
1. Découverte de protéines impliquées dans la fonction reproductrice par protéogénomique 179
2. Caractérisation de l'ovogénèse et identification des protéines de réserves du développement embryonnaire 211
3. Exploration du protéome des caeca hépatopancréatiques 237
4. Problématique du séquençage *de novo* du transcriptome 243
CHAPITRE IV : Impacts des xénobiotiques sur la reproduction masculine 255
CHAPITRE V : Synthèse et discussion générale 289
1. Les biomarqueurs chez les invertébrés : limites et nouvelles opportunités 291
1.1. Limites à l'utilisation des biomarqueurs 291

1.2. Intérêt des nouvelles techniques omiques 292
1.3. La protéogénomique pour la découverte de protéines chez les espèces non modèles 293
2. Mise en place d'une approche protéogénomique chez *Gammarus fossarum* 295
2.1. Création d'un catalogue de protéines 295
2.2. La base de données ORF : fondement pour une bonne annotation structurale 296
2.2.1. Détection d'erreurs d'assemblage du transcriptome 296
2.2.2. Stratégies pour une meilleure annotation structurale des protéines 297
2.3. Intérêt de la ressource protéogénomique 300
3. Développement d'outils pour l'évaluation de la reprotoxicité 302
3.1. Une première identification par annotation fonctionnelle automatique 302
3.2. Identification de protéines impliquées dans la physiologie reproductrice 303
3.2.1. Protéomique comparative des tissus reproducteurs mâles et femelles 303
3.2.2. Protéines impliquées dans l'ovogénèse 305
3.2.3. Protéines impliquées dans la spermatogénèse 308
3.3. Impact de la perturbation endocrine sur la reproduction chez le mâle 310
3.3.1. Suivi de la qualité spermatique pour l'évaluation de l'infertilité masculine 310
3.3.2. Modifications du protéome face à la pression toxique 313
CONCLUSION & PERSPECTIVES 317
BIBLIOGRAPHIE 321
ANNEXES 339

LISTE DES ABREVIATIONS

20-HE : 20-hydroxyecdysone

2D-DIGE : Differential Gel Electrophoresis

2D-PAGE : électrophorèse sur gel bidimensionnelle

AChE : acétylcholinestérase

AGH : Androgen Gland Hormone

BD : base de données

BLAST : Basic Local Alignement Search Tool

CAT : catalase

Cd : cadmium

CHH : crustacean hyperglycemic hormone

CYP450 : cytochromes P450

DCE : directive cadre sur l'eau

DDT : dichlodiphényltrichloroéthane

EC : Effective Concentration

EcR : ecdysone receptor

ELISA : enzyme-linked immunosorbent assay

EROD : ethoxyresorufin-o-deethylase

ESI : électrospray

EST : expressed sequence tag

G3PDH : glycéraldéhyde-3-phosphate déshydrogénase

GA : glande androgène

GIH : Gonado-Inhibiting Hormone

GSH : Gonado-Stimuling Hormone

GST : glutathion S-transférase

HSP: Heat Shock Protein

IAA : iodoacétamide

ICAT : Isotope-Coded Affinity Tag

iTRAQ : Isobaric tags for relative and absolute quantitation

JH : hormone juvénile

KEGG : Kyoto Encyclopedia of Genes and Genomes

LC : chromatographique liquide

LLTP : large lipid transfert protein

MALDI : Matrix Assisted Laser Desorption Ionisation

Met : Methoxyfenozide

MF : méthyle-farnesoate

MIH : Molt Inhibiting Hormone

NOEC : No Observed Effect Concentration

MOIH : Mandibular Organ Inhibiting Hormone

MS : spectrométrie de masse

MS/MS : spectrométrie de masse en tandem

MudPIT : Multidimensional Protein Identification Technology

m/z : rapport masse-sur-charge

NCBI : National Center for Biotechnology Information

NGS : Next Generation Sequencing

OM : organe mandibulaire

Orbitrap : trappe orbitale

ORF : Open Reading Frame (cadres ouverts de lecture)

OX-GS : organe X/glande du sinus

OY : organe Y

PCR : Polymerase Chain Reaction

PE : perturbateurs endocriniens

pI : point isoélectrique

PMF : peptide mass fingerprinting

POD : peroxydase

Pyr : pyriproxyfen

Q : analyseur quadripôle

RNA-Seq : RNA-sequencing

SC : spectral count

SDS-PAGE : sodium dodecyl sulfate polyacrylamide gel electrophoresis

SOD : superoxyde dismutase

TFA : acide trifluoroacétique

TOF : analyseur « Time of flight »

Trap : trappe ionique

USP : ultraspiracle

VIH : Vitellogenesis Inhibiting Hormone

VSOH : Vitellogenesis- Stimuling Ovarienne Hormone

Vtg : vitellogénine

LISTE DES FIGURES ET DES TABLEAUX

CHAPITRE I : Démarches et outils pour l'évaluation écotoxicologique des milieux aquatiques – effets sur la reproduction

Figure I-1 : Représentation des différents niveaux d'organisation biologique et exemples de paramètres mesurables pour l'évaluation de la qualité environnementale (Clements 2000; Lagadic et al. 2008) 9
Figure I-2 : Démarches dans l'évaluation du risque écotoxicologique et représentation séquentielle de réponses induites au sein d'un individu en réponse à un stress toxique (d'après Lagadic et al. 2008) 10
Figure I-3 : Relation entre le niveau d'expression du biomarqueur et l'état de santé de l'individu face à un stress toxique d'intensité croissante (adapté de Depledge and Fossi 1994) 14
Figure I-4 : Vue de la phylogénie animale chez les invertébrés basée sur des données moléculaires (Halanych 2004) 16
Figure I-5 : Vue d'ensemble de la phylogénie des pancrustacés d'après des données moléculaires (d'après von Reumont *et al.* 2012) 18
Tableau I-1 : Classification de *Gammarus fossarum* (Martin and Davis 2001) 19
Figure I-6 : Vue latérale d'un Gammaridae (d'après Chevreux and Fage 1970; Roux 1970) 19
Figure I-7 : Uropode 3 de *Gammarus fossarum* (A) et de *Gammarus pulex* (B) (Source : Quéau H., Irstea Lyon) 20
Figure I-8 : Répartition géographique de *Gammarus fossarum*(selon Barnard and Barnard 1983) 21
Tableau I-2 : Récapitulatif des différentes études employant des biomarqueurs protéiques chez les gammares 24
Tableau I-3 : Caractérisation de l'influence des facteurs confondants sur la réponse des biomarqueurs chez le gammare 30
Tableau I-4 : Caractérisation *in-situ* de la variabilité intrinsèque de la réponse du biomarqueur chez le gammare 31
Figure I-9 : Représentation schématique en vue latérale de la position des principaux organes du système reproducteur et endocrinien chez un amphipode mâle et femelle du sous-ordre *Gammaridea* (modifié d'après Hyne 2011) 35
Figure I-10 : *Gammarus fossarum* mâle et femelle, schéma d'une coupe transversale de *Gammaridae* et des appareils génitaux en coupe longitudinale (d'après Charniaux-Cotton 1965) 37
Figure I-11 : Stades embryonnaires corrélés aux stades de mues de la femelle *G. fossarum* (Xuereb 2009) 38
Figure I-12 : Changements intertégumentales à l'extrémité des périopodes chez *Gammarus fossarum* au cours du cycle de mue (Geffard *et al.* 2010) 39
Figure I-13 : Micrographies de spermatide (A) et de spermatozoïde (B) de *G. fossarum* en microscopie électronique à balayage utilisé en mode wet-STEM (x 2022) (Lacaze 2011) 40
Figure I-14 : Evolution du nombre de spermatozoïdes et de spermatides contenu dans la gonade en fonction du nombre de jour après la copulation chez *G. fossarum* (Lacaze 2011) 41
Figure I-15 : *Gammarus fossarum* mâle et femelle en précopulât (photo : Quéau H., Irstea) 42
Figure I-16 : Schéma représentant les différentes voies de signalisation hormonales (adapté de Luconi 2006) 43
Figure I-17 : Structure chimique du précurseur de la biosynthèse des ecdysteroids (cholestérol), de l'ecdysone et de la principale forme actives (20-hydroxyecdysone) (adapté de Petryk *et al.* 2003) 45
Figure I-18 : Voies de synthèse de l'hormone juvénile chez les arthropodes (adapté de Hui *et al.* 2010) 46
Figure I-19 : Modèle linéaire représentant la structure de la pré hormone produite par la glande androgène du crabe *Cherax quadricarinatus* (modifé de Ventura *et al.* 2011). 47
Tableau I-5 : Récapitulatif des hormones impliquées dans le processus reproductif chez les amphipodes et effets physiologiques associés 52
Figure I-20 : Représentation schématique de l'hypothèse du coût métabolique (adapté selon Coen *et al.* 1997) 57
Figure I-21 : Du gène à la protéine : synthèse de protéines à partir d'un gène codant chez un eucaryote (adapté d'après Nakazawa *et al.* 2008) 62
Figure I-22 : Chronologie des premiers projets de séquençage à large échelle du génome (modifié d'après Lander *et al.* 2001) 63
Figure I-23 : Principe de l'assemblage *de novo* du transcriptome 65

Figure I-24 : Représentation des propriétés diverses des protéines chez une cellule eucaryote (adapté de Patterson and Aebersold 2003) 66
Figure I-25 : Correction des erreurs d'annotation grâce aux données de spectrométrie de masse (d'après Armengaud 2009) 68
Figure I-26 : Les grandes étapes de l'analyse protéomique 69
Figure I-27 : Stratégie générale pour l'étude du protéome (modifié d'après Armengaud and Bland 2014) 71
Figure I-28 : Principe de l'OFF-Gel 72
Figure I-29 : Schéma du principe de la spectrométrie de masse (Allain 2014) 74
Figure I-30 : Schéma du principe de la spectrométrie de masse MS/MS (Allain 2014) 75
Tableau I-6 : Vue d'ensemble des sources d'ionisation et des analyseurs de masse employés en protéomique (d'après Armengaud 2013) 78
Figure I-31 : Approches pour l'identification des protéines (d'après Armengaud and Bland 2014) 79
Tableau I-7 : Méthodes de quantification employées en analyse protéomique massive 82
Figure I-32 : Exemple de recherche de similarité par BLAST sur une séquence d'ARN hélicase 16 S de *Bacillus licheniformis* 84
Tableau I-8 : Etudes protéomiques en écotoxicologie aquatique réalisées chez les mollusques 85
Tableau I-9 : Etudes protéomiques en écotoxicologie aquatique réalisées chez les crustacés et polychètes. 87

Publication n°1

Figure 1: Overview of biomarker applications using *Gammaru*s species 96
Figure 2: The ten most ferquently occuring proteins in proteomics studies in aquatic ecotoxicology using invertebrates 106

Publication n°2

Figure 1: Number of species presently cataloged in NCBI taxonomy and Gold databases (2013/11/19) 125
Figure 2: Traditional proteomics approaches for non-model organisms 131
Figure 3: Proteogenomics for non-model organisms 137

Figure I-33 : Stratégie générale pour le développement de biomarqueurs de reprotoxicité chez *Gammarus fossarum*.151

CHAPITRE II : Matériels et méthodes

Figure II-1 : Localisation géographique du site de prélèvement de la Tour du Pin 155
Figure II-2 : Prélèvement d'organismes épibenthiques au troubleau suivi par le tri des gammares sur une colonne de tamis (Lacaze 2011) 156
Figure II-3 : Matériel biologique employé pour la création d'un catalogue de protéines par protéogénomique 157
Figure II-4 : Dispositif expérimental en laboratoire 159
Tableau II-1 : Détails des expérimentations réalisées en laboratoire pour l'étude de la spermatogénèse et de l'infertilité masculine 160
Figure II-5 : Substances insecticides testées pour l'étude de l'infertilité masculine 160
Figure II-6 : Déroulement expérimental pour l'étude de l'infertilité masculine 161
Figure II-7 : Représentation schématisée de la procédure utilisée pour déterminer le stade d'intermue des organismes (Xuereb 2009) 162
Figure II-8 : Principe du séquençage Illumina-Solexa (Metzker 2010) 165
Figure II-9 : Stratégies expérimentales mises en œuvre pour l'étude du protéome 166

Figure II-10 : Principales étapes de la digestion enzymatique en solution et en gel 168
Figure II-11 : Principe de l'OFFGel (Agilent Technologies 2009) 170
Figure II-12 : Electrophorèse 1D SDS-PAGE 171
Figure II-13 : Représentation schématique du spectromètre de masse LTQ-Orbitrap 172
Figure II-14 : Traduction d'une séquence en acides nucléiques suivant les six cadres de lecture du code génétique 174
Figure II-15 : Interface de l'algorithme d'interprétation des spectres MS/MS MASCOT 175

CHAPITRE III : Découverte de protéines impliquées dans la reproduction de *Gammarus fossarum*

Publication n°3

Figure 1: Images of *G. fossarum* in amplexus and sampled tissues. 189
Figure 2: General proteogenomic strategy for discovering key proteins from *G. fossarum* 190
Figure 3: Analysis of the 1873 MS/MS identified proteins 192
Figure 4: Proteome dynamics of the male reproductive tissue from *G. fossarum* 194
Figure 5: Comparative analysis of the male and female reproductive tissue 197
Table 1: The most differentially detected proteins in female and male reproductive tissues 198
Figure 6: Clustered temporal abundance variations of proteins during spermatogenesis, before and after copulation. 200

Publication n°4

Figure 1: Female *G. fossarum* with developping embryos within its ventral pouch 217
Figure 2: General strategy for the identification of major egg yolk proteins from *G. fossarum*. 220
Figure 3: Clustered temporal abundance variations of proteins during oogenesis. 222
Figure 4: Clustered temporal abundance variations of proteins during embryogenesis. 223
Figure 5: Identification of major egg yolk proteins precursors in *G. fossarum*. 224
Table 1: List of precursors of major egg yolk protein in *Gammarus fossarum*. 226
Figure 6: Phylogenetic tree showing the relationship in the LLTP family. 227

Note 1

Tableau III-1 : Protéines identifiées dans les *cæca* hépatopancréatiques 241

Note 2

Figure III-1 : Stratégie pour l'amélioration de l'annotation structurale de la protéine d'intérêt par l'intégration des données MS 246

CHAPITRE IV : Impacts des xénobiotiques sur la reproduction masculine

Publication n°5

Figure 1: General strategy for male reproductive disorders assessment in *G. fossarum.* 267
Figure 2: Microscopic observations of *G. fossarum* male reproductive cells (x100). 267
Figure 3: Number of male reproductive cells in *G. fossarum* gonads after 15 days exposure to pollutants..... 269
Figure 4: Analysis of reproductibility between replicates for each conditions in terms of total number of assigned spectra per sample..... 271
Figure 5: Differential expression of the proteins that showed clustered temporal variations in response to xenobiotic exposure..... 272
Figure 6: Examples of clustered temporal variations of proteins in response to Cd, Pyr and Met exposure. 273
Table 1: Proteins with clustered temporal variations among toxic exposure previously identified as sex-specific..... 279

CHAPITRE V : Synthèse et discussion générale

Figure V-1 : Stratégie protéogénomique adoptée pour la découverte de protéines chez *Gammarus fossarum* 295
Figure V-2 : Homologie des protéines identifiées chez *G. fossarum* par l'interrogation de la base de données NCBI (Evalue = 0,01)..... 296
Figure V-3 : Origine des protéines chimériques prédites par le séquençage du transcriptome 297
Figure V-4 : Synthèse et dégradation de l'hormone juvénile chez les insectes..... 300
Tableau V-1 : Première identification de candidats biomarqueurs impliqués dans la reproduction sur la base des prédictions fonctionnelles..... 302
Figure V-5 : Analyse comparative de l'abondance des protéines issues des tissus reproducteurs mâles et femelles..... 304
Figure V-6 : Protéines modulées au cours de l'ovogénèse selon deux profils distincts 305
Figure V-7 : Protéines présentant un profil d'évolution au cours de l'ovogénèse et de l'embryogénèse cohérant avec un rôle de réserve du développement embryonnaire..... 306
Tableau V-2 : Protéines identifiées comme vitellogénines chez *G. fossarum* 307
Tableau V-3 : Protéines modulées au cours de la spermatogénèse..... 308
Figure V-8 : Exemples de profils d'évolution des protéines au cours de la spermatogénèse 309
Figure V-9 : Nombre de spermatozoïdes dans les gonades après 15 jours d'exposition aux polluants..... 311
Figure V-10 : Exemples de profil d'évolution de l'abondance des protéines chez les gammares mâles exposés aux polluants 314

CONCLUSION & PERSPECTIVES

Figure i : Stratégie pour le développement de nouveaux biomarqueurs chez *Gammarus fossarum* par des approches protéomiques..... 319

LISTE DES PUBLICATIONS ET DES COMMUNICATIONS

PUBLICATIONS

Publication n°1

Next-generation proteomics: towards customized biomarkers for environmental biomonitoring. (2014) ***Environmental Science and Technology*** 48 (23) : 13560-13572. Trapp J., Armengaud J., Chaumot A., Salvador A., Geffard O.

Publication n°2

Non-model organisms, a species endangered by proteogenomics. (2014) ***Journal of Proteomics*** 105 (0) : 5-18. Armengaud J., Trapp J., Pible O., Geffard O., Chaumot AI, Hartmann EM.

Publication n°3

Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods**.** (2014) ***Molecular and Cellular Proteomics*** 13 (12) : 3612-3625. Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J.

Publication n°4

High-throughput proteome dynamics for discovery of key proteins in non-model species: unsuspected vitellogenins diversity in the crustacean *Gammarus fossarum*! Soumise à ***Environmental Science and Technology***. Trapp J., Armengaud J., Imbert G., Gaillard J-C., Chaumot A., Geffard O.

Publication n°5

Proteomic investigation of male *Gammarus fossarum*, a freshwater crustacean, in response to endocrine disruptors. (2015) ***Journal of Proteome Research*** 14 (1) : 292-303. Trapp J., Armengaud J., Habtoul Y., Imbert G., Gaillard J-C., Abbaci K., Chaumot A., Geffard O.

COMMUNICATIONS ORALES

- **Colloques nationaux**

Recherche protéogénomique de biomarqueurs chez la crevette d'eau douce *Gammarus fossarum*, sentinelle clef en écotoxicologie**.** Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. ***Séminaire de Toxicologie nucléaire environnementale & humaine, Paris, Septembre 2013***

Le gammare, sentinelle de la contamination du milieu aquatique**.** Trapp J., Armengaud J., Chaumot A., Geffard O. ***Journée scientifique de Marcoule, CEA de Marcoule, 10-12 Juin 2014.***

Proteogenomics definition of reproductive disorder biomarkers from the freshwater crustacea *Gammarus fossarum*, a sentinel organism in ecotoxicology. Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. ***Congrès de Spectrométrie de Masse et d'Analyse Protéomique 2014, Lyon, 30 Juin-2 Juillet 2014.***

- **Colloques internationaux**

Proteogenomics, a straightforward approach for protein discovery in unsequenced test organisms in ecotoxicology. Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. ***23nd Society of Environmental Toxicology and Chemistry (SETAC) Europe Annual meeting, 12-16 April 2013, Scotland-Glasgow***

Documenting the unique biology of gammarids by means of high-throughput proteogenomics**.** Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. Plenary lecture, ***15t^{h} International Colloquium on Amphipoda, 2-7 September 2013, Poland-Szczawnica.***

Deciphering the major egg yolk proteins from the amphipod *Gammarus fossarum* by proteogenomics: key lessons regarding prediction of the functions of ecotoxicology-relevant proteins. Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. ***24nd Society of Environmental Toxicology and Chemistry (SETAC) Europe Annual meeting, Switzerland-Basel. 11-15 May 2014.***

COMMUNICATIONS AFFICHEES

- **Colloques nationaux**

Unravel the reproductive biology of *Gammarus fossarum* by comparative proteomics. Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. ***Colloque d'écophysiologie animale (CEPA), 7-9 Novembre 2013. Lyon.***

Développement d'outils pour le suivi de la perturbation endocrine chez le gammare mâle. Trapp J., Habtoul Y., Abbaci K., Chaumot A., Armengaud J., Geffard O. ***Société d'Ecotoxicologie Fondamentale et Appliquée (SEFA). 1-2 Juillet 2014. Besançon.***

- **Colloques internationaux**

Comparative proteomics of the freshwater amphipod *Gammarus fossarum*: focus on endocrine regulation. Trapp J., Geffard O., Imbert G., Gaillard J-C., Davin A-H., Chaumot A., Armengaud J. ***23nd Society of Environmental Toxicology and Chemistry (SETAC) Europe Annual meeting, 12-16 April 2013, Scotland-Glasgow***

INTRODUCTION

La prise de conscience des problèmes liés à la dispersion des polluants dans l'environnement s'est amorcée dans les années 1960, avec la parution de l'ouvrage « the Silent Spring » par Rachel Carson qui exposait les problèmes environnementaux liés à l'épandage d'un pesticide organochloré, le DDT (dichlodiphényltrichloroéthane). L'ouvrage décrit avec précision l'entrée du DDT dans la chaîne alimentaire, son accumulation dans les tissus adipeux des animaux et l'apparition des dommages génétiques et carcinogènes associés. Considérée comme une des œuvres littéraires majeures du 20ème siècle, the « Silent Spring » a contribué à l'initiation du mouvement écologiste et la formation de groupes activistes militant pour l'interdiction et la limitation de l'usage des pesticides (Gore 2009). En parallèle, le lien entre pollution industrielle et santé humaine est définitivement établi lorsque les troubles neurologiques dont souffrait la population de la baie de Minamata au Japon trouvent leur explication dans la consommation de produits de la mer contaminés au mercure, mercure provenant du déversement d'effluents industriels (Takizawa 2000).

Face à la pression sociétale, les gouvernements se sont dotés d'outils législatifs plus contraignants, dans le but de limiter et de surveiller la contamination environnementale. Les écosystèmes aquatiques dulçaquicoles font l'objet d'une attention particulière auprès des législateurs. En effet, la ressource en eau est liée à de nombreux enjeux économiques, sanitaires, écologiques et même politiques. Les milieux aquatiques sont des écosystèmes particulièrement vulnérables, étant le réceptacle ultime de la majorité des contaminants. Les voies d'entrées sont diverses, de manière directe par les rejets domestiques, industriels ou hospitaliers et de manière indirecte par les retombées atmosphériques ou le lessivage des sols agricoles, urbains ou industriels. En conséquence, la contamination y est multiple et complexe : métaux lourds, hydrocarbures, pesticides, rejets pharmaceutiques ou autres substances émergentes (Holt 2000). Aussi, 60 % des réserves d'eaux douces mondiales sont partagées entre au moins deux pays, donnant le jour à une multitude de conventions régissant leurs partages ou leurs gestions (Transboundary Freshwater Dispute Database 2003).

Au niveau européen, une approche communautaire pour la gestion des cours d'eau est essentielle, plusieurs fleuves comme le Rhin, la Meuse ou le Danube traversent différents pays membres. En 2000, la directive cadre sur l'eau (DCE) n°2000/60/CE est adoptée afin d'appliquer une législation unique sur l'ensemble des pays membres. Elle définit un cadre pour la gestion et la protection des eaux par grand bassin hydrographique et fixe des objectifs de préservation et de restauration des masses d'eaux, articulées autour de la notion de bon état chimique et écologique des cours d'eau. Ce bon état est déterminé par le respect des normes de qualité environnementale par le biais de valeurs seuils de substances chimiques dites « prioritaires » et également sur la base d'indices biotiques, construits sur la présence et l'abondance d'organismes de la faune et de la flore aquatique (diatomées, macrophytes, macro invertébrés). Toutefois, ces deux approches ne permettent pas de couvrir un continuum de causes à effets, garant d'une analyse robuste du risque environnemental ainsi que de la mise en place de programmes d'action et de restauration efficaces. L'approche chimique comporte des limites telles que l'impossibilité de pouvoir quantifier l'ensemble des polluants présents ainsi que leurs produits de dégradation associés et de leurs effets conjugués (effets additifs, antagonistes, synergiques). De même, elle ne permet pas d'évaluer l'impact des contaminants sur les organismes autochtones, notamment en raison de la difficulté à estimer leur fraction biodisponible. De façon complémentaire, l'analyse de l'état écologique permet d'apporter une appréciation globale de l'état de santé d'un écosystème. Néanmoins, elle ne permet pas de signaler de manière précoce une dégradation de la qualité des cours d'eau, ni d'identifier le type de pression anthropiques, pouvant être d'ordre chimique, physique (*e.g*. modification de l'habitat) ou biologique (*e.g*. compétition, prédation) (Depledge and Fossi 1994, Attrill and Depledge 1997, Clements 2000). Or, l'identification de ces pressions est nécessaire pour mettre en place une stratégie efficace de gestion et de restauration des milieux aquatiques.

C'est pourquoi, dans le cas de l'évaluation des pressions liées à l'altération chimique des masses d'eau, il est nécessaire de disposer d'outils précoces mais également capables de faire le lien entre la présence de xénobiotiques et l'apparition d'effets toxiques associés. L'analyse du risque écotoxicologique repose sur le suivi de biomarqueurs, se définissant comme tous paramètres biochimiques, moléculaires, cellulaires, histologiques ou physiologiques qui peuvent être observés ou mesurés dans les cellules, tissus, fluides corporels ou dans l'organisme, mettant en évidence l'exposition ou les effets d'un ou plusieurs polluants chimiques (Depledge and Fossi 1994, Lagadic *et al.* 2008). Ces derniers couvrent donc une large gamme de réponses : dommages à l'ADN,

induction des systèmes de détoxification, altérations des défenses immunitaires ou de la fonction reproductrice (Galloway *et al.* 2004, Depledge and Galloway 2005, Lagadic *et al.* 2008).

En raison de son importance capitale dans le maintien de l'espèce, l'évaluation de l'intégrité de la fonction reproductrice fait l'objet d'une attention soutenue auprès de la communauté scientifique. Cette fonction peut être altérée par une large gamme de xénobiotiques présents dans l'environnement et de manière spécifique, par des substances dites « perturbateurs endocriniens » (PE). Un PE se définit comme « *un agent exogène capable d'interférer avec la synthèse, la sécrétion, le transport, la liaison, l'action ou l'élimination des hormones naturelles qui sont responsables de la maintenance, de l'homéostasie, de la reproduction, du développement, du comportement* » (European Commission 1999). Historiquement, la prise de conscience des phénomènes de reprotoxicité liés à l'exposition à des xénobiotiques a débuté avec l'ouvrage de Rachel Carlson, cette dernière associant directement l'épandage locale de DDT à une diminution massive des populations d'oiseaux. Pour les milieux aquatiques, un des cas les plus connus est celui des alligators du lac Apopka en Floride où une diminution importante des populations locales, associée à un accroissement des anomalies de l'appareil reproducteur masculin (micro-pénis, anomalies testiculaires) ont été documentés dans les années 1990 suite à l'exposition à un produit de dégradation du DDT, le dichlorodiphényldichlororéthylène (Semenza *et al.* 1997). Parmi les autres exemples d'altérations des caractères sexuels primaires, on peut citer la masculinisation de gastéropodes femelles (imposex) induite par la présence de tributylétain (Sousa *et al.* 2014) ou encore la féminisation des poissons mâles induite par la présence de substances oestrogéniques (Jobling *et al.* 1998).

En raison de la grande connaissance de la physiologie et de l'endocrinologie des poissons, c'est chez ces derniers que les causes et les effets de la perturbation endocrine sont les plus documentés en écotoxicologie aquatique. Au fil des années, l'étude de la perturbation endocrine chez les poissons a fait l'objet de nombreux projets de recherche et de publications, documentant des altérations au niveau des caractères sexuels primaires et secondaires, de la sécrétion des hormones sexuelles stéroïdiennes ou encore de l'induction de la vitellogénine chez le mâle (Matthiessen 2003). Ces travaux ont permis le développement de biomarqueurs robustes pouvant être directement utilisés *in-situ* (biomonitoring). L'un des biomarqueurs les plus populaires est le suivi de l'induction de la vitellogénine (Vtg) chez le mâle, dont la mesure a été normalisée auprès de l'organisation internationale de standardisation (ISO 23893-3:2013). Chez les organismes ovipares, cette protéine

est utilisée comme précurseur dans la production du vitellus, qui fournit les réserves énergétiques nécessaires aux embryons pour assurer leur développement. Chez les vertébrés, elle est progressivement accumulée dans les oocytes au cours de l'ovogénèse, sous le contrôle hormonal de l'œstrogène. Bien que le gène encodant la Vtg soit présent chez le mâle, son expression reste faible en conditions physiologiques normales, mais peut être induit *via* la présence d'œstrogènes, conduisant à sa proposition comme biomarqueur d'exposition à des substances oestrogènomimétiques (Kime *et al.* 1999).

En revanche, pour les invertébrés, bien que ces derniers représentent 95 % des espèces animales et par conséquent jouent un rôle essentiel dans le fonctionnement des écosystèmes, il n'existe pas d'outils robustes pour évaluer l'effet des PE sur leur reproduction. Cette absence est la conséquence directe de la grande méconnaissance des hormones et protéines impliquées dans la régulation de cette fonction chez ces organismes, et notamment chez les crustacés (LeBlanc 2007). De plus, les molécules impliquées dans la fonction reproductrice se sont fortement diversifiées au cours de l'évolution, conduisant à des systèmes endocriniens très divers et différents de ceux connus ou caractérisés chez les vertébrés. Alors que la caractérisation de ces biomolécules faisait appel à des techniques coûteuses et difficiles à mettre en place, dont seul un nombre restreint d'organismes *i.e*. espèces modèles, pouvaient bénéficier, aujourd'hui et grâce aux récentes avancées technologiques dans le domaine du séquençage de l'information génétique et protéique, la découverte rapide et à large échelle de nouvelles protéines chez des espèces non-modèles est à présent possible.

Dans ce contexte, les travaux présentés dans ce manuscrit sont centrés sur la découverte chez une espèce d'intérêt en écotoxicologie, le crustacé d'eau douce *Gammarus fossarum*, de protéines impliquées dans sa reproduction. L'objectif est de mieux comprendre la régulation de la fonction reproductrice ainsi que les mécanismes de la perturbation endocrinienne chez les crustacés, permettant *in fine* de proposer de nouveaux biomarqueurs applicables dans le cadre de réseaux de surveillance de la qualité des eaux courantes. Ce document de thèse se structure en cinq chapitres. Le **chapitre I** est consacré à une synthèse bibliographique, centrée sur les démarches et outils pour l'évaluation écotoxicologique des milieux aquatiques, avec un intérêt particulier pour l'évaluation de la perturbation de la reproduction chez les crustacés. Ces travaux de recherche bibliographique ont notamment été valorisés sous la forme de deux articles. Ce chapitre se conclu par la présentation des objectifs de thèse. Le **chapitre II** détaille les méthodes

expérimentales mises en œuvre au cours de ce travail, notamment en termes de techniques analytiques et d'expérimentations en laboratoire**.** Les résultats obtenus sont présentées dans les **chapitre III** et **IV** sous forme de publications et de notes pour complément d'information. Le **chapitre V** est une synthèse générale des principaux résultats obtenus au cours de ces travaux. Enfin, le manuscrit se termine par une conclusion ouvrant sur les perspectives envisagées.

CHAPITRE I : Démarches et outils pour l'évaluation écotoxicologique des milieux aquatiques – effets sur la reproduction

Le premier chapitre de ce travail de thèse est une synthèse bibliographique des connaissances actuelles relatives à l'utilisation des biomarqueurs pour l'évaluation de la qualité biologique du milieu aquatique, l'accent étant mis plus particulièrement sur les marqueurs de reprotoxicité chez les crustacés.

La première partie (I-1) a pour objectif de définir la notion de biomarqueur, de présenter l'utilité de ces outils dans l'évaluation de la qualité biologique du milieu aquatique ainsi que de faire le bilan de leur utilisation dans un contexte de biosurveillance, en s'appuyant sur des exemples concrets avec les crustacés du genre *Gammarus*, modèles biologiques populaires en écotoxicologie aquatique.

La seconde partie (I-2) est un état de l'art concernant la reproduction et l'endocrinologie des crustacés, les sources et les effets de la perturbation de la fonction reproductrice ainsi que les outils permettant de l'évaluer.

La partie I-3 s'intéresse à l'apport en écotoxicologie, des nouveaux outils omiques et plus particulièrement de la protéomique, en présentant les différents outils employés.

La partie I-4 présente la valorisation de ces travaux de recherche bibliographique sous la forme de deux articles (**Publications 1** et **2**).

Enfin la dernière partie (I-5) expose les objectifs de ce travail de thèse ainsi que la stratégie expérimentale mise en œuvre.

1. Les biomarqueurs pour l'évaluation de la qualité biologique des hydrosystèmes

1.1. Approches pour la surveillance de la qualité environnementale

Face à la dégradation des milieux aquatiques et aux effets néfastes associés, il est nécessaire de mener une évaluation du risque environnemental afin de pouvoir [1] minimiser les menaces sur la santé humaine ainsi que sur les ressources alimentaires et en eau [2] assurer l'intégrité du fonctionnement des écosystèmes et [3] assurer qu'une espèce à haute valeur écologique ou commerciale ne décline pas (Depledge and Fossi 1994).

Dans le cas d'une altération de la qualité chimique des masses d'eau, l'évaluation du risque repose sur une double approche de diagnostic de l'état chimique et biologique. Les analyses chimiques peuvent être réalisées dans différents compartiments de l'écosystème : colonne d'eau, sédiments, particules en suspension. De même, l'évaluation de l'état biologique peut se faire à différents niveaux d'organisation, allant de l'individu à l'écosystème (Figure I-1). L'analyse de la qualité biologique du milieu repose sur l'utilisation du biote pour renseigner la présence et l'incidence des polluants dans l'écosystème, on parle alors de biomonitoring ou biosurveillance.

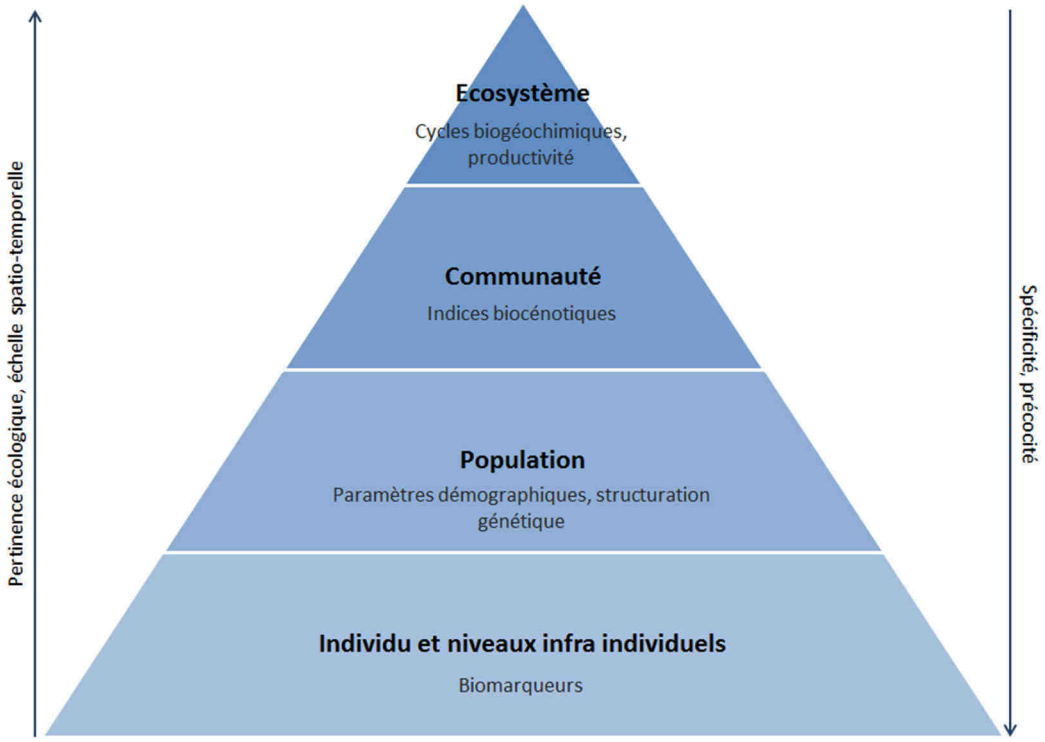


Figure I-1 : Représentation des différents niveaux d'organisation biologique et exemples de paramètres mesurables pour l'évaluation de la qualité environnementale (Clements 2000, Lagadic *et al.* 2008)

L'analyse de l'état chimique et écologique, actuellement employée dans le cadre de la DCE ne permet pas de couvrir un continuum de causes à effets, garant d'une analyse robuste du risque environnemental et de la mise en place de programmes d'action et de restauration efficaces. Il est donc nécessaire de proposer une approche complémentaire permettant de pouvoir lier la présence d'un contaminant avec l'apparition d'effets délétères associés. L'évaluation du risque en écotoxicologie repose sur la double approche chimique et biologique (Figure I-2), *via* la caractérisation de la contamination des organismes et l'apparition d'effets toxiques associés.

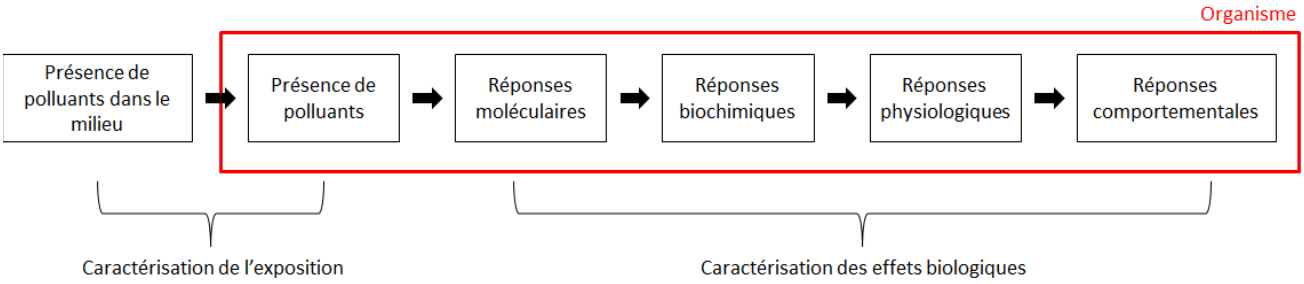


Figure I-2 : Démarches dans l'évaluation du risque écotoxicologique et représentation séquentielle de réponses induites au sein d'un individu en réponse à un stress toxique (d'après Lagadic *et al.* 2008)

Face aux matrices physiques, la quantification des polluants au sein du biote est une mesure intégrative permettant d'évaluer la fraction biodisponible, plus pertinente dans la compréhension des effets toxiques (Besse *et al.* 2013). Les effets toxiques se traduisent notamment par des troubles de l'homéostasie, l'activation des mécanismes de défense ou encore l'apparition de dommages au niveau moléculaire (ADN) ou tissulaire. Le suivi de l'altération de ces variables biologiques chez une espèce qualifiée de « **sentinelle** », est alors employé comme outil indicateur de la toxicité du milieu, *i.e.* les biomarqueurs.

1.2. Le rôle des biomarqueurs dans l'évaluation écotoxicologique

1.2.1. Classification

Un biomarqueur peut être distingué selon deux catégories : défense et dommage. La première catégorie indique la mise en place dans l'organisme, de mécanismes de régulation du métabolisme en réponse à la contamination et permet de confirmer l'exposition des organismes à une catégorie de xénobiotiques précis, liant l'exposition dans le milieu à l'internalisation de la substance

(McCarthy & Munkittrick, 1996). A l'opposé, les biomarqueurs de dommage constituent des indicateurs d'un dépassement des capacités de régulation de l'organisme, avec l'apparition d'effets délétères aux différents niveaux d'organisation biologique (Lagadic 2002). Cependant, dans la littérature, cette subdivision entre biomarqueurs d'exposition et d'effet est poreuse et fluctue selon les différents auteurs (van der Oost *et al.* 2003). Par exemple, la mesure de l'inhibition de l'acétylcholinestérase (AChE) est un des biomarqueurs les plus couramment employés en écotoxicologie. Cette enzyme est responsable de la dégradation d'un neurotransmetteur, l'acétylcholine. En présence de pesticides de type organophosphorés et carbamates, l'AChE est inhibé, provoquant des phénomènes de neurotoxicité par une surexcitation du système nerveux (Roméo and Giambérini 2012). Ce biomarqueur peut être employé comme indicateur spécifique d'une exposition à un groupe de xénobiotiques mais également comme indicateur de manifestation d'effets délétères sur l'organisme. C'est pourquoi, afin d'illustrer les différents types de biomarqueurs employés en écotoxicologie, leur présentation se fera suivant l'ordre séquentiel des réponses au sein de l'individu comme présenté en Figure I-2, par grande catégorie de réponses.

1.2.2. Exemples de biomarqueurs couramment employés en écotoxicologie

1.2.2.1. Détoxification des xénobiotiques

Dans un premier temps, les molécules internalisées au sein de la cellule peuvent en être excrétées par des systèmes d'efflux membranaire tel que le **Mécanisme de Défense MultiXénobiotique** dont le suivi de l'activité peut être employé comme biomarqueur (Pain *et al.* 2005). Concernant la métabolisation des xénobiotiques, les enzymes de la phase I permettent la « fonctionnalisation chimique » de ces derniers par leur hydrolyse, oxydation ou réduction afin d'augmenter leur solubilité. Les enzymes des **cytochromes P450** (CYP450) sont les principaux agents de cette réaction et leurs activités peuvent être mesurées par le suivi de la dégradation d'un substrat, l'éthoxyrésorufine (James and Boyle 1998). Par la suite, les enzymes de la phase II assurent la conjugaison du xénobiotique avec de petites molécules endogènes polaires, facilitant ainsi l'excrétion du complexe. Parmi ces dernières, le biomarqueur le plus populaire est celui mesurant l'activité de la **glutathion S-transférase** (GST) (McLoughlin *et al.* 2000, Maltby and Hills 2008, Karaouzas *et al.* 2011).

1.2.2.2. Défense antioxydative

Une fois internalisée dans l'organisme, une large gamme de xénobiotiques (HAP, pesticides ou métaux) peut conduire à la production d'espèces réactives à l'oxygène, provoquant un stress oxydatif. Ce stress oxydatif engendre des effets néfastes par la détérioration des lipides membranaires, l'oxydation des protéines et/ou des dommages à l'ADN. En absence de réparation efficace, ces dommages peuvent alors altérer le fonctionnement cellulaire ou tissulaire. La défense face au stress oxydatif est réalisée par un ensemble de molécules permettant d'empêcher et de contrôler la formation des radicaux libres ou d'inhiber leur réactivité avec les structures biologiques. En écotoxicologie, le suivi de l'activité des systèmes enzymatiques impliqués dans ces processus est couramment employé : **catalase (CAT)**, **peroxydase (POD)** ou **superoxyde dismutase (SOD)** (Sroda and Cossu-Leguille 2011).

1.2.2.3. Homéostasie

Afin d'assurer le bon fonctionnement physiologique, l'organisme met en place une série de processus afin de maintenir les différentes constantes de son milieu intérieur (*i.e*. homéostasie). La maintenance des paramètres physico-chimiques de l'organisme porte notamment sur le pH, l'osmolarité ou la composition ionique des fluides biologiques. Chez les organismes d'eau douce, les fluides corporels sont hyperosmotiques face à l'environnement aqueux. Les ions sont prélevés par l'alimentation et dans le milieu par les branchies. Or, de par leur rôle de surface d'échange, les branchies sont parmi les premiers organes impactés par un stress chimique et ne peuvent plus assurer leur rôle crucial dans l'iono-osmorégulation de l'organisme. L'activité branchiale du transporteur ionique **Na^+/K^+ATPase**, responsable du passage du sodium à l'hémolymphe est un biomarqueur couramment employé pour évaluer l'intégrité de cette fonction chez des organismes soumis à un stress (Felten *et al.* 2008). Par ailleurs, la régulation homéostasique des métaux essentiels est réalisée par des protéines non enzymatiques, les **métallothionéines**. Leurs inductions permettent de mettre en évidence une exposition de l'organisme à des métaux (Amiard *et al.* 2006).

1.2.2.4. Dommages aux biomolécules

On retrouve dans cette catégorie les indicateurs de génotoxicité (*e.g.* adduits à l'ADN, cassures de l'ADN par le **test des micronoyaux** ou des **comètes**) (Lacaze *et al.* 2010), d'atteintes aux lipides (peroxydation des lipides membranaires par la **production de malondialdéhyde**) (Sroda and Cossu-Leguille 2011) ou de protéotoxicité par l'induction des **Heat Shock Protein** (HSP), des protéines chaperonnes assurant la protection et la dégradation des protéines dénaturées (Dhainaut *et al.* 1997, Shatilina *et al.* 2010).

1.2.2.5. Physiologie de l'organisme

Cette dernière catégorie de biomarqueurs s'intéresse à l'évaluation de l'intégrité des grandes fonctions de l'organisme, directement liées à la survie et la reproduction de l'organisme. L'acquisition de l'énergie par l'alimentation peut être mesurée par la mesure de l'activité des **enzymes digestives,** l'évolution des **réserves énergétiques (lipides, sucres, protéines)** ou du **taux d'alimentation** (Dedourge-Geffard *et al.* 2012). La locomotion est un paramètre intégrateur car il est lié à l'échappement à la prédation, à la recherche de ressources alimentaires ou de partenaires sexuels et peut être évalué par le **suivi de l'activité locomotrice** (Felten *et al.* 2008). Le bon fonctionnement du système immunitaire est capital pour se défendre de l'invasion d'agents extérieurs pathogènes mais également pour assurer l'élimination des cellules endogènes malignes. Or, un stress toxique peut notamment provoquer des phénomènes d'immunodépression. L'immunocompétence peut être évaluée par le suivi de l'activité de **phagocytose**, des **lysozymes** ou le **dénombrement des cellules du système immunitaire** (*e.g.* hémocytes) (Brousseau *et al.* 2012). L'intégrité de la fonction reproductrice peut être mesurée par une large gamme de biomarqueurs moléculaires (**induction de la Vtg**) (Jubeaux *et al.* 2012), cellulaires (**production de gamètes**), tissulaires (**histopathologie des gonades**) ou individuels (**production de juvéniles**) (Schirling *et al.* 2006, Geffard *et al.* 2010). Alors que les réponses au niveau physiologique sont généralement peu spécifiques d'un type de stress toxique, le système immunitaire et reproducteur peut être également ciblé de manière spécifique par certains xénobiotiques présents dans l'environnement, tels que les perturbateurs endocriniens.

1.2.3. Intérêt et pertinence des biomarqueurs

Dans un contexte d'évaluation de la qualité environnementale, les biomarqueurs doivent permettre de répondre à deux questions :

- Les milieux sont-ils contaminés ?
- Cette contamination provoque-t-elle des effets néfastes sur les niveaux d'organisation biologique supérieurs ?

En réponse, Lagadic (2002) propose une qualification des biomarqueurs selon leur visée **diagnostique** et/ou **prédictive**. Certains biomarqueurs tels que l'inhibition de l'AChE ou l'induction des métallothionéines répondent à une gamme restreinte de substances chimiques. Bien qu'ils ne renseignent pas sur la nature spécifique des molécules, leur emploi permet de réduire le champ d'investigation à un nombre plus restreint de substances, facilitant l'analyse chimique des différentes matrices environnementales. Cependant, la majorité des biomarqueurs témoignant de l'exposition à des xénobiotiques possède un faible pouvoir prédictif, car des mécanismes de compensation peuvent se mettre en place, prévenant l'apparition d'effets biologiques néfastes. Toutefois, une caractérisation fine de la relation dose-réponse peut permettre d'améliorer le pouvoir prédictif de ce type de réponse, afin de corréler le niveau d'expression du biomarqueur à l'état de santé de l'individu (Figure I-3).

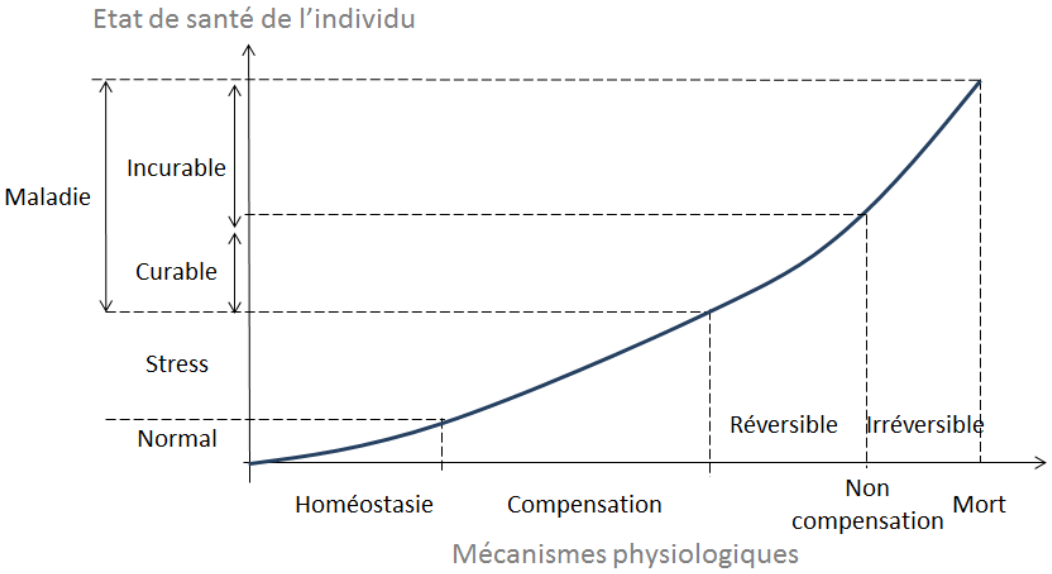


Figure I-3 : Relation entre le niveau d'expression du biomarqueur et l'état de santé de l'individu face à un stress toxique d'intensité croissante (adapté de Depledge and Fossi 1994)

Par exemple, chez l'amphipode d'eau douce *Gammarus fossarum* exposé à deux pesticides à différentes concentrations, une corrélation entre l'inhibition de l'AChE et l'apparition de troubles comportementaux *i.e*. alimentation et locomotion a été mis en évidence, permettant de modéliser la relation entre ces différents paramètres (Xuereb *et al.* 2009). Aussi, pour le même organisme, la relation entre le taux de dommages à l'ADN sur les spermatozoïdes et les anomalies du développement embryonnaire a pu être modélisée (Lacaze *et al.* 2011).

Enfin, pour une évaluation intégrative de la qualité environnementale, une approche dite « **multi-biomarqueurs** » doit être envisagée afin de pouvoir couvrir une large de gamme de réponses biologiques, alliant les critères de diagnostic, de précocité et de prédictivité.

1.3. Bilan de l'utilisation des biomarqueurs en biosurveillance environnementale : exemple avec les organisme sentinelles du genre *Gammarus*

Parmi les indices biocénotiques disponibles au sein de la DCE, l'indice « macroinvertébrés » est un des plus utilisés en raison de la forte représentativité écologique de ces organismes. En effet, les invertébrés représentent 95 % de la biodiversité animale (Depledge and Galloway 2005). Une synthèse de cette diversité est présentée en Figure I-4, s'appuyant sur une analyse comparative de séquences génomiques d'après Halanych (2004).

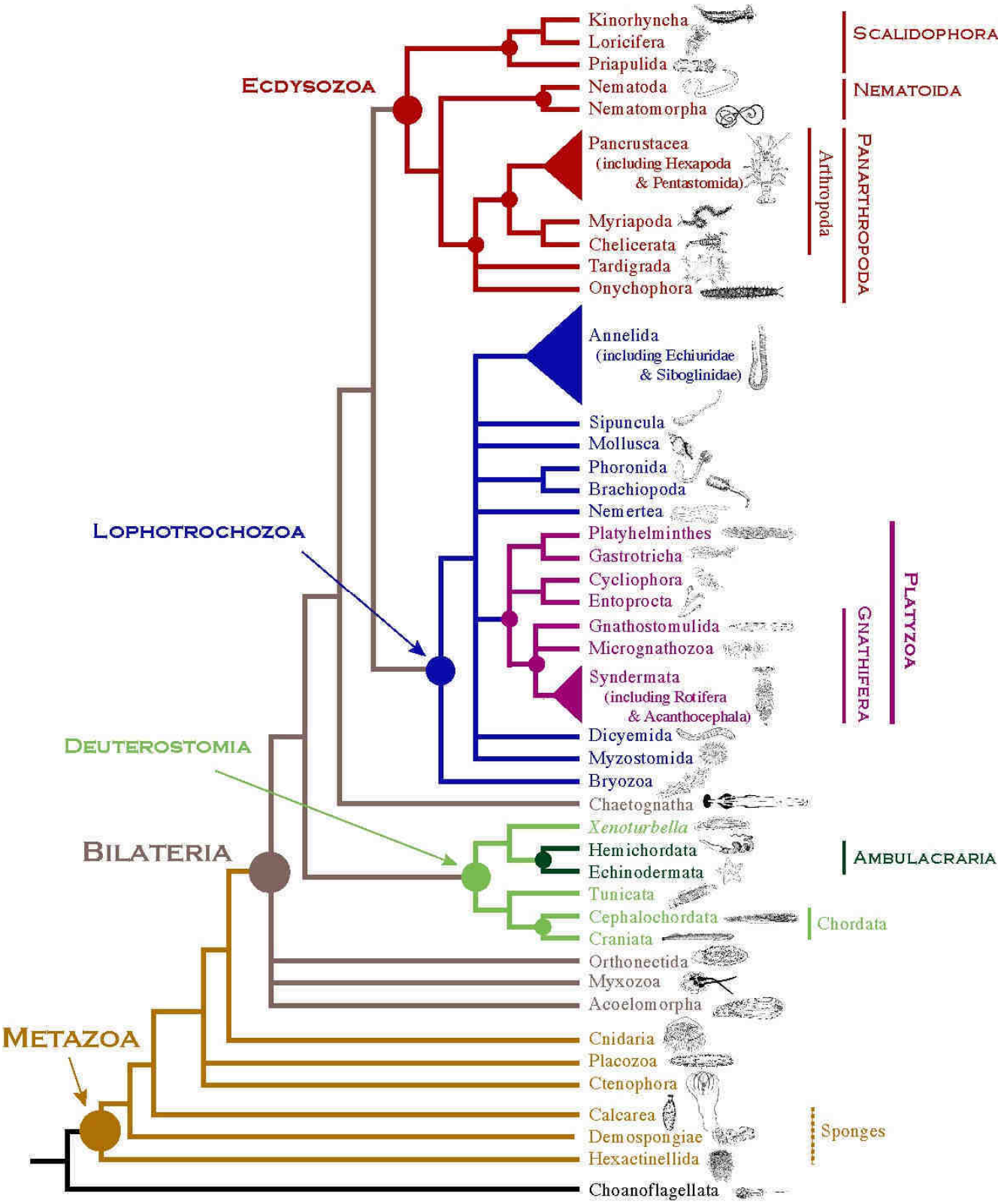


Figure I-4 : Vue de la phylogénie animale chez les invertébrés basée sur des données moléculaires (Halanych 2004)

Les invertébrés constituent un groupe paraphylétique qui diverge très tôt au cours de l'évolution des métazoaires (eucaryotes pluricellulaires hétérotrophes), il y a environ 800 millions d'années. Ils se distribuent selon plusieurs embranchements : [1] les métazoaires non-bilatériens avec les cnidaires (*e.g.* méduses, coraux et anémones) et les embranchements d'éponges et [2] les métazoaires bilatériens : le clade des lophotrochozoaires (*e.g*. mollusques et annélides) et des ecdysozoaires (*e.g.* arthropodes et nématodes) et la lignée des deutérostomiens (e.g. chordés et échinodermes). Suite aux longs processus évolutifs et adaptatifs opérés sur plusieurs millions d'années leur permettant de coloniser de nombreuses niches écologiques, les invertébrées présentent une grande diversité au niveau morpho-anatomique, biologique, physiologique et moléculaire.

Dans le cadre de l'évaluation de la qualité du milieu, les macroinvertébrés (taille >0,5 cm) sont des organismes de choix, ces derniers occupant une position clé dans la chaine alimentaire mais également pour des raisons plus pragmatiques (*i.e.* leur taille). De plus, d'un point de vue éthique, leur utilisation en substitut à des organismes vertébrés, soulève moins d'inquiétudes vis-à-vis de la sphère publique (Lagadic and Caquet 1998, Galloway *et al.* 2004, Depledge and Galloway 2005). Dans l'évaluation écotoxicologique des masses d'eau continentales, le choix de l'espèce sentinelle fluctue selon le contexte géographique (choix d'espèce native/autochtone) et également selon le compartiment étudié : le chironome (phase de vie larvaire de l'insecte), les annélides ou la Corbicule (bivalve) pour le compartiment sédimentaire, la moule zébrée (bivalve) ou la limnée (gastéropode) pour la colonne d'eau ou les crustacés amphipodes du genre *Gammarus* pour la phase épibenthique.

1.3.1. Utilisation de l'organisme sentinelle *Gammarus fossarum*

1.3.1.1. Classification et morphologie

Au sein des ecdysozoaires, l'embranchement des arthropodes est le plus représenté en termes de nombre d'espèce. On y retrouve le sous-embranchement des pancrustacés incluant les crustacés et les insectes, apparu entre -400 et -500 millions d'années qui présente à l'heure actuelle 66 000 espèces de crustacés. La Figure I-5 est une classification récente des pancrustacés selon une analyse phylogénétique de von Reumont *et al.* (2012).

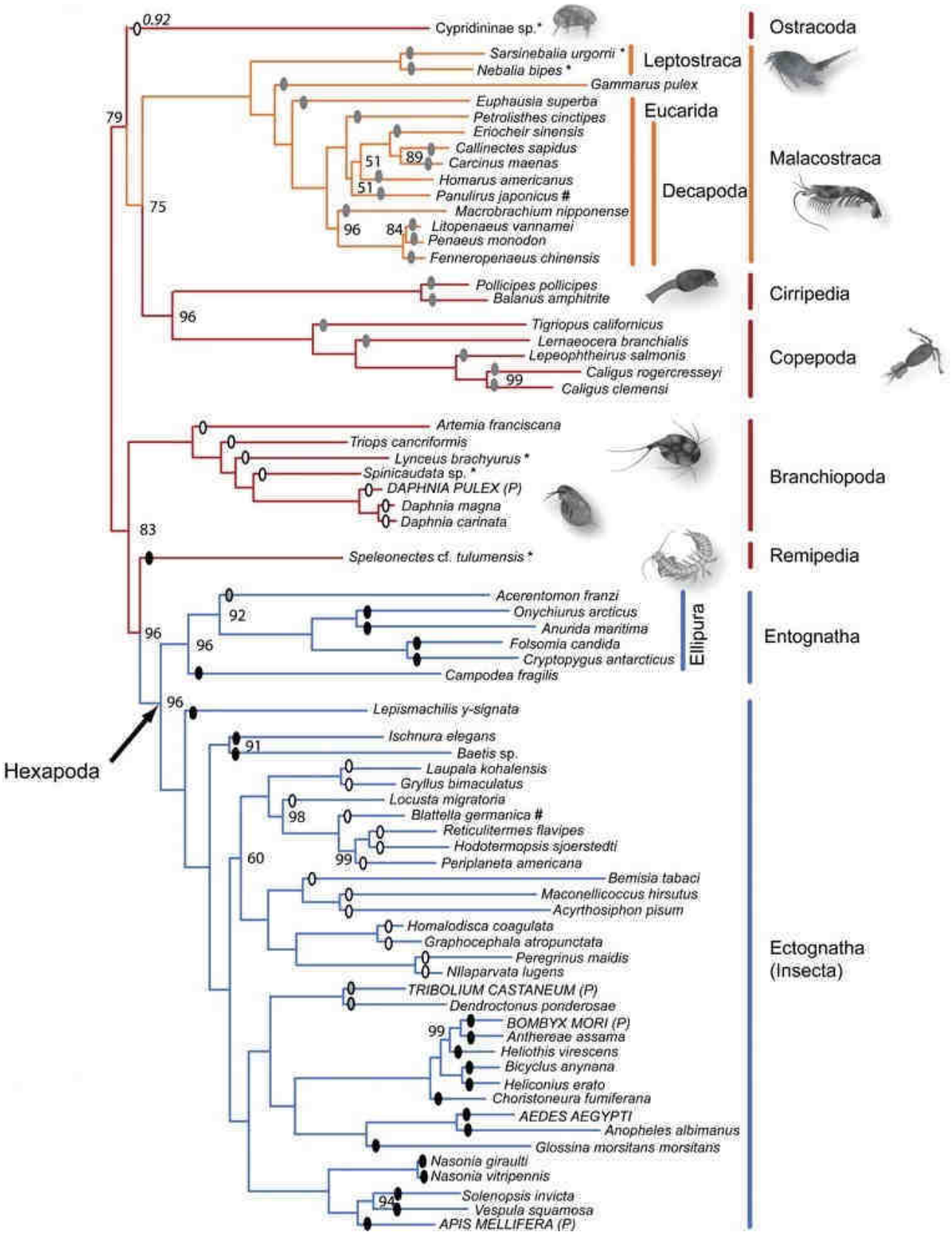


Figure I-5 : Vue d'ensemble de la phylogénie des pancrustacés d'après des données moléculaires (d'après von Reumont *et al.* 2012)

Bien que l'hypothèse selon laquelle les pancrustacés forment un taxon monophylétique semble acceptée au sein de la communauté scientifique, les relations entre les différents groupes en son sein sont toujours controversées. La systématique de notre modèle d'étude *Gammarus fossarum* est présentée en Tableau I-1.

Tableau I-1 : Classification de *Gammarus fossarum* (Martin and Davis 2001)

Embranchement	Arthropode
Classe	Malacostracé
Sous-classe	Eumalacostracé
Ordre	Amphipode
Sous-ordre	*Gammaridae*
Genre	*Gammarus*
espèce	*fossarum* (Koch, 1836)

Les *Gammaridae* sont les espèces les plus représentées et les plus largement répandues parmi les amphipodes, occupant une large gamme d'habitats marin ou dulçaquicole. Le genre *Gammarus*, réparti dans tout l'hémisphère nord, comprend plus d'une centaine d'espèces dont la majorité sont dulçaquicoles (résumé par Kunz *et al.* 2010).

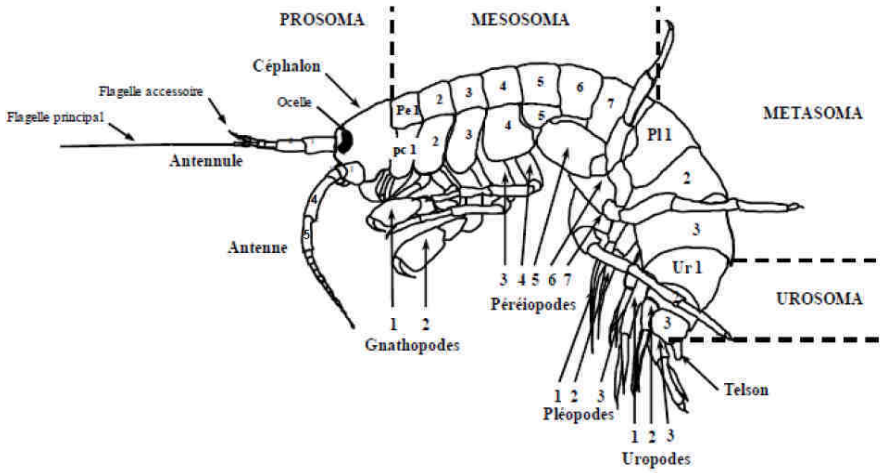


Figure I-6 : Vue latérale d'un Gammaridae (d'après Chevreux and Fage 1970, Roux 1970)

Pc : plaque coxales ; Pe : péréiomères ; Pl : pléomères ; Ur : uropomères.

C'est un crustacé muni de deux paires d'antennes. Chaque anneau porte, côté ventral, une paire d'appendices articulés et dont le tégument de chitine calcifié forme l'exosquelette. La morphologie générale d'un gammaridé est présentée en Figure I-6 et son corps est divisé en quatre régions distinctes comprenant de l'avant vers l'arrière, le prosoma (*i.e*. tête), le mésosoma (*i.e*. thorax), ainsi que le métasoma et l'urosoma qui forment l'abdomen (revu par Chevreux and Fage 1970, Roux 1970).

Sa taille maximale moyenne est de 20 mm et sa morphologie est caractérisée par une convexité dorsale régulière. Le mésosoma, composé de 7 périomères, porte 2 paires de gnathopodes permettant au gammare de se fixer sur un substrat et 4 paires de péreiopodes ayant une fonction de locomotion. Au niveau du métasoma, les pléopodes, appendices biramés, sont continuellement agités pour ventiler la cavité branchiale. Ils jouent également un rôle important dans la locomotion. *Gammarus fossarum* se caractérise par une cuticule non carénée et l'absence de rayures sur un corps plutôt lisse (Roux 1970, Karaman 1977). Les organismes appartenant à *Gammarus fossarum* et *Gammarus pulex* sont morphologiquement très proches les uns des autres et sont parfois référencées sous le terme «complexe *fossarum-pulex* ». La distinction peut notamment être réalisée au niveau de l'urosoma qui porte 3 paires d'uropodes, la rame interne faisant 1/3 de la rame externe au niveau de l'uropode 3 (Roux 1970) (Figure I-7).

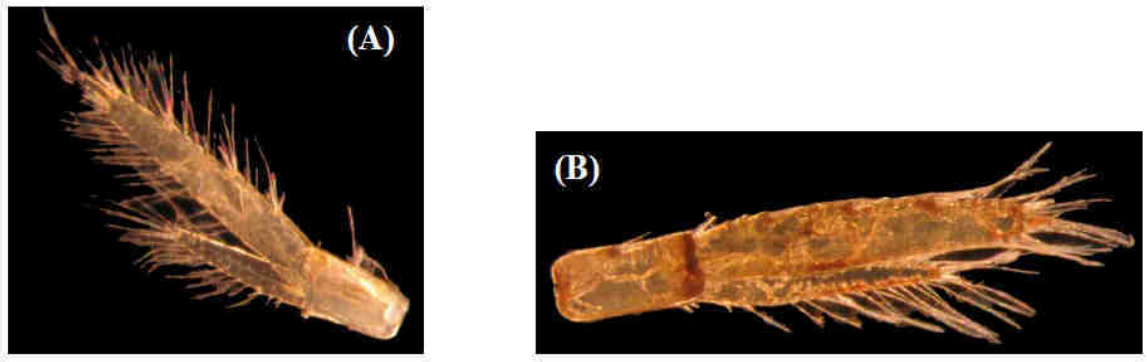


Figure I-7 : Uropode 3 de *Gammarus fossarum* (A) et de *Gammarus pulex* (B) (Source : Quéau H., Irstea Lyon)

1.3.1.2. Biologie et écologie

Le gammare est un organisme à croissance discontinue par mues successives, la rapidité de succession de celles-ci est dépendante de la température et du sexe. Il peut se reproduire tout au long de l'année et sa durée de vie est d'environ 2 ans. Sa respiration est branchiale et son appareil circulatoire, composé d'hémolymphe est largement ouvert. L'hémolymphe est un fluide unique extracellulaire dans lequel baignent des cellules nucléées, qui circule dans tout le corps, étant ainsi en contact direct avec la majorité des organes (Chapman, 1998). Les détails concernant la reproduction de l'organisme sont présentés dans la partie 2.1.1.

Les organismes du complexe *fossarum-pulex* peuplent les ruisseaux et les rivières de plaine et de moyenne montagne, écosystèmes caractérisés par un fort courant et une forte teneur en oxygène (Tachet *et al.* 2000). Regroupés en populations de forte densité pouvant atteindre plusieurs milliers d'organismes au m^2 (Felten 2003), les gammares affectionnent les substrats grossiers comme les galets ou les graviers et les substrats organiques comme les bryophytes, les végétaux rivulaires et la végétation morte. La Figure I-8 présente l'aire de répartition géographique de *G. fossarum*, allant des Pyrénées au milieu des Balkans (Barnard and Barnard 1983).

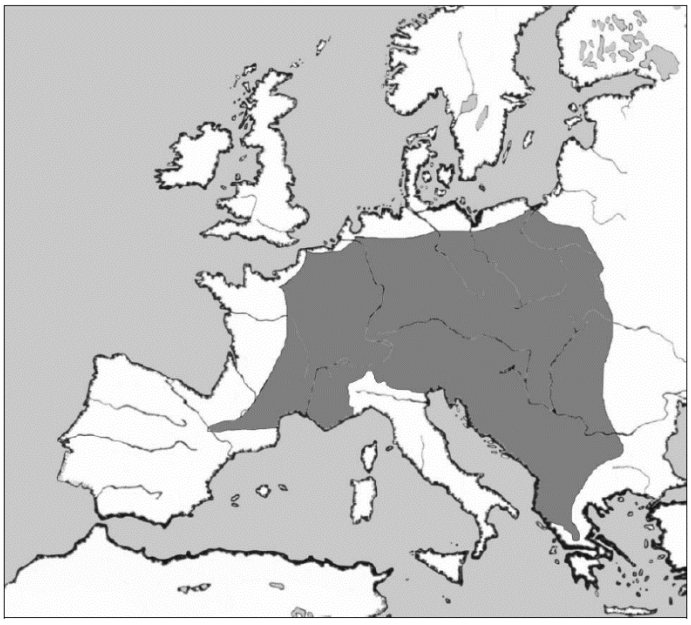


Figure I-8 : Répartition géographique de *Gammarus fossarum*(selon Barnard and Barnard 1983)

Au sein de l'écosystème, le gammare représente une importante source de nourriture pour différentes espèces de macroinvertébrés, de poissons, d'amphibiens et d'oiseaux (revu par Macneil

et al. 1999). Les gammares affichent également un vaste répertoire trophique, étant herbivores, détritivores et carnivores (revu par Macneil *et al.* 1997). Par leur statut de «shredders» (organismes déchiqueteurs), ils se nourrissent essentiellement de matière détritique telle que des feuilles. Ils jouent donc un rôle clef dans l'incorporation du matériel organique terrestre, en le fixant dans les réseaux trophiques d'eau douce, redistribuant ainsi la matière et l'énergie (Dangles and Guérold 2001, Maltby *et al.* 2002).

1.3.1.3. Utilisation en écotoxicologie

Les crustacés amphipodes sont fréquemment utilisés en écotoxicologie aquatique. Alors que le genre Gammarus est particulièrement employé en Europe, d'autres espèces d'amphipodes sont utilisées dans d'autres zones biogéographiques. Par exemple, en Amérique du Nord et Centrale, *Hyalella azteca* est employé depuis 25 ans en écotoxicologie aquatique continentale (Major *et al.* 2013), alors qu'en Australie, l'amphipode estuarien *Melita plumulosa* est employé pour l'évaluation de la toxicité des sédiments (Hyne *et al.* 2005). Cet engouement s'explique par leur statut d'espèce clé (pertinence écologique) mais également pour des critères pratiques, ces organismes étant facilement identifiables, échantillonnables et maintenables en laboratoire. De même, ils sont sensibles mais résistants à une large gamme de polluants (revu par Kunz *et al.* 2010).

Chez les organismes du genre *Gammarus*, l'impact du stress toxique a été étudié à différentes échelles d'organisation biologique [1] la **population** à travers la modélisation de sa dynamique (Coulaud *et al.* 2014), sa réparation en terme de micro-habitat, le suivi de classes d'âge et du sexe-ratio (Felten 2003), [2] de l'**individu** *via* le suivi de marqueurs comportementaux tels que le taux d'alimentation (Maltby *et al.* 2002, Coulaud *et al.* 2011, Charron *et al.* 2013) ou la locomotion et la ventilation (Lawrence and Poulter 1998, Felten *et al.* 2008, Felten *et al.* 2008, Xuereb *et al.* 2009), [3] au niveau **sub-individuel** avec l'étude de l'histologie des gonades ou des *caeca* hépatopancréatiques (Schirling *et al.* 2006, Scheil *et al.* 2008, Mazurová *et al.* 2010) ou encore par la mesure de biomarqueurs tels que le suivi de la génotoxicité (Lacaze *et al.* 2010, Lacaze *et al.* 2011), des réserves énergétiques (Charron *et al.* 2014), des dommages aux biomolécules (peroxydation des lipides)(Sroda and Cossu-Leguille 2011, Sroda and Cossu-Leguille 2011) ou encore l'osmolarité de l'hémolymphe (Felten *et al.* 2008, Felten *et al.* 2008).

Parmi les biomarqueurs les plus employés, on retrouve en majorité des biomarqueurs protéiques. Le Tableau I-2 référence pour les organismes du genre *Gammarus*, les études qui les emploient. Six espèces différentes ont été identifiées, quatre en eaux douces *G. fossarum*, *G. pulex* et *G. roeseli* en Europe et *G. lacustris* dans la région paléarctique et deux en eaux marines, *G. oceanicus* en Nord-Atlantique dans la zone sub-littorale et *G. locustra* en Nord-Est Atlantique dans la zone marnage. Sur un total de 29 publications, 21 biomarqueurs différents ont été employés. La majorité des biomarqueurs utilisés rendent compte d'une réponse générale au stress : activité de détoxification (10 %), défense antioxydative (20 %) et réponse générale au stress (22 %). De même, parmi les études présentées, 75 % ont été conduites en laboratoire. Les biomarqueurs sont peu employés dans une démarche de biosurveillance des masses d'eaux.

Tableau I-2 : Récapitulatif des différentes études employant des biomarqueurs protéiques chez les gammares

Contexte	Type de stress	Biomarqueurs employés	Espèce	Référence
Laboratoire	Stress thermal aigu	CAT, HSP70, sHSP, POD et GST	*G. oceanicus*	Bedulina *et al.* (2010)
Laboratoire	Matière organique naturelle	CAT, POD, HSP70 et sHSP	*G. pulex*	Bedulina *et al.* (2010)
Laboratoire	Salinité	Na^+/K^+ATPase	*G. duebeni et G.marinus*	Brooks and Lloyd Mills (2006)
Biomonitoring actif	Etude large échelle en contexte naturel	Enzymes digestives : Amylase, cellulase et trypsine	*G. fossarum*	Charron *et al.* (2013)
Laboratoire	Sédiment enrichis en cuivre	métallothionéines	*G. locusta*	Correia *et al.* (2001)
Biomonitoring actif	Etude en contexte naturel caractérisée par un gradient polymétallique	Enzymes digestives : amylase, cellulase et trypsine	*G. fossarum*	Dedourge-Geffard *et al.* (2013)
Biomonitoring actif	Etude en contexte naturel caractérisée par un gradient polymétallique	Enzymes digestives : amylase, estérase, endoglucanase, β-galactosidase et β-glucosidase	*G. fossarum*	Dedourge-Geffard *et al.* (2009)
Laboratoire	Cadmium	Na^+/K^+ATPase	*G. pulex*	Felten *et al.* (2008)
Laboratoire	Cadmium et parasitisme	HSP70 et métallothionéines	*G. fossarum*	Frank *et al.* (2013)
Laboratoire	Cadmium et parasitisme	métallothionéines	*G. roeseli*	Gismondi *et al.* (2012)
Laboratoire	Cadmium	Na^+/K^+ATPase	*G. fossarum*	Issartel *et al.* (2010)
Laboratoire	Hormones d'arthropodes	vitellogénine	*G. fossarum*	Simon *et al.* (2010)

Contexte	Type de stress	Biomarqueurs employés	Espèce	Référence
Laboratoire	Large gamme de xénobiotiques : hormones, pesticides et médicaments	vitellogénine	*G. fossarum*	Jubeaux *et al.* (2012)
Biomonitoring actif	Etude large échelle en contexte naturel	vitellogénine	*G. fossarum*	Jubeaux *et al.* (2012)
Laboratoire	Effluents d'usines produisant de l'huile d'olive et de citron	vitellogénine	*G. pulex*	Karaouzas *et al.* (2011)
Laboratoire	Cuivre	Enzymes digestives : N-acetyl-β-glucosaminidase, β-galactosidase et β-glucosidase	*G. pulex*	Lebrun *et al.* (2012)
Biomonitoring actif	Epandage de pesticides organiques	GST et AChE	*G. pulex*	Maltby and Hills (2008)
Laboratoire	Cinq classes de xénobiotiques : zinc, surfactant, organochloré, organophosphorés et pyrethroïde	GST et AChE	*G. pulex*	McLoughlin *et al.* (2000)
Laboratoire	Sédiments d'estuaire	métallothionéines	*G. locusta*	Neuparth *et al.* (2005)
Laboratoire	Filtre UV 3-benzylidene-camphor	HSP70	*G. fossarum*	Scheil *et al.* (2008)
Laboratoire	Pulse de cadmium	HSP70	*G. fossarum*	Schill *et al.* (2003)
Biomonitoring passif	Effluant de stations d'épuration	HSP70 et HSP90	*G. fossarum*	Schirling *et al.* (2005)
Laboratoire	Bisphenol A	HSP70 et HSP90	*G. fossarum*	Schirling *et al.* (2006)

Contexte	Type de stress	Biomarqueurs employés	Espèce	Référence
Laboratoire	Cuivre	CAT et GPxtot, SeGPx	*G. roeseli*	Sroda and Cossu-Leguille (2011)
Laboratoire	Stress thermique	GST, CAT, POD et sHSP	*G. lacustris*	Timofeyev *et al.* (2009)
Biomonitoring passif	Etude en contexte naturel suite à une mortalité massive d'écrevisses	HSP70	*G. pulex*	Triebskorn *et al.* (2002)
Laboratoire	Bloom de cyanobactéries et benzo[a]pyrène	GST, GPxtot, CAT et SOD	*G. oceanicus*	Turja *et al.* (2014)
Laboratoire	Chlorpyrifos	AChE	*G. fossarum*	Xuereb *et al.* (2009)

AChE : acetylcholinesterase ; CAT : catalase ; GST : glutathion S-transférase; GPxtot : glutathion peroxydase total; HSP70/90/sHSP : Heat Shock Protein 70 kDa, 90 kDa et small HSP, POD : peroxydase et SeGPx : Sélénium - dépendant glutathion peroxydase.

1.3.2. Contraintes actuelles pour l'utilisation des biomarqueurs en biosurveillance

L'évaluation écotoxicologique, contrairement au suivi de la qualité chimique et écologique, n'est pas intégrée au niveau de la DCE. En effet, différentes contraintes limitent l'utilisation routinière des biomarqueurs en biosurveillance.

1.3.2.1. Absence de développements spécifiques pour les invertébrés

Dans le cadre de l'évaluation des masses d'eau continentales par l'approche biomarqueur, le modèle poisson est l'un des plus employé (Narbonne 1998). En effet, cet organisme se montre intéressant pour différentes raisons. [1] Il peut être employé en tant qu'espèce « proxy » pour évaluer les risques sur les populations humaines, le poisson étant le seul vertébré exclusivement aquatique. [2] Il possède une importante valeur économique et concentre des enjeux sociétaux, sa consommation pouvant provoquer des effets néfastes sur la santé humaine et [3] en raison de leur place en haut de la chaine alimentaire, ils peuvent être considérés comme des intégrateurs de l'ensemble des conditions environnementales (Adams *et al.* 1993). Un des premiers biomarqueurs ayant été reconnu par la communauté scientifique pour l'évaluation écotoxicologique est le suivi de l'activité EROD (ethoxyresorufin-o-deethylase) chez le poisson, en réponse à une exposition à des hydrocarbures aromatiques polycycliques ou des dioxines. Ce biomarqueur étant spécifique et sensible, son application a par la suite été proposée chez les invertébrés. Cependant, chez les crustacés, bien que l'activité EROD ait été mesurée en réponse à l'exposition à des hydrocarbures aromatiques polycycliques, aucune induction suivant une relation dose-réponse n'a pu être démontrée. Chez *G. pulex* exposé à des polychlorobiphényles, le suivi de l'activité EROD n'a pas été concluant (Leroy *et al.* 2010). En effet, en raison de la distance phylogénétique, il est possible que le substrat utilisé pour le suivi de l'activité du CYP450 soit inadapté et que sa dégradation puisse être expliquée par l'intervention de CYP issus d'autres familles (James and Boyle 1998, Koenig *et al.* 2012). Il n'existe à l'heure actuelle aucun consensus sur la pertinence du suivi de l'activité EROD chez les crustacés.

Actuellement, la majorité des biomarqueurs employés chez les invertébrés résulte de la transposition directe de méthodes ayant fait leur preuve chez des organismes vertébrés. En raison de la distance phylogénétique, ces outils peuvent se révéler peu sensibles et discriminants face au stress toxique. Un autre exemple, illustrant cette problématique, concerne le suivi de la Vtg chez les crustacés et est présenté en I-2.3. Dans les faits, peu de développements de biomarqueurs

spécifiques aux invertébrés ont été réalisés, notamment en raison du manque de connaissance fondamentale de leur régulation au niveau moléculaire. Actuellement, une des seules alternatives existantes est le transfert d'outils des vertébrés aux invertébrés.

Dans le domaine médical, le suivi en routine de biomarqueurs protéiques se base sur le principe de la réaction entre l'anticorps et l'antigène par le biais du test ELISA (Enzyme-linked immunosorbent assay). Ce test permet la détection et la quantification de manière sensible et spécifique d'une protéine d'intérêt au sein d'un échantillon complexe (Simon *et al.* 2010). Cependant le développement de ce type de test nécessite d'importants investissements afin de purifier la protéine, produire un anticorps et s'assurer de sa spécificité (absence de réaction croisée non spécifique) or comme mentionné précédemment, peu de développements spécifiques de biomarqueurs ont été réalisés chez les invertébrés. De même, en raison du haut degré de spécificité entre l'anticorps et l'antigène, les tests ELISA sont sensibles à la distance phylogénétique et par conséquent peu transférables d'une espèce à l'autre. En l'absence de méthodologie pour la quantification directe de protéine, seules des approches indirectes peuvent être employées. Par exemple, dans le cas de la Vtg, le dosage de phosphates alkali-labiles, basé sur la quantification des phosphates labiles libérés par la Vtg après son hydrolyse dans un tampon alcalin, a été proposé pour mesurer la protéine (Gagné *et al.* 2005, Matozzo *et al.* 2008). Cependant cette méthode considère que cette protéine est la seule source de phosphore (Volz *et al.* 2002). Or chez les arthropodes, ce besoin est moins grand que chez les vertébrés et par conséquent, les Vtgs de ces derniers n'ont pas nécessairement évolués comme transporteurs de phosphates. Une autre alternative est la mesure de l'activité de la protéine, limitant fortement l'emploi de biomarqueurs non enzymatiques. Effectivement, parmi les biomarqueurs employés chez les gammares (Tableau I-2), les 2/3 portent sur le suivi d'une activité enzymatique. De même, chaque type de biomarqueur enzymatique requiert une procédure analytique spécifique en termes de tampon d'homogénéisation, de milieu réactionnel ou de longueur d'onde de mesure, le tout étant consommateur de matériel biologique. Or dans le cadre d'une approche multi-biomarqueur pour le suivi de la qualité des masses d'eau, cette multiplication de protocoles entraine une hausse des coûts par la mobilisation des ressources humaines et techniques ainsi que par la multiplication du matériel biologique.

1.3.2.2. Verrous dans l'interprétation des réponses

Idéalement, la réponse du biomarqueur ne serait modulée que par la pression toxique. Cependant, dans le cadre de l'utilisation de biomarqueurs liés à la physiologie de l'organisme, ces derniers sont en tout premier lieu modulés par la biologie de l'organisme. Le Tableau I-3 présente les différentes études réalisées sur le genre *Gammarus*, investiguant l'impact des paramètres biotiques ou abiotiques (dits facteurs confondants) sur la réponse des biomarqueurs.

Par exemple, le suivi de l'activité des enzymes digestives et de l'évolution des réserves énergétiques chez les femelles *G. fossarum* a mis en évidence une variation de ces deux paramètres au cours du cycle reproductif. Durant la première phase du cycle, les valeurs de deux paramètres augmentent, l'accumulation d'énergie étant nécessaire à la croissance des ovocytes alors que vers la fin du cycle, les réserves énergétiques et l'activité de digestion baissent (Charron *et al.* 2014). Chez les amphipodes, de nombreux paramètres biotiques ont une influence sur la réponse des biomarqueurs : le sexe, le stade de développement, le statut physiologique ou encore les phénomènes d'acclimatation locale sur les populations d'amphipodes autochtones. De plus, la variation de certains paramètres abiotiques tels que la température, le pH, la conductivité ou le taux d'oxygénation impacte également la réponse étudiée. Le suivi de l'activité des enzymes digestives chez les mâles *G. fossarum* en réponse à la température montre une baisse de l'activité à 7°C par rapport à 12°C et 16°C (Charron *et al.* 2013).

Afin de limiter l'impact des facteurs confondants biologique sur la réponse du biomarqueur, une stratégie de biosurveillance dite « active » peut être mise en place. Cette dernière repose sur l'implantation *in-situ* par encagement d'organismes « standards calibrés » (Coulaud *et al.* 2011, Lacaze *et al.* 2011). Cette démarche s'oppose au concept de biosurveillance « passive » où les réponses biologiques sont mesurées sur des organismes autochtones. L'approche active permet donc de contrôler le temps d'exposition des organismes mais également de réduire la variabilité intrinsèque du biomarqueur. Pour une meilleure définition des organismes « standards » en termes de taille, sexe, ou encore de statut reproducteur, l'influence de ces facteurs biotiques peut être préalablement explorée en laboratoire, comme présenté dans le Tableau I-3.

Tableau I-3 : Caractérisation de l'influence des facteurs confondants sur la réponse des biomarqueurs chez le gammare

Biomarqueur	Facteurs confondants	Espèce	Référence
AChE	Facteurs biotiques : Sexe, statut reproducteur de la femelle et masse de l'individu sur organismes acclimatés en conditions de laboratoire. Facteurs abiotiques : température 6, 12, 18 and 24 °C sur des organismes males après 2 et 15 jours d'exposition	*G. fossarum*	Xuereb *et al.* (2009)
Amylase	Sexe et statut reproducteur après 24 h d'acclimatation et disponibilité alimentaire après 19 h.	*G. palustris*	Borowsky and Guarna (1989)
Amylase	Température: 7, 12 et 16 °C et conductivité 200, 500 and 800 µS.cm-1 sur des organismes mâles après 20 jours d'exposition	*G. fossarum*	Charron *et al.* (2013)
Catalase	Âge/Taille et sexes sur organismes élevés en laboratoire	*G. locusta*	Correia *et al.* (2003)
Cellulase	Température: 7, 12 et 16 °C et conductivité 200, 500 and 800 µS.cm-1 sur des organismes mâles après 20 jours d'exposition	*G. fossarum*	Charron *et al.* (2013)
SeGPx	Âge/Taille et sexes sur organismes élevés en laboratoire	*G. locusta*	Correia *et al.* (2003)
SOD	Âge/Taille et sexes sur organismes élevés en laboratoire	*G. locusta*	Correia *et al.* (2003)
Trypsine	Température: 7, 12 et 16 °C et conductivité 200, 500 and 800 µS.cm-1 sur des organismes mâles après 20 jours d'exposition	*G. fossarum*	Charron *et al.* (2013)
Vtg	Statut reproducteur du mâle et de la femelle	*G. fossarum*	Jubeaux *et al.* (2012), Jubeaux *et al.* (2012)

Tableau I-4 : Caractérisation *in-situ* de la variabilité intrinsèque de la réponse du biomarqueur chez le gammare

Biomarqueur	Conditions d'expérimentations	Espèce	Référence
AChE	Impact saisonnier et spatial sur des organismes autochtones échantillonnés 1x/mois pendant 1 an sur une station de référence	*G. fossarum*	Xuereb *et al.* (2009)
Amylase	Organismes standards mâles encagés pendant 7 jours à 7 reprises sur 16 mois sur une station de référence	*G. fossarum*	Dedourge-Geffard *et al.* (2013)
Catalase	Impact du sexe et de la saison sur des organismes autochtones échantillonnés 1x/mois pendant 1 an sur une station de référence	*G. roeseli*	Sroda and Cossu-Leguille (2011)
Cellulase	Organismes standards mâles encagés pendant 7 jours à 7 reprises sur 16 mois sur une station de référence	*G. fossarum*	Dedourge-Geffard *et al.* (2013)
GPxtot	Impact du sexe et de la saison sur des organismes autochtones échantillonnés 1x/mois pendant 1 an sur une station de référence	*G. roeseli*	Sroda and Cossu-Leguille (2011)
Métallothionéines	Impact saisonnier et spatial sur des organismes autochtones échantillonnés 1x/mois pendant 1 an sur 3 stations et impact du sexe et de la masse sur 1 station	*G. pulex*	Geffard *et al.* (2007)
SeGPx	Impact du sexe et de la saison sur des organismes autochtones échantillonnés 1x/mois pendant 1 an sur une station de référence	*G. roeseli*	Sroda and Cossu-Leguille (2011)
Trypsine	Organismes standards mâles encagés pendant 7 jours à 7 reprises sur 16 mois sur une station de référence	*G. fossarum*	Dedourge-Geffard *et al.* (2013)
Vtg	Organismes standards mâles encagés pendant 21 jours au cours des 4 saisons sur 2 stations de référence	*G. fossarum*	Jubeaux *et al.* (2012)

Cependant, bien que cette approche permette de réduire la variabilité liée aux facteurs biotiques, elle ne permet pas de limiter l'influence des facteurs abiotiques, liés de manière intrinsèque au contexte naturel. Par conséquent, les études de biosurveillance sont couramment conduites sur un même réseau hydrographique, dans un contexte amont-aval par rapport à un point de rejet (Coulaud *et al.* 2011). En amont du rejet, une station est définie comme étant de « référence » et les valeurs mesurées des biomarqueurs comme des valeurs de référence relatives pour la comparaison avec les autres stations en l'aval. Cette approche se base sur l'hypothèse contestable que les conditions hydrogéologiques et abiotiques sont identiques au sein du même hydrosystème (Coulaud *et al.* 2011). En l'absence de valeur de référence absolue, valeur qui intègre la variabilité naturelle de la réponse, il est difficile de conduire une biosurveillance à large échelle, différentes stations ne pouvant être comparées au niveau spatial et temporel. Cette valeur de référence peut être définie en suivant la variabilité annuelle ou saisonnière mais également spatiale afin de créer une ligne de base de la réponse des biomarqueurs intégrant l'influence des facteurs abiotiques (Hanson *et al.* 2010). Idéalement, cette étape devrait être conduite sur des stations de références ayant des caractéristiques hydrogéologiques constatés (Xuereb *et al.* 2009, Jubeaux *et al.* 2012). Le Tableau I-4 présente différentes études réalisées sur le genre *Gammarus* caractérisant la variabilité intrinsèque naturelle de la réponse des biomarqueurs. Chez *G. fossarum*, cette stratégie a permis de définir des valeurs de références absolues pour le suivi de l'AChE, des enzymes digestives et de la génotoxicité et par conséquent, de valider leur pertinence et sensibilité en biosurveillance (Xuereb *et al.* 2009, Lacaze *et al.* 2011, Charron *et al.* 2013).

Cette stratégie de définition de valeur de référence absolue associée à une expérimentation à large échelle dans différents bassins versants présentant des profils de contamination variés, permet de statuer sur le caractère discriminant du biomarqueur dans un contexte de biosurveillance. Par exemple, afin de valider l'utilisation de la Vtg comme un biomarqueur de perturbation endocrine chez le gammare mâle, Jubeaux et *al*. ont encagé des organismes de taille comprise entre 8 et 11 mm pendant 21 jours (Jubeaux *et al.* 2012) au cours des quatre saisons dans deux stations de référence, l'une présentant une dureté de l'eau faible et l'autre moyenne. En parallèle, suivant la même procédure d'encagement, les auteurs ont étudié la réponse des organismes sur 21 stations ayant différents profils de contamination : 5 stations de référence, une impactée par une contamination métallique, une impactée par des composés organiques persistants et les 14 autres multi-contaminées. Alors que les expérimentations en laboratoire étaient en faveur de son

utilisation comme biomarqueur de la perturbation endocrine, l'application *in-situ* n'a pas permis de définir une ligne de base de la réponse en raison d'une importante variabilité interindividuelle ou de valider sa pertinence, des inductions indépendantes du contexte de contamination ayant été mesurées (Jubeaux *et al.* 2012).

2. Endocrinologie et perturbations de la reproduction chez les crustacés

En raison de son importance capitale dans le maintien de l'espèce dans son milieu, l'évaluation de l'intégrité de la fonction reproductrice fait l'objet d'une attention soutenue par la communauté scientifique. Cette fonction peut être altérée par une large gamme de xénobiotiques présents dans l'environnement et de manière spécifique, par les perturbateurs endocriniens. En écotoxicologie aquatique, alors que plusieurs outils robustes pour détecter leurs présence et effets chez les poissons sont disponibles, de tels outils sont inexistants chez les invertébrés. Au même titre que chez les vertébrés, les invertébrés régulent leurs grandes fonctions physiologiques, telles que la croissance, la reproduction ou le développement, *via* des hormones de signalisation dont la régulation peut être modulée par la présence de contaminants (Le Blanc, 2007). Cependant, alors que les deutérostomiens (Figure I-4) utilisent des stéroïdes (*e.g*. androgènes, œstrogènes et progestérone) en terminaison des cascades neuro-hormonales, les ecdysozoaires se caractérisent par l'utilisation d'ecdystéroïdes et de terpenoïdes dont la synthèse est initiée par des neurohormones spécifiques (Lafont 2000, LeBlanc 2007, Mazurová *et al.* 2008, Christie 2011, Mykles 2011, Webster *et al.* 2012).

Cette partie a pour objectif d'établir un bilan, chez les crustacés, des connaissances relatives à leur reproduction et leur endocrinologie, de recenser les différentes atteintes à ces processus ainsi que leurs causes connues et enfin les outils actuellement disponibles pour évaluer la reprotoxicité.

2.1. Reproduction et endocrinologie des crustacés

La régulation de la fonction reproductrice chez les crustacés a fait l'objet de différentes synthèses bibliographiques (LeBlanc 2007, Rodríguez *et al.* 2007, Mazurová *et al.* 2008, Hyne 2011). Cependant, chez les arthropodes, la majorité des études en endocrinologie a été réalisée sur la mouche *Drosphila melanogaster*. Chez les crustacés, l'identification et la caractérisation spécifique du système neuroendocrinien ont été principalement effectué chez les décapodes (crabe, crevettes, homard) en raison de leur valeur économique et de leur taille, les amphipodes n'ayant été que peu étudiés. En revanche chez *Gammarus fossarum*, sa reproduction au niveau physiologique et comportementale a été finement caractérisée, permettant de focaliser la prochaine partie sur notre modèle d'étude.

2.1.1. Reproduction

2.1.1.1. Anatomie du système reproducteur et endocrinien

Les gammares sont des organismes **gonochoriques**, l'anatomie de leur système reproducteur et endocrinien est présentée en Figure I-9.

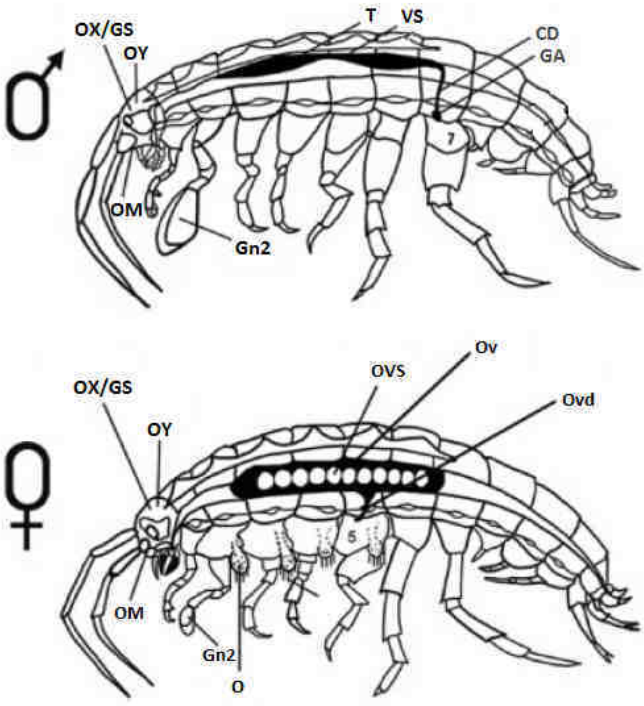


Figure I-9 : Représentation schématique en vue latérale de la position des principaux organes du système reproducteur et endocrinien chez un amphipode mâle et femelle du sous-ordre *Gammaridea* (modifié d'après Hyne 2011)

CD : canal déférent; GA : glande androgène ; Gn2 : gnathopode secondaire ; O : oostégite ; OM : organe mandibulaire ; Ov : ovaire ; Ovd : oviducte ; OVS : oocytes en vitellogénèse secondaire ; OX/GS : organe X/ glande du sinus ; OY : organe Y ; T : testicule ; VS : vésicule séminale

En terme de sécrétion hormonale, quatre sites ont été identifiés comme étant impliqués dans la régulation neuroendocrine de ces organismes : [1] Le complexe « organe X/glande du sinus », [2] l'organe Y, [3] l'organe mandibulaire et [4] la glande androgène uniquement présente chez le mâle. Chez les amphipodes, le complexe **organe X/glande du sinus** (OX-GS) se situe en périphérie des lobes optiques (Blanchet-Tournier 1980) alors que chez les décapodes il est localisé dans les pédoncules oculaires. L'**organe Y** (OY) ou glande de mue, est constitué d'une paire de glande épitheloïdes située au niveau du céphalothorax.

L'**organe mandibulaire** (OM) situé à proximité du point d'attache des muscles mandibulaires postérieurs, est constitué de glandes hautement vascularisées. La **glande androgène** (GA), uniquement présente chez le mâle est attachée à la partie postérieure du vas déférent.

L'**appareil génital** est composé de deux organes fusiformes indépendants, situés sous le cœur, au-dessus du tube digestif et des *caeca* hépatopancréatiques, s'étendant dans le mésoma (Figure I-10). Chez le mâle, les organes présentent un étranglement séparant la région antérieure ou testicule, d'une vésicule séminale se prolongeant par un canal déférent. Les deux vas déférents, où la gonade androgène est attachée en région sub-terminale, aboutissent séparément au niveau de la cavité ventrale aux papilles génitales servant au dépôt du sperme. Chez la femelle, les ovaires occupent le même emplacement que les testicules chez les mâles. Les oviductes se détachent du tiers inférieur de chaque ovaire et aboutissent séparément à la base des oostégites, de part et d'autre de la ligne médiane ventrale. Les oostégites forment une cavité ventrale appelée marsupium où les œufs fécondés sont déposés et se développent jusqu'à éclosion (Tachet *et al.* 2000).

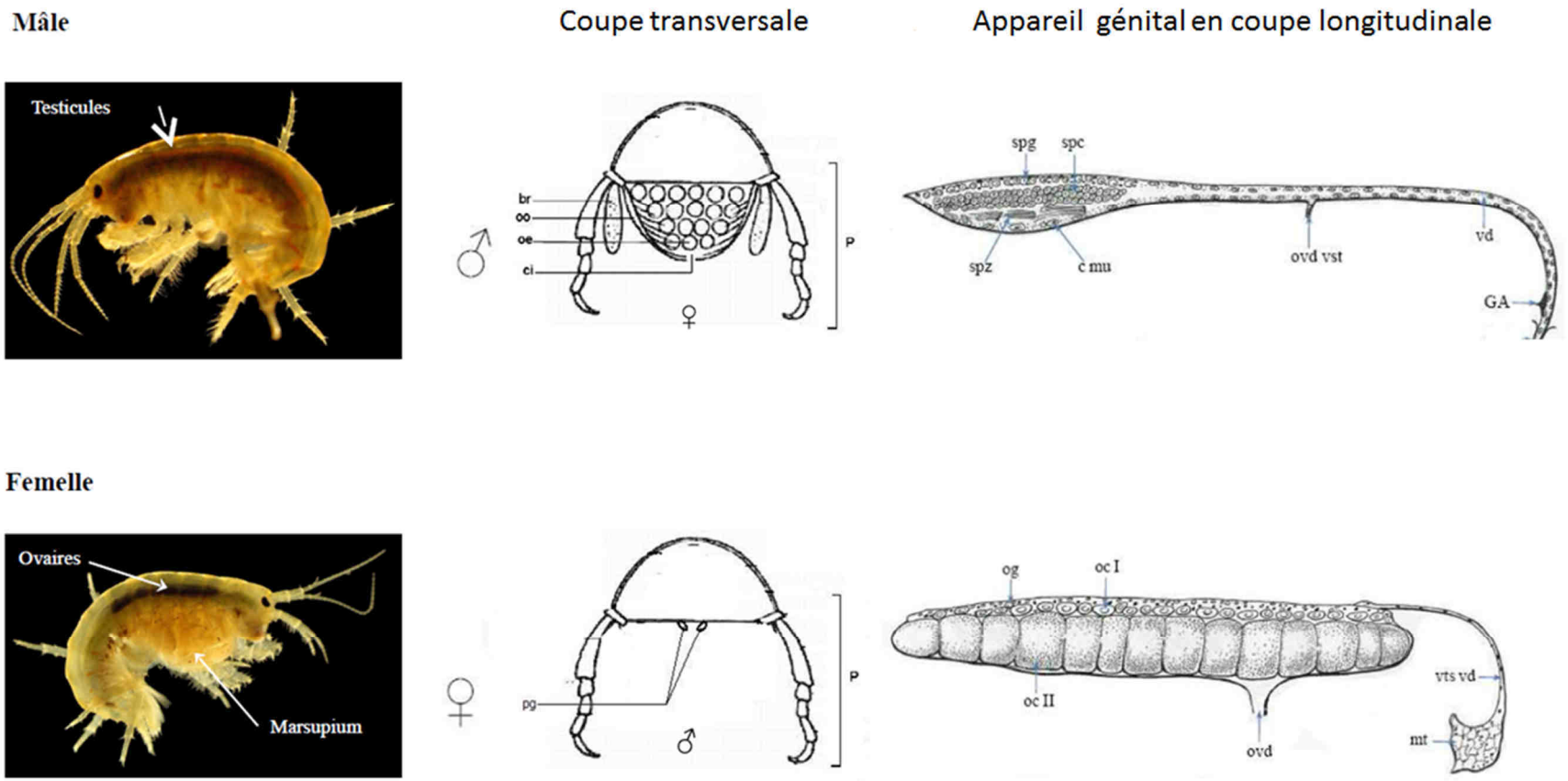


Figure I-10 : *Gammarus fossarum* mâle et femelle, schéma d'une coupe transversale de *Gammaridae* et des appareils génitaux en coupe longitudinale (d'après Charniaux-Cotton 1965)

br, branchie ; c mu : cellules à mucus ; GA : glande androgène ; ci, chambre incubatrice ; mt : tissu mésenchimatique indifférencié ; oc I : ovocyte en vitellogénèse primaire ; oc II : ovocyte en vitellogénèse secondaire ; oe, œufs ; og : ovogonie ; oo :oostégite ; ovd : oviducte ; ovd vst : oviducte vestigial ; g, papille génitale ; P, péréiopode ; spc : spermatocyte ; spg : spermatogonie ; spz : spermatozoïde ; vd : vas déférent ; vts : vestige du vas déférent, vs : vas déférent

Chez les crustacés, différentes fonctions physiologiques sont impliquées dans le processus reproductif : la différenciation sexuelle, la croissance et le développement, la mue, la gamétogénèse et la vitellogénèse.

2.1.1.2. Cycle reproducteur chez la femelle

Chez les crustacés amphipodes, le cycle de reproduction des femelles est rythmé par le cycle d'intermue (Charniaux-Cotton 1973, Blanchet-Tournier 1980). Chez les femelles sexuellement actives, la maturation des gonades (*i.e*. vitellogenèse) et le développement des embryons dans le marsupium se déroulent de manière parfaitement synchrone, à chaque cycle d'inter-mue. L'exuviation peut ainsi être considérée comme le point de départ et d'aboutissement de la maturation gonadique ainsi que du développement embryo-larvaire dans le marsupium. Les juvéniles éclos, issus de la ponte précédente, sortent du marsupium peu de temps avant l'exuviation de leur mère. Suite à cette exuviation, la femelle pond les œufs dans le marsupium qui seront aussitôt fécondés par un mâle. Parallèlement, dans la gonade, un nouveau lot d'ovocytes primaires entre en vitellogénèse.

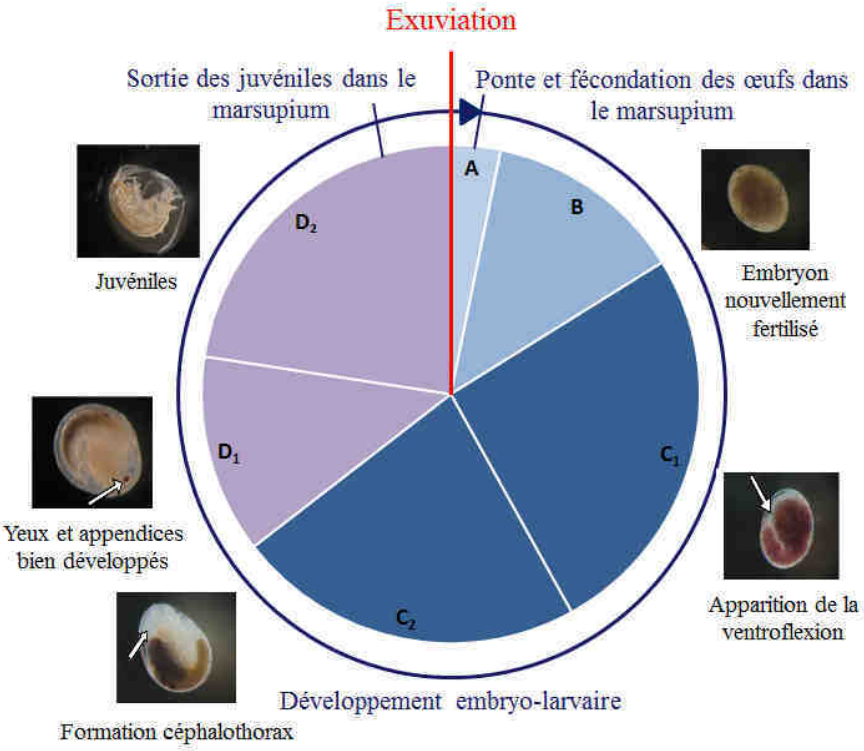


Figure I-11 : Stades embryonnaires corrélés aux stades de mues de la femelle *G. fossarum* (Xuereb 2009)

Chez la femelle adulte *G. fossarum*, les cycles de mue et de reproduction ont été caractérisés précisément par Geffard *et al.* (2010) (Figure I-11). Le cycle de mue dure environ 30 jours, à 12°C et se découpe en 3 phases (*e.g*. post-mue, intermue et pré-mue) comprenant 6 stades dont la détermination se base sur l'observation microscopique du développement de la nouvelle cuticule et du décollement de l'ancienne au niveau de la griffe des péréiopodes 3 et 4 (Figure I-12).

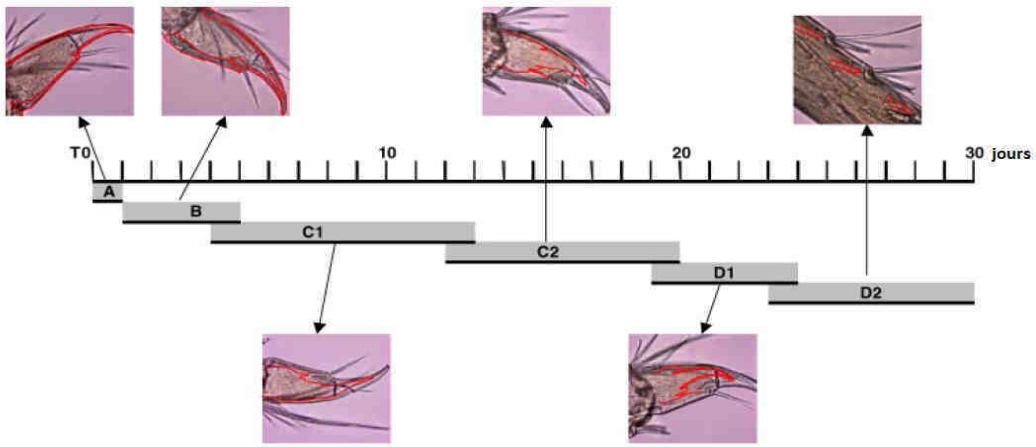


Figure I-12 : Changements intertégumentales à l'extrémité des périopodes chez *Gammarus fossarum* au cours du cycle de mue (Geffard *et al.* 2010)

La phase de **post-mue** comprend le stade A et B. Le stade A, se déroule sur un temps très réduit (1 jour) et commence à la fin de l'exuviation de l'exosquelette. La stade B est caractérisé par un durcissement de la cuticule et se déroule sur 4 jours. Au cours de cette phase, les ovocytes sont d'apparence blanche et présentent peu de globules lipidiques et de vésicules vitellines. L'enveloppe folliculaire est inégale. En parallèle, le 1er stade embryonnaire correspond à des œufs nouvellement fertilisés, de forme ovale et indifférenciée.

La phase d'**intermue** se déroule au cours du stade C1 et C2 et est caractérisée au niveau de la cuticule, par une rétractation de l'épiderme et dure au total 17 jours. Entre le stade C1 et C2, la vitellogénèse secondaire prend place. Le nombre moyen d'ovocytes baisse, passant d'une moyenne de 21 oocytes/femelle à 13. En parallèle, une importante croissance folliculaire a lieu et les vésicules vitellines croient en nombre et en taille. Dans le même temps, les embryons acquièrent une forme de virgule (ventroflexion) en stade 2 puis le céphalothorax se forme en stade 3.

Finalement la phase de **pré-mue** débute au 19ème jour, et est caractérisée par la synthèse de la nouvelle cuticule en cours du stade D1 et D2. Les ovocytes possèdent une structure identique à celle du stade C2, avec une surface folliculaire augmentée. Au niveau de l'embryogénèse, au cours de l'avant dernier stade (stade 4), les yeux et les appendices se développent pleinement, le dernier stade (5) correspondant à des juvéniles nouvellement éclos.

2.1.1.3. Spermatogénèse

Alors que chez la femelle, la maturation des ovocytes se fait de manière cyclique en parallèle de la mue, la maturation de gamètes chez le mâle est un processus constant et non synchrone, les gamètes présents dans les testicules étant à différents stades de développement. A l'origine, une spermatogonie secondaire entre en méiose, se transforme en spermatocyte primaire diploïde qui produit deux spermatocytes secondaires évoluant en spermatides puis en spermatozoïdes (Reger 1966). La Figure I-13 présente les différences morphologiques entre les spermatozoïdes et les spermatides chez *G. fossarum*.

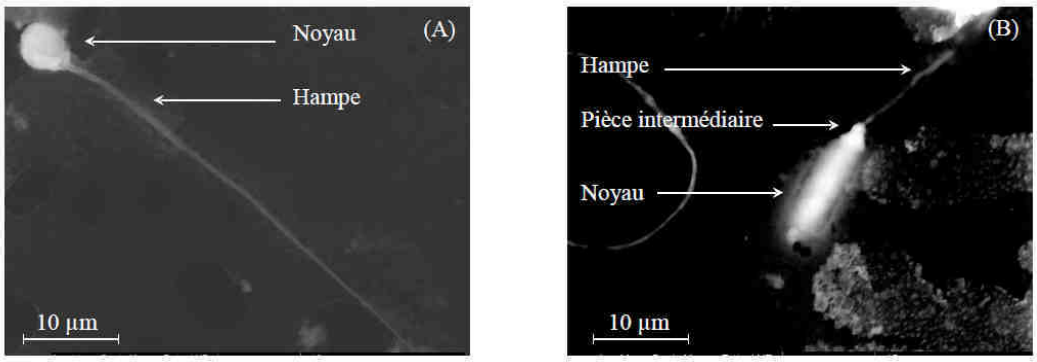


Figure I-13 : Micrographies de spermatide (A) et de spermatozoïde (B) de *G. fossarum* en microscopie électronique à balayage utilisé en mode wet-STEM (x 2022) (Lacaze 2011)

Chez les crustacés, l'organisation générale des cellules reproductrices masculines diffère des vertébrés chez qui les cellules sont flagellées et mobiles. Chez les amphipodes, la forme du noyau des spermatozoïdes est oblongue alors que les spermatides ont un noyau ovoïde, les deux étant pseudo-flagellés (Lacaze 2011). En effet, la structure pouvant être confondue avec un flagelle est une hampe, résultant d'une invagination de la membrane cellulaire et non mobile. De même, la tête allongée est caractéristique d'un spermatozoïde mature, correspondant à l'allongement de la vésicule acrosomique au cours de la maturation (Reger 1966).

Bien que le processus de production des spermatozoïdes soit continu, il peut également se dérouler par vagues, notamment après la copulation, afin de reconstituer les réserves spermatiques. Chez *G. fossarum*, Lacaze (2011) a décrit la cinétique de reconstitution des réserves spermatiques à 12 °C, suite à la copulation (désignée spermatogénèse dans le reste du manuscrit), présentée en Figure I-14.

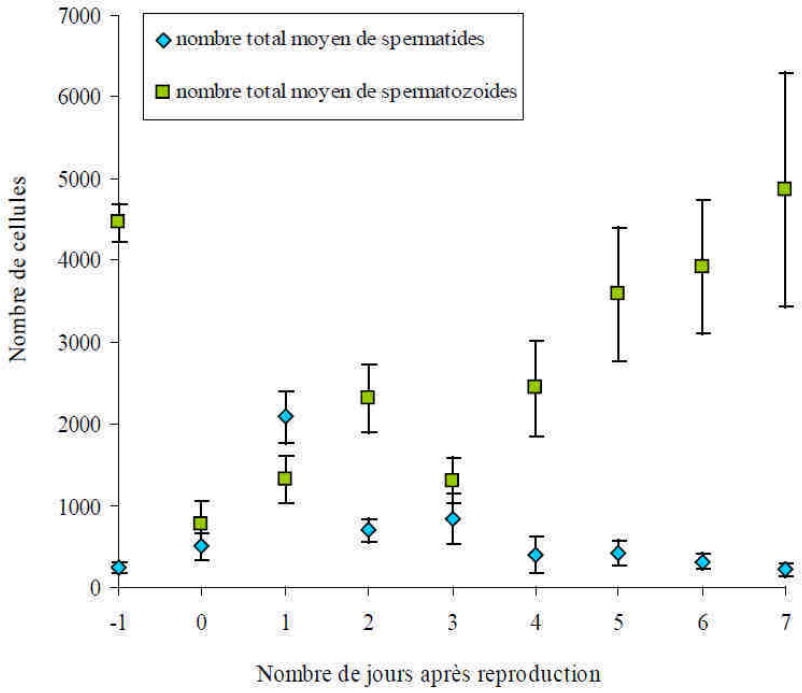


Figure I-14 : Evolution du nombre de spermatozoïdes et de spermatides contenu dans la gonade en fonction du nombre de jour après la copulation chez *G. fossarum* (Lacaze 2011)

J-1 : mâles en précopulât

La durée de la spermatogénèse est de 7 jours, la quantité de spermatozoïdes étant maximale et similaire à celle d'un organisme en précopulât. Pour les spermatides, dans le jour qui suit la copulation, leur quantité est maximale puis à J2, cette quantité baisse et reste constante au cours du temps.

2.1.1.4. Accouplement

Alors que la femelle est en phase de pré-mue, le mâle s'agrippe sur son dos au moyen de ses gnathopodes, et la maintient en place sous son corps. Le couple forme alors une figure appelée amplexus précopulatoire (ou précopulât) présenté en Figure I-15.



Figure I-15 : *Gammarus fossarum* mâle et femelle en précopulât (photo : Quéau H., Irstea)

En raison de la brève période de réceptivité de la femelle au cours du cycle, ce comportement serait une stratégie évolutive face à la compétition entre les mâles pour l'accès aux femelles réceptives. Cette étape dure quelques jours jusqu'à la mue de la femelle et sa durée est sous dépendance du sexe-ratio au sein de la population d'amphipode (Bollache and Cézilly 2004). L'exuvie de la femelle est facilitée par la présence du mâle et elle est suivie par la ponte des œufs. Le mâle se positionne alors ventralement, en travers, de façon à ce que ses papilles génitales se retrouvent en vis-à-vis du marsupium de la femelle. Le sperme est alors déposé sur les œufs fraîchement pondus. Cette opération est renouvelée deux à trois fois (Le Roux 1933). Chez *G. fossarum*, le mâle investit environ 80 % de ses réserves spermatiques dans la copulation.

La période de reproduction s'étend presque tout au long de l'année, avec des pics observés au printemps et à la fin de l'été (Le Roux 1933, Felten 2003).

2.1.2. Endocrinologie

2.1.2.1. Fonctionnement du système endocrinien

En réponse à des stimuli physiologiques ou environnementaux, le système nerveux, par le biais de structures neurosécrétrices, initie une cascade biochimique afin de communiquer avec le système endocrinien. Cette interaction, nommée régulation neuroendocrine, est initiée par des peptides neuroendocriniens. Ces peptides régulent directement le processus physiologique associé ou modulent la production et sécrétion des hormones. Ces messagers chimiques sécrétés par les glandes endocrines sont transportés par le système circulatoire jusqu'à l'organe cible afin de réguler les processus biologiques endogènes de l'organisme. L'action des hormones se base sur la présence de récepteurs nucléaires protéiques hautement spécifiques au niveau des cellules cibles et la formation d'un complexe récepteur-hormone qui médiate l'activité biochimique dans le tissus cible (*e.g*. voie génomique ; Figure I-16). La réponse physiologique obtenue au niveau des multiples sites d'action dépend alors de la balance hormonale c'est-à-dire de l'équilibre entre les différents médiateurs chimiques libérés.

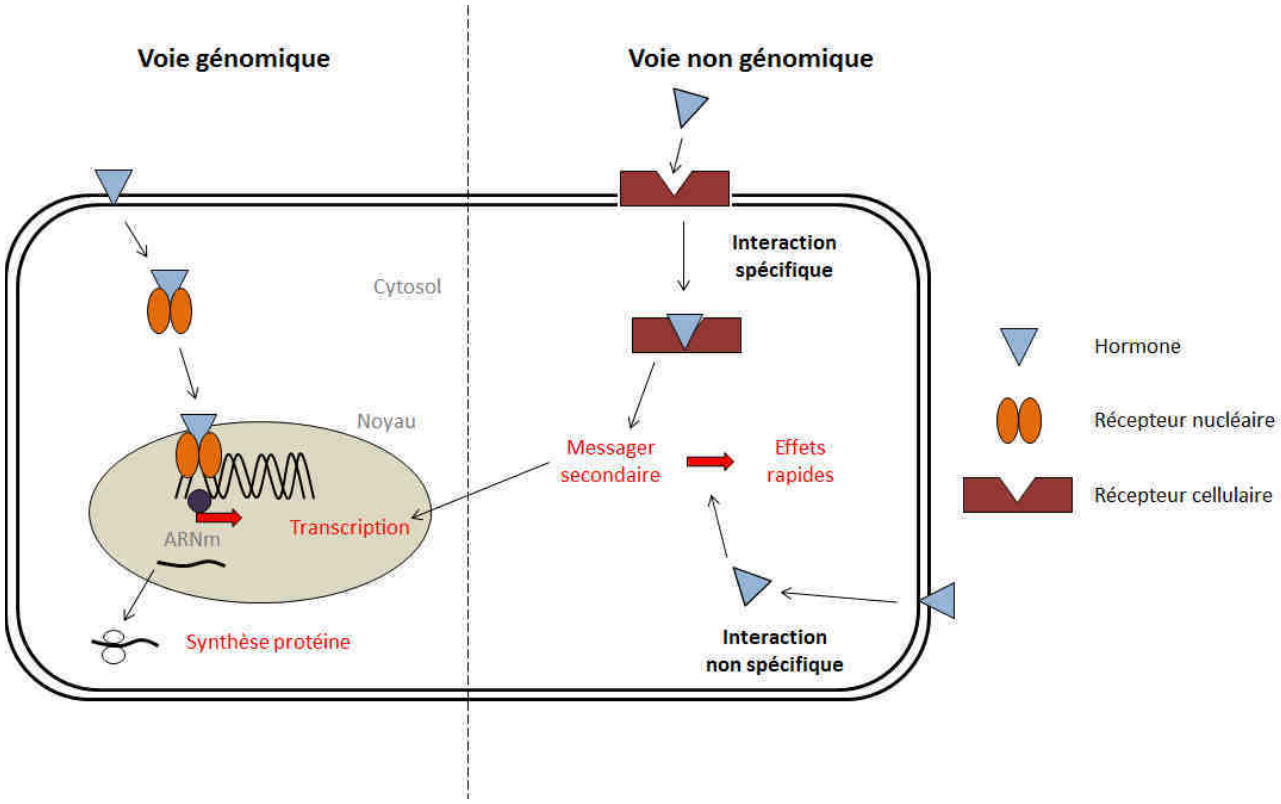


Figure I-16 : Schéma représentant les différentes voies de signalisation hormonales (adapté de Luconi 2006)

En addition de ce mode d'action caractérisé de « génomique », un mode « non génomique » existe, se caractérisant par une réponse rapide de l'organisme en réponse à l'hormone (quelques secondes) et une insensibilité aux inhibiteurs de la transcription et de la synthèse protéique (De Loof *et al.* 2013). Ces effets rapides peuvent être médiatés suivant deux modes, comme présenté dans la Figure I-16 : [1] une interaction spécifique par la formation d'un complexe avec récepteur cellulaire aux propriétés pharmacologiques distinctes de celles du récepteur nucléaire et [2] une interaction non spécifique (Falkenstein *et al.* 2000, Luconi 2006). Cette dernière interaction est encore relativement mal comprise mais son existence est appuyée par le fait que certains organes concentrent de très fortes concentrations en hormones, or une interaction spécifique se caractérise par une haute affinité entre le récepteur et le ligand, ne nécessitant que de faibles concentrations en ligand (De Loof *et al.* 2013).

Chez les crustacés, quatre classes d'hormones/neuro-hormones ont été identifiées: [1] les neuropeptides, [2] les hormones stéroïdiennes, [3] les hormones terpenoïdes et [4] les glycopeptides, synthétisées par quatre sites principaux (Figure I-9) : [1] Le complexe « organe X/glande du sinus », [2] l'organe Y, [3] l'organe mandibulaire et [4] la glande androgène.

Le complexe OX/GS est le centre de contrôle neuroendocrine, site de synthèse des **peptides neuroendocriniens**. L'OX, considéré comme un organe de sécrétion, est formé de groupes de cellules neurosécrétrices dont les terminaisons axonales entourent la GS adjacente. En réalité le terme « glande » est impropre car cet organe est en réalité une structure neurohémale comportant de grandes lacunes hémolymphatiques, occupant une fonction de stockage et de diffusion des neuropeptides (Stuenkel 1983). Les neuro-hormones synthétisées par l'OX/GS sont de la superfamille des « crustacean hyperglycemic hormone » (CHH), régulant un large spectre de fonctions biologiques. Ces hormones, composées d'environ 80 acides aminés, présentent une forte homologie de séquences. En fonction de leur précurseur et de leur structure primaire, elles peuvent être classées en deux groupes : type I les CHH *senso stricto* et les peptides transporteurs d'ions et le type II, les hormones inhibitrices de la mue (MIH), de l'organe mandibulaire (MOIH), de la vitellogénèse (VIH) et de la gonade (GIH) (Nakatsuji *et al.* 2009, Webster *et al.* 2012).

L'organe Y est le site principal de biosynthèse des **ecdystéroïdes** à partir du cholestérol (Figure I-17).

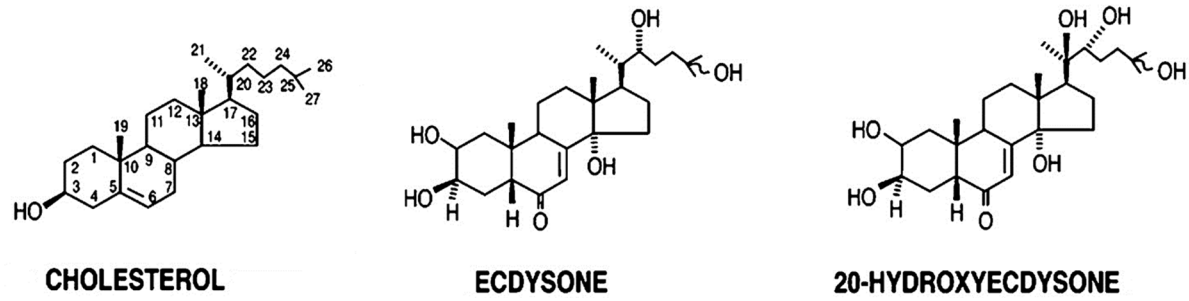


Figure I-17 : Structure chimique du précurseur de la biosynthèse des ecdysteroids (cholestérol), de l'ecdysone et de la principale forme actives (20-hydroxyecdysone) (adapté de Petryk *et al.* 2003)

En l'absence de synthèse *de novo*, le cholestérol, assimilé par l'alimentation, est transporté dans l'hémolymphe par des lipoprotéines jusqu'à l'OY. Bien que des différentes étapes restent relativement méconnues chez les crustacés, la biosynthèse des ecdystéroïdes semble être catalysée par des réductases et mono-oxygénases P450 (cytochrome). Les ecdystéroïdes, notamment l'ecdysone, sont par la suite convertis en forme biologiquement active dans les tissus périphériques. Une des principales formes est la 20-hydroxyecdysone (20-HE), aussi appelée hormone de mue (Mykles 2011). Les effets de la 20-HE sont majoritairement médiatés par un hétérodimère de deux récepteurs nucléaires : l'EcR (ecdysone receptor) et l'USP (ultraspiracle). L'EcR-USP a été principalement étudié chez *Drosophila melanogaster*, où il est impliqué dans le processus de métamorphose (Nakagawa and Henrich 2009). La synthèse d'ecdysone est sous le contrôle négatif du neuropeptide MIH sécrété par l'OX/GS. De même, chez le homard *Homarus americanus*, la présence de l'EcR a été mise en évidence dans le cytoplasme de cellules du complexe OX/ GS, suggérant la présence d'une boucle de rétrocontrôle (El Haj *et al.* 1994).

Alors que les principaux sites de synthèse des hormones chez les arthropodes sont reconnus comme étant les quatre autres organes précédemment cités, il a été prouvé que les ovaires ainsi que les œufs contiennent une forte teneur en ecdystéroïdes (résumé par Subramoniam 2000). Bien que l'hypothèse généralement retenue est une séquestration de ces hormones stéroïdiennes à partir de hémolymphe, des études chez le crabe *Carcinus maenas* suggèrent que les ovaires sont capables de synthèse *de novo* (Styrishave *et al.* 2008).

L'organe mandibulaire (OM) synthétise, à partir d'acide farnesoïque, des **hormones terpenoïdes**, nommées de manière générique chez les arthropodes « hormones juvénoïdes » (Nagaraju 2007, Hyne 2011) (Figure I-18). Aussi, chez les insectes, il a été mis en évidence que la GA contient

également de grande quantité d'hormone juvénile, par conséquent, cet organe pourrait également être un site majeur de synthèse des hormones terpenoïdes (De Loof *et al.* 2013).

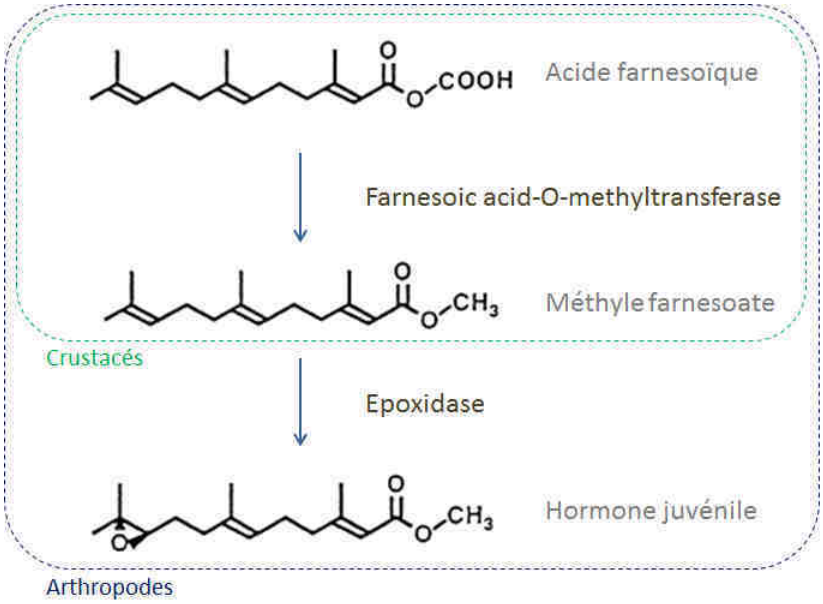


Figure I-18 : Voies de synthèse de l'hormone juvénile chez les arthropodes (adapté de Hui *et al.* 2010)

Chez les crustacés, le produit final de la synthèse est le méthyle-farnesoate (MF) formé par la méthylation de l'acide farnesoïque par l'action d'une méthyltransférase. En revanche, chez les insectes, le MF est un précurseur de l'hormone juvénile. En effet, à la suite du recrutement d'une epoxidase au cours de l'évolution, l'hormone juvénile est considérée comme une nouveauté évolutive, unique à ce groupe (Hui *et al.* 2010, Kenny *et al.* 2013). Une fois synthétisées par le MO, les hormones juvénoïdes sont transportées par des lipoprotéines jusqu'aux tissus cibles. L'activité de synthèse et de sécrétion de la MF est sous le contrôle négatif d'un neuropeptide sécrété par l'OX/GS, le MOIH (Wainwright *et al.* 1996).

Au cours des années 50-60, Charniaux-Cotton a réalisé sur l'amphipode marin *Orchestia gammarella* une série d'ablations et d'implantations croisées des gonades mâles et femelles. De ces expériences, l'hypothèse d'une hormone secrétée par la GA a été émise car cet organe est essentiel pour le maintien des caractéristiques sexuelles primaires et secondaires mâles et en son absence, le crustacé se développe spontanément en femelle (Charniaux-Cotton 1954). Presque un siècle plus

tard, la caractérisation de l'**hormone androgène** (AGH) et du gène l'encodant a été réalisée chez l'isopode terrestre *Armadillidium vulgare*.

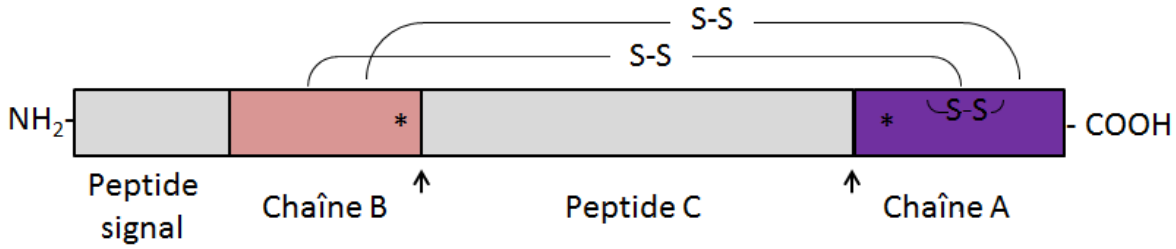


Figure I-19 : Modèle linéaire représentant la structure de la pré hormone produite par la glande androgène du crabe *Cherax quadricarinatus* (modifé de Ventura *et al.* 2011).

Les flèches indiquent les sites de clivages donnant l'hormone mature, constituée des chaines A et B reliées entre elles par des ponts disulfures (S-S). La chaine A contient également un pont disulfure. Les astérisques représentent les sites de glycosylation prédits.

Encodé comme une préhormone, l'AGH présenté en Figure I-19, est sous sa forme active, une hormone de nature peptidique avec des sites de glycosylation, caractérisée d'insuline-like (Okuno *et al.* 1999). Par la suite, les séquences de l'AGH ont été caractérisées chez plusieurs espèces de décapodes (Ventura *et al.* 2011).

2.1.2.2. Régulation hormonale des processus physiologiques

La différenciation sexuelle est un processus biologique capital qui affecte la différenciation des gonades mais induit également des différences au niveau morphologiques, physiologiques et comportementales. Les crustacés amphipodes sont généralement des espèces gonochoriques c'est-à-dire que les juvéniles sont sexuellement indifférenciés à l'éclosion et que l'acquisition des caractères sexuels se fait au cours du développement, par une succession de mues (Charniaux-Cotton and Payen 1988). Au niveau de la machinerie génétique, les mécanismes de la différenciation sexuelle chez les arthropodes ont été majoritairement étudiés chez la drosophile *Drosophila melanogaster*. Au début de la cascade, on retrouve le gène « Sex lethal » et en fin de cascade, le gène « doublesex ». Ce dernier a été identifié comme un facteur de transcription qui encode des transcrits spécifiques du mâle ou de la femelle, et qui ciblent directement les gènes conférant le dimorphisme sexuel. Les homologues de ces gènes ont été retrouvés chez différentes

espèces d'insectes et le rôle du gène « doublesex » semble être conservé (Sánchez 2008). Plus récemment, chez la daphnie *Daphnia magna*, deux homologues de ce gène ont été identifiés dont un semble être un facteur clé pour la détermination du phénotype mâle. Cependant, cette détermination sexuelle dite « génétique », réside dans la ségrégation des gènes qui initie les différentes voies de différenciation. Or, chez les crustacés, la différenciation sexuelle et en général la régulation des processus physiologiques liés à la reproduction, sont relativement plastiques et fortement influencées par les conditions environnementales (photopériode, température, disponibilité alimentaire). La daphnie étant un organisme principalement parthénogénétique, le passage à une reproduction sexuée par la production de mâle semble être principalement contrôlé par des facteurs environnementaux (dégradation des conditions), ces derniers agissant par des voies de signalisations alternatives à celles de la différenciation génétique (Kato *et al.* 2011).

Alors que la composante génétique dans la différenciation sexuelle reste très méconnue chez les crustacés, l'action des hormones est mieux comprise. Les premiers travaux ont été initiés sur l'amphipode *Orchestia gammarella* par des expériences d'ablation et d'implantation de la glande androgène(Charniaux-Cotton 1954) et ont permis de démontrer son rôle crucial comme facteur majeur de différenciation sexuel masculin. Par la suite, des observations similaires ont été réalisées chez d'autres ordres de crustacés (isopodes, décapodes) et ont démontré que le développement des ovaires se fait par autodifférenciation en absence de l'AGH (résumé par Ford 2008). La différenciation et de le développement de la GA chez les organismes génétiquement mâles se fait sous le contrôle du système neuro endocrinien (protocérébron) par une hormone gonado-stimulatrice (GSH : Gonado-Stimulating hormone). Par la suite, l'AGH est synthétisée par la glande associée et induit la différenciation des caractères sexuels primaires et secondaires et réprime le processus de vitellogénèse. Cependant, l'AGH ne semble pas être un déterminant sexuel car l'expression de certains caractères sexuels secondaires chez la femelle est sous la dépendance d'hormones ovariennes. En effet, l'ablation des ovaires chez la femelle *Orchestia gammarella* a conduit à la perte des caractères sexuels secondaires permanents (oostégites) et temporaires (développement de soies au cours des périodes de reproduction), suggérant l'existence d'un contrôle par des hormones ovariennes, dont la nature chimique reste à déterminer.

Chez plusieurs espèces d'arthropodes, il a été démontré que la GA contient de fortes teneurs en hormone juvénile (De Loof *et al.* 2013). Plusieurs travaux chez la Daphnie ont permis de démontrer que les terpenoïdes sont également un facteur de différenciation sexuelle. L'exposition en

laboratoire, de femelles au MF et à des composés analogues stimule la production de mâles (Olmstead and LeBlanc 2003, LeBlanc 2007, Lampert *et al.* 2012). De même, chez le crabe *Carcinus maneas* une corrélation entre le niveau de MF dans l'hémolymphe et le développement des testicules a été mise en évidence (Nagaraju and Borst 2008).

Le contrôle de la **spermatogénèse** chez les crustacés est également relativement méconnu. Des expérimentations d'inférence du gène codant pour l'AGH réalisées chez le crabe *Cherax quadricarinatus* ont provoqué une baisse de la production de sperme associée à l'apparition du processus de vitellogénèse (Ventura *et al.* 2011). Cependant chez la crevette d'eau douce géante *Macrobrachium rosenbergii,* il a été démontré que l'ultrastructure de l'AG est stable au cours de la spermatogénèse. Ces résultats suggèrent que l'AGH est secrétée à des niveaux constants et qu'elle occupe préférentiellement une fonction de maintien des caractères sexuels plutôt que de régulation du processus de spermatogénèse (Okumura and Hara 2004). Alors que l'influence du cycle de mue sur le cycle reproductif de la femelle est reconnue, elle est en revanche beaucoup moins documentée chez les mâles, ces derniers étant considérés comme pouvant se reproduire en permanence. Chez *Macrobrachium rosenbergii*, le nombre de spermatogonies augmente en fin de stade de pré-mue, puis ces cellules évoluent sous la forme de spermatocytes au cours de la post-mue puis sous forme de spermatozoïdes pendant la phase d'intermue et en début de stade de pré-mue (Okumura and Hara 2004). Chez la crevette *Litopenaeus vannamei*, les spermatophores sont remplacés de manière cyclique, ils disparaissent 12 h avant la mue et se reforment tout de suite après. Les auteurs de l'étude suggèrent une destruction programmée strictement reliée au cycle de mue et sous possible contrôle par des hormones ecdystéroïdiennes (Parnes *et al.* 2006).

La régulation de l'**ovogénèse** est, par comparaison à celle de la spermatogénèse, plus documentée. Elle est contrôlée au niveau neuroendocrinien par la libération de neuropeptides stimulateurs (*i.e*. GSH : Gonado-Stimuling Hormone et VSOH : Vitellogenesis- Stimuling Ovarienne Hormone) ou inhibiteurs la GIH (*i.e.* Gonado-Inhibiting Hormone), également dénommée VIH (*i.e*. Vitellogenesis-Inhibiting Hormone). Cependant, la nature du contrôle par ces neuropeptides (contrôle direct ou indirect) reste méconnue. L'existence de la GIH/VIH a été postulée par des expérimentations d'ablation du complexe OX/GS chez les décapodes, entrainant une accélération de la vitellogénèse secondaire, due à une augmentation de la synthèse et de l'endocytose de la vitellogénine. Par la suite, sa nature chimique a été déterminée chez les homards *Homarus gammarus* et *americanus* ainsi que chez l'isopode *A. vulgare* (résumé par Webster *et al.* 2012). A

l'inverse, cette recherche bibliographique n'a pas permis de mettre en évidence l'existence d'études sur la caractérisation précise de la GSH. Chez les décapodes, cette dernière serait membre de la famille des CHH, sécrétée par les ganglions thoraciques et aurait un rôle dans la stimulation de la vitellogénèse (Nagaraju 2011). Chez *Orchestia gammarella,* les cellules folliculaires secondaires sont des cellules endocrines, sécrétant la VSOH qui contrôle le processus de vitellogénèse. Cependant, l'existence de ce peptide semble être une spécificité des amphipodes (Charniaux-Cotton 1965).

En addition du rôle des neuropeptides dans la régulation de l'ovogénèse, celle-ci est également régulée par les hormones terpenoïdes et ecdystéroïdiennes. Différentes études chez des décapodes ont démontrés [1] que l'OM est plus actif et sécrète plus de MF au cours du développement ovarien, [2] que l'implantation de l'OM de femelles sexuellement matures chez des femelles immatures entraine une stimulation de la croissance ovarienne, [3] que l'administration de MF *in vivo* induit la synthèse de la vitellogénine et stimule la maturation des ovocytes et [4] que le MF stimule la croissance d'ovocytes dans les ovaires cultivés *in vitro* (résumé par Nagaraju 2007).

Concernant les ecdystéroïdes, chez la crevette d'eau douce *Macrobrachium nipponense*, une corrélation entre leur taux dans l'hémolymphe et la maturation ovarienne a été mise en évidence. Un pic d' ecdystéroïdes a été observé au moment de l'initiation de la vitellogénèse suivi d'un pic plus marqué juste avant la mue et la ponte (Okumura 1992). Alors que cette corrélation positive pourrait s'expliquer par la synchronisation de l'ovogénèse et de la mue et non par une action directe de l'ecdysone sur le processus de maturation des gamètes, la présence de récepteur de l'hormone (EcR) dans l'ovaire du crabe *Uca pugilator* renforce l'hypothèse d'un contrôle direct (Durica *et al.* 2002).

La **mue** est également un élément majeur dans la reproduction des crustacés et fait intervenir de nombreux processus physiologiques : prolifération des cellules épithéliales, sécrétion de la nouvelle cuticule ou encore extraction et stockage du calcium contenu dans l'ancienne cuticule. Le contrôle de ces processus se fait majoritairement grâce à l'ecdysone (Durica *et al.* 1999) dont la teneur fluctue au cours du cycle de mue, régulant les variations biochimiques et physiologiques associées à ce processus (Chang and O'Connor 1978). De façon générale, le niveau en ecdysone est faible durant la période de post-mue et d'inter-mue alors que dans la phase de pré-mue, les concentrations dans l'hémolymphe augmentent et atteignent un pic avant

l'exuviation, déclenchant les processus initiant la mue. Cette augmentation est suivie d'une diminution rapide déclenchant l'exuviation (Mykles 2011).

En parallèle, les dosages hémolymphiques de MIH chez des décapodes montrent que la concentration de cette hormone reste constante au cours du cycle avec une forte baisse durant les premières phases de pré-mue, concomitant avec le pic d'ecdysteroids, confirmant son rôle dans l'inhibition de la synthèse d'ecdysone (Nakatsuji *et al.* 2009).

Les effets de la forme active principale la 20-HE, sont majoritairement médiatés par un hétérodimère de deux récepteurs nucléaires : l'EcR et l'USP, principalement étudié chez *Drosophila melanogaster*, où il est impliqué dans le processus de métamorphose (King-Jones and Thummel 2005). L'EcR possède trois isoformes, divergentes au niveau N-terminal et jouant des rôles distincts dans l'activation de la transcription. Par conséquent, ces isoformes sont exprimées de manière différentielle selon les tissus (tissus apoptotiques ou se différentiant en réponse à l'ecdysone). L'USP ne possède qu'une seule isoforme et ne présente pas de fonction d'activation mais facilite la liaison du complexe avec le ligand et l'ADN (Cakouros *et al.* 2004). A l'heure actuelle, l'EcR-USP a fait l'objet de peu d'études chez les crustacés, qui ont mis en évidence sa présence sur différents tissus : les branchies, l'hépatopancréas ou encore l'épiderme chez le crabe *Uca pugilator* (Durica *et al.* 1999).

A l'instar de l'ecdysone, les niveaux hémolymphatiques de MF fluctuent au cours du cycle de mue, selon le même patron que l'ecdysone. L'administration de MF chez différentes espèces de décapodes a engendré une accélération de la mue (résumé par Nagaraju 2007) alors que chez la crevette *Litopenaeus vannamei*, des expériences d'inférence de gènes impliqués dans la synthèse de MF ont entrainé un retard de mue, renforçant l'hypothèse du rôle de cette hormone dans le contrôle de la mue et la croissance (Hui *et al*, 2008).

Un récapitulatif des différentes hormones synthétisées chez les amphipodes ainsi que les effets physiologiques associés est présenté en Tableau I-5.

Tableau I-5 : Récapitulatif des hormones impliquées dans le processus reproductif chez les amphipodes et effets physiologiques associés

Nature chimique de l'hormone	Hormone	Site de production	Organe cible	Effet physiologique
Neuropeptide	Molt-inhibiting hormone (MIH)	Organe X / Glande du sinus	Organe Y	Contrôle (-) de la synthèse d'ecdysone
	Mandibular Organ-Inhibiting Hormone (MOIH)	Organe X / Glande du sinus	Organe Mandibulaire	Contrôle (-) de la synthèse de méthyl-farnesoate
	Gonado-Stimuling Hormone (GSH)	Protocérébron	Gonades	Contrôle (-) du développement des gonades
	Gonado-Inhibiting Hormone / Vitellogenesis-Inhibiting Hormone (GIH/VIH)	Organe X / Glande du sinus	Gonades	Contrôle (+) du développement des gonades
	Vitellogenesis- Stimuling Ovarienne Hormone (VSOH)	Ovaires	Ovaires	Contrôle (+) de la vitellogénèse
Stéroïdienne	Ecdysone	Organe Y	Ovaires Organe X / Glande du sinus Testicules ? Autres tissus	Contrôle de la croissance Contrôle de la mue Régulation de l'ovogénèse Régulation de la spermatogénèse ?
Terpenoïde	Méthyl-farnesoate (MF)	Organe Mandibulaire	Gonades Autres tissus	Différenciation sexuelle Maintien des caractères sexuels primaires et secondaires masculins Régulation de la mue Régulation (+) de la vitellogénèse
Glycopeptide	Hormone androgène (AGH)	Glande androgène	Gonades	Différenciation sexuelle Maintien des caractères sexuels primaires et secondaires masculins

2.2. Perturbations de la fonction reproductrice chez les crustacés

L'étude des perturbations de la reproduction chez les crustacés a fait l'objet de nombreux articles et revues ces dernières années (LeBlanc 2007, Rodríguez *et al.* 2007, Mazurová *et al.* 2008, Hyne 2011). Ces études s'articulent majoritairement autour de la notion d'évaluation du risque par une approche pronostique en laboratoire concernant : [1] les composés reconnus comme étant PE chez les vertébrés et [2] les composés spécifiquement conçus pour interférer avec la croissance et la reproduction des arthropodes (insecticides régulateurs de croissance), les crustacés étant des espèces non-cibles. La littérature concernant l'étude d'altérations de la reproduction en contexte naturel et l'identification des mécanismes associés est en revanche beaucoup plus rare.

2.2.1. Exemples d'altération de la fonction reproductrice chez des populations naturelles

Le maintien d'une espèce dans son milieu dépend en partie de sa capacité à produire une descendance viable. C'est pourquoi, la mesure de la fécondité permet d'appréhender au plus proche l'effet sur la dynamique de population. Cependant, elle reste peu étudiée en milieu naturel. Chez l'amphipode marin *Echniogammarus marinus*, Ford et *al.* (2003) ont montré que la taille moyenne de la poche ventrale, décroit selon un gradient croissant de contamination. De même, chez *Gammarus fossarum*, l'implantation de gammares encagés sur un site français historiquement contaminé en cadmium a conduit à une réduction de 58 % la fertilité (nombre moyen d'embryons dans le marsupium) par rapport au site de référence (Coulaud *et al.* 2014).

Les gamètes possèdent la fonction unique de propager l'ADN parental pour le développement de la future génération. Par conséquent, la production de gamètes de bonne qualité (nombre et potentiel de fertilisation) est vitale. Chez *Echniogammarus marinus*, la collecte d'organismes sur un site impacté par des activités industrielles a mis en évidence une baisse de 20 % du nombre de spermatozoïdes par rapport aux sites de référence. Les travaux de Lacaze portant sur le développement d'un marqueur de génotoxicité chez *Gammarus fossarum* en laboratoire et dans un contexte d'encagement d'organismes *in situ*, ont démontré que par rapport aux ovocytes et aux hémocytes, les spermatozoïdes sont particulièrement sensibles aux dommages génotoxiques (Lacaze *et al.* 2010, Lacaze *et al.* 2011) . De plus, ces dommages ont des conséquences au niveau de la dynamique de population, une corrélation entre le niveau de génotoxicité des spermatozoïdes et les anomalies de développement embryonnaire ayant été mise en avant (Lacaze *et al.* 2011).

Au sein de la littérature, un des phénomènes les plus documentés d'altération de la fonction reproductive chez les crustacés porte sur l'apparition de phénomènes d'intersexualité. L'intersexualité se caractérise par la possession de caractères sexuels mâles et femelles et a été documentée chez de nombreuses espèces (LeBlanc 2007) . Bien que ce phénomène soit courant chez les crustacés gonochoriques, sa fréquence augmente au sein de sites contaminés. Par exemple, chez *Echniogammarus marinus*, la proportion d'individus intersexués est de 30 % sur un site impacté par des activités industrielles alors qu'elle est de 6 % dans le site de référence (Yang *et al.* 2008). De même, chez *Gammarus fossarum*, une étude de deux ans sur différentes rivières (5 sites) en Allemagne a mis en évidence l'existence de phénomènes d'intersexualité chez cette espèce. Trois sites en particulier présentaient des taux d'intersexualité compris entre de 7 et 24 %, alors que deux autres présentaient des taux inférieur à 1 %. Bien que la relation entre le taux d'intersexualité et la contamination des milieux n'a pu être établie, une implantation d'organismes originaires des sites peu impactés sur les sites fortement impactés a entrainé une augmentation de la fréquence d'intersexualité au cours du temps, suggérant l'influence d'une pression anthropique non identifiée (Jungmann *et al.* 2004).

2.2.2. Origine de la perturbation de la fonction reproductrice

Les perturbations de la fonction reproductrice ont des origines multiples, de nature chimique ou biologique, qu'il convient de différencier: [1] internalisation de molécules ciblant spécifiquement le système endocrinien (perturbateurs endocriniens), [2] internalisation de molécules provoquant des effets toxiques non cibles et [3] facteurs environnementaux abiotiques (photopériode, disponibilité alimentaire) ou biotiques (parasites, densité de population).

Les **perturbateurs endocriniens** ont la capacité de moduler la régulation hormonale selon quatre mode d'action, en :

- Inhibant ou stimulant la synthèse, la libération et/ou le transport des hormones
- Modulant le métabolisme des hormones afin d'augmenter ou de diminuer leur élimination
- Interférant avec les récepteurs hormonaux via une action agoniste ou antagoniste
- Modifiant la synthèse des récepteurs hormonaux

Les composés spécifiquement conçus pour interférer avec la croissance et la reproduction des arthropodes en tant qu'insecticides régulateurs de croissance ciblent directement la régulation des hormones ecdystéroïdiennes et terpenoïdes et peuvent être par conséquent classifiés comme PE chez les arthropodes. Parmi les pesticides les plus répandus, on retrouve principalement les insecticides analogues aux hormones juvénoïdes, capables de perturber le développement embryonnaire et d'interférer avec la métamorphose. Différentes études reportent chez les crustacés, les effets de ces insecticides sur le développement embryo-larvaire et/ou la mue (daphnie : Olmstead and Leblanc 2002, crevette et crabe : Tuberty and McKenney Jr 2005, myside : Ghekiere *et al.* 2006). De même, les insecticides analogues aux hormones juvénoïdes peuvent perturber la reproduction chez les femelles (crabe : Lee and Noone 1995, myside : Ghekiere *et al.* 2006, daphnie : Tokishita *et al.* 2006). Enfin, chez la daphnie, ces composés peuvent altérer le sexe-ratio de la progéniture en augmentant la production de mâles (Borst *et al.* 1987, Olmstead and LeBlanc 2003).

En comparaison des insecticides analogues aux hormones juvénoïdes, les insecticides mimétiques des ecdysteroids ont reçu moins d'attention de la part de la communauté scientifique. Ces derniers se présentent sous la forme d'antagoniste de la 20-HE, qui bloquent l'accès au récepteur EcR et empêchent le déclanchement du processus de métamorphose chez les insectes (LeBlanc 2007). Les effets de ces molécules ont principalement été étudiés chez les espèces cibles. Cependant, chez les larves du crabe *Rhithropanopeus harrisii*, Clare et *al*. (1992) ont montré que le tébufénozide provoque une accélération de la mue alors que chez la crevette *Artemia* et la daphnie, cette molécule est associée à une mortalité au cours de la mue ou post-mue avec présence de l'ancienne exuvie partiellement attachée à la nouvelle cuticule (Song *et al.* 1997).

A l'heure actuelle, la recherche sur la perturbation endocrine en écotoxicologie se concentre essentiellement sur le mode d'action génomique *via* l'interaction avec les récepteurs nucléaires (Figure I-16). Chez les crustacés, c'est les mimétiques des hormones terpenoïdes qui sont principalement étudiés.

Des nombreux travaux ont été consacrés à l'étude de l'impact des PE avérés chez les vertébrés tels que les hormones naturelles ou de synthèse, les hydrocarbures aromatiques polycycliques ou les surfactants. Parmi eux, les composés oestrogéniques ont fait l'objet d'une attention particulière, avec des résultats contradictoires. Par exemple, chez les amphipodes *G. pulex* exposés durant 100 jours à différentes concentrations de 17α-ethinylestrediol, Watts *et al.*

(2002) ont observé une augmentation du nombre de femelles au sein de la population mais pas d'effet sur la formation de précopulât ou sur le nombre de mâles alors que chez des mâles *Hyallela azteca* exposés aux mêmes concentrations pendant 10 jours, des aberrations tissulaires au niveau du tractus reproductif (intersexualité, perturbation de la spermatogénèse) ont été mises en évidence (Vandenbergh *et al.* 2003). Par ailleurs, chez le copépode *Nitracra spinipes*, l'exposition à différents composés oestrogéniques (7-β-oestradiol et 17-α-ethynylestradiol) n'a pas provoqué d'effets sur la reproduction (fécondité, sexe-ratio) ni sur le développement (Breitholtz and Bengtsson 2001). Chez les vertébrés, la régulation de la reproduction repose principalement sur l'action de la testostérone et de l'œstrogène, alors que chez les invertébrés, leurs implications sont controversées. En effet, la présence d'une activité de métabolisation de ces molécules ainsi que des récepteurs associés peuvent suggérer leur implication dans cette fonction (James and Boyle 1998, Köhler *et al.* 2007). Néanmoins à l'heure actuelle, il n'existe pas de preuves d'une synthèse endogène de ces hormones, rejetant l'hypothèse d'une implication dans la réponse endocrine. Scott (2012) suggère que chez les invertébrés, la présence d'une activité de biotransformation et de récepteurs « vertebrostéroïdes » serait en fait liée à une réponse adaptative de détoxification face à ces substances exogènes ubiquitaires dans l'environnement car sécrétées en grande quantité par les vertébrés.

Outre la perturbation endocrine, une large gamme d'agents chimiques est capable d'impacter la fonction reproductrice en provoquant des effets délétères au niveau des macromolécules biologiques, du fonctionnement cellulaire ou encore de l'intégrité tissulaire. Par exemple, de nombreux xénobiotiques (alkylphénols, nonylphénol, bisphénol A, DTT, Cadmium) sont capables de perturber l'homéostasie cellulaire du calcium, en inhibant les transporteurs cellulaires (Ca^{2+}ATPase) (revue par Kirk *et al.* 2003). Or, il a été démontré chez le rat, que ces transporteurs ioniques sont abondants dans les cellules de Sertoli, cellules qui contrôlent la spermatogénèse. L'exposition de ces cellules aux alkylphenols entraine une perturbation des concentrations intracellulaires en calcium conduisant à des phénomènes d'apoptose (Hughes *et al.* 2000).

De manière générale, dans le cas de l'augmentation de la charge interne en toxiques, des mécanismes de compensation se mettent en place afin de préserver l'organisme. Selon l'hypothèse du coût métabolique, l'organisme dispose d'un budget énergétique, utilisé pour son métabolisme, sa croissance et sa reproduction (Figure I-20).

Dans le cas de la mise en place des mécanismes de compensation, une augmentation de la demande énergétique afin de maintenir le métabolisme basal est réalisée au détriment de la croissance et de la reproduction (réallocation énergétique) (Coen *et al.* 1997).

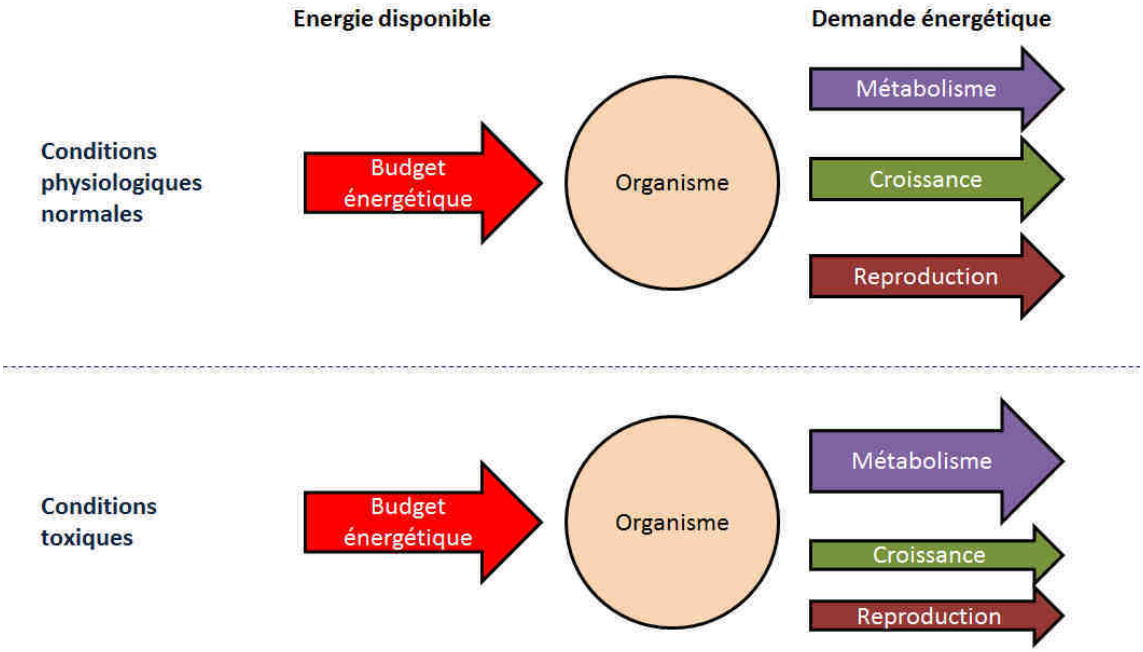


Figure I-20 : Représentation schématique de l'hypothèse du coût métabolique (adapté selon Coen *et al.* 1997)

Cette hypothèse du coût métabolique est également applicable pour tous changements environnementaux impactant le budget énergétique de l'organisme tels que la privation alimentaire. Geffard et *al.* (Geffard *et al.* 2010) ont montré que chez *Gammarus fossarum*, une réduction de 40 % de la ressource alimentaire entraine une perturbation globale de la reproduction chez la femelle par une baisse du nombre d'ovocytes produits, un retard de mue et une possible inhibition de la vitellogénèse.

Certains **facteurs environnementaux** impactent également de manière spécifique la régulation de la fonction reproductrice. Chez la daphnie, deux voies de détermination sexuelle existent, une génétique et une autre environnementale. Cette dernière est initiée par différents signaux : température, photopériode ou densité de population (Kato *et al.* 2011). De même, les phénomènes d'intersexualité chez les amphipodes sont en partie imputables à la présence de parasites féminisants (Ford 2008). Comme la majorité des espèces d'arthropodes, les tissus reproducteurs

des gammares peuvent être infectés par des parasites de type microsporidie à transmission verticale. Les hôtes mâles, ne pouvant transmettre le parasite à leurs descendances, sont considérés comme des impasses en termes de transmission. Par conséquent, les microsporidies ont développé des stratégies afin d'augmenter la proportion de femelles dans la population tels que l'avortement d'embryons mâles ou encore la conversion de mâles en femelles. De cette conversion résulte les phénomènes d'intersexualité, qui ont été documentés chez *G. roeseli* (Haine *et al.* 2004) et *G. duebeni* (Rodgers-Gray *et al.* 2004). Chez ce dernier, il a été démontré que le parasite *Nosema granulosis* manipule le sexe de l'hôte en prévenant la différenciation de la GA et de la synthèse d'AGH. Bien que la présence de parasites féminisants dans les populations d'amphipodes soit un phénomène courant, les pressions anthropiques peuvent provoquer une immunodépression de l'hôte favorisant ainsi la prolifération et la transmission des parasites (Ford 2012).

2.3. Outils pour l'évaluation de la reprotoxicité chez les crustacés

La majorité des travaux visant à évaluer l'impact des PE est réalisée en laboratoire et se focalise principalement sur les réponses au niveau individuel, le test de toxicité chronique chez *Daphnia magna* étant particulièrement employé (*i.e*. mesure du sexe ratio et de la fécondité). Il est donc difficile, en se basant sur ces observations, de décrire précisément les mécanismes de toxicité impliqués. A l'heure actuelle, il n'existe que très peu d'outils capables de statuer sur la capacité d'une substance à interagir spécifiquement avec le système endocrinien. Dans le cadre d'études *in-situ*, ce manque se traduit notamment par une impossibilité d'identifier les facteurs responsables de la perturbation de la reproduction *i.e*. présence de PE, présence de parasites, manifestation d'une toxicité générale. Par exemple, Jungmann *et al.* (2004) ont mis en évidence l'existence de phénomènes d'intersexualité chez différentes populations naturelles de *Gammarus fossarum* en Allemagne. Une population en particulier, située dans une réserve naturelle dépourvue de rejets d'effluents, présentait un taux d'individus intersexués de 24 % mais aucun lien n'a pu être établi entre les xénobiotiques détectés et l'occurrence de ce phénomène.

Cependant, certaines réponses individuelles peuvent être pertinentes pour l'évaluation du mode d'action des xénobiotiques. A partir de la description détaillée du cycle reproductif de la femelle *Gammarus fossarum*, Geffard et *al.* (2010) ont développé un test de reprotoxicité sur 21 jours chez cette espèce. Alors que l'impact des polluants peut être évalué par comparaison avec les

organismes témoins, une explication de leur mode d'action peut être avancée par la désynchronisation des processus de mue, de vitellogénèse et de développement embryonnaire. Par exemple, chez les organismes soumis à une privation alimentaire, une inhibition globale de ces différents processus a été observée, suggérant un effet indirect tel qu'une réallocation énergétique. En revanche, chez les organismes exposés au nonylphénol, seule une augmentation du nombre d'anomalies embryonnaires a été détectée, la mue et la vitellogénèse étant non perturbées, suggérant un effet spécifique de ce composé sur le développement embryonnaire.

Le dosage de la synthèse et/ou des concentrations en hormones circulantes, la quantification des récepteurs spécifiques ou des protéines impliqués dans le métabolisme des hormones constituent des biomarqueurs de choix pour les approches d'évaluations du risque en laboratoire et de biosurveillance des écosystèmes aquatiques. Cependant, les méthodes analytiques existantes sont principalement employées chez les décapodes, crustacés à forte valeur économique. A l'opposé, chez les espèces d'intérêt en écotoxicologie (amphipodes, copépodes ou mysidacés), l'utilisation de ces biomarqueurs se conforte à différentes limites [1] analytique, la quantité d'échantillon étant extrêmement réduite et [2] biologique en raison du manque de connaissances fondamentales sur le fonctionnement de leur système endocrinien. Toutefois, quelques biomarqueurs ont été proposés, permettant de mieux appréhender les mécanismes mis en œuvre dans la perturbation de la reproduction.

Chez les crustacés, la mue peut être suivie au niveau biochimique par la mesure de l'activité d'enzymes impliquées dans la dégradation de la chitine, composante majeure de l'exosquelette (Spindler-Barth *et al.* 1990, Zou and Fingerman 1999, Knotz *et al.* 2006, Duchet *et al.* 2011, Mesquita *et al.* 2011). La chitine, polymère de N-acetyl-β-glucosamine, est dans un premier temps dégradée par la chitinase en oligomères, qui sont à leur tour hydrolysés en monomères par la N-acetyl-β-glucosaminidase (chitobiase) (Jeuniaux 1963). L'activité de la **chitobiase** peut être suivie pour évaluer l'intégrité du cycle de mue face à une pression toxique. Elle a été suivie chez deux espèces de crustacés : le crabe épi benthique *Carcinus maenas* en réponse à la fluoxetine (antidépresseur) et le crabe *Uca pugilator* en réponse à différents composés œstrogèno-mimétique (Aroclor 1242, diethylstilbestrol, endosulfan, et 4-octylphenol). Alors que chez *C. maenas*, aucune perturbation de l'activité de la chitobiase n'a été mesurée, chez *U. pugilator*, cette dernière a été inhibée en réponse à l'endosulfan, l'Aroclor 1242 et au diethylstilbestrol.

Par analogie avec les travaux réalisés chez les poissons, le suivi de la synthèse de la **vitellogénine** en tant que biomarqueur de la perturbation endocrinienne chez les crustacés a été proposé par de nombreux auteurs (revue par Matozzo *et al.* 2008) et également chez *Gammarus fossarum* (Simon *et al.* 2010, Xuereb *et al.* 2011, Jubeaux *et al.* 2012, Jubeaux *et al.* 2012). Le suivi de son induction et de la production de la Vtg chez les poissons mâles est employé comme biomarqueur de la perturbation endocrinienne en réponse à l'exposition à des composés œstrogèno-mimétique (Kime *et al.* 1999, Matthiessen 2003). Mais alors que ce biomarqueur s'est montré robuste et spécifique chez le poisson, les études s'intéressant à sa modulation chez les invertébrés donnent des résultats plus contradictoires (résumé par Matozzo *et al.* 2008).

Chez les crustacés, des altérations de la Vtg ont été documentées suite à des expositions aux xéno-œstrogènes (Ghekiere *et al.* 2006, Xuereb *et al.* 2011), à des fongicides (Soetaert *et al.* 2006, Jubeaux *et al.* 2012) ou des métaux lourds (Martin-Diaz *et al.* 2005, Vandenbrouck *et al.* 2009). La majorité des études s'est intéressée à la modulation de la Vtg en réponse à une exposition à des composés œstrogèno-mimétiques afin de la proposer comme biomarqueur d'exposition à ces substances (Matozzo *et al.* 2008). Cependant, comme mentionné en I-2.2.2, il existe une grande divergence entre le système endocrinien vertébré et invertébré, et les PE reconnus chez les vertébrés n'ont pas nécessairement d'interaction directe avec le système endocrinien des invertébrés. Au regard de la bibliographie, peu d'études ont porté sur l'effet des PE reconnus chez les arthropodes sur la modulation de la Vtg (Tokishita *et al.* 2006, Hannas *et al.* 2011, Jubeaux *et al.* 2012). Or, chez la Daphnie, dans la région intergénique de la Vtg, la présence de séquences *cis* réactives à l'ecdysone et aux hormones juvénoïdes, a été mise en évidence, suggérant un contrôle de son expression par ces hormones (Tokishita *et al.* 2006).

Aussi, nombre des études réalisées sur les crustacés ont utilisé des organismes femelles (Ghekiere *et al.* 2005, Martin-Diaz *et al.* 2005, Ghekiere *et al.* 2006) ou n'ont pas distingué le sexe des individus (Gagné and Blaise 2004, Gagné *et al.* 2005). Or, en raison des variations de la teneur de cette protéine au cours du cycle, il est essentiel de distinguer le statut reproductif de l'organisme afin de clairement discriminer les variations naturelles, de celles induites par la présence d'un toxique (voir I-1.3.2.2), distinction qui pourtant rarement réalisée dans la littérature scientifique. Peu d'études ont employé des organismes mâles alors que la Vtg est principalement mesurée chez les poissons mâles. Les seules études trouvées chez les crustacés mâles portent sur le décapode *Palaemon elegans* (Sanders *et al.* 2005), les amphipodes *G. fossarum* (Xuereb *et al.* 2011, Jubeaux

et al. 2012, Jubeaux *et al.* 2012) et *E. marinus* (Short *et al.* 2014). Toutefois, une étude *in-situ* à large échelle avec des organismes *G. fossarum* mâles encagés n'a pas permis de conclure de son intérêt en tant que biomarqueur d'exposition aux PE. En effet, une grande variabilité interindividuelle a été mesurée aussi bien au sein des sites de référence que dans les sites contaminés et sur seulement 2 stations, une faible induction de la Vtg a été mesurée, inductions ne pouvant être imputées à une contamination spécifique. L'étude de la structure du gène chez la Daphnie a mis en évidence la présence, dans la région intra-génique, d'un domaine SOD en position NH_2-terminale, suggérant une potentielle activité anti-oxydative (Kato *et al.* 2004, Tokishita *et al.* 2006). Cette activité a été confirmée chez l'abeille mellifère *Apis melifera* où la synthèse de Vtg chez les ouvrières les protège du stress oxydatif, allongeant ainsi leur durée de vie (Seehuus *et al.* 2006). Ces observations pourraient expliquer l'induction de la Vtg face à un large spectre de polluants, le stress oxydatif étant une manifestation de l'effet toxique de nombreuses molécules. En raison de ses fonctions pléiotropes chez les arthropodes, la Vtg ne serait pas un bon candidat pour révéler spécifiquement l'exposition aux PE chez les crustacés (Jubeaux *et al.* 2012).

L'exemple de la vitellogénine illustre la nécessité en écotoxicologie, de stopper la transposition des biomarqueurs développés chez les vertébrés aux invertébrés et de définir des biomarqueurs de perturbation spécifiques du modèle d'étude.

3. Apports des nouveaux outils omiques en écotoxicologie

3.1. La toxicologie environnementale à l'ère des omiques

Les technologies dites « omiques » permettent la caractérisation ainsi que la quantification de l'ensemble des biomolécules d'une même famille, impliquées dans la structure et le fonctionnement d'un organisme. A l'origine le terme « génome» a été créé par Hans Winkler (1920), se référant à l'ensemble des chromosomes et aux gènes associés. Par la suite, McKusick and Ruddle (1987) ont enrichi la terminologie en cofondant le journal « genomics » *e.g.* : génomique, portant sur le séquençage de l'ADN et l'identification des gènes. Ce concept s'est par la suite étendu aux autres biomolécules exprimées à partir des gènes (Figure I-21), donnant naissance à la transcriptomique portant sur l'analyse des ARN matures (ARNm) et à la protéomique, portant sur l'analyse de la synthèse des protéines.

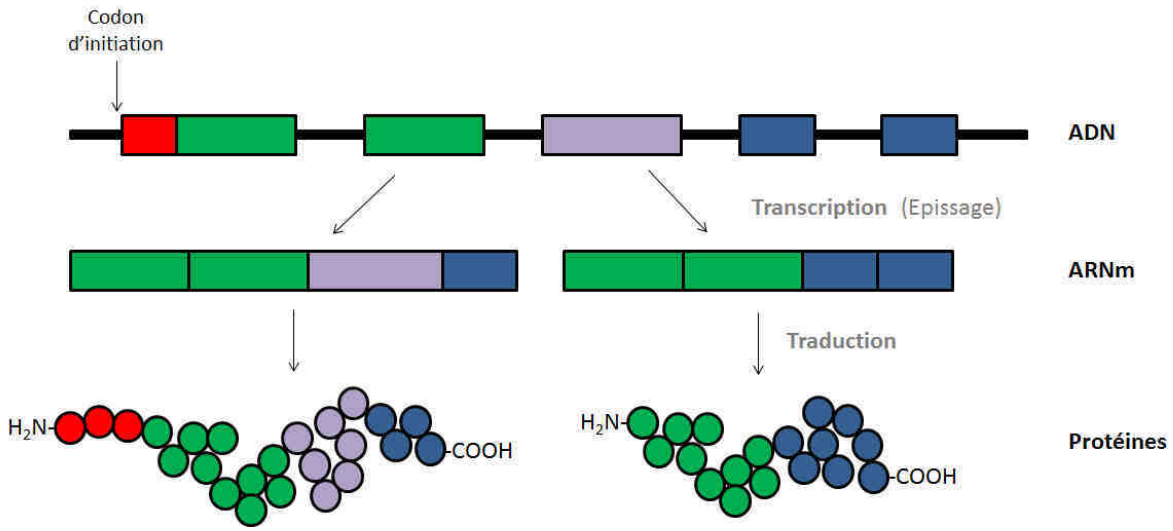


Figure I-21 : Du gène à la protéine : synthèse de protéines à partir d'un gène codant chez un eucaryote (adapté d'après Nakazawa *et al.* 2008)

Les deux protéines sont exprimées suivant différents mécanismes de transcription. Cette illustration ne prend pas en compte les étapes de maturation après leur traduction (modifications post-traductionnelle).

En toxicologie environnementale, l'analyse à large échelle des gènes et de ses produits, est employée pour étudier la réponse d'un organisme à un stress, afin de comprendre les mécanismes de toxicité et les effets néfastes associés.

3.1.1. Génomique et transcriptomique

Le séquençage de l'information génétique, par réplication de l'ADN à l'aide d'ADN polymérase, a été rendu possible par les travaux de Sanger *et al.* (1977). En 1995, le premier **génome** complet séquencé est celui de la bactérie *Haemophilus influenzae* Rd KW20 (Fleischmann *et al.* 1995) et ce sont chez les bactéries que l'on retrouve le plus grand nombre de génomes séquencés. En effet, face aux eucaryotes, leur génome est de taille réduite et ne présente pas de séquences non codante (intron), facilitant la mise en œuvre des projets de séquençage. Chez les eucaryotes, l'information génétique a été décryptée dans un premier temps sous la forme de «expressed sequence tag» (EST) représentant seulement la partie codante de l'ADN puis par la suite, le génome dans son intégralité a pu être obtenu (Figure I-22) (Lander *et al.* 2001).

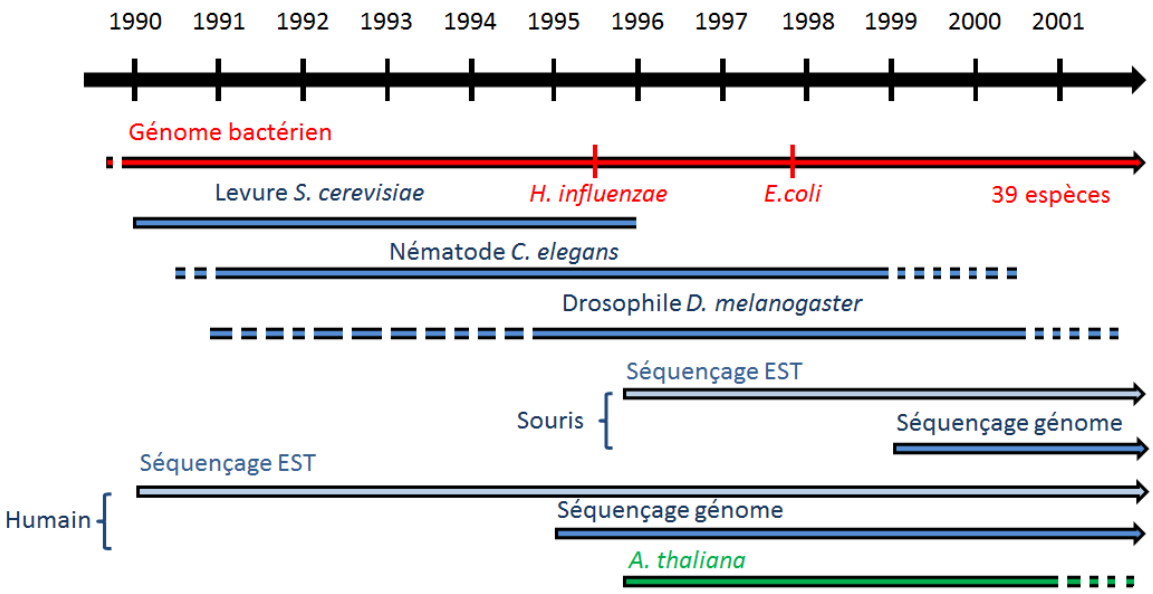


Figure I-22 : Chronologie des premiers projets de séquençage à large échelle du génome (modifié d'après Lander *et al.* 2001)

En 2005, de nouvelles techniques de séquençage (NGS : Next Generation Sequencing) ont fait leur apparition, caractérisées par une augmentation du débit et de la qualité de séquençage (Stranneheim and Lundeberg 2012). A l'heure actuelle, le génome complet de 177 organismes eucaryotes est disponible (août 2014, www.ebi.ac.uk/genomes/index.html). Parmi les arthropodes, on retrouve une dizaine de génomes, majoritairement ceux d'espèces d'intérêt en agronomie, dont

celui de la drosophile *Drosphila melanogaster*, l'abeille mellifère *Apis melifera* ou encore celui du coléoptère *Tribolium castaneum*.

Cependant, bien que les techniques NGS permettent le séquençage massif et en parallèle de fragments d'ADN de courte longueur, leur assemblage afin de reconstituer le génome reste problématique et constitue un frein dans les projets de séquençage de génome. Par conséquent, à l'heure actuelle, seules deux espèces d'intérêt en écotoxicologie aquatique ont leur génome séquencé : le poisson zèbre *Danio rerio* (Howe *et al.* 2013) et la daphnie *Daphnia pulex* (Colbourne *et al.* 2011). Lors de la publication du génome de la daphnie, les auteurs se sont principalement intéressés à sa structure et à son histoire évolutive *e.g*. phénomène de duplication, divergence face aux autres arthropodes, démontrant que un tiers de gènes prédits ne possèdent pas d'homologues au sein du vivant (gènes orphelins). La publication intégrait également le séquençage d'EST chez des daphnies mâles et femelles et chez des daphnies femelles soumises à différents challenges biotique (kairomone de prédateur *Chaoborus*) et toxique (Cadmium). Cette démarche a permis de mettre en évidence, pour chaque transcriptome, qu'en moyenne 80 % de ses transcrits étaient exprimés de manière spécifique. De même, une surreprésentation des transcrits associés aux gènes orphelins a pu y être observée (Colbourne *et al.* 2011).

Les avancées technologiques dans le domaine du séquençage des acides nucléiques peuvent également être exploitées pour l'étude du **transcriptome**. Cette technologie, dénommé « RNA-sequencing » (RNA-Seq), permet l'étude des éléments fonctionnels du génome en fonction de l'état physiologique ou en réponse à un stimulus donné. Alors qu'auparavant l'étude du transcriptome était basée sur l'hybridation des biomolécules sur un catalogue restreint des cibles (puce à ADN) ou sur la génération d'EST, les techniques RNA-Seq permettent d'identifier sans *a priori* les transcrits présents (ARNm, ARN non-codants ou de petite taille) et de les quantifier suivant une gamme dynamique plus large (Wang *et al.* 2009). A l'instar du génome, le séquençage des acides nucléiques est réalisé sur des fragments courts et par conséquent, il est nécessaire de réaliser une étape d'assemblage. Cet assemblage est facilité par la présence d'un génome de référence de la même espèce ou d'une espèce phylogénétiquement proche. Or, pour la majorité des organismes employés en écotoxicologie, ce génome de référence est inexistant. Toutefois, l'assemblage *de novo* est possible en raison d'une plus large couverture du transcriptome et de l'importante profondeur de séquençage conférées par les avancées techniques NGS (Ekblom and Galindo 2011). En effet, ces deux caractéristiques permettent d'augmenter les zones de chevauchement entre les

différents fragments séquencés afin de créer une séquence plus grande appelée « *contig* », comme présenté en Figure I-23.

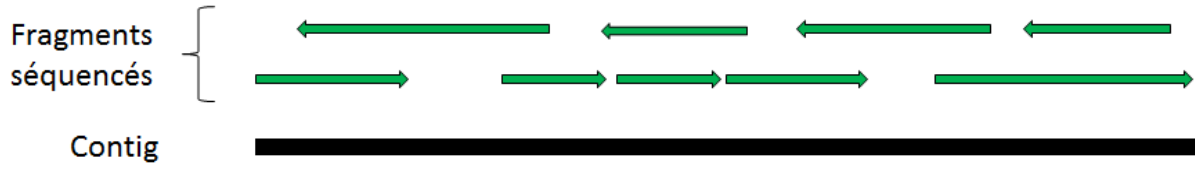


Figure I-23 : Principe de l'assemblage *de novo* du transcriptome

La technologie RNA-Seq a été employée chez plusieurs espèces d'intérêt en écotoxicologie, pour l'étude des ARNm : chez la lymnée *Lymnaea stagnalis* en réponse à un pesticide, le diquat (Bouetard *et al.* 2012), chez l'amphipode estuarien *Melita plumulosa* sur des organismes exposés à des sédiments contaminés (Hook *et al.* 2014), chez l'algue *Elodea nuttallii* en réponse au mercure et au cadmium (Regier *et al.* 2013) ou encore chez le chironome *Chironomus riparius* à différents stades de développement et exposés à différents xénobiotiques (Marinković *et al.* 2012). Cependant, les ARNm sont les intermédiaires aux processus de synthèse des protéines et la modification de leur expression ne se traduit pas nécessairement par un changement dans l'expression des protéines, ces dernières reflétant davantage le statut fonctionnel de la cellule ou de l'organisme.

3.1.2. Protéomique

Les protéines jouent un rôle primordial dans le fonctionnement cellulaire, ces dernières participent à l'ensemble des processus biologiques et présentant des propriétés variées (Figure I-24). Elles font également le lien entre le code génétique et la physiologie de l'organisme. Parmi les fonctions biologiques remplies par les protéines, on retrouve des fonctions structurelles (*e.g*. protéines du cytosquelette), motrices (*e.g*. contraction musculaire, mouvement des flagelles), de transport (*e.g*. hémoglobine pour la fixation de l'oxygène) ; réceptrices et régulatrices (*e.g*. récepteurs hormonaux) ou encore de catalyse (*e.g*. enzymes). Par conséquent, la **protéomique** est une source d'information de premier choix pour appréhender le fonctionnement des systèmes biologiques.

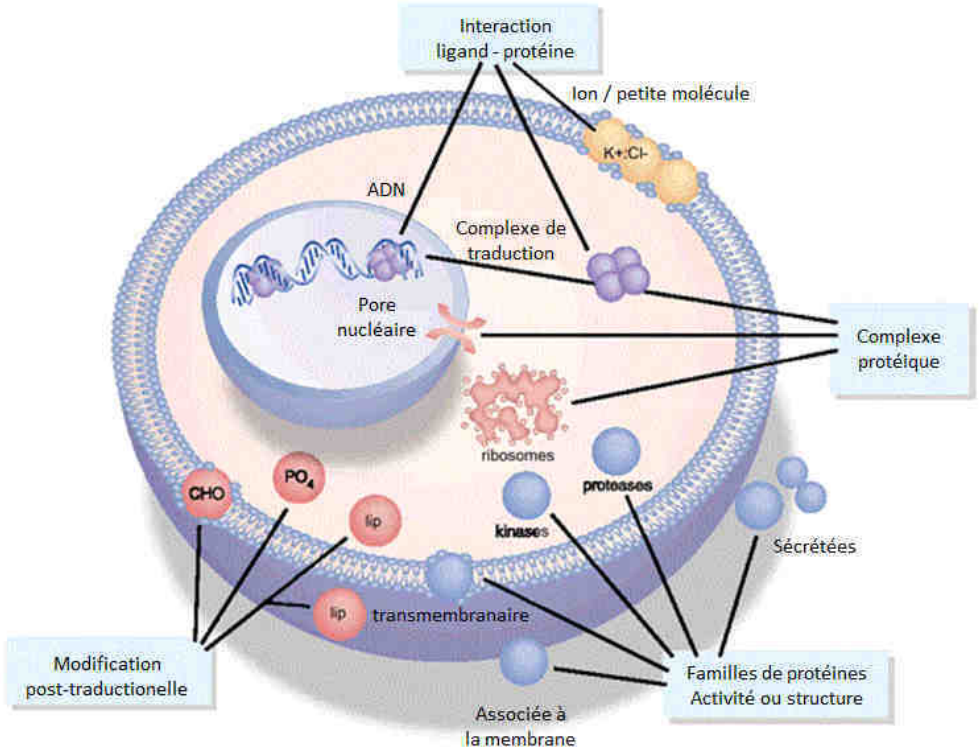


Figure I-24 : Représentation des propriétés diverses des protéines chez une cellule eucaryote (adapté de Patterson and Aebersold 2003)

La section d'une cellule eucaryote est présentée, mettant en lumière différents exemple représentatifs des fonctions et propriétés des protéines. Ceci implique différentes localisation subcellulaire, les interactions avec les biomolécules d'autres natures et les modifications post-traductionnelles avec les carbohydrates (CHO), phosphate (PO_4) et lipide (lip).

L'essor de la protéomique au début des années 1990 est la conséquence directe des avancées technologiques obtenues par les projets de séquençage de génome mais également en chimie, physique et informatique. D'après Patterson and Aebersold (2003), elle tire son émergence des avancées en :

- techniques séparatives de mélanges complexes protéiques ou peptidiques (électrophorèse et chromatographie)
- spectrométrie de masse (MS) appliquée aux protéines et peptides
- séquençage de l'information en acides nucléiques pour la constitution de bases de données (BD) de séquences nucléiques et protéiques
- mathématiques par le développement d'algorithmes corrélant les données générées par la spectrométrie de masse avec les BD protéiques
- informatique par l'augmentation de la puissance de calcul informatique

L'objectif initial de la protéomique était l'identification de toutes les protéines exprimées au sein d'une cellule, tissu ou d'un organisme. Par la suite, la protéomique a évolué vers une gamme d'applications plus large pour la détermination de diverses propriétés de la protéine : séquence (structure primaire), quantité, modifications post-traductionnelles, interactions avec d'autres protéines ou encore localisation subcellulaire (Patterson and Aebersold 2003). Actuellement, en écotoxicologie, la protéomique est majoritairement employée pour identifier les protéines sur-exprimées ou sous-exprimées selon les différentes conditions. Son utilisation dans ce domaine ainsi que les différentes étapes de l'analyse protéomique seront développées dans les parties I-3.2 et I-3.3.

3.1.3. Vers une approche multi-omiques

«*In the wonderland of complete sequences, there is much that genomics cannot do, and so the future belongs to proteomics: the analysis of complete complements of proteins*» (Fields 2001).

Alors que les avancées des technologies NGS ont permis l'obtention des génomes complets, leur valeur biologique ne peut être pleinement estimée sans l'identification des gènes codant pour les protéines (annotation structurale) ainsi que leur fonction biologique (annotation fonctionnelle). Chez les eucaryotes, la portion du génome codant pour les protéines est estimée inférieure à 25 % (Ansong *et al.* 2008). L'annotation structurale est réalisée de manière automatisée par des outils informatiques et les séquences codantes sont identifiées selon leur composition en bases, la présence de codon d'initiation ou encore la recherche de séquences homologues précédemment annotées chez d'autres organismes. Ces programmes permettent donc une prédiction des gènes et peuvent engendrer des erreurs notables (Reeves *et al.* 2009). L'intégration de données expérimentales permet donc d'améliorer ce procédé, par l'assimilation des données transcriptomiques obtenues par RNA-Seq (Li *et al.* 2011) ou des données protéomiques (Armengaud 2009).

Dès 1995, Yates *et al.* (1995) proposa de corréler les spectres de fragmentation MS/MS des peptides directement avec les bases de nucléotides générées à partir d'EST et traduites selon les six cadres ouverts de lecture (ORF : Open Reading Frame), afin de confirmer et de corriger l'annotation des gènes préexistante. Cette stratégie a également l'avantage d'améliorer le taux d'identification

des protéines, les données du transcriptome permettant de mieux estimer les potentialités en termes de production de protéines (Wang *et al.* 2012). Par la suite, avec la multiplication des génomes complétés, Jaffe *et al.* (2004) ont proposé de placer directement sur le génome, les peptides séquencés afin de produire une « cartographie protéogénomique » (Figure I-25). Aujourd'hui, la **protéogénomique** est la méthode de choix pour améliorer l'annotation de nouveaux génomes, elle permet [1] la validation des gènes prédits, [2] la correction des sites prédits d'initiation et de terminaison de la transcription (codon start et stop) et des ORF et [3] l'identification de gènes non prédits (Ansong *et al.* 2008).

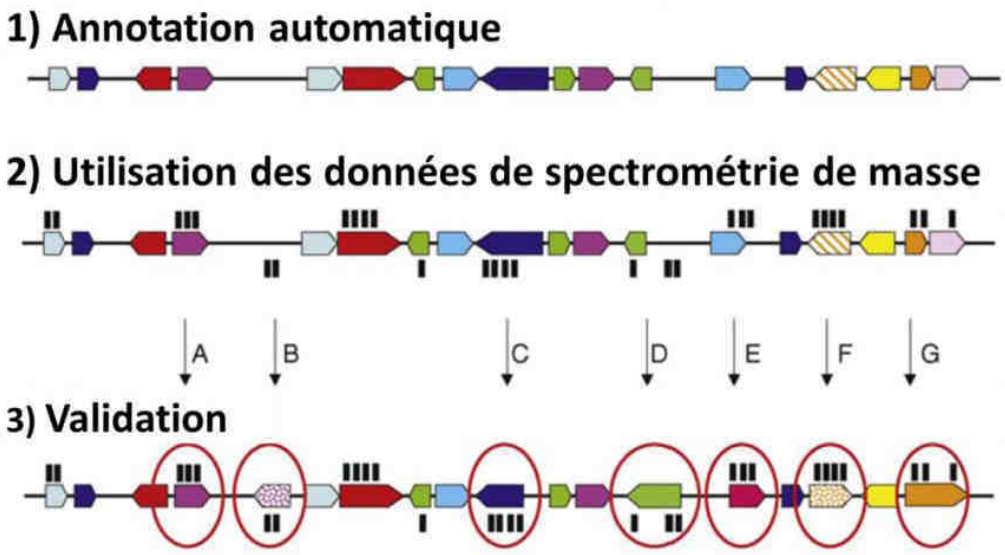


Figure I-25 : Correction des erreurs d'annotation grâce aux données de spectrométrie de masse (d'après Armengaud 2009)

1) annotation automatique du génome séquencé, 2) alignement des données peptidiques de spectrométrie de masse sur la séquence d'acides aminés, 3) validation manuelle des différents cas pour les erreurs d'annotation. Les différents peptides détectés en spectrométrie de masse sont indiqués par des rectangles noirs soit au-dessus (sens de lecture de la gauche vers la droite) soit au-dessous de la séquence (sens inverse de lecture). (A). L'existence d'un gène prédit est confirmée (B). Un gène non-prédit a été découvert. Le codon d'initiation du gène est localisé en aval (C) ou en amont (D) du site annoté auparavant. La détection de trois peptides montre que le gène doit être annoté sur un autre cadre de lecture (E). L'identification de quatre peptides indique que le gène doit être réorienté étant donné qu'il est traduit à partir de l'autre brin d'ADN (F). La détection de plusieurs peptides provenant de la même protéine permet de révéler l'erreur d'annotation commise avec la présence d'un codon stop anticipé ou le décalage de cadre de lecture (G). Les cas d'épissage et d'épissage alternatif pour les eucaryotes ne sont pas représentés.

A l'heure actuelle, chez les eucaryotes, ce type d'approche est essentiellement dévoué à des organismes modèles. Chez la plante *Arabidopsis thaliana*, le séquençage du protéome de différents tissus a permis la validation de 12 769 protéines prédites, la réannotation de 695 et la découverte de

778 (Castellana *et al.* 2008). Chez l'humain, une stratégie identique a permis de couvrir 84 % du protéome total prédit avec la confirmation de l'existence de 17 294 protéines et la découverte de 808 nouvelles annotations (Kim *et al.* 2014). Actuellement, l'analyse protéogénomique chez des organismes exposés à des stimuli environnementaux a été essentiellement réalisée en microbiologie. Chez la bactérie halotolérante *Tistlia consotensis*, l'analyse de son protéome en conditions hypo- et hyper-osmotique a permis de détecter la présence de 120 protéines non exprimées en conditions optimales de croissance, validant ainsi leur existence chez cet organisme (Rubiano-Labrador *et al.* 2014). Ainsi, l'annotation des génomes peut fortement bénéficier de travaux dans le domaine de l'écophysiologie ou de la toxicologie pour la validation et la découverte de gènes.

3.2. Outils et étapes de l'analyse protéomique

Comparée à l'analyse des acides nucléiques, l'analyse protéomique est caractérisée par une complexité accrue en raison des propriétés physico-chimique variées des protéines et de leur importante gamme de concentration. De même, alors que le séquençage des acides nucléiques utilise une réaction enzymatique (polymérisation par l'ADN polymérase) avec un nombre restreint de possibilités (quatre bases nucléotidiques différentes), le séquençage des protéines fait appel à des principes physiques (mesure de la masse moléculaire) et se confronte à une multitude de possibilités en termes de composition des protéines (vingt-deux acides aminés différents ainsi que les modification post-traductionnelles). L'identification à large échelle de protéines est réalisée selon une stratégie « bottom-up » car elle porte sur l'analyse des peptides provenant de la digestion d'un mélange de protéines. Elle fait opposition à la stratégie « top-down » où l'analyse est réalisée sur des protéines entières.

L'analyse protéomique est généralement composée de quatre étapes (Figure I-26).



Figure I-26 : Les grandes étapes de l'analyse protéomique

Dans le cas de la **préparation d'un échantillon** protéique, seules l'extraction et la solubilisation des protéines sont réalisées alors que dans le cas de la préparation d'échantillon peptidique, une digestion enzymatique est également exécutée. Cette digestion est effectuée par une protéase (en général la trypsine) et permet d'obtenir des peptides de taille analysable par le spectromètre de masse.

En raison de l'importante gamme dynamique et des propriétés physico-chimiques variées des protéines, l'étape de préparation des échantillons conditionne la fraction du protéome pouvant être étudié. En général, les protéines hydrophiles et abondantes sont les plus représentées (Bodzon-Kulakowska *et al.* 2007). Une couverture complète du protéome ne peut être atteinte qu'en combinant différentes stratégies de préparation de l'échantillon en terme de fraction subcellulaire ou encore de choix des tampons d'extraction (Omasits *et al.* 2013). Après l'étape de préparation de l'échantillon, celui-ci est fractionné par des méthodes séparatives basées sur l'électrophorèse et la chromatographie. Les mélanges peptidiques sont analysés par spectrométrie de masse et les données sont interprétées grâce à des logiciels bioinformatiques.

3.2.1. Les méthodes séparatives

Le développement des techniques séparatives pour décomplexifier l'échantillon protéique avant l'analyse par MS ont contribué à l'essor de la protéomique. La séparation se réalise au niveau protéique, en général par des techniques d'électrophorèse et au niveau peptidique par des techniques de chromatographie. Les différentes approches séparatives actuellement employées pour l'étude exhaustive du protéome sont présentées en Figure I-27.

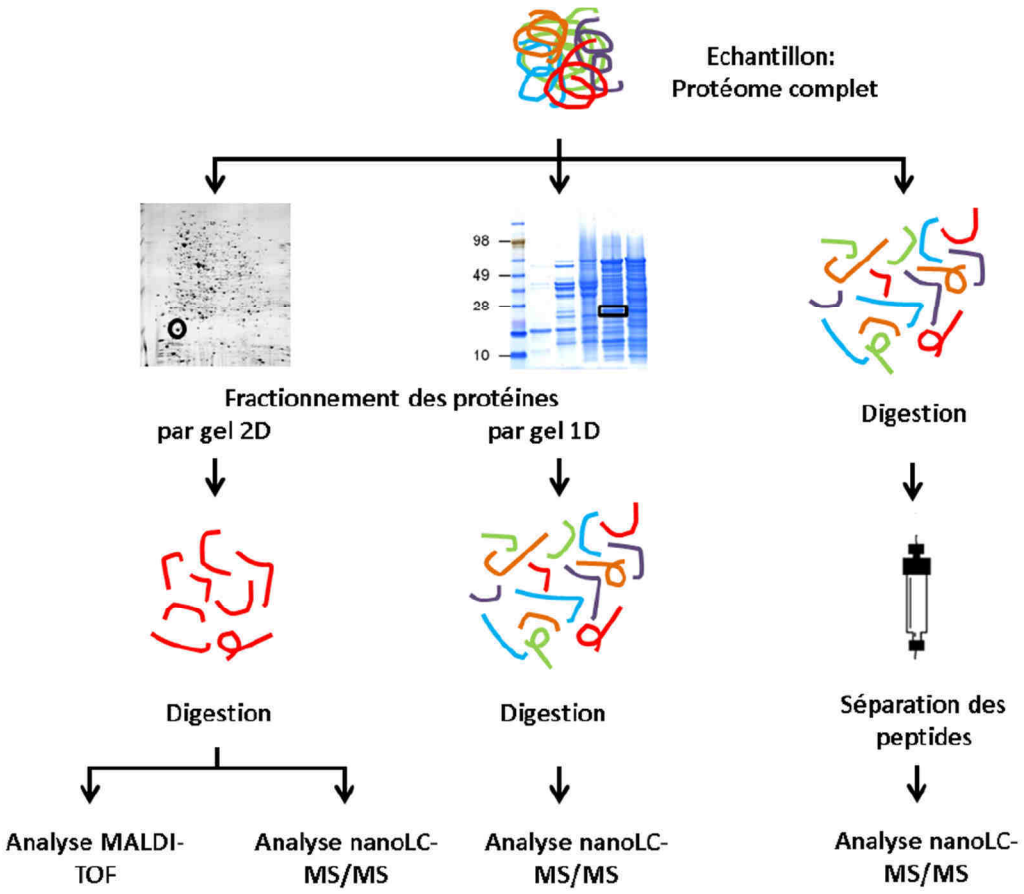


Figure I-27 : Stratégie générale pour l'étude du protéome (modifié d'après Armengaud and Bland 2014)

Une des méthodes les plus communément utilisées pour fractionner les protéines est la séparation par électrophorèse sur **gels SDS-PAGE** (sodium dodecyl sulfate polyacrylamide gel electrophoresis). Les protéines sont séparées suivant une dimension (1 D) selon leur poids moléculaire. Par la suite, le gel est coloré et les bandes protéiques sont excisées avant de procéder à la digestion enzymatique directement dans le gel. Afin d'augmenter le fractionnement du protéome, une deuxième dimension dans la séparation peut être ajoutée par l'approche **d'électrophorèse sur gel bidimensionnelle** (2D-PAGE). La première séparation est basée sur la différence de point isoélectrique (pI) et la seconde dimension sur le poids moléculaire. En plus de d'être une technique séparative, l'électrophorèse sur gel 2D permet la visualisation de la protéine après coloration du gel sous forme de spot protéique. Si un gel de référence a été préalablement établi avec des protéines identifiées, il est possible en superposant les images d'identifier directement la protéine d'intérêt sans recourir à l'analyse par MS. De même, cette technique permet également de comparer l'abondance des protéines entre les différents échantillons biologiques, par exemple entre des organismes soumis à un stress et des organismes témoins, afin

de définir les protéines sur- et sous-exprimées en réponse à la perturbation. Par la suite, seuls les spots protéiques d'intérêt sont analysés par MS (Rabilloud *et al.* 2010). Cependant, plusieurs inconvénients limitent l'efficacité des méthodes séparatives utilisant des gels notamment en termes de rendement analytique *e.g*. coloration, excision manuelle des spots et d'exhaustivité *e.g*. faible gamme dynamique et étude de protéines aux propriétés physico-chimiques restreintes. Une alternative à l'électrophorèse en gel est l'utilisation d'une séparation des protéines en solution suivant leur pI avec la méthode **OFF-Gel**.

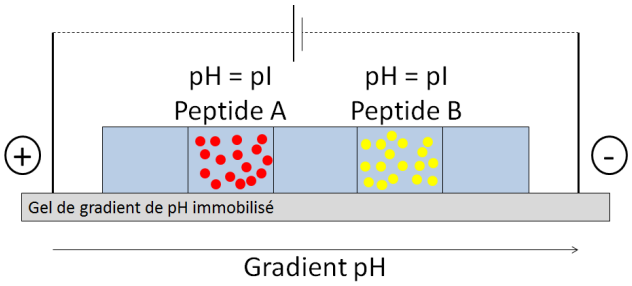


Figure I-28 : Principe de l'OFF-Gel

Le principe, présenté en Figure I-28, repose sur l'équi-répartition de l'échantillon au sein de puits contenant du tampon de migration et qui reposent sur une bande de gel avec gradient de pH immobilisé. Une tension est par la suite appliquée aux deux extrémités de la bande de gel et les protéines ou peptides migrent jusqu'à atteindre une position où le pH est égal à leur pI. A l'issue de l'électrophorèse, les biomolécules sont directement récupérées en solution au sein des différents puits. Cette méthode est davantage employée sur les mélanges peptidiques, les peptides ayant une plus large gamme de pI que les protéines, augmentant ainsi la résolution du protéome (Chenau *et al.* 2008).

Les méthodes de chromatographie liquide (LC) sont employées en couplage direct avec un spectromètre de masse en tandem ayant une source d'ionisation par électrospray (voir partie I-3.2.2.2). La séparation des peptides est réalisée suivant différentes propriétés physiques telles que la charge ou l'hydrophobicité. Une des méthodes les plus populaires est la **nano-LC**, une chromatographie en phase inverse à très faible débit. La colonne de chromatographie est composée de silice greffée à des chaînes linéaires de 8 ou 18 atomes de carbones (C8 et C18) et permet de séparer les peptides. Elles séparent les peptides selon leur hydrophobicité grâce à un gradient de solvant organique (généralement acétonitrile/eau). Les peptides hydrophiles sont

rapidement élués alors que les hydrophobes sont retenus plus longtemps dans la colonne (Gaspari and Cuda 2011). A l'instar de l'électrophorèse, une deuxième dimension peut être ajoutée à la séparation des peptides, en ajoutant en amont de la colonne en phase inverse, une colonne de type « échangeuse de cations » où les peptides sont élués selon leur charge électrique. Lorsque les deux colonnes sont physiquement couplées, le terme **MudPIT** (Multidimensional Protein Identification Technology) est employé (Washburn *et al.* 2001) alors que quand elles sont dissociées, on parle de **2D-LC** (Nagele *et al.* 2004).

La Figure I-27 présente les différentes approches pour l'analyse du protéome. Dans le cas de la réalisation de gel 2D, le mélange peptidique est peu complexe et est analysé directement par MS suite à la digestion enzymatique. En revanche pour les autres stratégies dites « **protéomique shotgun** », le mélange peptidique est plus complexe, nécessitant la réalisation d'étapes séparatives après digestion, en général par LC.

3.2.2. L'analyse par spectrométrie de masse

3.2.2.1. Principe général

La spectrométrie de masse permet de détecter et d'identifier des molécules d'intérêt chargées, nommés ions, par leur rapport masse-sur-charge (***m/z***). Par raccourci de langage, on parle de la masse de l'ion, cette dernière étant déterminée à partir du pic observé sur le spectre de masse. La MS permet également de caractériser la structure chimique des molécules en les fragmentant. Un spectromètre de masse est constitué de trois parties, comme illustré en Figure I-29 :

- Une **source d'ionisation** : chargeant électriquement les molécules introduites et permettant leur émission en phase gazeuse
- Un **analyseur** : séparant les ions en phase gazeuse selon leur rapport *m/z*
- Un **détecteur** : transformant le courant ionique en courant électrique, puis le traduit en spectre de masse où l'intensité du signal est mise en relation avec le rapport *m/z*

Un système d'enregistrement informatique est également présent permettant le traitement et la visualisation des données enregistrées.

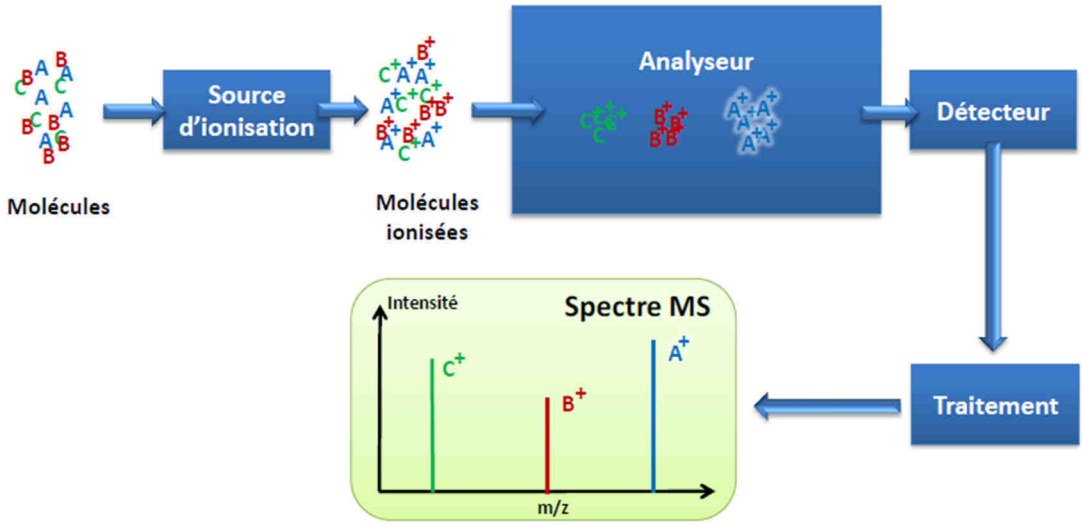


Figure I-29 : Schéma du principe de la spectrométrie de masse (Allain 2014)

Afin de compléter la mesure de la masse par des informations structurales, *i.e*. séquençage du peptide, deux analyseurs peuvent être combinés, on parle alors de **spectrométrie de masse en tandem** notée **MS/MS**. A partir du 1er spectre MS, les *n* peptides ionisés de la plus forte intensité sont isolés séparément selon une place autour de leur m/z comme présenté en Figure I-30. Chaque peptide isolé (ion parent ou précurseur) est fragmenté dans une cellule de collision, au niveau des liaisons peptidiques. Ceci aboutit à la formation d'ions fragments (ou ion fils), nommés b et y selon si la charge est conservé sur le fragment N-terminal ou C-terminal. Les ions fils sont analysés dans un second analyseur, générant un spectre de fragmentation, nommé **spectre MS/MS**. Le peptide est donc caractérisé selon sa masse et sa séquence (Biemann 1988).

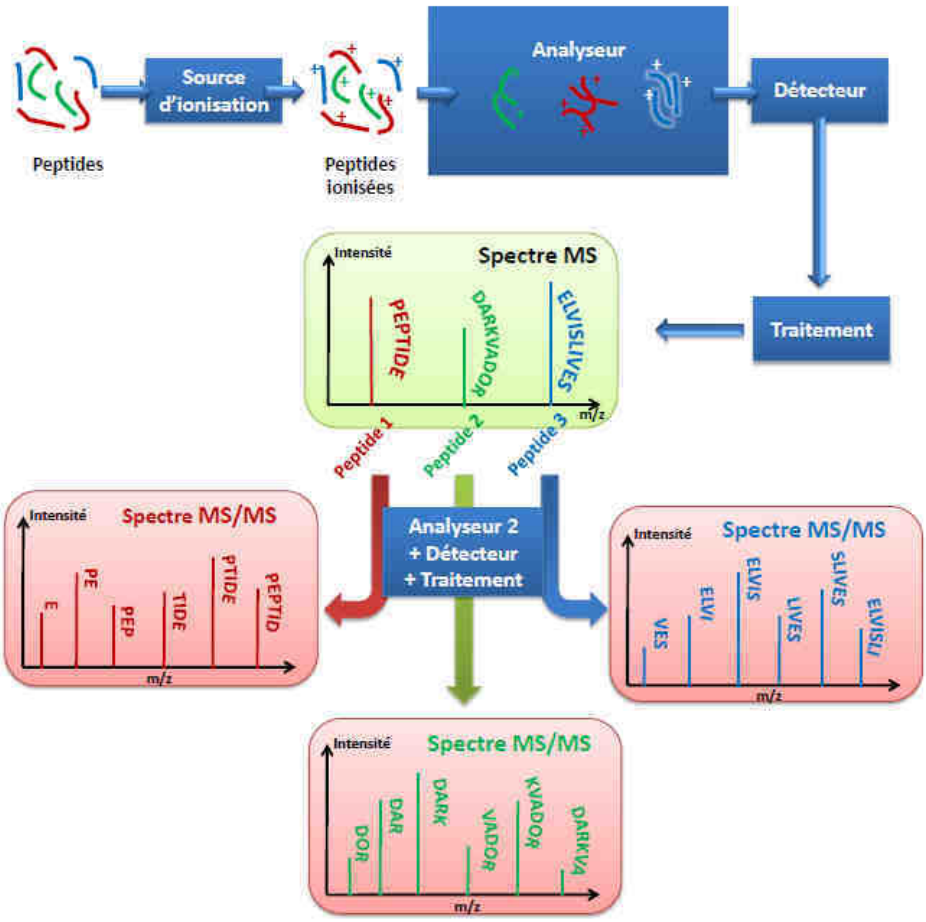


Figure I-30 : Schéma du principe de la spectrométrie de masse MS/MS (Allain 2014)

Après ionisation, les peptides sont transférés à un premier analyseur qui mesure leurs masses sélectionne les n plus intenses (n = 3 ici). Une fois isolé, chaque peptide est fragmenté dans une cellule de collision et les ions fragments générés sont analysés dans un second analyseur, générant ainsi un spectre MS/MS.

3.2.2.2. Outils pour l'analyse par spectrométrie de masse

Afin de répondre aux diverses problématiques scientifiques soulevées dans le domaine de la protéomique, différentes sources d'ionisation et différents analyseurs aux caractéristiques variées peuvent être employés (de Hoffmann and Stroobant 2005).

Les **sources d'ionisation MALDI** (Matrix Assisted Laser Desorption Ionisation) (Karas and Hillenkamp 1988) et l'électrospray **ESI** (Fenn *et al.* 1989) sont les plus couramment utilisées. Elles sont non dénaturantes car elles permettent le transfert de biomolécules intactes en phase gazeuse.

Pour la source MALDI, un dépôt cristallin contenant les peptides est réalisé sur une matrice (ou cible) contenant un composé absorbant dans le domaine UV. Un faisceau à laser pulsé est dirigé vers la cible afin d'exciter thermiquement la matrice, cette dernière sert de vecteur d'énergie entre le laser et les analytes. Les analytes sont alors entrainés dans un plasma d'expansion où se produit leur ionisation. Les ions monochargés produits passent ensuite sous forme gazeuse pour être analysés.

L'ionisation ESI nécessite l'introduction de l'échantillon en solution, en général en sortie de chromatographie liquide, les peptides étant élués directement dans la source d'ionisation suivant un gradient d'hydrophobicité. Ce mode d'ionisation permet la génération d'ions multichargés sous l'effet d'un fort champ électrique, qui passent par la suite en phase gazeuse. Ce processus se décompose en trois étapes [1] la formation de petites gouttelettes de solvant chargées contenant les analytes [2] fission des gouttelettes chargées en goulettes filles et [3] désolvatation progressive des gouttelettes filles et émission des ions en phase gazeuse.

Les analyseurs peuvent être caractérisés suivant différents paramètres de performances : sa résolution (capacité à dissocier deux ions ayant un rapport *m/z* proche), sa justesse de mesure de masse, sa gamme de masse ou encore sa sensibilité. Parmi les différents analyseurs disponibles, on retrouve l'analyseur à temps de vol (TOF), les quadripôles (Q) et ses dérivés, les trappes ioniques (trap) et orbitale (Orbitrap) (de Hoffmann and Stroobant 2005).

Pour l'**analyseur TOF**, à la sortie de la source, une tension est appliquée sur les ions, leur conférant une énergie cinétique. Les ions « volent » jusqu'au détecteur, et le temps nécessaire pour parcourir cette distance est fonction de leur masse. C'est pourquoi, à partir du chronométrage de leur temps de vol, il est possible de déduire le rapport *m/z*.

L'**analyseur quadripôle** est composé de deux paires de barres métalliques parallèles, soumises deux à deux à un potentiel comprenant une tension continue U et une tension alternative V. Les ions entrant dans l'analyseur subissent l'effet du champ quadripolaire. Selon les valeurs de U et V, certains ions adoptent des trajectoires stables et peuvent être détectés tandis que d'autres ont des

trajectoires instables et sont collectés par les barres du quadripôle. Par conséquent, ce champ électrique ne laisse passer que des ions ayant un rapport *m/z* donné. Toutefois, cet analyseur permet une grande vitesse de balayage par le contrôle des tensions U et V.

La **trappe ionique** est constituée de trois électrodes et constitue un « piège à ions » où s'effectuent successivement l'analyse et la détection des ions. Comme pour le quadripôle, un champ électrique est créé et les ions suivent une trajectoire oscillante à l'intérieur du piège jusqu'à être expulsés de la trappe (mode MS). En mode MS/MS, après piégeage des ions, ceux à fragmenter sont isolés puis fragmentés et les ions fils sont éjectés vers le détecteur. Cet analyseur a l'avantage d'être très sensible et robuste mais présente une résolution faible et ne tolère qu'un nombre limité d'ions.

La **trappe orbitale** permet un piégeage des ions, constitué d'une électrode creuse à l'intérieur de laquelle est placée co-axialement une deuxième électrode en forme de fuseau permettant la création d'un champ électrostatique. Les ions entrent dans une trajectoire circulaire autour de l'électrode centrale, avec des oscillations axiales. Le courant induit par ces oscillations permet, grâce à la transformée de Fourier, d'accéder aux mesures de *m/z*.

Désormais, il existe une large panoplie de **spectromètres de masse hybrides**, combinant différents types d'analyseurs. Une synthèse des différents systèmes disponibles est présentée en Tableau I-6.

Tableau I-6 : Vue d'ensemble des sources d'ionisation et des analyseurs de masse employés en protéomique (d'après Armengaud 2013)

	Abréviation	Nom complet	Caractéristiques
Techniques d'ionisation	MALDI	Matrix-Assisted Laser Desorption Ionization	Ions monochargés, ionisation pulsée
	ESI	ElectroSpray Ionization	Ions multichargés
Analyseur de masse	TOF	Time-Of-Flight	Analyseur ultra rapide
	Trap	Ion trap	Possibilités d'accumuler les ions
	Q	Quadrupole	Filtration des ions suivant leur masse
	Orbitrap	Orbit trap	Détermination précise de la masse, important pouvoir de résolution et grande gamme dynamique
Spectromètre de masse hybride	Q-TOF		Identification à haut débit de protéines
	Q-Trap		Protéomique quantitative ciblée
	Q-Orbitrap		Protéomique quantitative ciblée à haut débit
	TOF-TOF		Identification à haut débit de spots 2D-PAGE
	Trap-Orbitrap		Identification à haut débit de protéines

3.2.3. Stratégies pour l'interprétation des données protéomiques

3.2.3.1. Identification des protéines

L'étape centrale dans le traitement des données MS est la comparaison entre données expérimentales et celles théoriques issues de BD de protéines, réalisée à l'aide d'algorithmes. Deux grandes stratégies, présentée en Figure I-31 sont actuellement employées.

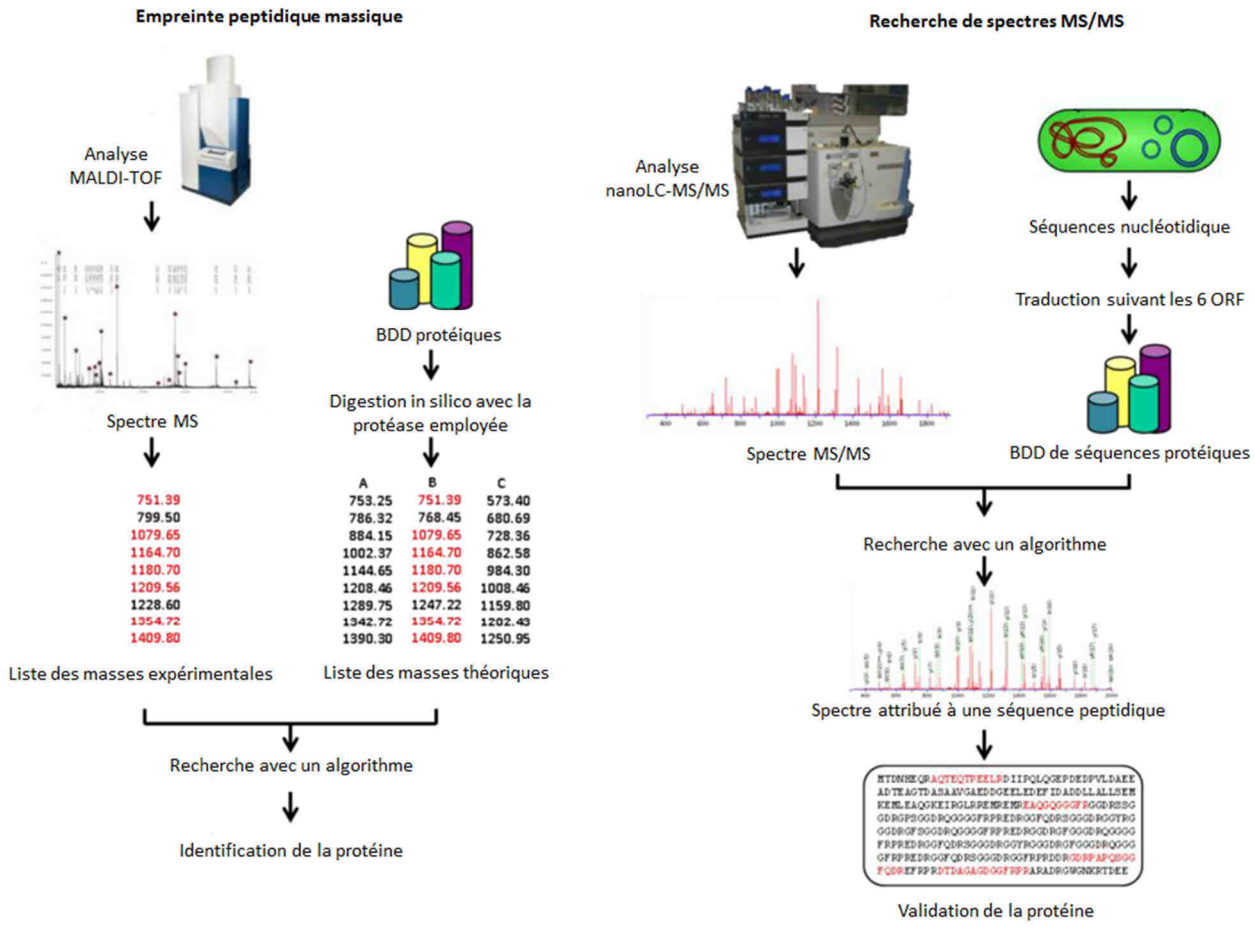


Figure I-31 : Approches pour l'identification des protéines (d'après Armengaud and Bland 2014)

La première, appelée **empreinte peptidique massique** (peptide mass fingerprinting, **PMF**) repose généralement sur une séparation du protéome par gel 2D puis une analyse des protéines par l'utilisation d'une source d'ionisation MALDI couplée à un analyseur de type TOF (MALDI-TOF-MS). Les masses expérimentales des peptides générées sont comparées aux masses théoriques des peptides obtenus par digestion *in silico* de l'ensemble des séquences de la BD protéique (Henzel *et al.* 1993). Toutefois, l'efficacité de cette méthode repose sur une bonne séparation du protéome afin de garantir la précision de mesure de la masse et nécessite une importante étape de préparation de l'échantillon. L'utilisation d'appareillage de LC, couplée à une source d'ionisation ESI et à un analyseur générant des spectres MS/MS (protéomique shotgun) est une méthode plus rapide et plus sensible par l'identification des protéines, identification par **recherche de spectres MS/MS**. Dans ce cas-ci, les séquences des peptides sont identifiées en corrélant les spectres de fragmentation expérimentaux avec les spectres théoriques issus de la digestion *in silico* des protéines présentes dans la BD (Eng *et al.* 1994). Ces attributions sont faites en tenant compte de

différents paramètres définis par l'utilisateur tels que l'enzyme de digestion utilisée, les modifications des acides aminées autorisées ou encore la tolérance sur la masse du précurseur. La confiance attribuée à l'identification peut être évaluée par l'observation du score de corrélation, mesurant le degré de similarité entre les données expérimentales et théoriques. Généralement, seul le peptide qui a le meilleur score de corrélation est conservé par la suite.

A partir des séquences peptidiques ou les peptides identifiés, il est possible de remonter aux protéines contenant ces peptides afin d'établir une liste de protéines identifiées. Toutefois, en protéomique shotgun, cette étape est ardue car dans certains cas un peptide peut être attribué simultanément à différentes protéines (protéines homologues ou redondantes dans la BD). En général, seule l'attribution correspondant à la protéine qui a été identifiée avec le plus grand nombre de peptides est gardée au moment de la création de la liste de protéines observées, en se basant sur le principe de parcimonie. Une protéine est généralement validée quand elle a été identifiée par au moins deux peptides distincts. De même, plus le taux de couverture de la protéine est important, plus la confiance dans son identification est grande, un critère qui favorise la découverte des protéines les plus longues et les plus abondantes (résumé par Cottrell 2011).

En absence de BD protéiques, les spectres MS/MS peuvent également être interprétés *de novo* grâce à des algorithmes spécifiques. Toutefois, cette approche est peu pertinente dans une démarche de caractérisation à large échelle du protéome, étant très chronophage.

3.2.3.2. Analyse quantitative

Dans les approches d'étude à large échelle du protéome, le terme « quantification » dite relative se réfère à l'estimation de l'abondance d'une protéine, en général par comparaison entre différentes conditions expérimentales (témoins et exposés). Elle s'oppose à la démarche de quantification « absolue » utilisée en spectrométrie de masse ciblée. Les stratégies pour évaluer l'abondance d'une protéine sont multiples et dépendantes de la méthode analytique choisie. Elles se distinguent notamment par le marquage ou l'absence de marquage des biomolécules, respectivement sous le terme de « label » et « label-free » (résumés par Martyniuk *et al.* 2012, Otto *et al.* 2012, Armengaud 2013).

Dans le cas d'une analyse protéomique sur gel 2D, les gels sont colorés au nitrate d'argent ou au bleu de Coomassie et la quantification est réalisée par **densitométrie**. Par la suite avec l'introduction de la 2D-DIGE (Differential Gel Electrophoresis), les échantillons protéiques à analyser sont marqués avec des fluorochromes différents possédant la même masse mais ayant des longueurs d'onde d'excitation et d'émission différents. Les échantillons marqués sont par la suite mélangés et les protéines séparés sur un même gel. Pour les méthodes basées sur l'utilisation de gel 2D, la quantification est réalisée avant l'analyse MS, contrairement aux méthodes « gel-free » où elle est réalisée en parallèle à l'interprétation des spectres MS/MS. Le Tableau I-7 synthétise les différentes méthodes disponibles en protéomique massive pour l'évaluation de l'abondance des protéines.

Différentes techniques de marquage par des étiquettes chimiques sont disponibles pour les méthodes gel-free. La technique **ICAT** (Isotope-Coded Affinity Tag) porte sur la greffe au niveau des cystéines des fragments peptidiques, d'étiquettes composées d'une fonction réagissant avec les thiols, d'un bras contenant les isotopes stables légers ou lourds d'un même élément et d'un groupement biotine. La technique **iTRAQ** (Isobaric tags for relative and absolute quantitation) se base sur le marquage des extrémités N-terminale des peptides par des étiquettes isobariques, contenant un ion rapporteur et un bras servant de balance. La masse globale de l'étiquette reste stable alors que celles du rapporteur et de la balance fluctuent selon les différents marquages. Pour les deux techniques, les échantillons peptidiques marqués sont mélangés et les peptides séparés par LC. Dans le cas de l'iTRAQ, les peptides de séquence identique mais issus des différents échantillons sont éluées au même moment et le 1er analyseur va les laisser passer simultanément, ces derniers présentant la même masse. En revanche, après fragmentation, la liaison entre la balance et le rapporteur est rompue et la quantification est réalisée en comparant l'aire du pic de chaque type de rapporteur. Dans le cas de l'ICAT, les peptides de séquences identiques ne possèdent pas la même masse, étant marqués par différents isotopes et ne sont donc pas analysés simultanément. Toutefois, après identification du peptide par interprétation des spectres MS/MS, la comparaison des différentes intensités de l'ion précurseur permet d'estimer l'abondance du peptide au sein de chaque échantillon. Il existe également des méthodes de marquage dites métaboliques, où les organismes croissent sur des milieux de culture enrichis avec différents isotopes (en général ^{14}N et ^{15}N). Toutefois, cette démarche est restreinte à des modèles biologiques de faible complexité, *ex* : lignées cellulaires.

Tableau I-7 : Méthodes de quantification employées en analyse protéomique massive

Type	Nom	Principe	Multiplexage	Type d'information nécessaire	Information quantitative	Avantages	Inconvénients
Analyse en gel 2D							
Label-free	2D-PAGE	Analyse différentielle d'image	Comparaison inter-gel	Intensité du signal / Coloration	Protéine	Mesure au niveau de la protéine	Etude de protéines de gamme restreinte en pI et hydrophobicité
Label	2D-DIGE	Analyse différentielle d'image sur un même gel à différentes longueurs d'onde	3	Intensité du signal / Coloration	Protéine	Mesure au niveau de la protéine et comparaison des échantillons sur un même gel	Etude de protéines de gamme restreinte en pI et hydrophobicité
Analyse « gel-free »							
Label	iTRAQ	Mesure de l'intensité des différents rapporteurs (aire sous le pic)	2-8	MS/MS	Protéine/ Peptide	Grande capacité de multiplexage	
Label	ICAT	Mesure de l'intensité des différentes étiquettes isotopiques	2-3	MS	Protéine/ Peptide	Permet de détecter de faible changement d'abondance	Marque uniquement les peptides contenant des cystéines
Label-free	Spectral Count	Comptage du nombre de spectre MS/MS attribué à une protéine	Comparaison inter-échantillon	MS/MS	Peptide	Simplicité de préparation de l'échantillon	Sous- représentation de protéines de petite taille
Label-free	Aire du pic	Mesure de l'aire sous le pic	Comparaison inter-échantillon	MS/MS	Peptide	Simplicité de préparation de l'échantillon	Bruit de fond et normalisation de l'abondance des pics

Enfin, les méthodes label free comprennent la méthode de mesure d'aire sous le pic et de décompte des spectres (spectral count, SC).

3.2.3.3. Annotation fonctionnelle

La question centrale relative aux projets « omiques » est l'identification de la fonction biologique assurée par les gènes et produits associés. Face à la quantité de données produites aux cours de ces études, des procédures automatisées « d'inférence de fonction » ont été développées. Alors que les méthodes analytiques pour le séquençage en acides nucléiques et en acides aminés diffèrent, les outils employés pour l'annotation fonctionnelle sont, pour la majorité, partagés par l'ensemble de la communauté « omiques ». De manière générale, c'est l'information en acides aminés qui est employée pour réaliser cette étape du traitement des données. Les outils développés pour l'annotation fonctionnelle reposent sur la disponibilité de séquences préalablement annotées sur des BD publiques telles que NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/protein) ou UniProt/SwissProt (http://www.uniprot.org/) (Hawkins *et al.* 2008). L'alimentation de ces bases est assurée par la communauté scientifique au travers des différents projets de séquençage.

Une des méthodes les plus employées pour prédire la fonction d'un gène est d'utiliser les outils de recherche par homologie tels que BLAST (Basic Local Alignement Search Tool ou PSI-BLAST (Position-Specific Iterative-BLAST). L'annotation du gène d'intérêt (query) se fait alors sur la base des similarités avec les séquences caractérisées antérieurement contenues dans ces bases de données (Altschul *et al.* 1990). Les résultats obtenus, illustrés en Figure I-32, compilent les séquences présentant la plus importante similarité, avec leur identifiant (accession), un descriptif contenant leur fonction biologique et l'espèce associée, un score renseignant sur la qualité de l'alignement, la couverture de l'alignement face au query ou encore la E-value (assimilable à la *p*-value). Les séquences retenues sont celles présentant un score significatif et une valeur d'E-value inférieure à la valeur seuil. En absence d'homologues pertinents, la protéine est qualifiée d'orpheline.

Description	Max score	Total score	Query cover	E value	Ident	Accession
RNA helicase [Bacillus licheniformis S 16]	983	983	100%	0.0	100%	EWH23647.1
RNA helicase [Bacillus licheniformis]	953	953	100%	0.0	96%	WP_025810549.1
MULTISPECIES: RNA helicase [Bacillus]	953	953	100%	0.0	96%	WP_009329965.1
RNA helicase [Bacillus licheniformis]	947	947	100%	0.0	96%	WP_016886328.1
MULTISPECIES: RNA helicase [Bacillus]	945	945	100%	0.0	96%	WP_023856890.1
RNA helicase [Bacillus licheniformis]	944	944	100%	0.0	96%	WP_016885643.1
ATP-dependent RNA helicase DeaD [Bacillus licheniformis]	944	944	100%	0.0	96%	WP_020453516.1
RNA helicase [Bacillus licheniformis]	943	943	100%	0.0	96%	WP_003177757.1
RNA helicase [Bacillus licheniformis]	933	933	100%	0.0	95%	WP_021837826.1

Figure I-32 : Exemple de recherche de similarité par BLAST sur une séquence d'ARN hélicase 16 S de *Bacillus licheniformis*

Afin d'obtenir des informations complémentaires, d'autres BD peuvent être consultées. Celle du consortium Gene Ontology (http://geneontology.org/) permet d'enrichir l'annotation fonctionnelle par des informations concernant trois propriétés : le compartiment cellulaire, la fonction moléculaire (catalyse, structure…) ou le processus biologique (respiration, fermentation …) (Ashburner *et al.* 2000). La BD KEGG Pathway (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/pathway.html) propose d'intégrer visuellement les données au sein de systèmes biologiques telles que les voies métaboliques ou encore les processus cellulaires (Kanehisa and Goto 2000).

Enfin certaines BD telle que Pfam (http://pfam.xfam.org/), permettent la recherche de régions fonctionnelles (domaines) au sein de la protéine. L'identification de ces domaines permet d'affiner la compréhension de son rôle fonctionnel *e.g.* présence d'un domaine d'interaction protéine - protéine, d'un domaine transmembranaire ou encore d'un domaine catalytique (Bateman *et al.* 2002).

3.3. Applications et limites des outils protéomiques en écotoxicologie aquatique

Le Tableau I-8 et le Tableau I-9 présentent les différentes études protéomiques réalisées en écotoxicologie aquatique, respectivement chez les mollusques et crustacés/polychètes.

Tableau I-8 : Etudes protéomiques en écotoxicologie aquatique réalisées chez les mollusques

Organisme	Contexte	Conditions	Méthode séparative	Spectromètre de masse	Taux d' identification	Identifications principales (max. 10)	Reference
Estuaire							
Scorbicularia plana	Pollution marine	*In-situ*	2D-PAGE	MALDI-TOF MS/MS : nESI –LCQ-IT	2/19	hypoxanthine-guanine phosphoribosyltransférase et G3PDH	Romero-Ruiz *et al.* (2006)
Saccostrea glomerata	Exposition aux chlorures de Cu, Zn et Pb	Laboratoire	2D-PAGE	MS/MS : nanoLC-nESI LTQ-XL	18/20	cristalline, vitellogénine, protéine ribosomale 40S, tropomyosine, myosine, beta-tubuline, ATP synthase, triose-phosphate isomérase, poly protéine et nongradient byssal precursor	Thompson *et al.* (2011)
Eaux marines							
Crassostrea hongkongensis	Pollution métallique	*In-situ*	2D-PAGE	MALDI-TOF/TOF	19/173	adenosylmethionine synthase, beta-actine, actine, mitochondrial H^+ ATPase, prohibitin, phosphorylase phosphatase, sous-unité du protéasome 26S, enolase, CRE-PGK-1, acyl-CoA déshydrogénase et hydroxyphenylpyruvate dioxygénase.	Liu and Wang (2012)
hamaelea gallina	Exposition à l'Aroclor, Cu, As et tributylétain	Laboratoire	2D-PAGE	MALDI-TOF ESI Q-TOF	4/15	Tropomyosine, myosine, actine et Act87E	Rodriguez-Ortega *et al.* (2003)
Chlamys nobilis, Ruditapes philippinarum, Perna viridis et Saccostrea cucullata	Exposition au Cd et comparaison interspécifique	Laboratoire	2D-PAGE	MALDI-TOF/TOF	8/?	Actine, galectine, tubuline et SOD	Liu *et al.* (2012)
Mytilus edulis	Exposition au pétrole, alkyl phénols et HAP.	Laboratoire	2D-PAGE	MALDI-TOF nanoLC-nESI-Q-Trap	19/?	HSP70, Glutamyl tRBA synthétase, actine, facteur de réplication, arginase, cycline D1, fructose-bi phosphate aldolase, amino-peptidase, inhibiteur de protéase et tropomyosine.	Manduzio *et al.* (2005)

Mytilus edulis	Exposition aux phtalates, polybromodiphényléthers et bisphénol A	Laboratoire	2D-DIGE	MALDI-TOF et ESI-Q-TOF	X	Hydroxy-acide oxydase 1, GST, aldéhyde déshydrogénase 4A1, anhydrase carbonique, Mn-SOD, ATP synthase, peroxine 10, cytochrome P450 2A6, enoyl-CoA hydratase, phospholipase A2 et sous unité du cytochrome c oxydase.	Apraiz *et al.* (2006)
Mytilus edulis	Pollution marine	*In-situ*	Chromatographie échangeuse d'ion et 2D-PAGE	HPLC-ESI-MS/MS	10/13	Cathepsine B, AMP binding protéinase, acyl-CoA déshydrogénase, GST, phosphoglycérate kinase, Pex1, HSC70, HSC71, fascine et aldéhyde déshydrogénase 1A2	Amelina *et al.* (2007)
Mytilus edulis	Exposition au pétrole, alkyl phénols et HAP.	Laboratoire	2D-PAGE	MALDI-TOF nanoLC-nESI-Q-Trap	19/?	HSP70, Glutamyl tRBA synthétase, actine, facteur de réplication, arginase, cycline D1, fructose-bi phosphate aldolase, amino-peptidase, inhibiteur de protéase et tropomyosine.	Manduzio *et al.* (2005)
Perna viridis	Exposition au Cd et au peroxyde d'hydrogène. Comparaison inter-tissulaire	Laboratoire	2-DE	MALDI-TOF/TOF	15/26	Cu/Zn-SOD, calréticuline, disulfure isomérase, disulfure isomérase, enolase, ATP synthase, Hsps-p23, calponine, arginine kinase et myosine	Leung *et al.* (2011)
Ruditapes decussatus	Exposition au Cd	Laboratoire	2-DE	microLC-ESI-LTQ/FT-Orbitrap	X	Actine, beta-tubuline, aldéhyde déshydrogénase, inhibiteur Rab GTPase et acyl CoA déshydrogénase	Chora *et al.* (2009)
Haliotis diversicolor (gastéropode)	Exposition au phtalates de diallyle et bisphénol A	Laboratoire	2-DE	MALDI-TOF	18/24	Aldéhyde déshydrogénase, GST, Mn-SOD, Beta-tubuline, HSP70, thiorédoxine peroxydase, catalase, sex-hormone binding globine, 7-dehydrocholesterol réductase et ATP synthase.	Zhou *et al.* (2010)
Eaux douces							
Dreissena polymorpha	Exposition au benzo(alpha)pyrène et comparaison intra-sexe	Laboratoire	2D-PAGE	MALDI-TOF/TOF	7/19	Phosphatase, N-acétyle-transférase, actine, aspartate aminotransférase, alcool déshydrogénase et peroxyrédoxine-6	Riva *et al.* (2011)
Dreissena polymorpha	Exposition au triclosan	Laboratoire	2D-PAGE	MALDI-TOF/TOF et LC-ESI-LTQ/FT-Orbitrap	12/12	HSP70, tubuline, cristalline, G3PDH, myosine, malate déshydrogénase, calcium-binding protéine, annexine B9 et protéine 14-3-3	Riva *et al.* (2012)

Tableau I-9 : Etudes protéomiques en écotoxicologie aquatique réalisées chez les crustacés et polychètes.

Organisme	Contexte	Conditions	Méthode séparative	Spectromètre de masse	Taux d' identification	Identifications principales (max. 10)	Reference
Crustacés							
Eriocheir sinensis (estuaire)	Exposition au Cd	Laboratoire	2D-PAGE	nanoLC-nESI-Q-TOF	15/25	Alpha tubuline, GST, crustacean calcium-binding protein, disulfure isomérase, ATP synthase, cathepsine D, G3PDH, sous-unité du protéasome, thiorédoxine et arginine kinase	Silvestre *et al.* (2012)
Gammarus pulex (eaux douces)	Exposition aux PCB	Laboratoire	2D-PAGE	MALDI-TOF/TOF	14/21	Facteur d'élongation, actine, alpha-spectrine, G3PDH, enolase, hémocyanine, Arginine kinase, glutamate déshydrogénase, dihydropyrimidine déshydrogénase et fonction inconnue	Leroy *et al.* (2010)
Hyalella azteca et Diporeia sp. (eaux douces)	Exposition à l'atrazine et comparaison interspécifique	Laboratoire	2D-PAGE	MALDI-TOF/TOF	*Hyalella*: 19/180 *Diporeia* : 4/20	G3PDH, alcool déshydrogénase, alanine glyoxylate aminotransférase, enolase, ATP synthase, arginine kinase, créatine kinase et phosphoglycérate kinase	Ralston-Hooper *et al.* (2011)
Polychètes							
Ophelina sp.1 (eaux marines)	Pollution métallique	*In-situ*	2D-DIGE	MALDI-TOF/TOF	24/46	Protéine cytosquelette, troponine-C, beta actine, para myosine, actine, peptidase s8 et s53, subtilisine protéinase, hémoglobine et protéine LIM.	Neave *et al.* (2012)

FT : Fourier transform ion cyclotron resonance ; G3PDH : glycéraldéhyde-3-phosphate déshydrogénase ; IT : ion trap ; LCQ-IT Quadrupole ion trap mass et LTQ : Linear Ion Trap

En majorité, ces études ont été réalisées en laboratoire (3/4 des études), l'objectif principal étant d'élucider le mode d'action des xénobiotiques, par la comparaison du protéome d'organismes témoins et d'organismes exposés, ainsi que de découvrir des biomarqueurs spécifiques de la perturbation.

En termes d'équipement analytique, toutes les études (n=18) utilisent la 2D-PAGE pour la séparation des protéines et leur visualisation. Seules deux études emploient une révélation des spots par DIGE, les autres se basant sur la densitométrie. En conséquence, la source d'ionisation des peptides la plus employée est la source MALDI, couplé à un analyseur TOF (n=13).

L'emploi de la 2D-PAGE confère différentes limites à l'analyse protéomique en termes de cadence analytique ou d'exhaustivité (voir I-3.2.1). Par conséquent, un nombre restreint de spots peuvent être analysés (~44 spots par étude). L'identité de la protéine est obtenue en moyenne pour 54 % des spots protéiques analysés par MS. Ce taux s'explique par l'absence de BD protéiques spécifique à l'organisme d'étude pour l'interprétation des spectres MS/MS (Figure I-31). En effet, la recherche de spectres se fait contre l'intégralité de la BD NCBI ou SwissProt, comprenant une grande variété d'organismes. Cette démarche est qualifiée d'identification par homologie. Elle présente deux désavantages majeurs : [1] l'utilisation de large BD abaisse la sensibilité en terme d'identification des peptides (plus de bruit de fond, identification non spécifique) (Wang *et al.* 2012) et [2] seules les protéines les plus conservées et les plus ubiquitaires au sein du vivant peuvent être identifiées. En effet, dans les études présentées en Tableau I-8 et Tableau I-9, les protéines principalement identifiées sont celles impliquées dans le métabolisme énergétique (ATP synthase, arginine kinase ou glyceraldhyde-3-phosphate déshydrogénase) ou dans le maintien du cytosquelette (actine, tubuline, myosine). Or, lors du séquençage d'EST chez des daphnies soumises à différents stress toxique ou biotique, une surreprésentation des transcrits associés aux gènes spécifiques à la daphnie a été mise en évidence (Colbourne *et al.* 2011). Par conséquent, en écotoxicologie, l'identification des protéines par homologie ne permet pas de détecter les protéines les plus pertinentes pour comprendre le mode d'action des polluants ou de proposer de nouveaux biomarqueurs. Une alternative à cette approche est donc nécessaire.

4. Valorisation sous la forme de revue bibliographique

Cette partie présente deux articles de synthèse bibliographique réalisés au cours de ces travaux de thèse.

La première publication (**Publication 1**) est une revue critique s'intéressant, chez les invertébrés, à l'utilisation des biomarqueurs en surveillance environnementale ainsi que les différentes limites freinant leur utilisation en routine. Les avancées dans le domaine des omiques permettent de lever certains de ces obstacles, notamment par la possibilité de développer par protéogénomique de nouveaux biomarqueurs spécifiques de ces organismes d'intérêt. De même, la validation et l'application de ces outils peut être portée par une approche de protéomique ciblée.

La deuxième publication (**Publication 2**) se place dans un contexte plus large de l'utilisation des espèces non-modèles en protéomique. En effet, les limites précédemment énoncées pour l'interprétation des données de spectrométrie de masse en l'absence de bases de données protéiques impactent de nombreux champs de recherche en biologie : microbiologie, agronomie ou écologie. Cette partie, valorisée sous la forme d'une revue bibliographique dans un numéro thématique « Proteomics of non-model organisms » du Journal of Proteomics, s'intéresse aux opportunités offertes par la protéogénomique pour l'étude du protéome des organismes non-modèles.

4.1. Les outils protéomiques pour le renouveau de l'approche biomarqueurs en écotoxicologie

Publication 1 :

Next-generation proteomics : towards customizied biomakers for environmental biomonitoring

Version auteur de l'article publié dans *Environmental Science and Technology*, 2014, 48 (23) : 13560-13572
La version finale est disponible sur http://pubs.acs.org/doi/abs/10.1021/es501673s

Judith Trapp [a,b], Jean Armengaud [b], Arnaud Salvador [c], Arnaud Chaumot [a], Olivier Geffard [a]
[a] Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, CS70077, F-69626 Villeurbanne, France
[b] CEA, DSV, IBEB, Lab. Biochim. System. Perturb., Bagnols-sur-Ceze, F-30207, France
[c] UMR 5280, Institut des Sciences Analytiques, Université de Lyon 1, F-69100 Villeurbanne, France

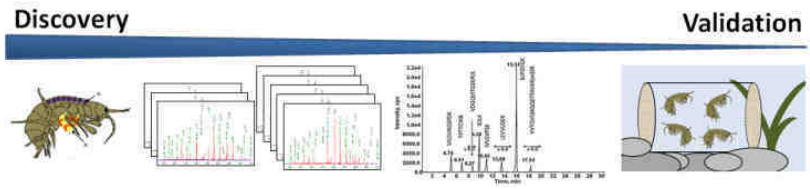
Résumé :

En raison de leur représentativité écologique, les invertébrés sont des organismes couramment employés dans le cadre de l'évaluation écotoxicologique. Pourtant jusqu'à présent, les biomarqueurs employés chez ces organismes résultent d'un transfert direct d'outils développés chez les vertébrés, et cela malgré une importante distance évolutive. Afin d'améliorer le suivi de la qualité biologique des milieux, des biomarqueurs spécifiques de ces organismes d'intérêt doivent être développés. Les technologies de protéomique nouvelle génération permettent l'identification de protéines impliquées dans des fonctions physiologiques clés ou encore dans la défense face à un stress environnemental, spécifiques de l'organisme d'intérêt. Toutefois, l'investissement nécessaire au niveau analytique restreint leurs usages à une approche de découverte. La validation et l'utilisation en routine de biomarqueurs dans le cadre de réseau de biosurveillance requièrent l'utilisation de techniques de spectrométrie de masse plus conventionnelles. Cependant, l'utilisation des approches protéomiques sur les organismes d'intérêt en écotoxicologie reste un défi, en raison de l'absence de bases de données protéiques, limitant l'analyse à l'observation de protéines ubiquitaires et conservées au sein de l'arbre du vivant. Ces limites ainsi que les différentes stratégies permettant d'y répondre sont discutées. L'utilisation de nouveaux outils, tels que la protéogénomique ou la protéomique ciblée devraient permettre le développement de nouveaux biomarqueurs, spécifique des organismes d'intérêt en écotoxicologie et applicable pour un suivi en routine des milieux.

Mots clés : Biomarqueurs, biomonitoring, protéomique, spectrométrie de masse, invertébrés.

Next-generation proteomics: towards customized biomarkers for environmental biomonitoring

Abstract



Due to their ecological representativeness, invertebrates are commonly employed as test organisms in ecotoxicological assessment. However, to date, biomarkers employed for these species were the result of a direct transposition from vertebrates, despite deep evolutionary divergence. In order to gain efficiency in the diagnostics of ecosystem health, specific biomarkers must be developed. In this sense, next-generation proteomics enables the specific identification of proteins involved in key physiological functions or defence mechanisms, which are responsive to ecotoxicological challenges. However, the analytical investment required restricts use in biomarker discovery. Routine biomarker validation and assays rely on more conventional mass spectrometers. Here, we describe how proteomics remains a challenge for ecotoxicological test organisms because of the lack of appropriate protein sequences databases, thus restricting the analysis on conserved and ubiquitous proteins. These limits, and some strategies used to overcome them, are discussed. These new tools, such as proteogenomics and targeted proteomics, should result in new biomarkers, specific to relevant environmental organisms and applicable to routine ecotoxicological assessment.

Keywords: Biomarker, biomonitoring, proteomics, mass spectrometry, invertebrates

1. INTRODUCTION

As pioneers in biosurveillance, Kolkwitz and Marsson[1] proposed, in 1909, a saprobic classification system for freshwater ecosystems, based on the presence/absence of indicator species with differing organic pollution tolerances. Biocenotic indices are now the main metric used for global ecosystem quality assessment. However, because these indices reflect established modifications of community structure, such as species representations, they cannot supply early signals of disturbance. Furthermore, these metrics are not able to provide information about the mechanistic processes of disturbance, nor to clearly establish their origins (*e.g.*, habitat modification or chemical quality degradation)[2-5]. While chemical analysis is a useful tool to address this question, it is still impossible to quantify all chemical compounds and their associated degradation products, to assess their bioavailabilities and to predict their conjugate effects on biota. Hence, biological indicators have been proposed that can be assessed at the sub-individual level, in order to link the presence of chemical compounds and their effects on biota by detecting sub-lethal changes. These indicators are designed to be an early warning signal for ecosystem degradation. Similar to human medical diagnoses, key proteins involved in the molecular response mechanisms related to xenobiotic toxicity or homeostasis can be used to provide evidence of exposure to, and/or effects of, one or more chemical pollutants on sentinel organisms. Biomarkers based on protein measurements are particularly interesting because proteins are the molecular effectors of biological processes. Their modulation is thus more likely to be linked to the effective impact on the physiology of the organism.

During the past three decades, considerable efforts have been made to develop and apply biomarkers in environmental risk assessment, but criticisms of their relevance or usefulness have been raised by the scientific community[6]. Nevertheless, during the last decade, ecotoxicological hazard assessment using protein-targeted approaches has evolved considerably, due mainly to the development of proteomic techniques. Next-generation proteomics allows the characterisation of an entire set of proteins (proteomes) from cell lines, tissues, or even organisms[7]. In ecotoxicology, these techniques have been employed for the determination of protein sequences and abundances, and for the inference of biological functions.

This review attempts to provide a critical analysis of current approaches based on protein-targeted methodologies in ecotoxicology, with a specific viewpoint on biomonitoring perspectives. The paper focuses on invertebrates, a major component of biocenosis, representing 95% of all animal

species[8] and consequently playing a major role in ecosystem function. Illustrated with numerous examples of protein biomarkers developed for the amphipod *Gammarus*, an ecotoxicologically relevant genus, we describe the current use of biomarkers in environmental management. Proteomic methodologies applied to aquatic invertebrates are reviewed and, finally, a viewpoint is developed on how novel strategies of next-generation proteomics should result in the generalisation of specific biomarker use in the assessment of ecosystem health.

2. CURRENT USE OF BIOMARKERS IN ECOTOXICOLOGICAL ASSESSMENT

We used research conducted on the *Gammarus* genus to exemplify the possibilities in biomarker assessment, as i) a significant number of publications on this genus, covering a wide range of stressor and biomarker typologies, are available, allowing us to sketch a comprehensive picture of the use of invertebrate-related biomarkers in ecotoxicology [9-37], and ii) a robust, active biomonitoring strategy, based on caged organisms, has been developed for this biological model[14, 20, 38-41]. Based on the twenty-nine publications detailed in **Table S1**, a quantitative overview of biomarkers applications is provided, also illustrated in **Figure 1**.

Regarding biomarker categories (**Figure 1 Panel A**), biomarkers of effect have been documented predominantly, accounting for 70% of the different applications. While some biomarkers allow the specific identification of exposure to a class of xenobiotics or alterations of physiological function, the majority of biomarker applications monitor a general response to disturbance: 10% detoxification activity, 20% antioxidative defence and 22% general stress response. Three-quarters of the experiments in this domain have been conducted in controlled conditions in the laboratory. Regarding the stressors that have been tested, four-fifths of the studies have been conducted in a perspective of the environmental hazard assessment of chemical contamination, while the remainder have been carried out on ecophysiological challenges, in a more ecological perspective such as understanding environmental factors controlling species distribution[10, 33]. Thus, biomarkers are mainly applied in a predictive approach in the laboratory. Indeed, while the biomarker approach has been incorporated into several pollution-monitoring programmes, their systematic and large scale application is rare. Several authors have already attempted to explain why biomarkers have not been applied to environmental assessments[6, 42, 43].

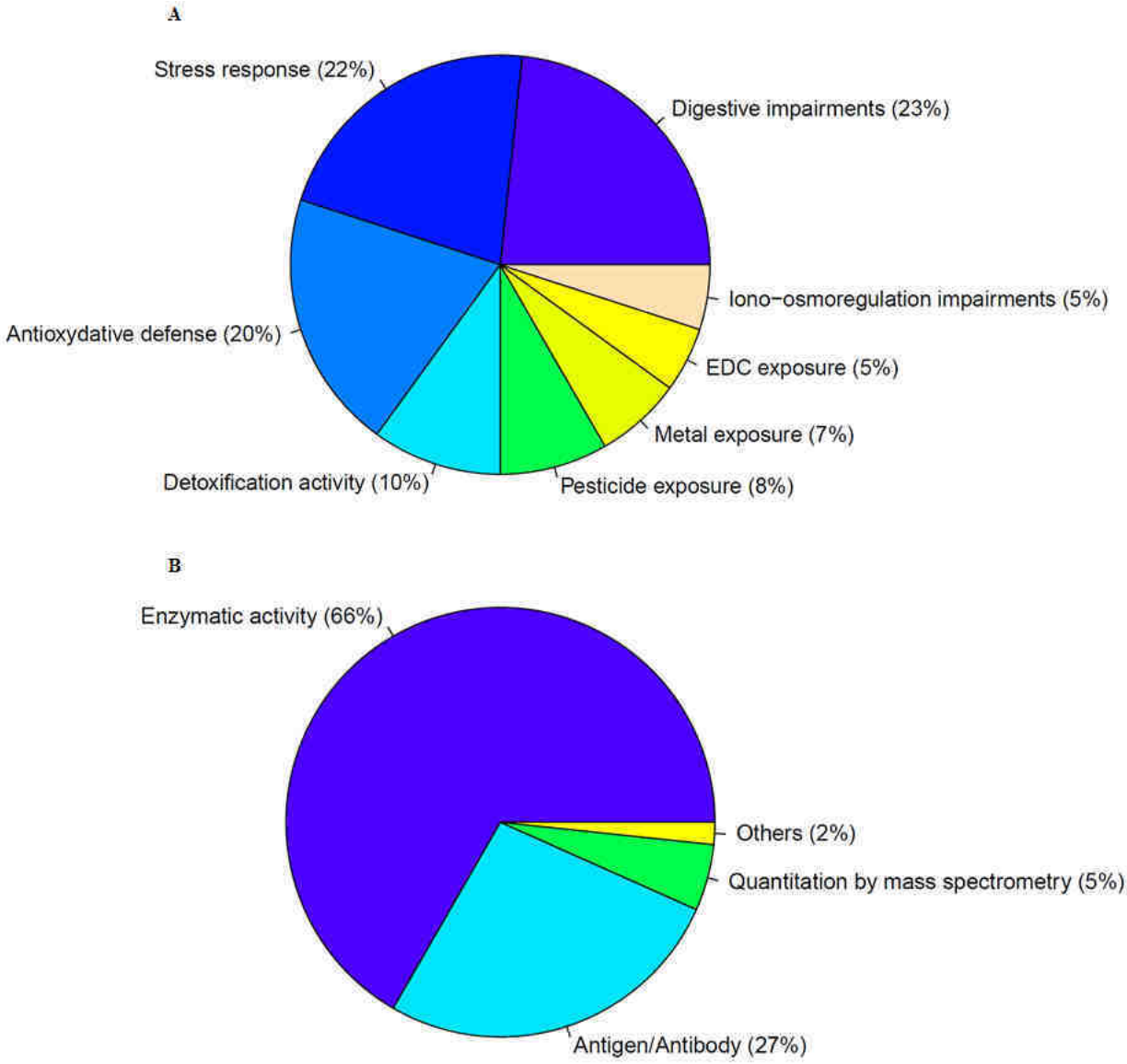


Figure 1: Overview of biomarker applications using *Gammarus* species

(A) Ratio of employed biomarkers by category among the different examples (n=60). (B) Methodology for analytical measurement among the different examples (n=60). Publications (n=29) employed in this meta-analysis are listed in Table S1.

Methodological constraints

Sensitivity and specificity: Enzyme-linked immunosorbent assays (ELISA tests) are the gold standard for routine measurement of biomarkers. Based on the antigen/antibody reaction, they are highly specific and sensitive for the detection and quantification of a targeted protein in a complex mixture, but the molecular structure of the protein must be characterized and known. Moreover,

the development of a robust ELISA test requires substantial resources to purify the protein, produce a specific antibody and validate its specificities. The return on investment of this type of development clearly depends on the number of tests to be performed. While economic interests for vertebrate toxicology related to the biomedical and agronomic fields offer resources to develop this kind of method, funding opportunities for invertebrate test organisms in ecotoxicology are limited. The ELISA assay is poorly transferable between species because of its high specificity towards protein conformation, making it sensitive to sequence divergence across phylogenetic distance. For gammarids, ELISA tests are not available for the measurement of any biomarkers. Nevertheless, one-quarter of the biomarker applications (**Figure 1, Panel B**) are based on the antigen/antibody reaction: heat-shock proteins (HSP) proteins by western blot analysis (22%); metallothionein-like proteins (MTLP) by differential pulse polarography (3%); or the sodium pump Na^+/K^+ATPase (NKA) by immunocytochemistry (2%). Remarkably, reaction specificities and cross-reactivities have not been characterised for these biomarkers, challenging their reliabilities. Ultimately, direct protein quantification is rare and indirect assays for monitoring protein activity remain one of the most widely used alternatives for measuring protein biomarkers. Indeed, these account for two-thirds of the biomarkers developed for *Gammarus* (**Figure 1, Panel B**). This alternative strongly limits the possibilities for developing non-enzymatic biomarkers such as vitellogenin-like protein (Vtg), MTLP or hormonal receptors.

With a solid knowledge of the molecular players involved in vertebrate biology, specific and robust biomarkers were developed. Later, when these were proven to be useful for monitoring specific exposures, or early warning indicators of specific diseases, they were applied directly to invertebrate models, with no systematic biochemical and pharmacological characterisation. However, distinct enzyme isoforms may have different sensitivities to substrates and/or inhibitors. One of the most fully documented examples is the measurement of the activity of cytochrome P450 (CYP450)-dependent monooxygenase. This enzyme is involved in phase I of xenobiotic biotransformation in vertebrates and its activity can be monitored by the degradation of ethoxyresorufin by ethoxyresorufin-O-deethylase (EROD). For fish, EROD activity in response to polycyclic aromatic hydrocarbons or dioxins has been studied intensively and proved to be responsive[44]. Along the lines successfully demonstrated for fish, it was also monitored for crustaceans exposed to polycyclic aromatic hydrocarbons. Although some EROD activities were reported, no crystal-clear induction has been observed and there is currently no consensus on this

biomarker relevance for crustaceans; the metabolism of the substrate may be explained by the involvement of other CYP families[45, 46].

Reproducibility: Reproducibility is one of the main principles of scientific methodology, and is essential for the reliable, large-scale biomonitoring of water bodies. To date, only two biomarker assays have been standardised by the International Organization for Standardization: vitellogenin (ISO 23893-3:2013) and EROD activity (ISO/TS 23893-2:2007). Both of these assays are designed for monitoring in fish, by means of ELISA in the plasma and fluorimetry in the hepatic microsomal fraction, respectively. In the absence of a standardised analytical protocol, which is mandatory for experimental consistency and data analysis, an accurate comparison of different studies, sampling dates and sites is challenging.

Multi-biomarker feasibility: Among the different biomarker applications using gammarids (*n*=29), more than half were conducted using more than one biomarker. Indeed, for an integrative ecosystem health assessment, a multi-biomarker approach, associated with a wide range of biological responses, and representing a continuum from effects of exposure to damage, must be used[47]. In the perspective of a multi-biomarker deployment in routine biomonitoring, a multiplexing methodology allowing the monitoring of a panel of biomarkers in a single biological sample, in a single run, would achieve a high-throughput and be cost-effective. However, for any biomarker, a specific analytical procedure in terms of homogenisation buffer, reaction medium and/or wavelength measurement is required. The multiplication of the protocols employed is extremely resource-consuming (time, personnel, finances and organisms).

Interpretation drawbacks

Relevance for impact prediction: Basic toxicology is based on the definition of a dose–response relationship, and a biomarker response should reveal the magnitude of the response of the organism to the contaminant[48]. In this respect, the dose–response relationship must be characterized in order to distinguish alterations that reflect homeostasis from those that are early indicators of an adverse effect, but this is rarely performed.

One cornerstone of the biomarker philosophy is to provide an early warning for individual or higher-organisation-level disorder. First, biomarkers are intended to detect sub-lethal changes (capacity for precocity) and therefore candidates must be sensitive to the presence of contaminants. To be qualified as “ecological biomarkers”, the biomarkers must then have a strong

relation to individual fitness, given that their measurement should offer a useful indicator of ecologically relevant effects (capacity for predictiveness)[6]. Several biomarker responses show a clear relationship with individual fitness, as described by Xuereb *et al.* [37], where compartmental (locomotion and feeding activity) alterations were linked to acetylcholinesterase (AChE) inhibition. Thus, a multi-biomarker approach is necessary for an integrative assessment from exposure to pathological disorder. In addition, prior to application in the field, a substantial technical investment in the laboratory is required for the accurate interpretation of the biomarker response.

Functional role: While most of the biomarkers currently employed in ecotoxicology using invertebrate models have resulted from a transposition of vertebrate toxicology, few studies have been directed towards validating the functional role of proteins before they are applied as biomarkers in new test organisms[20]. One of the most explicit examples of this is the measurement of yolk protein as a biomarker of the impairment of reproduction. Yolk protein is synthesised and progressively accumulates in oocytes during oogenesis (vitellogenesis), in response to oestrogens in vertebrates. While the gene encoding Vtg is present in males, it remains silent or poorly expressed under normal physiological conditions, and its induction has thus, for several years, been monitored as a biomarker of endocrine disruption in fish[49]. Consequently, Vtg induction in aquatic invertebrates has been proposed and employed as a biomarker of exposure to oestrogenic compounds, but this has produced contradictory results (reviewed by Matozzo *et al* [50]). Indeed, because of the high divergence of the endocrine system, chemicals that act as endocrine disruptors in vertebrates may have no direct homology of effects on the hormonal system of invertebrates. For crustaceans, the hormonal machinery is based on two hormones: ecdysone and methyl farnesoate, while the oestrogen-signalling pathway is not relevant in this phylum[51]. Furthermore, as suggested by molecular studies, invertebrate Vtgs are quite different from vertebrate Vtgs, both in terms of their sequences and in terms of their functional specialisation as egg-yolk proteins within the large lipid-transfer-protein superfamily which, for instance, also ensures functions of defence, healing and growth in crustacean decapods[52, 53]. In a large-scale field study using *G. fossarum* males, Jubeaux *et al.*[20] showed that the pleiotropic function of Vtg precludes its use as a specific male endocrine-disruption biomarker. Similar observations were made for the marine amphipod *Echinogammarus marinus*, where expression of the orthologous gene of *G. fossarum* Vtg-like protein was monitored in females, males and intersex males[54]. These results clearly demonstrated the limits to the crude exploitation of invertebrates as vertebrate surrogates for ecotoxicological

assessment, and to the direct transfer of the associated tools. As demonstrated in **Figure 1 Panel A**, the majority of the biomarkers used are well conserved and ubiquitous proteins in the animal kingdom (stress proteins, antioxidative or digestive enzymes), but despite preservation of the protein function, physiological manifestations of biomarker modulation can also differ between species. For example, although in mammals AChE inhibition leads to asphyxia because of muscle tetany, in aquatic organisms, oxygen uptake is more passive (*i.e.* oxygen transfer through the gills) and AChE inhibition is instead associated with deleterious effects on mobility rather than mortality[37]. In conclusion, for an accurate assessment of the health of a sentinel organism, biomarker transposition from vertebrates to invertebrates must be better supported with strong molecular data in invertebrates, and invertebrate-specific biomarker candidates should be identified and characterised to gain insights into more specific and accurate biomonitoring.

Interpretation and validation: Our survey shows the lack of biomarker utilisation in a biomonitoring context. We noted above that the application of biomarkers is technically challenging, and their interpretation in natural conditions is also arduous. Prior to the use of a specific biomarker in routine biomonitoring, several criteria must be defined in terms of i) discrimination potential, ii) intrinsic variability and iii) reference level, for a robust assessment of anthropogenic contamination. Ideally, a biomarker should only be modulated by contaminants, but abiotic and biotic factors also influence the outcome of the assay, making it difficult to clearly discriminate a stressor-induced response from intrinsic variability. Thus, most of the biomonitoring studies (60 %) listed here have been conducted “actively”. Active biomonitoring is based on in situ implantation of standard calibrated organisms by a caging procedure, which reduces the variability compared to a field population. In order to define “standard” organisms with the lowest biomarker variability, the influence of biotic factors such as sex, reproductive status or size/age[12, 20, 21, 55, 56] should be explored.

Although the caging procedure is an attractive method for obtaining the lowest possible variability related to biotic factors, the influence of abiotic factors is intrinsically related to the field context and therefore cannot be controlled, but should be taken into account. For large-scale biomonitoring programme that compare multiple stations in time and space, a baseline level of the biomarker response as an absolute value must be defined by monitoring the full annual and spatial variability in reference stations[57]. Ideally, this survey must be conducted in different reference stations with contrasted hydrogeological characteristics[20, 58]. While Jubeaux et *al.*[21] hypothesise a

possible use of Vtg as an endocrine-disruptor-exposure biomarker for *G. fossarum* males, after laboratory exposure, its application in the field turns out to be inconclusive because of high inter-individual variability, such as induction by unidentified environmental factors [20]. On the other hand, for AChE, digestive enzymes or genotoxicity, this strategy made it possible to define absolute reference values and confirmed their discrimination potentials[12, 58, 59].

Existing methodologies of biomarker measurement restrict their use and interpretation and consequently are not favourable to their application in routine biomonitoring. Moreover, assessment of an organism's health using biomarkers is a hypothesis-driven approach, as few of the parameters studied are intended to be responsive to the disturbance, mechanistic conclusion usually being derived from vertebrate organisms. As a result, significant disturbance signals can be overlooked. To date, knowledge on invertebrate gene products has impeded specific biomarker development for these organisms. Consequently, one of the only alternatives has been to employ biomarkers developed in vertebrates. However, molecular ecotoxicology is evolving with the recent development of new techniques, especially those from the metagenomic, transcriptomic, and next-generation proteomic areas. Proteomics allows the specific characterisation of an entire protein set (proteomes) of a given cell line, as well as tissues and, more recently, whole organisms[7]. Compared to genomics and transcriptomics, proteomics identifies and measures the quantities of the real effectors of the physiological responses. It can, consequently, offer a hypothesis-free investigation into stress-response mechanisms, *i.e.* cataloguing, with no *a priori*, molecular targets responsive to specific stressors.

3. CURRENT ECOTOXICOLOGICAL ASSESSMENT USING PROTEOMICS TOOLS

Improvements in hazard assessment*:* Proteomics emerged in the 1980s with the development of high-resolution gel electrophoresis, two-dimensional polyacrylamide gel electrophoresis (2DE-PAGE), which allowed the separation and visualisation of proteins in complex mixtures[7, 60]. The conventional approach for proteomic studies in ecotoxicology is based on the classical 2DE-PAGE technique and a comparison of the most abundant proteins present in samples from controls and environmentally challenged organisms. Technically, the proteins are resolved in a first dimension at their isoelectric point, then in the second dimension based on their molecular weight. Once

separated, the protein spots are selected based on quantitative criteria (change in their abundance between the experimental conditions).

A biomonitoring case study conducted by Vioque-Fernandez *et al.* [61] investigated the discriminating potential between the responses of well-established biomarkers and proteomic investigations using 2DE-PAGE. *Procambarus clarkii* crayfish were collected at eight sites in the Donana National Park (Spain) and twelve digestive-gland biomarkers were assayed: antioxidant enzymes (catalase, glutathione peroxidase, glucose-6-phosphate dehydrogenase), phase I and phase II biotransformation enzymes (EROD, glutathione S-transferase *i.e.* GST, reduced/oxidised glutathione, glyoxalase II), esterases (AChE, carboxylesterase), metallothionein and lipoperoxidation. In parallel, the authors analysed gill protein expression on 2DE-PAGE. While seven out of twelve biomarkers roughly discriminated reference sites from contaminated sites, the proteomic approach allowed a more discriminative classification from clean/low to highly polluted with thirty-five significantly altered proteins, establishing its superiority as a diagnostic tool. This result is logically due to the massive screening of this experimental methodology. Moreover, the proteins resolved in 2DE-PAGE are truly specific to the test organism, while this affirmation cannot be applied to the biomarkers used. However, this approach was only descriptive and did not directly explain the molecular mechanisms related to the disturbance, nor did it clearly discriminate its origins, because in this study the proteins remained unidentified.

Later, the development of mass spectrometry (MS) technology in protein chemistry, combined with large-scale nucleotide sequencing (expressed sequence tags and genomic DNA), allowed the development of whole-genome proteomics[62]. Briefly, a first protein MS identification strategy was based on “peptide mass fingerprinting”. Protein spots were selected and excised from the 2DE-PAGE gel, then proteins were digested with trypsin and analysed using MALDI-TOF MS. A list of experimental peptide masses with their corresponding intensities was produced and compared with theoretical peptide masses, which are generated from *in silico* digestion of protein sequences present in the database and annotated from DNA sequencing information[7, 60]. More recently, tandem MS has made it possible to measure the peptide masses as well as to obtain information relating to the amino-acid composition of each peptide. In this case, the MS/MS spectra are assigned to peptide sequences based on fragmentation prediction and protein sequence databases.

Table S2 lists the latest achievements in discovery proteomics focussed on aquatic invertebrates and ecotoxicology uses. Among the twenty-two studies displayed in our analysis, the majority of these took place in a context of the risk assessment of marine water bodies (60%), a quarter of

them (27%) in a freshwater context and the remainder in an estuarine context. Molluscs, especially bivalve[63-79], have been by far the most-employed taxa (77%), while only one study using a gastropod has been reported[80]. This is not surprising because marine environments are over-represented in our analysis and the utilisation of bivalve organisms is predominant for this ecosystem[81]. Finally, three studies (14%) employed crustaceans [82-84] and one a polychaete[85]. Most of the studies aimed to elucidate the stressor mode of action and the molecular mechanisms displayed by the organism in order to manage the disturbance, by comparing proteomes from exposed and unexposed organisms. In addition, unknown stressor-specific biomarkers can be identified[48, 84, 86]. Proteomic investigations using laboratory workhorses such as *Daphnia magna* and *Caenorhabditis elegans* have also been carried out, but rather for the generation of protein datasets[87, 88] or physiological investigations[89] than for ecotoxicological outcom*es.* Also*,* in this review, special emphasis has been given to "non-model organism" with few molecular resources available.

While proteomics allows non-hypothesis-driven experiments, ideal for an exhaustive ecotoxicological assessment, four-fifths of studies have been conducted in laboratory-controlled conditions. Only passive biomonitoring studies have been conducted, with the associated interpretation limitations described above. Proteomics holds promise for improved ecotoxicological assessment at the sub-individual level. However, comprehensive studies have extensively evaluated the applications of proteomics and have highlighted their limits in ecotoxicology[48, 90, 91]. Here, using the studies listed in **Table S2**, we briefly summarise the current limitations for practical applications in a biomonitoring perspective.

Substantial workflow*:* One of greatest challenges in proteomics is the substantial workflow inherent to sample and data processing, which includes several steps: i) protein extraction, ii) proteome separation, iii) data acquisition and iv) data analysis[92]. In our review, we have found that all studies employed 2DE-PAGE. This technique is low throughput because it includes a large amount of sample handling: staining, protein spot excision and trypsin digestion. As a result, proteomic studies usually focus on a single-timepoint experiment. As demonstrated for the classic biomarker approach, low-throughput analytical methods do not allow progress in the realm of response interpretation and cannot reasonably be considered for routine applications. However, there is plenty of room for improvement given that shotgun proteomics has become increasingly popular in other biological fields and has already been employed in fish ecotoxicology[93, 94]. Shotgun

proteomics allows the identification of proteins in complex mixtures, without prior protein separation on 2DE-PAGE. It consists of trypsin proteolysis of the whole protein mixture, resolving the peptides by high-performance liquid chromatography (HPLC), and then peptide identification with high-throughput tandem MS[62]. This methodology considerably simplifies the workload and allows the routine identification of hundreds to thousands of proteins.

Protein quantification: When 2DE-PAGE is employed, quantification is performed prior to protein identification. This technique is poorly adapted for protein quantification as i) several proteins may contribute to a single spot, ii) it has poor general reproducibility, and iii) it has poor proteome coverage, while the quantitative patterns are essential for predicting the impact on the organism's physiology. Differences in staining of a given protein spot may result from the contribution of different proteins present in the same spot that will not be deconvoluted. Expert quantitative approaches based on the use of fluorescent dyes are difficult to perform and are restricted to the most abundant proteins visible on the gel. Moreover, the gel approach is limited to some of the soluble proteins, since the membrane and basic proteins are excluded from such analysis[62]. In fourth-fifths of the studies, quantification was performed by densitometry, *i.e*. colloidal Coomassie blue and silver staining, which rules out exact quantification, such as the detection of low-abundance proteins[95]. In 9% of the studies, fluorescence staining was employed, providing a more sensitive quantification than the other classical dye staining [96]. Finally, while 2DE-PAGE was employed in 91% of the studies, the remaining 9% used differential gel electrophoresis (2D-DIGE). 2D-DIGE allows the co-separation of proteins isolated from different treatments in a single gel, where each sample is labelled with a specific fluorophore before the samples are mixed together. In parallel with the development of shotgun proteomics, other quantification methods have been developed using labelling and non-labelling techniques (reviewed in Armengaud[62]). In this case, protein quantification is performed in parallel with protein identification. In any case, without absolute quantification available in targeted proteomics, these discovery-based approaches remain largely descriptive and are not suited to water quality diagnosis.

Protein identification: The number of proteins identified is highly dependent on the analytical method employed. New developments in shotgun proteomics enable a deeper characterization of proteomes. For example, Washburn *et al.*[97] performed one of the largest proteome analyses of the yeast *Saccharomyces cerevisiae* using a method coupling multi-enzymatic digestion, two-dimensional chromatography and tandem mass spectrometry. While previous studies using 2-DE

methods allowed the identification of a few hundred proteins (400 spots)[98], the number of proteins identified by Washburn *et al.* was around 1500. Furthermore, protein classes that are usually under-represented in proteome analysis were detected. Although, for model organisms, the protein identification rate has been enhanced by new analytical developments, this remains a great challenge for non-model organisms, such as those employed in ecotoxicology. Indeed, protein identification using MS is based on a database where the exact protein sequences are listed. Unfortunately, for most of the species used in ecotoxicology, this information is lacking. Homology-driven proteomics using protein cross-species matching is one alternative to the lack of a properly annotated genome, but this approach only detects the most conserved proteins, which are generally related to protein synthesis and folding, general metabolism, and structural proteins. Many studies conducted over the last decade have run into problems matching peptide masses, because of low scores or false-positive matches[62]. Among the studies considering protostome invertebrates displayed in **Table S2**, the average rate of protein identification has been about 54%, a reasonable yield that tends to improve each year. **Figure 2** presents the ten most frequently detected proteins in our literature survey. Proteins involved in ATP supply and in the maintenance of cytoskeletal structure are highly represented. Furthermore, as demonstrated by transcriptomic studies on *Daphnia pulex*, lineage-specific genes have been shown to be amongst the most responsive to environmental challenges[99]. Consequently, cross-species matching for protein identification does not make it possible to decipher relevant and specific modes of action in ecophysiological and ecotoxicological studies. Monsinjon and Knigge[91] even consider that, in ecotoxicology, protein identifications are first determined by protein abundance, sequence conservation and prevalence in databases, and then by a specific association with a certain condition. It is clear that a new alternative to homology-driven proteomics should be developed.

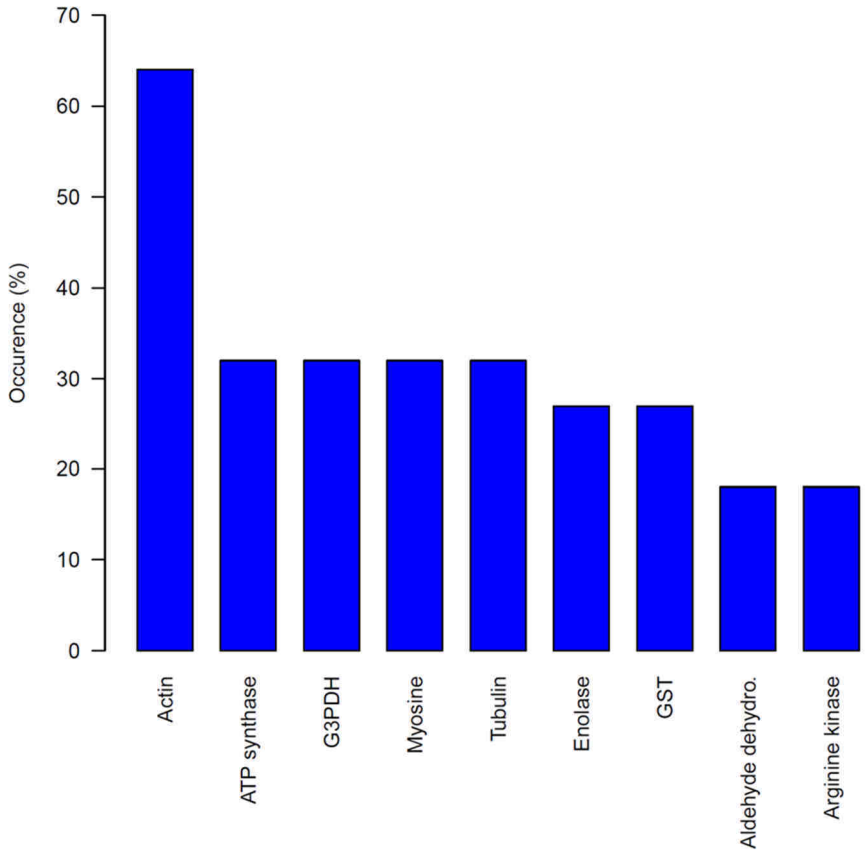


Figure 2: The ten most ferquently occuring proteins in proteomics studies in aquatic ecotoxicology using invertebrates

Studies (n=22) are displayed in Table S2. Aldehyde dehydro., aldehyde dehydrogenase; G3PDH: glyceraldehyde-3-phosphate dehydrogenase

Data interpretation: The identification of proteins by mass spectrometry has proven to be robust as soon as several peptides have been identified with low p-values. The significance of the results can be easily assessed from the estimation of false-positive rates, taking into consideration a decoy database methodology for example. Strategies, in terms of the statistical evaluation of quantitative proteomic data, are numerous and are highly dependent on the methodology employed: gel-based *vs.* gel-free separations, label *vs.* label-free strategies. A growing collection of software packages is available and data can be processed by modular building blocks or monolithic analysis software. For 2DE-PAGE, normalisation of protein volume and intensity is performed to compare spots from different gels and then a fold change is calculated to select the most interesting spots deserving mass spectrometry-based identification. In label-free shotgun proteomics, quantification is based on spectral counting, *i.e*. the number of MS/MS spectra assigned to a protein. For comparing two conditions or for following the proteome dynamics along a kinetic experiment, the PatternLab software has been proposed, with different statistical tests[100]. Nevertheless, one of the major

downsides to proteomic studies is the limited number of replicates, because of the substantial workflow and associated financial constraints. Thus, statistical power is lowered, exacerbating the influence of confounding factors.

Tools for the biological interpretation of proteomic data are mostly shared with the transcriptomic community, such as gene ontology analysis or pathway focus interpretation. Protein function is predicted on the basis of sequence preservation by combining available sequences and functional information from various taxonomic groups. This information is, for the most part, directly derived from vertebrates, for which annotated genomes are more numerous. However, for invertebrates, confidence in these predictions is relatively limited because of high diversification rates and longer evolutionary times; therefore, homologues in distant organisms may have acquired different functions[101-103].

4. TARGETED PROTEOMICS IN SUPPORT OF BIOMARKER-BASED METHODOLOGY

The development of tandem MS and new concepts in protein quantification by the "targeted-proteomics" approach have allowed efficient and sensitive biomarker-based biomonitoring. In this section, we will develop two important themes: absolute quantification and cross-species transferability of biomarkers by tandem MS.

Absolute quantification: Protein quantification by MS was first developed for medical purposes. Although ELISA assays are commonly used for quantifying proteins, antibodies are not always available for new candidate protein markers discovered by means of "omics" methods[104, 105]. Shotgun proteomics is a powerful method for discovering candidates in complex protein mixtures but, because of its exhaustive nature, this approach does not allow the precise quantification of a given biomarker. Targeted proteomics aims to specifically measure the abundance of a given set of peptides. These approaches have evolved considerably over the last decade, with new mass spectrometers adapted to such strategies and also bioinformatic pipelines. Selected reaction monitoring (SRM) is an LC-MS-based technique for the quantitation of a specific protein of interest. SRM has the ability to detect low amounts (to the pg/ml range) of a targeted analyte, with peptides as multi-indicators of proteins. Absolute quantification can be performed by the addition of an isotopically distinguishable internal standard (reviewed by Gallien *et al*[106]). This technique has already been applied to the quantitation of Vtg in *G. fossarum*[31] (**Figure 1 Panel B**). From the translated Vtg gene sequence obtained by polymerase chain reaction amplification of the genome

copy and DNA sequencing, putative proteolytic peptides and associated transitions (m/z values associated with the precursor and fragment ion of some peptides) were designed *in silico*. Relevant peptides were selected according to analytical criteria, such as specificity and sensitivity, and then validated by experimental measurement. Finally, when the best peptide to monitor using LC-MS/MS was selected and its biological relevance was confirmed, an isotopically labelled reference peptide was synthesised for absolute quantification[31].

Multiplexing: Targeted proteomics can also be performed using the multiple reaction monitoring (MRM) mode, where multiple peptides per protein and/or multiple proteins from the same biological sample can be simultaneously quantified. For example, Anderson and Hunter [107] designed a quantitative LC-MS/MS for tryptic peptide monitoring of 53 high- and medium-abundance proteins in a single human plasma sample. Multiplexing by mass spectrometry is in this case straightforward and offers interesting perspectives in ecotoxicology due to the low cost of the methodological development. Indeed, low-throughput and non-specific analytical methods severely limit the development of biomarkers by research laboratories as well as their use in routine biomonitoring by environmental agencies. Remarkably, a multiplexing methodology uses less biological material and consequently lower resource investment in terms of organism and experimental maintenance. With a high-throughput analytical method, questions during the development phase regarding the definition of the dose–response relationship or the reference range can be straightforwardly investigated. Where three large scale field experiments were required for the definition of reference values for Vtg[20], AChE[58] and digestive enzymes[12] in *G. fossarum*, a multiplexing methodology will make it possible to perform this step in a single procedure. When the development phase has been completed, ecotoxicological assessment using a multi-biomarker approach can be technically possible. Furthermore, the possibility of monitoring a wide range of biological responses, covering a continuum from various xenobiotic exposures to damage, provides the opportunity to improve environmental hazard assessment.

Cross-species transferability: As part of ecological risk assessment and environmental management, ecotoxicology needs to address the challenge of biodiversity. Each taxon possesses a particular sensitivity to toxicants and consecutive effects must be assessed in order to predict effects and changes on species composition[108, 109]. As a result, monitoring the responses of different species from distinct phylogenetic lineages is mandatory. This issue is highly challenging in invertebrates, for which the molecular divergence between lineages is more significant than in vertebrates[101]. However, recent investigations using comparative quantitative proteomics of the

nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster* showed that the abundance of orthologous proteins across species correlates remarkably well, better than protein abundance versus transcript abundance within each organism, or transcript abundance across organisms. Schrimpf *et al.*[110] hypothesised that orthologous proteins that were not found in similar abundance may reflect an evolutionary diversification of the orthologue roles. Consequently, with the possible evolutionary conservation of important pathways among the orthologues, mass spectrometry appears to be a promising way to propose time-saving and cost-effective biomarkers across species. Jubeaux *et al.* [111] developed a promising strategy for the development of shared biomarkers among several invertebrate taxa by the quantitation of Vtg-like protein. At first, they searched Vtg-sequenced invertebrate organisms in a public database. By comparing available Vtg sequences, conserved tryptic peptides were identified and those with shared MS/MS transitions that would give the best sensitivity were selected. Functional validation was performed by comparing proteomic signals according to the inter-gender ratio. They then demonstrated that the assay developed from Vtg-sequenced species could be applied to Vtg from unsequenced organisms. In a perspective of ecological quality monitoring, the common peptide sequences will overcome polymorphisms within species and variability across species, unlocking the need for new developments for each local population used as sentinel species.

5. PROTEOGENOMICS FOR PROTEIN DISCOVERY IN SENTINEL SPECIES

While MS-based proteomics is a promising technique for discovering new biomarkers and contributes to increasing biomarker-based biomonitoring, protein identification remains difficult for sentinel species because of the scarcity of specific protein sequence databases for most taxonomic groups. While sequencing the genome of organisms is the most comprehensive approach for creating a complete protein sequence database, its accurate annotation remains fastidious, especially when eukaryotes are considered. For these organisms, sequencing only mature RNAs is an attractive alternative for the rapid identification of protein sequences. RNA-Seq, a next-generation sequencing technology, is a cost-effective technique for the quick identification of protein-encoding genes amongst mature RNAs through a comprehensive coverage of the transcriptome by deep-sequencing [112]. This technique has been employed in several test organisms in ecotoxicology, such as the amphipod *Melita plumulosa*[113], the pond snail *Lymnaea stagnalis*[114] and the midge *Chironomus riparius*[115]. Thus, a protein database can be generated by six-reading-frame translation of the RNA-Seq nucleic acid sequences. The combination of genomics and

proteomics, namely proteogenomics, first emerged for genome annotation. Indeed, after sequencing, gene annotation, *i.e*. the prediction of open reading frames (ORF), is generally processed *via* an informatics pipeline and is therefore prone to errors. Thus, the generation of a "proteogenomic map" by mapping MS-certified peptides onto the genome allows the more accurate determination of genome structure by correcting ORFs, for example detecting unassigned reading frames[116]. Later, the term "proteogenomics" evolved, as it was employed in studies where proteomics and genomics were linked but not intended for genome annotation. Proteogenomics can now be employed for multi-omics projects, including the use of six-frame translation. Consequently, the use of an RNA-Seq-derived database is a straightforward strategy for discovering proteins in non-model organisms[112]. Furthermore, as for any organism without a genomic backbone, the *de novo* assembly of the transcriptome can lead to nucleotide sequence errors, resulting from chimeric sequences or the presence of premature termination codons. Hence, the use of biochemical data allows for the experimental validation of the existence of the gene products such as, for proteins of interest, specific data mining in order to remove errors arising from the sequencing and assembly of nucleotide sequences.

While proteogenomics allows protein discovery and enhances structural annotation, the inference of protein function is still problematic. Reliability regarding functional prediction by sequence similarity was soon challenged by the scientific community. In the 1990s, Eisen[117] proposed the improvement of functional prediction by reconstructing the evolutionary history of gene products. The usefulness of phylogenetic analysis can be illustrated with examples of invertebrate Vtgs, which are quite different from vertebrate Vtgs. Phylogenetic analysis resulted in the re-classification of decapods Vtgs as apolipocrustaceins because they are more closely related to vertebrate apolipoprotein B than to vertebrate Vtgs. On the other hand, decapods clotting proteins are true Vtg orthologs. Thus, direct assignment on the basis of sequence similarity can lead to false mechanistic hypotheses[118]. Another strategy for functional inference, based on the recent work of Shrimpf *et al.*[110], is to rely on comparative proteomics and the conservation of protein abundances among orthologs sharing similar biological functions. While this observation was first made two invertebrates, it was later confirmed in other organisms, ranging from bacteria to humans[119]. Thus, for a newly characterized organism, comparative analysis of homologs abundance can be informative for functional inference.

A downside of these prediction methods is the necessity to use only gene products with experimentally determined functions as the basis. Furthermore, at present, functional predictions

based on sequence comparisons may have limited impact in environmental studies and biomarker development. Although the proportion of lineage-specific genes in daphnia is expected to decrease when new crustacean molecular resources become available, it has been shown that daphnia orphans genes respond preferentially to environmental disturbance and, thus, are ideally suited as biomarker candidates. As highlighted by Nikinmaa and Rytkonen[103], combining "omics" studies with functional aspects is rarely performed in ecotoxicology, while these authors strongly encourage this, supporting the idea that it could make "*aquatic toxicologists forerunners in functional genomics as a whole*". Thus, researchers in ecotoxicology should first inspire themselves with strategies developed by physiologists, such as the comparative proteomics of physiological tissues or developmental stages. For example, Chen et *al*.[120] performed a comparative analysis on the silkworm *Bombyx mori* gonads in order to characterise proteins involved in sexual dimorphism and gametogenesis, resulting in the identification of 194 testis-specific proteins and 113 ovary-specific proteins. In the freshwater prawn *Macrobrachium rosenbergii*, significant work was carried out on its reproduction and development; for example, gene silencing of the putative androgenic hormone affected phenotypical gender differences such as spermatogenesis[121]. More recently, transcriptome sequencing was carried out on several larval and post-larval stages. Over-expressed transcripts in either stage were identified and further characterized by targeted analysis (*i.e*. RT-PCR)[122]. For an ecotoxicologically relevant organism, such strategies should result in the discovery of interesting biomarker candidates for reprotoxicity and Endocrine Disrupting Chemicals (EDC) exposure assessment. **Figure 3** proposes a conceptual framework for the development of biomarkers using proteomic tools, illustrated with our biological model, *Gammarus fossarum*. This proof of concept is oriented with the objective of developing biomarkers related to reproductive function as, in the past few years, significant work has been carried out by our research team for this purpose[20, 21, 59, 111, 123]. During the discovery phase, physiological tissues related to reproductive function (reproductive tissues and endocrine glands) were employed to establish an exhaustive protein catalogue using a combination of RNA-Seq and shotgun proteomics. A strategy using comparative proteomics between different tissues should then be used to uncover, without any *a priori*, new proteins involved in reproductive function[124]. As the physiological events related to reproductive features (*i.e.* embryogenesis, oogenesis, moult cycle, spermatogenesis) are well described, temporal proteome dynamics over each process should be characterised to indicate proteins with a consistent evolution along phenotypic modifications.

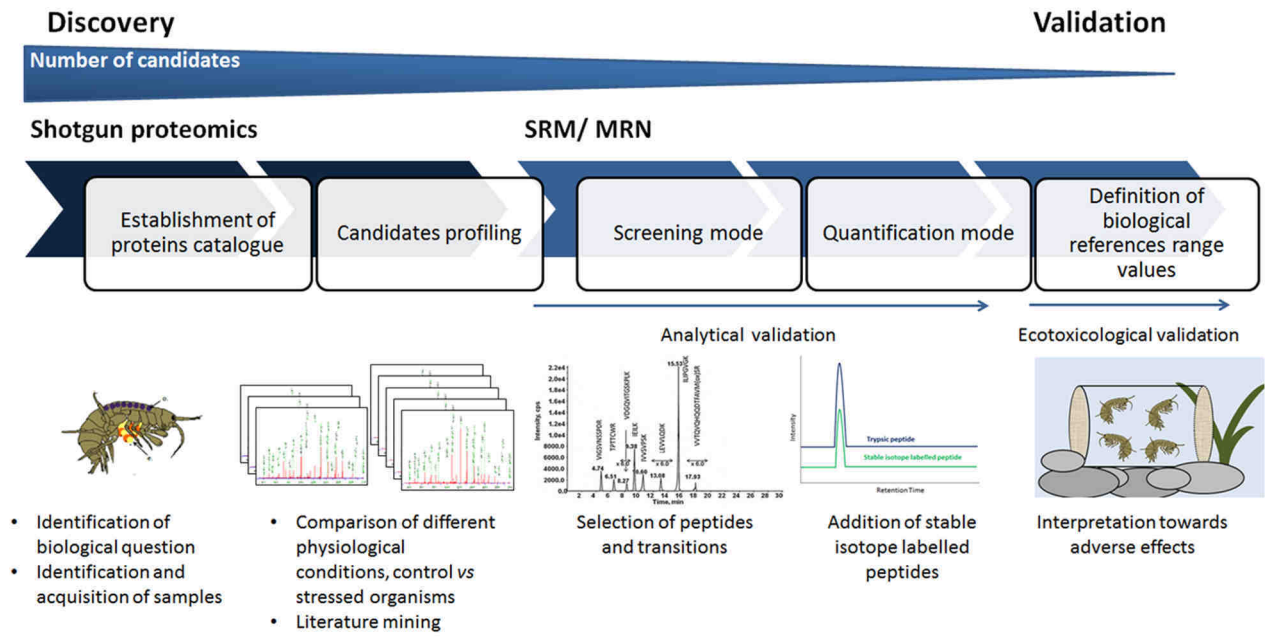


Figure 3: Conceptual framework for biomarker development using proteomics tools

Next, the comparison of proteomes from control organisms and organisms exposed to reprotoxic substances can be carried out. Ideally, this step should be conducted in association with additional investigations on physiological parameters (gamete production, histopathology) to identify adverse effects on reproductive function, a biological level of organisation relevant to ecological risk assessment[125]. At the end of this exhaustive discovery phase, a list of candidate biomarkers will be established for the development of targeted proteomic approaches. During the screening mode, relevant peptides should be selected according to analytical criteria (specificity, sensitivity, repeatability), as well as their potential in terms of transferability. Later, when the best peptides have been selected, stable isotope-labelled peptides can be synthesised and used for absolute quantification. Finally, when a robust analytical method is available, ecotoxicological validation should be conducted in the laboratory, in terms of the interpretation of adverse effects and the characterisation of confounding factors, and finally *in situ* for the determination of biomarker discrimination potential and reference value ranges.

The combination of novel strategies in next-generation proteomics now enables the consideration of multi-biomarker approaches based on direct and specific protein measurement and easily transferable to other test species. Initially, when fundamental knowledge of invertebrate molecular regulation was limited, few specific biomarkers had been identified for these test organisms.

Available biomarkers were, rather, the result of a direct transposition of vertebrates biomarkers. Because of deep evolutionary divergence, their potencies in discriminating contamination were limited. Currently, for virtually any organism, proteogenomics allows the discovery of specific proteins, and thus the development of customised biomarkers for ecotoxicological risk assessment. The inference of associated biological function is based on a combination of sequence comparison, comparative quantitative proteomics, phylogenetic analysis and experimental evidence. However, because of analytical requirements, such as the stochastic nature, global proteome profiling is restricted to the discovery phase. Targeted analysis by SRM is a powerful technique for the sensitive detection and quantification of a specific protein. Its application is becoming increasingly popular for biomarker validation. In ecotoxicology, one of the major problems for biomarker-based biomonitoring is the absence of a high-throughput analytical method for routine application to pollution monitoring programmes. As a result, the development of MRM assays is a promising strategy for overcoming feasibility problems related to the multi-biomarker approach. Despite these analytical developments, significant investments in terms of biomarker validation, *i.e*. characterisation towards impact prediction and *in situ* determination of intrinsic variability and reference levels, must be performed, but should gain in efficiency with increased yields of data acquisition.

Associated content: Details on the publications employed for the construction of metrics regarding biomarker applications and proteomics applications in aquatic ecotoxicology are displayed respectively in Table S1 and Table S2. This material is available free of charge via the Internet at http://pubs.acs.org

Acknowledgements: This research was funded by the Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (France) and the Commissariat de l'Energie Atomique et aux Energies Alternatives (France) through the Programme Transversal de Toxicologie. We thank the reviewers that provided helpful comments on our manuscript.

Abbreviations: 2DE-PAGE, two-dimensional polyacrylamide gel electrophoresis; AChE, acetylcholinesterase; CYP, cytochromes; DIGE, differential in-gel electrophoresis; EDC, Endocrine Disrupting Chemicals; ELISA, enzyme-linked immunosorbent assay; EROD, ethoxyresorufin-O-deethylase; G3PDH: glyceraldehyde-3-phosphate dehydrogenase; GST, glutathione S-transferase;

HSP, heat-shock protein; LC, liquid chromatography; MRM, multiple reaction monitoring; MS/MS, tandem mass spectrometry; MTLP, metallothionein-like proteins; NKA, Na^+/K^+ATPase; SRM, selected reaction monitoring; Vtg, vitellogenin.

REFERENCES

1. Kolkwitz, R.; Marsson, M., Ökologie der tierischen Saprobien. *Int. Rev. Ges. Hydrobiol. Hydrograph* **1909,** *2*, 126-152.
2. Attrill, M. J.; Depledge, M. H., Community and population indicators of ecosystem health: Targeting links between levels of biological organisation. *Aquat Toxicol* **1997,** *38*, (1-3), 183-197.
3. Clements, W. H., Integrating effects of contaminants across levels of biological organization: An overview. *J. Aquat. Ecosyst. Stress Recovery* **2000,** *7*, (2), 113-116.
4. Power, M., Recovery in aquatic ecosystems: an overview of knowledge and needs. *J. Aquat. Ecosyst. Stress Recovery* **1998,** *6*, (3), 253-257.
5. Depledge, M. H.; Fossi, M. C., The role of biomarkers in environmental assessment (2). Invertebrates. *Ecotoxicology* **1994,** *3*, (3), 161-72.
6. Forbes, V. E.; Palmqvist, A.; Bach, L., The use and misuse of biomarkers in ecotoxicology. *Environ Toxicol Chem* **2006,** *25*, (1), 272-280.
7. Graves, P. R.; Haystead, T. A., Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* **2002,** *66*, (1), 39-63; table of contents.
8. Barnes, D., *Invertebrate zoology*. 2nd ed.; Saunders: Philadelphia, 1968.
9. Bedulina, D. S.; Timofeyev, M. A.; Zimmer, M.; Zwirnmann, E.; Menzel, R.; Steinberg, C. E., Different natural organic matter isolates cause similar stress response patterns in the freshwater amphipod, Gammarus pulex. *Environ Sci Pollut Res Int* **2010,** *17*, (2), 261-9.
10. Bedulina, D. S.; Zimmer, M.; Timofeyev, M. A., Sub-littoral and supra-littoral amphipods respond differently to acute thermal stress. *Comp Biochem Physiol B Biochem Mol Biol* **2010,** *155*, (4), 413-8.
11. Brooks, S. J.; Lloyd Mills, C., Gill Na(+), K(+)-ATPase in a series of hyper-regulating gammarid amphipods: enzyme characterisation and the effects of salinity acclimation. *Comp Biochem Physiol A Mol Integr Physiol* **2006,** *144*, (1), 24-32.
12. Charron, L.; Geffard, O.; Chaumot, A.; Coulaud, R.; Queau, H.; Geffard, A.; Dedourge-Geffard, O., Effect of water quality and confounding factors on digestive enzyme activities in Gammarus fossarum. *Environ Sci Pollut Res Int* **2013,** *20*, (12), 9044-56.
13. Correia, A. D.; Costa, F. O.; Neuparth, T.; Diniz, M. E.; Costa, M. H., Sub-lethal Effects of Copper-Spiked Sediments on the Marine Amphipod *Gammarus locusta* : Evidence of Hormesis ? . *Ecotoxicol. Environ. Res.* **2001**, (2), 32-38.
14. Dedourge-Geffard, O.; Charron, L.; Hofbauer, C.; Gaillet, V.; Palais, F.; Lacaze, E.; Geffard, A.; Geffard, O., Temporal patterns of digestive enzyme activities and feeding rate in gammarids (Gammarus fossarum) exposed to inland polluted waters. *Ecotoxicol Environ Saf* **2013,** *97*, 139-46.
15. Dedourge-Geffard, O.; Palais, F.; Biagianti-Risbourg, S.; Geffard, O.; Geffard, A., Effects of metals on feeding rate and digestive enzymes in Gammarus fossarum: an in situ experiment. *Chemosphere* **2009,** *77*, (11), 1569-76.
16. Felten, V.; Charmantier, G.; Mons, R.; Geffard, A.; Rousselle, P.; Coquery, M.; Garric, J.; Geffard, O., Physiological and behavioural responses of Gammarus pulex (Crustacea: Amphipoda) exposed to cadmium. *Aquat Toxicol* **2008,** *86*, (3), 413-25.
17. Frank, S. N.; Godehardt, S.; Nachev, M.; Trubiroha, A.; Kloas, W.; Sures, B., Influence of the cestode Ligula intestinalis and the acanthocephalan Polymorphus minutus on levels of heat shock proteins (HSP70) and metallothioneins in their fish and crustacean intermediate hosts. *Environ Pollut* **2013,** *180*, 173-9.
18. Gismondi, E.; Rigaud, T.; Beisel, J. N.; Cossu-Leguille, C., Microsporidia parasites disrupt the responses to cadmium exposure in a gammarid. *Environ Pollut* **2012,** *160*, (1), 17-23.

19. Issartel, J.; Boulo, V.; Wallon, S.; Geffard, O.; Charmantier, G., Cellular and molecular osmoregulatory responses to cadmium exposure in Gammarus fossarum (Crustacea, Amphipoda). *Chemosphere* **2010,** *81*, (6), 701-10.
20. Jubeaux, G.; Simon, R.; Salvador, A.; Lopes, C.; Lacaze, E.; Queau, H.; Chaumot, A.; Geffard, O., Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: inconclusive experience. *Aquat Toxicol* **2012,** *122-123*, 9-18.
21. Jubeaux, G.; Simon, R.; Salvador, A.; Queau, H.; Chaumot, A.; Geffard, O., Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. *Aquat Toxicol* **2012,** *112-113*, (0), 72-82.
22. Karaouzas, I.; Cotou, E.; Albanis, T. A.; Kamarianos, A.; Skoulikidis, N. T.; Giannakou, U., Bioassays and biochemical biomarkers for assessing olive mill and citrus processing wastewater toxicity. *Environ Toxicol* **2011,** *26*, (6), 669-76.
23. Lebrun, J. D.; Perret, M.; Geffard, A.; Gourlay-France, C., Modelling copper bioaccumulation in Gammarus pulex and alterations of digestive metabolism. *Ecotoxicology* **2012,** *21*, (7), 2022-30.
24. Maltby, L.; Hills, L., Spray drift of pesticides and stream macroinvertebrates: experimental evidence of impacts and effectiveness of mitigation measures. *Environ Pollut* **2008,** *156*, (3), 1112-20.
25. McLoughlin, N.; Yin, D. Q.; Maltby, L.; Wood, R. M.; Yu, H. X., Evaluation of sensitivity and specificity of two crustacean biochemical biomarkers. *Environ Toxicol Chem* **2000,** *19*, (8), 2085-2092.
26. Neuparth, T.; Correia, A. D.; Costa, F. O.; Lima, G.; Costa, M. H., Multi-level assessment of chronic toxicity of estuarine sediments with the amphipod Gammarus locusta: I. Biochemical endpoints. *Mar Environ Res* **2005,** *60*, (1), 69-91.
27. Scheil, V.; Triebskorn, R.; Kohler, H. R., Cellular and stress protein responses to the UV filter 3-benzylidene camphor in the amphipod crustacean Gammarus fossarum (Koch 1835). *Arch Environ Contam Toxicol* **2008,** *54*, (4), 684-9.
28. Schill, R. O.; Gorlitz, H.; Kohler, H. R., Laboratory simulation of a mining accident: acute toxicity, hsc/hsp70 response, and recovery from stress in Gammarus fossarum (Crustacea, Amphipoda) exposed to a pulse of cadmium. *BioMetals* **2003,** *16*, (3), 391-401.
29. Schirling, M.; Jungmann, D.; Ladewig, V.; Ludwichowski, K. U.; Nagel, R.; Kohler, H. R.; Triebskorn, R., Bisphenol A in artificial indoor streams: II. Stress response and gonad histology in Gammarus fossarum (Amphipoda). *Ecotoxicology* **2006,** *15*, (2), 143-56.
30. Schirling, M.; Jungmann, D.; Ladewig, V.; Nagel, R.; Triebskorn, R.; Kohler, H. R., Endocrine effects in Gammarus fossarum (Amphipoda): influence of wastewater effluents, temporal variability, and spatial aspects on natural populations. *Arch Environ Contam Toxicol* **2005,** *49*, (1), 53-61.
31. Simon, R.; Jubeaux, G.; Chaumot, A.; Lemoine, J.; Geffard, O.; Salvador, A., Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum). *J Chromatogr A* **2010,** *1217*, (31), 5109-15.
32. Sroda, S.; Cossu-Leguille, C., Effects of sublethal copper exposure on two gammarid species: which is the best competitor? *Ecotoxicology* **2011,** *20*, (1), 264-73.
33. Timofeyev, M. A.; Shatilina, Z. M.; Protopopova, M. V.; Bedulina, D. S.; Pavlichenko, V. V.; Kolesnichenko, A. V.; Steinberg, C. E. W., Thermal stress defense in freshwater amphipods from contrasting habitats with emphasis on small heat shock proteins (sHSPs). *J Therm Biol* **2009,** *34*, (6), 281-285.
34. Triebskorn, R.; Adam, S.; Casper, H.; Honnen, W.; Pawert, M.; Schramm, M.; Schwaiger, J.; Kohler, H. R., Biomarkers as diagnostic tools for evaluating effects of unknown past water quality conditions on stream organisms. *Ecotoxicology* **2002,** *11*, (6), 451-465.
35. Turja, R.; Guimaraes, L.; Nevala, A.; Kankaanpaa, H.; Korpinen, S.; Lehtonen, K. K., Cumulative effects of exposure to cyanobacteria bloom extracts and benzo[a]pyrene on antioxidant defence biomarkers in Gammarus oceanicus (Crustacea: Amphipoda). *Toxicon* **2014,** *78*, 68-77.
36. Xuereb, B.; Noury, P.; Felten, V.; Garric, J.; Geffard, O., Cholinesterase activity in Gammarus pulex (Crustacea Amphipoda): characterization and effects of chlorpyrifos. *Toxicology* **2007,** *236*, (3), 178-89.

37. Xuereb, B.; Lefevre, E.; Garric, J.; Geffard, O., Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda): linking AChE inhibition and behavioural alteration. *Aquat Toxicol* **2009,** *94*, (2), 114-22.
38. Besse, J. P.; Coquery, M.; Lopes, C.; Chaumot, A.; Budzinski, H.; Labadie, P.; Geffard, O., Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: towards the determination of threshold values. *Water Res* **2013,** *47*, (2), 650-60.
39. Lacaze, E.; Devaux, A.; Mons, R.; Bony, S.; Garric, J.; Geffard, A.; Geffard, O., DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment. *Environ Pollut* **2011,** *159*, (6), 1682-91.
40. Coulaud, R.; Geffard, O.; Xuereb, B.; Lacaze, E.; Queau, H.; Garric, J.; Charles, S.; Chaumot, A., In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. *Water Res* **2011,** *45*, (19), 6417-29.
41. Maltby, L.; Clayton, S. A.; Wood, R. M.; McLoughlin, N., Evaluation of the Gammarus pulex in situ feeding assay as a biomonitor of water quality: Robustness, responsiveness, and relevance. *Environ Toxicol Chem* **2002,** *21*, (2), 361-368.
42. Handy, R. D.; Galloway, T. S.; Depledge, M. H., A proposal for the use of biomarkers for the assessment of chronic pollution and in regulatory toxicology. *Ecotoxicology* **2003,** *12*, (1-4), 331-343.
43. Sanchez, W.; Porcher, J. M., Fish biomarkers for environmental monitoring within the Water Framework Directive of the European Union. *Trac-Trend Anal Chem* **2009,** *28*, (2), 150-158.
44. Goksøyr, A.; Förlin, L., The cytochrome P-450 system in fish, aquatic toxicology and environmental monitoring. *Aquat Toxicol* **1992,** *22*, (4), 287-311.
45. James, M. O.; Boyle, S. M., Cytochromes P450 in crustacea. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* **1998,** *121*, (1-3), 157-72.
46. Koenig, S.; Fernandez, P.; Sole, M., Differences in cytochrome P450 enzyme activities between fish and crustacea: relationship with the bioaccumulation patterns of polychlorobiphenyls (PCBs). *Aquat Toxicol* **2012,** *108*, 11-7.
47. Depledge, M. H.; Galloway, T. S., Healthy animals, healthy ecosystems. *Front Ecol Environ* **2005,** *3*, (5), 251-258.
48. Lemos, M. F. L.; Soares, A. M. V. M.; Correia, A. C.; Esteves, A. C., Proteins in ecotoxicology - How, why and why not? *Proteomics* **2010,** *10*, (4), 873-887.
49. Kime, D. E.; Nash, J. P.; Scott, A. P., Vitellogenesis as a biomarker of reproductive disruption by xenobiotics. *Aquaculture* **1999,** *177*, (1-4), 345-352.
50. Matozzo, V.; Gagne, F.; Marin, M. G.; Ricciardi, F.; Blaise, C., Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. *Environ Int* **2008,** *34*, (4), 531-45.
51. LeBlanc, G. A., Crustacean endocrine toxicology: a review. *Ecotoxicology* **2007,** *16*, (1), 61-81.
52. Hayward, A.; Takahashi, T.; Bendena, W. G.; Tobe, S. S.; Hui, J. H., Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans. *FEBS Lett* **2010,** *584*, (6), 1273-8.
53. Cheng, W.; Tsai, I. H.; Huang, C. J.; Chiang, P. C.; Cheng, C. H.; Yeh, M. S., Cloning and characterization of hemolymph clottable proteins of kuruma prawn (Marsupenaeus japonicus) and white shrimp (Litopenaeus vannamei). *Dev Comp Immunol* **2008,** *32*, (3), 265-74.
54. Short, S.; Yang, G.; Kille, P.; Ford, A. T., Vitellogenin is not an appropriate biomarker of feminisation in a crustacean. *Aquat Toxicol* **2014,** *153*, (0), 89-97.
55. Borowsky, R.; Guarna, M. M., Excess Amylase in Gammarus-Palustris (Crustacea, Amphipoda) - Its Release into and Possible Roles in the Environment. *Marine Biology* **1989,** *101*, (4), 529-534.
56. Correia, A. D.; Costa, M. H.; Luis, O. J.; Livingstone, D. R., Age-related changes in antioxidant enzyme activities, fatty acid composition and lipid peroxidation in whole body Gammarus locusta (Crustacea : Amphipoda). *J Exp Mar Biol Ecol* **2003,** *289*, (1), 83-101.
57. Hanson, N.; Forlin, L.; Larsson, A., Spatial and annual variation to define the normal range of biological endpoints: an example with biomarkers in perch. *Environ Toxicol Chem* **2010,** *29*, (11), 2616-24.

58. Xuereb, B.; Chaumot, A.; Mons, R.; Garric, J.; Geffard, O., Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda) Intrinsic variability, reference levels, and a reliable tool for field surveys. *Aquat Toxicol* **2009,** *93*, (4), 225-33.
59. Lacaze, E.; Devaux, A.; Jubeaux, G.; Mons, R.; Gardette, M.; Bony, S.; Garric, J.; Geffard, O., DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level. *Sci Total Environ* **2011,** *409*, (17), 3230-6.
60. Gevaert, K.; Vandekerckhove, J., Protein identification methods in proteomics. *Electrophoresis* **2000,** *21*, (6), 1145-1154.
61. Vioque-Fernandez, A.; Alves de Almeida, E.; Lopez-Barea, J., Assessment of Donana National Park contamination in Procambarus clarkii: integration of conventional biomarkers and proteomic approaches. *Sci Total Environ* **2009,** *407*, (5), 1784-97.
62. Armengaud, J., Microbiology and proteomics, getting the best of both worlds! *Environ Microbiol* **2013,** *15*, (1), 12-23.
63. Apraiz, I.; Mi, J.; Cristobal, S., Identification of proteomic signatures of exposure to marine pollutants in mussels (Mytilus edulis). *Mol Cell Proteomics* **2006,** *5*, (7), 1274-85.
64. Rodriguez-Ortega, M. J.; Grosvik, B. E.; Rodriguez-Ariza, A.; Goksoyr, A.; Lopez-Barea, J., Changes in protein expression profiles in bivalve molluscs (Chamaelea gallina) exposed to four model environmental pollutants. *Proteomics* **2003,** *3*, (8), 1535-43.
65. Amelina, H.; Apraiz, I.; Sun, W.; Cristobal, S., Proteomics-based method for the assessment of marine pollution using liquid chromatography coupled with two-dimensional electrophoresis. *J Proteome Res* **2007,** *6*, (6), 2094-104.
66. Liu, F. J.; Wang, W. X., Proteome pattern in oysters as a diagnostic tool for metal pollution. *J Hazard Mater* **2012,** *239*, (0), 241-248.
67. Binelli, A.; Marisa, I.; Fedorova, M.; Hoffmann, R.; Riva, C., First evidence of protein profile alteration due to the main cocaine metabolite (benzoylecgonine) in a freshwater biological model. *Aquat Toxicol* **2013,** *140-141*, 268-78.
68. Chora, S.; Starita-Geribaldi, M.; Guigonis, J. M.; Samson, M.; Romeo, M.; Bebianno, M. J., Effect of cadmium in the clam Ruditapes decussatus assessed by proteomic analysis. *Aquat Toxicol* **2009,** *94*, (4), 300-8.
69. Dondero, F.; Negri, A.; Boatti, L.; Marsano, F.; Mignone, F.; Viarengo, A., Transcriptomic and proteomic effects of a neonicotinoid insecticide mixture in the marine mussel (Mytilus galloprovincialis, Lam.). *Sci Total Environ* **2010,** *408*, (18), 3775-86.
70. Ji, C.; Wu, H.; Wei, L.; Zhao, J.; Yu, J., Proteomic and metabolomic analysis reveal gender-specific responses of mussel Mytilus galloprovincialis to 2,2',4,4'-tetrabromodiphenyl ether (BDE 47). *Aquat Toxicol* **2013,** *140-141*, 449-57.
71. Leung, P. T.; Wang, Y.; Mak, S. S.; Ng, W. C.; Leung, K. M., Differential proteomic responses in hepatopancreas and adductor muscles of the green-lipped mussel Perna viridis to stresses induced by cadmium and hydrogen peroxide. *Aquat Toxicol* **2011,** *105*, (1-2), 49-61.
72. Liu, F.; Wang, D. Z.; Wang, W. X., Cadmium-induced changes in trace element bioaccumulation and proteomics perspective in four marine bivalves. *Environ Toxicol Chem* **2012,** *31*, (6), 1292-300.
73. Manduzio, H.; Cosette, P.; Gricourt, L.; Jouenne, T.; Lenz, C.; Andersen, O. K.; Leboulenger, F.; Rocher, B., Proteome modifications of blue mussel (Mytilus edulis L.) gills as an effect of water pollution. *Proteomics* **2005,** *5*, (18), 4958-63.
74. Maria, V. L.; Gomes, T.; Barreira, L.; Bebianno, M. J., Impact of benzo(a)pyrene, Cu and their mixture on the proteomic response of Mytilus galloprovincialis. *Aquat Toxicol* **2013,** *144-145*, 284-95.
75. Riva, C.; Binelli, A., Analysis of the Dreissena polymorpha gill proteome following exposure to dioxin-like PCBs: mechanism of action and the role of gender. *Comp Biochem Physiol Part D Genomics Proteomics* **2014,** *9*, (1), 23-30.
76. Riva, C.; Binelli, A.; Rusconi, F.; Colombo, G.; Pedriali, A.; Zippel, R.; Provini, A., A proteomic study using zebra mussels (D. polymorpha) exposed to benzo(α)pyrene: The role of gender and exposure concentrations. *Aquat Toxicol* **2011,** *104*, (1–2), 14-22.

77. Riva, C.; Cristoni, S.; Binelli, A., Effects of triclosan in the freshwater mussel Dreissena polymorpha: a proteomic investigation. *Aquat Toxicol* **2012,** *118-119*, (0), 62-71.
78. Romero-Ruiz, A.; Carrascal, M.; Alhama, J.; Gomez-Ariza, J. L.; Abian, J.; Lopez-Barea, J., Utility of proteomics to assess pollutant response of clams from the Donana bank of Guadalquivir Estuary (SW Spain). *Proteomics* **2006,** *6 Suppl 1*, S245-55.
79. Thompson, E. L.; Taylor, D. A.; Nair, S. V.; Birch, G.; Haynes, P. A.; Raftos, D. A., A proteomic analysis of the effects of metal contamination on Sydney Rock Oyster (Saccostrea glomerata) haemolymph. *Aquat Toxicol* **2011,** *103*, (3-4), 241-9.
80. Zhou, J.; Cai, Z. H.; Li, L.; Gao, Y. F.; Hutchinson, T. H., A proteomics based approach to assessing the toxicity of bisphenol A and diallyl phthalate to the abalone (Haliotis diversicolor supertexta). *Chemosphere* **2010,** *79*, (5), 595-604.
81. Besse, J. P.; Geffard, O.; Coquery, M., Relevance and applicability of active biomonitoring in continental waters under the Water Framework Directive. *Trac-Trend Anal Chem* **2012,** *36*, (0), 113-127.
82. Leroy, D.; Haubruge, E.; De Pauw, E.; Thome, J. P.; Francis, F., Development of ecotoxicoproteomics on the freshwater amphipod Gammarus pulex: identification of PCB biomarkers in glycolysis and glutamate pathways. *Ecotoxicol Environ Saf* **2010,** *73*, (3), 343-52.
83. Ralston-Hooper, K. J.; Sanchez, B. C.; Adamec, J.; Sepulveda, M. S., Proteomics in aquatic amphipods: can it be used to determine mechanisms of toxicity and interspecies responses after exposure to atrazine? *Environ Toxicol Chem* **2011,** *30*, (5), 1197-203.
84. Silvestre, F.; Gillardin, V.; Dorts, J., Proteomics to assess the role of phenotypic plasticity in aquatic organisms exposed to pollution and global warming. *Integr Comp Biol* **2012,** *52*, (5), 681-94.
85. Neave, M. J.; Streten-Joyce, C.; Nouwens, A. S.; Glasby, C. J.; McGuinness, K. A.; Parry, D. L.; Gibb, K. S., The transcriptome and proteome are altered in marine polychaetes (Annelida) exposed to elevated metal levels. *J Proteomics* **2012,** *75*, (9), 2721-35.
86. Bradley, B. P.; Shrader, E. A.; Kimmel, D. G.; Meiller, J. C., Protein expression signatures: an application of proteomics. *Mar Environ Res* **2002,** *54*, (3-5), 373-7.
87. Dircksen, H.; Neupert, S.; Predel, R.; Verleyen, P.; Huybrechts, J.; Strauss, J.; Hauser, F.; Stafflinger, E.; Schneider, M.; Pauwels, K.; Schoofs, L.; Grimmelikhuijzen, C. J., Genomics, transcriptomics, and peptidomics of Daphnia pulex neuropeptides and protein hormones. *J Proteome Res* **2011,** *10*, (10), 4478-504.
88. Fröhlich, T.; Arnold, G. J.; Fritsch, R.; Mayr, T.; Laforsch, C., LC-MS/MS-based proteome profiling in Daphnia pulex and Daphnia longicephala: The Daphnia pulex genome database as a key for high throughput proteomics in Daphnia. *BMC Genomics* **2009,** *10*.
89. Paik, Y. K.; Jeong, S. K.; Lee, E. Y.; Jeong, P. Y.; Shim, Y. H., C. elegans: an invaluable model organism for the proteomics studies of the cholesterol-mediated signaling pathway. *Expert Rev Proteomics* **2006,** *3*, (4), 439-53.
90. Dowling, V. A.; Sheehan, D., Proteomics as a route to identification of toxicity targets in environmental toxicology. *Proteomics* **2006,** *6*, (20), 5597-604.
91. Monsinjon, T.; Knigge, T., Proteomic applications in ecotoxicology. *Proteomics* **2007,** *7*, (16), 2997-3009.
92. Carpentler, S. C.; Panis, B.; Vertommen, A.; Swennen, R.; Sergeant, K.; Renaut, J.; Laukens, K.; Witters, E.; Samyn, B.; Devreese, B., Proteome analysis of non-model plants: A Challenging but powerful approach. *Mass Spectrom Rev* **2008,** *27*, (4), 354-377.
93. Martyniuk, C. J.; Alvarez, S.; Denslow, N. D., DIGE and iTRAQ as biomarker discovery tools in aquatic toxicology. *Ecotoxicol Environ Saf* **2012,** *76*, (2), 3-10.
94. Martyniuk, C. J.; Denslow, N. D., Towards functional genomics in fish using quantitative proteomics. *Gen Comp Endocrinol* **2009,** *164*, (2-3), 135-41.
95. Gorg, A.; Weiss, W.; Dunn, M. J., Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **2004,** *4*, (12), 3665-85.
96. Tonge, R.; Shaw, J.; Middleton, B.; Rowlinson, R.; Rayner, S.; Young, J.; Pognan, F.; Hawkins, E.; Currie, I.; Davison, M., Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **2001,** *1*, (3), 377-96.

97. Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001,** *19*, (3), 242-7.
98. Perrot, M.; Sagliocco, F.; Mini, T.; Monribot, C.; Schneider, U.; Shevchenko, A.; Mann, M.; Jeno, P.; Boucherie, H., Two-dimensional gel protein database of Saccharomyces cerevisiae (update 1999). *Electrophoresis* **1999,** *20*, (11), 2280-98.
99. Colbourne, J. K.; Pfrender, M. E.; Gilbert, D.; Thomas, W. K.; Tucker, A.; Oakley, T. H.; Tokishita, S.; Aerts, A.; Arnold, G. J.; Basu, M. K.; Bauer, D. J.; Caceres, C. E.; Carmel, L.; Casola, C.; Choi, J. H.; Detter, J. C.; Dong, Q.; Dusheyko, S.; Eads, B. D.; Frohlich, T.; Geiler-Samerotte, K. A.; Gerlach, D.; Hatcher, P.; Jogdeo, S.; Krijgsveld, J.; Kriventseva, E. V.; Kultz, D.; Laforsch, C.; Lindquist, E.; Lopez, J.; Manak, J. R.; Muller, J.; Pangilinan, J.; Patwardhan, R. P.; Pitluck, S.; Pritham, E. J.; Rechtsteiner, A.; Rho, M.; Rogozin, I. B.; Sakarya, O.; Salamov, A.; Schaack, S.; Shapiro, H.; Shiga, Y.; Skalitzky, C.; Smith, Z.; Souvorov, A.; Sung, W.; Tang, Z.; Tsuchiya, D.; Tu, H.; Vos, H.; Wang, M.; Wolf, Y. I.; Yamagata, H.; Yamada, T.; Ye, Y.; Shaw, J. R.; Andrews, J.; Crease, T. J.; Tang, H.; Lucas, S. M.; Robertson, H. M.; Bork, P.; Koonin, E. V.; Zdobnov, E. M.; Grigoriev, I. V.; Lynch, M.; Boore, J. L., The ecoresponsive genome of Daphnia pulex. *Science* **2011,** *331*, (6017), 555-61.
100. Carvalho, P. C.; Fischer, J. S.; Chen, E. I.; Yates, J. R., 3rd; Barbosa, V. C., PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics* **2008,** *9*, 316.
101. Peterson, K. J.; Lyons, J. B.; Nowak, K. S.; Takacs, C. M.; Wargo, M. J.; McPeek, M. A., Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci U S A* **2004,** *101*, (17), 6536-41.
102. Studer, R. A.; Robinson-Rechavi, M., How confident can we be that orthologs are similar, but paralogs differ? *Trends Genet* **2009,** *25*, (5), 210-6.
103. Nikinmaa, M.; Rytkonen, K. T., Functional genomics in aquatic toxicology-do not forget the function. *Aquat Toxicol* **2011,** *105*, (3-4 Suppl), 16-24.
104. Kuhn, E.; Wu, J.; Karl, J.; Liao, H.; Zolg, W.; Guild, B., Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. *Proteomics* **2004,** *4*, (4), 1175-1186.
105. Muraoka, S.; Kume, H.; Watanabe, S.; Adachi, J.; Kuwano, M.; Sato, M.; Kawasaki, N.; Kodera, Y.; Ishitobi, M.; Inaji, H.; Miyamoto, Y.; Kato, K.; Tomonaga, T., Strategy for SRM-based verification of biomarker candidates discovered by iTRAQ method in limited breast cancer tissue samples. *J Proteome Res* **2012,** *11*, (8), 4201-10.
106. Gallien, S.; Duriez, E.; Domon, B., Selected reaction monitoring applied to proteomics. *J Mass Spectrom* **2011,** *46*, (3), 298-312.
107. Anderson, L.; Hunter, C. L., Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics* **2006,** *5*, (4), 573-88.
108. von der Ohe, P. C.; Liess, M., Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. *Environ Toxicol Chem* **2004,** *23*, (1), 150-6.
109. Celander, M. C.; Goldstone, J. V.; Denslow, N. D.; Iguchi, T.; Kille, P.; Meyerhoff, R. D.; Smith, B. A.; Hutchinson, T. H.; Wheeler, J. R., Species extrapolation for the 21st century. *Environ Toxicol Chem* **2011,** *30*, (1), 52-63.
110. Schrimpf, S. P.; Weiss, M.; Reiter, L.; Ahrens, C. H.; Jovanovic, M.; Malmstrom, J.; Brunner, E.; Mohanty, S.; Lercher, M. J.; Hunziker, P. E.; Aebersold, R.; von Mering, C.; Hengartner, M. O., Comparative functional analysis of the Caenorhabditis elegans and Drosophila melanogaster proteomes. *PLoS Biol* **2009,** *7*, (3), e48.
111. Jubeaux, G.; Audouard-Combe, F.; Simon, R.; Tutundjian, R.; Salvador, A.; Geffard, O.; Chaumot, A., Vitellogenin-like proteins among invertebrate species diversity: potential of proteomic mass spectrometry for biomarker development. *Environ Sci Technol* **2012,** *46*, (11), 6315-23.
112. Armengaud, J.; Trapp, J.; Pible, O.; Geffard, O.; Chaumot, A.; Hartmann, E. M., Non-model organisms, a species endangered by proteogenomics. *J Proteomics* **2014,** *105*, (0), 5-18.
113. Hook, S. E.; Twine, N. A.; Simpson, S. L.; Spadaro, D. A.; Moncuquet, P.; Wilkins, M. R., 454 pyrosequencing-based analysis of gene expression profiles in the amphipod Melita plumulosa: transcriptome assembly and toxicant induced changes. *Aquat Toxicol* **2014,** *153*, (0), 73-88.

114. Bouetard, A.; Noirot, C.; Besnard, A. L.; Bouchez, O.; Choisne, D.; Robe, E.; Klopp, C.; Lagadic, L.; Coutellec, M. A., Pyrosequencing-based transcriptomic resources in the pond snail Lymnaea stagnalis, with a focus on genes involved in molecular response to diquat-induced stress. *Ecotoxicology* **2012,** *21*, (8), 2222-34.
115. Marinković, M.; de Leeuw, W. C.; de Jong, M.; Kraak, M. H. S.; Admiraal, W.; Breit, T. M.; Jonker, M. J., Combining Next-Generation Sequencing and Microarray Technology into a Transcriptomics Approach for the Non-Model Organism Chironomus riparius. *PLoS ONE* **2012,** *7*, (10).
116. Jaffe, J. D.; Berg, H. C.; Church, G. M., Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **2004,** *4*, (1), 59-77.
117. Eisen, J. A., Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* **1998,** *8*, (3), 163-7.
118. Avarre, J. C.; Lubzens, E.; Babin, P. J., Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate apolipoprotein B. *BMC Evol Biol* **2007,** *7*, 3.
119. Laurent, J. M.; Vogel, C.; Kwon, T.; Craig, S. A.; Boutz, D. R.; Huse, H. K.; Nozue, K.; Walia, H.; Whiteley, M.; Ronald, P. C.; Marcotte, E. M., Protein abundances are more conserved than mRNA abundances across diverse taxa. *Proteomics* **2010,** *10*, (23), 4209-4212.
120. Chen, J. E.; Li, J. Y.; You, Z. Y.; Liu, L. L.; Liang, J. S.; Ma, Y. Y.; Chen, M.; Zhang, H. R.; Jiang, Z. D.; Zhong, B. X., Proteome analysis of silkworm, Bombyx mori, larval gonads: characterization of proteins involved in sexual dimorphism and gametogenesis. *J Proteome Res* **2013,** *12*, (6), 2422-38.
121. Ventura, T.; Manor, R.; Aflalo, E. D.; Weil, S.; Raviv, S.; Glazer, L.; Sagi, A., Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. *Endocrinology* **2009,** *150*, (3), 1278-86.
122. Ventura, T.; Manor, R.; Aflalo, E. D.; Chalifa-Caspi, V.; Weil, S.; Sharabi, O.; Sagi, A., Post-Embryonic Transcriptomes of the Prawn Macrobrachium rosenbergii: Multigenic Succession through Metamorphosis. *PLoS ONE* **2013,** *8*, (1), e55322.
123. Geffard, O.; Xuereb, B.; Chaumot, A.; Geffard, A.; Biagianti, S.; Noel, C.; Abbaci, K.; Garric, J.; Charmantier, G.; Charmantier-Daures, M., Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. *Environ Toxicol Chem* **2010,** *29*, (10), 2249-59.
124. Trapp, J.; Geffard, O.; Imbert, G.; Gaillard, J. C.; Davin, A. H.; Chaumot, A.; Armengaud, J., Proteogenomics of Gammarus fossarum to document the reproductive system of amphipods. *Mol Cell Proteomics* **2014**.
125. Ankley, G. T.; Bennett, R. S.; Erickson, R. J.; Hoff, D. J.; Hornung, M. W.; Johnson, R. D.; Mount, D. R.; Nichols, J. W.; Russom, C. L.; Schmieder, P. K.; Serrrano, J. A.; Tietge, J. E.; Villeneuve, D. L., Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* **2010,** *29*, (3), 730-41.

4.2. La protéogénomique au secours des espèces non-modèles

Publication 2 :

Non-model organisms, a species endangered by proteogenomics

Version auteur de l'article publié dans *Journal of proteomics*, 2014, 105 (0) 5-18
La version finale est disponible sur http://www.sciencedirect.com/science/article/pii/S1874391914000190

Jean Armengaud [a], Judith Trapp [a,b], Olivier Pible [a], Olivier Geffard [a], Arnaud Chaumot [a], Erica M. Hartmann [a]

[a] CEA, DSV, IBEB, Lab. Biochim. System. Perturb., Bagnols-sur-Ceze, F-30207, France
[b] Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, CS70077, F-69626 Villeurbanne, France

Résumé :

Alors que l'application à large échelle des techniques protéomiques était possible seulement chez les organismes dont le génome a été séquencé (*i.e*. espèces modèles), l'apparition des techniques de séquençage nouvelle génération bouleverse ce statuquo. La protéogénomique, l'utilisation des données protéomiques pour affiner l'annotation du génome, permet un plus haut niveau d'intégration des données omiques. Par extension, combiner les données génomiques et protéomiques devient une démarche routinière au sein de nombreux projets de recherche. Les approches dérivées de la protéogénomique sont actuellement en pleine expansion, permettant un niveau d'étude sans précédent sur les organismes non-modèles au niveau moléculaire. Aujourd'hui, la première ébauche d'un génome peut être rapidement obtenue et à un coût raisonnable, par le biais des séquenceurs nouvelle génération. Ces données peuvent être directement employées pour interpréter les spectres de masse produits en protéomique, sans étape préalable d'annotation du génome. De même, l'utilisation des techniques RNA-Seq pour caractériser les séquences en acides nucléiques, directement traduites en séquences protéiques est une alternative séduisante. Toutefois, certaines limites subsistent dans l'utilisation des techniques RNA-Seq, nécessitant d'être améliorer pour gagner en efficacité. Dans cet article, l'intérêt de travailler sur les organismes non-modèles est présenté ainsi que les méthodes protéomiques jusqu'ici employées et les améliorations conférées par la protéogénomique. Ces nouveaux développements menacent la distinction entre organismes modèles et non-modèles, du moins en protéomique !

Mots clés : Organismes non modèles, protéomique massive, technologies de séquençage nouvelles générations, ébauche de génome, RNA-Seq, protéogénomique.

Non-model organisms, a species endangered by proteogenomics

Abstract

Previously, large-scale proteomics was possible only for organisms whose genomes were sequenced, meaning the most common model organisms. The use of next-generation sequencers is now changing the deal. With "proteogenomics", the use of experimental proteomics data to refine genome annotations, a higher integration of omics data is gaining ground. By extension, combining genomic and proteomic data is becoming routine in many research projects. "Proteogenomic"-flavored approaches are currently expanding, enabling the molecular studies of non-model organisms at an unprecedented depth. Today draft genomes can be obtained using next-generation sequencers in a rather straightforward way and at a reasonable cost for any organism. Unfinished genome sequences can be used to interpret tandem mass spectrometry proteomics data without the need for time-consuming genome annotation, and the use of RNA-seq to establish nucleotide sequences that are directly translated into protein sequences appears promising. There are, however, certain drawbacks that deserve further attention for RNA-seq to become more efficient. Here, we discuss the opportunities of working with non-model organisms, the proteomic methods that have been used until now, and the dramatic improvements proffered by proteogenomics. These put the distinction between model and non-model organisms in great danger, at least in terms of proteomics!

Keywords: models, non-model organisms, high-throughput proteomics, next-generation sequencing, draft genome, genome annotation, RNA-seq, proteogenomics, proteome

1. INTRODUCTION

Biological systems have adapted to a large variety of habitat conditions found previously and currently on Earth. Taxonomists have delineated these specialized life forms within species, the basic unit of biological classification, defined by significantly different traits. For multicellular eukaryotes, primarily animals and plants, the frontier between two species is the ability to interbreed and produce fertile offspring; however, the definition is more ambiguous for unicellular organisms [1]. The current estimate of diversity is still controversial and probably far from reality because the tools to assess this diversity are still under development and unexplored biomes are awaiting characterization. The recent expansion of molecular techniques highlighted important genetic heterogeneities surpassing the richness of species diversity previously predicted from morphological similarities [2]. Today, the panoply of life on Earth, *i.e.*, the entire biosphere, is immensely rich in species with between 2 and 50 million non-microbial species [3-5] and between 10 million and 1 billion prokaryotic species [6, 7]. The complete census of a soil sample, which can contain upwards of 1 billion bacterial cells per gram, is currently impossible, especially given that estimates of its diversity vary over three logs of magnitude up to 100,000 species, with most of these species having never been observed or otherwise detected (**Figure 1**). A glimpse of the huge diversity of microorganisms was revealed in metagenome sequencing surveys that resulted in the definition of more than 60 novel bacterial phyla, half of which have no cultivable representatives and are currently considered “microbial dark matter” because of the difficulties their characterization poses [8]. An initial assessment of the diversity of microorganisms can be obtained using shotgun DNA sequencing, and the recent development of whole-cell matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry allows the rapid identification of isolates and so could foster the discovery of new cultivable organisms from the poorest characterized branches of the tree of life [9]. Metaproteomics is also of interest to assess microbial diversity, in addition to its primary goal of characterizing microbial community functions [10]. Metaorganisms appear well characterized in terms of taxonomy, but at the molecular level much work remains. The existence of 0.4 million plants, 1.5 million fungi and 8.7 million animals has been estimated [5]. Among animals, the largest diversity arises from arthropods (90,000 myriapods, 150,000 crustaceans, 600,000 arachnids, and 5 million insects), while chordata (5,500 mammals, 10,000 birds, 10,000 reptiles, 15,000 amphibians and 40,000 fishes) are less diverse. However, the number of sequenced genomes falls far short of the number of known organisms and poorly reflects the true diversity of life on Earth (**Figure 1**).

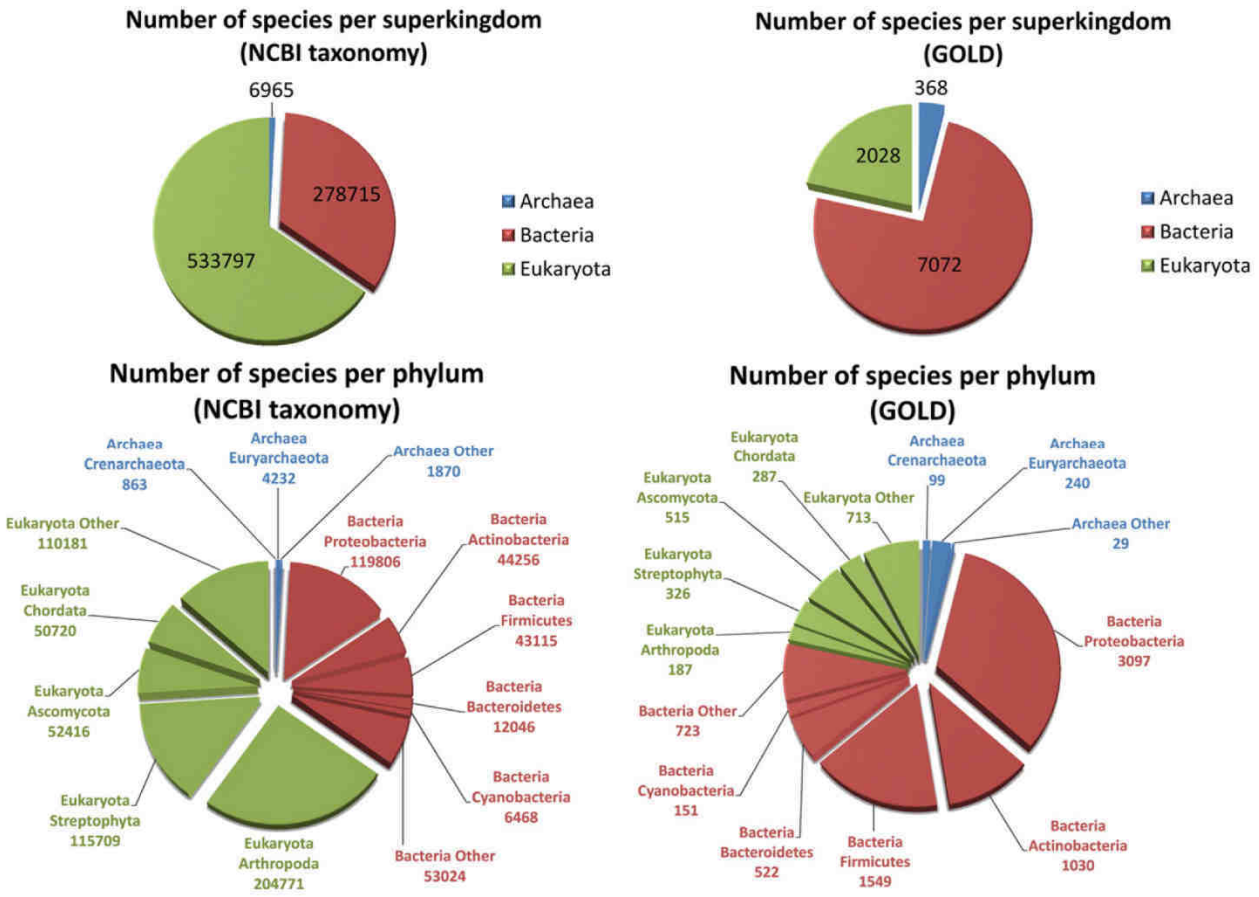


Figure 1: Number of species presently cataloged in NCBI taxonomy and Gold databases (2013/11/19)

The number of species presently indexed in the NCBI taxonomy database (www.ncbi.nlm.nih.gov/taxonomy) are indicated on the left; the number of species for which at least one genome has been sequenced or being sequenced as catalogued in the Genomes OnLine Database (GOLD, www.genomesonline.org), a resource for centralized monitoring of genome and metagenome projects conducted by the whole scientific community [98], are presented on the right. Upper pie charts show the number per superkingdom; lower pie charts indicate the number per phylum

The legacy of Charles Darwin, his formulation of the central principles of evolutionary biology in *On the Origin of Species* in 1859, is based on the obvious existence of shared traits between organisms. Molecular biology approaches applied to a few model organisms from the tree of life confirmed the common base of shared principles between all living forms of life on Earth and the unity of basic cellular mechanisms. They also helped to define the conditions and mechanisms of the evolution of life. However, even today, the extensive characterization of a simple unicellular organism, such as a bacterium or yeast, is a tremendous task and that requires the coordinated efforts of a large specialized research community. Integration of descriptive experiments with the latest high-throughput technologies, such as genome sequencing and annotation, transcriptome and proteome dynamics, together with hypothesis-driven experiments, relies on expertise and tools

available to research consortia only. For a restricted number of these cellular models, an integrated view of their content and metabolism is available, but their individual components and their respective functions are not yet fully described [11]. Regarding multicellular organisms, plant and animal cell models are still being studied to decipher the key elements in each tissue during development.

While the model species approach enables the decryption of the universal mechanisms of biology, its inherent reductionist view is incapable of unlocking the mechanisms by which biological functions cope with the complexity of environmental conditions in which life develops. Thus, in addition to being selected on criteria that may bias their ecological representativeness (discussed below), model species only represent particular cases within species diversity. Our comprehension of any given organism does not necessarily advance our understanding of other organisms or life on the whole. Famously, the fruitfly *Drosophila melanogaster*, which has been very heavily studied, now appears to belong to a highly derived insect lineage whose genomic composition and function are quite marginal among insects and arthropods [12]. In addition to the fundamental understanding that can be gained from basic research, increasing the diversity of experimental organisms is also mandatory in applied science, as it is the only way to understand the potentialities of adaptation and resilience of biological systems to changing environmental conditions (global change, anthropogenic pressure, etc.). Environmental pollution offers a suitable example to illustrate this issue. In order to assess chemical contaminant effects on ecosystems, toxic responses are extrapolated from a limited number of model species for all species in the environment; this simplistic approach is a fundamental principle of regulatory ecotoxicity testing and ecological risk assessment [13]. This issue is challenging because test species are rarely selected purely on biological grounds, for instance the inclusion of a range of ecologies, developmental and reproductive modes, and life history traits. The divergence between native species and test species chosen as surrogates is also critical to environmental monitoring. Indeed, the particular model chosen as the sentinel species is a major determinant of the ecological relevance of the results obtained from any experimental approach to supply information about environmental effects on resident biological communities [14]. The inter-taxon heterogeneity regarding sensitivity to environmental contaminants, which is partly explained by the molecular diversity that emerged during the evolution of species, makes it mandatory to develop multi-species approaches in ecotoxicology. This specific example in environmental management underlines the increasing

necessity to make biological knowledge and resources available for non-model species both for basic and applied scientific goals.

In this review, we discuss the differences between model and non-model organisms, especially as they pertain to proteomics. We describe the traditional methods for identifying proteins from non-model organisms and give recent examples of such studies. We present the recent emergence of proteogenomics and show how the alliance of next-generation sequencing and next-generation mass spectrometry is currently fostering the in-depth characterization of non-model organisms.

2. MODEL VERSUS NON-MODEL ORGANISMS

Defining model, and conversely non-model, organisms. To properly define non-model organisms, we should first examine the characteristics that define model organisms. Subtle differences exist between model organisms and experimental organisms in general, making this definition somewhat tricky. Some organisms are popular models studied for decades by thousands of scientists from a large range of disciplines, have a high representational scope, and contribute to an integrative understanding of more complex systems [15, 16], but some are more niche systems used by a very restricted number of research teams. As discussed by Ankeny & Leonelli [16], the term "model organism" does not apply to all organisms used for experimental research, but only to those with a representational scope and target. Whether two popular experimental organisms, the fruitfly *Drosophila melanogaster* and the worm *Caenorhabditis elegans*, should be considered model organisms *per se* or not is subjected to debate because they are taxonomic outliers, and thus cannot directly serve as models for other organisms while they allowed to shed light on some fundamental biological processes. Although not all experimental organisms are model organisms, all model organisms are experimental organisms. To be a useful laboratory experimental organism, an organism should be pertinent to a key biological question and amenable to research in the laboratory, and as such should fulfill three main characteristics: i) "cultivable" as a stable isolated species in controlled laboratory conditions, ii) phenotypically "characterizable", and iii) genetically "engineerable" to identify the function of proteins following the manipulation of the corresponding genes. The first characteristic is a prerequisite to propagate the organism, design experiments in multiple conditions with equivalent biological starting material, and ultimately share the experimental organism with others in the scientific community to compare results on the same

basis. As a result, experimental organisms with a small size and rapid life cycle have been selected to facilitate breeding. The organism should also have well characterized phenotypes with optimized laboratory growth conditions and an in-depth description of its specific traits. Finally, establishing its genome sequence and annotating the nucleic acid patrimony is the first obligatory step for the modern characterization of the molecular elements of a widespread experimental organism. Today, genomic information is considered a standard scientific background and a starting point for any experiments involving high-throughput exploratory omic methods (transcriptomics, proteomics, metabolomics) or targeted genetic engineering for *in vivo* or *in vitro* validation studies (mutagenesis, two-hybrid interactomics, heterologous expression...).

Because of their scientific interest for numerous researchers, a wealth of knowledge is available for the most model organisms. From a historical perspective, some organisms have been chosen as models by leading research teams and quickly became worldwide laboratory workhorses for investigation of basic, fundamental biological processes as well as for the development of new molecular biology-based methods [15]. Among the most popular models [17], the budding yeast *Saccharomyces cerevisiae* is remarkable in being the target of many innovative methods: first entire chromosome (chromosome III) sequenced from any organism (reported in 1992), chromosome per chromosome DNA sequencing, systematic open reading frame knock-out, systematic green fluorescent protein tagging of every coding sequences for subcellular localization, systematic interaction screening using two-hybrid constructs, array-based transcriptomics, RNA-sequencing, metabolomics, high-throughput shotgun proteomics (MudPIT), phosphoproteomics, and complete mass spectrometric mapping are some examples. Of great industrial importance, this organism was cited in 103,279 publications referenced in Pubmed (2013/12/24). Remarkably, its rather simple genome (~6,000 coding sequences) and the dynamic range of its eukaryotic proteome made *S. cerevisiae* a perfect playing field for validating proteomic methods and a perfect standard for evaluating their performance. Many experiments aimed at improving human health use stable human cell lines to provide the material for molecular biology experiments, and these can be also considered as models. In addition, the Gram-negative bacterium *Escherichia coli* has held the attention of molecular biologists because of its fast growth rate and easy genetic engineering. Other popular model organisms are the bacteria *Bacillus subtilis* and *Deinococcus radiodurans*, the green algae *Chlamydomonas reinhardtii*, the angiosperm plant *Arabidopsis thaliana*, the mice *Mus musculus*, and the fish *Danio rerio.* Because these models are popular in terms of molecular biology

and cell biology, protocols, methods and commercially available kits have been optimized for each, facilitating further investigations.

Model organisms are extensively studied, reduced complexity representatives for larger, more complex phenomena, but that does not imply that non-model organisms are unrepresentative. Rather, most non-model organisms are either not cultivable as a stable isolated species in controlled laboratory conditions or simply not well characterized at the molecular biology level, thus requiring huge efforts in protocol adaptation and method development, or the target of investigation of a relatively small number of scientists whatever its taxonomy is. As a result, non-model organisms often make difficult experimental organisms.

Key biological questions require the investigation of non-model organisms. Non-model organisms are fundamentally interesting because they differ from current models. The interest in non-model species constitutes the heart of comparative biology to expose mechanisms underlying biological diversity. As direct inheritance from the "pre-Darwin" pioneering works in comparative anatomy (*e.g.*, Tyson, Buffon, Cuvier), evolutionary development and comparative genomics investigate diverse species, chosen because of their peculiar phylogenetic positions or morphological adaptations. This penchant for studying disparate organisms explains the present availability of genomes for unexpected organisms such as limpets, leeches, lancelets, water striders, etc. Similarly, ecophysiological studies focus on species living in extreme conditions, such as polar ecosystems, subterranean environments, deserts, and evolutionary ecology also uses cross-species comparisons to decipher how evolution supports species adaptation to their biotic and abiotic environment. The specialized and unique biology of such diversified species requires the development of specific methods and the adaptation of strategies conceived primarily for model organisms. In addition, the study of "exotic" species for their peculiar adaptations, or because they are amenable to study a particular biological phenomenon, has led to extraordinary practical and theoretical advances in biology, which could not have been achieved using only canonical model species. There are many examples of the contributions of "exotic" organisms to biology, such as the advances in molecular biology permitted by the discovery of the Taq polymerase in the bacterium *Thermus aquaticus* from hot springs, the contribution of *Agrobacterium tumefasciens* to vegetal transgenesis techniques, the demonstration of hormonal control of genetic expression by ecdysone in the insect mud worm *Chironomus*, etc. Non-model species can also be of interest because of their particular functional roles in ecosystems. Indeed, although each species contributes to the

overall integrity of a functional ecosystem, some species are central players that provide key services, as illustrated by the pollinator honey bees. Their disappearance often triggers the loss of other resident species in a domino effect, and for this reason the bees are called a “keystone species”. To understand the immediate effects and long-term consequences of human activities on the environment, these species are used as surrogates for a specific ecosystem. Studies of the molecular mechanisms of their responses when exposed to elevated temperature, acidification or pollution aim to understand their ability to cope with environmental alterations. For example, the Antarctic krill *Euphausia superba* plays a pivotal role in Antarctic ecosystem [18]. This crustacean is a significant consumer of phytoplankton blooms during the austral summer and algae found under the sea ice during the winter, and in turn prey for the southern ocean megafauna such as whales, seals and sea birds. Ecological data show the significant decline of population densities, suggesting that this species is affected by climate changes [18]. Until now, no proteomic studies have been conducted on such a keystone animal. In contrast to keystone species, other species, such as alien invasives, can be a direct threat to native biodiversity. These species aggressively overcome most native species that they encounter during their invasion, leading to biotic homogenization. Biological invasion affects virtually all ecosystems, often promoted by habitat degradation and assisted by the absence of natural enemies. Among alien species, the zebra mussels (*Dreissena polymorpha*) introduced to the US Great Lakes, threaten native communities by occupying their habitat [19]. Preliminary proteomic data have been obtained with this non-model organism [20, 21], but they are insufficient to explain interactions with anthropogenic conditions. In wetlands, the common reed *Phragmites australis* is spreading throughout the world, affecting communities, altering sediment dynamics and water cycling [22]. Understanding the molecular players contributing to the success of the invasion process could be a valuable resource for engineering efficient programs for biological control and ecosystem restoration. Furthermore, non-model organisms could be a tremendous source of bioproducts for human health, and their study could be as prolific as the development of drugs from medicinal plants employed in indigenous societies [23]. There are few model extremophiles, but studies of these organisms reveal molecular mechanisms related to their adaptation to extreme conditions of life. Uncultured microorganisms are also highly interesting as they are numerous and play a vital role in the global carbon cycle but are extremely difficult to characterize because of the lack of starting material. Metaproteomics is revealing the dominant metabolic processes occurring within marine bacterial communities [24] but further characterization of individual members of these communities should be investigated.

3. TRADITIONNAL TOOLS AND STRATEGIES FOR PROTEOMICS OF NON-MODEL ORGANISMS

Traditional tools for discovery proteomics, developed and validated using well characterized model organisms, have been adapted with relatively few modifications for non-model organisms (**Figure 2**). The greatest challenge tends to be the bioinformatic data analysis for unsequenced organisms, especially those with no sequenced close relatives, rather than the sample preparation, which is based on more universal biochemical properties to purify and separate proteins to varying degrees.

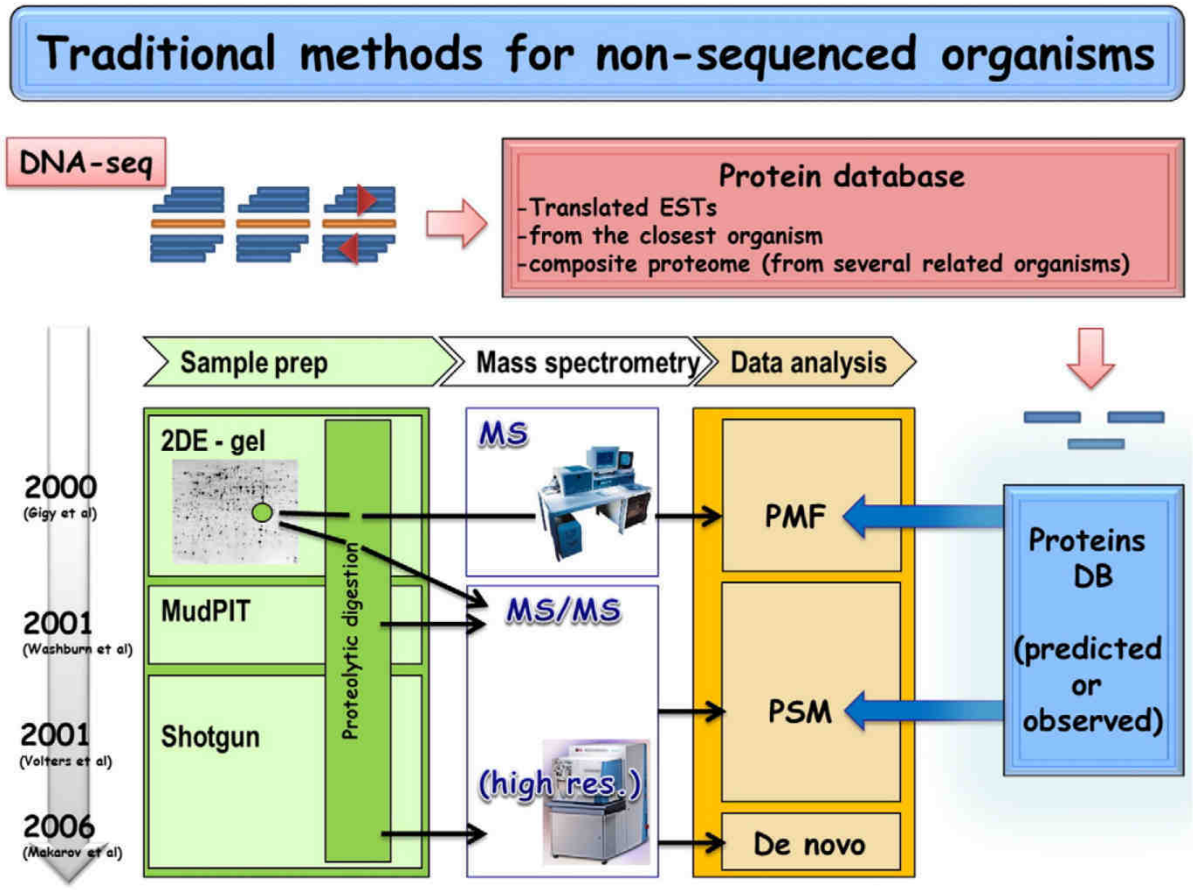


Figure 2: Traditional proteomics approaches for non-model organisms

The protein database used for the interpretation of data is constructed from annotated genome(s), either from the non-model organism itself or from the most-closely related organism that has been sequenced, or from a series of closely-related organisms (composite proteomics), or from a partial or complete expressed sequence tag (EST) information obtained by sequencing cloned cDNA. Abbreviations are peptide mass fingerprint (PMF), peptide spectral matching (PSM), two dimensional gel electrophoresis (2DE), multidimensional protein identification technology (MUD-PIT), high resolution (high res.). A time scale is indicated on the right together with the most emblematic references [99-102].

Sample preparation. In the absence of sophisticated chromatographic separation and mass spectrometry-based identification and quantitation, the relatively simplistic combination of

polyacrylamide gel electrophoresis (PAGE) and MALDI-TOF MS can be an efficient way to identify prominent proteins from most organisms [25], but peptide assignment can become more difficult for complex eukaryotic proteomes. In this approach, proteins are separated based on their molecular weight, empirical information that can also be used to identify proteins of interest. Once separated in the gel, protein spots can be selected based on quantitative or qualitative criteria (*e.g.*, changes in abundance as measured by spot intensity under certain experimental conditions) or based on the predicted molecular weight of the target(s). Trypsin digested proteins are analyzed using mass spectrometry, often but not necessarily MALDI-TOF MS, or tandem mass spectrometry. One step up the complexity ladder, 2-dimensional electrophoresis incorporates separation by isoelectric focusing, which enhances the ability to distinguish between proteins of similar size. With difference gel electrophoresis, independently generated samples are labeled with complementary fluorescence dyes and subsequently run on the same gel, reducing inter-gel variability and making it easier to perform relative quantitation. As with 1-dimensional PAGE, protein spots are excised, digested, and analyzed using mass spectrometry [26]. Gel-free methods, wherein proteins are generally digested prior to digestion, are also widely used. The classic example is multidimensional protein identification technology (MudPIT), wherein peptides are separated based on charge and hydrophobicity in a 2-dimensional liquid chromatography column containing a strong cation exchange resin followed by a reverse phase C18 resin [27]. Because this method already entails liquid chromatography, it can be directly interfaced with electrospray ionization, often coupled to tandem mass spectrometers. Quantitation can be performed *via* the addition of internal standards, isobaric tags, or label-free methods such as spectral counting [28]. With the ever increasing instrument power, shotgun proteomics can also bypass complex separation techniques in favor of a single chromatography step to separate peptides that are then analyzed using high resolution tandem mass spectrometry. While this method is less labor intensive, orthogonal information, such as a protein's molecular weight and isoelectric point, is not collected and thus cannot be used to validate identifications.

Data analysis*.* Regardless of the sample preparation and data acquisition methods, the resulting spectra, which for single MS consist of a list of peptide masses with their corresponding intensities and for MS/MS the additional fragment ion spectra from the chosen precursor ions, are interpreted with a limited number of bioinformatic techniques, all of which, with the exception of *de novo*, depend on sequence databases. For organisms whose genomes have been sequenced and

accurately annotated, we have the luxury of using databases exclusively comprising protein-coding sequences from that organism. Alternatively, databases can be constructed from transcripts, as in the NCBI expressed sequence tags (EST) database. The EST database, designed from cDNA sequences, has been used to create customized protein databases for several eukaryotic non-model species, resulting in improved protein identification compared to public databases as previously illustrated with the example of the trematode *Opisthorchis viverrini* [29]. The advantage of using databases generated from cDNA as opposed to genomes is that they tend to contain fewer irrelevant entries, *e.g.*, incorrectly annotated open reading frames, non-coding sequences, incorrect splice variants. However, ESTs are short and thus the protein sequence truncated, which decreases the probability of observing the corresponding peptide, and the collection ESTs for any given organism may be incomplete. Although the ideal is to have a sequence database corresponding as close as possible to the phenotypic reality of the experimental organism, *i.e.*, the database contains only expressed proteins from that organism, proteins can sometimes be identified by homology. In its simplest form, homology-driven proteomics entails the use of a sequence database from one or more close relatives to the experimental organism, permitting the identification of conserved proteins [30, 31]. More intricate strategies use, for example, similarity-based searching as opposed to identity-based searching [32, 33] to account for more mismatches between the empirical data and the reference sequences or composite databases [34, 35] that allow for ever greater sequence variation by allowing the identification of a protein from a combination of sequence variants found in multiple closely related organisms. In the absence of any database information, MS/MS spectra can be sequenced *de novo* by correlating the intervals between fragment ion masses with amino acids [36]. However, *de novo* peptide sequencing is time-consuming and requires very high quality spectra, preferably with complete fragment ion series. Finally, it is worth noting that for organisms with polyploid genomes, protein identification is not trivial, and methods which allow distinguishing unambiguous from shared peptides should be taking into consideration.

4. A SURVEY OF SOME RECENT ACHIEVEMENTS WITH TRADITIONAL APPROACHES

Non-model plants. Land plants related to human nutrition are interesting for both fundamental and applied research. Plants show a remarkable resistance to environmental challenges, such as drought, infection, and parasites. Despite the difficulties of such studies, non-model plants are of high interest to understand this resistance and its potential adaptation to global climate changes, as

illustrated by the following selected examples. The *Solidago canadensis* from the *Asteracea* family is an introduced plant species native to North America that successfully invaded Europe and Asia. Spontaneous colonization of this plant in multiplycontaminated former cokery soil exhibiting high contents of heavy metals and polycyclic aromatic hydrocarbons, as well as high salinity, was documented in France. The molecular mechanisms underlying this remarkable tolerance of ecosystem anthropization were investigated using 2D-DIGE comparative proteomics of leaves from control and exposed populations [31]. Over 1800 protein spots were detected, among which 96 were found differentially expressed. These were analyzed by MALDI-TOF/TOF and proteins were identified *via* queries of the NCBI protein database restricted to the *Viridiplantae* (green plant) and *Asteridae* (subclass) EST database. A total of 63 proteins were identified based on the relative similarities of these proteins to those already catalogued in the databases [31]. The American beech *Fagus grandifolia* is an ecologically important species in many North American forests and is currently impacted by Beech bark disease, an insect-fungus complex. In a long-term affected forest, only one tree in one hundred remains disease free. To decipher resistance biomarkers to orient breeding programs, protein expression of healthy and infected organisms was investigated using 2-DE coupled with nanoLC-MS/MS. Because of high genetic complexities due to high heterozygosis, proteomics of these trees is particularly challenging. While protein identification using SwissProt database was unproductive, search with a custom database compiling 24 individual EST libraries of the *Fagacea* family, *i.e.*, Red Oak (*Quercus rubra* L.), White Oak (*Quercus alba* L.), American chestnut (*Castanea dentata* (Marsh.) Borkh) and Chinese chestnut (*Castanea mollissima* Blume), proved to be much more efficient [37]. The pomegranate, botanical name *Punica granatum*, is a small tree with atypical edible fruit commonly found throughout the Mediterranean region. The proteome of this non-model plant was investigated using a gel-free shotgun proteomics approach [30]. Because its genome sequence is not yet established, the MS/MS data were analyzed with three distinct plant protein databases, and proteins were assigned by homology with other plant species, resulting in a convincing list of 1,488 proteins.

Non-model animals*.* While many model laboratory animal lines have been established, there are still more non-model animals of great interest. For example, tropical parasites found in poorer countries have been often neglected by biomedical research and are thus poorly characterized. In spite of the difficulties posed by working with non-model animals, it was determined that infection with *Opisthorchis viverrini* was at the origin of a certain liver cancer. Proteins secreted by the

parasite are believed to play an important role in predisposing the liver cells to cancer. These proteins were resolved using OFFGEL isoelectrofocusing, followed by nanoLC-MS/MS with multiple reaction monitoring and identification based on EST sequences [29]. Waterfleas, also known as daphniids, are attractive biological subjects in ecology, evolution and environmental science. While *Daphnia pulex* can be now considered a model organism because of the plethora of omics data available and the large scientific community involved in its study, closely related species such as *D. longicephala* are relatively poorly understood. This species, characterized by its unusual morphological plasticity, is an excellent potential model to study phenotypic plasticity. However, protein identification by homology-proteomics remains difficult, even if the sequence of a closely related species is available. Indeed, only 317 proteins could be identified in *D. longicephala* using the draft version of the *D. pulex* genome from a large dataset comprising 407,778 peptide MS/MS spectra [38]. The marine mussel *Mytilus edulis* is an aquatic invertebrate with external fertilization subjected to intensive aquaculture. The protein expression in eggs and sperm of this non-model organism was investigated using 2-DE [39, 40]. In this case, protein identifications were first attempted with MALDI-TOF with poor results due to the paucity of *Mytilus* database entries, but MS/MS significantly increased the number of identifications along with their confidence. The exploration of the venom proteome from snakes is complicated due to the absence of genome sequences in most cases and the presence of multiple isoforms of proteinases. Nevertheless, such studies have been performed for *Crotalus atrox* by direct analysis and following treatment with solid-phase combinatorial peptide ligand libraries for enrichment of low abundant proteins, 2-DE separation of proteins and MS characterization of in-gel digested protein bands [41].

Pathogens, parasites, and exotic organisms take the stage. While homology-based proteomic studies are successful in higher organisms, many algae, fungi, and other unicellular organisms which are really diverse in terms of phylogeny due to their ancient evolution, may be more recalcitrant to such strategies because of the lack of genomic and EST sequences of closely related organisms. *Phytophthora cinnamomi* (oomycetes class) is a soil-borne water mould that infects plants causing root rot or dieback. It is considered one of the world's most invasive plant pathogens. The responses of *Quercus ilex*, the holm oak tree native to the Mediterranean region, to *P. cinnamomi* have been studied by 2-DE and MALDI-TOF/TOF identification of proteins. Ninety seven proteins that changed abundance in response to the inoculation were successfully identified and correlated with the physiology of plants subjected to drought stress [42]. The boundaries of life are an

attractive area of research for novel biotechnology applications as well as to understand the origin and evolution of life. To that end, different extremophiles have been analyzed in terms of genomics and proteomics [43]. Psychrophiles such as the bacterium *Shewanella lingstonensis* [44], radiation tolerant thermophiles such as the archaeon *Thermococcus gammatolerans* [45], or the remarkable arsenic tolerant *Exiguobacterium* sp. S17 [46] are some intriguing examples of how proteomic approaches can explain the adaptive response of life.

Megaviruses masquerading as normal cellular organisms. The recent discovery of giant viruses without any resemblance to any other virus families challenges our view of the viral and cellular worlds. These giant viruses are on the order of a micrometer in size and their genome complexity is similar to that of archaea and bacteria with at least 1,000 genes. First, the mimivirus was found infecting an amoeboid protozoa, namely *Acanthamoeba polyphaga* [47]. Its genome sequence revealed the existence of 1,262 putative open reading frames, 90% of which did not exhibit any similarity to proteins of known function. Later, a virophage called Sputnik was found parasitizing the giant mimivirus [48]. Numerous novel taxa of giant viruses have since been documented. For example, pandoraviruses have been isolated from marine sediments and mud [49]. Although it is difficult to obtain sufficient quantities of these megaviruses, a shotgun proteomic analysis of purified pandoraviruses particles confirmed the use of the universal genetic code and demonstrated the presence of novel open reading frames with no known homologs [49].

5. PROTEOGENOMICS, SENSU STRICTO AND SENSU LATO

Proteogenomics, *sensu stricto*. Strictly speaking, proteogenomics is the use of proteomic data to improve the annotation of genomes (**Figure 3**). Generally, genome annotation proceeds *via* pipelines constructed using our current understanding of gene annotation and therefore prone to errors and inconsistencies [50], and at present pan-genome annotation using whole genome multiple alignment is limited to well documented phyla [51]. First coined by [52], the adjective “proteogenomic” describes the mapping of peptides identified by tandem mass spectrometry during a proteome analysis to the specific locus on the nucleic acid sequence coding for these peptides. In this case, the MS/MS spectra are assigned using a database comprising all possible open reading frames encoded by a six frame translation of the nucleic acid molecules.

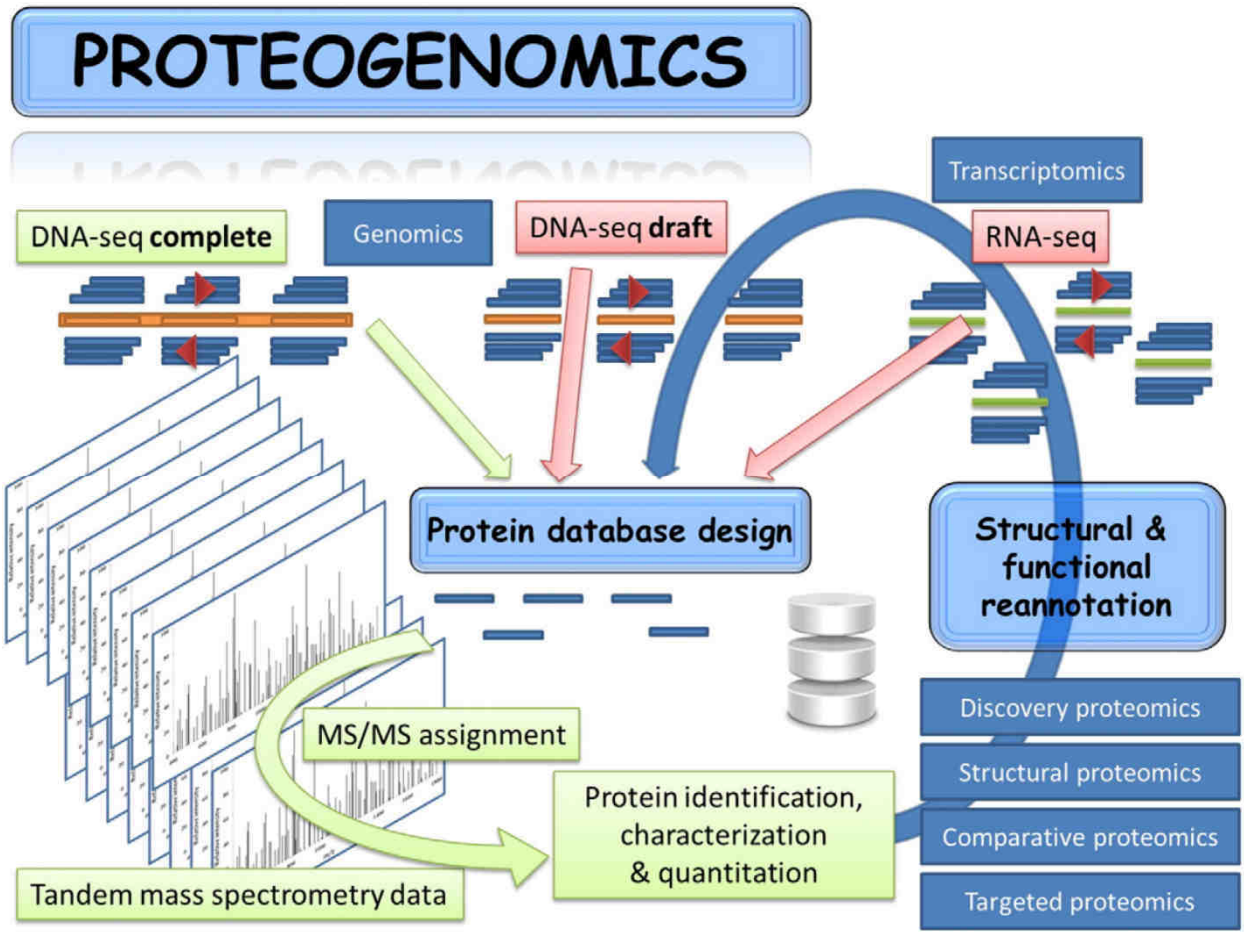


Figure 3: Proteogenomics for non-model organisms

Typically, a six reading frame translation of the nucleic acid information (either complete or draft DNA-seq or RNA-seq data) generates a first protein database. The MS/MS spectra recorded by tandem mass spectrometry are assigned to peptide sequences and the resulting mass spectrometry-certified proteins can be further structurally and functionally characterized, and thus can improve the protein database. These can be the targets of any proteomic strategy: discovery proteomics [103], comparative proteomics [104, 105], structural proteomics [106], targeted proteomics [107],

An extensive proteogenomic map obtained from a dataset comprising 9,709 unique peptides revealed 16 new open reading frames not originally predicted by genomic methods from *Mycoplasma pneumonia* [52]. Since that study, numerous proteogenomic reannotations of genomes have been reported (see for recent reviews: [53, 54]. The largest study to date was performed by Venter et al. [55] who reported the proteogenomic analysis of 46 organisms at once. Experimental evidence was found for 682 novel proteins and highlighted the exceptionally high misannotation rate of the *Deinococcus radiodurans* genome published in 1999. For some bacteria and archaea, proteogenomic annotation has been performed in parallel with genome sequencing:

Mycoplasma mobile [56], *Deinococcus deserti* [57], *Thermococcus gammatolerans* [45], *Campylobacter concisus* [58], and *Acholeplasma laidlawii* [59]. In addition to the discovery of novel coding sequences, refinement of the structural annotation of genes is obtained by systematically cataloging the N-terminal peptides, whereby the exact translation initiation codon or the precise peptide signal maturation can be identified [60]. The largest reported study on N-terminal protein processing was performed on 57 bacterial species from an outsized MS/MS dataset comprising 112 million spectra [61]. To reduce the dependence on such large datasets and thus data processing time, specific methods to label N-terminal peptides have been proposed, such as the use of light/heavy permanently positively charged reagents [62], as well as their immunocapture [63]. The use of non-canonical start codons for translation in *D. deserti* has been evidenced with this N-terminal peptide-oriented strategy [64].

Recent trends in proteogenomics*.* The proteogenomic field is actively improving with constant development in bioinformatics and statistics [65]. GFS was a precursor program to match MS/MS-derived peptide sequences to genomic loci [66]. Peppy is a software tool that, in addition to tracking peptide loci on the genome, is able to perform decoy database searches to restrict identification to a specified false discovery rate threshold [67]. GenoSuite is a recently developed proteogenomic pipeline to validate, refine, and discover protein coding genes from prokaryotes with peptide sequences established using MS/MS [68]. This tool was used to efficiently reannotate the bacterium *Bradyrhizobium japonicum*. Another pipeline, PepLine, was proposed for mapping MS/MS data onto genomic sequences from eukaryotes and for detecting the intron/exon structure of genes taking into account peptide tags [69]. Another automated proteogenomic program was developed and recently applied to reveal 165 novel genes and update 741 genes in *Zea mays* [70]. The PG Nexus pipeline was created to simultaneously visualize proteomic data with genomes and transcriptomes [71]. The database used in proteogenomic analyses is a six frame translation of the nucleic acid sequence and thus contains many spurious protein sequences. The abnormal database size tends to underestimate the confidence of peptide assignments [72], while large proteomic datasets inflate the number of false positives as demonstrated by the deep coverage of the *E. coli* proteome [73]. Jagtap et al. [74] proposed to perform a two-step database search to improve the number of peptide sequence matches with a primary search for exploring the space and create a sub-database for a more confident second search. Regardless, stringent parameters and statistics should be set in any proteogenomic workflow in order to avoid any novel annotation errors and

their propagation when new, relatively close genomes are annotated. Moreover, the workflow should rely on experimental data acquired with high mass accuracy instruments and from perfectly reliable biological samples.

Proteogenomics, *sensu lato*. Over the course of time, the word "proteogenomics" has been used to describe research projects where proteomics and genomics are intimately linked, albeit not directly aimed at genome annotation or reannotation (**Figure 3**). For example, the phytopathogenic mollicute *Spiroplasma melliferum* is the provocative agent of spiroplasmosis in the bee *Apis mellifera*. Proteomic analysis was performed after sequencing its genome, which was assembled into 4 megacontigs and annotated [75]. This study simultaneously presented the genome, proteome, and metabolome of this bacterium. The authors mentioned the word "proteogenomics" in the title of their study but not in the sense of the original definition related to genome reannotation. We consider that some recent approaches fit a loose definition of "proteogenomics". We recently analyzed the key proteins involved in the maintenance of osmotic pressure from the Andean alphaproteobacterium *Tistlia consotensis* [76]. In this case, the genome draft was created using high-throughput DNA sequencing without the intention of annotating the genome but rather to create a translation database for the interpretation of MS/MS spectra. DNA sequencing resulted in a large number of contigs. This genomic information was used to create a database of open reading frames through a six frame translation, which allowed the establishment of a proteome catalog under various physiological conditions. Without the exact genome of the protozoan parasite *Leishmania donovani*, Pawar et al. [77] used a proteogenomic-flavored strategy taking into account the three most related genomes, *i.e.* those of *Leishmania major*, *Leishmania infantum*, and *Leishmania braziliensis*, translated into the six possible reading frames. A total of 3,711 proteins were identified and the dataset was used to refine certain genes from the reference genomes. We also explored a mass spectrometry-based method to screen new bacterial isolates of a given genus [78] or even clade [9]. In this case, proteogenomics was used to identify the most reliable biomarkers for screening, along with their post-translational modifications. Based on these examples, we propose that proteogenomics *sensu lato* should be used to qualify projects based on multi-omics data including closely linked nucleic acid and protein information, including typically the use of six frame translation.

6. THE WINNING ALLIANCE OF GENOMICS AND PROTEOMICS FOR NON-MODEL ORGANISMS

Unfinished genome sequences bypass time-consuming sequence assembly and annotation. Homology-driven proteomics is tedious when closely related genome sequences are not available, and in any case the results are restricted to the most conserved peptides from the most conserved proteins, so sequencing the genome of a non-model organism is the most straightforward approach to create the most appropriate protein sequence database. Sequencing efforts should be directly proportional to the expected complexity of the genome and whether related organisms have been already sequenced, as they could serve as possible scaffold for genome assembly. While a nearly finished genome sequence is usually expected to begin the structural and functional annotation of genes, with the addition of proteomic information, annotation can actually be performed on unfinished genome sequences with low genome coverage. As pioneered by Nanduri et al. [79] with the *Mannheimia haemolytica* bacterium, proteomic analysis can be performed with a partial database. The sequence coverage should be reasonably good to obtain comprehensive coverage of the theoretical proteome [80]. This "quick and dirty" genome approach allows fast and reliable protein identification in shotgun analyses, as exemplified with the *Tistlia consotensis* proteogenomic analysis [76]. In this case, the global sequence coverage was 10.4 fold, but the high G+C content (70.4 %) led to a poor assembly of the genome sequence. The 2,377 resulting contigs were merged and treated as a single chimeric chromosome. The whole nucleic acid sequence can be either annotated to extract protein coding sequences (CDS) or directly translated into open reading frames in the six possible frames. These predicted CDS or other open reading frames are then used to interpret the MS/MS data, and only the tandem mass spectrometry-identified open reading frames are then functionally annotated *via* homologs obtained from Blast queries. This approach is especially attractive for non-model organisms with small genomes. Besides, striving to finalize a genome annotation should be considered as really worth for specific representatives, especially for the first genome of a given lineage, because the recent large increase in the number of draft genome sequences led to the propagation of annotation errors.

RNA-seq generates directly informative mature protein sequences. In contrast to prokaryotes, eukaryotic genomes contain larger amounts of intragenic and intergenic non-protein coding sequences, with a large ratio of "junk" DNA containing highly repetitive sequences, and as a result only a small portion of the genome encodes proteins. Moreover, due to alternative splicing the protein database generation from the genome sequence will be more difficult for eukaryotic

organisms. Consequently, sequencing only mature RNAs is an attractive alternative for the rapid identification of protein sequences. Compared to partial coverage EST libraries first established for some eukaryotic non-model species, high-throughput RNA sequencing using next-generation sequencer technologies (RNA-Seq) allows deep and cost-effective transcriptome coverage [81]. RNA-seq has been successfully applied to several non-model species, such as the tropical forest tree *Shorea beccariana* [82], the lotus *Nelumbo nucifera* [83], the fungus *Amanita exitialis* [84], the Atlantic Cod *Gadus morhua* [85], and the marine copepod *Calanus finmarchicus* [86]. RNA-seq has been used to improve the annotation of genomes by revealing new gene models and alternative splicing events as pioneered by Denoeud et al. [87]. Essentially, protein identification can be performed on the basis of RNA-seq nucleic acid sequence information without the added inconvenience of whole genome sequencing and annotation [88-91]. Moreover, protein abundances observed by mass spectrometry and RNA-seq estimates correlate well [92, 93], especially if translating mRNAs, *i.e.*, mRNA attached ribosome-nascent chain complexes, are analyzed [94]. This correlation is exemplified by the characterization of both the transcriptome and proteome of the common reed *Phragmites australis* [22] performed to better understand its invasion process. In this study, the transcriptome was sequenced using a combination of Illumina and 454 pyrosequencing technologies, yielding a total of 124,450 unique transcripts. The transcripts were translated into open reading frames, which allowed the identification of 1,280 proteins detected by tandem mass spectrometry [22]. Moreno et al [95] reported the RNA-seq analysis of the murine parasite *Heligmosomoides polygyrus*, a nematode. The analysis resulted in the generation and annotation of 33,641 transcripts, which were used to create an open reading frame database and identify 209 proteins by shotgun proteomics. The proteome of the domesticated tomato, *Solanum lycopersicum*, was also efficiently analyzed using a RNA-seq derived database [96]. RNA-seq technologies are inexpensive and effective for non-model eukaryotes. The main drawback of the approach is the large size of the transcript database resulting from the assembly of the sequence reads. Because the reading sense is unknown, each transcript needs to be translated into all six possible reading frames. Moreover, the large redundancy of small reads increases nucleotide uncertainties and assembling errors that result in protein sequence errors, penalizing frameshifts and premature terminations of open reading frames. Wang & Zhang [97] proposed a tool to generate customized databases from RNA-seq for proteomics. Novel approaches to handle such databases are currently under development, such as that proposed by Woo et al. [91] with compression of redundant data and splice graph compiling of exon junctions. To design the most

appropriate protein sequence database from RNA-seq, the mRNAs should comprehensively represent the protein space and thus, ideally, the same sample should be for both RNA-seq and proteomics analyses.

7. PERSPECTIVES AND CONCLUDING REMARKS

The vast technical improvements in proteomics that have occurred over the last two decades have largely been developed with or gauged against *S. cerevisiae*. While this progress certainly owes a lot to this extensively well studied model organism, the giant leaps of the future will be taken using more challenging non-model organisms. It is now relatively easy to get a draft genome sequence of any isolated microorganism or alternatively a draft transcriptome sequence using high-throughput RNA-seq. However, the difficulties in assembling short sequencing reads as well as managing and storing huge amounts of informatics data delay the use of these sequences. Incompleteness notwithstanding, this draft information allows the design of the most appropriate protein database from which valid coding sequencing and specific proteins of interest can be determined using tandem mass spectrometry. The recent achievements in bioinformatics for the field of proteogenomics are paving the way for novel integrative pipelines where genome, RNA-seq transcriptome, and proteome information will be handled and exploited in the most convenient and fruitful way. At present, single-cell proteogenomics is still fully *terra incognita*, mainly because the analytical approaches to perform sequencing of nucleic acid molecules and identification of proteins are so different. However, such long-term perspectives could be really rewarding if applied to the wide number of non-cultivable microorganisms, especially unicellular eukaryotes. Those who specialize in protein biochemistry and mass spectrometry and those who specialize in genomics should profit from one another to fully take advantage of all that proteogenomics has to offer for non-model organisms. Non-model organisms can still be difficult experimental organisms to handle, but their molecular mechanisms can be deciphered quicker than ever thanks to next-generation nucleic acid sequencers and next-generation mass spectrometers. The inter-species differences that had such an impression on Charles Darwin, but also intra-species differences (population polymorphism), are still as important as ever, but with the increased power of proteogenomics, we are in a position to respect those differences and study more individual organisms insteading of relying on the proxy of model organisms. With the union of next-generation of sequencers and mass spectrometers, non-model organisms are a species endangered by proteogenomics!

Acknowledgements: This work was supported by the Commissariat à l'Energie Atomique et aux Energies Alternatives via the CEA transverse toxicology program, the Agence Nationale de la Recherche (ANR-12-BSV6-0012-01), the Région Languedoc-Roussillon (label « Chercheur d'Avenir Confirmé » 2010), and IRSTEA. EMH was supported by a Fulbright grant.

REFERENCES

[1] Koonin EV, Wolf YI. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? Front Cell Infect Microbiol. 2012;2:119.
[2] Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, et al. Cryptic diversity as a window on diversity and conservation. Trends in Ecology and Evolution. 2007;22:148-55.
[3] Costello MJ, May RM, Stork NE. Can we name Earth's species before they go extinct? Science. 2013;339:413-6.
[4] Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. How many species are there on Earth and in the ocean? PLoS Biol. 2011;9:e1001127.
[5] Scheffers BR, Joppa LN, Pimm SL, Laurance WF. What we know and don't know about Earth's missing biodiversity. Trends Ecol Evol. 2012;27:501-10.
[6] Schloss PD, Handelsman J. Status of the microbial census. Microbiol Mol Biol Rev. 2004;68:686-91.
[7] Schloss PD, Handelsman J. Toward a census of bacteria in soil. PLoS Comput Biol. 2006;2:e92.
[8] Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013;499:431-7.
[9] Christie-Oleza JA, Miotello G, Armengaud J. Proteogenomic definition of biomarkers for the large roseobacter clade and application for a quick screening of new environmental isolates. J Proteome Res. 2013;12:5331-9.
[10] Becher D, Bernhardt J, Fuchs S, Riedel K. Metaproteomics to unravel major microbial players in leaf litter and soil environments: challenges and perspectives. Proteomics. 2013;13:2895-909.
[11] Armengaud J. Proteogenomics and systems biology: quest for the ultimate missing parts. Expert Rev Proteomics. 2010;7:65-77.
[12] Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, et al. The genome of the model beetle and pest Tribolium castaneum. Nature. 2008;452:949-55.
[13] Celander MC, Goldstone JV, Denslow ND, Iguchi T, Kille P, Meyerhoff RD, et al. Species extrapolation for the 21st century. Environ Toxicol Chem. 2011;30:52-63.
[14] Baird DJ, Brown SS, Lagadic L, Liess M, Maltby L, Moreira-Santos M, et al. In situ-based effects measures: determining the ecological relevance of measured responses. Integr Environ Assess Manag. 2007;3:259-67.
[15] Ahrens CH, Schrimpf SP, Brunner E, Aebersold R. Model organism proteomics. J Proteomics. 2010;73:2051-3.
[16] Ankeny RA, Leonelli S. What is so special about model organisms? Studies in the history and the phylosophy of science. 2011;42:313-23.
[17] Muller B, Grossniklaus U. Model organisms--A historical perspective. J Proteomics. 2010;73:2054-63.
[18] Clark MS, Thorne MA, Toullec JY, Meng Y, Guan LL, Peck LS, et al. Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. PLoS One. 2011;6:e15919.
[19] Pejchar L, Mooney HA. Invasive species, ecosystem services and human well-being. Trends Ecol Evol. 2009;24:497-504.
[20] Pedriali A, Riva C, Parolini M, Cristoni S, Sheehan D, Binelli A. A redox proteomic investigation of oxidative stress caused by benzoylecgonine in the freshwater bivalve Dreissena polymorpha. Drug Test Anal. 2013;5:646-56.

[21] Riva C, Binelli A, Rusconi F, Colombo G, Pedriali A, Zippel R, et al. A proteomic study using zebra mussels (D. polymorpha) exposed to benzo(alpha)pyrene: the role of gender and exposure concentrations. Aquat Toxicol. 2011;104:14-22.
[22] He R, Kim MJ, Nelson W, Balbuena TS, Kim R, Kramer R, et al. Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, Phragmites australis (Poaceae), reveals genes involved in invasiveness and rhizome specificity. Am J Bot. 2012;99:232-47.
[23] Newman DJ, Cragg GM, Snader KM. The influence of natural products upon drug discovery. Nat Prod Rep. 2000;17:215-34.
[24] Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF, et al. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. ISME J. 2011;5:856-65.
[25] Hartmann EM, Colquhoun DR, Halden RU. Identification of putative biomarkers for toluene-degrading Burkholderia and pseudomonas by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and peptide mass fingerprinting. Biosci Biotechnol Biochem. 2010;74:1470-2.
[26] Winnik WM, Dekroon RM, Jeong JS, Mocanu M, Robinette JB, Osorio C, et al. Analysis of proteins using DIGE and MALDI mass spectrometry. Methods Mol Biol. 2012;854:47-66.
[27] Webb KJ, Xu T, Park SK, Yates JR, 3rd. Modified MuDPIT separation identified 4488 proteins in a system-wide analysis of quiescence in yeast. J Proteome Res. 2013;12:2177-84.
[28] Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem. 2012;404:939-65.
[29] Mulvenna J, Sripa B, Brindley PJ, Gorman J, Jones MK, Colgrave ML, et al. The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke Opisthorchis viverrini. Proteomics. 2010;10:1063-78.
[30] Capriotti AL, Caruso G, Cavaliere C, Foglia P, Piovesana S, Samperi R, et al. Proteome investigation of the non-model plant pomegranate (Punica granatum L.). Anal Bioanal Chem. 2013;405:9301-9.
[31] Immel F, Renaut J, Masfaraud JF. Physiological response and differential leaf proteome pattern in the European invasive Asteraceae Solidago canadensis colonizing a former cokery soil. J Proteomics. 2012;75:1129-43.
[32] Junqueira M, Spirin V, Balbuena TS, Thomas H, Adzhubei I, Sunyaev S, et al. Protein identification pipeline for the homology-driven proteomics. J Proteomics. 2008;71:346-56.
[33] Liska AJ, Sunyaev S, Shilov IN, Schaeffer DA, Shevchenko A. Error-tolerant EST database searches by tandem mass spectrometry and multiTag software. Proteomics. 2005;5:4118-22.
[34] Fagerquist CK. Amino acid sequence determination of protein biomarkers of Campylobacter upsaliensis and C. helveticus by "composite" sequence proteomic analysis. J Proteome Res. 2007;6:2539-49.
[35] Fagerquist CK, Yee E, Miller WG. Composite sequence proteomic analysis of protein biomarkers of Campylobacter coli, C. lari and C. concisus for bacterial identification. Analyst. 2007;132:1010-23.
[36] Chong KF, Leong HW. Tutorial on de novo peptide sequencing using MS/MS mass spectrometry. J Bioinform Comput Biol. 2012;10:1231002.
[37] Mason ME, Koch JL, Krasowski M, Loo J. Comparisons of protein profiles of beech bark disease resistant and susceptible American beech (Fagus grandifolia). Proteome Sci. 2013;11:2.
[38] Frohlich T, Arnold GJ, Fritsch R, Mayr T, Laforsch C. LC-MS/MS-based proteome profiling in Daphnia pulex and Daphnia longicephala: the Daphnia pulex genome database as a key for high throughput proteomics in Daphnia. BMC Genomics. 2009;10:171.
[39] Diz AP, Dudley E, MacDonald BW, Pina B, Kenchington EL, Zouros E, et al. Genetic variation underlying protein expression in eggs of the marine mussel Mytilus edulis. Mol Cell Proteomics. 2009;8:132-44.
[40] Diz AP, Dudley E, Skibinski DO. Identification and characterization of highly expressed proteins in sperm cells of the marine mussel Mytilus edulis. Proteomics. 2012;12:1949-56.
[41] Calvete JJ, Fasoli E, Sanz L, Boschetti E, Righetti PG. Exploring the venom proteome of the western diamondback rattlesnake, Crotalus atrox, via snake venomics and combinatorial peptide ligand library approaches. J Proteome Res. 2009;8:3055-67.

[42] Sghaier-Hammami B, Valero-Galvan J, Romero-Rodriguez MC, Navarro-Cerrillo RM, Abdelly C, Jorrin-Novo J. Physiological and proteomics analyses of Holm oak (Quercus ilex subsp. ballota [Desf.] Samp.) responses to Phytophthora cinnamomi. Plant Physiol Biochem. 2013;71:191-202.
[43] Burg D, Ng C, Ting L, Cavicchioli R. Proteomics of extremophiles. Environ Microbiol. 2011;13:1934-55.
[44] Park J, Kawamoto J, Esaki N, Kurihara T. Identification of cold-inducible inner membrane proteins of the psychrotrophic bacterium, Shewanella livingstonensis Ac10, by proteomic analysis. Extremophiles. 2012;16:227-36.
[45] Zivanovic Y, Armengaud J, Lagorce A, Leplat C, Guerin P, Dutertre M, et al. Genome analysis and genome-wide proteomics of Thermococcus gammatolerans, the most radioresistant organism known amongst the Archaea. Genome Biol. 2009;10:R70.
[46] Belfiore C, Ordonez OF, Farias ME. Proteomic approach of adaptive response to arsenic stress in Exiguobacterium sp. S17, an extremophile strain isolated from a high-altitude Andean Lake stromatolite. Extremophiles. 2013;17:421-31.
[47] La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. Science. 2003;299:2033.
[48] La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, et al. The virophage as a unique parasite of the giant mimivirus. Nature. 2008;455:100-4.
[49] Philippe N, Legendre M, Doutre G, Coute Y, Poirot O, Lescot M, et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013;341:281-6.
[50] Poptsova MS, Gogarten JP. Using comparative genome analysis to identify problems in annotated microbial genomes. Microbiology. 2010;156:1909-17.
[51] Angiuoli SV, Dunning Hotopp JC, Salzberg SL, Tettelin H. Improving pan-genome annotation using whole genome multiple alignment. BMC Bioinformatics. 2011;12:272.
[52] Jaffe JD, Berg HC, Church GM. Proteogenomic mapping as a complementary method to perform genome annotation. Proteomics. 2004;4:59-77.
[53] Armengaud J, Hartmann EM, Bland C. Proteogenomics for environmental microbiology. Proteomics. 2013;13:2731-42.
[54] Renuse S, Chaerkady R, Pandey A. Proteogenomics. Proteomics. 2011;11:620-30.
[55] Venter E, Smith RD, Payne SH. Proteogenomic analysis of bacteria and archaea: a 46 organism case study. PLoS One. 2011;6:e27587.
[56] Jaffe JD, Stange-Thomann N, Smith C, DeCaprio D, Fisher S, Butler J, et al. The complete genome and proteome of Mycoplasma mobile. Genome Res. 2004;14:1447-61.
[57] de Groot A, Dulermo R, Ortet P, Blanchard L, Guerin P, Fernandez B, et al. Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium Deinococcus deserti. PLoS Genet. 2009;5:e1000434.
[58] Deshpande NP, Kaakoush NO, Mitchell H, Janitz K, Raftery MJ, Li SS, et al. Sequencing and validation of the genome of a Campylobacter concisus reveals intra-species diversity. PLoS One. 2011;6:e22170.
[59] Lazarev VN, Levitskii SA, Basovskii YI, Chukin MM, Akopian TA, Vereshchagin VV, et al. Complete genome and proteome of Acholeplasma laidlawii. J Bacteriol. 2011;193:4943-53.
[60] Armengaud J. A perfect genome annotation is within reach with the proteomics and genomics alliance. Curr Opin Microbiol. 2009;12:292-300.
[61] Bonissone S, Gupta N, Romine M, Bradshaw RA, Pevzner PA. N-terminal protein processing: a comparative proteogenomic analysis. Mol Cell Proteomics. 2013;12:14-28.
[62] Bertaccini D, Vaca S, Carapito C, Arsene-Ploetze F, Van Dorsselaer A, Schaeffer-Reiss C. An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: dN-TOP. J Proteome Res. 2013;12:3063-70.
[63] Bland C, Bellanger L, Armengaud J. Magnetic Immunoaffinity Enrichment for Selective Capture and MS/MS Analysis of N-Terminal-TMPP-Labeled Peptides. J Proteome Res. 2013; in press.
[64] Baudet M, Ortet P, Gaillard JC, Fernandez B, Guerin P, Enjalbal C, et al. Proteomics-based refinement of Deinococcus deserti genome annotation reveals an unwonted use of non-canonical translation initiation codons. Mol Cell Proteomics. 2010;9:415-26.

[65] Castellana N, Bafna V. Proteogenomics to discover the full coding content of genomes: a computational perspective. J Proteomics. 2010;73:2124-35.
[66] Holmes MR, Giddings MC. Using GFS to identify encoding genomic loci from protein mass spectral data. Curr Protoc Bioinformatics. 2008;Chapter 13:Unit 13 9.
[67] Risk BA, Spitzer WJ, Giddings MC. Peppy: proteogenomic search software. J Proteome Res. 2013;12:3019-25.
[68] Kumar D, Yadav AK, Kadimi PK, Nagaraj SH, Grimmond SM, Dash D. Proteogenomic Analysis of Bradyrhizobium japonicum USDA110 Using Genosuite, an Automated Multi-algorithmic Pipeline. Mol Cell Proteomics. 2013;12:3388-97.
[69] Ferro M, Tardif M, Reguer E, Cahuzac R, Bruley C, Vermat T, et al. PepLine: a software pipeline for high-throughput direct mapping of tandem mass spectrometry data on genomic sequences. J Proteome Res. 2008;7:1873-83.
[70] Castellana NE, Shen Z, He Y, Walley JW, Cassidy CJ, Briggs SP, et al. An Automated Proteogenomic Method Utilizes Mass Spectrometry to Reveal Novel Genes in Zea mays. Mol Cell Proteomics. 2013; in press.
[71] Pang CN, Tay AP, Aya C, Twine NA, Harkness L, Hart-Smith G, et al. Tools to Covisualize and Coanalyze Proteomic Data with Genomes and Transcriptomes: Validation of Genes and Alternative mRNA Splicing. J Proteome Res. 2013; in press.
[72] Blakeley P, Overton IM, Hubbard SJ. Addressing statistical biases in nucleotide-derived protein databases for proteogenomic search strategies. J Proteome Res. 2012;11:5221-34.
[73] Krug K, Carpy A, Behrends G, Matic K, Soares NC, Macek B. Deep Coverage of the Escherichia coli Proteome Enables the Assessment of False Discovery Rates in Simple Proteogenomic Experiments. Mol Cell Proteomics. 2013;12:3420-30.
[74] Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, et al. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. Proteomics. 2013;13:1352-7.
[75] Alexeev D, Kostrjukova E, Aliper A, Popenko A, Bazaleev N, Tyakht A, et al. Application of Spiroplasma melliferum proteogenomic profiling for the discovery of virulence factors and pathogenicity mechanisms in host-associated spiroplasmas. J Proteome Res. 2012;11:224-36.
[76] Rubiano-Labrador C, Bland C, Miotello G, Guerin P, Pible O, Baena S, et al. Proteogenomic insights into salt tolerance by a halotolerant alpha-proteobacterium isolated from an Andean saline spring. J Proteomics. 2013; in press.
[77] Pawar H, Sahasrabuddhe NA, Renuse S, Keerthikumar S, Sharma J, Kumar GS, et al. A proteogenomic approach to map the proteome of an unsequenced pathogen - Leishmania donovani. Proteomics. 2012;12:832-44.
[78] Christie-Oleza JA, Pina-Villalonga JM, Guerin P, Miotello G, Bosch R, Nogales B, et al. Shotgun nanoLC-MS/MS proteogenomics to document MALDI-TOF biomarkers for screening new members of the Ruegeria genus. Environ Microbiol. 2013;15:133-47.
[79] Nanduri B, Lawrence ML, Vanguri S, Pechan T, Burgess SC. Proteomic analysis using an unfinished bacterial genome: the effects of subminimum inhibitory concentrations of antibiotics on Mannheimia haemolytica virulence factor expression. Proteomics. 2005;5:4852-63.
[80] Sun J, Wang W, Hundertmark C, Zeng AP, Jahn D, Deckwer WD. A protein database constructed from low-coverage genomic sequence of Bacillus megaterium and its use for accelerated proteomic analysis. J Biotechnol. 2006;124:486-95.
[81] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57-63.
[82] Kobayashi MJ, Takeuchi Y, Kenta T, Kume T, Diway B, Shimizu KK. Mass flowering of the tropical tree Shorea beccariana was preceded by expression changes in flowering and drought-responsive genes. Mol Ecol. 2013;22:4767-82.
[83] Kim MJ, Nelson W, Soderlund CA, Gang DR. Next-Generation Sequencing-Based Transcriptional Profiling of Sacred Lotus 'China Antique'. Tropical plant biology. 2013;6:161-79.

[84] Li P, Deng WQ, Li TH, Song B, Shen YH. Illumina-based de novo transcriptome sequencing and analysis of Amanita exitialis basidiocarps. Gene. 2013;532:63-71.
[85] Lanes CF, Bizuayehu TT, de Oliveira Fernandes JM, Kiron V, Babiak I. Transcriptome of Atlantic Cod (Gadus morhua L.) Early Embryos from Farmed and Wild Broodstocks. Mar Biotechnol (NY). 2013;15:677-94.
[86] Christie AE, Roncalli V, Wu LS, Ganote CL, Doak T, Lenz PH. Peptidergic signaling in Calanus finmarchicus (Crustacea, Copepoda): in silico identification of putative peptide hormones and their receptors using a de novo assembled transcriptome. Gen Comp Endocrinol. 2013;187:117-35.
[87] Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, et al. Annotating genomes with massive-scale RNA sequencing. Genome Biol. 2008;9:R175.
[88] Ning K, Nesvizhskii AI. The utility of mass spectrometry-based proteomic data for validation of novel alternative splice forms reconstructed from RNA-Seq data: a preliminary assessment. BMC Bioinformatics. 2010;11 Suppl 11:S14.
[89] Omasits U, Quebatte M, Stekhoven DJ, Fortes C, Roschitzki B, Robinson MD, et al. Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome. Genome Res. 2013;23:1916-27.
[90] Wang X, Slebos RJ, Wang D, Halvey PJ, Tabb DL, Liebler DC, et al. Protein identification using customized protein sequence databases derived from RNA-Seq data. J Proteome Res. 2012;11:1009-17.
[91] Woo S, Cha SW, Merrihew G, He Y, Castellana N, Guest C, et al. Proteogenomic Database Construction Driven from Large Scale RNA-seq Data. J Proteome Res. 2013; in press.
[92] Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics. 2009;10:161.
[93] Ning K, Fermin D, Nesvizhskii AI. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res. 2012;11:2261-71.
[94] Zhong J, Cui Y, Guo J, Chen Z, Yang L, He QY, et al. Resolving Chromosome-Centric Human Proteome with Translating mRNA Analysis: A Strategic Demonstration. J Proteome Res. 2013; in press.
[95] Moreno Y, Gros PP, Tam M, Segura M, Valanparambil R, Geary TG, et al. Proteomic analysis of excretory-secretory products of Heligmosomoides polygyrus assessed with next-generation sequencing transcriptomic information. PLoS Negl Trop Dis. 2011;5:e1370.
[96] Lopez-Casado G, Covey PA, Bedinger PA, Mueller LA, Thannhauser TW, Zhang S, et al. Enabling proteomic studies with RNA-Seq: The proteome of tomato pollen as a test case. Proteomics. 2012;12:761-74.
[97] Wang X, Zhang B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. Bioinformatics. 2013;29:3235-7.
[98] Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 2012;40:D571-9.
[99] Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci U S A. 2000;97:9390-5.
[100] Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, et al. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal Chem. 2006;78:2113-20.
[101] Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19:242-7.
[102] Wolters DA, Washburn MP, Yates JR, 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73:5683-90.
[103] Malmstrom J, Lee H, Aebersold R. Advances in proteomic workflows for systems biology. Curr Opin Biotechnol. 2007;18:378-84.
[104] Christie-Oleza JA, Pina-Villalonga JM, Bosch R, Nogales B, Armengaud J. Comparative proteogenomics of twelve Roseobacter exoproteomes reveals different adaptive strategies among these marine bacteria. Mol Cell Proteomics. 2012;11:M111 013110.

[105] Toueille M, Mirabella B, Guerin P, Bouthier de la Tour C, Boisnard S, Nguyen HH, et al. A comparative proteomic approach to better define Deinococcus nucleoid specificities. J Proteomics. 2012;75:2588-600.
[106] Walzthoeni T, Leitner A, Stengel F, Aebersold R. Mass spectrometry supported determination of protein complex structure. Curr Opin Struct Biol. 2013;23:252-60.
[107] Maiolica A, Junger MA, Ezkurdia I, Aebersold R. Targeted proteome investigation via selected reaction monitoring mass spectrometry. J Proteomics. 2012;75:3495-513.

5. Objectifs des travaux

Cette thèse est le fruit de la collaboration de deux structures de recherches : le laboratoire d'écotoxicologie de l'Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (Irstea) situé à Villeurbanne sous la direction de Olivier Geffard et le laboratoire de Biochimie des Systèmes Perturbés (LBSP) du Commissariat à l'énergie atomique et aux énergies alternatives (CEA) situé à Marcoule sous la direction de Jean Armengaud.

Les activités de recherche du laboratoire d'écotoxicologie s'intéressent à la compréhension des réponses biologiques aux contaminations du milieu aquatique. Dans le cadre d'appui aux politiques publiques portées notamment par la DCE, une des missions du laboratoire porte sur le développement d'outils de mesure de la qualité des milieux continentaux. Parmi les différents modèles biologiques utilisés au sein du laboratoire, le gammare est particulièrement employé et son utilisation répond à deux types d'objectifs : [1] le développement d'indicateurs de la contamination chimique biodisponible par le biais d'encagement d'organismes standards pour une biosurveillance active des hydrosystèmes (Besse *et al.* 2013) et [2] le développement de marqueurs de toxicité en laboratoire dans une finalité de biosurveillance active. Ces marqueurs se basent sur le développement de bioessais *i.e.* alimentation (Coulaud *et al.* 2011) et reproduction (Geffard *et al.* 2010) ou de marqueurs moléculaires *i.e.* neurotoxicité (Xuereb *et al.* 2009), génotoxicité (Lacaze *et al.* 2011) ou reprotoxicité (Jubeaux *et al.* 2012).

Les activités du LBSP portent sur le développement de nouvelles méthodologies protéomiques traitant de [1] la chimie des protéines, notamment par le marquage des extrémités N-terminales pour l'annotation des codons d'initiation ou d'immuno enrichissement de protéines/peptides d'intérêt (Bland *et al.* 2014), [2] la spectrométrie de masse pour le séquençage à haut débit des protéines par des approches nanoLC-MS/MS et [3] l'intégration des données omiques (carte protéogénomique, expression différentielle des protéines, détection de contamination des séquences génomiques). Ces techniques sont particulièrement mises en application dans le domaine de la microbiologie environnementale : annotation protéogénomique de génomes bactériens (Armengaud 2013), identification de toxine chez une bactérie du clade des *Roseobacter* (Durighello *et al.* 2014), compréhension des mécanismes de dégradation des dioxines par *Sphingomonas wittichii* (Hartmann and Armengaud 2014) ou des mécanismes de résistance aux

conditions extrêmes pour la bactérie halotolérante *Tistlia consotensis* (Rubiano-Labrador *et al.* 2014) ou radiotolérante *Thermococcus gammatolerans* (Zivanovic *et al.* 2009).

Ce travail de thèse propose une approche innovante pour la découverte de biomarqueurs protéiques chez le crustacé amphipode *Gammarus fossarum* par l'emploi des outils omiques. La reproduction est une fonction cruciale pour le maintien de l'espèce dans son milieu et chez les invertébrés, en raison de la grande méconnaissance de sa régulation, les phénomènes de reprotoxicité sont peu étudiés. Ces travaux se sont focalisés sur la découverte de protéines impliquées dans le processus reproductif ainsi que sur l'étude de la perturbation endocrinienne, afin de proposer de nouveaux biomarqueurs, applicables *in fine* dans le cadre de réseaux de surveillance des hydrosystèmes. La stratégie expérimentale employée se compose de trois grandes étapes, présentées en Figure I-33.

Création d'un catalogue de protéines par protéogénomique

Afin de répondre à l'absence de BD en séquences protéiques spécifique à *Gammarus fossarum*, un catalogue de protéines été créé par protéogénomique, en combinant le séquençage du transcriptome et du protéome (Chapitre III, **Publication 3**). Pour cela, quatre tissus d'intérêt ont été choisis de par leur importance dans la physiologie de l'organisme : le céphalon, les testicules et les ovaires en raison de leur implication dans la reproduction. Additionnellement, les *cæca* hépatopancréatiques ont également été étudiés car ce tissu constitue le centre métabolique de l'individu et est impliqué dans différentes réponses d'intérêt en écotoxicologie : acquisition d'énergie, défense immunitaire, reproduction ou détoxification des xénobiotiques (résumé par Zhou *et al.* 2009). Dans un premier temps, le transcriptome de ces tissus a été séquencé par RNA-Seq afin de générer une BD en séquences protéiques à partir des séquences d'ARNm. Puis, les protéines issues des différents tissus ont à leur tour été séquencées par nanoLC-MS/MS.

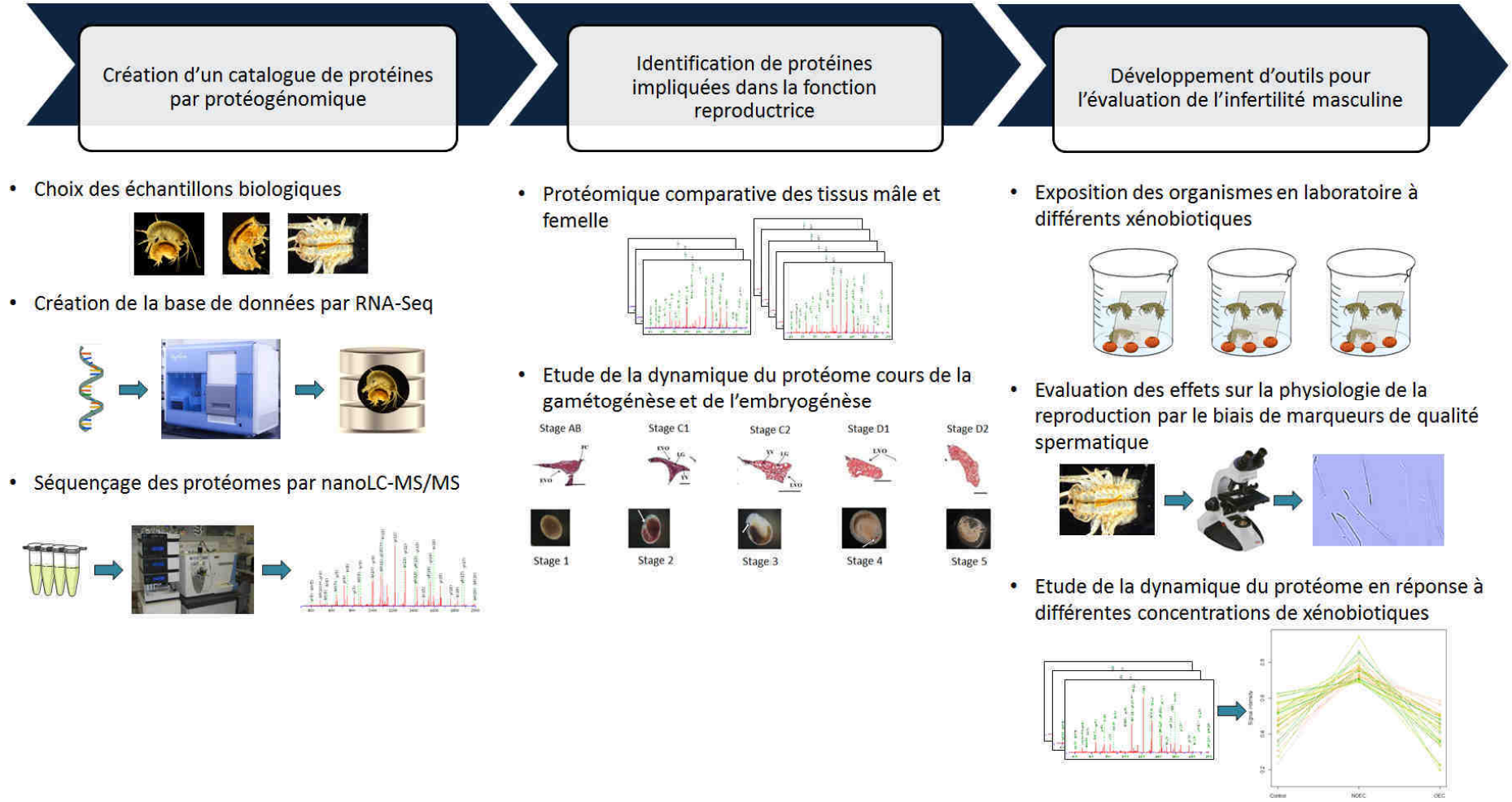


Figure I-33 : Stratégie générale pour le développement de biomarqueurs de reprotoxicité chez *Gammarus fossarum*

Recherche de protéines impliquées dans la fonction reproductrice

Une première approche pour identifier des protéines potentiellement impliquées dans la reproduction de l'organisme est de comparer l'abondance des protéines entre les gonades mâles et femelles afin d'identifier celles présentant un dimorphisme sexuel prononcé. Cette comparaison a été réalisée à partir des données quantitatives obtenues par le séquençage du protéome au cours de la réalisation du catalogue de protéine (Chapitre III, **Publication 3**).

La seconde stratégie repose sur l'étude de l'évolution de l'abondance des protéines au cours de différents processus physiologiques liés à la reproduction, démarche rendue possible chez le gammare grâce à la connaissance de sa physiologie reproductive. La dynamique du protéome des testicules a été étudiée au cours de la spermatogénèse, à sept différents stades (Chapitre III, **Publication 3**). Aussi, la dynamique du protéome des ovaires au cours de l'ovogénèse a été étudiée. Afin d'identifier spécifiquement les protéines de réserves du développement embryonnaire, l'évolution de l'abondance des protéines au cours de l'embryogénèse a également été étudiée (Chapitre III, **Publication 4**).

Développement d'outils pour l'évaluation de l'infertilité masculine

En raison du manque d'études sur les phénomènes d'infertilité masculine, les travaux concernant la découverte de biomarqueurs de reprotoxicité se sont focalisés sur les organismes mâles (Chapitre IV, **Publication 5**). Pour cela, les organismes mâles ont été exposés au cours de deux spermatogénèses successives (15 jours) à trois différents xénobiotiques, le Cadmium et deux pesticides, potentiels perturbateurs endocriniens le methoxyfenozide, ciblant la voie des ecdystéroïdes et le pyriproxyfen ciblant la voie des juvénoïdes. Au même titre que l'étape précédente d'annotation fonctionnelle, cette démarche couple l'étude du protéome des gonades et de la physiologie reproductive par la mesure de marqueurs de qualité spermatique.

CHAPITRE II : Matériels et méthodes

Ce chapitre permet de détailler les choix expérimentaux qui ont été mis en œuvre au cours de ces travaux.

La première partie (II-1) concerne la manipulation du modèle d'étude *G. fossarum*, son prélèvement en milieu naturel et son maintien en laboratoire ainsi que les méthodes de dissection pour la récupération des tissus d'intérêt. Aussi, elle présente les expérimentations effectuées en laboratoire.

La deuxième partie (II-2) se concentre sur les techniques et stratégies mises en œuvre pour les analyses moléculaires transcriptomique et protéomique.

Enfin la dernière partie (II-3) présente le traitement informatique des données de séquençage.

1. Biotests chez *Gammarus fossarum*

1.1. Matériel biologique

1.1.1. Prélèvement des organismes

A l'heure actuelle, les techniques de production de gammares en laboratoire ne sont pas maitrisées. C'est pourquoi, depuis plusieurs années, le laboratoire d'écotoxicologie de l'Irstea utilise une population naturelle, qualifiée de population référence sur un site peu contaminé, présentant d'importante densité d'organismes, échantillonnable tout au long de l'année. Cette population se situe en amont du bassin de la Bourbe, au niveau de la commune de la Tour du Pin en Isère (Lat: 45°569'442"; Long: 5°459'115") (Figure II-1).

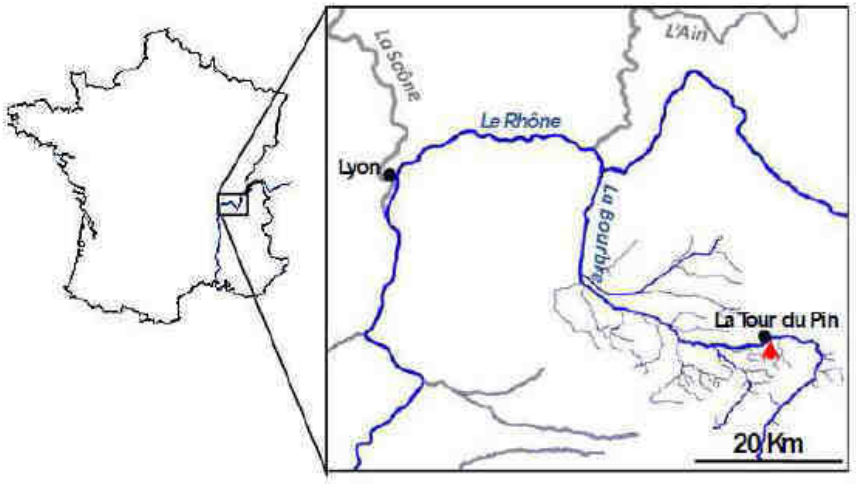


Figure II-1 : Localisation géographique du site de prélèvement de la Tour du Pin

Les organismes sont prélevés selon la méthode du « *kick-sampling* » à l'aide d'un troubleau (base rectangulaire 25x18 cm, maille 630 µm). Cette technique, adaptée aux cours d'eau lotique de faible profondeur, consiste à remuer le substrat afin d'entrainer les gammares dans le troubleau, placé à l'aval. Un premier tri est réalisé directement sur site en vidant le contenu du troubleau sur une colonne de tamis de maille, avec un dernier filtre de maille de 2 mm à 2,5 mm afin de sélectionner des gammares adultes de taille homogène (Figure II-2).

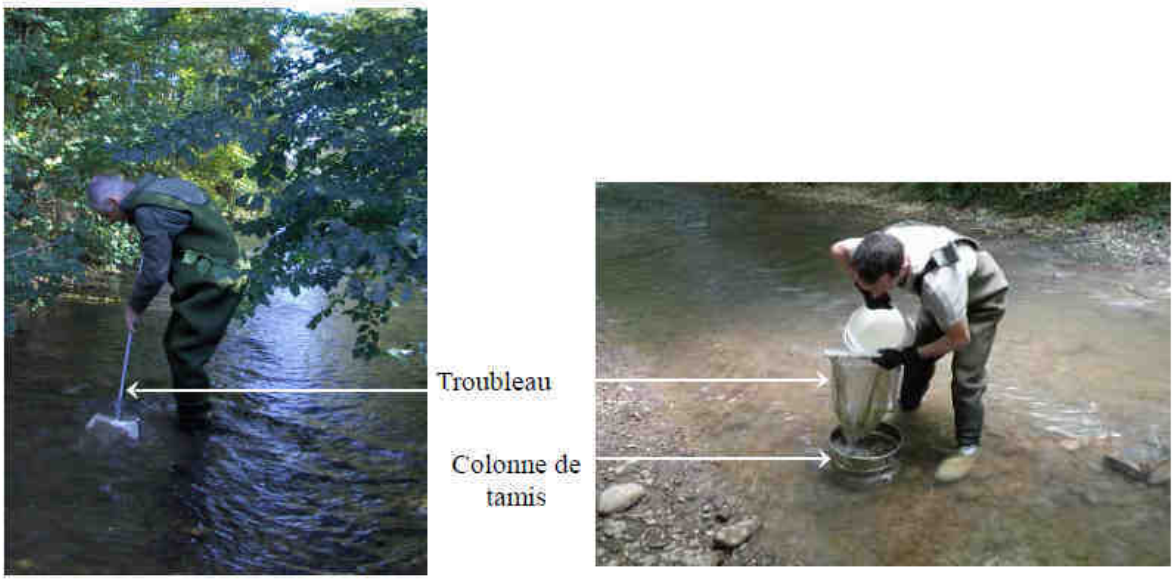


Figure II-2 : Prélèvement d'organismes épibenthiques au troubleau suivi par le tri des gammares sur une colonne de tamis (Lacaze 2011)

Suite au tri, les organismes sont distribués dans des sceaux remplis d'eau du site. Jusqu'à l'arrivée au laboratoire, les sceaux sont stockés dans une glacière afin de limiter les variations de température.

1.1.2. Maintien des organismes en laboratoire

De retour au laboratoire, un nouveau tri est effectué afin d'éliminer les débris organiques et les espèces non-cibles. Avant toute utilisation, les organismes sont acclimatés aux conditions de laboratoire (*i.e.* stabulation) pendant une période minimum de 15 jours. Les gammares sont répartis dans des aquariums de 30 L placés dans un bain thermorégulé de 12 ± 1 °C. L'eau des aquariums est constituée d'un mélange d'eau de forage et d'eau osmosée à un pH de 7,4 ± 0,2 et une conductivité de 500 ± 20 $\mu S.cm^{-1}$ et est renouvelée selon un débit de 2,25 $L.h^{-1}$. Un système de bullage permet de maintenir un taux d'oxygène dissous à saturation. La photopériode est fixée à 16 h jour/8 h de nuit, sous une intensité lumineuse comprise en 500 et 1000 lux. Les organismes sont nourris *ad libitum* avec des feuilles d'aulne (*Alnus glutinosa*), récoltées en fin d'automne dans une région peu anthropisée dans la localité des Ardillats (Rhône).

1.1.3. Identification et acquisition des échantillons biologiques

Les études protéomiques ont été directement réalisées sur différents organes : le céphalon, les tissus reproducteurs mâles et femelles ainsi que les *cæca* hépatopancréatiques (Figure II-3). Pour la réalisation du catalogue de protéines, seuls des organismes en amplexus ont été utilisés alors que pour les autres expérimentations, les organismes ont été choisis à différents stades reproductifs.

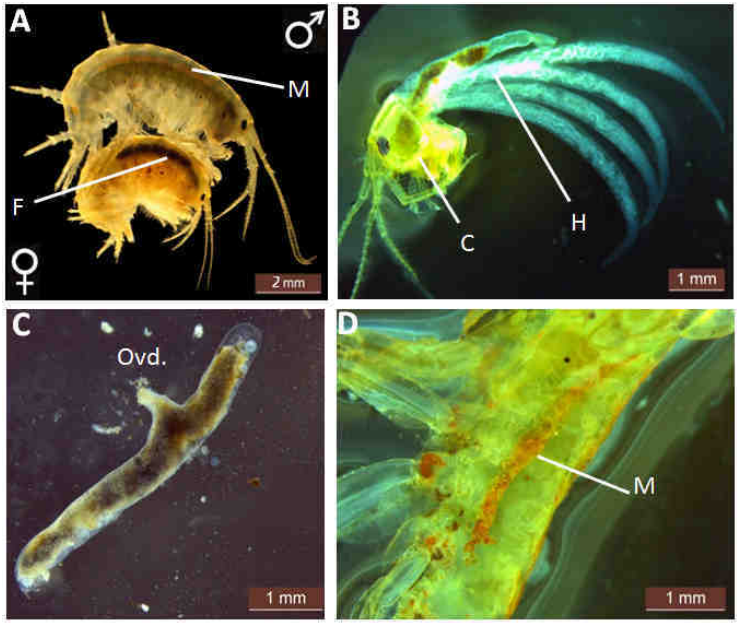


Figure II-3 : Matériel biologique employé pour la création d'un catalogue de protéines par protéogénomique

(A) Couple en précopulat ; (B) cephalon avec les *cæca* hépatopancréatiques attachés ; (C) un ovaire et (D) coupe transversale de gammare mâle avec mes testicules en orange.

Pour l'étape d'identification de protéines impliquées dans la reproduction, les embryons présents dans la poche marsupiale ont également été utilisés.

Les organismes sont placés sous loupe binoculaire pour récupérer les différents tissus. La partie terminale de l'urosoma est coupée permettant, à l'aide de fins forceps (forceps Dummont) de détacher facilement le céphalon avec les *cæca* hépatopancréatiques attachés. Avec de fins ciseaux (ciseaux Vannas), les *cæca* sont coupés du céphalon, et chaque type de tissu est placé dans un microtube contenant le tampon adéquat en fonction du traitement analytique réalisé par la suite. Afin de récupérer les gonades, les cuticules dorsales et ventrales sont découpées en veillant à ce que les lames du ciseau restent le plus près possible de la cuticule et ne s'enfoncent pas dans les

organes. Les gonades sont prélevées avec les forceps et placées dans un microtube. Dans le cas de l'analyse protéomique ou transcriptomique, les échantillons ont été cryogénisés par azote liquide et stockés à -80°C jusqu'à utilisation.

1.2. Expérimentations mises en œuvre pour répondre aux objectifs de la thèse

La stratégie expérimentale mise en œuvre dans cette thèse se déroule en trois grandes étapes : [1] la réalisation d'un catalogue de protéines sur le céphalon, les testicules, les ovaires et les *cæca* hépatopancréatiques sur organismes en amplexus [2] l'étude de la dynamique du protéome des testicules, des ovaires et des embryons respectivement à différents stades de spermatogénèse, d'ovogénèse et d'embryogénèse et [3] l'étude de la dynamique du protéome des testicules en réponse à différentes substances modèles reprotoxiques, en parallèle avec la mesure de marqueurs de qualité spermatique.

Deux types d'expérimentations ont été réalisés au cours de cette thèse : [1] des expositions en continu, en absence de contaminant, afin de créer le catalogue de protéines et d'étudier les cycles reproductifs et [2] des expositions en condition semi-statique avec des contaminants pour l'étude de l'infertilité masculine.

Pour la réalisation du catalogue de protéines ainsi que pour l'étude de l'ovogénèse et de l'embryogénèse, les organismes ont été directement prélevés dans les aquariums de stabulation (II-1.1.2).

Dans le cas de l'étude de la spermatogénèse, le seul critère disponible est l'observation de la date de copulation de l'individu. Il est donc nécessaire d'individualiser les couples et de relever quotidiennement les mâles que se sont reproduits, copulation mise en évidence par la séparation du mâle et de la femelle. A partir de la description faite par Lacaze (2011) (Figure I-14), différents stades ont été choisis : les gammares mâles « J-1 » correspondant à des individus matures en amplexus et proche de l'accouplement, pour lesquels la production de spermatozoïdes est complète, les gammares mâles « J0 » correspondant à des individus venant de se reproduire, les gammares mâles « J1 » ; « J2 » ; « J3 » et « J4 » post-copulation et enfin les gammares mâles « J7 »,

correspondant à des individus qui ont finalisé la production de spermatozoïdes, semblables aux mâles « J-1 ». Les gammares mâles, initialement choisis en amplexus avec une femelle D2, sont placés et maintenus dans des béchers en polyéthylène (un couple par bécher) et l'occurrence de la copulation inspectée quotidiennement.

Les béchers sont alimentés de la même eau que celle alimentant les aquariums de stabulation, par le biais d'une pompe péristatique, puisant dans une réserve d'eau renouvelée quotidiennement et oxygénée en permanence. Les béchers sont placés dans un bain thermorégulé et ces derniers sont équipés d'une surverse obstruée par de la toile à bluter (maille de 300 µm) afin d'empêcher la fuite des organismes (Figure II-4). De même, une pièce de toile à bluter (maille 500 µm, longueur x largeur : 6 x 5 cm) est disposée dans chaque bécher pour fournir une surface d'accroche aux organismes. Les conditions physico-chimiques (pH, conductivité, température) sont relevées quotidiennement. Le détail des deux expérimentations est présenté en Tableau II-1.

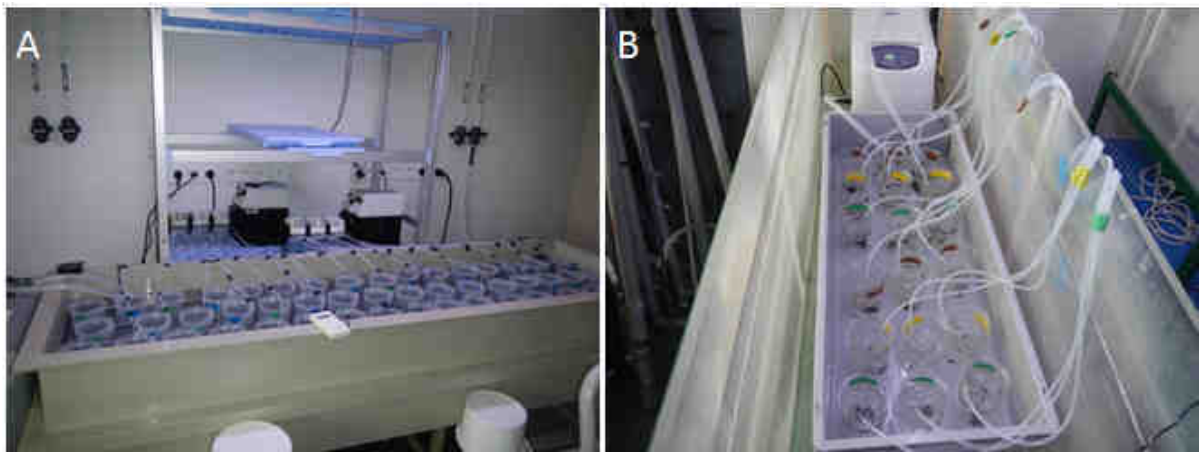


Figure II-4 : Dispositif expérimental en laboratoire

(A) : condition continue pour l'étude de la spermatogénèse et (B) condition semi-statique pour l'étude de l'infertilité

Dans le cas de l'étude de l'infertilité masculine, le dispositif expérimental est similaire mais l'exposition est faite en en condition semi-statique avec un renouvellement du milieu toutes les 48 h, avec un nouvel apport de feuilles d'aulne. Les gammares exposés au cadmium sont placés dans des béchers de 500 ml en verre afin d'éviter l'adsorption du contaminant sur les parois.

Tableau II-1 : Détails des expérimentations réalisées en laboratoire pour l'étude de la spermatogénèse et de l'infertilité masculine

Cd : cadmium, Pyr : pyriproxyfen et Met : methoxyfenozide

	Etude de la spermatogénèse	Etude de l'infertilité masculine
Durée (jours)	7	15
Température (°C)	12±1 °C	12±1 °C
Conductivité (µS.cm-1)	500±20 µS.cm^{-1}	500±20 µS.cm^{-1}
pH	7,4±0,2	8,3±0,2
Rythme jour/nuit	16 h/ 8 h	16 h/ 8 h
Renouvellement du milieu	Continu (500 ml/6 h)	Semi-statique (500 ml/48 h)
Aération	Constante	Constante
Nourrissage	*Ad libitum* (feuille d'aulne)	*Ad libitum* (feuille d'aulne)
Contaminants et concentrations	Aucun	Témoin eau et solvant (acétone 0,05 %) Cd (0.3, 1 et 3 µg.L^{-1}) Pyr (0.5, 5 et 50 µg.L^{-1}) Met (0.001, 0.1 et 10 µg.L^{-1})
Nombre de réplicat / condition	5	Protéomique : 5 Marqueurs de qualité spermatique : 7-10
Paramètres mesurés (Temps et fréquence)	Survie (quotidien) Protéome (-1, 0, 1, 2, 3, 4 et 7 j)	Survie (quotidien) Comportement reproducteur (5–8 jours) Marqueurs de qualité spermatique (15 j) Protéome (15 j)

Les organismes mâles sont exposés au cours de deux cycles de spermatogénèse successifs pendant 15 jours à plusieurs substances modèles : cadmium (Cd), pyriproxyfen (Pyr) et methoxyfenozide (Met) pour un total de 11 conditions différentes (Figure II-6). Le Met et le Pyr sont respectivement des insecticides analogues aux hormones ecdystéroïdiennes et juvénoïdes, et sont potentiellement des perturbateurs endocriniens chez les arthropodes (Carlson *et al.* 2001, Serafini 2001) (Figure II-5).

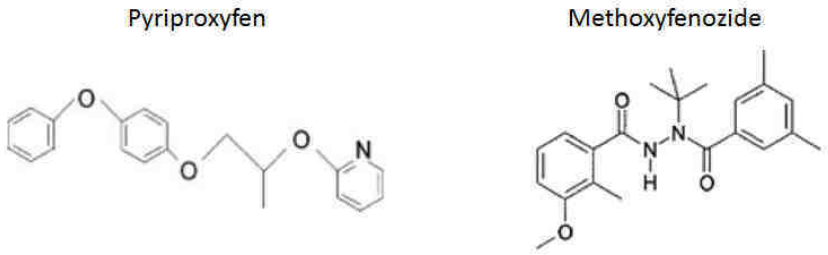


Figure II-5 : Substances insecticides testées pour l'étude de l'infertilité masculine

Pour chaque condition testée, les organismes sont répartis dans trois béchers, contenant chacun 7 gammares. Comme pour l'étude de la spermatogénèse, les gammares mâles sont choisis en amplexus avec une femelle en D2, et isolés. L'occurrence de la copulation est vérifiée quotidiennement, marquant le début de l'expérimentation en condition toxique (J0). Après 5 jours d'exposition, des femelles en stade D2 (*n* = nombre de gammares mâles par bécher) sont rajoutées et le comportement reproducteur (formation de l'amplexus, copulation) vérifié quotidiennement. Les femelles fécondées sont quotidiennement enlevées. Une fois que 70 % des couples se sont reproduits au sein d'un même bêcher (jour 8), ceux encore en amplexus sont éliminés et l'expérimentation arrêtée sept jours plus tard (Jour 15), permettant de compléter un second cycle de spermatogénèse.

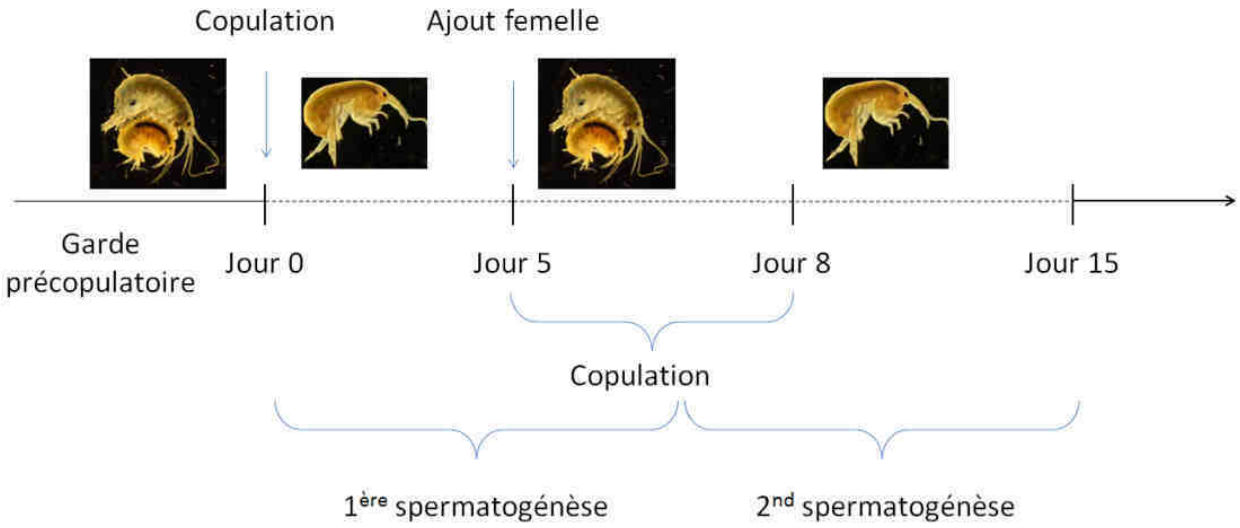


Figure II-6 : Déroulement expérimental pour l'étude de l'infertilité masculine

1.3. Détermination des stades de mue et du développement embryonnaire

Pour l'étude de la dynamique du protéome au cours de l'ovogénèse et de l'embryogénèse, il est nécessaire d'identifier au préalable les différents stades de développement des organismes. Le cycle reproductif de la femelle comprenant le cycle de mue, l'ovogénèse ainsi que l'embryogénèse a été décrit par Geffard *et al.* (2010) (Figure I-11). L'ovogénèse est synchronisée avec le cycle de mue, c'est pourquoi c'est le stade de mue qui est déterminé, ce dernier étant plus facilement observable sur la base de critères morphologiques.

Les organismes sont prélevés dans les aquariums de stabulation (II-1.1.2), disposés dans un cristallisoir et triés sur table lumineuse. Le stade de mue des femelles est déterminé par l'observation microscopique des tissus de l'extrémité des périopodes des troisième et quatrième paires. Les appendices sont sectionnés avec de fins ciseaux et montées entre lames et lamelles dans du milieu de stabulation (Figure II-7). L'observation microscopique (microscope DM 2500® (Leica), grossissement x400 et x630) du décollement de l'épiderme et de l'apparition du nouvel exosquelette permet de déterminer le stade de mue (Figure I-12). L'observation du stade de développement embryonnaire se fait selon des critères macroscopiques présentés en Figure I-11.

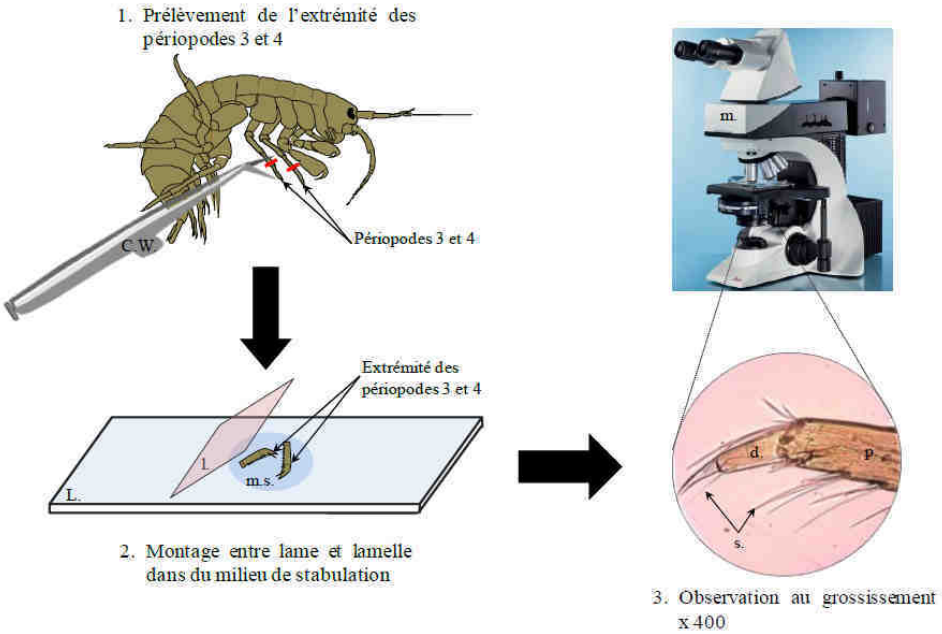


Figure II-7 : Représentation schématisée de la procédure utilisée pour déterminer le stade d'intermue des organismes (Xuereb 2009)

c.w. ciseaux de Wecker ; d : dactilopode ; L : lame ; l : lamelle ; m : microscope ; m.s. : milieu de stabulation ; p : protopodite ; s : soie.

1.4. Dénombrement spermatique

L'étude de l'infertilité masculine couple le séquençage du protéome des gonades et le suivi de l'intégrité de la fonction reproductive par la mesure de marqueurs de qualité spermatique. Ces marqueurs se basent sur la numération des spermatides et des spermatozoïdes, adapté selon le protocole de Lacaze (2011). Après dissection, les testicules d'un organisme sont dilacérés à l'aide d'un scalpel et déposés au sein d'un microtube contenant 30 µL de tampon phosphate salin (10 mM, pH 7,4 + 0,3 % de sérum d'albumine bovine), plongé dans la glace. La suspension cellulaire est homogénéisée par 20 aspiration-refoulements avec une micropipette et 20 µL déposés sur un hémocytomètre. La numération cellulaire est réalisée par microscopie optique (Leica DM 2500, magnification x100) en comptant 13 rectangles quadrillés. Les deux types cellulaires sont reconnus à partir de critères morphologiques : cellule pseudo -flagellée ovoïde pour les spermatides et oblongue pour les spermatozoïdes (Figure I-13).

Le nombre total estimé de gamètes est calculé en multipliant le nombre de gamètes par µL par le volume de la suspension cellulaire. Les analyses statistiques sont réalisées par le logiciel de traitement statistique R (R Development Core Team 2008). Afin de définir les conditions d'application des tests statistiques, l'homogénéité de variance des données est testée par un test de Bartlett alors que l'hypothèse de normalité est éprouvée par un test de Shapiro sur les résidus de l'ANOVA. Selon les conditions d'application des tests statistiques, l'existence de différences significatives entre les différents groupes expérimentaux est testée par une ANOVA (test paramétrique) ou un test de Kruskal-Wallis (test non paramétrique). En cas de différence significative ($p < 0,05$), les tests post-hoc associés de Tukey HSD et de Kruskal-Wallis post-hoc sont réalisés, afin de pouvoir clairement identifier les groupes expérimentaux différents les uns des autres ($p < 0,05$) face aux paramètres étudiés.

2. Analyses moléculaires

2.1. Analyse transcriptomique

L'analyse transcriptomique se décline en cinq grandes étapes : [1] le conditionnement des échantillons biologiques, [2] l'extraction des ARN totaux pour chaque type de tissus, [3] la création de l'échantillon composite à séquencer, [4] la création de la libraire d'ADNc et [5] le séquençage de l'ADNc par la technique Illumina-Solexa. Les deux dernières étapes ont été réalisées par un prestataire de service, ACOBIOM (anciennement Skuldtech) (http://acobiom.com/) basé à Montpellier.

Les tissus utilisés pour le séquençage du transcriptome sont les mêmes que ceux employés pour la création du catalogue de protéines. Suite à des pré-tests pour déterminer la quantité d'ARN totaux apportée par chaque type de tissus, le nombre d'organismes utilisés est 11, 9, 57 et 173 pour le céphalon, les *cæca* hépatopancréatiques, les ovaires et les testicules. Une fois prélevés, les tissus sont regroupés par type et placés dans un microtube contenant du tampon stabilisant RNAlater (Qiagen), plongé dans de la glace. Après une nuit d'incubation à 4°C, le RNAlater est éliminé par pipetage et les tissus stockés à sec à -80°C.

Pour chaque type de tissus, les ARN totaux sont isolés en utilisant le kit RNeasy Mini Kit (Quiagen) selon les instructions du fabricant. Une fois extraits, la qualité de chaque échantillon tissulaire est vérifiée grâce au bioanalyseur Agilent 2100 RNA Nanochip (Agilent Technologies, Santa Clara, California) couplé au logiciel Agilent 2100 Expert Software, en utilisant le kit associé. Par la suite, la concentration en ARN est également définie en déposant 1 µL de l'échantillon sur le spectrophotomètre NanoDrop ND-1000 UV/VIS (Thermo Scientific, Wilmington, Delaware, USA). Un échantillon « équimassique », est constitué à partir des quatre types de tissus, puis confié au prestataire de service.

La technique NGS choisie est celle du séquençage Illumina sur la plateforme Illumina/Solexa Genome Analyzer. Cette technique est la plus employée en raison de son faible coût et de sa rapidité (van Dijk *et al.* 2014). La préparation de la librairie d'ADNc est adaptée à cette technique et est réalisée suivant les recommandations du fabricant. Dans un premier temps, les ARNm sont purifiés par le biais de leur queue poly-A en utilisant des fragments nucléotidiques T attachés à des billes magnétiques. L'ARNm purifié est alors fragmenté par dénaturation thermique à 94°C dans un

thermocycleur pendant 5 min en présence de tampon de fragmentation (Illumina, Hayward, CA) afin d'obtenir des fragments de taille moyenne de 250 bp. Puis par une quinzaine de cycle de PCR, la librairie d'ADNc est créée, puis purifiée sur colonne (QIAquick PCR ; Quiagen). La qualité des fragments est à nouveau contrôlée grâce au bioanalyseur Agilent 2100 RNA Nanochip. Des adaptateurs spécifiques sont ajoutés au cours de l'amplification par PCR, à l'extrémité des fragments d'ADNc. Ces derniers permettent la fixation des fragments d'ADNc simple brins sur une surface solide, où une nouvelle phase d'amplification peut être réalisée. Cette amplification est réalisée au niveau local, à proximité du brin initial et permet la création de « cluster » unique sur la surface solide. A l'issue de cette étape, le séquençage dont le principe est présenté en Figure II-8, est initié.

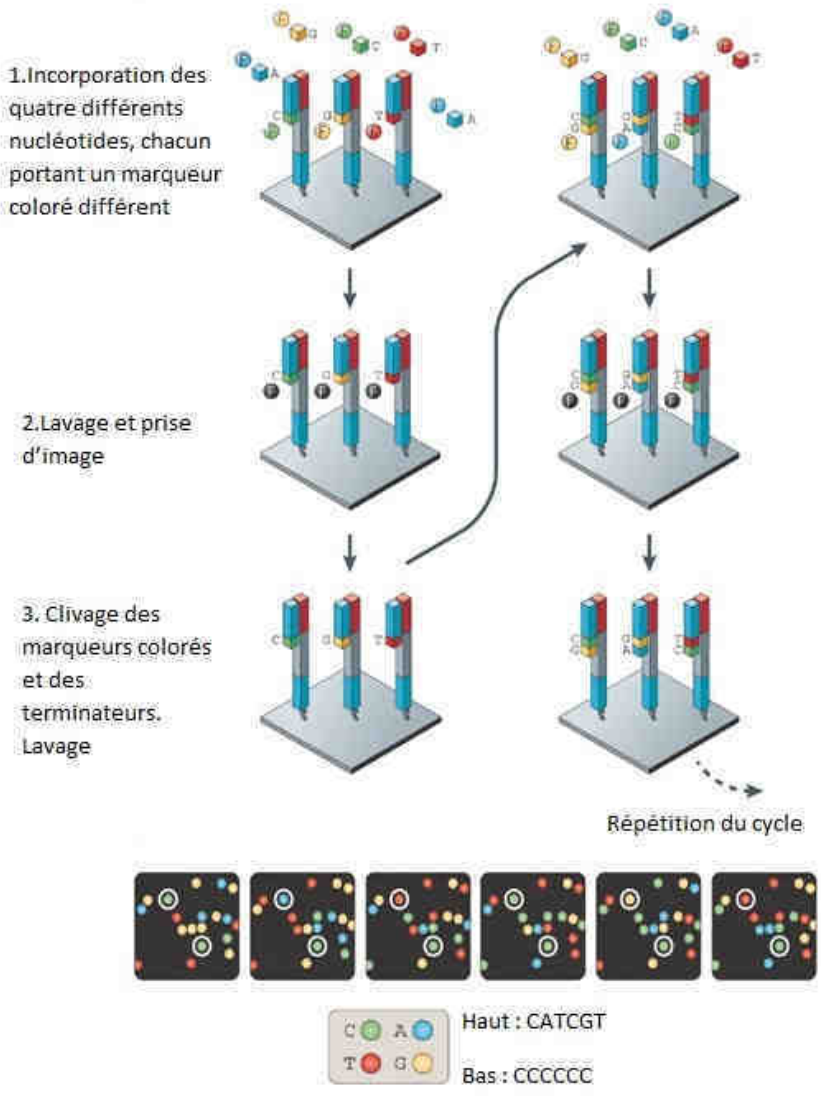


Figure II-8 : Principe du séquençage Illumina-Solexa (Metzker 2010)

2.2. Analyse protéomique

La première expérimentation de création de catalogue se base sur une couverture importante du protéome en employant plusieurs techniques de séparation des protéines et une importante quantité de tissus, nécessitant plusieurs organismes pour un même échantillon biologique. En revanche, pour l'étude de la dynamique du protéome au cours du cycle reproductif ou en réponse à des substances reprotoxiques, le protocole employé est plus simple à mettre en œuvre et la mesure est individuelle. Les différentes stratégies sont présentées en Figure II-9.

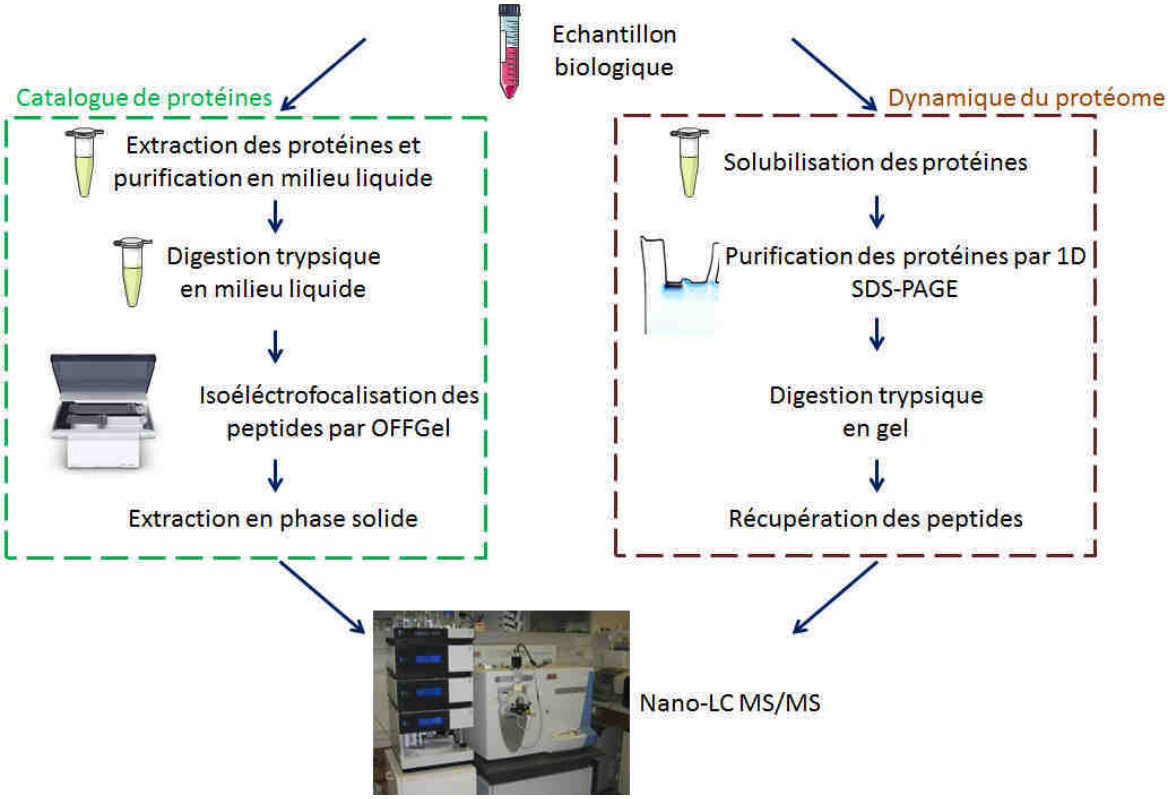


Figure II-9 : Stratégies expérimentales mises en œuvre pour l'étude du protéome

2.2.1. Extraction des protéines

- Catalogue de protéines

Les extractions de protéines ont été faites à partir d'ovocytes, de testicules, de céphalon et de *cæca* hépatopancréatiques, provenant respectivement de 44, 64, 26 et 30 organismes. Pour les gonades, 3 réplicas biologiques ont été réalisés, contre deux réplicas biologiques pour les autres tissus. Après récupération, les tissus sont déposés dans un microtube contenant 20 µl de tampon

de lyse (50 mM Tris, pH 7.8 + 6 M urée, 2 M thio-urée, 0.1 % (v/v) Triton X-100, 4 % CHAPS, 65 mM DTT et 2 % d'inhibiteur de protéase, Mini, EDTA-free, Roche) et conservés à – 80°C jusqu'à utilisation.

Pour l'extraction des protéines, réalisée avec les échantillons plongés dans de la glace fondante, les échantillons sont homogénéisés grâce à la sonde TissueRuptor (Quiagen) dans 2 ml de tampon de lyse. Après 15 min de repos, une sonde à ultrasons Helscher UP50H est plongée pendant 3 min (cycle 0,3 et amplitude 30 %) dans les homogénats. Puis, ces derniers sont centrifugés à 12 000 x g pendant 15 min à 4°C. Pour un volume de surnageant, trois volumes de mélange éthanol/diéthyléther (1 :1, v/v) sont ajoutés. Après 10 min de repos, les échantillons sont centrifugés à 16 000 x g pendant 15 min à 4°C. Le surnageant est éliminé partiellement puis 200 µL d'acétone froid sont ajoutés. Les échantillons sont centrifugés à 16 000 x g pendant 5 min à 4°C et l'opération précédente est répétée 2 fois. Enfin, les culots sont placés dans un évaporateur (Savant SpeedVac® SPD121P) pendant 5 min. Les culots sont conservés jusqu'à -20°C jusqu'à leur utilisation.

- Dynamique du protéome

Les extractions de protéines ont été faites à partir d'ovocytes, de testicules et d'embryons. Pour chaque expérimentation, 5 réplicas biologiques ont été réalisés par condition, à partir de gonades provenant d'un seul individu et pour les embryons, d'un pool de six embryons. Après récupération, les tissus sont déposés dans un microtube contenant 20 µL de Tampon Tris 50 mM et conservés à – 80°C jusqu'à utilisation. Les échantillons sont décongelés sur lit de glace fondante. Pour les gonades, les protéines sont directement solubilisées par l'ajout de 40 µL de tampon LDS 1 X. Pour les embryons, en raison de la présence de cuticule, les échantillons sont broyés avec la sonde TissueRuptor (Quiagen) dans 180 µL de tampon LDS 1 x. Après avoir été mélangés par vortex, les échantillons sont placés dans un bain à ultrasons transonic 780H pendant 1 min et laissés au repos dans la glace pendant 15 min. Enfin, les échantillons sont chauffés pendant 5 min à 95°C sous agitation constante à 800 rpm et les protéines directement purifiées par électrophorèse SDS-PAGE comme décrit en II-2.2.3.

2.2.2. Digestion enzymatique des protéines

La digestion enzymatique permet l'hydrolyse des protéines en fragments peptidiques. Elle est précédée d'une phase de réduction des ponts disulfures par à l'action du DTT puis de l'alkylation des cystéines par l'iodoacétamide (IAA) qui permet l'ajout de groupements carboxyamidométhyles sur les fonctions thiols réduites des cystéines. Cela empêche la reformation des ponts disulfures qui pourraient gêner l'interprétation des résultats de spectrométrie. Lors de l'étape de digestion enzymatique, la trypsine est l'enzyme la plus utilisée dans les études protéomiques. Elle coupe spécifiquement après les résidus d'Arginine (R) et de Lysine (K) (résumé par Cañas *et al.* 2007).

Les principales étapes analytiques pour la réalisation du catalogue de protéines (digestion en solution) et pour l'étude de la dynamique du protéome (digestion en gel) sont présentées en Figure II-10.

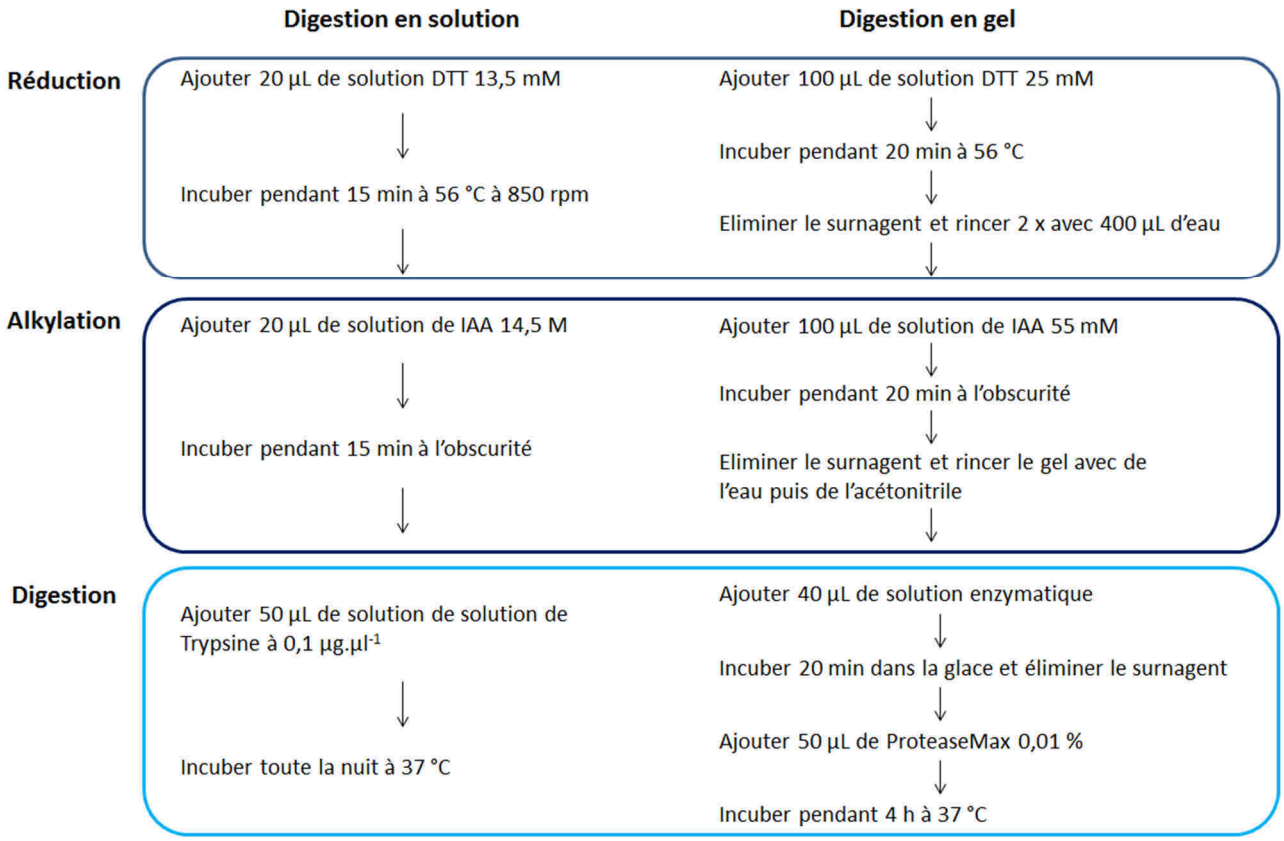


Figure II-10 : Principales étapes de la digestion enzymatique en solution et en gel

- Digestion en milieu liquide

Afin d'obtenir la meilleure couverture du protéome, la quantité de protéines digérée pour chaque échantillon est de 250 µg (Ernoult *et al.* 2008). La concentration en protéines des échantillons est déterminée en microplaque par le kit CooAssay Standard Protein Assay Kit (Interchim) en suivant les instructions du fabricant. Les échantillons sont dilués dans du tampon de digestion enzymatique (1 M d'urée, 0,001 % de Protéase Max, dissous dans 50 mM de tampon bicarbonate) afin d'obtenir une concentration finale de 2 µg de protéines.L^{-1}. Les différentes solutions de DTT et IAA sont également préparées en utilisant le tampon de digestion enzymatique alors que la solution de trypsine (Roche) est réalisée en ajoutant de l'acide trifluoroacétique (TFA) 0,01 %. La digestion est arrêtée en plaçant les échantillons à -20°C.

- Digestion en gel

La digestion en gel est réalisée suite à l'analyse des protéines par électrophorèse 1D SDS-PAGE (II-2.2.3). Avant la digestion, une étape de décoloration du gel doit être réalisée, avec l'ajout de 200 µL de mélange méthanol : ammonium bicarbonate 50mM (1:1 v/v). Après avoir laissé incuber pendant 1 min à 300 rpm, le surnageant est éliminé et le gel déshydraté avec 200 µL de acétonitrile : ammonium bicarbonate 50mM (1:1 v/v). Après une incubation pendant 5 min à 300 rpm, le surnageant est éliminé et le gel à de nouveau déshydraté avec 200 µL de acétonitrile. Après incubation pendant 1 min à 300 rpm, les échantillons sont placés en évaporateur (Savant SpeedVac® SPD121P) pendant 2 – 5 min. Les différentes solutions de DTT, IAA et de ProteaseMax (Promega) 0,01 % sont préparées en utilisant du tampon de bicarbonate d'ammonium 50 mM. Au cours de l'étape de rinçage du gel entre l'alkylation et la digestion, après avoir été rincé deux fois avec 400 µL d'eau, le gel est déshydraté de la même manière que pendant l'étape de décoloration. La solution de digestion enzymatique est composée de 10ng.$µl^{-1}$ de trypsine dans une solution de ProteaseMax 0,01%.

A l'issue de la digestion, 40 µL d'échantillon sont déposés dans un insert contenant 4 µL de TFA 5 %. Après mélange par aspiration-refoulement, les échantillons sont stockés à -20°C jusqu'à analyse MS.

2.2.3. Techniques séparatives du protéome

- Isoéléctrofocalisation par OFFGel

La séparation du protéome pour la création du catalogue de protéines a été réalisée sur les peptides digérés par la trypsine. Les peptides sont séparés en utilisant l'appareil Agilent 3100 OFFGEL Fractionator (Agilent, G3100 AA) avec un kit haute résolution selon une configuration 24 puits *i.e*. pH 3-10 (Agilent, 5188-6424) en suivant les instructions du fabricant. Les principales étapes sont présentées en Figure II-11

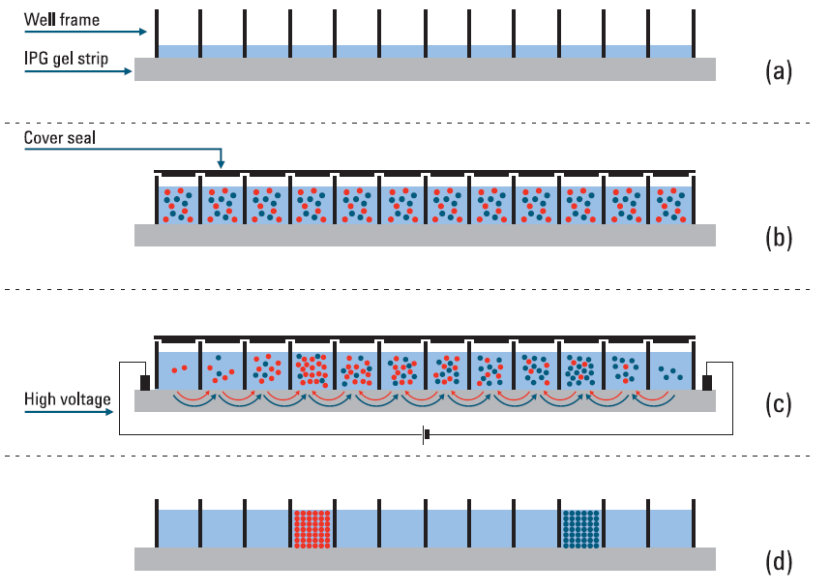


Figure II-11 : Principe de l'OFFGel (Agilent Technologies 2009)

(A) Réhydratation du gel avec le tampon de migration, (B) distribution à part égale de l'échantillon dans les puits et mise en place du couvercle, (C) application d'un haut voltage aux extrémités du gel et migration des biomolécules au travers du gel à une position où le pH est égal au pI de la molécule et (D) après séparation, les molécules restent dans la solution tampon et peuvent être facilement collectés par pipetage.

Les peptides sont soumis à un courant de 50 µA et à une puissance de 200 mW jusqu'à arriver à une tension de 50 kVh. Suite à la migration, les différentes fractions obtenues sont dessalées par extraction en phase solide en utilisant les colonnes Micro SpinColumns (Harvard, #74-4601) selon les instructions du fabricant. Les peptides sont élués de la phase solide avec une solution eau : acétonitrile (1 :1 ; v/v) contenant 0,1 % de TFA.

- Electrophorèse à une seule dimension SDS-PAGE

La réalisation de gel d'électrophorèse 1D SDS-PAGE pour l'étude de la dynamique du protéome a été employée en tant que technique de purification des protéines et non pour leur séparation suivant leur poids moléculaire. Suite à la dénaturation des protéines à 95°C, 35 µL d'échantillon sont déposés dans les puits du gel (gradient 4–12 % 10-puits NuPAGE (Invitrogen)) (Figure II-12). La migration est réalisée à 200 V, dans du tampon de migration MOPS 1 X (Invitrogen) et arrêtée quand tout l'échantillon a pénétré dans le gel (~5min). Le gel est rincé 4-5 X à l'eau milliQ puis coloré avec du colorant Coomassie Blue Safe (Invitrogen) pendant 30 min. Après quatre rinçages à l'eau milliQ, le gel a été stocké pendant une nuit à 4°C dans sa dernière eau de rinçage.

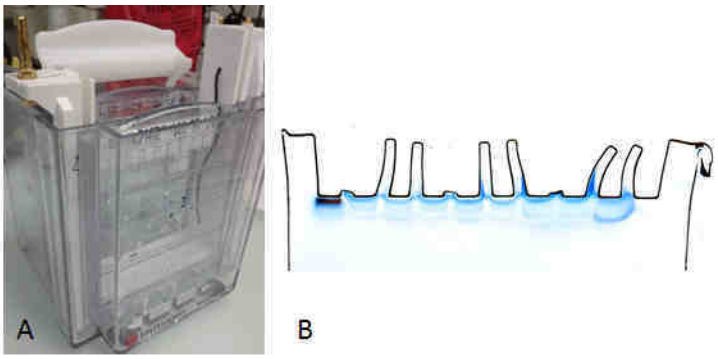


Figure II-12 : Electrophorèse 1D SDS-PAGE

(A) Dispositif d'électrophorèse SDS-PAGE et (B) Gel d'électrophorèse après coloration

Le lendemain, après numérisation du gel (Figure II-12), les bandes contenant les protéines sont récupérées en coupant au scalpel le gel placé sur une plaque de verre. Les bandes elles-mêmes sont découpées en petits cubes, placées dans une microplaque et conservées à -20°C jusqu'à digestion enzymatique.

- Chromatographie liquide

La colonne de chromatographie liquide LC UltiMate 3000 (Dionex-LC Packings) est placée en amont du spectromètre de masse. Les mélanges de peptides sont chargés (injection 2 µL) et dessalés en ligne sur une précolonne phase inverse Acclaim Pepmap 100 C18 (5µm taille de bille, 100Å taille de pore, 5mm x 300µm) de LC Packings. Ils sont ensuite passés sur une nanocolonne Acclaim Pepmap 100 C18 (3µm taille de bille, 100Å taille de pore, 15cm x 75µm) de LC Packings à

un débit de 0.3µL/min. En utilisant un solvant A (0,1 % d'acide méthanoïque) et un solvant B (0.1% d'acide méthanoïque /80% d'acétonitrile), les peptides sont séparés en utilisant un gradient de 96 minutes allant de 5 à 60 % de solvant B.

2.2.4. Séquençage du protéome par spectrométrie de masse

Le spectromètre de masse employé est un LTQ-Orbitrap XL de Thermo Scientific. Il est utilisé en mode « *data-dependant* » suivant la stratégie « TOP3 ». Sa gamme dynamique est comprise entre 300 et 1800 m/z. Après ionisation par nano spray, les ions sont piégés dans la trappe linéaire (LTQ), éjectés axialement et collectés dans la trappe en forme de C, pour être stockés jusqu'à injection dans le piège orbital. Une fois dans cet analyseur, les mouvements d'oscillation axiales autour de l'électrode centrale sont décomposés par transformée de Fournier pour obtenir le spectre MS. Les trois peptides les plus intenses sont alors sélectionnés, isolés, fragmentés et analysées dans la LTQ pour obtenir le profil MS/MS.

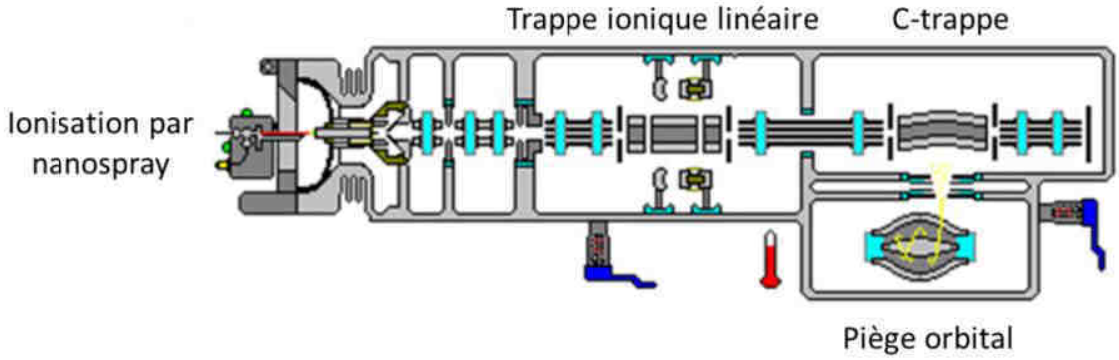


Figure II-13 : Représentation schématique du spectromètre de masse LTQ-Orbitrap

3. Traitement informatique des données de séquençage

3.1. Assemblage *de novo* du transcriptome

A l'issue du séquençage Illumina, les séquences obtenues à partir des fragments d'ADNc sont traitées par différents logiciels bioinformatiques suivant quatre étapes : [1] le nettoyage des brins adaptateurs ajoutés au moment de la création de la librairie d'ADNc (II-2.1), [2] l'assemblage du transcriptome, [3] l'estimation de la qualité du transcriptome, [4] la cartographie du transcriptome (Schliesky *et al.* 2012).

L'assemblage est réalisé en deux étapes, en utilisant deux types d'algorithmes le « De Bruijn graphs » avec Velvet et le « overlap-layout-consensus » avec MIRA. Le logiciel Velvet est appliqué dans un premier temps afin de simplifier et comprimer les données en fusionnant les séquences similaires (Miller *et al.* 2010). Puis, le logiciel d'assemblage MIRA reconstruit les séquences en recherchant les zones chevauchantes. Afin d'obtenir des séquences consensus de haute qualité, les paramètres de pourcentage de recouvrement et taille minimale de recouvrement sont respectivement fixés à 90 % et 30 bp. A l'issue de l'assemblage, deux types de séquences sont produites, les contigs qui résultent d'assemblage multiples et les singletons qui n'ont pas réussi à être intégrés dans l'assemblage.

La qualité du transcriptome est estimée suivant trois paramètres : la taille moyenne et médiane des contigs ainsi que le N50 qui rapporte la taille du contig alors que la moitié des bases qu'il contient ont été comptées. Toutefois, ces paramètres ont été développés pour les assemblages de génome et sont peu adaptés à l'estimation de la qualité d'assemblage d'un transcriptome (Schliesky *et al.* 2012). La cartographie du transcriptome a été réalisée par la recherche d'homologues comme présentée en II-3.3.

Enfin, dans l'optique de la création d'une BD de séquences protéiques pour l'interprétation des spectres MS/MS, les contigs sont traduits suivant les six cadres de lectures (Figure II-15), contenant tous les cadres ouverts de lecture (ORF).

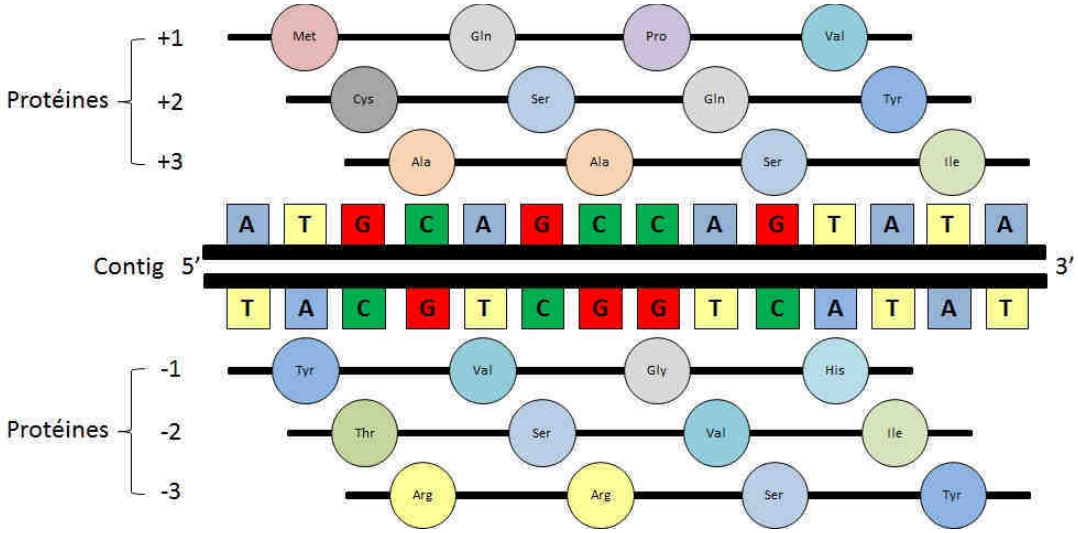


Figure II-14 : Traduction d'une séquence en acides nucléiques suivant les six cadres de lecture du code génétique

3.2. Interprétation des spectres MS/MS et identification des séquences peptidiques

En sortie du spectromètre de masse, un fichier RAW contenant une collection des spectres MS et MS/MS générés par échantillon est créé. Ces fichiers RAW sont alors convertis en fichier *Mascot Generic File* (MGF) en utilisant le logiciel extract_msn (ThermoScientific) selon le paramétrage suivant : 400 (minimum mass), 5.000 (maximum mass), 0 (grouping tolerance), 0 (intermediate scans), 10 (minimum peaks), 2 (extract MSn) et 1.000 (threshold) afin de pouvoir être traités par l'algorithme d'interprétation de spectres Mascot. Le paramétrage de MASCOT est réalisé par le biais de l'interface MASCOT Daemon (Figure II-15) avec les paramètres suivants : l'enzyme utilisée est la trypsine, le nombre de coupures défectueuses est fixé à 2, la tolérance sur la masse de l'ion parent est fixée à 5 ppm, la tolérance sur les ions fragments est fixée à 0.5 Da, la carbamidomethylation des cystéines comme modification fixe et l'oxydation des méthionines comme modification variable. Aussi en parallèle, suivant les mêmes paramètres de recherche, une recherche en mode « decoy » est réalisée pour estimer le taux de fausse découverte.

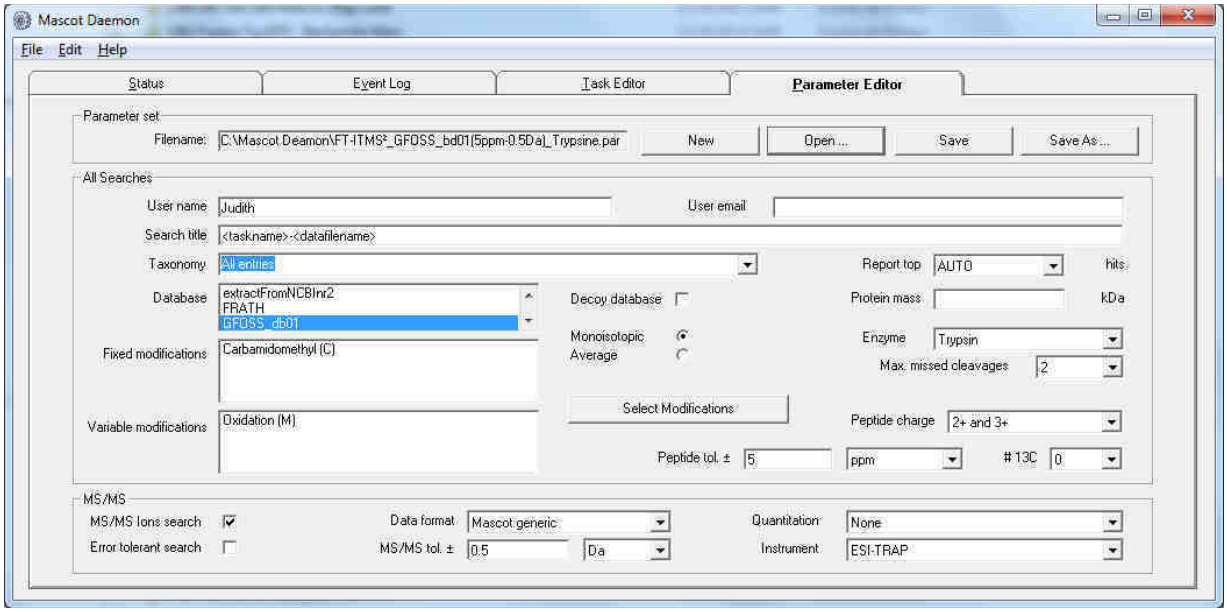


Figure II-15 : Interface de l'algorithme d'interprétation des spectres MS/MS MASCOT

Deux BD de séquences protéiques ont été interrogées pour l'interprétation des spectres MS/MS, la BD NCBInr et la BD GFOSS obtenue à partir du séquençage du transcriptome. A l'issue de ce traitement, des fichiers DAT sont générés et traités par l'outil IRMA 1.30.4 (Dupierris *et al.* 2009), permettant le filtrage des résultats d'identification MASCOT selon une *p*-value fixée à 0,05. Les résultats, contenant l'intégralité des peptides séquencés ainsi que les protéines identifiées sont exportés sous format Excel, facilitant l'exploitation des résultats. Le traitement des résultats sur Excel passe notamment par l'emploi d'une macro permettant de ranger par score les protéines et recalculer le nombre de spectral count lorsque le peptide est attribué à une protéine de score supérieure. Au final, une protéine est considérée comme étant validée quand deux peptides distincts lui sont attribués.

3.3. Interprétation des données

Une première approche pour identifier des protéines d'intérêt est de comparer leur abondance selon différentes conditions physiologiques ou en réponse à un stress toxique. Le logiciel PatternLab a été employé à cette fin, ce dernier ayant été développé pour traiter les données de protéomique shotgun et notamment les données de spectral count (Carvalho *et al.* 2008). Ce logiciel en libre-service (http://pcarvalho.com/patternlab/) comprend plusieurs modules. Le premier module employé est celui du TFold test qui permet de comparer deux conditions

distinctes, sur la base du test-t. Cette approche a été utilisée pour comparer l'abondance des protéines entre les tissus reproducteurs mâles et femelles afin d'identifier les protéines présentant un dimorphisme sexuel prononcé. Les données de SC obtenues à partir des tissus reproducteurs mâles et femelles au cours de la création du catalogue de protéines ont été utilisées à cette fin. La *p*-value est fixée à 0,05, la valeur de Fold change minimale à 2 et le taux de fausse découverte (Benjamini-Hochberg q-value) à 10 %. Pour l'étude de la dynamique du protéome, le module TrendQuest, basé sur l'algorithme développé par Carvalho *et al.* (2008) a été employé. Ce module permet de grouper en « cluster » les protéines qui présentent le même profil de variation temporelle. Différents critères de qualité doivent être renseignés tels le nombre minimum de réplicas dans lesquels la protéine doit être présente ou encore le nombre minimum de protéines par cluster.

Une seconde approche pour identifier des protéines d'intérêt se base sur l'étude de leur séquence. Dans un premier temps, sur l'ensemble des protéines identifiées au cours des différentes expérimentations, la recherche d'homologues par BLASTp sur NCBI est réalisée par le biais d'un programme développé au CEA de Marcoule. L'algorithme BlastX (version # 2.2.15, GenBank release number #166) est employé avec une E-value seuil fixée à 1^{E-10}. Cette analyse permet également de prédire les fonctions des protéines et par conséquent orienter l'interprétation du rôle biologique des protéines séquencées. Pour certaines protéines d'intérêt identifiées sur la base des précédentes approches, une analyse phylogénétique a été réalisée en intégrant d'autres protéines issues de la même famille. Les différents jeux de séquences sont traités en utilisant la plateforme SeaWiew version 4 (Gouy *et al.* 2010). Les alignements multiples sont réalisés par le programme ClustalW version 2 en utilisant les paramètres par défaut (Larkin *et al.* 2007).

CHAPITRE III : Découverte de protéines impliquées dans la reproduction de *Gammarus fossarum*

Le chapitre III regroupe les travaux relatifs à la création d'un catalogue de protéines par protéogénomique chez *Gammarus fossarum* et la recherche de protéines impliquées dans sa reproduction. Ces travaux sont présentés sous la forme de deux publications (**Publications 3** et **4**) et de notes (**Notes 1** et **2**).

La première partie III-1 traite de la création du catalogue de protéines sur trois tissus d'intérêt, le céphalon et les tissus reproducteurs mâles et femelles. Par la suite, afin d'identifier des protéines présentant un dimorphisme sexuel prononcé et donc potentiellement impliquées dans la reproduction, une comparaison de leur abondance entre les organes reproducteurs mâles et femelles a été réalisée par la technique label-free. Aussi, afin d'identifier des protéines ayant un rôle dans le processus reproductif chez le mâle, la dynamique du protéome de son tissu reproducteur a été suivie au cours de la spermatogénèse.

La partie III-2 se focalise sur la découverte de protéines jouant un rôle dans l'ovogénèse par le suivi de la dynamique du protéome des ovaires au cours de cinq stades différents de la vitellogénèse. De même, afin d'identifier de manière spécifique les protéines occupant la fonction de vitellogénine chez *Gammarus fossarum*, l'abondance des protéines au cours de l'embryogénèse a été étudiée afin de réaliser une identification croisée entre les protéines présentant un profil d'accumulation au cours de l'ovogénèse et de consommation au cours de l'embryogénèse.

Alors que la base de données ORF pour l'interprétation des spectres MS/MS a été construite à partir de quatre types de tissus, la partie III-3 s'intéresse à l'exploration du protéome des *cæca* hépatopancréatiques.

Enfin la dernière partie III-4 traite de la problématique de l'assemblage *de novo* du transcriptome et fournit une note technique sur la reconstruction des séquences en acides aminés par l'utilisation des données expérimentales obtenues en spectrométrie de masse en tandem.

1. Découverte de protéines impliquées dans la fonction reproductrice par protéogénomique

Publication 3:

Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods

Version auteur de l'article publié dans *Molecular and Cellular Proteomics*, 2014, 13 (12) : 3612-3625
La version finale est disponible sur http://www.mcponline.org/content/13/12/3612.abstract

Judith Trapp [a,b], Olivier Geffard [a] , Gilles Imbert [b], Jean-Charles Gaillard [b], Anne-Hélène Davin [b], Arnaud Chaumot [a] ,Jean Armengaud [b]

[a] Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, CS70077, F-69626 Villeurbanne, France
[b] CEA, DSV, IBEB, Lab. Biochim. System. Perturb., Bagnols-sur-Ceze, F-30207, France

Résumé :

En raison de leur importance écologique, les crustacés amphipodes sont des organismes tests populaires en évaluation du risque environnemental. Alors que la protéomique permet de révéler de manière inédite les mécanismes moléculaires en réponse à un stress, de telles investigations sont rares chez ces organismes d'intérêt, en raison de l'absence de bases de données en séquences protéiques. Dans cette étude, nous proposons chez *Gammarus fossarum*, espèce clé des hydrosystèmes dulcicoles européens, une approche protéogénomique destinée à identifier de manière spécifique les protéines de cet amphipode. Dans un premier temps, une base de données ORF a été créée par séquençage RNA-Seq. Par la suite, une analyse protéomique shotgun a été réalisée sur trois tissus impliqués dans la reproduction : les tissus reproducteurs mâles et femelles et le céphalon, afin d'identifier de nouvelles protéines d'intérêt par le biais de la base ORF. Un total de 1873 protéines, dont 218 orphelines, ont été certifiées par spectrométrie de masse, représentant la plus grande ressource protéomique actuellement existante pour un crustacé. Une analyse comparative de l'abondance des protéines au sein des tissus mâle et femelle a permis d'identifier celles présentant un fort dimorphisme sexuel. Enfin, au cours de la spermatogénèse, des profils d'expression ont été réalisés sur sept différents stades reproductifs permettant de mettre en évidence des protéines impliquées dans différents aspects de la reproduction masculine.

Mots clés : Amphipode, protéogénomique, RNA-Seq, protéomique shotgun, profils d'expression

Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods

Abstract

Because of their ecological importance, amphipod crustacea are employed worldwide as test species in environmental risk assessment. While proteomics allows new insights into the molecular mechanisms related to the stress response, such investigations are rare for these organisms due to the lack of comprehensive protein sequence databases. Here, we propose a proteogenomic approach for identifying specific proteins of the freshwater amphipod, *Gammarus fossarum*, a keystone species in European freshwater ecosystems. After deep RNA sequencing, we created a comprehensive ORF database. We identified and annotated the most relevant proteins detected, through a shotgun tandem mass spectrometry analysis carried out on the proteomes from three major tissues involved in the organism's reproductive function: the male and female reproductive systems, and the cephalon, where different neuroendocrine glands are present. The 1,873 mass-spectrometry-certified proteins represent the largest crustacean proteomic resource to date, 218 proteins being lineage specific. Comparative proteomics between the male and female reproductive systems indicated key proteins with strong sexual dimorphism. Protein expression profiles during spermatogenesis at seven different stages highlighted the major gammarid proteins involved in the different facets of reproduction.

Keywords: Amphipod, Expression profile, Proteogenomics, RNA-Seq, Shotgun proteomics

1. INTRODUCTION

Next-generation proteomics, relying on ultra-rapid and sub-ppm mass spectrometry analyzers, is able to offer in-depth insights into the molecular players sustaining the physiology of complex organisms. Identifying and quantitating thousands of proteins has become, over the last decade, a routine task for most proteomic platforms, with the development of high-throughput shotgun proteomics. The interpretation of large-scale MS/MS data is only possible if a high quality database of nucleic acid sequences is available. Homology-driven proteomics using cross-species matching is a first alternative if the genome of interest is unknown, while *de novo* sequencing, *i.e*. interpretation of MS/MS data to establish the exact sequence of each peptide from scratch, is another possibility. However, major drawbacks of these two approaches lead to a scarcity of results as soon as a non-model organism, distantly related to a sequenced organism, is analyzed. Indeed, only highly conserved and ubiquitous proteins will be identified and carefully annotated with such approaches. It is noteworthy that lineage-specific genes are more likely to be linked to the organism's unique biology, as demonstrated by the characterization of the genome of *Daphnia pulex*, the water flea, a freshwater microcrustacean whose orphan genes have been shown to be amongst the most ecoresponsive (1). Consequently, gene products from non-model organisms responding to environmental challenges are overlooked.

Among amphipods, the genus, *Gammarus*, represents the highest number of epigean freshwater species distributed throughout the Northern Hemisphere. In Europe, two closely related species, *Gammarus pulex* and *Gammarus fossarum,* are keystone species for freshwater ecosystems. They are commonly used as sentinel species in freshwater risk assessment, for several reasons (2). First, they are widespread and found throughout a large habitat range, where they often occur at high densities. Second, they occupy a large trophic repertoire: herbivores, predators, and detritivores playing a major role in the leaf-litter breakdown processes (3). They also constitute a food reserve for macroinvertebrates and fish. Finally, gammarids can be easily maintained in the laboratory or used for *in situ* bioassays (2), where the impact of pollutants can be assessed by measuring molecular markers related to diverse modes of action, such as neurotoxicity (4), as well as using life-history-trait reproductive features (5). Alterations of sexual phenotype (intersexuality) have been reported *in situ* (6) as well as alterations by xenobiotics of various physiological parameters related to reproductive success, *i.e.* gametogenesis, embryogenesis, fecundity or molt (5). However, the molecular mechanisms involved in these reproductive impairments are unknown. A major reason for this is that hormones and proteins involved in the regulation of reproductive function in *G. fossarum*, and in crustacea in general, have until recently been largely

misunderstood (7). While *Gammarus* is a relevant ecotoxicological animal model, comprehensive genomic resources are lacking for this genus, and also for crustacea within the tree of life in general. To date, only the genome of the zooplanktonic branchiopod, *Daphnia pulex*, a laboratory-cultured organism with clonal reproduction, has been released (1). This organism is a popular ecotoxicological model because of its ease of rearing, but the large phylogenetic distance between crustacea, and the ecology of other classes of crustacea, question the insightfulness of this model (8). Furthermore, in the case of assessing reprotoxicity, different life-history strategies *i.e*. parthenogenic *vs.* sexual reproduction, may have different consequences concerning vulnerability to pollutants – for example, different physiological manifestations.

A novel approach that intimately combines genomics and proteomics, namely proteogenomics, has emerged (reviewed in (9-11)) after the pioneering work of Yates et *al.* (12), which used combined EST databases translated into six reading frames for the interpretation of MS/MS spectra. Proteogenomics *sensu stricto* consists of better annotating genomes by means of high-throughput proteomic data, and this approach has been exemplified for several bacteria (13, 14) and eukaryotes (15, 16). Numerous protein-encoding genes can be identified using mass spectrometry data when missed by automatic annotation software programs, or their structures can be corrected by experimental validation of their N-terminal translational starts. Recently, proteogenomics-style approaches, where genome-sequencing data are used directly to interpret large proteomic datasets, and vice-versa, have flourished, albeit not directly aimed at genome annotation or reannotation (17-19). On this basis, a *sensu lato* definition of proteogenomics has been discussed to qualify projects based on multi-omics data, with closely linked nucleic acid and protein information, including, typically, the use of six-frame translation (20). In this vein, Ning *et al*. (21), Nagarai *et al.* (22), Wang *et al*. (23) and Woo *et al*. (24) noted the interest of sample-specific protein databases established from RNA-seq data, which can enable a straightforward analysis of MS/MS data. While comprehensive coverage of transcriptomes by RNA-Seq has been achieved for several marine species, studies on amphipods remain scarce, restricted to *Parhyale hawaiensis* (25, 26) and *Melita plumulosa* (27).

Here, we have documented gammarid reproductive function through the identification and characterization of novel proteins. For this, we used a proteogenomic approach aimed at quickly pinpointing the proteins from *G. fossarum,* a model of interest in terms of ecotoxicology. After deep RNA-Seq of total RNAs of specific organs from individuals sampled in a wild population of gammarids, we created a six-reading-frame translation protein database for interpreting a large shotgun tandem mass spectrometry dataset obtained in *G. fossarum* specimens. We have detailed

and discussed the proteome of key physiological organs: the female reproductive system, the male reproductive system, and the cephalon, where several neuroendocrine glands are located. To highlight protein candidates involved in reproductive physiology, we first performed comparative proteomics between the male and female reproductive systems to pinpoint key proteins with strong sexual dimorphism. Subsequently, protein expression profiles during spermatogenesis at seven different stages were analyzed to uncover new proteins that could be directly related to spermatogenesis in *G. fossarum*.

2. EXPERIMENTAL PROCEDURES

Sampling and maintenance of animals. The amphipods were sampled at La Tour du Pin (Lat: 45°569'442"; Long: 5°459'115"), upstream of the Bourbre River (mid-eastern France). This site has a good water quality according to data records from the Réseau National de Bassin (French Watershed Biomonitoring Network; http://sierm.eaurmc.fr/eaux-superficielles/index.php), and accommodates high densities of gammarids. Sexually mature organisms were collected by kick sampling using a net. They were sieved (> 2 mm), stored in plastic bottles containing ambient freshwater, and quickly transported to the laboratory. During an acclimatization period of at least 15 days, the organisms were kept in 30 l tanks continuously supplied with aerated, uncontaminated groundwater adjusted to the sampling site conductivity (*i.e.* 600 µS.cm^{-1}). A 16/8 h light/dark photoperiod was maintained and the temperature was kept at 12±1 °C. The gammarids were fed *ad libitum* with alder leaves (*Alus glutinosa*), supplemented with freeze-dried worms (*Tubifex tubifex*) twice a week.

Preparation of biological samples. Gammarids in amplexus were selected to cover a wide distribution range of body-size (from 1 to 2 cm) and dissected under stereomicroscope magnification, as described by Lacaze *et al*.(28). The third segment of the urosoma was cut with microscissors, the cephalon was removed from the body with fine forceps, and the attached *ceaca* were cut. Dorsal and ventral cuticles were then excised. The male gonads, including the seminal vesicle, were gently removed with fine forceps, as for oocytes. For RNA library preparation, each tissue was collected as one tissue-type sample using 11, 9, 57 and 173 organisms for cephalons, *ceaca*, oocytes and testes, respectively. For the proteomes, only adult organisms sampled using 2 and 2.5 mm sieves, and determined as *Gammarus fossarum* by phenotypic criteria, were selected. Each tissue was collected as one tissue-type sample using 44, 26 and 64 organisms for cephalon, oocyte and testis tissues, respectively. Proteomes were investigated in duplicate for cephalons and in triplicate for reproductive tissues. All pooled samples were immediately frozen in liquid

nitrogen, either in the presence of RNAlater RNA Stabilization Reagent (Qiagen) or cold protein lysis buffer (50 mM Tris, pH 7.8 + 6 M urea, 2 M thiourea, 0.1% (v/v) Triton X-100, 4% CHAPS, 65 mM DTT and 2% protease inhibitor complete, Mini, EDTA-free, Roche), and stored at –80 °C until required.

Construction of a paired-end RNA-Seq library and Illumina sequencing. Tissues were disrupted with the TissueRuptor instrument (Qiagen) and total RNA was isolated using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Total RNAs for each tissue sample were quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, Delaware, USA) and their qualities were verified using an Agilent Bioanalyzer 2100 RNA Nanochip (Agilent Technologies, Santa Clara, California). Similar amounts of total RNAs from each tissue sample were pooled in order to create an equimolar sample for library construction, representing equally the four tissues. The cDNA library for Illumina mRNA-Sequencing (Illumina) was constructed with 5 µg total RNA. The manufacturer's protocol was carried out with the following modifications. In brief, poly-A-containing mRNA molecules were purified using poly-T oligos that were attached to magnetic beads. The purified mRNA was fragmented by addition of 5× fragmentation buffer (Illumina, Hayward, CA) and heated to 94 °C in a thermocycler for 5 min. This fragmentation time is the default setting yielding nucleic acid fragments of ~250 bp. First-strand cDNA was synthesized using random primers to eliminate the general bias towards the 3' end of the transcripts. Second-strand cDNA synthesis was performed by adding GEX second-strand buffer (Illumina, Hayward, CA), dNTPs, RNaseH and DNA polymerase I, followed by incubation for 2.5 h at 16 °C. Second-strand cDNA was further subjected to end repair, A-tailing, and adapter ligation, as recommended. Purified cDNA templates were enriched by 15 cycles of PCR for 10 s at 98 °C, 30 s at 65 °C, and 30 s at 72 °C using PE1.0 and PE2.0 primers and Fastart Taq DNA polymerase (Roche). The samples were cleaned using QIAquick PCR purification columns and eluted in 30 µl Elution Buffer (QIAGEN, CA). Purified cDNA libraries were quantified using a 2100 Bioanalyzer DNA 100 Chip (Agilent). Sequencing was performed on the Illumina/Solexa Genome Analyzer according to the manufacturer's instructions.

Contig assembly and creation of a customized protein sequence database. EST sequence analysis and assembly were performed by the Skuldtech Company (http://www.skuldtech.com/index.php). ESTs were assembled into clusters using the Velvet and MIRA software programs. In order to obtain highly reliable consensus sequences, overlapping identity percentage and minimum overlapping length parameters were set to 90% and 30 bp, respectively. Contigs resulting from the

assembly of multiple sequences are referred to as unique sequences and sequences that do not fit into any assemblies are called singletons. Singletons are mainly chimeras or artefacts from cDNA synthesis, sequencing or assembly, resulting in an increase of protein sequence errors, inappropriate frameshifts, and premature ORF terminations. Thus, only the unique sequences were translated into six reading frames, resulting in a customized protein sequence database, named GFOSS, and containing 1,311,444 sequences totaling 289,084,257 amino acids.

Protein extraction and trypsin digestion. For protein extraction, tissues were disrupted with 2 ml cold protein lysis buffer with the TissueRuptor instrument (Qiagen). Homogenates were sonicated for 3 min using a Hielscher Compact Lab Homogenizer (amplitude 30% and cycle 0.3). Samples were centrifuged for 15 min at 12,000 g and 4 °C. Clear supernatants were collected and delipidated by adding 3 volumes of ethanol/diethylether solution (1:1, vol/vol) for one volume of sample, as previously reported by Simon *et al*. (29). The mixtures were vortexed and incubated on ice for 10 min. Samples were then centrifuged at 16,000 g for 15 min at 4 °C. The resulting supernatants were removed. The pellets were washed twice with 200 µl cold acetone and then resuspended in 300 µl trypsin buffer (50 mM ammonium bicarbonate, 1 M urea, and 0.01% ProteaseMAX surfactant (Promega)). During the extraction procedure, samples were kept on ice. The protein concentration of the samples was determined in microplates using the CooAssay Standard Protein Assay Kit (Interchim) with bovine serum albumin as standard. The protein samples were adjusted to 2 $\mu g.\mu l^{-1}$ by dilution with trypsin buffer. Samples were treated with DTT (1 mM final) for 15 min at 56 °C under constant agitation to reduce cysteines. These were alkylated with 3.75 mM iodoacetamide for 15 min at room temperature in the dark. Proteolysis was performed overnight with 2% sequencing grade trypsin (Roche) dissolved in 0.01% trifluoroacetic acid.

Isoelectric focusing of peptides. After digestion, peptides were resolved by their isoelectric point, according to manufacturer's instructions, using an Agilent 3100 OFFGEL Fractionator (Agilent, G3100 AA) with a High Resolution Kit, i.e. pH 3-10 (Agilent, 5188-6424). A total of 250 µg digested peptides was loaded into the 24-well fractionation device and focused at 50 µA current and 200 mW power until 50 kVh was reached. After focusing, salts were removed with Micro SpinColumns (Harvard, #74-4601) following the manufacturer's instructions. Peptides were eluted with 50 µl 50% aqueous acetonitrile solution containing 0.1% trifluoroacetic acid.

NanoLC-MS/MS analysis. NanoLC-MS/MS experiments were performed on an LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings). Peptide samples (2 µl) were loaded and desalted online on a reverse-phase precolumn, C18 PepMap 100 column (LC Packings), and then resolved on a nanoscale C18 PepMap^{TM} 100 capillary column (LC Packings) at a flow rate of 0.3 µl/min with a gradient of CH_3CN, 0.1% formic acid prior to injection into the ion trap mass spectrometer. Peptides were separated using a 90 min gradient from 5 to 60% solvent B (0.1% HCOOH, 80% CH_3CN). Solvent A was 0.1% HCOOH, 100% H_2O. Full-scan mass spectra were measured from *m/z* 300 to 1800 with the LTQ-Orbitrap XL mass spectrometer in data-dependent mode using the TOP3 strategy. In brief, a scan cycle was initiated with a full scan of high mass accuracy in the Orbitrap followed by MS/MS scans in the linear ion trap on the three most abundant ions.

Proteomic profiles of male reproductive tissues during spermatogenesis. Expression profiles were performed on *G. fossarum* males at seven different spermatogenesis stages, which have been described previously (28): before copulation (male in amplexus) and after copulation (Day 0, 1, 2, 3, 4 and 7). Precopulates were individually selected and isolated in 500 ml polyethylene beakers under the same experimental conditions as mentioned above. All beakers were checked daily to determine the date of reproduction, attested by the female ecdysis (5). For each stage, five biological replicates were performed. Male reproductive tissues were sampled individually, immediately frozen in liquid nitrogen and stored at –80 °C until required. Proteins were directly dissolved in 40 µl LDS sample buffer (Invitrogen). Samples were sonicated for 1 min in a transonic 780H sonicator and boiled for 5 min at 95 °C. Proteins were resolved by SDS-PAGE with a short migration of 10 min at 150 V on 4–12% gradient, 10-well NuPAGE (Invitrogen) gels run with MES buffer (Invitrogen). Gels were stained with Coomassie Blue Safe stain (Invitrogen). After overnight destaining with water, the whole protein content from each well was extracted as a sole polyacrylamide band and processed for further destaining and iodoacetamide treatment. The samples were then proteolyzed with sequencing grade trypsin (Roche) using 0.01% ProteaseMAX surfactant (Promega). The resulting peptide mixtures were diluted 1:20 in 0.1% trifluoroacetic acid and analyzed by nanoLC-MS/MS.

MASCOT database mining. Peak lists were generated with the Mascot Daemon software (version 2.3.2; Matrix Science) using the extract_msn.exe data import filter (Thermo). Data import filter options were set to 400 (minimum mass), 5000 (maximum mass), 0 (grouping tolerance), 0 (intermediate scans), and 1000 (threshold), as described previously by Christie-Oleza *et al.* (30).

MS/MS spectra were assigned to peptide sequences with the Mascot Daemon 2.3.2 search engine (Matrix Science) against the NCBInr protein sequence database or the GFOSS database. Searches for tryptic peptides were performed with the following parameters: full-trypsin specificity, a mass tolerance of 5 ppm on the parent ion and 0.5 Da on the MS/MS, static modifications of carboxyamidomethylated Cys (+57.0215), and dynamic modifications of oxidized Met (+15.9949). The maximum number of missed cleavages was set at 2. All peptide matches with a MASCOT peptide score below a *p* value of 0.05 were filtered by the IRMa 1.30.4 software. A protein was considered to be validated when at least two different peptides were detected in the same experiment. The false-positive rate for protein identification was estimated, by search with a reverse decoy database, to be below 0.1% using the same parameters in both the NCBInr protein sequence database and the GFOSS database.

Statistical data treatment for comparative proteomics. Spectral counts (number of spectra recorded per protein) were extracted from the spectra-to-peptide dataset and the normalized spectral abundance factor (NSAF) for each protein was calculated as described previously (31). For comparative proteomics between male and female reproductive tissues, the data were analyzed with the Tfold module of the PatternLab software program. No specific normalization of spectral counts was applied, as strictly equal quantities of protein samples were analyzed. Fold change and p-value cutoffs were set at 2 and 0.05, respectively. The false discovery rate (Benjamini-Hochberg q-value) was 10%. The F-Stringency factor was systematically optimized. For analyzing the proteome dynamics along the different spermatogenesis stages, the data were clustered with the TrendQuest module of PatternLab. We retained in the final list of proteins for temporal profiling the proteins validated with at least two distinct peptides when merging all the data. Again, no specific normalization of spectral counts was applied. Proteins observed in at least 3 of the replicate samples were considered for clustering; the minimum signal required was 5. Cluster health was set to 0.9 and the minimum items per cluster to 3.

Sequence analysis. BLASTp search was performed with the NCBI website facilities (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the nonredundant protein sequences. The BlastX algorithm (version # 2.2.15, GenBank release number #166) was used with a threshold E-value set at $1E^{-10}$.

Data repository. This whole transcriptome project has been deposited in the EMBL Nucleotide Sequence Database under the accession number, PRJEB5098 (http://www.ebi.ac.uk/ena/data/view/PRJEB5098). The mass spectrometry proteomics data have

been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org/) *via* the PRIDE partner repository (32) with the dataset identifier, PXD000576.

3. RESULTS

General proteogenomic strategy for discovering key proteins involved in reproduction from a nonmodel complex organism.

To represent as accurately as possible the real ecological of gammarids, we sampled heterogeneously sized organisms from a field population located in eastern central France. Because our focus was on proteins involved in reproduction, only sexually mature organisms were chosen, by selecting those in precopulatory mate guarding (**Figure 1, Panel 1**). We first established a database of nucleic-acid-derived protein coding sequences that could be a generic platform for ecotoxicology-related proteomic studies on gammarids, as shown in **Figure 2**.

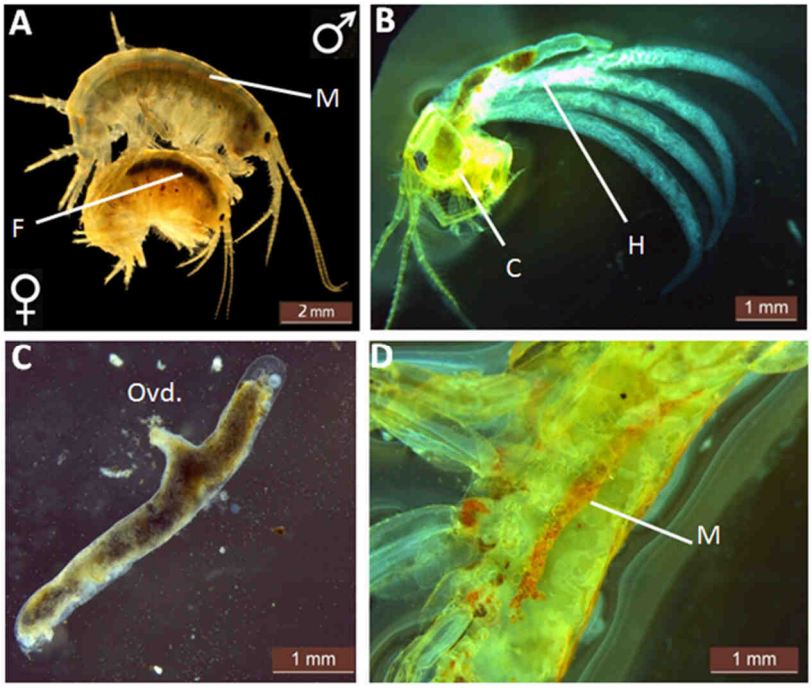


Figure 1: Images of *G. fossarum* in amplexus and sampled tissues.

(A) Sexually mature organisms in precopulatory mate guarding. (B) The cephalon and the four attached hepatopancreatic caeca. (C) A female oocyte. (D) Transversal view of a male gammarid with attached reproductive tissue. M: male reproductive tissue (orange); F: female reproductive tissue (black); C: cephalon; H: hepatopancreatic caeca; ovd.: oviduct

For this, we extracted total RNAs from four key tissues: the male and female reproductive tissues (**Figure 1, Panels 3 & 4**), the cephalon where several neuroendocrine glands are located (**Figure 1, Panel 2**), and the hepatopancreatic caeca (**Figure 1, Panel 2**), involved in xenobiotic detoxification and energy acquisition (digestive enzyme secretion, food absorption and nutrient storage).

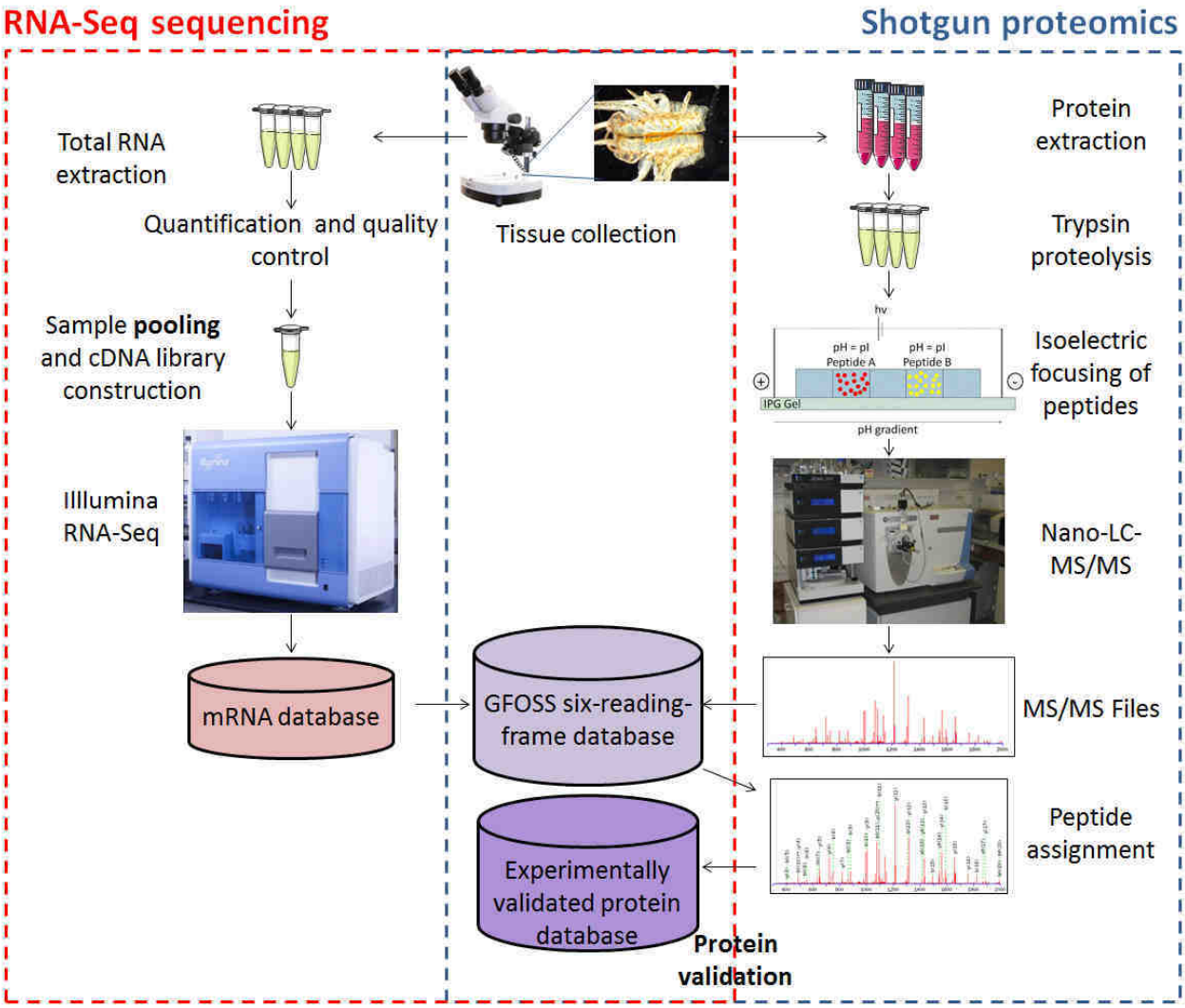


Figure 2: General proteogenomic strategy for discovering key proteins from *G. fossarum*

Equal amounts of RNA from each tissue were pooled and sequenced in a single Illumina sequencing run (**Figure 2**). The resulting nucleic acid information was used to create a protein database named GFOSS, which could be used for proteogenomic identification of proteins. For this, each mRNA contig was systematically translated into the six possible reading frames, considering all possible putative ORFs. As in previous proteogenomic studies, tandem mass spectrometry data could be assigned with such a database (23). The protein catalogue established from experimental evidence allows the selection of the appropriate reading direction and frame of

each protein-coding mRNA contig. As indicated in **Figure 2**, we recorded a large dataset on the proteomes from three tissues by resolving each sample by OFFGEL electrophoresis and performing 24 nanoLC-MS/MS runs per sample. In order to identify proteins related to reproductive function, we first performed automatic functional prediction based on cross-species homology. To discover new proteins involved in this function, we then compared the male and female reproductive tissues with a label-free comparative approach. We also focused our attention on the proteome dynamics of the male reproductive tissue to indicate new proteins related to spermatogenesis in *G. fossarum.*

Transcriptome characterization by RNA-Seq.

A normalized cDNA library was constructed from the pool of mRNAs extracted from the male and female reproductive tissues, the hepatopancreatic *caeca* and the cephalon of sexually mature gammarids. The library generated 323,731,722 paired-end reads of approximately 100 bp. After low quality read filtering, d*e novo* assembly of the Illumina sequences was carried out with a two-stage strategy. First, 1,406,956 sequences were preassembled with an N50 of 334 nt using the Velvet program. A dataset of 218,574 contigs with an N50 and N25 of 799 nt and 1221 nt, respectively, was then created with the Mira software program. The average GC content was 47.9%. The total length of mRNA sequences was 145,008,280 nt. A total of 386,238 reads, referred to as singletons, were not incorporated into the final assembly. Amongst the assembled contigs, 127,332 (58%) produced significant hits against the nonredundant NCBI database when taking an E-value threshold of $1E^{-10}$ and using the BLASTp search tool (**Supplementary Data, Table S1**). We extracted the taxonomic information associated with the best protein homologue for each significant hit. The largest number of contigs (8,927, *i.e.* 4%) showed the highest similarity to protein sequences from the crustacean, *Daphnia pulex*, while another significant number (4,053, *i.e.* 2%) were similar to proteins from the beetle, *Tribolium castaneum*. It is noteworthy that few similarities were found to proteins annotated from the amphipod, *Parhyale hawaiensis*, using this metric (221 contigs) although the maternal and embryonic transcriptomes are available for this organism (26).

In addition, 33 BLASTp results are annotated as *cytochrome c oxidase subunit* 1 (*CO*1) from diverse species, mainly crustacea (**Supplementary Data, Table S1**). Because this gene has been proposed as a genetic marker for species assignment in *Gammarus* (33), we performed a thorough search for *CO*1 sequences from GFOSS by BLASTn using 376-bp sequences available for organisms from a regional study covering the area-of-origin of our population (accession numbers X97163–X97185;

(34)). Phylogenetic tree reconstruction (based on the classification proposed by Meyran *et al*. (34)) revealed the presence of different *G. fossarum* lineages concomitant with *G. pulex* lineages. The wild population sampled for RNA-Seq is, thus, an assembly of different lineages of the *G. pulex/fossarum* species complex. Our RNA-Seq-derived database thus constitutes a resource that is potentially of interest for studies in gammarids other than the *G. fossarum* species.

Establishment of protein catalogues in *Gammarus fossarum*.

Our proteogenomic experiment led to a global dataset of 1,148,367 MS/MS spectra recorded. These indicated a total of 7,644 peptide sequences and certified 1,873 protein sequences. A total of 218 (12%) proteins are lineage-specific proteins (*i.e.* orphans) and are described for the first time. This relatively high number of mass-spectrometry-certified orphans indicates the phylogenetic distance of *G. fossarum* amongst the already-sequenced animal models, and that crustacea in general are poorly characterized in terms of molecular biology. Considering the taxonomic information obtained through the species identity of the most closely related homologous sequence, a relatively low number of proteins showed significant sequence similarities to protein sequences from *Daphnia pulex* (159, 8%) and from the beetle, *Tribolium castaneum* (70, 4%). A summary plot reflecting the global proteome is presented in **Figure 3**.

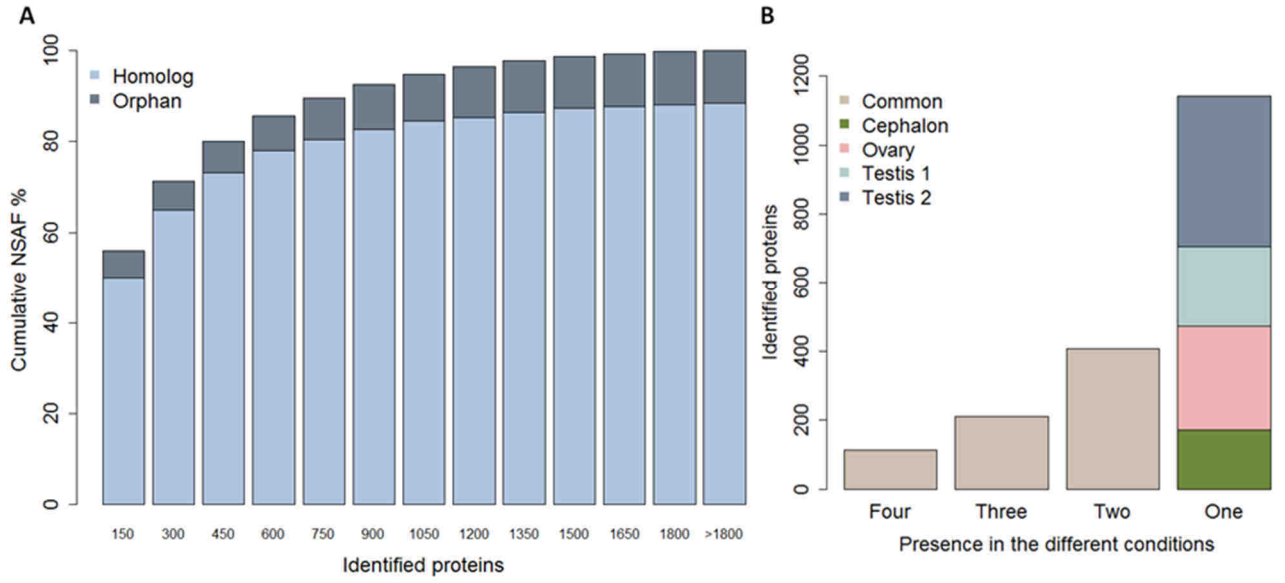


Figure 3: Analysis of the 1873 MS/MS identified proteins

(A) Distribution of identified proteins by class of 150 proteins according to their cumulative NSAF and respective contribution of homolog and orphan proteins. (B) Number of proteins identified in respectively, four, three, two and one biological tissue, male reproductive tissue is represented twice in the global analysis (Testis 1: protein catalogue and Testis 2: spermatogenesis experiment).

As shown in Panel A, where the distribution of the proteins in terms of label-free quantitation is shown for groups of proteins ranked from the most abundant to the least abundant, the ratio of orphans is remarkably constant. Panel B indicates the occurrence of the MS/MS-detected proteins in the different tissues. The specific study of these proteins appears worthy because of the relatively low overlap between tissues. The relatively low number of RNA contigs certified by mass spectrometry in the study may be explained by differences in i) the dynamic range between proteome and transcriptome, ii) the nature of the samples analyzed, as the hepatopancreatic *caeca* were not included in the proteomic study, and iii) the redundancy of RNA sequences that could be due to site-specific variants resulting from the heterogeneity of the animal population, diversity in terms of splicing variants, sequencing errors with the Illumina high-throughput approach, and some assembly errors resulting from the fact that no genome-backbone was available for the assembly. The assignment rate of the MS/MS spectra is low, especially with the OFF-GEL approach (5%) compared to the gel-based shotgun approach when analyzing the same tissue, *i.e.* testis tissues, but this figure agrees with a previous interpretation of data acquired for two other complex animals (35). Indeed, the high dynamic range of the proteome in these complex organisms, and the diversity of the proteome, lead to noisy MS/MS spectra for low abundance proteins. Post-translational modifications could be an additional factor of complexity.

The cephalon proteome. When using the NCBInr database, a relatively low number of MS/MS spectra (3103) could be assigned amongst the 231,098 MS/MS spectra recorded. The 878 unique peptide sequences (**Supplementary Data, Table S2**) proved the presence of 135 proteins (**Supplementary Data, Table S3**). Meanwhile, using our customized, RNA-Seq-derived database, 10,647 MS/MS spectra could be assigned, indicating the existence of 2,576 distinct peptide sequences (**Supplementary Data, Table S4**). In this case, a total of 546 mRNA-translated products (**Supplementary Data, Table S5**) were identified. Using the BLASTp tool, we observed that 505 (92%) exhibited significant sequence similarities with an already-functionally-annotated protein from another sequenced organism (E-value below $1E^{-10}$). Of note, compared with the results obtained with the NCBInr database, our customized database led to a large increase of identified peptides (~3-fold) and consequently, proteins (~4-fold).

Ovary and testis proteomes. Using the GFOSS database, among the 700,078 MS/MS spectra recorded, a total of 36,516 could be assigned (**Supplementary Data, Table S6**). These validated a large list of peptides (5,435 unique peptide sequences), which allowed the identification of 1,219 mRNA-translated products (**Supplementary Data, Table S7**). Among these, 1,069 proteins (88%)

showed significant similarities to previously annotated proteins from other organisms. The other 150 proteins (12%) can be referred as orphans. The level of orphans is higher (1.5x) for these reproductive tissues than has been observed for the cephalon proteome, highlighting the lack of knowledge regarding the key molecular players for both reproductive tissues. Furthermore, among the five proteins contributing 10% of the ovary and testis proteomes in terms of quantity, as estimated by the normalized spectral abundance factor, one (ID 68651) does not have any functional annotation. **Figure 4** shows the complementary strategy followed for establishing the protein profiles of male reproductive systems.

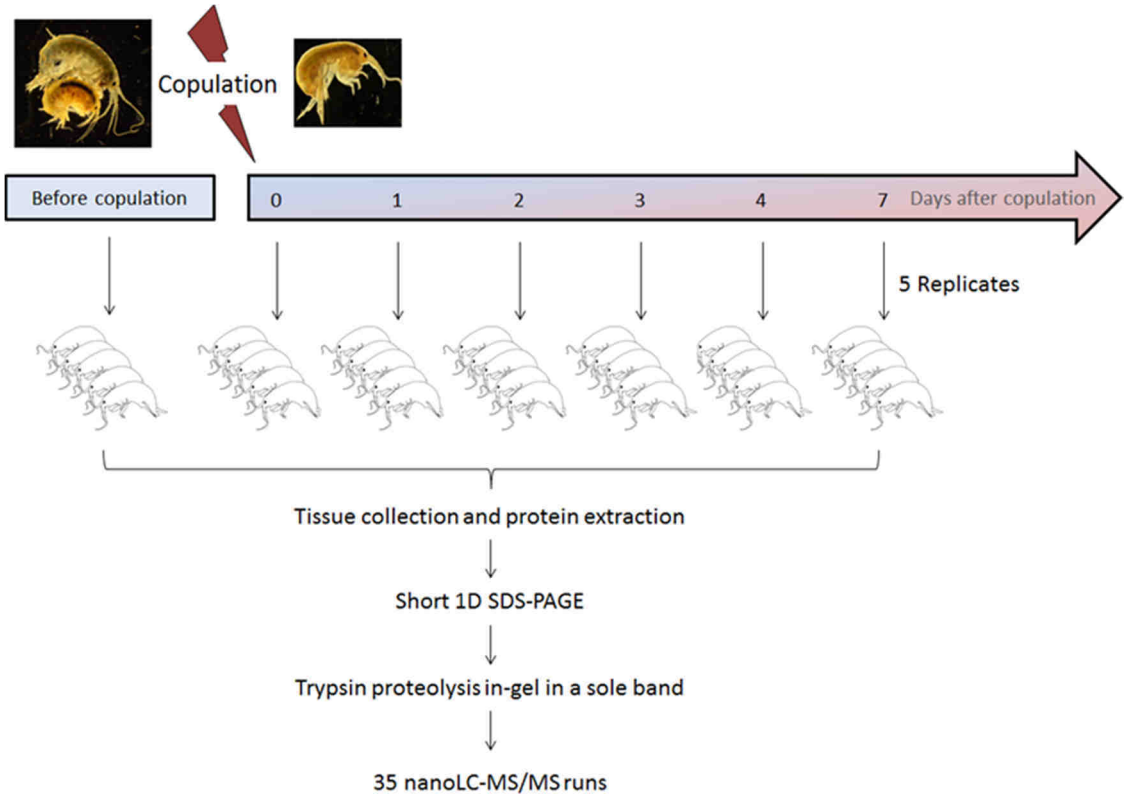


Figure 4: Proteome dynamics of the male reproductive tissue from *G. fossarum*

Five replicates were sampled at seven different spermatogenesis stages. The specific tissue was excised from individual organisms and analyzed as indicated.

These were analyzed at seven different spermatogenesis stages, on five biological replicates. Using the GFOSS database, among the 217,191 MS/MS spectra recorded, a total of 50,793 could be assigned (**Supplementary Data, Table S8**). These validated a large list of peptides (2,699 unique peptide sequences) and allowed label-free quantitation of 914 translated contigs (**Supplementary Data, Table S9**). Compared with the previous analysis of the male reproductive tissue proteome, 437 further products were newly identified here. Amongst the 914 proteins, 853 items (93%) showed significant similarities to previously annotated proteins from other organisms. The other 61 proteins (7%) can be referred as orphans. In this global dataset, only three proteins contribute

to 10% of the proteome in terms of quantity: two histones (ID_13250 histone H2A and ID_65834 histone H3.2-like) and one protein (ID_100349) for which the function could not be predicted.

Identification of *G. fossarum* proteins associated with reproductive function.

Automatic functional annotation. The functional annotation of the 1,873 identified proteins was performed by cross-species prediction based on sequence similarity searches. The function of a total of 1,216 proteins (65%) could be annotated with relative confidence (threshold E-value set at $1E^{-10}$), as marked biological differences between phylogenetically distant organisms may lead to the acquisition by sequence-related proteins of very different functions, which can only be differenciated by functional studies and not by sequence comparison alone. We identified 66 mRNA-translated products that may contribute to the *G. fossarum* reproductive process, on the basis of this functional annotation complemented by our biological expertise in crustacean physiology. These have been grouped into families according to the biological processes in which they may be involved (**Supplementary Data, Table S10**).

A total of six contigs are markedly related to enzymes involved in the hormonal regulation of steroids and isoprenoid methyl farnesoate (MF) synthesis. Four were identified as having a high probability of being involved in methyl farnesoate metabolism. Two isoforms of farnesoic acid O-methyltransferase (FAMeT) were identified in our survey, one exclusively detected in the male reproductive system while the other was also detected in the cephalon. Based on the phylogenetic analysis by Hui *et al*. (36), both contigs are integrated among the crustacean FAMeT family (**Supplemental Data File S1**) sustaining their involvement in MF biosynthesis. Additionally, two proteins that may possibly be involved in sexual hormone catabolism were sequenced from testis. While steroids are substrates for cytochromes P450 (37), two CYP isoforms were identified in the reproductive tissues. One member of the CYP12A2 cytochrome P450 family was identified in the female reproductive tissue and one member of the CYP4C39 cytochrome P450 was identified in male reproductive tissue. In the cuticle degradation process, closely linked to the molt cycle, which is intricately synchronized with female reproduction in amphipods (5), seven contigs annotated as chitinase, an enzyme involved in the hydrolysis of chitin glycosidic bonds, were found in our *G. fossarum* protein catalogue. Two isoforms were identified in the cephalon, while one isoform was observed in the female reproductive tissue. The four other contigs were identified in testis at different spermatogenesis stages. Among the seven contigs, five are related to the protein family analyzed previously by Huang *et al*. (38) (**Supplemental Data File S1**), from which four belong to

protein groups unrelated to the molting process (chitinase 4 and 1) and one (ID_12415) belongs to the chitinase 2 group.

Remarkably, a large number of oogenesis-related proteins were detected by tandem mass spectrometry. Among these mRNA-translated products, 45 are lipoproteins, *i.e*. major yolk proteins or potential components of major yolk proteins, and five are receptors, involved in the incorporation of the yolk proteins. Based on the prediction of function, the existence of twenty mRNA-translated products annotated as vitellogenin (VTG) was identified, seven with close similarity to an existing sequence from *G. fossarum* (29). Some contigs indicated the same protein, leading to a total of fourteen distinct proteins. Apart from these vitellogenins, a protein showing the highest similarity to a decapod apolipocrustacein (the major egg yolk protein in decapods according to Avarre et *al*. (39) was detected. The 23 other contigs were named under the generic term of clottable proteins, and some indicated the same protein, leading to a total of twelve distinct proteins. We retained this "clotting" annotation in decapods as indicative of proteins potentially involved in *Gammarus* female reproduction, because clotting proteins in decapods result from the neofunctionalization of a protein belonging to the groups of metazoan VTG, including non-decapod crustacean VTGs (39). While the majority of lipoproteins were identified in the female reproductive tissue, seventeen mRNA-translated products were also identified in the other tissues. Finally, four contigs with the term "sperm" in their predicted functions were identified. Three isoforms of the epididymal sperm-binding protein were observed in testis at different spermatogenesis stages. Finally, one nuclear autoantigenic sperm protein was detected in the female reproductive tissue.

Comparative proteomics of male and female reproductive tissues. In order to quickly pinpoint new candidates involved in the reproductive process, a cross comparison of protein abundances in male and female reproductive tissues was performed with a label-free shotgun strategy. Among the 1,219 mRNA-translated products detected in the reproductive tissues, 401 (33%) are shared by both males and females, as shown in **Figure 5 Panel A**. Subsequently, for a detailed comparison of the most abundant proteins between both tissues, spectral-count-based quantitative proteomics was performed. A multiple-hypothesis-testing Student's t-test was applied on our MS/MS dataset by means of the Tfold module of the PatternLab software program (40). **Supplementary Data, Table S7** lists the calculated protein variations between male and female reproductive tissues and their p-value confidence.

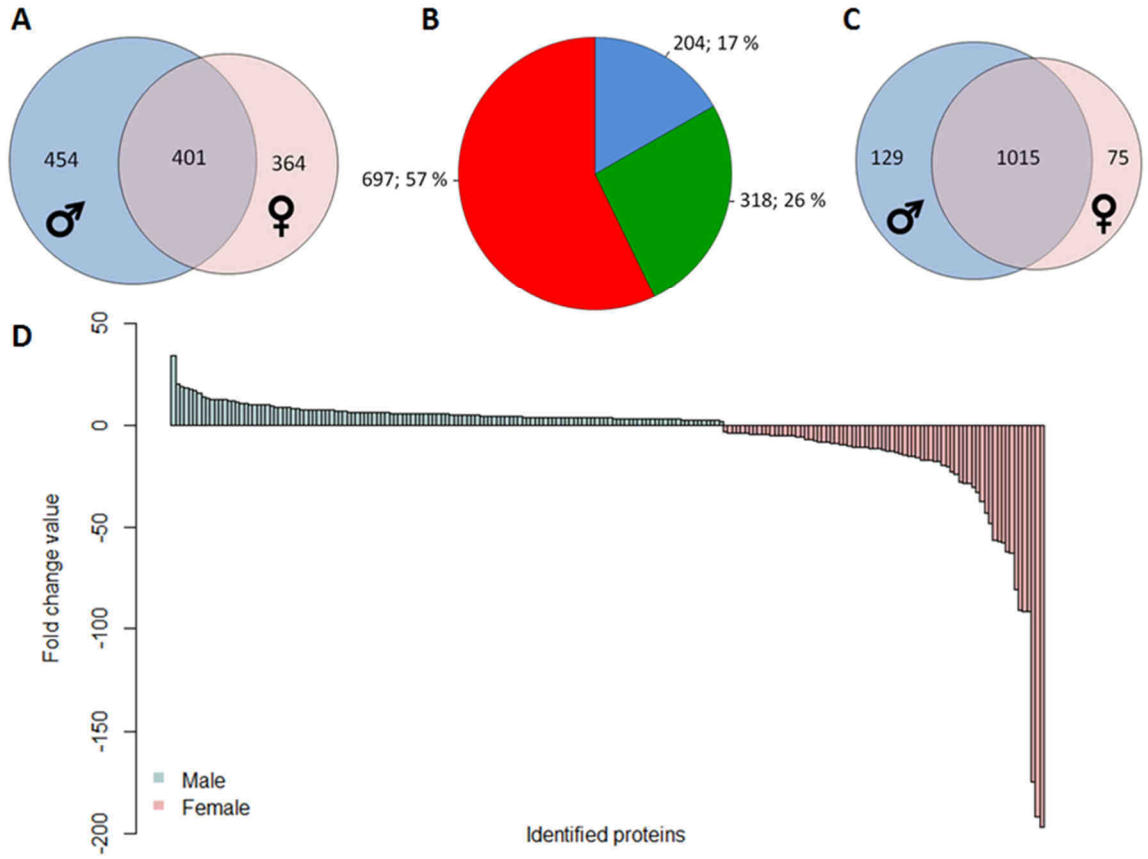


Figure 5: Comparative analysis of the male and female reproductive tissue

(A) Total number of proteins identified in male and female reproductive tissues. (B) Distribution of the proteins among the different categories obtained after data analysis using the Tfold module of the PatternLab software program, *i.e.* Blue class (fold change > 2 and p-value < 0.05), Green class (fold change > 2 and p-value > 0.05), and Red class (fold change > 2 and/or p-value > 0.05). (C) Number of proteins with significant sexual dimorphism (fold change > 2 and p-value < 0.05) according to the Tfold analysis. (D) Distribution of proteins with significant sexual dimorphism according to their fold change value.

Figure 5 Panel B presents the distribution of the proteins according to their fold change and p-value. Taking into account a fold change of at least 2.0 and a confidence threshold of 0.95 (p-value < 0.05), 129 contigs were significantly more frequently detected in the male tissue than the female tissue, while 75 were significantly more highly detected in the female than the male tissue (**Figure 5, Panels C & D**).

Amongst the contigs more highly detected in the female reproductive tissue, 40% (31) are orphans. Thirty-three percent (25) belong to the large lipid transfer protein (LLTP) superfamily, including 14 vitellogenin (11 proteins), one apolipocrustacein and 10 clottable proteins (7 proteins). These were all classified as major yolk in the previous section but these mRNA-translated products proved to be female specific. Amongst other female-specific products, the previously mentioned cytochromes P450, and three heat-shock proteins (70 kDa and 21 kDa),

were identified. We also observed proteins with housekeeping functions, such as energy production (two ATP synthases), amino-acid metabolism (delta-1-pyrroline-5-carboxylate dehydrogenase), and cytoskeletal structure maintenance (tubulin).

Table 1 presents the characteristics of the five contigs with the strongest sexual dimorphism for each sex. In females, the contig 19261-encoded protein (frame 6) that exhibits the strongest sexual dimorphisms (a fold change between female and male tissues of 175) is an orphan, and the others are part of the LLTP family.

Table 1: The most differentially detected proteins in female and male reproductive tissues

Contig ID	Probable function	Fold change	p value	Number of MS/MS spectra	Number of peptide sequences
Female reproductive tissue					
25127	Clottable protein	197	$0.4E^{-2}$	589	8
277	Vitellogenin	192	$0.8E^{-2}$	574	37
19261	No predicted function	175	$3.8E^{-2}$	873	20
276	Clottable protein	91	$0.3E^{-2}$	271	23
194758	Apolipocrustacein	91	$0.7E^{-2}$	271	29
Male reproductive tissue					
2029	Projectin	34	$3.7E^{-2}$	100	28
6444	I-connectin	20	$1.1E^{-2}$	58	11
3157	No predicted function	19	$0.1E^{-2}$	55	9
22255	Myosin	19	$2.1E^{-2}$	132	5
608	No predicted function	18	$2.7E^{-2}$	51	8

In male reproductive tissue, most of the contigs (41%) could not be functionally annotated due to the low sequence similarities with already-annotated proteins. Remarkably, a set of 40 contigs (31%) are involved in various aspects of cellular shaping by cytoskeletal structure maintenance. Additional proteins take part in energy metabolism (7 contigs), immunity (3 contigs), amino-acid metabolism (3 contigs) and hydrolysis of glycosidic linkage (3 contigs). One isoform of the juvenile hormone epoxide hydrolase (JHEH) proved to be specific to the male reproductive tissue. The membrane protein, flotillin-1 (ID 5059), which is included in the five contigs contributing to one-tenth of the reproductive proteome, is specific to this tissue. Of note, seven contigs are calcium-

dependent or involved in calcium homeostasis. Compared with the proteome of the female organ, a lower number (10) of proteins (8%) are orphans. Finally, among the contigs that were significantly more frequently detected in the testis (**Table I**), three are involved in cytoskeletal structure maintenance while the two others have no predicted function.

***Identification of spermatogenesis-related proteins by means of protein dynamics*.** In order to indicate new proteins that could be directly related to spermatogenesis in *G. fossarum*, we analyzed the proteome dynamics of the male reproductive tissue. For this, we correlated the temporal patterns of proteins detected in testis sampled at seven different spermatogenesis stages: Day -1, *i.e.* pre-copulatory mate guarding, Day 0, Day 1, Day 2, Day 3, Day 4 and Day 7 after copulation. Among the 914 proteins identified in this analysis, a total of 17 proteins presented interesting clustered temporal variations. Of these, nine proteins are devoid of functional annotation (53%) and three were more frequently detected in male reproductive tissue in the previous comparative analysis: one protein with no predicted function (ID 608), a calcium transporter (ID 4227) and a hemocytin-like protein (ID 461). Five distinct trends were delineated (**Figure 6**). The characteristics of the corresponding proteins from each cluster are listed in **Supplemental Data Table S9.** These different trends demonstrate the complexity of the dynamic variation patterns of the proteome during spermatogenesis. As shown in **Figure 6**, trend A cluster, comprising 3 proteins, was defined by a general steady state until Day 3, then a decreased expression at Day 4, followed by a maximal expression at Day 7. Trend B cluster, also comprising three proteins, showed an expression maximum at Day 0, followed by a continuous decrease until Day 3 and then an increase until Day 7. Trend C cluster highlighted three proteins whose expression decreased until Day 1, followed by a continuous increase until Day 3. The three proteins from the trend D cluster exhibited an increasing expression until Day 1 or Day 2, and then a decrease from Day 2 to Day 4. Trend E is the dominant pattern, with five proteins, including one male-specific protein with no predicted function (ID 608), with expression continuously decreasing until Day 2, and then increasing from Day 3 to Day 4. Both Trend D and Trend E are somehow correlated with spermatogenesis in *G. fossarum*, as the protein increase at day 0 or over the two first days corresponds to the beginning of spermatogenesis. The maxima of the three other trends (at day 3 or day 4) correlate with the change in spermatozoon production rate.

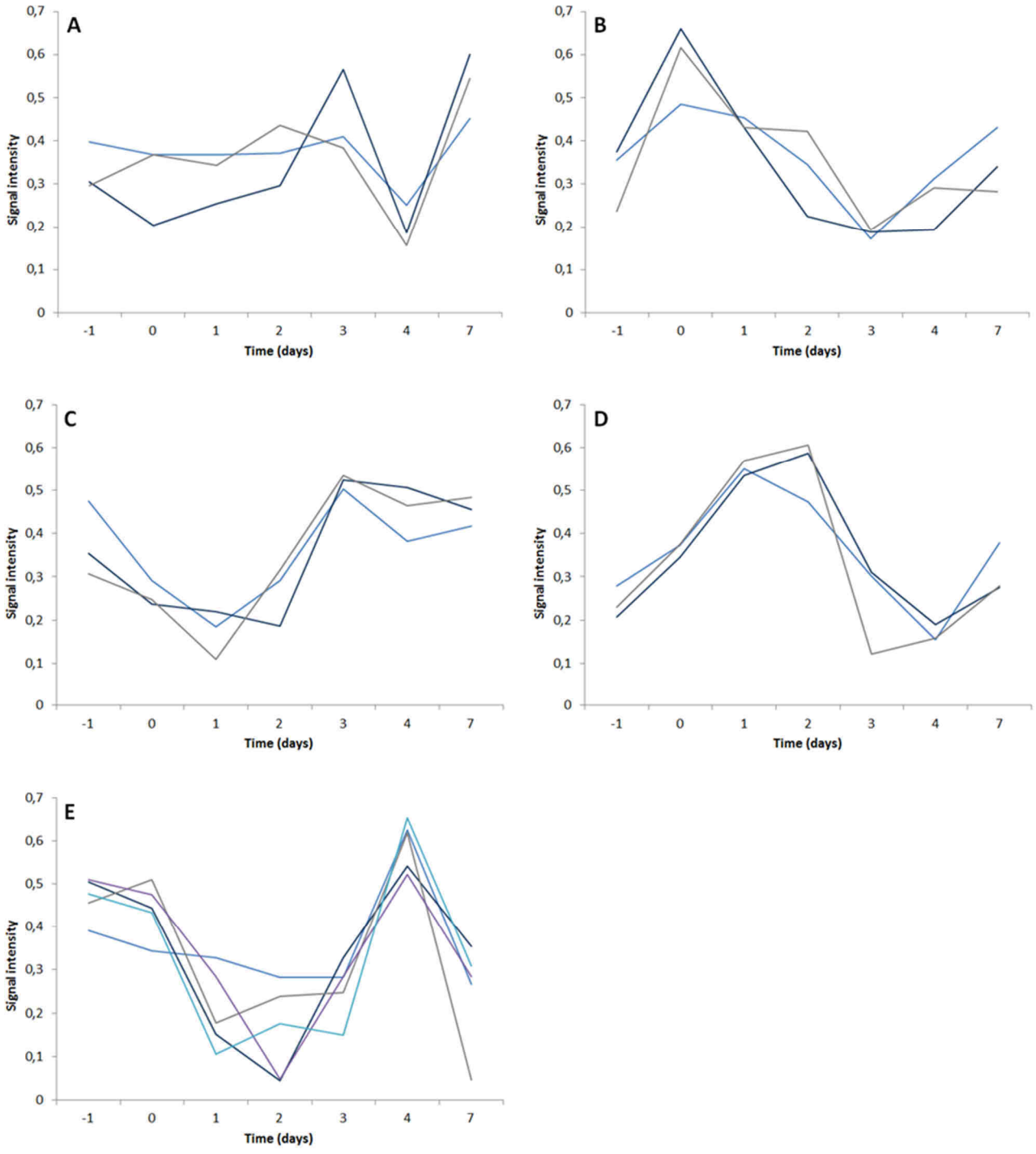


Figure 6: Clustered temporal abundance variations of proteins during spermatogenesis, before and after copulation.

The different trends were obtained from clustering the spectral count using the TrendQuest module of the PatternLab program, and are displayed in Panels A, B, C, D and E. Cluster health was set to 0.9 and the minimum items per cluster to 3.

4. DISCUSSION

Although crustacea are one of the largest and most specious groups of animals, inhabiting all aquatic and terrestrial environments, little comprehensive molecular information is available for this major arthropod group. Today, the only available crustacean genome is that of *Daphna magna*, a Branchiopod Cladocera (1). According to pancrustacean phylogeny, branchipoda are more closely related to hexapoda (insects) than to malacostraca (including decapods and amphipods) (41). Consequently, the large wealth of information obtained for the amphipod, *Gammarus*, both at the transcriptomic and proteomic levels, is really unique and will allow further comparisons when other malacostraca are studied. Moreover, such molecular information is a key resource for understanding amphipod biology and enables the development of new biomarkers for *in situ* toxicological bioassays (42). Indeed, while daphnids, with their short generation time, are laboratory workhorses, their ecology (lentic ecosystem) and their small size hinder their application in the field in running water. On the other hand, the ecology of *G. fossarum* makes it ideally suited for river biomonitoring. Furthermore, its size enables the direct targeting of specific organs, allowing the study of specific physiological functions, such as reproduction. We identified 218,574 contigs in the *Gammarus* transcriptome. We observed that ~40% of the contigs could be *Gammarus*-specific genes as they present poor sequence similarity with previously known genes. Thus, as shown with the cephalon proteome, a shotgun proteomic approach based on a dedicated RNA-Seq-derived database greatly improved MS/MS data interpretation compared with the use of current protein sequence databases from public resources. Here, the experimental difficulty of our approach is the dissection of relatively large quantities of organs from small animals (~1 mm size) to extract sufficient mRNA and protein material. However, the interpretation of data is relatively straightforward, as the proteins identified and quantified by means of the shotgun nanoLC-MS/MS approach give a relatively comprehensive view of the *G. fossarum* molecular players. We recently proposed a quantitative LC-MS/MS assay for the quantitation of vitellogenin as a biomarker of endocrine disruption (29). The extensive protein catalog described here will allow further development of toxicological biomarkers, with the aim of a rapid evaluation of the impact of pollutants in freshwaters (42).

In amphipods, molting and oogenic cycles are closely linked. While the new exoskeleton is still flexible after ecdysis, mature oocytes are spawned through the oviducts into the marsupium and are immediately fertilized by the male. The molting process is controlled by steroid molting hormones, called ecdysteroids. Ecdysteroid biosynthetic reactions are catalyzed by P450 mono-oxygenases and associated reductases (43). In this proteogenomic survey, a cytochrome P450

mono-oxygenase isoform (CYP450, ID_100255) was identified in the male reproductive tissue, while another one (ID_122081) was shown to be tissue-specific for the female organ. On the basis of their organ localization, the two identified proteins may be involved in ecdysteroid metabolism. In addition to ecdysteroid action, the hormonal system of arthropods (insects and crustacea) relies on sesquiterpenoid molecules such as farnesoic acid, methyl farnesoate (MF) and the juvenile hormone (JH). In crustacea, the end product of the sesquiterpenoid pathways seems to be MF, formed after methylation of farnesoic acid through the action of farnesoic acid-O-methyltransferase (FAMeT) (36). Here two FAMeT isoforms were identified, one exclusively detected in the male reproductive system while the other was also detected in the cephalon. The functions of MF in crustacea are, to date, elusive, but this hormone can be considered to be a determinant of male sex, inducing the production of male offspring in daphnids (44). It also stimulates gonadal maturation, MF being proven to stimulate testicular development in the crab, *Carcinus maneas* (45). Moreover, in the shrimp, *Litopenaeus vannamei*, RNA interference of putative FAMeT suggests that this enzyme is essential for the molt process by regulating molt-related genes (46). Although the mechanism of biosynthesis of MF is well established, little is known about its catabolism. In insects, JH inactivation is processed by JH carboxyesterase and epoxide hydrolase. Here, homologues of both enzymes have been identified in the male reproductive tissue, but the presence of JH in *G. fossarum* remains to be established experimentally. For pest control in agriculture, the JH/MF hormonal machinery is targeted directly by molecular analogues of this hormone. Consequently, these act as endocrine-disrupting chemicals for non target organisms such as *G. fossarum.* The identification of proteins involved in MF metabolism will allow the further development of endocrine disruption biomarkers in crustacea (42).

Regarding morphogenesis, the major constituent of the arthropod exoskeleton is chitin, a linear polymer of N-acetyl-D-glucosamine. During the premolt stage of crustacea and insects, the organisms secrete enzymes for chitin degradation in order to recycle the old cytoskeleton while synthesizing a new one. In crustacea, measurement of the activity of chitinolytic enzymes has already been performed in crabs, as a biomarker of molting (47). Here, seven contigs annotated as chitinase were detected by mass spectrometry. According the protein family classification by Huang et *al.* (38), the one belonging to the chitinase 2 group, is an interesting candidate biomarker for molting. This observation is in accordance with the tissue distribution of this protein, as this contig was sequenced in the cephalon while the others were detected in the male and female reproductive tissues, suggesting a physiological function unrelated to the molt process.

Among the most abundant proteins involved in the female reproductive process identified in our study, most are members of the large lipid transfer protein (LLTP) superfamily. These proteins are structural scaffolds for the assembly of lipoproteins, as they bind a multitude of lipid species. Apolipoproteins, clottable proteins and vitellogenin (Vtg) are members of the LLTP superfamily. Generally, Vtg is considered to be the major yolk protein in most oviparous organisms (39). In the present study, we identified a total of 45 mRNA-translated products belonging to the LLTP superfamily, mainly on the basis of sequence similarities to decapod protein homologues, for which functional annotation is confusing because of neofunctionalization of orthologous VTG gene to clotting function (39). Among all these candidates potentially involved in the female reproductive process, only 25, pinpointing the existence of 18 distinct proteins, were found to be preferentially expressed in female reproductive tissue using comparative proteomics. Evidence of multiple copies of vitellogenin genes had already been obtained through targeted PCR amplification: two copies in *D. magna* (48) and four copies in the mosquito, *Culex tarsalis* (49). Our study, the first ever performed for an arthropod at such a large scale at the protein level, reveals unsuspected diversity of LLTP members.

Most of the female-specific proteins are orphans (40%), including the most abundant one (ID_19261), with no putative conserved domains detected. This suggests that the major yolk protein may not be structurally related to LLTPs in gammarids. This hypothesis is reinforced by molecular data obtained from other species. The major yolk protein is a transferrin-like, iron-binding protein in sea urchins (50), while it is a lipase-related protein in higher diptera (51). The existence of many gammarid-specific proteins, such as the large variety of LLTPs revealed by our analysis, may suggest that major yolk proteins are diverse and possibly genus specific, this diversity being explained by the variety of carried lipid species. It is noteworthy that the female organisms used in our experiment were in the last phase of their reproductive cycle but the abundances of the different proteins detected here should change during the course of the cycle. According to the essential role of yolk proteins to ensure the development of a new generation, these results, together with our knowledge of the *G. fossarum* ovarian cycle and embryonic development (5), generate multiple research hypotheses that require further experiments to ascertain their functions.

Regarding proteins involved in male reproductive function, we also employed an additional strategy for uncovering molecular players involved in the spermatogenesis process. While the first protein catalogue from testis tissue was carried out on a specific spermatogenesis stage,

proteome dynamics of this tissue was performed on seven different stages, leading to the identification of 437 additional proteins. Proteome dynamics during spermatogenesis revealed five distinct temporal trends, suggesting that spermatogenesis is a highly complex temporal event. The majority of the proteins delineated in these trends are devoid of functional prediction. Indeed, the spermatozoon cell is the most diverse cell type in the animal kingdom and few investigations have been carried out on genes that regulate spermatogenesis in crustacea. Thus, these proteins are interesting candidates for a potential involvement, directly or indirectly, in male reproductive function. Several proteins that were functionally annotated in our study are crucial for the integrity of reproductive function. For example, flotillins, members of a ubiquitous family of membrane proteins, are involved in numerous cellular processes and are also essential for spermatozoon functionality. In spermatozoa, the acrosome is a cap-like structure that enables fusion with the egg, and proteins from the flotillin family have been shown to be involved in its biogenesis and functionality (52). In our study, flotillin-1 (ID 5059) proved to be abundant, as it was included in the five most detected proteins from the reproductive proteome, and classified as male specific. Among acrosomal enzymes, glycohydrolases are believed to digest the extracellular glycocalyx surrounding the egg (53). The three glycohydrolases classified as male specific are potentially acrosomal enzymes involved in sperm-egg fusion.

In order to ensure sperm viability, a “programed degradation” of spermatozoa before they become obsolescent is carried out by the immune system. In the male shrimp, *Litopenaeus vannamei*, progressive melanization of spermatophores has been documented for elimination through an innate immune response (54). Melanization is initiated by the activation of prophenoloxidase, a protein that was found to be male specific. Two other male-specific proteins involved in innate immunity were identified: a hemocytin-like protein and a transglutaminase (55). We also detected several isoforms of epididymal sperm-binding 1 during our spermatogenesis kinetics experiment, which has been proposed as an indicator of sperm defectiveness (56).

In our proteogenomic survey, we identified a strong prevalence in the male reproductive tissue of calcium-dependent proteins and proteins involved in calcium homeostasis. Indeed, regulation of the Ca^{2+} internal concentration is believed to be crucial for sperm maturation, capacitation and sperm-egg interaction, and is consequently subject to strict spatiotemporal control. One sarcoplasmic calcium-binding protein and three isoforms of sarcoplasmic/endoplasmic reticulum calcium ATPase were detected, with one isoform (ID_4227) that showed dynamic changes during spermatogenesis. These proteins, previously identified in the acrosome from vertebrates, are believed to sequester Ca^{2+} within this organelle, supporting the hypothesis that the acrosome acts

as an intracellular Ca^{2+} storage compartment (57). In addition, three proteins identified as calcium dependent were classified as male specific: transglutaminase and two isoforms of copine-8. The functions of copine-8 are elusive, but in humans the gene is predominantly expressed in prostate and testis, suggesting an important role in the development and regulation of male sexual characteristics (58).

In conclusion, we have generated a proteogenomic dataset for the amphipod, *Gammarus fossarum*, using Illumina-Solexa pyrosequencing and shotgun proteomics, which paves the way for several studies relating to this ecotoxicological sentinel organism. Here, we have focused our analysis on the reproductive process in order to propose new biomarkers related to its disturbance. To our knowledge, this represents the first extra-large proteogenomic resource for a crustacean. At this stage, we were able to certify by tandem mass spectrometry the presence of 1,873 proteins in *G. fossarum*. Further analyses are needed to go deeper into this catalogue. While the results of functional annotation by sequence similarities were quite limited because of the absence of other datasets for this lineage, we could highlight interesting candidates involved in various processes related to reproduction, such as hormonal metabolism, morphogenesis, sperm viability and calcium homeostasis. We also found a high diversity of vitellogenin-like proteins in the female organ. However, confidence in these functional annotations is low, as for any new model organism. For this reason, we performed a comparative analysis of the abundance of proteins in male and female reproductive tissues and quickly pinpointed specific candidates involved in the reproductive process. Interestingly, we uncovered a substantial proportion of orphan proteins, with a significantly higher ratio in reproductive tissue compared to the cephalon. Reproductive proteins are known to evolve more rapidly because of divergent natural selection leading to molecular reproductive isolation in closely related species (59). Finally, several orphans were found in abundance in male and female reproductive tissues, and classified as sex specific, and thus potentially involved, directly or indirectly, in reproductive function. In order to better understand their functions, and the respective molecular mechanisms, new investigations should now be performed with the protein sequence knowledge detailed in the present study as a starting point.

Abbreviations: CO1, cytochrome c oxidase subunit 1; CYP, cytochrome P450; FAMeT, farnesoic acid O-methyltransferase; JH, juvenile hormone; JHEH, juvenile hormone epoxide hydrolase; LLTP, large lipid transfer protein; MF, methyl farnesoate; Vtg, vitellogenin.

Acknowledgements: J. Trapp was supported by a fellowship from the Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (France) and from the 'programme transversal de toxicologie' of the Commissariat de l'Energie Atomique et aux Energies Alternatives (France). We thank the Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (France) and the Commissariat de l'Energie Atomique et aux Energies Alternatives (France) for financial support. We also thank K. Abbaci, G. Le Goff, G. Miotello, G. Steinmetz, H. Quéau and the PRIDE Team for their technical assistance, as well as P.C. Carvalho (COPPE, Rio de Janeiro, Brazil) for adapting the export option of the PatternLab program.

REFERENCES

1. Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Caceres, C. E., Carmel, L., Casola, C., Choi, J. H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., Frohlich, T., Geiler-Samerotte, K. A., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E. V., Kultz, D., Laforsch, C., Lindquist, E., Lopez, J., Manak, J. R., Muller, J., Pangilinan, J., Patwardhan, R. P., Pitluck, S., Pritham, E. J., Rechtsteiner, A., Rho, M., Rogozin, I. B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y. I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J. R., Andrews, J., Crease, T. J., Tang, H., Lucas, S. M., Robertson, H. M., Bork, P., Koonin, E. V., Zdobnov, E. M., Grigoriev, I. V., Lynch, M., and Boore, J. L. (2011) The ecoresponsive genome of Daphnia pulex. *Science* 331, 555-561
2. Kunz, P. Y., Kienle, C., and Gerhardt, A. (2010) Gammarus spp. in aquatic ecotoxicology and water quality assessment: toward integrated multilevel tests. *Rev Environ Contam Toxicol* 205, 1-76
3. Dangles, O., and Guerold, F. (2001) Linking shredders and leaf litter processing: Insights from an acidic stream study. *Int Rev Hydrobiol* 86, 395-406
4. Xuereb, B., Chaumot, A., Mons, R., Garric, J., and Geffard, O. (2009) Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda) Intrinsic variability, reference levels, and a reliable tool for field surveys. *Aquat Toxicol* 93, 225-233
5. Geffard, O., Xuereb, B., Chaumot, A., Geffard, A., Biagianti, S., Noel, C., Abbaci, K., Garric, J., Charmantier, G., and Charmantier-Daures, M. (2010) Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. *Environ Toxicol Chem* 29, 2249-2259
6. Jungmann, D., Ladewig, V., Ludwichowski, K. U., Petzsch, P., and Nagel, R. (2004) Intersexuality in Gammarus fossarum Koch - A common inducible phenomenon? *Arch hydrobio* 159, 511-529
7. LeBlanc, G. (2007) Crustacean endocrine toxicology: a review. *Ecotoxicology*, pp. 61-81
8. Kulkarni, D., Gergs, A., Hommen, U., Ratte, H. T., and Preuss, T. G. (2013) A plea for the use of copepods in freshwater ecotoxicology. *Environ Sci Pollut Res Int* 20, 75-85
9. Ansong, C., Purvine, S. O., Adkins, J. N., Lipton, M. S., and Smith, R. D. (2008) Proteogenomics: needs and roles to be filled by proteomics in genome annotation. *Brief Funct Genomic Proteomic* 7, 50-62
10. Renuse, S., Chaerkady, R., and Pandey, A. (2011) Proteogenomics. *Proteomics* 11, 620-630
11. Armengaud, J., Hartmann, E. M., and Bland, C. (2013) Proteogenomics for environmental microbiology. *Proteomics* 13, 2731-2742
12. Yates, J. R., 3rd, Eng, J. K., and McCormack, A. L. (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem* 67, 3202-3210
13. Gallien, S., Perrodou, E., Carapito, C., Deshayes, C., Reyrat, J. M., Van Dorsselaer, A., Poch, O., Schaeffer, C., and Lecompte, O. (2009) Ortho-proteogenomics: multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Res* 19, 128-135

14. Venter, E., Smith, R. D., and Payne, S. H. (2011) Proteogenomic analysis of bacteria and archaea: a 46 organism case study. *PLoS ONE* 6, e27587
15. Castellana, N. E., Payne, S. H., Shen, Z., Stanke, M., Bafna, V., and Briggs, S. P. (2008) Discovery and revision of Arabidopsis genes by proteogenomics. *Proc Natl Acad Sci U S A* 105, 21034-21038
16. Castellana, N. E., Shen, Z., He, Y., Walley, J. W., Cassidy, C. J., Briggs, S. P., and Bafna, V. (2014) An automated proteogenomic method uses mass spectrometry to reveal novel genes in Zea mays. *Mol Cell Proteomics* 13, 157-167
17. Rubiano-Labrador, C., Bland, C., Miotello, G., Guerin, P., Pible, O., Baena, S., and Armengaud, J. (2014) Proteogenomic insights into salt tolerance by a halotolerant alpha-proteobacterium isolated from an Andean saline spring. *J Proteomics* 97, 36-47
18. Wilmes, P., Andersson, A. F., Lefsrud, M. G., Wexler, M., Shah, M., Zhang, B., Hettich, R. L., Bond, P. L., VerBerkmoes, N. C., and Banfield, J. F. (2008) Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J* 2, 853-864
19. Christie-Oleza, J. A., Miotello, G., and Armengaud, J. (2013) Proteogenomic definition of biomarkers for the large Roseobacter clade and application for a quick screening of new environmental isolates. *J Proteome Res* 12, 5331-5339
20. Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., and Hartmann, E. M. (2014) Non-model organisms, a species endangered by proteogenomics. *J Proteomics* 105, 5-18
21. Ning, K., and Nesvizhskii, A. I. (2010) The utility of mass spectrometry-based proteomic data for validation of novel alternative splice forms reconstructed from RNA-Seq data: a preliminary assessment. *BMC Bioinformatics* 11 Suppl 11, S14
22. Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., and Mann, M. (2011) Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 7, 548
23. Wang, X., Slebos, R. J. C., Wang, D., Halvey, P. J., Tabb, D. L., Liebler, D. C., and Zhang, B. (2012) Protein identification using customized protein sequence databases derived from RNA-seq data. *J Proteome Res* 11, 1009-1017
24. Woo, S., Cha, S. W., Merrihew, G., He, Y., Castellana, N., Guest, C., MacCoss, M., and Bafna, V. (2014) Proteogenomic database construction driven from large scale RNA-seq data. *J Proteome Res* 13, 21-28
25. Blythe, M. J., Malla, S., Everall, R., Shih, Y. H., Lemay, V., Moreton, J., Wilson, R., and Aboobaker, A. A. (2012) High through-put sequencing of the Parhyale hawaiensis mRNAs and microRNAs to aid comparative developmental studies. *PLoS ONE* 7, e33784
26. Zeng, V., Villanueva, K. E., Ewen-Campen, B. S., Alwes, F., Browne, W. E., and Extavour, C. G. (2011) De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis. *BMC Genomics* 12, 581
27. Hook, S. E., Osborn, H. L., Spadaro, D. A., and Simpson, S. L. (2014) Assessing mechanisms of toxicant response in the amphipod Melita plumulosa through transcriptomic profiling. *Aquat Toxicol* 146, 247-257
28. Lacaze, E., Devaux, A., Jubeaux, G., Mons, R., Gardette, M., Bony, S., Garric, J., and Geffard, O. (2011) DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level. *Sci Total Environ* 409, 3230-3236
29. Simon, R., Jubeaux, G., Chaumot, A., Lemoine, J., Geffard, O., and Salvador, A. (2010) Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum). *J Chromatogr A* 1217, 5109-5115
30. Christie-Oleza, J. A., Miotello, G., and Armengaud, J. (2012) High-throughput proteogenomics of Ruegeria pomeroyi: seeding a better genomic annotation for the whole marine Roseobacter clade. *BMC Genomics* 13, 73
31. Christie-Oleza, J. A., Pina-Villalonga, J. M., Bosch, R., Nogales, B., and Armengaud, J. (2012) Comparative proteogenomics of twelve Roseobacter exoproteomes reveals different adaptive strategies among these marine bacteria. *Mol Cell Proteomics* 11, M111 013110
32. Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D.,

Wang, R., and Hermjakob, H. (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* 41, D1063-1069

33. Weiss, M., Macher, J. N., Seefeldt, M. A., and Leese, F. (2014) Molecular evidence for further overlooked species within the Gammarus fossarum complex (Crustacea: Amphipoda). *Hydrobiologia* 721, 165-184

34. Meyran, J. C., Monnerot, M., and Taberlet, P. (1997) Taxonomic status and phylogenetic relationships of some species of the genus Gammarus (Crustacea, Amphipoda) deduced from mitochondrial DNA sequences. *Mol Phylogenet Evol* 8, 1-10

35. Fröhlich, T., Arnold, G. J., Fritsch, R., Mayr, T., and Laforsch, C. (2009) LC-MS/MS-based proteome profiling in Daphnia pulex and Daphnia longicephala: The Daphnia pulex genome database as a key for high throughput proteomics in Daphnia. *BMC Genomics* 10

36. Hui, J. H., Hayward, A., Bendena, W. G., Takahashi, T., and Tobe, S. S. (2010) Evolution and functional divergence of enzymes involved in sesquiterpenoid hormone biosynthesis in crustaceans and insects. *Peptides* 31, 451-455

37. James, M. O., and Boyle, S. M. (1998) Cytochromes P450 in crustacea. *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol* 121, 157-172

38. Huang, Q. S., Yan, J. H., Tang, J. Y., Tao, Y. M., Xie, X. L., Wang, Y., Wei, X. Q., Yan, Q. H., and Chen, Q. X. (2010) Cloning and tissue expressions of seven chitinase family genes in Litopenaeus vannamei. *Fish Shellfish Immunol* 29, 75-81

39. Avarre, J. C., Lubzens, E., and Babin, P. J. (2007) Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate apolipoprotein B. *BMC Evol Biol* 7, 3

40. Carvalho, P. C., Fischer, J. S., Chen, E. I., Yates, J. R., 3rd, and Barbosa, V. C. (2008) PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics* 9, 316

41. von Reumont, B. M., Jenner, R. A., Wills, M. A., Dell'ampio, E., Pass, G., Ebersberger, I., Meyer, B., Koenemann, S., Iliffe, T. M., Stamatakis, A., Niehuis, O., Meusemann, K., and Misof, B. (2012) Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. *Mol Biol Evol* 29, 1031-1045

42. Trapp, J., Armengaud, J., Salvador, A., Chaumot, A., and Geffard, O. (2014) Next-generation proteomics: towards customized biomarkers for environmental biomonitoring *Environ Sci Technol* revised version submitted (August 2014, 14th)

43. Mykles, D. L. (2011) Ecdysteroid metabolism in crustaceans. *J Steroid Biochem Mol Biol* 127, 196-203

44. Lampert, W., Lampert, K. P., and Larsson, P. (2012) Induction of male production in clones of Daphnia pulex by the juvenoid hormone methyl farnesoate under short photoperiod. *Comp Biochem Physiol C Toxicol Pharmacol* 156, 130-133

45. Nagaraju, G. P., and Borst, D. W. (2008) Methyl farnesoate couples environmental changes to testicular development in the crab Carcinus maenas. *J Exp Biol* 211, 2773-2778

46. Hui, J. H., Tobe, S. S., and Chan, S. M. (2008) Characterization of the putative farnesoic acid O-methyltransferase (LvFAMeT) cDNA from white shrimp, Litopenaeus vannamei: Evidence for its role in molting. *Peptides* 29, 252-260

47. Mesquita, S. R., Guilhermino, L., and Guimaraes, L. (2011) Biochemical and locomotor responses of Carcinus maenas exposed to the serotonin reuptake inhibitor fluoxetine. *Chemosphere* 85, 967-976

48. Tokishita, S., Kato, Y., Kobayashi, T., Nakamura, S., Ohta, T., and Yamagata, H. (2006) Organization and repression by juvenile hormone of a vitellogenin gene cluster in the crustacean, Daphnia magna. *Biochem Biophys Res Commun* 345, 362-370

49. Chen, S., Armistead, J. S., Provost-Javier, K. N., Sakamoto, J. M., and Rasgon, J. L. (2010) Duplication, concerted evolution and purifying selection drive the evolution of mosquito vitellogenin genes. *BMC Evol Biol* 10, 142

50. Brooks, J. M., and Wessel, G. M. (2002) The major yolk protein in sea urchins is a transferrin-like, iron binding protein. *Dev Biol* 245, 1-12

51. Sappington, T. W. (2002) The major yolk proteins of higher Diptera are homologs of a class of minor yolk proteins in lepidoptera. *J Mol Evol* 55, 470-475

52. Wu, Y., Chen, X., Wang, S., Jiang, M., Zheng, B., Zhou, Q., Bi, Y., Zhou, Z., Huang, X., and Sha, J. (2012) Flotillin-2 is an acrosome-related protein involved in mouse spermiogenesis. *Journal of Biomedical Research* 26, 278-287

53. Tulsiani, D. R., Abou-Haila, A., Loeser, C. R., and Pereira, B. M. (1998) The biological and functional significance of the sperm acrosome and acrosomal enzymes in mammalian fertilization. *Exp Cell Res* 240, 151-164

54. Parnes, S., Raviv, S., Shechter, A., and Sagi, A. (2006) Males also have their time of the month! Cyclic disposal of old spermatophores, timed by the molt cycle, in a marine shrimp. *J Exp Biol* 209, 4974-4983

55. Vazquez, L., Alpuche, J., Maldonado, G., Agundis, C., Pereyra-Morales, A., and Zenteno, E. (2009) Review: Immunity mechanisms in crustaceans. *Innate Immun* 15, 179-188

56. D'Amours, O., Frenette, G., Bordeleau, L. J., Allard, N., Leclerc, P., Blondin, P., and Sullivan, R. (2012) Epididymosomes transfer epididymal sperm binding protein 1 (ELSPBP1) to dead spermatozoa during epididymal transit in bovine. *Biol Reprod* 87, 94

57. Jimenez-Gonzalez, C., Michelangeli, F., Harper, C. V., Barratt, C. L., and Publicover, S. J. (2006) Calcium signalling in human spermatozoa: a specialized 'toolkit' of channels, transporters and stores. *Hum Reprod Update* 12, 253-267

58. Maitra, R., Grigoryev, D. N., Bera, T. K., Pastan, I. H., and Lee, B. (2003) Cloning, molecular characterization, and expression analysis of Copine 8. *Biochem Biophys Res Commun* 303, 842-847

59. Marshall, J. L., Huestis, D. L., Garcia, C., Hiromasa, Y., Wheeler, S., Noh, S., Tomich, J. M., and Howard, D. J. (2011) Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. *Mol Biol Evol* 28, 423-435

2. Caractérisation de l'ovogénèse et identification des protéines de réserves du développement embryonnaire

Publication 4:

High-throughput proteome dynamics for discovery of key proteins in non-model species: unsuspected vitellogenins diversity in the crustacean *Gammarus fossarum*!

Soumise à *Environmental Science and Technology*

Judith Trapp [a,b], Jean Armengaud [b], Arnaud Chaumot [a], Gilles Imbert [b], Jean-Charles Gaillard [b], Olivier Geffard[a]

[a] Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, CS70077, F-69626 Villeurbanne, France
[b] CEA, DSV, IBEB, Lab. Biochim. System. Perturb., Bagnols-sur-Ceze, F-30207, France

Résumé :

En écotoxicologie, les techniques omiques sont largement employées pour l'identification des mécanismes moléculaires mis en place en réponse à un stress. Toutefois, pour tout organisme non-modèle, la véracité des fonctions biologiques prédites automatiquement pour les gènes et leurs produits doit être établie. En effet, ces inférences fonctionnelles sont réalisées en combinant les séquences et informations fonctionnelles d'organismes phylogénétiquement distants pouvant conduire à des prédictions erronées en raison de la distance évolutive avec l'organisme étudié. Le but de cette étude est d'identifier les protéines précurseurs du vitellus (*i.e*. vitellogénine) chez l'amphipode d'eau douce *Gammarus fossarum*. Chez les organismes ovipares femelles, la bonne production de la vitellogénine est vitale pour le développement embryonnaire, assurant le maintien de la population dans son milieu naturel. De même, chez les poissons mâles, sa synthèse est employée comme biomarqueur d'exposition aux substances oestrogènomimétiques. Dans cette étude, les ovaires et les embryons ont été caractérisés à cinq stades de développement par protéomique shotgun, permettant l'analyse des modifications du protéome au cours de l'ovogénèse et de l'embryogénèse. Ces données fournissent la première description moléculaire à large échelle du cycle reproducteur d'un crustacé alors que l'analyse temporelle a révélé respectivement trois et quatre profils de variations temporelles pour l'ovogénèse et l'embryogénèse. Cette approche a permis de proposer différentes hypothèses sur la fonction des protéines *e.g*. vitellogénèse primaire ou secondaire. Au final, huit protéines issues de différentes familles au sein de la superfamille des « large lipid transfer protein » ont été identifiées comme étant les « vraies vitellogénines » de *Gammarus fossarum*, mettant en évidence une diversité insoupçonnée de protéines impliquées dans la formation du vitellus.

Mots clés : Protéomique shotgun, profils d'expression, reproduction, crustacé, vitellogénine

.

High-throughput proteome dynamics for discovery of key proteins in non-model species: unsuspected vitellogenins diversity in the crustacean *Gammarus fossarum*!

Abstract

In ecotoxicology, omics-based approaches are widely used for the identification of gene products related to stress response. However, the reliability of automatic functional annotation of proteins is yet to be established for any new model organism. Indeed, functional predictions are derived by combining sequence and functional data from phylogenetically distant organisms, which can lead to inaccurate predictions due to quite different evolutionary scenarios. Here, we identified the major egg yolk protein precursors (*i.e*. vitellogenin) in the amphipod *Gammarus fossarum*, a sentinel species in freshwater risk assessment, for which the genome is not yet sequenced. In oviparous females, vitellogenin production is vital for embryonic development, ensuring population viability in the ecosystem, while its production in male organisms is commonly employed as a biomarker of exposure to xenoestrogens in fish. We conducted a large proteomic experiment on female ovaries and embryos at five different developmental stages and analyzed proteome temporal dynamics in oogenesis and embryogenesis. These data provide the first large-scale molecular description of a crustacean reproductive cycle, while the temporal analysis revealed three and four different protein trends in oogenesis and embryogenesis, respectively. This exhaustive proteomic dynamics approach allows several functional molecular hypotheses in the oogenesis process to be drawn, such as protein involvement in primary *vs.* secondary vitellogenesis. Finally, we identified eight proteins as “true vitellogenins” in *G. fossarum*, from different families of a large lipid transfer protein superfamily, revealing an unsuspected diversity of molecular players involved in yolk formation.

Keywords: Shotgun proteomics, Expression profile, Reproduction, Crustacean, Vitellogenin

1. INTRODUCTION

While research in translational medicine is based on few biological models, studies in environmental science is confronted to a multiplication of models and relevant species in regard to ecosystem functioning. Part of environmental risk assessment is based on the measurement of protein biomarkers on sentinel species, autochthonous or encaged organisms which are representative of their ecosystem. While specific development of enzyme-linked immunosorbent assays (ELISA tests) has been carried out on vertebrates' organism for monitoring protein biomarkers of environmental interest, the lack of knowledge regarding invertebrate molecular biology impedes such developments. The important molecular divergence between vertebrates and invertebrates or within invertebrates' lineages also hinders direct transposition and global use of the developed ELISA tests. In this context, specific and robust biomarkers available for invertebrates' species are rare (Trapp et *al*. 2014).

Thanks to the development of next generation sequencing technologies, in depth and high throughput molecular description of the biology of non-model species is now possible. Proteogenomics approach, combining genomics and proteomics data, is an up and coming strategy for protein discovery in non-model species (Wang et *al.* 2012, Armengaud et *al*. 2014, Woo et *al.* 2014). Yet, while proteogenomics strategy enables specific protein discovery in a non-model species, the inference of protein function currently based on sequences similarities is still problematic. Automatic pipelines predict gene function based on sequence similarities. However, for organisms with important phylogenetic distance compared to the ones present in database, this can lead to unknown or inaccurate predictions due to quite different evolutionary scenario (Studer and Robinson-Rechavi 2009, Nikinmaa and Rytkonen 2011).

The crustacean amphipod Gammarus fossarum is commonly used as a sentinel species in freshwater risk assessment of environmental chemical contamination, employed in both laboratory experiments and in situ bioassays. Impacts on pollutants are currently assessed by measuring life-history traits such as locomotor behavior and feeding rate (Felten et *al*. 2008, Coulaud et *al*. 2011) and some mode of action can be decipher such as neurotoxicity (Xuereb et *al*. 2009), genotoxicity (Lacaze et *al*. 2011) and digestive impairments (Dedourge-Geffard et *al*. 2009). However and in regard to the lack of genomic data and accurate annotation in amphipods, no specific endocrine disruption biomarkers are available to assess and to diagnose the impact of these pollutants in aquatic systems. Recently, we conducted a large proteogenomic survey to discover the proteins involved in gammarid reproductive function (Trapp et *al*. 2014). We identified and annotated the most relevant proteins detected, through a massive shotgun tandem

mass spectrometry analysis carried out on the proteomes of the male and female reproductive systems, allowing to identify key proteins with strong sexual dimorphism i.e. 75 ovaries –specific proteins and 129 testis-specific proteins, and potential candidates in the reproductive process. Male reproductive physiology was investigated by performing protein expression profiles during spermatogenesis to pinpoint key proteins involved in this process (Trapp et *al*. 2014).

Here, in the same way as performed in male organisms, we aimed to gain new knowledge of G. fossarum female reproductive process i.e. oogenesis. This process is mainly characterized by vitellogenesis, the progressive accumulation of vitellogenin (Vtg) in oocytes. Vitellogenin is the precursor of major egg yolk protein, playing a key role in the reproduction of most oviparous species. This abundant, large molecular weight protein is cleaved during embryonic growth, supplying energy as well as building blocks to support the development of the new generation (Wu et *al*. 2013). In vertebrates, synthesis of Vtg is under the control of the estrogen pathway and induction in males has been widely employed as a biomarker of endocrine disruption in response to estrogen-mimetic exposure (Kime et *al.* 1999). Drastic evolutionary divergence in reproductive biology exists between vertebrates and arthropods. Arthropods hormonal machinery is not controlled by vertebrate-like steroids but by ecdysteroids and juvenoid hormones (LeBlanc 2007). Also, as suggested by molecular studies, invertebrate Vtgs are quite different from vertebrate Vtgs, in terms of both their sequences and their functional specialization as egg yolk protein precursors in the large lipid transfer protein (LLTP) superfamily, which also ensures clotting and growth functions in crustacean decapods (Avarre et *al.* 2007, Hayward et *al*. 2010). The LLTP family members, including the microsomal triglyceride transfer protein (MTP) family, the apolipoprotein-like (Apo-like) family and the Vtg-like protein family, originate from the duplication of a primitive protein, followed by functional diversification (Wu et *al*. 2013). Thus, regarding the intense evolutionary events among the LLTP superfamily, identification of proteins acting as Vtg in Gammarus fossarum cannot be achieved by automatic functional prediction by sequences similarities but rather by specific experimental observations and characterization. Indeed, in our previous experiments, we already evidenced the presence of proteins predicted as “vitellogenin” in testis and as highly abundant “clottable proteins” in ovaries, both from organisms in similar physiological conditions, suspecting erroneous functional prediction.

Based on the detailed physiological description of the complete female reproductive cycle – molting, follicle growth and embryonic development (Geffard et *al*. 2010), the study was designed: i) to uncover new molecular players in *Gammarus* female reproductive function and ii) to identify

the precursor of major egg yolk proteins in order to propose new and specific biomarkers of reproductive disorders. We performed a shotgun proteomic analysis on embryos and ovaries sampled at different stages of development. Protein dynamics over the different stages were analyzed to highlight key modulated proteins. Cross-comparison between proteins that accumulated during oogenesis, proteins with decreased amounts during embryogenesis, and those previously classified as female-specific, was performed to identify novel Vtg. A phylogenetic analysis of these Vtg shows that different LLTP subfamilies are involved in the yolk formation process. Our study evidenced the necessity to conduct specific functional studies based on empirical observations in non-model species and the power of shotgun proteomics to achieve such objective.

2. EXPERIMENTAL PROCEDURES

Amphipod collection and maintenance. The amphipods were collected upstream of the Bourbre River (south-eastern France) at La Tour du Pin (Lat: 45°569'442"; Long: 5°459'115"). Gammarids were collected by kick sampling using a net and were sieved (2–2.5 mm). The organisms were kept in 30-L tanks during an acclimatization period lasting at least 15 days. The temperature was kept at 12±1°C, with the photoperiod maintained at a 16/8-h light/dark cycle, and the tanks were continuously supplied with aerated, uncontaminated groundwater adjusted to the sampling site conductivity (i.e. 500 µS.cm-1). The gammarids were fed ad libitum on alder leaves (*Alnus glutinosa*), supplemented twice a week with freeze-dried worms (*Tubifex tubifex*).

Preparation of biological samples. Using phenotypic criteria (Karaman 1977), organisms determined as Gammarus fossarum were selected. At the beginning of each molt cycle, simultaneous oocyte maturation and embryo development occurs with embryos developing into the female ventral pouch (**Figure 1**). Using limb inter-tegmental change criteria (**Figure 2**), molt stages of female gammarids were classified into six different categories: post-molt stages A and B, inter-molt stages C1 and C2, and finally pre-molt stages D1 and D2 (Geffard et al. 2010). For each molt stage, a specific oocyte development stage is associated. As stage A lasts for 1 day, the two post-molt stages were grouped into one condition "AB", giving a total of five conditions. Accordingly, five different embryo development stages were delineated. For each female, six embryos were collected from the ventral pouch and the ovaries were excised as described by Lacaze et al. (Lacaze et al. 2010).

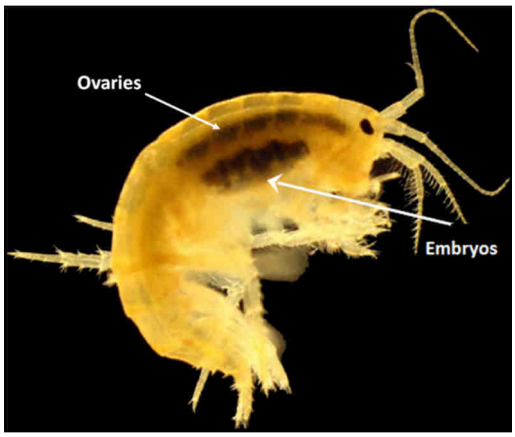


Figure 1: Female *G. fossarum* with developping embryos within its ventral pouch

For each stage, five biological replicates were performed, immediately frozen in liquid nitrogen and stored at −80°C until needed. Proteins from oocytes were directly dissolved in 40 µL LDS sample buffer (Invitrogen), while embryo samples were disrupted with the TissueRuptor instrument (Qiagen) in 180 µL LDS sample buffer. Samples were sonicated for 1 min. in a transonic 780H sonicator and boiled for 5 min. at 95°C. A 35-µL protein extract was resolved using SDS-PAGE with a short migration of 10 min. at 150 V on 4–12% gradient 10-well NuPAGE (Invitrogen) gels run with MES buffer (Invitrogen). Gels were stained with Coomassie Blue Safe stain (Invitrogen). After overnight destaining with water, the entire protein content from each well was extracted as a single polyacrylamide band and processed for further destaining and iodoacetamide treatment (Hartmann et al. 2014). The samples were proteolyzed with Trypsin Sequencing Grade (Roche) using 0.01% ProteaseMAX surfactant (Promega). The resulting peptide mixtures were diluted 1:20 in 0.1% trifluoroacetic acid and analyzed using nanoLC-MS/MS.

NanoLC-MS/MS analysis. NanoLC-MS/MS experiments were conducted on an LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings), essentially as previously described by Hartmann et al. (Hartmann and Armengaud 2014). Peptide samples (2 µL) were loaded and desalted online on a reverse-phase precolumn, C18 PepMap 100 column (LC Packings), and then resolved on a nanoscale C18 PepMapTM 100-capillary column (LC Packings) at a flow rate of 0.3 µL/min. with a gradient of CH3CN, 0.1% formic acid prior to injection into the ion trap mass spectrometer. Peptides were separated using a 90-min. gradient from 5 to 60% solvent B (0.1% HCOOH, 80% CH3CN). Solvent A was 0.1% HCOOH, 100% H2O. Full-scan mass spectra were measured from m/z 300 to 1,800 with the LTQ-Orbitrap XL mass spectrometer in data-dependent mode using the TOP3 strategy. In brief, a scan cycle was

initiated with a full scan of high mass accuracy in the Orbitrap followed by MS/MS scans in the linear ion trap on the three most abundant precursor ions.

MS/MS data assignment. Using the extract_msn.exe data import filter (Thermo), peak lists were generated with the Mascot Daemon software (version 2.3.2; Matrix Science). Data import filter options were set at 400 (minimum mass), 5,000 (maximum mass), 0 (grouping tolerance), 0 (intermediate scans) and 1,000 (threshold), as described previously by Christie-Oleza et al. (2012). MS/MS spectra were assigned to peptide sequences with the Mascot Daemon 2.3.2 search engine (Matrix Science) against our customized RNA-seq–derived database containing 1,311,444 sequences totaling 289,084,257 amino acids (Trapp et al. 2014). Searches for tryptic peptides were performed with the following parameters: full-trypsin specificity, a mass tolerance of 5 ppm on the parent ion and 0.5 Da on the MS/MS, static modifications of carboxyamidomethylated Cys (+57.0215), and dynamic modifications of oxidized Met (+15.9949). The maximum number of missed cleavages was set at 2. All peptide matches with a MASCOT peptide score below a p-value of 0.05 were filtered using the IRMa 1.30.4 software (Dupierris et al. 2009). A protein was considered valid when at least two different peptides were detected. The false-positive rate for protein identification was estimated by a search with a reverse decoy database to be below 0.1% using the same parameters.

Assessment of proteome dynamics. Spectral counts (number of spectra recorded per protein) were extracted from the spectra-to-peptide data set. The normalized spectral abundance factor (NSAF) for each protein was calculated (Christie-Oleza et al. 2012). To analyze the proteome dynamics along the different stages, proteins with the same trend of modulation along the cycle were clustered with the TrendQuest module of the PatternLab program (Fischer Jde et al. 2010). Spectral counts were normalized for missing values with a 0.5 increment. Proteins observed in at least three of the replicate samples were considered for clustering and the minimum signal required was 5. Cluster health was set at 0.9 and the minimum number of items per cluster was three.

Phylogenetic analysis. Egg yolk protein precursors unraveled by proteome dynamics were compared with other LLTP proteins known from diverse animal phyla. In the proteomic experiment, we employed an RNA-seq–derived database for MS/MS spectrum interpretation. De novo assembly of the transcriptome was carried out in absence of a genome backbone, thus leading to some inaccurate predictions of amino acid sequences. Indeed, we identified by

proteogenomics some assembling errors as evidenced by erroneous translation shift (data not shown). Consequently, using MS-sequenced peptides we reconstructed the Vtg and LLTP amino acid sequences as shown in Supplementary Data File S1. Based on the phylogenetic tree from Hayward et al. (2010), LLTP amino acid sequences (n=28) were retrieved from public databases (NCBI and JGI). All analyses were carried out using the multiplatform program SeaView version 4 (Gouy et al. 2010). Multiple sequence alignment was performed using the Clustal Omega program with default parameter values (Sievers et al. 2011). The N-terminal conserved amino acid sequence motifs of the LLT module were extracted and for high quality alignment the Gblocks program was applied to eliminate poorly aligned positions and divergent regions (Castresana 2000). Based on the LLT module alignment, the phylogenetic tree was constructed using the maximum likelihood (ML) method with the PhyML program, with default parameter values (Guindon et al. 2010). Branch support values were estimated using the approximate likelihood ratio test approach.

Data repository. The mass spectrometry proteomic data were deposited at the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al. 2013) with the data set identifiers PXD001002 and DOI 10.6019/PXD001002.

3. RESULTS

The protein catalogue of the Gammarus fossarum reproductive cycle.

Figure 2 shows the general strategy used to identify the precursors of major egg yolk proteins in G. fossarum. According to limb inter-tegumental change criteria, females with different molt stages were classified into five categories. From each female organism (n=5 per stage, i.e. 25 animals), the ovary pair and a pool of six embryos were sampled and their corresponding proteomes were extracted and analyzed separately. After electrophoresis under denaturing conditions, each of the 50 resulting samples was retrieved in a single polyacrylamide band. After in-gel trypsin digestion of these polyacrylamide bands, peptides were resolved on reverse phase chromatography and analyzed by tandem mass spectrometry with an Orbitrap mass spectrometer. This experiment gave an overall dataset of 273,526 MS/MS spectra. Interpretation of mass spectra data was made based on G. fossarum RNA-seq–derived database (GFOSS). The transcriptome was constructed from a pool of mRNAs extracted from G. fossarum, including the male and female reproductive tissues from sexually mature gammarids. After querying the GFOSS database, 1,884 unique

peptide sequences after querying a G. fossarum RNA-seq–derived database (GFOSS) and allowed the identification of 544 mRNA-translated products.

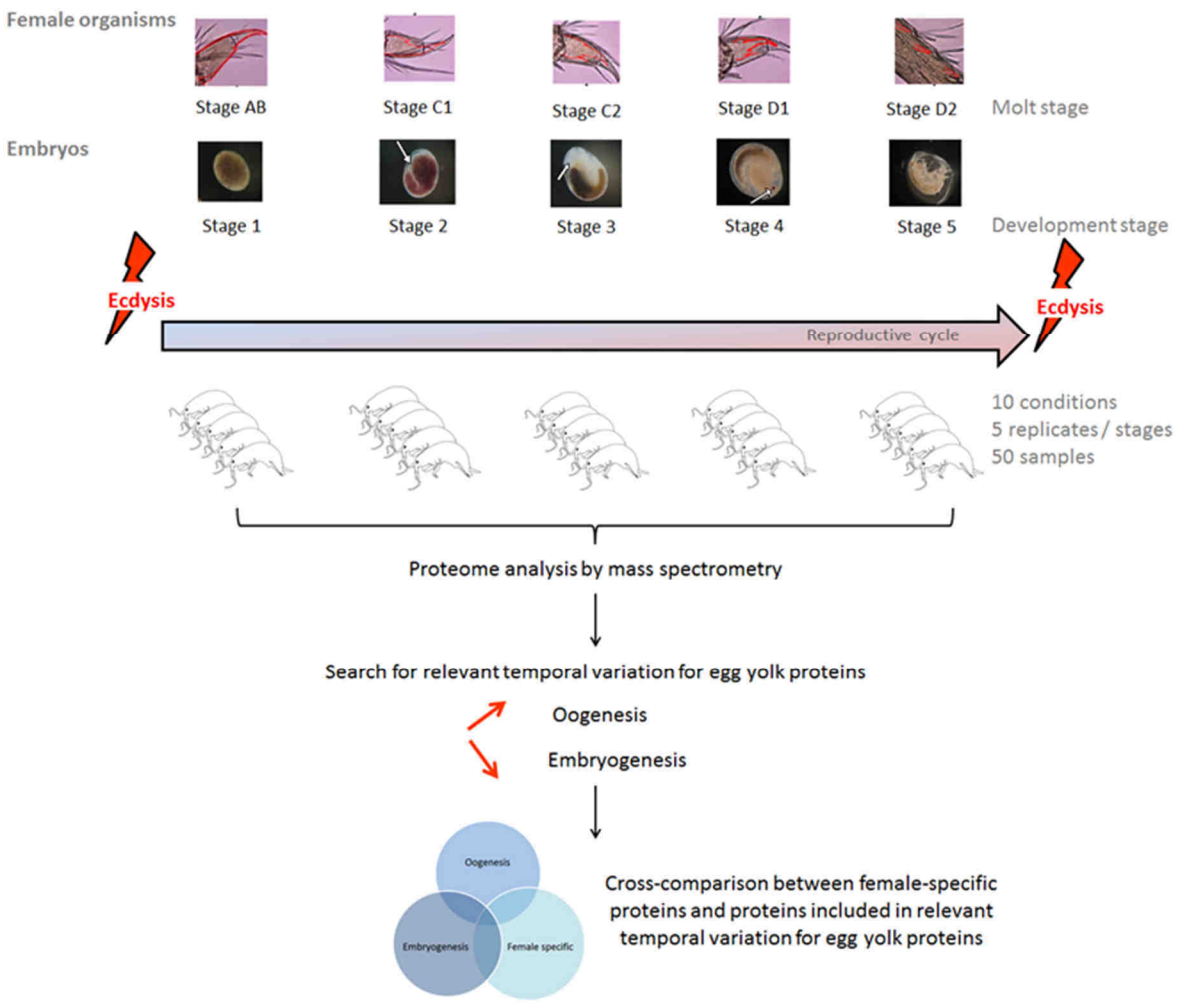


Figure 2: General strategy for the identification of major egg yolk proteins from *G. fossarum*.

Five female organisms were sampled at five different molt stages (Geffard et *al*. 2010). For an individual organism, the associated embryos (a pool of six) were collected from the ventral pouch and the ovaries were excised. After proteome analysis and clustering procedures to identify the temporal pattern during oogenesis and embryogenesis, relevant temporal patterns were established.

The ovary proteome. Amongst the 142,778 MS/MS spectra recorded for the ovary proteomes, 26,762 MS/MS spectra were assigned to peptide sequences (**Supplementary Data, Table S1**). These spectra validated a list of 1,466 peptides, which allowed the confident identification of 295 proteins (**Supplementary Data, Table S2**). Among these, 239 proteins (81%) show significant similarities to previously annotated proteins from other species while the other 56 proteins (19%) are lineage-specific. Remarkably, only two proteins, members of the large lipid transfer family annotated as “clotting protein precursor”, made up 10% of the total proteome in terms of

quantities assessed by spectral count: the protein (ID_39606; mean NSAF= 4.9±1.2 %) and the protein (ID_17046; mean NSAF= 5.9±1.4%).

Proteins identified in embryos. From the 130,748 MS/MS spectra recorded for the embryo samples, 23,144 MS/MS spectra were assigned (**Supplementary Data, Table S3**). These spectra validated a list of 2,149 peptide sequences, which allowed the identification of 425 proteins (**Supplementary Data, Table S4**). While 368 proteins (87%) show significant similarities to previously annotated proteins from other species, 57 proteins (13%) are orphans. Again, two proteins made up 10% of the total identified proteins in terms of quantity: the histone H2A (ID_13250; mean NSAF= 5.6±1.8 %) and the histone H4 (ID_100349; mean NSAF= 4.6±1.9%). In our previous proteogenomic survey, these two proteins proved to be abundant in male reproductive tissue, accounting for 7% of the testis proteome in terms of spectral count.

Proteome dynamics throughout the reproductive cycle.

To pinpoint key proteins involved in oogenesis and embryogenesis, the established protein catalogues from ovaries and embryos, respectively, were subjected to a clustering procedure using the PatternLab Trend Quest module (Carvalho *et al.* 2008), which groups similar protein temporal profiles.

Expression profiles during oogenesis. A total of 27 proteins (7% of the total proteome quantity) presented significant temporal variations during oogenesis (**Supplemental Data Table S2**). Among these proteins, 16 (59%) are members of the LLTP family classified under generic terms such as Vtg, clottable protein or apolipocrustaceins. Remarkably, ten proteins (37%) are devoid of functional annotation, including eight orphans. We identified three distinct profile trends, which are presented in **Figure 3.** The characteristics of the proteins belonging to these three groups are listed in **Supplemental Data Table S2.** Trend A comprises three proteins whose presence first increased from stage AB to stage C1, further increased from stage C2 to stage D1, and then dropped to the initial level at D2. Trend B is the dominant trend, comprising 16 proteins. This pattern shows proteins continuously accumulating from stage AB until reaching stage D1. Finally, the eight proteins included in trend C present a pattern of a substantial increase at the beginning of the cycle (transition from stage AB to stage C1), remaining at a generally steady level afterwards (**Figure 3**).

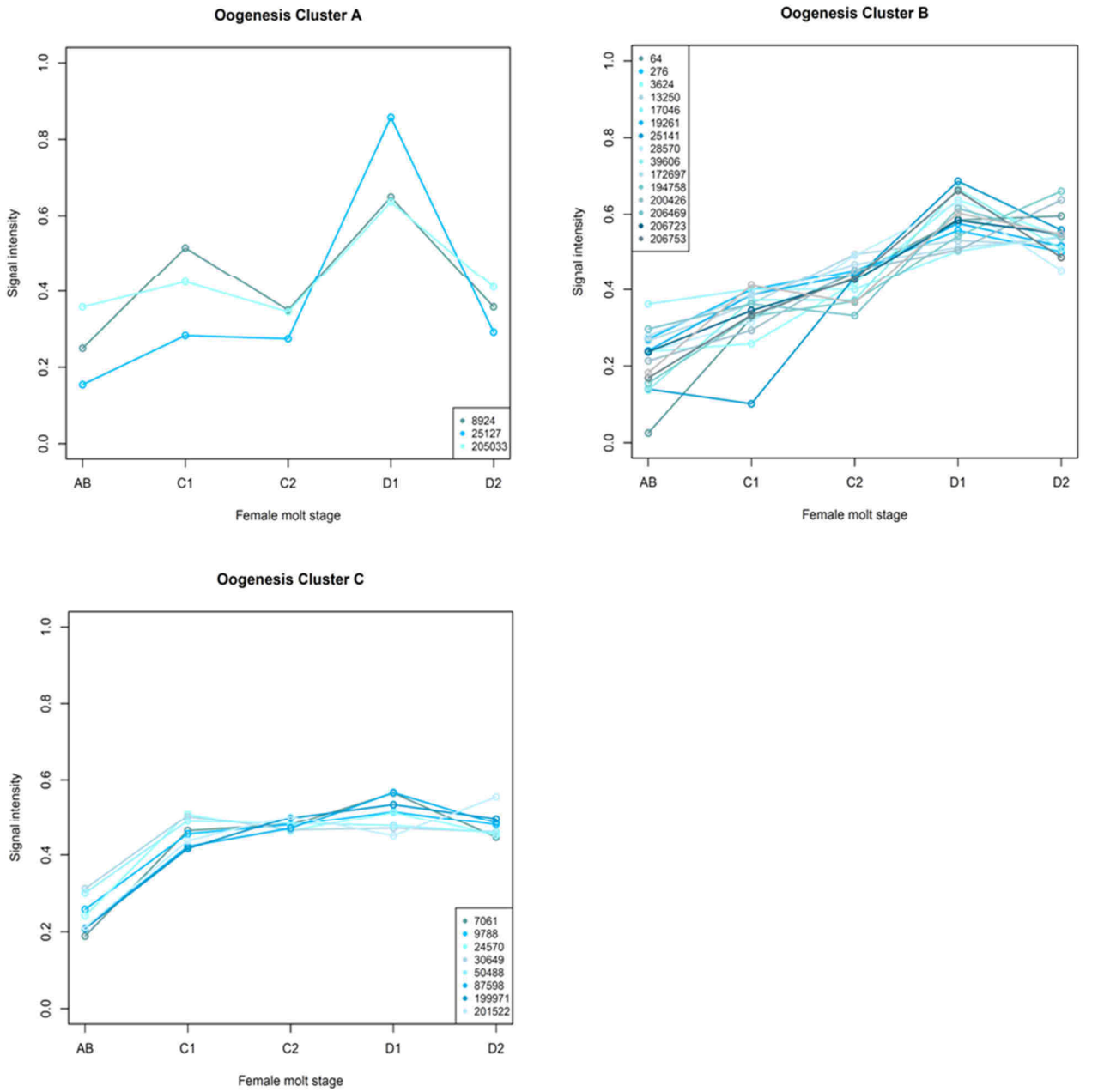


Figure 3: Clustered temporal abundance variations of proteins during oogenesis.

The different trends were obtained from clustering the spectral count using the TrendQuest module of the PatternLab program. Cluster health was set at 0.9 and the minimum number of items per cluster at three.

Expression profiles during embryogenesis. The trend quest clustering analysis of protein profiles recorded during embryogenesis revealed four distinct trends, shown in **Figure 4.** The proteins belonging to each group are listed in **Supplemental Data Table S4.** A total of 27 proteins (5% of the identified proteins) presented significant temporal variations. Among these proteins, 12 (44%) are members of the LLTP family and five proteins (19%) are devoid of functional annotation, including three orphans. The other proteins are involved in housekeeping functions, especially cytoskeletal structure maintenance.

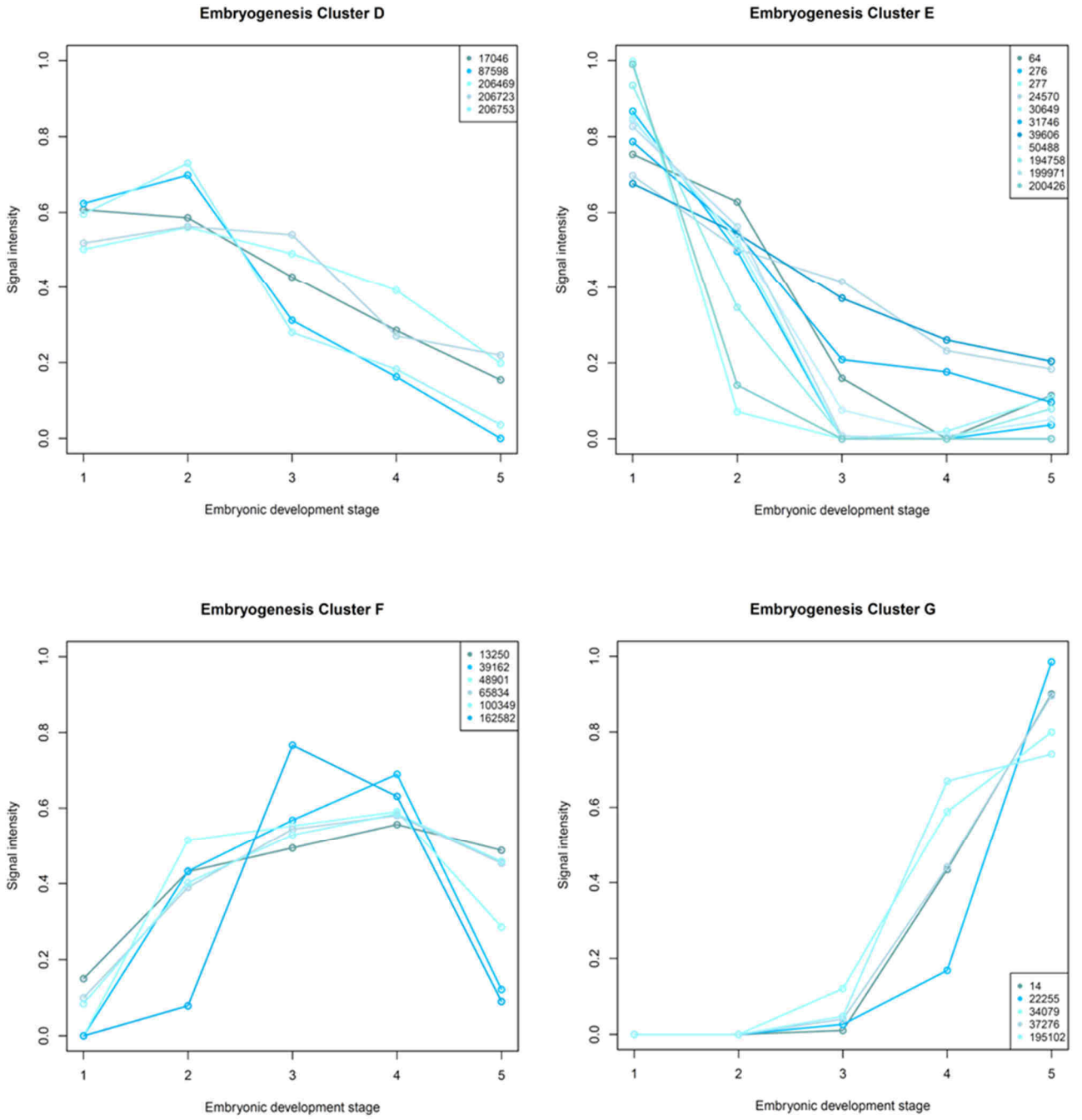


Figure 4: Clustered temporal abundance variations of proteins during embryogenesis.

The different trends were obtained from clustering the spectral count using the TrendQuest module of the PatternLab program. Cluster health was set at 0.9 and the minimum number of items per cluster at three.

Trend D comprises five proteins whose general abundance remained at a steady level until stage 2 and then continuously decreased throughout embryo development. Trend E is the dominant trend with 11 proteins. This pattern showed proteins whose presence continuously decreased over the cycle, with a very rapid rate of consumption at the early stages of development; after stage 3, some proteins were no longer detected (thus entirely consumed) while the presence of others decreased more slowly. Among proteins displaying clustered temporal changes, the whole set of

proteins with functional prediction related to the LLTP family, such as the orphans, are grouped in Trends D and E. The other two trends comprise proteins with general housekeeping functions. Trend F groups six proteins whose abundance continuously increased until reaching stage 3 or stage 4 and then decreased. Finally, the five proteins included in trend G present a pattern of a substantial increase after stage 3.

Deciphering the precursors of major egg yolk proteins.

Identification of relevant temporal patterns. Accumulation of Vtg takes place during secondary vitellogenesis in *G. fossarum*. This phase mainly occurs at the C2 molt stage with a net increase in the follicular area, yolk vesicles and lipid globules (Geffard *et al.* 2010). Consequently, the 16 proteins included in pattern B, whose expression continuously increased until reaching stage D1, exhibited a pattern similar to the one expected. These 16 proteins are therefore possible candidates for egg yolk protein precursors. In contrast, egg yolk proteins are consumed during embryogenesis, providing resources for organogenesis and growth. Therefore, a pattern where proteins decreased during embryogenesis should indicate proteins possibly involved in this process. Trend D and trend E behaved as expected in this case, leading to a selection of 16 proteins. **Figure 5** shows the cross-comparison between the 16 proteins from the oogenesis experiment (**Condition A**) and the 16 proteins from the embryogenesis experiment (**Condition B**), giving a list of 11 proteins.

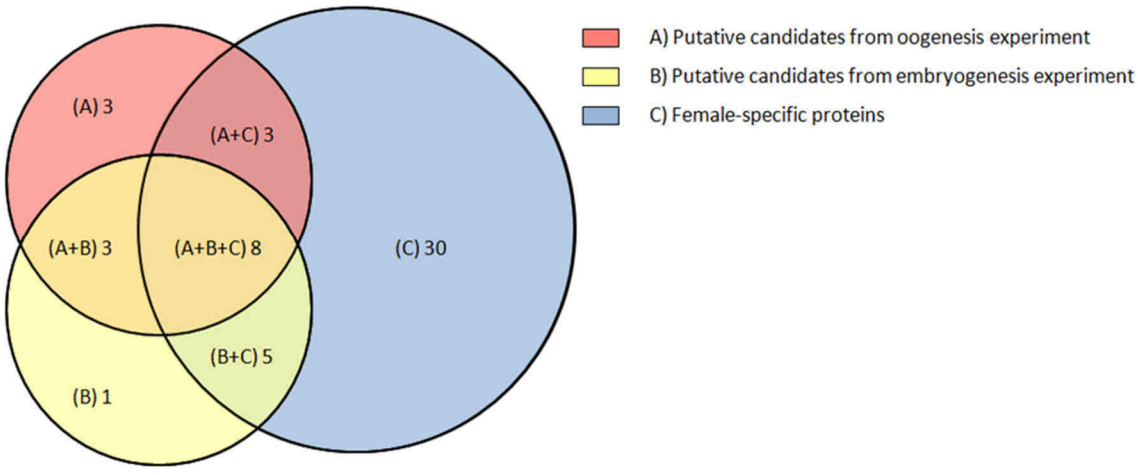


Figure 5: Identification of major egg yolk proteins precursors in *G. fossarum*.

A cross-comparison was performed between A) the 16 putative candidates included in the accumulation pattern from the oogenesis experiment; B) the 16 putative candidates included in the consummation pattern from the embryogenesis experiment and C) the 46 mRNA products previously classified as female-specific and present in this study.

Specificity towards female reproductive function. In a previous study (Trapp *et al.* 2014), we conducted a proteomic comparative study using male and female reproductive tissues to mine putative candidates involved in the reproductive process. A list of 75 mRNA-translated products were classified as specific for the female reproductive tissue. In the present study, 46 of these ovaries specific proteins were identified, as shown in **Figure 5 (Condition C)**. In addition, **Figure 5** displays the final cross-comparison between the female-specific proteins present in this study and those included in the temporal variation relevant to the functional role of egg yolk precursor, *i.e.* trends B, D and E. **Table 1** presents the eight mRNA-translated products that undoubtedly are *G. fossarum*-specific precursors of major egg yolk proteins, all members of the LLTP family. The abundance levels of these precursors range from a total of 2,975 to 489 MS/MS spectra recorded in the ovaries (ratio, 6:1) and from a total of 1,966 to 162 MS/MS spectra recorded in the embryos (ratio, 12:1). These proteins account for 21% and 12% of the total ovary and embryo proteomes, respectively. In both tissues, the two most abundant mRNA-translated products in terms of spectral count and NSAF are ID 17046 and ID 39606, both annotated as clottable proteins *via* decapod blast queries.

Table 1: List of precursors of major egg yolk protein in *Gammarus fossarum*.

			Oogenesis		Embryogenesis	
Contig ID	**Predicted function**	**Origin of the closest protein**	Total number of MS/MS spectra recorded	NSAF %	Total number of MS/MS spectra recorded	NSAF %
17046	Clottable protein 2	*Penaeus monodon*	2975	4.9	1966	3.4
206469	Clotting protein precursor	*Pacifastacus leniusculus*	1023	3.0	934	2.9
39606	Clotting protein precursor	*Pacifastacus leniusculus*	1662	5.9	1095	3.9
276	Hemolymph clottable protein	*Litopenaeus vannamei*	873	1.4	282	1.1
194758	Apolipocrustacein	*Penaeus semisulcatus*	859	1.3	439	0.7
277	Vitellogenin	*Pandalopsis japonica*	741	2.5	267	0.5
200426	Vitellogenin	*Metapenaeus ensis*	598	1.6	162	0.5
64	Vitellogenin	*Nasonia vitripennis*	489	0.9	189	0.4

Phylogenetic analysis. In order to investigate how the different LLTP families are involved in the yolk formation process in *G. fossarum*, the eight major egg yolk proteins precursors shown in **Table 1** were compared with LLTP gene products from diverse animal phyla, chosen on the basis of the phylogenetic analysis of Hayward et *al*. (Hayward *et al.* 2010). After Gblock analysis, the resulting LLT module alignment resulted in 276 amino acids (**Supplementary Data File S2)**. Among the eight proteins, four of these could be integrated into the alignment: ID 64, 17 046, 194 758 and 206 469. **Figure 6** shows the phylogenetic relationships among the LLTP superfamily. The proteins ID 17 046, 206 469 and 64 are true Vtg orthologs, branched with decapod clottable proteins, while ID 194 758 is included in the Apo family, a paralogous group from the Vtg family including precursors of egg yolk proteins identified in decapods.

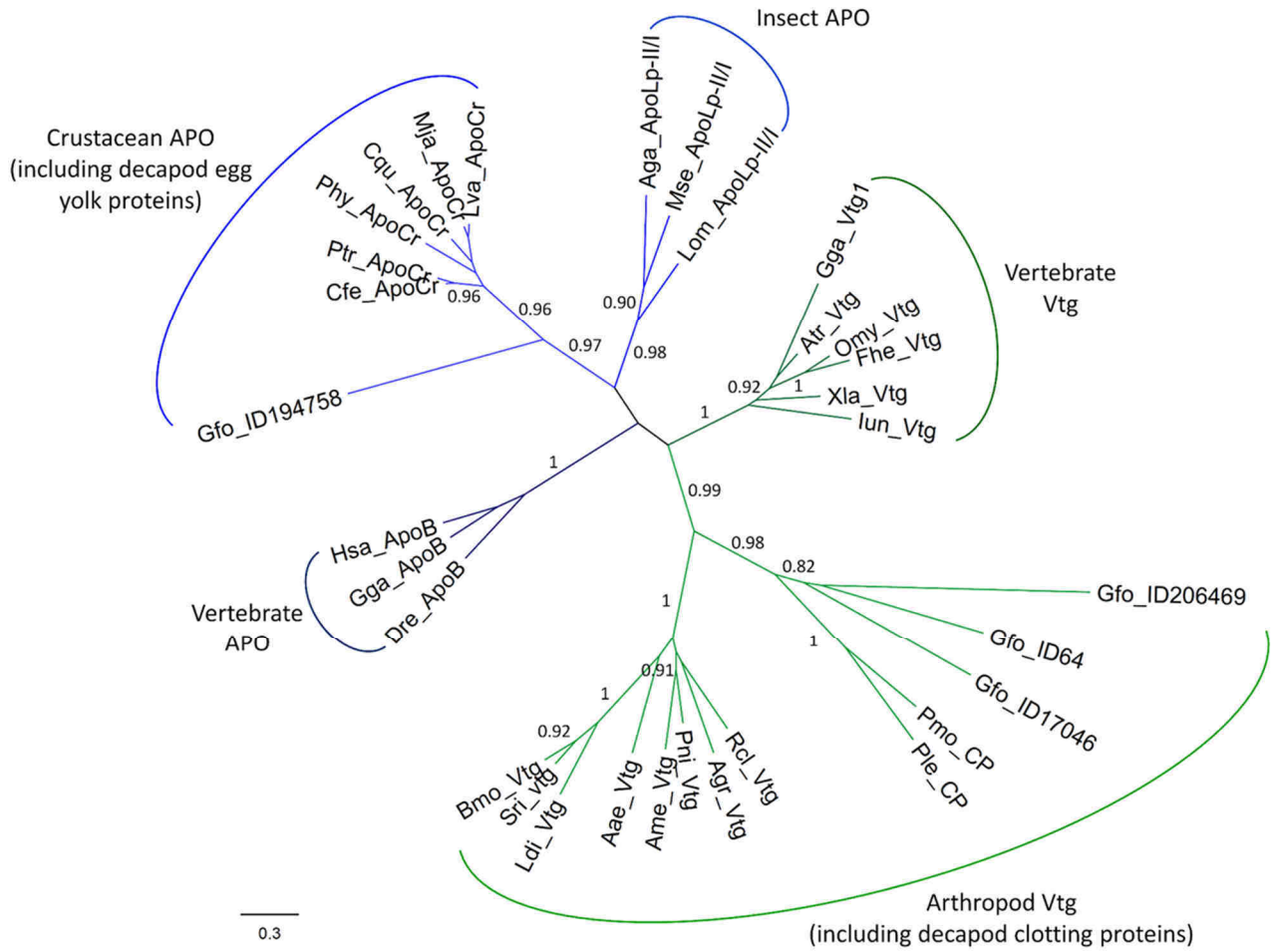


Figure 6: Phylogenetic tree showing the relationship in the LLTP family.

Amino acid sequences of other organisms were chosen according to the study of Hayward et al. (Hayward et al. 2010). Alignment was performed on 226 amino acids on the N-terminal conserved amino acid sequence motifs of the LLT module. Maximum likelihood (ML) tree reconstruction was performed using the PhyML program with default parameter values (Guindon et al. 2010). Branch support values were estimated using the approximate likelihood ratio test and only values above 0.8 are displayed.

4. DISCUSSION

The MS/MS spectrum data were interpreted based on our in-house ORF database constructed from RNA-seq data (Trapp *et al.* 2014). The transcriptome was established from the pool of mRNAs extracted from the male and female reproductive tissues, the hepatopancreatic *caeca* and the cephalons from sexually mature gammarids. Regarding the female reproductive process, in our previous comparative analysis, 40% of the female-specific proteins were classified as orphan, indicating that in amphipods, major egg yolk protein precursors may not be related to the LLTP family. In this study, the proteins finally classified as precursors of egg yolk protein are related members of this family, like the majority of oviparous organisms. Therefore, we can speculate that orphan proteins are not part of a key conserved function but rather are subject to divergent selection as molecular players in reproductive isolation or gamete recognition (Marshall *et al.* 2011).

Gammarus fossarum is an example of an amphipod whose molting and reproductive cycles occur concurrently in sexually mature females. The reproductive cycle is characterized by the maturation of oocytes with the development of the follicular surface around the nucleus and the oocyte diameter. This process allows yolk formation via the uptake of extra-ovarian proteins, nutrients and other maternal factors as a starting point for a new generation. For the oogenesis proteome, the pattern of protein accumulation is expected as candidates for yolk formation. Nevertheless, in our experiment, two distinct patterns of accumulation are documented (Figure 3, trends B and C). Notwithstanding, the process of oogenesis maturation can be further subdivided into two phases: primary and secondary vitellogenesis. In stage AB, oocytes are not enveloped by follicle cells and their cytoplasm shows few lipid globules and yolk vesicles. This step of endogenous vesicle development is related to primary vitellogenesis. In contrast, development of exogenous yolk vesicles occurs during secondary vitellogenesis. This process of egg yolk formation is particularly intensive at the C2 stage (Geffard *et al.* 2010). Consequently, proteins involved in egg yolk formation are included in trend B, exhibiting a continuous accumulation over the cycle, whereas in trend C, accumulation reaches a steady stage after stage C1, the profile being relevant with primary vitellogenesis and endogenous uptake of yolk vesicles.

Like the proteins included in trends B and C, the proteins included in trend A are also predominantly members of the LLTP family. Yet their temporal pattern is inconsistent with a storage function, with peak expression at C1 and D1. By their lipid-binding properties, LLTP members play a central role in lipid circulation, and one hypothesis is that these proteins could act as carriers of lipidic molecules into the ovaries for i) coordination of crucial physiological events in

oogenesis or ii) *de novo* synthesis of hormones. Arguments in favor of their role in oogenesis coordination are the occurrence of peak expression at C1 when oocytes enter into secondary vitellogenesis, and a sharper peak at D1, just before spawning of mature oocytes. Interestingly, in the freshwater prawn *Macrobrachium nipponense*, a correlation between circulating ecdysteroids, the arthropod molting hormone in the hemolymph, and ovarian maturation has been described (Okumura 1992). Indeed, a small ecdysteroid peak was recorded concurrently with the initiation of Vtg synthesis and a stronger and sharper peak was recorded at the D2 stage, prior to ecdysis and spawning. After biogenesis into the Y-organ, these lipidic hormones are released into the hemolymph and may form conjugate Vtg-like proteins (Subramoniam 2000). We hypothesize that some of the proteins included in trend A may carry ecdysteroids from the Y-organ to the ovaries, where they are released into the active form (20-hydroxyecysone) for the coordination of crucial physiological events related to oogenesis. In the crab *Uca pugilator*, expression of ecdysteroid receptors (EcR) in the ovary during ovarian maturation has been recorded, suggesting that this tissue is a potential target tissue for ecdysteroid action (Durica *et al.* 2002). This observation supports well our first hypothesis. Alternatively we hypothesize that these proteins act as carriers of the ecdysteroid precursor, cholesterol, for endogenous synthesis during secondary vitellogenesis and further transfer to the embryos as a maternal factor. Several crustacean species have been proven to accumulate large quantities of ecdysteroids within the ovary and these molecules, under free and conjugate forms, are passed to egg embryogenesis (reviewed by Subramoniam 2000). While the general hypothesis of the presence of these steroids in the ovaries is their sequestration from the hemolymph, studies on the *Carcinus maenas* crab suggest that the ovaries are capable of their *de novo* synthesis (Styrishave *et al.* 2008). Thus, cholesterol assimilated in the hepatopancreatic *caeca* from food sources may form conjugates with LLTP synthesized in this digestive organ (Subramoniam 2011) and then carried to the female reproductive tissue.

Regarding the identification of the egg yolk protein precursors in embryo proteomes, candidate proteins are clustered into trends D and E. Because of the pleiotropic function of LLTP members, specificity towards the female reproductive tissue was challenged. Indeed, some proteins within the trends are involved in the yolk formation but they can be found in other tissues such as the cephalon or the testis, evidencing that they play pleiotropic functions. A total of eight proteins, all members of the LLTP family, were classified as major egg yolk protein precursors and can be referred to as “functional Vtg” of *G. fossarum*. Evidence of multiple copies of Vtg genes has already been reported: four copies in the mosquito *Culex tarsalis* (Chen *et al.* 2010) discovered

through targeted PCR amplification and five contigs in the amphipod *Melita plumulosa* observed by RNA-Seq (Hook *et al.* 2014). However, these observations were based on automatic functional prediction, while our massive shotgun proteomic analysis allows specific identification, revealing an unsuspected diversity of Vtg. Indeed, as Vtg is essential for embryonic development, increased quantities of egg yolk may be beneficial to embryo survival and high Vtg diversity may be advantageous for reproduction of the organism (Chen *et al.* 2010). Based on the few omics data available on crustacean biodiversity, gene family expansion appears to be a common feature in this group (Colbourne *et al.* 2011, Zeng *et al.* 2011). Furthermore, the LLTP evolutionary history is characterized by the duplication of a primitive yolk protein, followed by functional diversification (Wu *et al.* 2013).

Analysis of phylogenetic relationships of *G. fossarum* functional Vtg showed that some forms are true Vtg orthologs branched with decapod clottable proteins, while others are included in the Apo family, a paralogous group from the Vtg family, which is branched with decapod apolipoproteins. The involvement of two different protein groups for yolk formation was also demonstrated in crustaceans. Interestingly, while in crustaceans phylogenetic analysis showed that major egg yolk precursor protein is a true Vtg ortholog in the branchiopod *Daphnia magna* and an Apo-like protein in decapods (Avarre *et al.* 2007), functional yolk proteins of the species assayed here originate from both families. Thus, amongst crustaceans, *Gammarus fossarum* is the first documented intermediary case where diverse LLTP proteins are involved in the yolk formation process.

Here, our study has clearly illustrated how direct functional assignment on the basis of sequence similarity can be misleading and how dangerous straightforward monitoring of the identified gene product as a biomarker is. Indeed, the automatic pipeline for functional prediction is based on the comparison of sequence and functional information from various organism groups, most particularly vertebrates in the animal world. Consequently, for invertebrates, which are organisms of prime interest in ecotoxicology, automatic functional annotation can lead to inaccurate predictions due to quite different evolutionary scenarios and false mechanistic hypotheses. Therefore, for any new model organism, confirmation in terms of organism-specific function should be mandatory before it is considered for ecotoxicological applications. Nikinmaa and Rytkönen (Nikinmaa and Rytkonen 2011) even consider that “combining genomic and functional aspects … could make aquatic toxicologists forerunners in functional genomics as a whole, not just its toxicological applications”. In our previous attempt to measure Vtg in *G. fossarum*, the

sequence employed was indeed selected after functional annotation by gene ontology (Xuereb *et al.* 2011) but before being directly applied as a biomarker of EDC exposure, functional validation turned out to argue in favor of involvement in the vitellogenesis process, with significant inter-sex differences and consistent evolution through the female reproductive cycle (Xuereb *et al.* 2011, Jubeaux *et al.* 2012). In the present study, two proteolytic peptides presenting the sequence used for the LC-MS/MS quantification method (Simon *et al.* 2010) were sequenced and attributed to two mRNA-translated products that proved to be abundant but were not included in the relevant clustered temporal pattern. While previous targeted analysis on a specific protein strongly supported its involvement in vitellogenesis, this massive shotgun proteomics experiment allowed the identification of more prevalent players in this process.

This first large-scale molecular description of a crustacean reproductive cycle in an ecotoxicologically relevant species allowed us to discover major proteins involved in gammarid reproduction and should result in further biomarker development of endocrine disruptor exposure for crustaceans. In addition, the identification of the true *G. fossarum* egg yolk protein precursors challenges the emerging consensus on the unsuitability of Vtg as a biomarker of endocrine exposure or feminization in crustaceans (Jubeaux *et al.* 2012, Short *et al.* 2014). In daphnids, the Vtg gene structure has been studied, demonstrating the presence of i) a superoxide dismutase-like domain at the NH_2-terminus, with antioxidant activity confirmed in worker bees (Seehuus *et al.* 2006) and ii) a *cis*-regulatory element responsive to the two arthropod hormones, *i.e*. ecdysteroids and juvenoid hormone, in the intergenic region (Kato *et al.* 2004, Tokishita *et al.* 2006). With regard to the diversity of LLTP proteins demonstrated in this study, modulation of specific forms towards the xenobiotic mode of action, *i.e*. i) antioxidant activity in response to general stress, or ii) specific induction when exposed to arthropod hormonal mimetics, is plausible. Future proteomic experiments using gammarids exposed to xenobiotics must be conducted to clearly determine if some forms are specifically modulated by arthropod EDCs or if Vtg also plays a versatile biological function and can thus be proposed as a specific biomarker of endocrine disruption.

The aim of this study was to gain better knowledge of gammarid reproduction by proteome dynamics in order to propose specific biomarkers of endocrine disruption. The study provides the first large-scale molecular description of the female gonads in crustaceans throughout their reproductive cycle. Protein members of the large lipid transfer family are key molecular players in the oogenesis process and fulfill diverse biological roles, a possible source of confounding factors

in ecotoxicological studies monitoring Vtg alterations. These results allow the identification of *G. fossarum*-specific major egg yolk proteins precursors, revealing an unsuspected diversity. Due to functional shift affecting LLTP in crustacean decapods, these proteins were functionally annotated as “clottable protein”, showing that data from omics studies must be further documented in terms of organism-specific function before being considered from an ecotoxicological perspective. Future research is necessary to study their modulation in the presence of toxic stressors, in order to clearly demonstrate whether or not Vtg is an appropriate biomarker of endocrine disruption in crustaceans.

Abbreviations: Apo, apolipoproteins; EDC, endocrine-disrupting chemical; HSP, heat shock protein; LLTP, large lipid transfer protein; MTP, microsomal triglyceride transfer protein; NSAF, normalized spectral abundance factor; RT-PCR, real-time reverse transcription polymerase chain reaction; Vtg, vitellogenin.

Acknowledgements: We thank the Institut National de Recherche en Sciences et Technologies pour l’Environnement et l’Agriculture (France), the Commissariat de l’Energie Atomique et aux Energies Alternatives (France) through the transversal toxicology program (PPTOX), and the ANR program “ProteoGam” (ANR-14-CE21-0006-02) for financial support. We also thank PRIDE Team for their technical assistance, as well as P.C. Carvalho (COPPE, Rio de Janeiro, Brazil) for adapting on-demand the export option of the PatternLab program.

REFERENCES

Armengaud, J., J. Trapp, O. Pible, O. Geffard, A. Chaumot and E. M. Hartmann (2014). "Non-model organisms, a species endangered by proteogenomics." J Proteomics **105**(0): 5-18.

Avarre, J. C., E. Lubzens and P. J. Babin (2007). "Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate apolipoprotein B." BMC Evol Biol **7**: 3.

Carvalho, P. C., J. S. Fischer, E. I. Chen, J. R. Yates, 3rd and V. C. Barbosa (2008). "PatternLab for proteomics: a tool for differential shotgun proteomics." BMC Bioinformatics **9**: 316.

Castresana, J. (2000). "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis." Mol Biol Evol **17**(4): 540-552.

Chen, S., J. S. Armistead, K. N. Provost-Javier, J. M. Sakamoto and J. L. Rasgon (2010). "Duplication, concerted evolution and purifying selection drive the evolution of mosquito vitellogenin genes." BMC Evol Biol **10**(1): 142.

Christie-Oleza, J. A., G. Miotello and J. Armengaud (2012). "High-throughput proteogenomics of Ruegeria pomeroyi: seeding a better genomic annotation for the whole marine Roseobacter clade." BMC Genomics **13**: 73.

Christie-Oleza, J. A., J. M. Pina-Villalonga, R. Bosch, B. Nogales and J. Armengaud (2012). "Comparative proteogenomics of twelve Roseobacter exoproteomes reveals different adaptive strategies among these marine bacteria." Mol Cell Proteomics **11**(2): M111 013110.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Caceres, L. Carmel, C. Casola, J. H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Frohlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V. Koonin, E. M. Zdobnov, I. V. Grigoriev, M. Lynch and J. L. Boore (2011). "The ecoresponsive genome of Daphnia pulex." Science **331**(6017): 555-561.

Coulaud, R., O. Geffard, B. Xuereb, E. Lacaze, H. Queau, J. Garric, S. Charles and A. Chaumot (2011). "In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring." Water Res **45**(19): 6417-6429.

Dedourge-Geffard, O., F. Palais, S. Biagianti-Risbourg, O. Geffard and A. Geffard (2009). "Effects of metals on feeding rate and digestive enzymes in Gammarus fossarum: An in situ experiment." Chemosphere **77**(11): 1569-1576.

Dupierris, V., C. Masselon, M. Court, S. Kieffer-Jaquinod and C. Bruley (2009). "A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa." Bioinformatics **25**(15): 1980-1981.

Durica, D. S., X. Wu, G. Anilkumar, P. M. Hopkins and A. C. Chung (2002). "Characterization of crab EcR and RXR homologs and expression during limb regeneration and oocyte maturation." Mol Cell Endocrinol **189**(1-2): 59-76.

Felten, V., G. Charmantier, R. Mons, A. Geffard, P. Rousselle, M. Coquery, J. Garric and O. Geffard (2008). "Physiological and behavioural responses of Gammarus pulex (Crustacea: Amphipoda) exposed to cadmium." Aquat Toxicol **86**(3): 413-425.

Fischer Jde, S., L. Liao, P. C. Carvalho, V. C. Barbosa, G. B. Domont, G. Carvalho Mda and J. R. Yates, 3rd (2010). "Dynamic proteomic overview of glioblastoma cells (A172) exposed to perillyl alcohol." J Proteomics **73**(5): 1018-1027.

Geffard, O., B. Xuereb, A. Chaumot, A. Geffard, S. Biagianti, C. Noel, K. Abbaci, J. Garric, G. Charmantier and M. Charmantier-Daures (2010). "Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment." Environ Toxicol Chem **29**(10): 2249-2259.

Gouy, M., S. Guindon and O. Gascuel (2010). "Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building." Mol. Biol. Evol. **27**(2): 221-224.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel (2010). "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." Syst Biol **59**(3): 307-321.

Hartmann, E. M., F. Allain, J. C. Gaillard, O. Pible and J. Armengaud (2014). "Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria." Methods Mol Biol **1197**: 275-285.

Hartmann, E. M. and J. Armengaud (2014). "Shotgun proteomics suggests involvement of additional enzymes in dioxin degradation by Sphingomonas wittichiiRW1." Environ Microbiol **16**(1): 162-176.

Hayward, A., T. Takahashi, W. G. Bendena, S. S. Tobe and J. H. Hui (2010). "Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans." FEBS Lett **584**(6): 1273-1278.

Hook, S. E., N. A. Twine, S. L. Simpson, D. A. Spadaro, P. Moncuquet and M. R. Wilkins (2014). "454 pyrosequencing-based analysis of gene expression profiles in the amphipod Melita plumulosa: transcriptome assembly and toxicant induced changes." Aquat Toxicol **153**(0): 73-88.

Jubeaux, G., R. Simon, A. Salvador, C. Lopes, E. Lacaze, H. Queau, A. Chaumot and O. Geffard (2012). "Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: inconclusive experience." Aquat Toxicol **122-123**: 9-18.

Jubeaux, G., R. Simon, A. Salvador, H. Queau, A. Chaumot and O. Geffard (2012). "Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males." Aquat Toxicol **112-113**(0): 72-82.

Karaman, G. S. (1977). "Freshwater Gammarus species from Europe, North Africa and adjacent regions of Asia (Crustacea-Amphipoda). Part I. Gammarus pulex-group and related species " Bijdragen Tot de Dierkunde **47**.
Kato, Y., S. Tokishita, T. Ohta and H. Yamagata (2004). "A vitellogenin chain containing a superoxide dismutase-like domain is the major component of yolk proteins in cladoceran crustacean Daphnia magna." Gene **334**(0): 157-165.
Kime, D. E., J. P. Nash and A. P. Scott (1999). "Vitellogenesis as a biomarker of reproductive disruption by xenobiotics." Aquaculture **177**(1-4): 345-352.
Lacaze, E., A. Devaux, R. Mons, S. Bony, J. Garric, A. Geffard and O. Geffard (2011). "DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment." Environ Pollut **159**(6): 1682-1691.
Lacaze, E., O. Geffard, S. Bony and A. Devaux (2010). "Genotoxicity assessment in the amphipod Gammarus fossarum by use of the alkaline Comet assay." Mutat Res-Gen Tox En **700**(1–2): 32-38.
LeBlanc, G. A. (2007). "Crustacean endocrine toxicology: a review." Ecotoxicology **16**(1): 61-81.
Marshall, J. L., D. L. Huestis, C. Garcia, Y. Hiromasa, S. Wheeler, S. Noh, J. M. Tomich and D. J. Howard (2011). "Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes." Mol Biol Evol **28**(1): 423-435.
Nikinmaa, M. and K. T. Rytkonen (2011). "Functional genomics in aquatic toxicology-do not forget the function." Aquat Toxicol **105**(3-4 Suppl): 16-24.
Okumura, T. (1992). "Changes in hemolymph vitellogenin and ecdysteroid levels during the reproductive and nonreproductive molt cycles in the freshwater prawn Macrobrachium nipponense." Zool. Sci. **9**: 37-45.
Seehuus, S. C., K. Norberg, U. Gimsa, T. Krekling and G. V. Amdam (2006). "Reproductive protein protects functionally sterile honey bee workers from oxidative stress." Proc Natl Acad Sci U S A **103**(4): 962-967.
Short, S., G. Yang, P. Kille and A. T. Ford (2014). "Vitellogenin is not an appropriate biomarker of feminisation in a crustacean." Aquat Toxicol **153**(0): 89-97.
Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**: 539.
Simon, R., G. Jubeaux, A. Chaumot, J. Lemoine, O. Geffard and A. Salvador (2010). "Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum)." J Chromatogr A **1217**(31): 5109-5115.
Studer, R. A. and M. Robinson-Rechavi (2009). "How confident can we be that orthologs are similar, but paralogs differ?" Trends Genet **25**(5): 210-216.
Styrishave, B., T. Lund and O. Andersen (2008). "Ecdysteroids in female shore crabs Carcinus maenas during the moulting cycle and oocyte development." J Mar Biol Assoc UK **88**(3): 575-581.
Subramoniam, T. (2000). "Crustacean ecdysteriods in reproduction and embryogenesis." Comp Biochem Physiol C Toxicol Pharmacol **125**(2): 135-156.
Subramoniam, T. (2011). "Mechanisms and control of vitellogenesis in crustaceans." Fisheries Science **77**(1): 1-21.
Tokishita, S., Y. Kato, T. Kobayashi, S. Nakamura, T. Ohta and H. Yamagata (2006). "Organization and repression by juvenile hormone of a vitellogenin gene cluster in the crustacean, Daphnia magna." Biochem Biophys Res Commun **345**(1): 362-370.
Trapp, J., J. Armengaud, A. Salvador, A. Chaumot and O. Geffard (2014). "Next-generation proteomics: toward customized biomarkers for environmental biomonitoring." Environ Sci Technol **48**(23): 13560-13572.
Trapp, J., O. Geffard, G. Imbert, J. C. Gaillard, A. H. Davin, A. Chaumot and J. Armengaud (2014). "Proteogenomics of Gammarus fossarum to document the reproductive system of amphipods." Mol Cell Proteomics.
Vizcaino, J. A., R. G. Cote, A. Csordas, J. A. Dianes, A. Fabregat, J. M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, G. O'Kelly, A. Schoenegger, D. Ovelleiro, Y. Perez-Riverol, F. Reisinger, D. Rios, R. Wang and H. Hermjakob (2013). "The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013." Nucleic Acids Res **41**(Database issue): D1063-1069.

Wang, X., R. J. C. Slebos, D. Wang, P. J. Halvey, D. L. Tabb, D. C. Liebler and B. Zhang (2012). "Protein identification using customized protein sequence databases derived from RNA-seq data." Journal of Proteome Research **11**(2): 1009-1017.

Woo, S., S. W. Cha, G. Merrihew, Y. He, N. Castellana, C. Guest, M. MacCoss and V. Bafna (2014). "Proteogenomic database construction driven from large scale RNA-seq data." J Proteome Res **13**(1): 21-28.

Wu, L. T., J. H. Hui and K. H. Chu (2013). "Origin and evolution of yolk proteins: expansion and functional diversification of large lipid transfer protein superfamily." Biol Reprod **88**(4): 102.

Xuereb, B., L. Bezin, A. Chaumot, H. Budzinski, S. Augagneur, R. Tutundjian, J. Garric and O. Geffard (2011). "Vitellogenin-like gene expression in freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization in females and potential for use as an endocrine disruption biomarker in males." Ecotoxicology **20**(6): 1286-1299.

Xuereb, B., A. Chaumot, R. Mons, J. Garric and O. Geffard (2009). "Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda) Intrinsic variability, reference levels, and a reliable tool for field surveys." Aquat Toxicol **93**(4): 225-233.

Zeng, V., K. E. Villanueva, B. S. Ewen-Campen, F. Alwes, W. E. Browne and C. G. Extavour (2011). "De novo assembly and characterization of a maternal and developmental transcriptome for the emerging model crustacean Parhyale hawaiensis." BMC Genomics **12**(1): 581.

3. Exploration du protéome des caeca hépatopancréatiques

Note 1 :

Chez les invertébrés, l'appareil digestif est un organe cible des polluants et par conséquent, concentre également de nombreux système de détoxification et de défense. De même, ce tissu est le centre métabolique de l'organisme, impliqué dans de nombreuses fonctions physiologiques clés telles que l'immunité ou la reproduction. En raison de son intérêt en écotoxicologie, cette étude porte sur l'exploration du protéome de la glande digestive, nommée *cæca* hépatopancréatiques, chez *G. fossarum*. Après séparation du protéome par isoélétrofocalisation OFFGel, les différentes fractions peptidiques ont été analysées par nanoLC-MS/MS. Un total de 102 106 spectres MS/MS ont été enregistrés, 593 d'entre eux ont pu être attribués pointant l'existence de 83 peptides uniques, soit un taux d'attribution de l'ordre de 0,6 %. En raison de la nature protéolytique de cet organe, les spectres enregistrés présentaient un important bruit de fond et leur taux d'attribution s'est montré particulièrement faible. En conclusion, les techniques protéomiques employées se révèlent inadéquates pour l'étude des *cæca* hépatopancréatiques et dans une perspective de poursuites de travaux sur cet organe, il est recommandé d'employer d'autres approches analytiques, basées notamment sur l'expression de gènes.

Introduction

Les organes digestifs concentrent de nombreux systèmes de détoxification et de défense, en teneur plus importante que dans le reste de l'organisme. Conventionnellement, en écotoxicologie, une large gamme de biomarqueurs est mesurée sur ces tissus. Parmi les rares biomarqueurs normalisés, on retrouve chez le poisson, le suivi de l'activité EROD dans la fraction microsomale hépatique (ISO/TS 23893-2:2007). Chez les bivalves, les activités antioxydantes, de métabolisation des xénobiotiques et la réponse lysosomale sont également communément mesurées au sein de la glande digestive (Canesi *et al.* 2010). De même, le système de détoxification MDMX et métallothionéines sont principalement identifiés au niveau des organes digestifs et des branchies (Pain *et al.* 2005; Amiard *et al.* 2006). Chez les crustacés, l'hépatopancréas joue un rôle majeur dans la défense immunitaire et est impliqué dans la synthèse des enzymes digestives et l'assimilation de l'énergie (résumé par Zhou *et al.* 2009). Cet organe occupe également un rôle dans la reproduction, ce dernier étant un lieu de synthèse majeur de protéines Vtg-like (Subramoniam 2011). Chez les amphipodes, les *cæca* hépatopancréatiques occupent la fonction d'appareil digestif, différant en termes de structure, ces derniers étant tubulaires. Différents biomarqueurs d'intérêt ont été mesurés sur ce tissu : l'activité des enzymes digestives amylase, cellulase et trypsine chez *Gammarus fossarum* (Charron 2014) ainsi que les réserves énergétiques et la défense lysosomale chez *Dikerogammarus villosus* (Guerlet et al. 2008). En raison de la fonction de centre métabolique qu'occupent les *cæca* hépatopancréatiques, la caractérisation du protéome de cet organe est donc hautement pertinente dans une démarche de découverte de protéine d'intérêt en écotoxicologie.

Objectif général

Alors que la BD transcriptomique présentée dans la **publication 3** comprend les transcrits issus de ce tissu, le protéome associé n'a pas été présenté au sein de la publication. L'objectif de cette note est donc de documenter le protéome des *cæca* hépatopancréatiques chez *Gammarus fossarum* en utilisant la même méthodologie que celle utilisée pour la création du catalogue de protéines sur les organes reproducteurs et le céphalon. Brièvement, les organismes ont été disséqués sous loupe binoculaire et 30 organismes ont été utilisés pour réaliser un échantillon biologique, l'expérience étant réalisée en duplicat. Par la suite, les protéines ont été extraites, purifiées et la digestion a été réalisée hors-gel. Le mélange peptidique a par la suite été fractionné par OFFGel

suivant la configuration 24 puits donnant un total de 48 échantillons à analyser par nanoLC-MS/MS.

Résultats et discussion

Un total de 102 106 spectres MS/MS a été enregistré, 593 d'entre eux ont pu être attribués pointant l'existence de 83 peptides uniques. Alors que pour les tissus reproducteurs et la cephalon, le taux d'attribution des spectres était de l'ordre de 5 %, pour les *cæca* hépatopancréatiques, il est ici de l'ordre de 0,6 %. De même, les spectres MS/MS enregistrés pour ce tissu présentent un important bruit de fond. Une observation similaire a été réalisée chez l'écrevisse de Louisiane *Procambarus clarkii* par Vioque-Fernandez *et al.* (2009). Dans cette étude, les auteurs cherchaient à comparer sur l'hépatopancréas les réponses de différents biomarqueurs tels que l'activité EROD, AChE ou GST et les réponses du protéome par 2D-PAGE. Toutefois, l'étude protéomique a été réalisée sur les branchies, en raison de l'activité protéolytique de l'organe digestif, interférant avec l'analyse. Dans notre étude, malgré l'utilisation d'inhibiteurs de protéases dans le tampon d'extraction, les enzymes digestives réalisent de nombreuses coupures, non spécifiques à la trypsine et donc non reconnues d'après nos critères d'attribution des spectres. De même, les *cæca* hépatopancréatiques sont également des organes de réserves énergétiques, stockant des sucres et des lipides en importante quantité, biomolécules pouvant être également source de bruit de fond.

Le Tableau III-1 présente les quatorze protéines identifiées dans l'étude du protéome des *cæca* hépatopancréatiques, dont dix ont été uniquement séquencées dans cet organe. Six d'entre-elles sont de la famille des LLTP, confirmant le rôle majeur de ce tissus dans la synthèse des protéines Vtg-like, quatre sont des peptidases et enfin trois sont dépourvues d'annotation fonctionnelle.

Tableau III-1 : Protéines identifiées dans les *cæca* hépatopancréatiques

N° du contig	Fonction prédite	Origine du plus proche homologue	E-value	Nombre de spectres MS/MS	Présence dans d'autres tissus
184	Aminopeptidase	*Crassostrea gigas*	1,00E-123	12	Non
22001	Clottable protein	*Marsupenaeus japonicus*	1,00E-07	6	Non
17046	Clottable protein 2	*Penaeus monodon*	1,00E-19	3	Oui
18500	Hemolymph clottable protein	*Penaeus monodon*	1,00E-14	32	Oui
133271	Metalloproteinase	*Aedes aegypti*	3,00E-16	23	Non
134359	No predicted function	*Drosophila ananassae*	1,00E-29	18	Non
181349	No predicted function	*Daphnia pulex*	9,00E-07	13	Non
724	No predicted function	*Daphnia pulex*	6,00E-84	2	Non
106932	Single-stranded DNA-binding protein	*Nasonia vitripennis*	5,00E-30	7	Non
24570	Vitellogenin-like protein precursor	*Gammarus fossarum*	1,00E-57	8	Non
37795	Vitellogenin-like protein precursor	*Gammarus fossarum*	4,00E-78	6	Oui
203698	Vitellogenin-like protein precursor	*Gammarus fossarum*	5,00E-16	65	Oui
47660	Zinc metalloproteinase nas-12	*Culex quinquefasciatus*	2,00E-04	8	Non
169712	Zinc metalloproteinase nas-12	*Culex quinquefasciatus*	2,00E-03	15	Non

Conclusion

Alors que les *cæca* hépatopancréatiques est un organe d'intérêt en écotoxicologie, son étude par les approches utilisées au cours de ces travaux se révèlent inadéquates. En effet, l'analyse du protéome n'a pas permis d'identifier un nombre conséquent de protéines en raison de la nature protéolytique de l'organe. C'est pourquoi, dans la possibilité de poursuites de travaux sur ce tissu, il est recommandé de développer des approches protéomiques spécifiques. Alors qu'au cours de ces travaux de thèse, la ressource transcriptomique a été générée à des fins d'annotation structurale des protéines, elle reste une source d'information biologique inexploitée. Une alternative serait également d'utiliser des techniques basées sur l'expression de gènes, notamment par la réalisation de puces à ADN pour la découverte de nouveaux biomarqueurs d'intérêt.

Références

Amiard, J. C., C. Amiard-Triquet, et al. (2006). "Metallothioneins in aquatic invertebrates: Their role in metal detoxification and their use as biomarkers." Aquatic Toxicology **76**(2): 160-202.

Canesi, L., C. Barmo, et al. (2010). "Effects of vibrio challenge on digestive gland biomarkers and antioxidant gene expression in Mytilus galloprovincialis." Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **152**(3): 399-406.

Charron, L. (2014). Biomarqueurs énégertiques chez un amphipode d'eau douce Gammarus fossarum : Développement, lien avec le succès reproducteurs et application in situ PhD, Université de Reims Champagne-Ardennes.

Guerlet, E., K. Ledy, et al. (2008). "Is the freshwater gammarid, Dikerogammarus villosus, a suitable sentinel species for the implementation of histochemical biomarkers?" Chemosphere **72**(5): 697-702.

Pain, S., S. Biaganti-Risbourg, et al. (2005). Relevance of the multixenobiotic defence mechanism (MXDM) for the biological monitoring of freshwaters - example of its use in zebra mussels. Trends in water pollution research. J. V. Livingston. New York, Nova Science Publishers**:** 203-220.

Subramoniam, T. (2011). "Mechanisms and control of vitellogenesis in crustaceans." Fisheries Science **77**(1): 1-21.

Vioque-Fernandez, A., E. Alves de Almeida, et al. (2009). "Assessment of Donana National Park contamination in Procambarus clarkii: integration of conventional biomarkers and proteomic approaches." Sci Total Environ **407**(5): 1784-1797.

Zhou, J., W.-N. Wang, et al. (2009). "Glutathione S-transferase in the white shrimp Litopenaeus vannamei: Characterization and regulation under pH stress." Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **150**(2): 224-230.

4. Problématique du séquençage *de novo* du transcriptome

Note 2 :

Les techniques NGS permettent un niveau inédit d'investigation au cœur de la biodiversité terrestre pour la compréhension de la diversité des processus biologiques ou évolutifs. Elles sont notamment employées pour obtenir un premier aperçu du fonctionnement au niveau moléculaire par la caractérisation du transcriptome d'espèces non-modèles. Toutefois, à l'issue du séquençage, les séquences d'ARN sont reconstruites en absence de génome de référence, engendrant des erreurs en termes d'annotation structurale des protéines. Cette annotation peut néanmoins être corrigée en intégrant les données expérimentales de spectrométrie de masse, comme présenté dans cette note.

Introduction

Les avancées technologiques dans le domaine du séquençage des acides nucléiques (NGS) ont permis la démocratisation des approches en biologie moléculaire chez des espèces non-modèles, trouvant une large palette d'applications en agronomie, microbiologie, écologie ou encore en écotoxicologie. Ces techniques sont majoritairement exploitées pour caractériser le transcriptome des organismes. Elles permettent également, comme illustré avec notre modèle d'étude *Gammarus fossarum*, la réalisation d'expérimentation protéomique à large échelle pour la découverte de protéines. Toutefois, alors que les technologies NGS permettent une caractérisation sans précédent de la biodiversité au niveau moléculaire, elles comportent leur part d'erreur au niveau de l'exactitude des séquences codantes, notamment en raison de l'absence de génome de référence pour l'assemblage des brins séquencés (reads). D'un point de vu biologique, ces erreurs peuvent entrainer des biais d'interprétation car ces séquences erronées ne trouvent pas d'homologues au sein des BD publiques et par conséquent entrainent une surestimation de la proportion de gènes orphelins chez l'organisme étudié (Gibson *et al.* 2013).

Suite au traitement des données MS/MS pour la création du catalogue de protéines (**publication 3**), nous avons pu observer au sein d'un même contig l'existence de différents ORF, phénomène observé pour 64 d'entre eux. L'objectif de cette note est d'expliciter les différentes raisons à l'origine de ces changements de cadre de lecture et de présenter un exemple concret d'utilisation des données expérimentales MS/MS pour améliorer l'annotation structurale des protéines.

Origines des changements de cadre de lecture

L'utilisation d'une BD de séquences protéiques restreinte permet une interprétation plus sensible des spectres MS/MS en protéomique shotgun (Wang *et al.* 2012). C'est pourquoi, au cours de ces travaux de thèse, seuls les contigs résultant de l'assemblage du transcriptome ont été utilisés pour la création de la BD ORF. En effet, les contigs sont issus de l'assemblage de séquences multiples alors que les singletons sont principalement chimériques, leur intégration peut provoquer de l'interférence dans l'interprétation des données MS/MS. Toutefois, le séquençage du protéome a mis en évidence que certains contigs étaient lus selon différents ORF, ces changements de cadre de lecture s'opérant toujours dans le même sens (sens ou anti-sens). Ces observations trouvent leur

origine au niveau [1] biologique par des processus post-traductionnels d'épissage et [2] technique par l'existence d'erreurs de séquençage *i.e.* addition ou délétion d'un nucléotide et d'erreurs au cours de l'assemblage *de novo* du transcriptome. Les proportions attribuables aux processus biologiques ou aux artéfacts analytiques ne peuvent être définies. Toutefois, l'absence de génome de référence affecte certainement la qualité d'assemblage du transcriptome.

Exemple d'amélioration de l'annotation structurale par les données de spectrométrie de masse

Pour certaines protéines d'intérêt identifiées au cours des différentes expérimentations protéomiques, un travail manuel de correction de séquence en acides aminés a été réalisé à partir de séquences peptidiques observées en spectrométrie de masse (Figure III-1). Ce travail a pour objectif de générer une séquence en acides aminés plus proche de la réalité biologique et a été notamment réalisé pour les protéines dont l'histoire phylogénétique a été reconstruite, afin d'améliorer la qualité de l'alignement et de l'arbre phylogénétique associé.

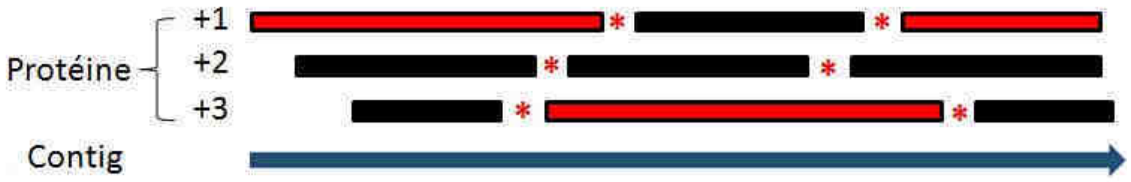


Figure III-1 : Stratégie pour l'amélioration de l'annotation structurale de la protéine d'intérêt par l'intégration des données MS

Les étoiles rouge * correspondent à des codons STOP sur la séquence prédite et les zones en rouge aux peptides qui ont été séquencés par MS.

Ici, l'exemple du contig 17046, identifié comme étant une Vtg de *G. fossarum* (**publication 4**) est illustré. Le brin sens en acides nucléiques est présenté, associé aux séquences protéiques prédites suivant les trois cadres de lectures +1, +2 et +3. Un total de 69 peptides uniques a été associé à ce contig : 37 peptides sur l'ORF +3, 36 peptides sur l'ORF +2 et 6 peptides sur l'ORF +1. La séquence en acides aminés reconstruite débute au niveau de la méthionine (introduite par le codon d'initiation) la plus proche du premier peptide séquencé et se termine au niveau du premier codon STOP en aval du dernier peptide séquencé. Au final, la séquence en acides aminés employée pour l'alignement est donnée en format FASTA.

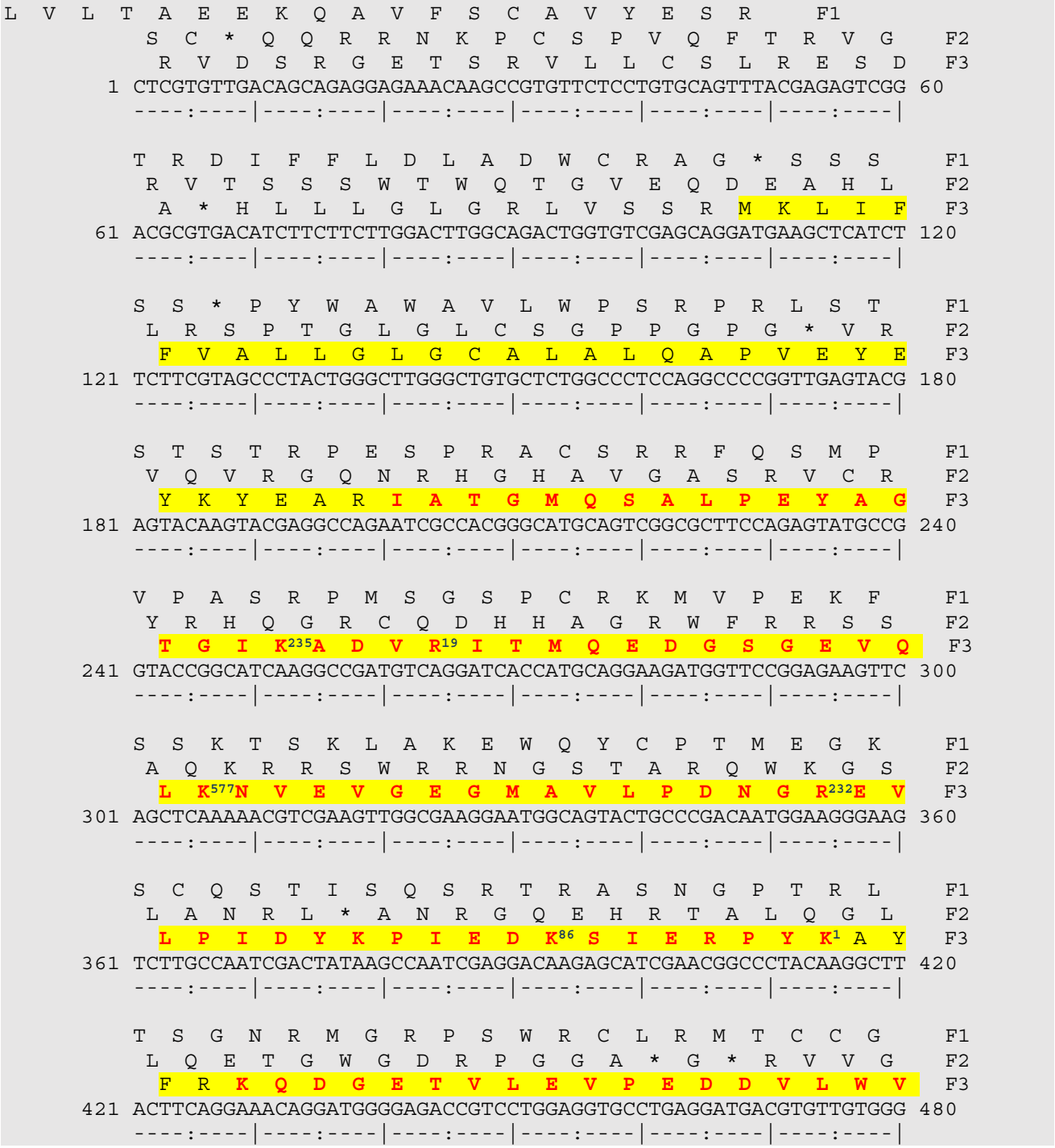
Légende

S F K: peptide certifié par MS/MS

M I N: séquence en acides aminées reconstruite

A I R^{40}: nombre de spectral count pour chaque peptide obtenu au cours de l'expérimentation (dans le cas de peptides chevauchants, seul le spectral count le plus grand est affiché).

* : présence d'un codon STOP



```
    *  P  T  S  G  E  E  S  P  T  S  S  V  S  P  R  S  S  S  P     F1
     D  Q  R  Q  A  R  S  R  Q  L  L  P  Y  R  P  D  R  P  A  R    F2
      T  N  V  R101 R83 G  V  A  N  F  F  R1 I  A  P  I  V  Q  P  G  F3
481 TGACCAACGTCAGGCGAGGAGTCGCCAACTTCTTCCGTATCGCCCCGATCGTCCAGCCCG 540
    ----:----|----:----|----:----|----:----|----:----|----:----|

    A  P  C  L  P  L  E  K  A  L  S  S  S  S  P  R  R  I  L  Q     F1
     L  R  A  S  P  S  R  R  L  *  A  P  R  R  L  V  G  Y  S  R    F2
      S  V  P  P  P  R150 E  G  S  E  L  L  V  A  S  S  D  T  P  G  F3
541 GCTCCGTGCCTCCCCCTCGAGAAGGCTCTGAGCTCCTCGTCGCCTCGTCGGATACTCCAG 600
    ----:----|----:----|----:----|----:----|----:----|----:----|

    D  Y  G  H  L  R  G  S  L  Q  S  R  V  H  H  H  S  C  C  H     F1
     I  M  D  T  F  A  G  R3 C  R  V  E  Y  T  I  T  P  A  A  I    F2
      L  W  T  P  S  R  V  V  A  E  S  S  T  P  S  L  L  L  P  S  F3
601 GATTATGGACACCTTCGCGGGTCGTTGCAGAGTCGAGTACACCATCACTCCTGCTGCCAT 660
    ----:----|----:----|----:----|----:----|----:----|----:----|

    L  R  R  Q  V  W  H  P  R  R  H  P  V  Q  K  G  S  *  G  T     F1
     S  D  V  R4 S  G  I  R  E  D  I  Q  S  R  R  A  P  K  G  P    F2
      P  T  S  G  L  A  S  A  K  T  S  S  P  E  G  L  L  R  D  P  F3
661 CTCCGACGTCAGGTCTGGCATCCGCGAAGACATCCAGTCCAGAAGGGCTCCTAAGGGACC 720
    ----:----|----:----|----:----|----:----|----:----|----:----|

    R  W  T  T  S  Q  T  R  D  D  L  H  V  V  Q  S  S  M  T  S     F1
     G  G  L  Q  V  K  L  E  M  T  Y  M  L  C  S  L  P  *  L  Q    F2
      V  D  Y  K  S  N  S  R  *  P  T  C  C  A  V  F  H  D  F  S  F3
721 CGGTGGACTACAAGTCAAACTCGAGATGACCTACATGTTGTGCAGTCTTCCATGACTTCA 780
    ----:----|----:----|----:----|----:----|----:----|----:----|

    V  P  A  P  T  G  C  S  S  H  R  L  V  L  V  T  *  L  T  S     F1
     Y  L  L  Q  Q  G  V  P  L  T  D  W  C  W  *  P  N  *  R  H    F2
      T  C  S  N  R  V  F  L  S  Q  T  G  A  G  N  L  T  N  V  T  F3
781 GTACCTGCTCCAACAGGGTGTTCCTCTCACAGACTGGTGCTGGTAACCTAACTAACGTCA 840
    ----:----|----:----|----:----|----:----|----:----|----:----|

    L  R  H  R  T  E  M  F  Q  P  G  H  T  L  E  S  T  S  L  K     F1
     F  V  I  V  P  R  C  F  S  Q  V  T  R  W  R  V  H  P  *  R    F2
      S  S  S  Y  R  D  V  S  A  R  S  H  V  G  E  Y  I  L  K  G  F3
841 CTTCGTCATCGTACCGAGATGTTTCAGCCAGGTCACACGTTGGAGAGTACATCCTTAAAG 900
    ----:----|----:----|----:----|----:----|----:----|----:----|

    A  S  Q  D  S  S  G  W  R  G  V  R  L  S  Q  Q  *  F  C  N     F1
     Q  A  R  T  V  Q  D  G  E  G  *  G  *  V  S  S  N  S  A  T    F2
      K  P  G  Q  F  R2 M  E  R  G  E  V  E  S  A  V  I  L  Q  P  F3
901 GCAAGCCAGGACAGTTCAGGATGGAGAGGGGTGAGGTTGAGTCAGCAGTAATTCTGCAAC 960
    ----:----|----:----|----:----|----:----|----:----|----:----|

    L  S  V  S  I  L  P  R  S  S  *  *  A  I  S  P  W  N  W  C     F1
     F  R  F  Q  F  C  Q  G  R  H  N  E  Q  S  V  H  G  T  G  V    F2
      F  G  F  N  S  A  K134 V  V  I  M  S  N  Q  S  M  E  L  V  S  F3
961 CTTTCGGTTTCAATTCTGCCAAGGTCGTCATAATGAGCAATCAGTCCATGGAACTGGTGT 1020
    ----:----|----:----|----:----|----:----|----:----|----:----|

    P  P  N  Q  S  P  A  M  *  G  L  Q  R  K  T  *  S  L  S  R     F1
     R  Q  T  S  P  P  Q  C  K  V  F  S  A  K  P  E  V  Y  Q  D    F2
      A  K  P  V  P  R  N  V  R  S  S  A  Q  N  L  K  S  I  K  T  F3
1021 CCGCCAAACCAGTCCCCCGCAATGTAAGGTCTTCAGCGCAAAACCTGAAGTCTATCAAGA 1080
    ----:----|----:----|----:----|----:----|----:----|----:----|
```

```
     P  G  T  I  N  *  K  A  Q  N  R  N  L  K  A  S  L  L  A  L     F1
      L  E  P  S  I  R  K  P  K  T  E  T  *  R  L  R  S  S  L  Y     F2
       W  N  H  Q  L  E  S  P  K  P  K  P  E  G  F  A  P  R83 F  I   F3
1081 CCTGGAACCATCAATTAGAAAGCCCAAAACCGAAACCTGAAGGCTTCGCTCCTCGCTTTA 1140
     ----:----|----:----|----:----|----:----|----:----|----:----|

     S  S  T  R  I  R  T  N  L  S  T  A  Q  S  S  T  S  D  C  *     F1
      P  V  Q  E  S  E  R  T  *  V  Q  P  N  L  L  P  A  I  V  D    F2
       Q  Y  K  N  Q  N  E  P  E  Y  S  P  I  F  Y  Q  R65 L  L  I   F3
1141 TCCAGTACAAGAATCAGAACGAACCTGAGTACAGCCCAATCTTCTACCAGCGATTGTTGA 1200
     ----:----|----:----|----:----|----:----|----:----|----:----|

     *  A  A  S  E  *  T  T  K  E  F  K  I  T  Y  W  T  G  L  R     F1
      K  R  L  P  N  E  R  R  R  S  S  K  *  L  I  G  Q  G  *  D    F2
       S  G  F  R  M  N  D  E  G  V  Q  N  N  L  L  D  R116V  K162I  F3
1201 TAAGCGGCTTCCGAATGAACGACGAAGGAGTTCAAAATAACTTATTGGACAGGGTTAAGA 1260
     ----:----|----:----|----:----|----:----|----:----|----:----|

     *  L  S  R  R  Q  L  S  D  A  G  Y  R  T  R  T  P  T  Q  S     F1
      N  Y  R  E  G  S  *  A  M  Q  V  T  E  Q  G  L  R  R  R  A    F2
       T  I  E  K  A  A  E  R  C  R  L  Q  N  K  D  S  D  A  E  P   F3
1261 TAACTATCGAGAAGGCAGCTGAGCGATGCAGGTTACAGAACAAGGACTCCGACGCAGAGC 1320
     ----:----|----:----|----:----|----:----|----:----|----:----|

     Q  P  H  S  G  A  P  K  M  V  A  *  P  K  A  *  W  M  P  C     F1
      N  H  I  Q  G  H  R  R  W  S  H  S  P  R  H  D  G  C  H  A    F2
       T  T  F  R100G  T  E  D  G  R11I  A  Q  G  M  M  D  A  M  H   F3
1321 CAACCACATTCAGGGGCACCGAAGATGGTCGCATAGCCCAAGGCATGATGGATGCCATGC 1380
     ----:----|----:----|----:----|----:----|----:----|----:----|

     T  C  C  A  P  S  T  M  K  R  F  R  N  W  T  R  R  C  R  S     F1
      P  A  A  R  P  Q  P  *  K  G  S  G  T  G  Q  E  D  A  G  A    F2
       L  L  R139A  L  N  H  E  K  V  Q  E  L  D  K  K  M  Q  E  H   F3
1381 ACCTGCTGCGCGCCCTCAACCATGAAAAGGTTCAGGAACTGGACAAGAAGATGCAGGAGC 1440
     ----:----|----:----|----:----|----:----|----:----|----:----|

     T  R  T  S  W  L  E  K  F  S  S  T  L  *  Q  L  L  P  L  D     F1
      P  G  Q  A  G  *  R  N  F  H  R  R  S  S  S  C  C  L  W  T    F2
       Q  D  K  L  V  R  E  I  F  I  D  A  L  A  A  A  A  S  G  P   F3
1441 ACCAGGACAAGCTGGTTAGAGAAATTTTCATCGACGCTCTAGCAGCTGCTGCCTCTGGAC 1500
     ----:----|----:----|----:----|----:----|----:----|----:----|

     Q  P  *  K  C  Y  C  G  K  S  E  T  M  N  Y  L  I  I  G  L     F1
      N  R  E  S  A  T  A  E  N  P  K  P  *  T  I  *  *  *  D  C    F2
       T  V  K  V144L  L  R49 K1  I  R  N  H  E  L  S  D  N  R  T  A  F3
1501 CAACCGTGAAAGTGCTACTGCGGAAAATCCGAAACCATGAACTATCTGATAATAGGACTG 1560
     ----:----|----:----|----:----|----:----|----:----|----:----|

     L  P  F  S  S  T  *  L  T  T  S  D  L  R  N  L  A  N  N  I     F1
      F  H  F  L  Q  L  S  *  Q  H  P  I  C  A  T  *  L  T  T  S    F2
       S  I  F  F  N  L  A  N  N  I  R185S  A  Q  L  S  *  Q  H  P   F3
1561 CTTCCATTTTCTTCAACTTAGCTAACAACATCCGATCTGCGCAACTTAGCTAACAACATC 1620
     ----:----|----:----|----:----|----:----|----:----|----:----|

     R  S  A  D  N  *  R  A  S  G  L  C  H  V  S  G  R  R  K  E     F1
      D  L  R  I  I  E  E  L  V  D  F  A  M  S  L  E  E  E  K  N    F2
       I  C  G  *  L  K  S  *  W  T  L  P  C  L  W  K  K  K  R  T   F3
1621 CGATCTGCGGATAATTGAAGAGCTAGTGGACTTTGCCATGTCTCTGGAAGAAGAAAAGAA 1680
     ----:----|----:----|----:----|----:----|----:----|----:----|

     R  H  C  *  S  L  G  Q  E  Q  R  C  L  D  C  P  S  H  T  H     F1
```

```
      V  T  A  S  L  L  V  K  N  N  V  A  L  I  V  R86 R3 I  R  I     F2
       S  L  L  V  S  W  S  R  T  T  L  P  *  L  S  V  A  Y  A  Y     F3
1681  CGTCACTGCTAGTCTCTTGGTCAAGAACAACGTTGCCTTGATTGTCCGTCGCATACGCAT  1740
      ----:----|----:----|----:----|----:----|----:----|----:----|

      M  R  Q  F  T  D  Q  E  I  L  S  H  D  V  R  I  W  *  Q  N     F1
       C  V  S  S  Q  I  K  K  Y  S  H  M  M  Y  A  F  G  S  K70 M     F2
       A  S  V  H  R  S  R  N  T  L  T  *  C  T  H  L  V  A  K  C     F3
1741  ATGCGTCAGTTCACAGATCAAGAAATACTCTCACATGATGTACGCATTTGGTAGCAAAAT  1800
      ----:----|----:----|----:----|----:----|----:----|----:----|

      V  *  *  N  Q  D  C  Q  Q  V  F  A  S  S  R  T  K  N  *  K     F1
       C  D  R  T  K  I  V  N  K  Y  L  P  H  L  E  Q  K7 I  E  S     F2
       V  I  E  P  R  L  S  T  S  I  C  L  I  S  N  K  K  L  K  V     F3
1801  GTGTGATAGAACCAAGATTGTCAACAAGTATTTGCCTCATCTCGAACAAAAAATTGAAAG  1860
      ----:----|----:----|----:----|----:----|----:----|----:----|

      F  Q  D  N  L  G  Q  D  L  E  R  D  G  C  C  *  S  W  R  P     F1
       S  K11 I  T  W  V  K52 I  L  N  V  M  A  A  A  D  L  G  A  R347    F2
       P  R  *  P  G  S  R  S  *  T  *  W  L  L  L  I  L  A  P  E     F3
1861  TTCCAAGATAACCTGGGTCAAGATCTTGAACGTGATGGCTGCTGCTGATCTTGGCGCCCG  1920
      ----:----|----:----|----:----|----:----|----:----|----:----|

      R  E  H  P  D  P  G  E  G  S  T  R  V  A  R  G  D  V  S  S     F1
       E  S  I  P  I  L  E  K20 V  A  R17 G  *  H  E  V  T  L  V  A     F2
       R  A  S  R  S  W  R  R  *  H  E  G  S  T  R  *  R  *  *  R     F3
1921  AGAGAGCATCCCGATCCTGGAGAAGGTAGCACGAGGGTAGCACGAGGTGACGTTAGTAGC  1980
      ----:----|----:----|----:----|----:----|----:----|----:----|

      D  I  F  V  R40 F  A  A  I  H  G  V  C  D  Q  L  *  Q  L  L     F1
       T  S  S  S  V  S  L  P  S  T  V  S  A  T  N  F  D  N  S  S     F2
       H  L  R  P  F  R  C  H  P  R  C  L  R  P  T  L  T  T  P  P     F3
1981  GACATCTTCGTCCGTTTCGCTGCCATCCACGGTGTCTGCGACCAACTTTGACAACTCCTC  2040
      ----:----|----:----|----:----|----:----|----:----|----:----|

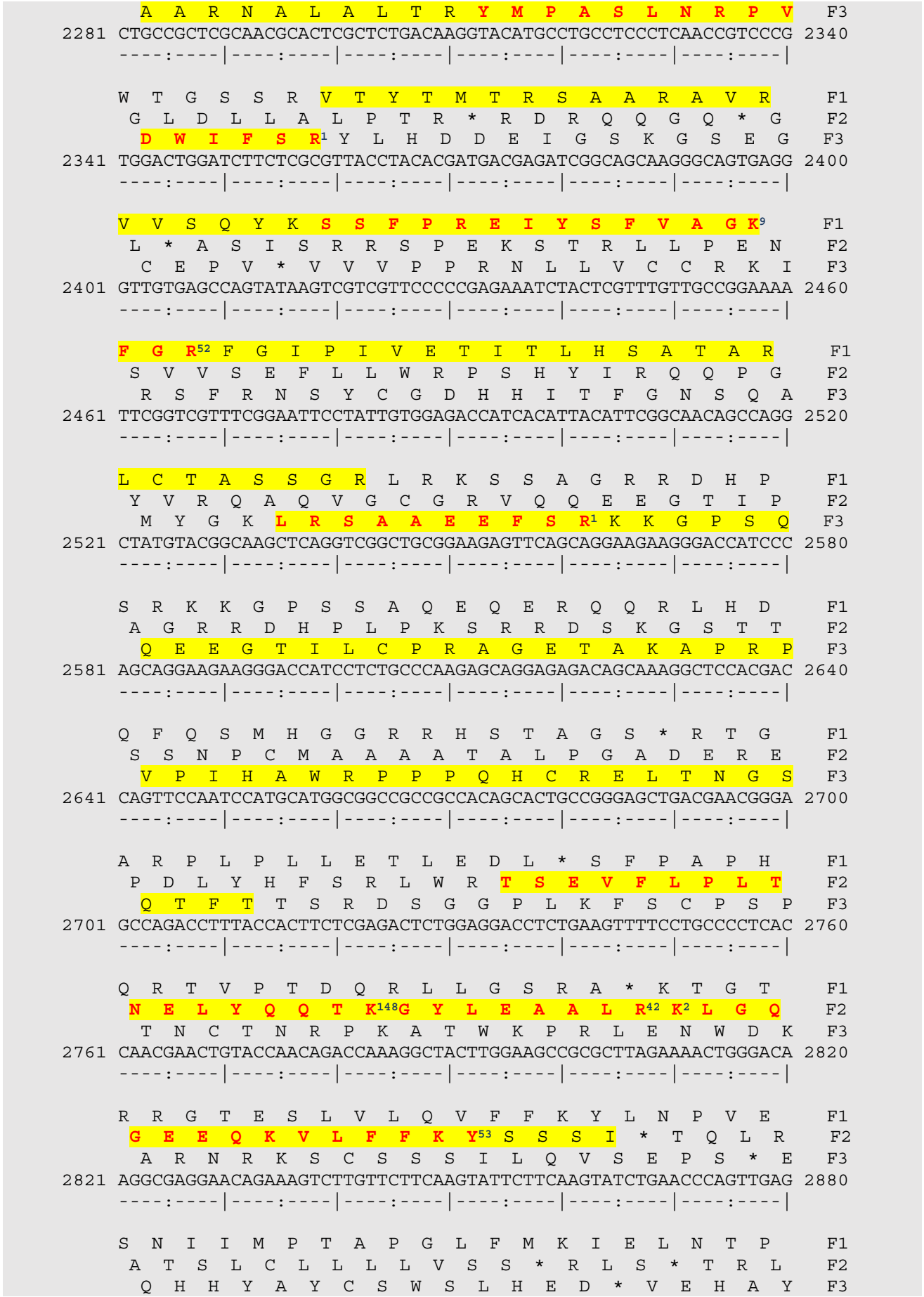
      P  *  *  D  L  Q  C  P  A  A  H  C  R  I  Q  G  R  R  L  P     F1
       R  D  K  I  F  N  V  L  Q  P  I  A  E  S  K85 V  E  D  Y  R71    F2
       V  I  R  S  S  M  S  C  S  P  L  P  N  P  R  *  K  I  T  A     F3
2041  CCGTGATAAGATCTTCAATGTCCTGCAGCCCATTGCCGAATCCAAGGTAGAAGATTACCG  2100
      ----:----|----:----|----:----|----:----|----:----|----:----|

      H  P  P  G  R  R  S  P  T  T  L  S  *  L  S  C  R  M  S  P     F1
       I  R  Q  V  A  G  R  L  R  L  S  P  D  F  H  A  E  C  H  L     F2
       S  A  R  S  P  V  A  Y  D  S  L  L  T  F  M  P  N  V  T  F     F3
2101  CATCCGCCAGGTCGCCGGTCGCCTACGACTCTCTCCTGACTTTCATGCCGAATGTCACCT  2160
      ----:----|----:----|----:----|----:----|----:----|----:----|

      S  S  R  E  *  S  S  L  L  T  T  S  L  H  N  R  S  K  A  T     F1
       H  P  E  N  D  L  L  F  L  R  R  A  F  T  T  G  P  K  L  H     F2
       I  Q  R  M  I  F  S  S  Y  D  E  P  S  Q  Q  V  Q  S  Y  I     F3
2161  TCATCCAGAGAATGATCTTCTCTTCTTACGACGAGCCTTCACAACAGGTCCAAAGCTACA  2220
      ----:----|----:----|----:----|----:----|----:----|----:----|

      Y  L  G  S  L  T  V  Y  P  A  L  K  I  P  D  T  D  F  Y  P     F1
       I  L  G  R  *  Q  S  I  P  H  *  R  S  Q  I  Q  T  S  I  H     F2
       S  W  V  V  D  S  L  S  R326I  E  D  P  R  Y  R  L  L  S  T     F3
2221  TATCTTGGGTCGTTGACAGTCTATCCCGCATTGAAGATCCCAGATACAGACTTCTATCCA  2280
      ----:----|----:----|----:----|----:----|----:----|----:----|

      L  P  L  A  T  H  S  L  *  Q  G  T  C  L  P  P  S  T  V  P     F1
       C  R  S  Q  R  T  R  S  D  K  V  H  A  C  L  P  Q  P  S  R     F2
```


A A R N A L A L T R Y M P A S L N R P V F3
2281 CTGCCGCTCGCAACGCACTCGCTCTGACAAGGTACATGCCTGCCTCCCTCAACCGTCCCG 2340
W T G S S R V T Y T M T R S A A R A V R F1
G L D L L A L P T R * R D R Q Q G Q * G F2
D W I F S R1 Y L H D D E I G S K G S E G F3
2341 TGGACTGGATCTTCTCGCGTTACCTACACGATGACGAGATCGGCAGCAAGGGCAGTGAGG 2400
V V S Q Y K S S F P R E I Y S F V A G K9 F1
L * A S I S R R S P E K S T R L L P E N F2
C E P V * V V V P P R N L L V C C R K I F3
2401 GTTGTGAGCCAGTATAAGTCGTCGTTCCCCCGAGAAATCTACTCGTTTGTTGCCGGAAAA 2460
F G R52 F G I P I V E T I T L H S A T A R F1
S V V S E F L L W R P S H Y I R Q Q P G F2
R S F R N S Y C G D H H I T F G N S Q A F3
2461 TTCGGTCGTTTCGGAATTCCTATTGTGGAGACCATCACATTACATTCGGCAACAGCCAGG 2520
L C T A S S G R L R K S S A G R R D H P F1
Y V R Q A Q V G C G R V Q Q E E G T I P F2
M Y G K L R S A A E E F S R1 K K G P S Q F3
2521 CTATGTACGGCAAGCTCAGGTCGGCTGCGGAAGAGTTCAGCAGGAAGAAGGGACCATCCC 2580
S R K K G P S S A Q E Q E R Q Q R L H D F1
A G R R D H P L P K S R R D S K G S T T F2
Q E E G T I L C P R A G E T A K A P R P F3
2581 AGCAGGAAGAAGGGACCATCCTCTGCCCAAGAGCAGGAGAGACAGCAAAGGCTCCACGAC 2640
Q F Q S M H G G R R H S T A G S * R T G F1
S S N P C M A A A A T A L P G A D E R E F2
V P I H A W R P P P Q H C R E L T N G S F3
2641 CAGTTCCAATCCATGCATGGCGGCCGCCGCCACAGCACTGCCGGGAGCTGACGAACGGGA 2700
A R P L P L L E T L E D L * S F P A P H F1
P D L Y H F S R L W R T S E V F L P L T F2
Q T F T T S R D S G G P L K F S C P S P F3
2701 GCCAGACCTTTACCACTTCTCGAGACTCTGGAGGACCTCTGAAGTTTTCCTGCCCCTCAC 2760
Q R T V P T D Q R L L G S R A * K T G T F1
N E L Y Q Q T K148G Y L E A A L R42 K2 L G Q F2
T N C T N R P K A T W K P R L E N W D K F3
2761 CAACGAACTGTACCAACAGACCAAAGGCTACTTGGAAGCCGCGCTTAGAAAACTGGGACA 2820
R R G T E S L V L Q V F F K Y L N P V E F1
G E E Q K V L F F K Y53 S S S I * T Q L R F2
A R N R K S C S S S I L Q V S E P S * E F3
2821 AGGCGAGGAACAGAAAGTCTTGTTCTTCAAGTATTCTTCAAGTATCTGAACCCAGTTGAG 2880
S N I I M P T A P G L F M K I E L N T P F1
A T S L C L L L L V S S * R L S * T R L F2
Q H H Y A Y C S W S L H E D * V E H A Y F3

```
2881 AGCAACATCATTATGCCTACTGCTCCTGGTCTCTTCATGAAGATTGAGTTGAACACGCCT 2940
     ----:----|----:----|----:----|----:----|----:----|----:----|

     T  V  L  F  G  E  V  K  M  H  A  K  I  S  P  L  I  N  S  P     F1
      R  S  C  S  E  K  L  R  C  T  P  R  Y  P  H  *  *  I  A  L    F2
       G  L  V  R  R  S  *  D  A  R  Q  D  I  P  T  D  K  *  P  F    F3
2941 ACGGTCTTGTTCGGAGAAGTTAAGATGCACGCCAAGATATCCCCACTGATAAATAGCCCT 3000
     ----:----|----:----|----:----|----:----|----:----|----:----|

     S  D  L  P  K  L  Q  K  I  N  F  R  A  E  G  I  V  K  A  I     F1
      Q  I  Y  R  S  F  R  K  S  T  S  G  P  K  E  L  S  K  P  S    F2
       R  S  T  E  A  S  E  N  Q  L  Q  G  R  R  N  C  Q  S  H  L    F3
3001 TCAGATCTACCGAAGCTTCAGAAAATCAACTTCAGGGCCGAAGGAATTGTCAAAGCCATC 3060
     ----:----|----:----|----:----|----:----|----:----|----:----|

     S  K  T  N  S  Y  S  L  V  M  A  P  W  T  N  R  T  V  L  C     F1
      R  R  P  T  A  T  P  S  S  W  R  R  G  P  T  G  P  Y  C  A    F2
       E  D  Q  Q  L  L  P  R  H  G  A  V  D  Q  P  D  R  I  V  R    F3
3061 TCGAAGACCAACAGCTACTCCCTCGTCATGGCGCCGTGGACCAACCGGACCGTATTGTGC 3120
     ----:----|----:----|----:----|----:----|----:----|----:----|

     G  V  D  V  E  K  T  V  S  L  P  L  K  Y  E  I  E  A  D  L     F1
      G  L  M  W  R  R  P  S  L  C  P  *  N  T  R  L  K  L  I  F    F2
       G  *  C  G  E  D  R  L  S  A  L  E  I  R  D  *  S  *  S  S    F3
3121 GGGGTTGATGTGGAGAAGACCGTCTCTCTGCCCTTGAAATACGAGATTGAAGCTGATCTT 3180
     ----:----|----:----|----:----|----:----|----:----|----:----|

     Q  T  P  A  Q  I  L  W  A  E  L  T  P  Q  F  N  E  T  G  E     F1
      R  R  Q  R  R  S  C  G  R  S  S  P  P  S  S  T  R  P  A  R    F2
       D  A  S  A  D  L  V  G  G  A  H  P  P  V  Q  R  D  R  R  G    F3
3181 CAGACGCCAGCGCAGATCTTGTGGGCGGAGCTCACCCCCCAGTTCAACGAGACCGGCGAG 3240
     ----:----|----:----|----:----|----:----|----:----|----:----|

     D  I  P  V  I  K  I  R  T  L  P  F  T  A  I  A  G  M  M  S     F1
      I  S  Q  S  S  R  S  A  P  C  L  S  R  P  L  L  A  *  C  L    F2
       Y  P  S  H  Q  D  P  H  P  A  F  H  G  H  C  W  H  D  V  Y    F3
3241 GATATCCCAGTCATCAAGATCCGCACCCTGCCTTTCACGGCCATTGCTGGCATGATGTCT 3300
     ----:----|----:----|----:----|----:----|----:----|----:----|

     T  L  E  R  H  P  T  R  S  G  K  G  A  D  T  D  G  C  Q  A     F1
      H  W  S  G  I  P  H  A  V  A  R  A  Q  T  P  M  A  A  R  L    F2
       T  G  A  A  S  H  T  Q  W  Q  G  R  R  H  R  W  L  P  G  Y    F3
3301 ACACTGGAGCGGCATCCCACACGCAGTGGCAAGGGCGCAGACACCGATGGCTGCCAGGCT 3360
     ----:----|----:----|----:----|----:----|----:----|----:----|

     M  Q  R  V  Q  W  L  R  G  T  E  N  W  F  K  E  E  W  T  F     F1
      C  N  E  C  N  G  S  E  A  R  R  T  G  S  K  K  S  G  L  S    F2
       A  T  S  A  M  A  Q  R  H  G  E  L  V  Q  R  R  V  D  F  Q    F3
3361 ATGCAACGAGTGCAATGGCTCAGAGGCACGGAGAACTGGTTCAAAGAAGAGTGGACTTTC 3420
     ----:----|----:----|----:----|----:----|----:----|----:----|

     S  K  P  N  S  L  S  A  K  F  D  Y  R  G  E  V  Q  P  P  F     F1
      A  S  P  T  A  S  A  Q  S  S  T  I  G  A  R  F  S  L  L  S    F2
       Q  A  Q  Q  P  Q  R  K  V  R  L  *  G  R  G  S  A  S  F  Q    F3
3421 AGCAAGCCCAACAGCCTCAGCGCAAAGTTCGACTATAGGGGCGAGGTTCAGCCTCCTTTC 3480
     ----:----|----:----|----:----|----:----|----:----|----:----|

     S  Q  S  N  V  L  R  F  L  R  E  P  S  R  L  R  R  G  S  A     F1
      V  R  A  T  C  C  A  S  S  E  N  L  V  A  *  G  E  V  Q  P    F2
       S  E  Q  R  V  A  L  P  Q  R  T  *  S  L  E  A  R  F  S  L    F3
3481 AGTCAGAGCAACGTGTTGCGCTTCCTCAGAGAACCTAGTCGCTTGAGGCGAGGTTCAGCC 3540
```

```
         ----:----|----:----|----:----|----:----|----:----|----:----|

         S  F  Q  S  E  P  E  Q  R  V  A  L  P  Q  R                  F1
          P  F  S  Q  S  Q  S  N  V  L  R  F  L  R  X                  F2
           L  S  V  R  A  R  A  T  C  C  A  S  S  E  X                  F3
    3541 TCCTTTCAGTCAGAGCCAGAGCAACGTGTTGCGCTTCCTCAGAGA 3585
         ----:----|----:----|----:----|----:----|----:
```

>Gam_VTG_ID17046

MKLIFFVALLGLGCALALQAPVEYEYKYEARIATGMQSALPEYAGTGIKADVRITMQEDG

SGEVQLKNVEVGEGMAVLPDNGREVLPIDYKPIEDKSIERPYKAYFRKQDGETVLEVPED

DVLWVTNVRRGVANFFRIAPIVQPGSVPPPREGSELLVGYSRIMDTFAGRCRVEYTITPA

AISDVRSGIREDIQSRRAPKGPGGLQVKLEMTYMLCSLPTCCAVFHDFSTCSNRVFLSQT

GAGNLTNVTSSSYRDVSARSHVGEYILKGKPGQFRMERGEVESAVILQPFGFNSAKVVIM

SNQSMELVSAKPVPRNVRSSAQNLKSIKTWNHQLESPKPKPEGFAPRFIQYKNQNEPEYS

PIFYQRLLISGFRMNDEGVQNNLLDRVKITIEKAAERCRLQNKDSDAEPTTFRGTEDGRI

AQGMMDAMHLLRALNHEKVQELDKKMQEHQDKLVREIFIDALAAAASGPTVKVLLRKIRN

HELSDNRTASIFFNLANNIRSAQLSLTTSDLRIIEELVDFAMSLEEEKNVTASLLVKNNV

ALIVRRIRICVSSQIKKYSHMMYAFGSKMCDRTKIVNKYLPHLEQKIESSKITWVKILNV

MAAADLGARESIPILEKVARGDVSSDIFVRFAAIHGVCDNSSRDKIFNVLQPIAESKVED

YRIRQVAYDSLLTFMPNVTFIQRMIFSSYDEPSQQVQSYISWVVDSLSRIEDPRYRLLST

AARNALALTRYMPASLNRPVDWIFSRVTYTMTRSAARAVRVVSQYKSSFPREIYSFVAGK

FGRFGIPIVETITLHSATARLCTASSGRLRSAAEEFSRKKGPSQQEEGTILCPRAGETAK

APRPVPIHAWRPPPQHCRELTNGSQTFTTSEVFLPLTNELYQQTKGYLEAALRKLGQGEE

QKVLFFKYSSSI

Conclusion

Chez les organismes modèles, les études protéogénomiques ont pour objectif de corriger l'annotation du génome par la validation des gènes prédits par bio-informatique, la découverte de gènes non annotés et la correction des codons d'initiation, de terminaison ou des ORF (Ansong *et al.* 2008). Chez les espèces dépourvues de génome de référence et uniquement étudiées par transcriptomique, l'intégration des données biochimiques permet d'améliorer l'annotation structurale de la protéine sur toute sa longueur. Par ailleurs, chez ces espèces, une autre stratégie

prometteuse réside dans l'amélioration de la qualité générale de l'assemblage par la combinaison de différentes techniques RNA-Seq. Cette approche a pour but d'augmenter la redondance de couverture de séquençage du transcriptome par l'utilisation de techniques du type Roche/454 GS FLX et Illumina-Solexa permettant respectivement le séquençage de brins longs et courts suivi d'un assemblage hydride (He et al. 2012).

Ansong, C., S. O. Purvine, et al. (2008). "Proteogenomics: needs and roles to be filled by proteomics in genome annotation." Brief Funct Genomic Proteomic **7**(1): 50-62.

Gibson, A. K., Z. Smith, et al. (2013). "Why so many unknown genes? Partitioning orphans from a representative transcriptome of the lone star tick Amblyomma americanum." BMC Genomics: 135.

He, R., M. J. Kim, et al. (2012). "Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, Phragmites australis (Poaceae), reveals genes involved in invasiveness and rhizome specificity." American Journal of Botany **99**(2): 232-247.

Wang, X., R. J. C. Slebos, et al. (2012). "Protein identification using customized protein sequence databases derived from RNA-seq data." Journal of Proteome Research **11**(2): 1009-1017.

CHAPITRE IV : Impacts des xénobiotiques sur la reproduction masculine

Publication 5 :

Proteomic investigation of male *Gammarus fossarum*, a freshwater crustacean, in response to endocrine disruptors.

Version auteur de l'article publié dans *Journal of Proteome Research*, 2015, 14 (1), 292-303
La version finale est disponible sur http://pubs.acs.org/doi/abs/10.1021/pr500984z

Judith Trapp [a,b], Jean Armengaud [b], Olivier Pible [b], Jean-Charles Gaillard [b], Khedidja Abbaci [a], Yassine Habtoul [a], Arnaud Chaumot [a], Olivier Geffard [a]

[a] Irstea, Unité de Recherche MALY, Laboratoire d'écotoxicologie, CS70077, F-69626 Villeurbanne, France
[b] CEA, DSV, IBEB, Lab. Biochim. System. Perturb., Bagnols-sur-Ceze, F-30207, France

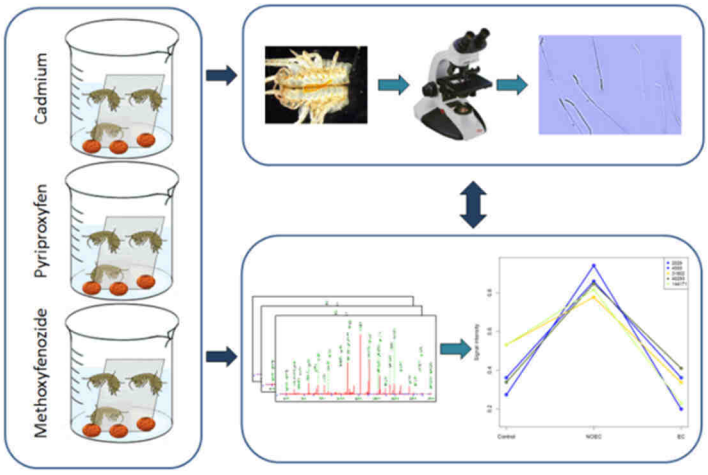
Résumé :

Alors que la baisse de la fertilité humaine masculine due à la pollution est un problème planétaire, peu d'attention est portée à ses manifestations et causes au travers du règne animal. Les arthropodes constituent le taxon le plus riche en termes d'espèces et sont présents au sein de tous les écosystèmes terrestres et aquatiques. Au cours de l'évolution, les molécules impliquées dans la régulation de leur système endocrinien ont fortement divergées face à leurs homologues vertébrés. Aussi, les arthropodes peuvent présenter des sensibilités différentes face aux perturbateurs endocriniens. Dans cette étude, l'altération de la fertilité masculine chez un crustacé, *Gammarus fossarum*, espèce populaire dans l'évaluation du risque des milieux aquatiques d'eaux douces, a été étudiée suite à une exposition en laboratoire à trois différents xénobiotiques, le cadmium, le methoxyfenozide et le pyriproxyfen. L'intégrité du processus reproductif a été évaluée par le biais de marqueurs de qualité spermatique. Pour chaque substance, le protéome des gonades a été caractérisé par protéomique shotgun afin d'étudier l'effet de ces xénobiotiques au niveau moléculaire. Les modifications sur un total de 871 protéines ont été mesurées en réponse à la pression toxique. Un effet drastique a été observé sur la production de spermatozoïdes, suivant une relation dose réponse. Alors que chez les vertébrés mâles, l'exposition à des PE induit généralement la synthèse de protéines spécifiques de la femelle, notre étude sur *G. fossarum* a plutôt mis en évidence la modulation de protéines spécifiques du mâle.

Mots-clés : Infertilité, perturbation endocrine, crustacé, protéomique shotgun, écotoxicologie

Proteomic investigation of male *Gammarus fossarum*, a freshwater crustacean, in response to endocrine disruptors.

Abstract



While the decrease in human sperm count in response to pollutants is a worldwide concern, little attention is being devoted to its causes and occurrence in the biodiversity of the animal kingdom. Arthropoda is the most species-rich phyla, inhabiting all aquatic and terrestrial ecosystems. During evolution, key molecular players of the arthropod endocrine system have diverged from the vertebrate counterparts. Consequently, arthropods may have different sensitivities towards endocrine disrupting chemicals (EDC). Here, alteration of sperm quality in a crustacean, *Gammarus fossarum*, a popular organism in freshwater risk assessment, was investigated after laboratory exposure to various concentrations of three different xenobiotics, cadmium, methoxyfenozide and pyriproxyfen. The integrity of the reproductive process was assessed by means of sperm quality markers. For each substance, semi-quantitative/relative proteomics based on spectral counting procedure was carried out on male gonads to observe the biological impact. The changes in a total of 871 proteins were monitored in response to toxic pressure. A drastic effect was observed on spermatozoon production, with a dose-response relationship. While exposure to EDCs leads to strong modulations of male-specific proteins in testis, no induction of female-specific proteins was noted. Also, a significant portion of orphans proved to be sensitive to toxic stress.

Keywords: Ecotoxicology, crustacean, shotgun proteomics, expression profile, endocrine disruption

1. INTRODUCTION

Because spermatozoa perform the unique and critical function of propagating DNA for the development of the next generation, the global decline in human sperm count observed over the past half century is of great concern (1), with environmental contamination being particularly incriminated. Compared to oocytes, spermatozoa are highly sensitive to xenobiotics, due to their lack of defensive systems such as DNA repair or antioxidant enzymes (2). Male reproductive disorders have been reported for several vertebrate phyla, as exemplified by testicular abnormalities and impaired fecundity in US Lake Apopka alligators (3), or by feminized fish found downstream of a wastewater treatment plant in the UK (4). Among all the contaminants that are able to alter sperm production, endocrine-disrupting chemicals (EDC), especially estrogenic-like or anti-androgen-like compounds, are diverse and abundant in the environment (5).

As highlighted by Lewis and Ford (6), invertebrate organisms have been considerably less well studied than vertebrates, despite their ecological importance. Arthropods (*i.e.* insects, crustaceans, and arachnids) make up more than 80% of all described animal species (7), inhabiting all aquatic and terrestrial environments. There is a patent lack of knowledge regarding endocrine regulation of reproduction in small aquatic invertebrates and in ecotoxicologically relevant species, such as amphipods, mysids and daphnids. As reported by Lye et *al*. (8), some evidence suggests that crustaceans are less vulnerable to known vertebrate EDCs, such as those that mimic estrogen compounds, compared to other invertebrates, including echinoderms and mollusks. The hormonal regulation of arthropods is not based on vertebrate-like steroids, but on ecdysteroids, juvenoid hormones (JH), and for males, the androgenic gland hormone (9). This high divergence of endocrine systems suggests that possible invertebrate-targeted EDCs are different and exhibit a more diverse chemical nature than vertebrate EDCs. In arthropod pest control management, some pesticides are designed to specifically interact with arthropod hormone nuclear receptors (*i.e.* ecdysteroids and JH nuclear receptors) through agonist or antagonist action. Thus these molecules act as endocrine disruptors in a variety of targeted insect species by disrupting embryogenesis and preventing metamorphosis. Different observations suggest that wild populations of aquatic invertebrates, especially crustaceans, can be affected as nontarget organisms (10-12). In the same way, field evidence of alterations of sexual characteristics has been reported, especially in amphipods. For example, Jungmann *et al.* (13) and Ford *et al.* (14) collected *Gammarus fossarum* and *Echinogammarus marinus* organisms, respectively, that exhibited intersex phenotypes, *i.e*. the co-

occurrence of male and female sexual characteristics. In the same vein, Schirling *et al.* (15) observed abnormal structures of oocytes in *Gammarus pulex* and *G. fossarum* collected from streams impacted by sewage treatment plant effluents. Crustacean amphipods are popular organisms in ecotoxicological risk assessment and are currently used as sentinel organisms because of their ecological relevance in relation to their key role in the food chain and their contribution to the decomposition process (leaf litter breakdown) (16). In addition to its use in the laboratory, these organisms may also be easily employed in *in situ* bioassays for sensing bioavailable pollutants and their effects (17) through the assessment of associated toxic effects, *i.e*. induction of detoxification mechanisms (18), neurotoxicity (19), or energy acquisition impairments (20). Furthermore, physiological parameters related to the gammarid reproductive process, *i.e.* spermatogenesis (21), and to the female reproductive cycle (22), have been well documented.

To date, no specific markers of exposure to endocrine disruptors were developed for aquatic invertebrates, and particularly for amphipods. In recent decades, two main strategies have been employed to improve the assessment of the impact of EDCs on crustaceans. First, bioassays have been proposed for measuring the impact of pollutants on various physiological processes that are under endocrine regulation, such as growth, molting, embryonic development, vitellogenesis, and fertility. In this context, for the water flea, *Daphnia magna* (Crustacea, Cladocera), with parthenogenic reproduction, the offspring sex ratio was proposed as an end point to screen chemicals for juvenoid-related endocrine activity (23). Geffard et *al.* (22) also developed a reproductive toxicity test in *G. fossarum*, for which some synchronous physiological events related to reproduction are measured, such as molting cycle, follicle growth, oocyte production and embryonic development. Geffard et *al.* proposed to use the specific effect on either molting or follicle growth, and their desynchronization, as an indicator of a potential specific exposure to EDCs. Another strategy has been to rely on the development of molecular markers. However, because of the lack of genomic and/or proteomic data for aquatic invertebrates, specific and direct methods for the identification and quantification of protein biomarkers are scarce. Until now, biomarkers employed for invertebrate species are rather the result of a direct transposition of assays previously developed and validated in vertebrates, for which molecular biology is better understood (24). Hence, based on developments in vertebrates, research groups in aquatic ecotoxicology have mainly focused their activities on the measurement of vitellogenin, in order to propose this protein as a specific EDC biomarker in invertebrates. Vitellogenin and vitellogenin-like proteins (Vtg) are

precursor molecules of yolk proteins that provide the energy reserves needed for embryonic development in oviparous organisms. In vertebrates, Vtg production is under the control of the estrogen pathway and its levels are high in sexually mature females, displaying fluctuations during gametogenesis (25). Conversely, the Vtg content is generally very low in juveniles and males, and in these organisms its induction has been employed as a robust biomarker for diagnosing the presence of estrogen mimetics and their associated effects, *i.e*. feminization (25). With regard to research performed in fish, Vtg has been monitored in several invertebrate species as a biomarker of EDC exposure (26) and recently, a selected reaction monitoring assay for the quantitation of *G. fossarum* Vtg was developed. Nevertheless, while the Vtg protein in *G. fossarum* females is a reliable indicator of egg quality demonstrating functionality of this protein in the reproductive process (27), its potential use as an EDC biomarker in males is compromised. Indeed, in laboratory experiments and in a large-scale field experiment using caged organisms, a strong inter-individual variability, very low or no induction under contamination exposure, and a strong impact of unidentified environmental factors, were shown (28). Similar observations were obtained for the marine amphipod, *Echniogammarus marinus*, where expression of the orthologous gene was monitored in females, males and intersex males (29). Such findings and phylogenetic analysis of vitellogenin in metazoans (30) highlight the difference between invertebrate and vertebrate Vtgs, demonstrating pleiotropic functions not strictly related to reproduction. These studies underline the need for the specific identification of new key molecular players in gammarid reproductive biology, for use as relevant biomarkers in ecotoxicology. We, thus, recently proposed a proteogenomic strategy for discovering proteins involved in *G. fossarum* reproductive function (31). A shotgun proteomic analysis was carried out on the proteomes of three major tissues involved in the organism's reproductive function: the male and female reproductive systems, and the cephalon. Proteins were identified based on a generic RNA-seq-derived protein sequence database. Based on the work of Chen and *al*. (32), key proteins with strong sexual dimorphism were listed by performing comparative proteomics between the male and female reproductive organs. Additionally, protein expression profiles over seven spermatogenesis stages were carried out to pinpoint new proteins involved in spermatogenesis, uncovering a substantial proportion of orphan proteins (31).

The present study aimed to discover potential new biomarkers for the monitoring of reproductive disorders and EDC exposure in male gammarids. In mammals, proteomics showed to be a valuable

tool for addressing normal and pathological spermatogenesis (33, 34) and in non-model fish, proteomic profiling proved to be a powerful approach for studying endocrine disruption (35). Thus, we simultaneously investigated, in response to a metal (cadmium, Cd) and two xenobiotics (methoxyfenozide, Met, and pyriproxyfen, Pyr), both the proteome modifications occurring in the male reproductive organ and the alterations in reproductive health by means of sperm quality parameters. Here, we employed a semi-quantitative/relative proteomics based on spectral counting procedure as it proved to be a straightforward strategy for characterizing proteome alteration in toxicoproteomics (36, 37). Cd is a nonessential heavy metal which is highly toxic to gammarids and is associated with deleterious effects on iono-osmoregulation (38), feeding rate, vitellogenesis, and embryonic development (22). Met and Pyr are potent arthropod EDCs, employed as growth regulators in agriculture, the former targeting ecdysteroid signaling (molt hormone) while the latter interferes with JH signaling. Experiments were conducted using three different concentrations of each pollutant. First, sperm quality markers in response to toxic exposure were assessed by the numeration of spermatids (*i.e*. immature reproductive cells) and spermatozoa. For each substance tested, based on the measured physiological endpoints, proteomic investigations were then carried out at two different concentrations, a no observed effect concentration (NOEC) and an effect concentration (EC) in terms of the reproductive physiology, and compared with the corresponding control condition. Protein dynamics over the different concentrations were analyzed for the identification of proteins modulated by toxic exposure. We identified several proteins as candidate biomarkers of reproductive disorders and EDC exposure – predominantly orphans – and we discuss their perspectives in terms of ecotoxicological risk assessment.

2. EXPERIMENTAL SECTION

Exposure of gammarids to toxic compounds. Gammarids were collected from a river in mid-Eastern France and acclimatized to laboratory conditions as previously described (31). Males of homogenous size (8-11 mm) paired with females at D2 molt stages (*i.e.* hatched juveniles in brood pouches, visible oocytes) (22) were selected and isolated in a 30 L tank. Just after copulation (separation of the amplexus), males at the beginning of spermatogenesis were collected for chemical exposure during two consecutive spermatogenesis cycles (15 days of exposure). Organisms were exposed to cadmium (Cd) (0.3, 1, and 3 $\mu g.L^{-1}$), methoxyfenozide (Met) (0.001, 0.1,

and 10 µg.L^{-1}) and pyriproxyfen (Pyr) (0.5, 5, and 50 µg.L^{-1}). The three concentration ranges were chosen based on previous work (22), (27), (39). Stock solutions for organic compounds were prepared in acetone. Contaminated media were obtained by dilution of stock solutions into uncontaminated, drilled groundwater. Solvent-free (*i.e.* water) and solvent (*i.e.* 0.005% acetone) controls were also included. For each condition tested, 21 males of similar size were placed into three 500 mL glass (Cd) or polyethylene (Met and Pyr) beakers (seven animals per beaker), under constant oxygenation. A piece of net (mesh size: 200 µm; 6 x 5 cm) was added into each beaker as a resting surface. Organisms were fed *ad libitum* with conditioned alder leaves (*Alnus glutinosa*). Exposure media and alder leaves were renewed every two days, while water quality parameters (pH, conductivity, temperature and dissolved oxygen) and survival were recorded daily. After five days of exposure, sexually mature females at D2 stage were added into the beakers (an equal number of females as surviving males). The occurrence of copulation was checked daily. When at least 70% of the males had copulated (Day 7 of the experiment), single females and males still in amplexus were discarded. At the end of the exposure (Day 15 of the experiment), organisms were culled. Male gonads were sampled under stereomicroscopic magnification using fine forceps. The testis of five animals were analyzed by shotgun proteomics, while the remaining organisms (n=7-10) were culled for the assessment of sperm quality markers.

Sperm quality indicators. Under stereomicroscopic magnification, gonads were dilacerated with a scalpel, transferred into a microtube containing 30 µL phosphate buffered saline (10 mM, pH 7.4 + 0.3% bovine serum albumin), kept on ice, and gently mixed 20 times with a micropipette, as described by Lacaze et al. (40). For spermatozoon numeration, 20 µL well-mixed spermatozoon suspension was deposited on the surface of a hemocytometer and observed under optic microscopy (Leica DM 2500, magnification x100). The total number of cells was calculated by multiplying the sperm count per µL by the volume of sperm suspension. Statistical analyses were carried out with R software (41). Normality and variance homogeneity were checked using the Shapiro-Wilk and Bartlett tests, respectively. Intergroup comparisons were then performed by the ANOVA test and the *post hoc* Tukey HSD test for parametric conditions, as well as the Kruskal-Wallis test, and associated *post hoc* test for nonparametric conditions.

Shotgun LC-MS/MS analysis. For shotgun proteomics, each testis to be analyzed was directly dissolved in 40 µL LDS sample buffer (Invitrogen). The sample was subjected to 1 minute of sonication (transonic 780H sonicator) and boiled for 5 minutes at 95 °C. The testis was totally

dissolved into the LDS sample buffer and 35 µL of each sample were then subjected to SDS-PAGE on a 10-well NuPAGE 4–12% gradient (Invitrogen) for 10 minutes at 150 V with MES buffer. Protein samples (35 µL) were then subjected to SDS-PAGE on a 10-well NuPAGE 4–12% gradient (Invitrogen) for 10 minutes at 150 V with MES buffer. Gels were stained with Coomassie Blue Safe stain (Invitrogen) and destained overnight with water. Protein concentration was not assessed as the testis proteins account only for 4 µg per organ in average. The variation in protein amount per gonad tissues is relatively low as judged from the Coomassie staining**.** The whole protein content from each well was extracted as a single polyacrylamide band and processed for further destaining and iodoacetamide treatment as previously described. Proteins were proteolyzed with sequencing grade trypsin (Roche) using 0.01% ProteaseMAX surfactant (Promega). The resulting peptide mixtures were analyzed in data-dependent mode with an LTQ-Orbitrap XL hybrid mass spectrometer (ThermoFisher) coupled to an UltiMate 3000 LC system (Dionex-LC Packings), operated as previously described. In brief, 2 µL from each sample were loaded and first desalted online on a reverse-phase precolumn, C18 PepMap 100 column (LC Packings). Then, the peptides were separated using a 90 min gradient on a nanoscale C18 PepMapTM 100 capillary column (LC Packings) at a flow rate of 0.3 µL/min with a gradient of CH_3CN, 0.1% formic acid prior to injection into the ion trap mass spectrometer. Full-scan mass spectra were measured from m/z 300 to 1800 using the TOP3 strategy. In brief, a scan cycle was initiated with a full scan of high mass accuracy in the Orbitrap at a resolution of 30,000 and with the lock mass option, which was followed by MS/MS scans in the linear ion trap on the three most abundant precursor ions with a signal of at least 15,000 and potential charge states of 2+ and 3+. A dynamic exclusion of previously selected ions was set for a 60 sec duration.

Protein identification and spectral counting quantitation. Peak lists were generated with the Mascot Daemon software (version 2.3.2; Matrix Science) using the extract_msn.exe data import filter (Thermo). Data import filter options were set to 400 (minimum mass), 5,000 (maximum mass), 0 (grouping tolerance), 0 (intermediate scans), and 1,000 (threshold), as described previously (42). MS/MS spectra were assigned to peptide sequences with the Mascot Daemon 2.3.2 search engine (Matrix Science) against the customized RNA-seq-derived database containing 1,311,444 sequences and totalling 289,084,257 amino acids, as previously described (31). MS/MS spectra were assigned with the following parameters: full-trypsin specificity, maximum of two missed cleavages, mass tolerances of 5 ppm on the parent ion and 0.5 Da on the MS/MS, static modification of

carboxyamidomethylated cysteine (+57.0215), and oxidized methionine (+15.9949) as dynamic modification. All peptide matches with a MASCOT peptide score below a p value of 0.05 were filtered and assigned to a protein according the principle of parsimony (42). A protein was considered to be validated when at least two different peptides were detected. The false discovery rate for protein identification was estimated by employing the decoy search option of MASCOT (Matrix Science) to be below 0.1%. Each xenobiotic-dependent proteome response was analyzed with the TrendQuest module of the PatternLab program (43). For this, spectral counts were not normalized. Proteins observed in at least five of the replicate samples were considered for clustering; the minimum signal required was five and the minimum items per cluster was four.

MS/MS data repository. The mass spectrometry proteomics data have been deposited with the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the PRIDE partner repository (44), with the dataset identifier PXD001267.

3. RESULTS

Experimental strategy for comparative proteomics of toxicant-exposed animals. Figure 1 presents the global strategy employed for the assessment of male reproductive disorders in *G. fossarum* in response to toxic pressure. For each xenobiotic, male organisms were exposed to three concentrations during two consecutive spermatogenesis cycles, *i.e.* two weeks, and compared with controls. During the exposure, no significant mortality was observed for the overall conditions, with survival values higher than 85%. At the end of the experiment, when the gonads were retrieved under stereomicroscope magnification, no intersex individuals were observed. We then constructed a two-step strategy for the analysis of the biological parameters. First, the relevance and sensitivity of sperm quality markers in response to toxicant were assessed by the numeration of spermatids and spermatozoa. Then, for each tested substance, we defined the most interesting condition to analyze the proteomes of the organisms where sperm quality markers revealed reproductive disorders, *i.e*. the EC. Furthermore, as modifications at the molecular level are supposed to occur earlier than those visible at the cellular level, we also included in our study a NOEC for investigating early proteome modifications in response to toxic stress. Proteomic analyses were carried out on the testis of individual animals with a semi-quantitative/relative proteomics

based on spectral counting procedure. Protein dynamics over the different concentrations were analyzed by clustering to reveal proteins modulated by toxic exposure.

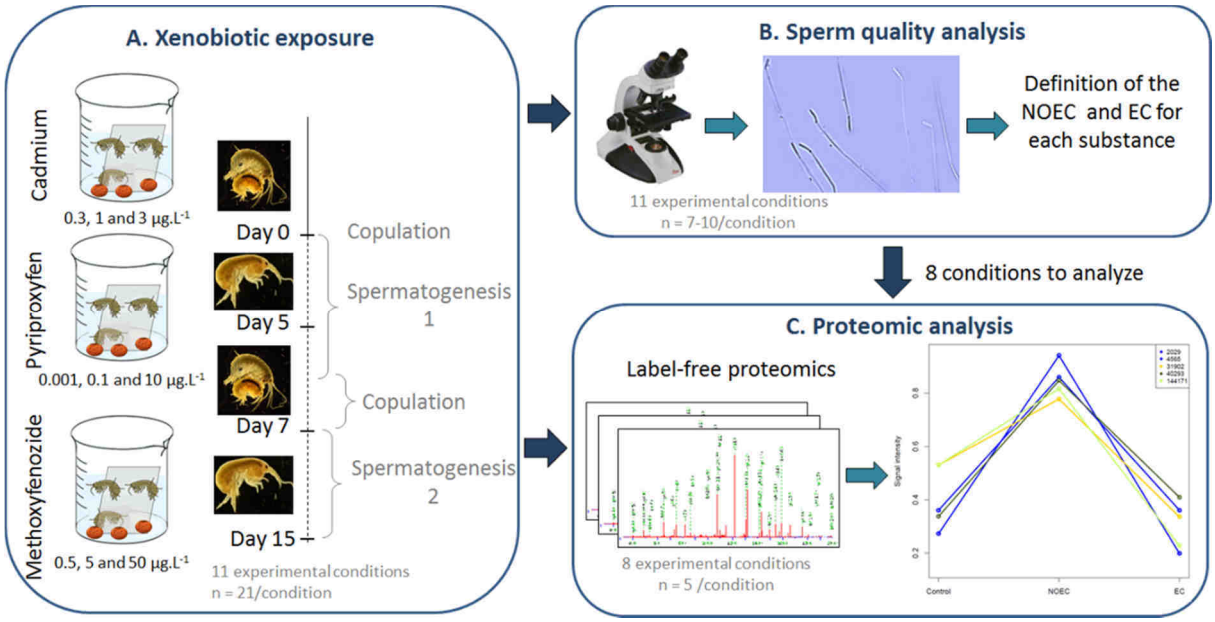


Figure 1: General strategy for male reproductive disorders assessment in *G. fossarum.*

(A) Male organisms are exposed during two consecutive spermatogenesis cycles to three different xenobiotics, each at three different concentrations, including one solvent control and one solvent-free control. (B) Sperm quality analysis by numeration of spermatids and spermatozoa. (C) Proteome analysis on the most interesting conditions for each chemical by shotgun proteomics and clustering procedures for the identification of proteome dynamics.

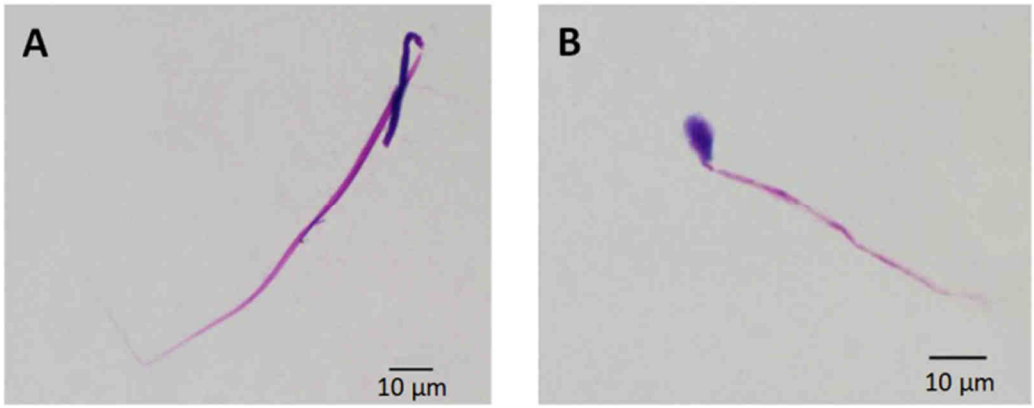


Figure 2: Microscopic observations of *G. fossarum* male reproductive cells (x100).

(A) Spermatozoa and (B) Spermatids after May-Grunwald-Giemsa staining.

Impacts of Cd, Met and Pyr on reproductive physiology.

Sperm quality markers. Two flagellated cell types were counted under optic microscopy: the egg-shaped spermatids and the oblong-shaped spermatozoa. Their respective morphologies are presented in **Figure 2 (Panel A** and **B**, respectively). No specific morphological abnormalities were observed upon toxic pressure.

The numbers of spermatids in males exposed to the three xenobiotics are presented in **Figure 3 (Panel A)**. In the control condition, the mean spermatids number was $9.5x10^2 \pm 2.3x10^2$ and the mean spermatozoon number was $100x10^2 \pm 28x10^2$. Thus, spermatids only represent about 9% of the flagellated cells in the gammarid gonad. Cd and Met did not impact significantly on spermatid production (Kruskal-Wallis rank sum test, $p>0.05$). Conversely, a significant impact on the production of spermatids was observed for the organisms exposed to Pyr (ANOVA, $p=0.03$). *Post hoc* comparisons showed that only males exposed to 50 $\mu g.L^{-1}$ Pyr had significantly fewer spermatids (mean ± S.E.M.: $6.7x10^2 \pm 4.5x10^2$) compared with the animals from the solvent control ($13x10^2 \pm 4.2x10^2$) (Tukey HSD, $p=0.05$). Whereas overall spermatid production was not significantly impacted upon exposure of any of the three toxicants, spermatozoon production was shown to be more sensitive. The numbers of spermatozoa in males exposed to the three xenobiotics are presented in **Figure 3 (Panel B).** In the control conditions, the average spermatozoon numbers (mean ± S.E.M) were $100x10^2 \pm 28x10^2$ and $88x10^2 \pm 10x10^2$ in the solvent-free control and solvent-control, respectively. For Met, sperm production was characterized by strong inter-individual variability, especially for the lowest concentration of toxicant (0.001 $\mu g.L^{-1}$, mean ± S.E.M.: $88x10^2 \pm 53x10^2$). Although a slight downward trend was observed, the spermatozoon count for organisms exposed to this substance was not significantly impacted (Kruskal-Wallis rank sum test, $p=0.07$). A significant impact on spermatozoon production was observed in organisms exposed to Cd and Pyr (Kruskal-Wallis rank sum test, $p=0.6x10^{-3}$ and $p=1.2x10^{-5}$, respectively), both in a dose-dependent relationship. Compared with control conditions, the mean inhibition of spermatozoon production ranged from 58% (1 $\mu g.L^{-1}$) to 75% (3 $\mu g.L^{-1}$) for Cd, while for Pyr it ranged from 40% (5 $\mu g.L^{-1}$) to 73% (50 $\mu g.L^{-1}$).

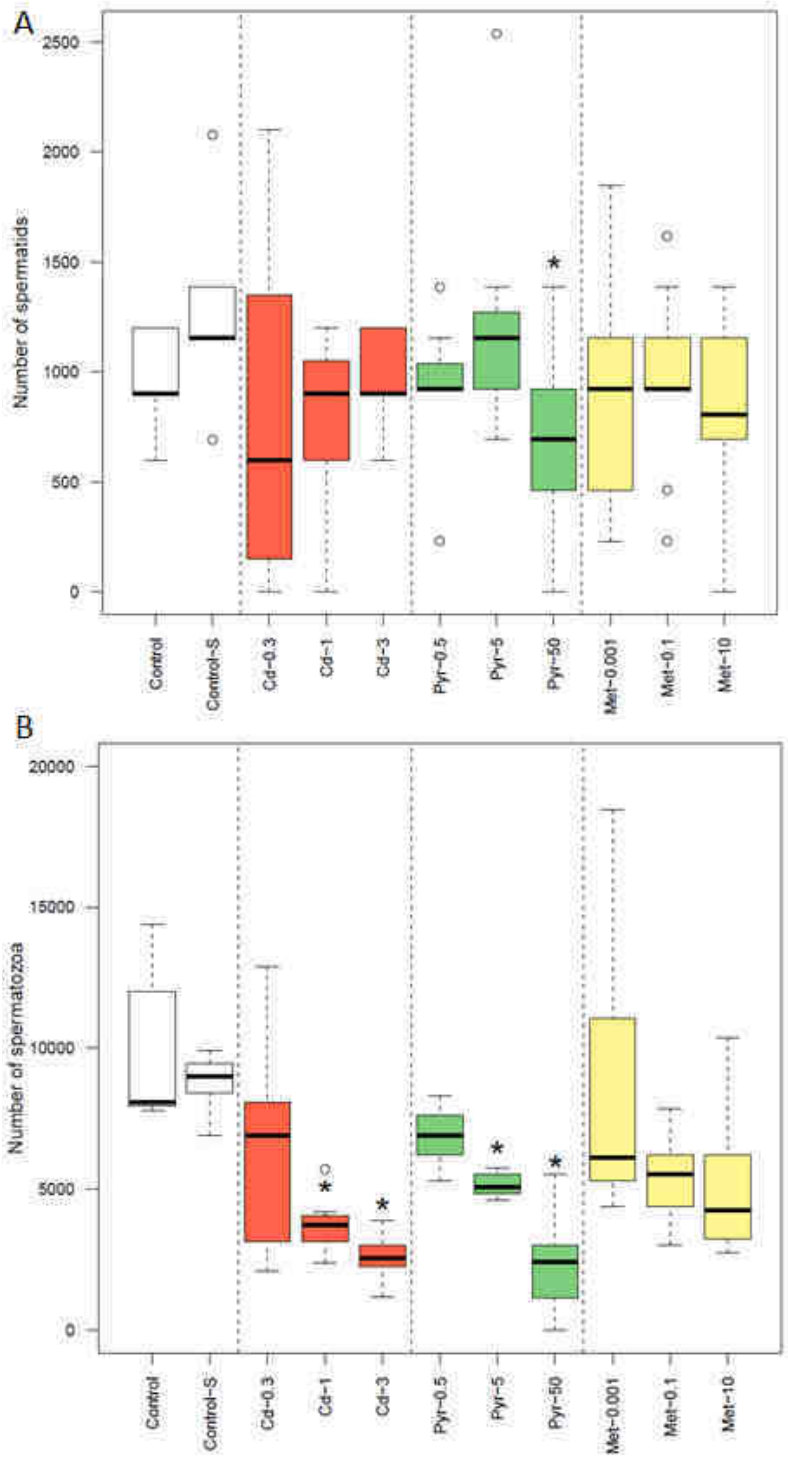


Figure 3: Number of male reproductive cells in *G. fossarum* gonads after 15 days exposure to pollutants.

Number of (A) Spermatid and (B) Spermatozoa. Control and solvent-free control (Control-S) in white; Cd (0.3, 1, and 3 µg.L-1) in red; Pyr (0.5, 5 and 50 µg.L-1) in green and Met (0.001, 0.1, and 10 µg.L-1) in yellow. Boxes extend from the first to the third quartile with a bold segment for the median value; the whiskers extend to the most extreme data points, no more than 1.5 times the interquartile range. Data points of more than 1.5 times the interquartile range are represented by empty circles. The p-values of the global effect of treatment are reported in the text. * = significant difference ($p < 0.05$) from the overall control groups (solvent-free control and solvent control). n = 7-10 males per condition.

Toxic compound concentrations for shotgun proteomics. For Cd, the lowest concentration of 0.3 $\mu g.L^{-1}$ was chosen as the NOEC on the basis of the observation of sperm quality markers, while the highest concentration of 3 $\mu g.L^{-1}$ was set as the EC. In the same way, for Pyr, the NOEC and EC were set as 0.5 $\mu g.L^{-1}$ and 50 $\mu g.L^{-1}$, respectively. For Met, such clear-cut responses were not observed, but rather there was a slight downward trend. Thus, we chose to study the testis proteomes of organisms exposed to two concentrations of Met, *i.e.* 0.001 and 0.1 $\mu g.L^{-1}$. As a result, accounting for the solvent-free and solvent controls, proteomic analysis was carried out on a total of eight conditions. For each condition, the male gonads of five exposed animals were analyzed in order to obtain statistically robust results.

Impact of pollutants on the testis proteome.

Global view of the testis proteome. After electrophoresis under denaturing conditions, the proteome of each of the forty samples was retrieved in a single polyacrylamide gel band and digested in the gel. The resulting peptide mixtures were analyzed by nanoLC-MS/MS with a high resolution Orbitrap hybrid mass spectrometer. This experiment led to a global dataset of 237,950 MS/MS spectra recorded. A total of 57,558 MS/MS spectra could be assigned after querying an RNA-seq-derived protein sequence database, indicating the existence of 3,020 unique peptide sequences. These allowed the identification of 871 mRNA-translated products, among which 66 (8%) orphans were listed. A complete inventory of peptides and proteins identified in this study is provided in the Supporting Information (Appendix 1 and 2), as well as their quantitation based on spectral counts and normalized spectral abundance factors (NSAF). Remarkably, four abundant proteins made up 10% of the organ proteome: the histones H2A (ID_13250; mean NSAF percentage 4.1 ± 1.1) and H4 (ID_100349; mean NSAF percentage 2.3 ± 0.6), an arginine kinase (ID_14404; mean NSAF percentage 2.7 ± 0.5), and a protein with no predicted function (ID_128511; mean NSAF percentage 2.0 ± 0.8). The reproducibility between each replicates was evaluated from the total number of assigned spectra per sample (Appendix 2), 1200 in average, as shown in **Figure 4**. Based on ANOVA analysis, no significant differences were observed between the different conditions ($p>0.05$).

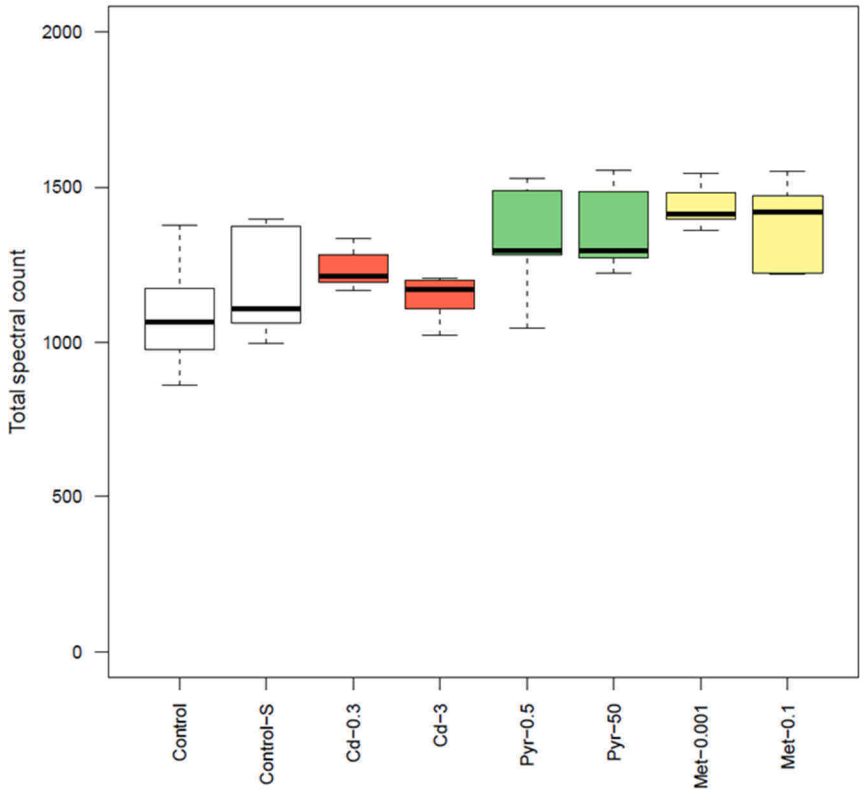


Figure 4: Analysis of reproductibility between replicates for each conditions in terms of total number of assigned spectra per sample

Boxes extend from the first to the third quartile with a bold segment for the median value; the whiskers extend to the most extreme data points, no more than 1.5 times the interquartile range. Data points of more than 1.5 times the interquartile range are represented by empty circles. n=5 males per condition.

Identification of temporal variations of testis proteomes. For each toxicant, the proteome dynamics along the different concentrations were analyzed using the clustering method specifically developed for shotgun proteomics (45). Across all the experiments, a total of 70 proteins showed significant clustered temporal variations. As displayed in **Figure 5**, among these proteins, 44 were common to the three toxicants, twelve were shared by two conditions (three Cd-Met, three Cd-Pyr and six Pyr-Met) and fourteen were unique to a single condition. For each substance, a total of six clustered temporal patterns were recorded. The characteristics of the corresponding proteins are listed in Appendix 2. Among the 70 proteins exhibiting temporal variations, 14 are devoid of functional annotation (20%) because of the low sequence similarity to known proteins, with three orphans. An important proportion is involved in cytoskeletal structure maintenance (22 proteins) and energy metabolism (10 proteins). The others are involved in calcium homeostasis (four proteins), heat-shock response (three proteins) or immunity (three proteins). Interestingly one

protein, functionally annotated as vitellogenin (ID_198421), is present in clusters Cd_A and Pyr_D. Based on the phylogenetic analysis reported by Hayward et *al.* (30), this protein is included among the decapod clottable protein family, which is involved in immunity.

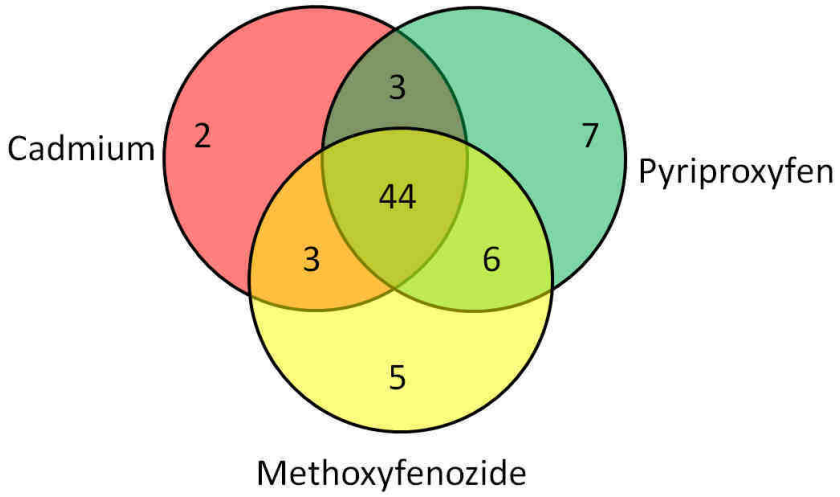


Figure 5: Differential expression of the proteins that showed clustered temporal variations in response to xenobiotic exposure.

In the Cd series, a total of 630 proteins were compared, resulting in six TrendQuest clusters comprising 55 proteins: fourteen in cluster A, ten in cluster B, nine in cluster C, nine in cluster D, five in cluster E and five proteins in cluster F. In the Pyr series, 722 proteins were compared, resulting in five clusters containing a total of 63 proteins: nineteen proteins in cluster A, sixteen in cluster B, twelve in cluster C, five proteins in cluster D, four in cluster E and four in cluster F. Finally, in the Met series, 716 proteins were compared, resulting in six clusters containing a total of 61 proteins: fifteen in cluster A, fifteen in cluster B, eleven in cluster C, eight in cluster D, five in cluster E and four in cluster F. **Figure 6** shows, for each xenobiotic, the cluster corresponding to the trend where proteins are first induced in response to chemical stress and then decrease when physiological effects appear, and thus which could be specific of an early effect of toxicants. The eighteen different clusters are all available in Appendix 3.

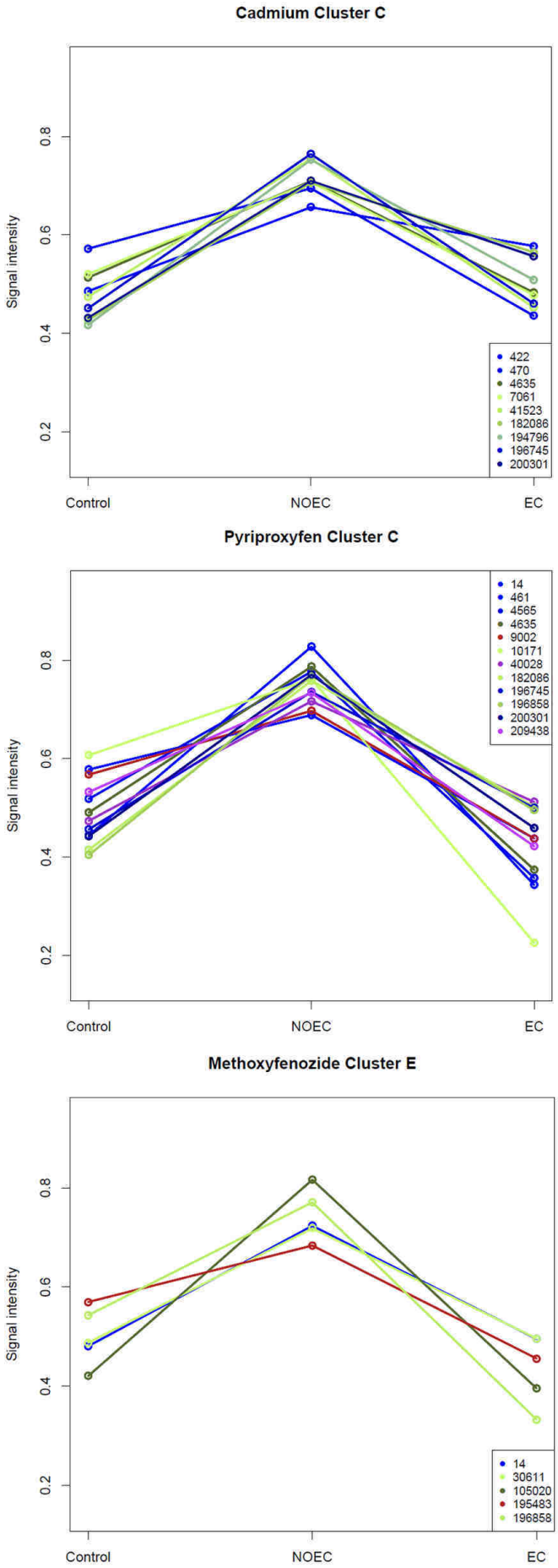


Figure 6: Examples of clustered temporal variations of proteins in response to Cd, Pyr and Met exposure.

The different trends were obtained after clustering the proteins with their spectral counts using the TrendQuest module of the PatternLab program, minimum items per cluster was set to three. A total of eighteen trends were delineated. All are displayed in Appendix 2. For Pyr and Met, the control condition is the solvent control. The NOEC was set at 0.3 $\mu g.L^{-1}$, 0.5 $\mu g.L^{-1}$, and 0.001 $\mu g.L^{-1}$ for Cd, Pyr, and Met, respectively. The EC was set at 3 $\mu g.L^{-1}$, 50 $\mu g.L^{-1}$, and 0.1 $\mu g.L^{-1}$ for Cd, Pyr, and Met, respectively. Color legends are: blue for proteins classified as male specific (30), green for proteins showing temporal variations among all xenobiotics, brown for proteins showing clustered temporal variations in Pyr and Cd, yellow for proteins showing clustered temporal variations in Cd and Met, purple for proteins showing clustered temporal variations in Met and Pyr, and red when the protein is uniquely modulated by the studied xenobiotic. *n* = 5 males per condition.

4. DISCUSSION

Using sperm parameters to assess reproductive health status. Male reproductive disorders in invertebrates have been an underexplored issue in environmental risk assessment. Here, we aimed to study male reproductive disorders in the amphipod, *G. fossarum*, a keystone organism in European freshwater ecosystems, in response to three toxicants. For this, we measured firstly sperm quality markers, and then global proteome alterations, in response to toxic pressure. Reproductive cell viability is commonly measured in the evaluation of sperm quality and has proved to be sensitive for fish (46), urchins, and bivalves (47). However, previous investigations on *G. fossarum* spermatozoa in response to genotoxic agents, or in field experiments, have shown no significant impact on cell viability (21, 48). Thus, in *G. fossarum*, the measurement of sperm viability does not appear to be a sensitive endpoint for the assessment of male reproductive disorders. Here, we choose to monitor the production of reproductive cells and to differentiate their associated maturation states.

Spermatogenesis takes place in seminiferous tubules and is divided into three phases: i) the proliferation phase, where intense mitotic division from spermatogonia to spermatocytes takes place, ii) the meiotic phase producing spermatids, and iii) the differentiation phase leading to spermatozoa (33). Apoptosis is a key phenomenon in the control of sperm production, surplus and abnormal cells being eliminated by this mechanism. In mammals, the phase that is the most susceptible to toxic insult is believed to be the meiotic phase (49). Apoptotic phenomena are most commonly observed in dividing spermatocytes, and less frequently in the other cell types (50). Thus, the production of both spermatids and spermatozoa usually decreases in response to toxicants. Surprisingly, the production of spermatids in *G. fossarum* was not significantly affected by exposure to any of the three toxicants, while spermatozoon production was impacted in a dose-response relationship. Thus, in crustaceans, the most susceptible phase toward toxic pressure is the differentiation phase, leading to a failure of spermatozoa maturation. Such a result has also been observed in the freshwater prawn, *Macrobrachium rosenbergii*, exposed to tributyltin (51). At a 10 $ng.L^{-1}$ treatment, a decrease only in spermatozoon production was recorded, while higher concentrations caused a decreased abundance of immature sperm cells concomitant with histological disorders in seminiferous tubules (51). The latter observation suggests a massive toxic impact on testis tissue, preventing the production of reproductive cells, rather than a specific control mechanism within cell development.

In crustaceans, investigations on sperm quality have essentially been carried out in an agronomic perspective, while few studies have been carried out on environmentally challenged organisms. The marine amphipod, *Echinogammarus marinus*, collected at industrially impacted sites in Scotland, showed a 15% decrease in sperm count compared to samples collected from reference sites (14). In the present study, we measured, in controlled laboratory conditions, an important dose-dependent decrease in spermatozoon production in response to Cd and Pyr, up to 75%. To our knowledge, no data regarding the environmental relevance of the Met concentrations used are available for reference, but for Pyr and Cd, the chosen concentrations are environmentally realistic (17, 39). From our results, we can deduce that such a drastic impact on sperm production can lead to dramatic effects at the demographic level. In the amphipod, *Gammarus duebeni*, the relationship between sperm count and fertilization success was modelled by Ford *et al.* (52), theorizing that a 10% reduction in the sperm count leads to population collapse in less than a decade. We recently proposed a robust population model for capturing *G. fossarum* annual population dynamics in response to modifications of juvenile survival, adult survival and female fertility (53). Thus, the spermatozoon count is a relevant parameter to implement for apprehending population vulnerability related to male reproductive disorders.

Impact of pollutants on reproductive physiology. The arthropod endocrine system relies on the action of hormones belonging to two distinct chemical classes: ecdysteroids also called the molting hormone, and terpenoids, known as the JH. Chemicals for arthropod pest control are designed to interfere with these signaling pathways by directly interacting with the hormone nuclear receptors through either agonist or antagonist actions. Met exhibits a high affinity to the ecdysone receptor complex (EcR:USP) of lepidopteran insects (caterpillar pest), as it is a potent mimic of one of the active forms of ecdysteroids, 20-hydroxyecdysone. Induced effects include molting irregularities in caterpillars, such as abnormal cuticle development or premature apolysis. This pesticide is believed to have an excellent margin of safety to nontarget organisms (54), although available data are limited. Remarkably, no significant impact on spermatozoon production was observed in our survey, whilst sperm counts from organisms exposed to Met were characterized by strong inter-individual variability. Under exposure to the lowest concentration, individual organisms could exhibit a two-fold (either greater or lesser) variation in sperm count as compared with the control condition (mean = 88 $x10^2$; min = 44 $x10^2$; max =190 $x10^2$; %RSD = 60 % compared to the control condition, mean = 100 $x10^2$; %RSD = 12 %). While ecdysteroids are known to control various

physiological processes related to reproduction, such as molting or embryonic development, knowledge about their actions on spermatogenesis is scarce. In the shrimp, *Litopenaeus vannamei*, Parnes *et al.* (55) showed that males go through reproductive cycles that are strictly associated with their molt cycles, suggesting a possible hormonal regulation by ecdysteroids. In the European corn worm, *Ostrinia nubilalis*, physiological doses of ecdysteroids were administrated onto *in vitro*-cultured testes to determine their effects on the final stage of spermatogenesis. A stimulatory effect was detected on testes from 4-5-day–post-ecdysis larvae, while the treatment produced no stimulation on 3-day–post-ecdysis larvae (56). Here, inter-individual variability can be explained by the occurrence of different molt stages among our pool of organisms, resulting in a different sensitivity towards Met. A temporal window of effects due to specific physiological status is a characteristic of EDCs.

Pyr is a member of the insecticidal JH analogues, which inhibit the larval development and maturation of many insects (ticks, mosquitos, ants, houseflies, and fleas) (57). In our survey, exposure to this pesticide inhibited sperm production in a dose-response relationship. In decapods, JH analogues have been proven to affect metamorphosis and to induce morphological abnormalities and reduce growth rates (reviewed by 9). In daphnids, exposure to Pyr reduces female fecundity and stimulates male production (12, 39). Compared with ecdysteroids, JH functions have to date been elusive. Moreover, the JH receptor is still uncharacterized. In males, this hormone is believed to play a role in reproduction and behavior. In decapods, hormonal treatment resulted in the stimulation of testis growth (58). One characteristic of hormone mimetics is their nonmonotonic dose-response curves, where exposure to low concentrations leads to physiological stimulatory effects, while higher doses may actually be inhibitory. Here, organisms exposed to Pyr showed clear inhibition of spermatozoon production. Thus, exposure to lower concentrations of toxicant may give a stimulatory response.

Finally, we showed that, in amphipods, sperm production is highly sensitive to Cd exposure. Freshwater crustaceans are among the most Cd-sensitive macroinvertebrate species, as shown by previous studies that have documented impairments of female reproduction, embryonic development, energy acquisition (22), and iono-osmoregulation (38) triggered by Cd exposure. Because of its electronic structure, Cd acts through a process of “ionic mimicry”, interacting with transporters involved in the uptake of calcium, iron and zinc, and thus leading to a wide range of adverse effects (reviewed by 59). As the integrity of calcium homeostasis is crucial for sperm

maturation, capacitation, and sperm-egg interaction, sperm production is thus strongly impacted by low Cd doses.

Identification of molecular responses triggered by toxic exposure. Based on the physiological endpoints, we analyzed the dynamics of testis proteomes for each xenobiotic at three different stages. Proteins for which modulation was only demonstrated between the NOEC and the EC were disregarded, because this pattern indicates that the organism's defensive capacity was already overwhelmed at this stage. Concerning exposure to pesticides, six proteins were modulated by both the Met and Pyr compounds, including i) an Na^+/K^+ ATPase, which suggests a disturbance of iono-osmoregulation (increase at the NOEC and decrease at the EC for Pyr, irrelevant pattern for Met), and ii) a factor of hemocyte aggregation, a hemocytin (60), which showed an induction followed by a decreased abundance in response to both pollutants. Five proteins were uniquely modulated upon Met exposure. Two heat shock proteins (HSP) were induced and a glucan-pattern-recognition protein, involved in the recognition of foreign invaders, was downregulated. In this survey, seven proteins were uniquely modulated by Pyr. Among these, in addition to housekeeping proteins, an HSP 70 and a protein without functional annotation (ID 17884) were induced, while another protein with no predicted function (ID 3678) and an orphan (ID 110918) were less abundant. Surprisingly, we have previously sequenced several proteins involved in JH metabolism in *G. fossarum* testes, but no variation of these occurred during spermatogenesis (31) or xenobiotic exposure, suggesting that this hormone may not directly control spermatogenesis but rather may maintain male sexual characteristics.

Only two proteins, involved in cytoskeletal structure maintenance were modulated uniquely by Cd exposure. In male reproduction, the effects of Cd have been intensively studied in model mammals. These effects include decrease in testicular growth rate, failure of spermatozoon maturation, induction of oxidative stress in sperm cells, and alteration of cytoskeletal protein assembly by junctional disruption. Sperm production was observed to be significantly altered in organisms exposed to Pyr and to Cd. Three proteins were modulated by these two xenobiotics: i) a protein with no predicted function (ID 135487) that was induced upon exposure to Cd and to Pyr, ii) another, male-specific, protein (ID 69913), which was repressed in response to Cd (irrelevant pattern for Pyr), and iii) a protein annotated as vitellogenin, repressed in response to Pyr (irrelevant pattern for Cd). Of note, this functional prediction is misleading as it is only based on phylogenetic analysis. Because of the absence of specificity towards the female reproductive tissue, this protein

is more likely to be a clottable protein, involved in innate immunity. In fact, in our survey, several proteins involved in immunity were repressed in response to pollutants. As is the case for other arthropods, the testis of gammarids can be infected by parasites that are known to actively convert males into females for enhancing the parasite transmission to the offspring, leading to a feminized phenotype (61). Thus, intersexuality in a field population from an anthropically impacted site can also be an indirect consequence of immunosuppression caused by pollutants.

Among the proteins modulated by all three pollutants, the majority are housekeeping proteins involved in cytoskeletal structure maintenance or energy metabolism. Four proteins are calcium-dependent or involved in calcium homeostasis, including three that were previously classified as male specific. Interestingly, these three proteins, including two Ca^{2+} transporters, are all induced in response to xenobiotic exposure. Hughes *et al.* (62) showed that a wide range of xenobiotics inhibit Ca^{2+} ATPase in Sertoli testicular cells and that dysregulation of Ca homeostasis leads to Ca^{2+}-dependent cell death *via* apoptosis. Thus, overexpression of Ca^{2+} transporters may be a compensatory mechanism. The other male-specific protein induced is annotated as Copine-8. In humans, the associated gene is predominantly expressed in prostate and testis, suggesting an important role in the development and regulation of male sexual characteristics, although its function is to date elusive (63). Finally, two orphans (ID 41523, 196558) and six proteins also without functional annotation (ID 9438, 7061, 4635, 128511, 5492, 40293) showed clustered temporal variations with all three xenobiotics. Three of them showed relevant pattern only in the Cd condition. Proteins ID 41523 and 7061 were induced when comparing the NOEC to the control, while the protein ID 128511 was repressed. The other remaining proteins, including one male specific (ID 9438), were all induced when comparing the NOEC to the control in response to all three xenobiotics.

Identification of endocrine disruptor biomarkers. Traditionally, endocrine disruption in crustaceans has been studied through the prism of vertebrate endocrine disruption, *i.e*. monitoring the induction of vitellogenin. However, this biomarker has been proven over the years to be inappropriate for this objective (28, 29). Previous comparisons of male and female reproductive proteomes allowed the identification of proteins with strong sexual dysmorphism: 129 and 75 proteins for males and females, respectively. **Table 1** shows the proteins with clustered temporal variation among toxic exposure that were previously identified as sex-specific and modulated during spermatogenesis (31).

Table 1: Proteins with clustered temporal variations among toxic exposure previously identified as sex-specific

Contig ID	Predicted function	Origin of the closest homolog	Blast E-value	Total Spectral Count	Proteogenomic experiment		Cluster		
					Spermatogenesis	Comparative	Cd	Pyr	Met
14	Slow muscle myosin S1 heavy chain	*Homarus americanus*	2,64E-123	1071		Male	B	C	E
422	Myosin heavy chain	*Bombyx mandarina*	4,23E-64	287		Male	C	A	B
461	Hemocytin-like	*Apis florea*	1,73E-19	221	X	Male		C	D
470	Spectrin beta chain-like	*Apis florea*	0,00E+00	218		Male	C		
2029	Projectin	*Procambarus clarkii*	0,00E+00	147		Male	F		
2853	Sarcoplasmic calcium-binding protein	*Procambarus clarkii*	3,22E-79	318		Male	B	B	C
4227	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum	*Acyrthosiphon pisum*	0,00E+00	350	X	Male	A	F	B
4565	Calmin-like protein	*Tigriopus californicus*	6,72E-21	333		Male	F	C	C
9438	No predicted function	*Bactrocera oleae*	1,62E-118	405		Male	B	F	D
22255	Myosin heavy chain	*Drosophila melanogaster*	1,35E-104	498		Male	B	A	B
23428	Heat shock cognate 70	*Aedes aegypti*	1,30E-140	132		Female			D
34845	Copine-8	*Crassostrea gigas*	5,64E-88	331		Male	D	B	C
69913	No predicted function	*Daphnia pulex*	6,28E-84	211		Male	D	A	
196745	Myosin heavy chain	*Marsupenaeus japonicus*	1,30E-51	382		Male	C	C	B
200301	Myosin heavy chain	*Gammarus duebeni*	1,33E-102	334		Male	C	C	B

In this survey, the modulation of fourteen male-specific proteins was recorded in response to EDCs exposure, while only one protein, an HSP 70, previsouly classified as more abundant in female tissue, was found induced. This chaperon protein has versatile roles and its accumulation in the ovaries may be a maternal factor involved in various aspects of embryonic development, such as cell movement, proliferation, morphogenesis, and apoptosis (64). Here, this protein, along with other modulated HSPs in the testis, may have been induced for the promotion of sperm cell survival and to ensure cellular homeostasis.

Compared to vertebrates, where the proportion of steroid hormones differs between males and females, crustacean sex determination is under the control of a specific hormone, the androgenic gland hormone, which controls the expression of male sexual characteristics. In the absence of this hormone, organisms spontaneously develop into females. Thus, in crustaceans, the intersexual phenotype is rather the result of a demasculinization process than of a feminization one (61). Here, exposure to EDC modulated several male-specific proteins, apart from housekeeping proteins. For example, two Ca^{2+} transporters and the protein, Copine-8, were upregulated. Additionally, homologs of a protein from *Daphnia pulex* and of a protein from the olive fruit fly, *Bactrocera oleae,* both with no predicted function, were induced in response to EDCs. It is worth noting that proteins devoid of functional prediction were preponderant in this study, accounting for 20% of the proteins showing temporal variations. In our global protein dataset, 8% of the identified proteins are orphans and a third were not functionally annotated. Reproductive proteins are highly diversified among the tree of life, as the result of sexual selection, and are consequently highly species specific. Thus, these proteins may be involved in either sperm navigation towards the egg or sperm-egg interactions. These surface receptors play an essential role in the fertilization process, consequently any defects or abnormalities may lower the organism's reproductive potential (65). Furthermore, compared to our previous spermatogenesis experiment, a significant portion of proteins without predicted function were newly detected, including 24 *Daphnia pulex* homologs and 18 orphans. Transcriptomic analysis on the *Daphnia pulex* genome has shown that orphan genes are amongst the most ecoresponsive (66), and thus they are interesting biomarker candidates. In any case, highly divergent proteins are generally present in low abundance in the proteomes of organisms. Standard shotgun proteomic strategies allow the preferential detection of the abundant proteins, for which the functions are more commonly shared across the tree of life (67). Additionally, in mammals spermatogenesis, "low copy number proteins" are believed to be crucial for germ cell differentiation (33). Thus, while our strategy enabled the discovery of these

proteins as potential biomarkers, measuring their sexual dysmorphism, such as monitoring their modulation towards the organism's physiological stage or xenobiotic exposure, requires more targeted proteomics approaches. Consequently, based on our earlier work for the quantitation of vitellogenin by LC-MS/MS (68), a multiple-reaction-monitoring strategy should be developed for further biomarker validation. This could also be applied for routine biomonitoring using *G. fossarum* as a sentinel organism (24).

Supporting Information: Supplementary data concerning raw peptide identification, protein identification and quantitation can be found respectively in Appendix 1 and 2. All clustered temporal trends in response to xenobiotic exposure are displayed in Appendix 3.

Abbreviations: EcR:USP: ecdysone receptor complex; EDC: endocrine-disrupting chemical; HSP: heat shock protein; JH: juvenile hormone; Met: methoxyfenozide; NOEC: No Observed Effect Concentration; LOEC: Lowest Observed Effect Concentration; Pyr: pyriproxyfen.

Acknowledgements: J. Trapp was supported by a fellowship from the Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (France) and from the 'programme transversal de toxicologie' of the Commissariat de l'Energie Atomique et aux Energies Alternatives (France). We thank the Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (France) and the Commissariat de l'Energie Atomique et aux Energies Alternatives (France) for financial support.

REFERENCES

1. Merzenich, H.; Zeeb, H.; Blettner, M., Decreasing sperm quality: a global problem? *BMC Public Health* **2010,** 10, 24.
2. Aitken, R. J.; Baker, M. A., Oxidative stress, sperm survival and fertility control. *Mol Cell Endocrinol* **2006,** 250, (1-2), 66-9.
3. Guillette, L. J., Jr.; Gross, T. S.; Masson, G. R.; Matter, J. M.; Percival, H. F.; Woodward, A. R., Developmental abnormalities of the gonad and abnormal sex hormone concentrations in juvenile alligators from contaminated and control lakes in Florida. *Environ Health Perspect* **1994,** 102, (8), 680-8.
4. Jobling, S.; Nolan, M.; Tyler, C. R.; Brighty, G.; Sumpter, J. P., Widespread sexual disruption in wild fish. *Environ Sci Technol* **1998,** 32, (17), 2498-2506.
5. Varghese, A. C.; du Plessis, S. S.; Agarwal, A., Male gamete survival at stake: causes and solutions. *Reprod Biomed Online* **2008,** 17, (6), 866-880.
6. Lewis, C.; Ford, A. T., Infertility in male aquatic invertebrates: a review. *Aquat Toxicol* **2012,** 120-121, 79-89.
7. Zhang, Z. Q., Animal biodiversity: An introduction to higher-level classification and taxonomic richness. *Zootaxa* **2011**, (3148), 7-12.

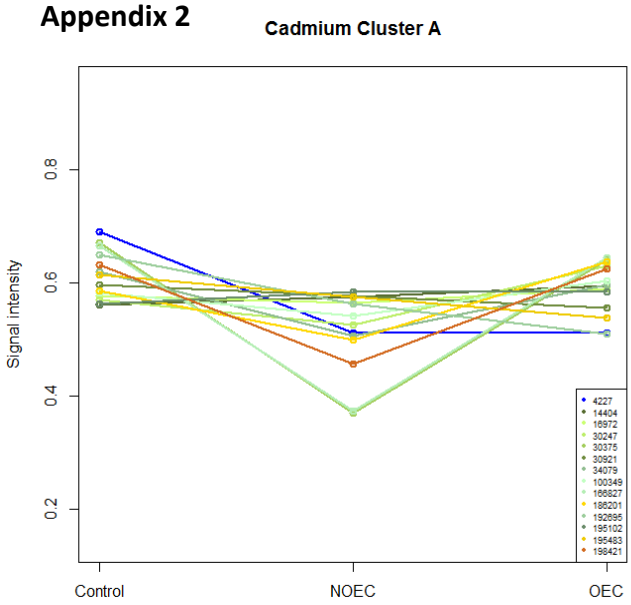
8. Lye, C. M.; Bentley, M. G.; Galloway, T., Effects of 4-nonylphenol on the endocrine system of the shore crab, Carcinus maenas. *Environ Toxicol* **2008,** 23, (3), 309-18.
9. LeBlanc, G. A., Crustacean endocrine toxicology: a review. *Ecotoxicology* **2007,** 16, (1), 61-81.
10. Wang, H. Y.; Olmstead, A. W.; Li, H.; Leblanc, G. A., The screening of chemicals for juvenoid-related endocrine activity using the water flea Daphnia magna. *Aquat Toxicol* **2005,** 74, (3), 193-204.
11. Ghekiere, A.; Verslycke, T.; Fockedey, N.; Janssen, C. R., Non-target effects of the insecticide methoprene on molting in the estuarine crustacean Neomysis integer (Crustacea: Mysidacea). *J Exp Mar Biol Ecol* **2006,** 332, (2), 226-234.
12. Olmstead, A. W.; LeBlanc, G. A., Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean Daphnia magna. *Environ Health Perspect* **2003,** 111, (7), 919-24.
13. Jungmann, D.; Ladewig, V.; Ludwichowski, K. U.; Petzsch, P.; Nagel, R., Intersexuality in Gammarus fossarum Koch - A common inducible phenomenon? *Arch Hydrobio* **2004,** 159, (4), 511-529.
14. Ford, A. T.; Fernandes, T. F.; Rider, S. A.; Read, P. A.; Robinson, C. D.; Davies, I. M., Endocrine disruption in a marine amphipod? Field observations of intersexuality and de-masculinisation. *Mar Environ Res* **2004,** 58, (2-5), 169-73.
15. Schirling, M.; Jungmann, D.; Ladewig, V.; Nagel, R.; Triebskorn, R.; Kohler, H. R., Endocrine effects in Gammarus fossarum (Amphipoda): influence of wastewater effluents, temporal variability, and spatial aspects on natural populations. *Arch Environ Contam Toxicol* **2005,** 49, (1), 53-61.
16. Dangles, O.; Guérold, F., Linking shredders and leaf litter processing: Insights from an acidic stream study. *Int Rev Hydrobiol* **2001,** 86, (4-5), 395-406.
17. Besse, J. P.; Coquery, M.; Lopes, C.; Chaumot, A.; Budzinski, H.; Labadie, P.; Geffard, O., Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: towards the determination of threshold values. *Water Res* **2013,** 47, (2), 650-60.
18. Geffard, A.; Queau, H.; Dedourge, O.; Biagianti-Risboug, S.; Geffard, O., Influence of biotic and abiotic factors on metallothionein level in Gammarus pulex. *Comp Biochem Physiol C Toxicol Pharmacol* **2007,** 145, (4), 632-40.
19. Xuereb, B.; Chaumot, A.; Mons, R.; Garric, J.; Geffard, O., Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda) Intrinsic variability, reference levels, and a reliable tool for field surveys. *Aquat Toxicol* **2009,** 93, (4), 225-33.
20. Dedourge-Geffard, O.; Palais, F.; Biagianti-Risbourg, S.; Geffard, O.; Geffard, A., Effects of metals on feeding rate and digestive enzymes in Gammarus fossarum: an in situ experiment. *Chemosphere* **2009,** 77, (11), 1569-76.
21. Lacaze, E.; Devaux, A.; Jubeaux, G.; Mons, R.; Gardette, M.; Bony, S.; Garric, J.; Geffard, O., DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level. *Sci Total Environ* **2011,** 409, (17), 3230-6.
22. Geffard, O.; Xuereb, B.; Chaumot, A.; Geffard, A.; Biagianti, S.; Noel, C.; Abbaci, K.; Garric, J.; Charmantier, G.; Charmantier-Daures, M., Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment. *Environ Toxicol Chem* **2010,** 29, (10), 2249-59.
23. Tatarazako, N.; Oda, S., The water flea Daphnia magna (Crustacea, Cladocera) as a test species for screening and evaluation of chemicals with endocrine disrupting effects on crustaceans. *Ecotoxicology* **2007,** 16, (1), 197-203.
24. Trapp, J.; Armengaud, J.; Salvador, A.; Chaumot, A.; Geffard, O., Next-generation proteomics: towards customized biomarkers for environmental biomonitoring. *Environ. Sci. Technol.* **2014**.
25. Kime, D. E.; Nash, J. P.; Scott, A. P., Vitellogenesis as a biomarker of reproductive disruption by xenobiotics. *Aquaculture* **1999,** 177, (1-4), 345-352.
26. Matozzo, V.; Gagne, F.; Marin, M. G.; Ricciardi, F.; Blaise, C., Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review. *Environ Int* **2008,** 34, (4), 531-45.
27. Jubeaux, G.; Simon, R.; Salvador, A.; Queau, H.; Chaumot, A.; Geffard, O., Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males. *Aquat Toxicol* **2012,** 112-113, (0), 72-82.

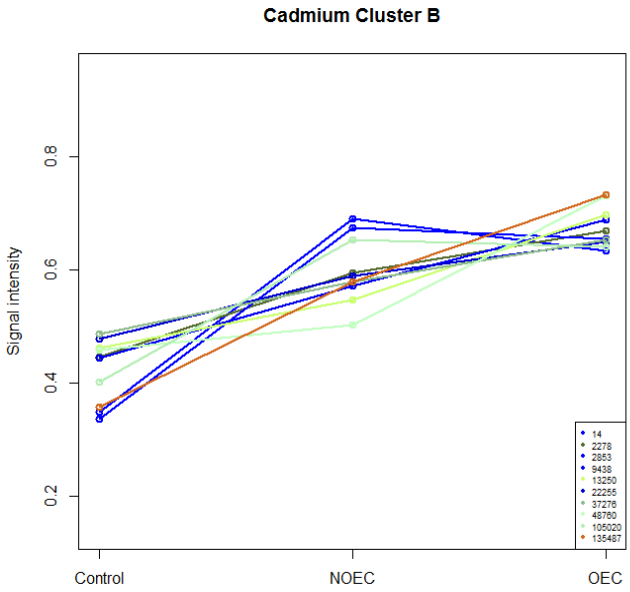
28. Jubeaux, G.; Simon, R.; Salvador, A.; Lopes, C.; Lacaze, E.; Queau, H.; Chaumot, A.; Geffard, O., Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: inconclusive experience. *Aquat Toxicol* **2012,** 122-123, 9-18.
29. Short, S.; Yang, G.; Kille, P.; Ford, A. T., Vitellogenin is not an appropriate biomarker of feminisation in a crustacean. *Aquat Toxicol* **2014,** 153, (0), 89-97.
30. Hayward, A.; Takahashi, T.; Bendena, W. G.; Tobe, S. S.; Hui, J. H., Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans. *FEBS Lett* **2010,** 584, (6), 1273-8.
31. Trapp, J.; Geffard, O.; Imbert, G.; Gaillard, J. C.; Davin, A. H.; Chaumot, A.; Armengaud, J., Proteogenomics of Gammarus fossarum to document the reproductive system of amphipods. *Mol Cell Proteomics* **2014**.
32. Chen, J. E.; Li, J. Y.; You, Z. Y.; Liu, L. L.; Liang, J. S.; Ma, Y. Y.; Chen, M.; Zhang, H. R.; Jiang, Z. D.; Zhong, B. X., Proteome analysis of silkworm, Bombyx mori, larval gonads: characterization of proteins involved in sexual dimorphism and gametogenesis. *J Proteome Res* **2013,** 12, (6), 2422-38.
33. Chocu, S.; Calvel, P.; Rolland, A. D.; Pineau, C., Spermatogenesis in mammals: proteomic insights. *Syst Biol Reprod Med* **2012,** 58, (4), 179-90.
34. Chocu, S.; Evrard, B.; Lavigne, R.; Rolland, A. D.; Aubry, F.; Jegou, B.; Chalmel, F.; Pineau, C., Forty-Four Novel Protein-Coding Loci Discovered Using a Proteomics Informed by Transcriptomics (PIT) Approach in Rat Male Germ Cells. *Biol Reprod* **2014**.
35. Martyniuk, C. J.; Alvarez, S.; McClung, S.; Villeneuve, D. L.; Ankley, G. T.; Denslow, N. D., Quantitative proteomic profiles of androgen receptor signaling in the liver of fathead minnows (Pimephales promelas). *J Proteome Res* **2009,** 8, (5), 2186-200.
36. Brunelli, L.; Llansola, M.; Felipo, V.; Campagna, R.; Airoldi, L.; De Paola, M.; Fanelli, R.; Mariani, A.; Mazzoletti, M.; Pastorelli, R., Insight into the neuroproteomics effects of the food-contaminant non-dioxin like polychlorinated biphenyls. *J Proteomics* **2012,** 75, (8), 2417-30.
37. Campagna, R.; Brunelli, L.; Airoldi, L.; Fanelli, R.; Hakansson, H.; Heimeier, R. A.; De Boever, P.; Boix, J.; Llansola, M.; Felipo, V.; Pastorelli, R., Cerebellum proteomics addressing the cognitive deficit of rats perinatally exposed to the food-relevant polychlorinated biphenyl 138. *Toxicol Sci* **2011,** 123, (1), 170-9.
38. Felten, V.; Charmantier, G.; Mons, R.; Geffard, A.; Rousselle, P.; Coquery, M.; Garric, J.; Geffard, O., Physiological and behavioural responses of Gammarus pulex (Crustacea: Amphipoda) exposed to cadmium. *Aquat Toxicol* **2008,** 86, (3), 413-25.
39. Ginjupalli, G. K.; Baldwin, W. S., The time- and age-dependent effects of the juvenile hormone analog pesticide, pyriproxyfen on Daphnia magna reproduction. *Chemosphere* **2013,** 92, (9), 1260-6.
40. Lacaze, E.; Geffard, O.; Bony, S.; Devaux, A., Genotoxicity assessment in the amphipod Gammarus fossarum by use of the alkaline Comet assay. *Mutat Res-Gen Tox En* **2010,** 700, (1–2), 32-38.
41. R Development Core Team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna, Austria, 2008.
42. Hartmann, E. M.; Allain, F.; Gaillard, J. C.; Pible, O.; Armengaud, J., Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods Mol Biol* **2014,** 1197, 275-85.
43. Fischer Jde, S.; Liao, L.; Carvalho, P. C.; Barbosa, V. C.; Domont, G. B.; Carvalho Mda, G.; Yates, J. R., 3rd, Dynamic proteomic overview of glioblastoma cells (A172) exposed to perillyl alcohol. *J Proteomics* **2010,** 73, (5), 1018-27.
44. Vizcaino, J. A.; Cote, R. G.; Csordas, A.; Dianes, J. A.; Fabregat, A.; Foster, J. M.; Griss, J.; Alpi, E.; Birim, M.; Contell, J.; O'Kelly, G.; Schoenegger, A.; Ovelleiro, D.; Perez-Riverol, Y.; Reisinger, F.; Rios, D.; Wang, R.; Hermjakob, H., The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **2013,** 41, (Database issue), D1063-9.
45. Carvalho, P. C.; Fischer, J. S.; Chen, E. I.; Yates, J. R., 3rd; Barbosa, V. C., PatternLab for proteomics: a tool for differential shotgun proteomics. *BMC Bioinformatics* **2008,** 9, 316.
46. Grzyb, K.; Rychlowski, M.; Biegniewska, A.; Skorkowski, E. F., Quantitative determination of creatine kinase release from herring (Clupea harengus) spermatozoa induced by tributyltin. *Comp Biochem Physiol C Toxicol Pharmacol* **2003,** 134, (2), 207-13.

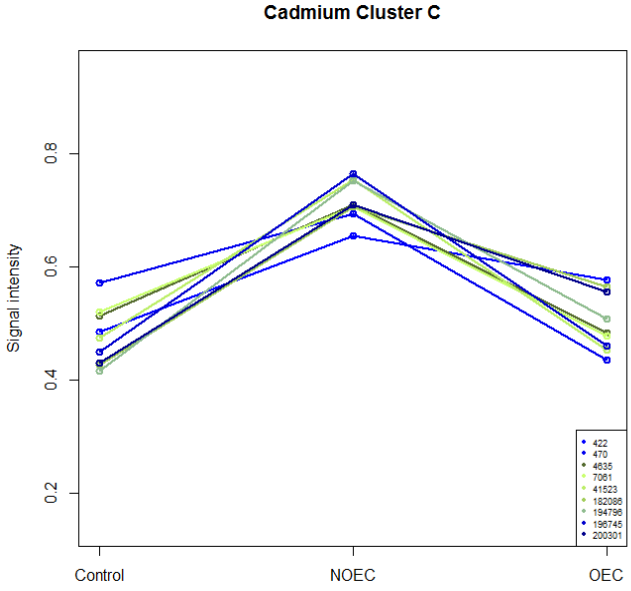
47. Favret, K. P.; Lynn, J. W., Flow-cytometric analyses of viability biomarkers in pesticide-exposed sperm of three aquatic invertebrates. *Arch Environ Contam Toxicol* **2010,** 58, (4), 973-84.
48. Lacaze, E.; Devaux, A.; Mons, R.; Bony, S.; Garric, J.; Geffard, A.; Geffard, O., DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment. *Environ Pollut* **2011,** 159, (6), 1682-91.
49. Evans, T. J.; Ganjam, V. K., Reproductive Anatomy and Physiology. In *Reproductive and Developmental Toxicology*, 2011; pp 7-32.
50. Sinha Hikim, A. P.; Swerdloff, R. S., Hormonal and genetic control of germ cell apoptosis in the testis. *Rev Reprod* **1999,** 4, (1), 38-47.
51. Revathi, P.; Iyapparaj, P.; Vasanthi, L. A.; Munuswamy, N.; Krishnan, M., Ultrastructural changes during spermatogenesis, biochemical and hormonal evidences of testicular toxicity caused by TBT in freshwater prawn Macrobrachium rosenbergii (De Man, 1879). *Environ Toxicol* **2014,** 29, (10), 1171-81.
52. Ford, A. T.; Martins, I.; Dunn, A. M., Insights into sperm-fertilisation relationships in the Arthropoda with ecological significance modelled in an amphipod. *Invertebr Reprod Dev* **2012,** 56, (1), 50-56.
53. Coulaud, R.; Geffard, O.; Coquillat, A.; Queau, H.; Charles, S.; Chaumot, A., Ecological modeling for the extrapolation of ecotoxicological effects measured during in situ assays in Gammarus. *Environ Sci Technol* **2014,** 48, (11), 6428-36.
54. Carlson, G. R.; Dhadialla, T. S.; Hunter, R.; Jansson, R. K.; Jany, C. S.; Lidert, Z.; Slawecki, R. A., The chemical and biological properties of methoxyfenozide, a new insecticidal ecdysteroid agonist. *Pest Manag Sci* **2001,** 57, (2), 115-9.
55. Parnes, S.; Raviv, S.; Shechter, A.; Sagi, A., Males also have their time of the month! Cyclic disposal of old spermatophores, timed by the molt cycle, in a marine shrimp. *J Exp Biol* **2006,** 209, (Pt 24), 4974-83.
56. Gelman, D. B.; Woods, C. W.; Borkovec, A. B., Effects of ecdysone and 20-hydroxyecdysone on apyrene spermiogenesis in the European corn borer, Ostrinia nubilalis. *J Insect Physiol* **1988,** 34, (7), 733-738.
57. Sullivan, J. J.; Goh, K. S., Environmental fate and properties of pyriproxyfen. *J Pest Sci* **2008,** 33, (4), 339-350.
58. Nagaraju, G. P. C., Is methyl farnesoate a crustacean hormone? *Aquaculture* **2007,** 272, (1-4), 39-54.
59. Thompson, J.; Bannigan, J., Cadmium: toxic effects on the reproductive system and the embryo. *Reprod Toxicol* **2008,** 25, (3), 304-15.
60. Kotani, E.; Yamakawa, M.; Iwamoto, S.-i.; Tashiro, M.; Mori, H.; Sumida, M.; Matsubara, F.; Taniai, K.; Kadono-Okuda, K.; Kato, Y.; Mori, H., Cloning and expression of the gene of hemocytin, an insect humoral lectin which is homologous with the mammalian von Willebrand factor. *Biochim Biophys Acta* **1995,** 1260, (3), 245-258.
61. Ford, A. T., Intersexuality in Crustacea: an environmental issue? *Aquat Toxicol* **2012,** 108, 125-9.
62. Hughes, P. J.; McLellan, H.; Lowes, D. A.; Kahn, S. Z.; Bilmen, J. G.; Tovey, S. C.; Godfrey, R. E.; Michell, R. H.; Kirk, C. J.; Michelangeli, F., Estrogenic alkylphenols induce cell death by inhibiting testis endoplasmic reticulum Ca(2+) pumps. *Biochem Biophys Res Commun* **2000,** 277, (3), 568-74.
63. Maitra, R.; Grigoryev, D. N.; Bera, T. K.; Pastan, I. H.; Lee, B., Cloning, molecular characterization, and expression analysis of Copine 8. *Biochem Biophys Res Commun* **2003,** 303, (3), 842-7.
64. Rupik, W.; Jasik, K.; Bembenek, J.; Widłak, W., The expression patterns of heat shock genes and proteins and their role during vertebrate's development. *Comp Biochem Physiol A Mol Integr Physiol* **2011,** 159, (4), 349-366.
65. du Plessis, S. S.; Kashou, A. H.; Benjamin, D. J.; Yadav, S. P.; Agarwal, A., Proteomics: A subcellular look at spermatozoa. *Reprod Biol Endocrinol.* **2011,** 9.
66. Colbourne, J. K.; Pfrender, M. E.; Gilbert, D.; Thomas, W. K.; Tucker, A.; Oakley, T. H.; Tokishita, S.; Aerts, A.; Arnold, G. J.; Basu, M. K.; Bauer, D. J.; Caceres, C. E.; Carmel, L.; Casola, C.; Choi, J. H.; Detter, J. C.; Dong, Q.; Dusheyko, S.; Eads, B. D.; Frohlich, T.; Geiler-Samerotte, K. A.; Gerlach, D.; Hatcher, P.; Jogdeo, S.; Krijgsveld, J.; Kriventseva, E. V.; Kultz, D.; Laforsch, C.; Lindquist, E.; Lopez, J.; Manak, J. R.; Muller, J.; Pangilinan, J.; Patwardhan, R. P.; Pitluck, S.; Pritham, E. J.; Rechtsteiner, A.; Rho, M.; Rogozin, I. B.; Sakarya, O.; Salamov, A.; Schaack, S.; Shapiro, H.; Shiga, Y.; Skalitzky, C.; Smith, Z.; Souvorov, A.; Sung, W.; Tang, Z.; Tsuchiya, D.; Tu, H.; Vos, H.; Wang, M.; Wolf, Y. I.; Yamagata, H.; Yamada, T.; Ye, Y.; Shaw, J. R.; Andrews, J.;

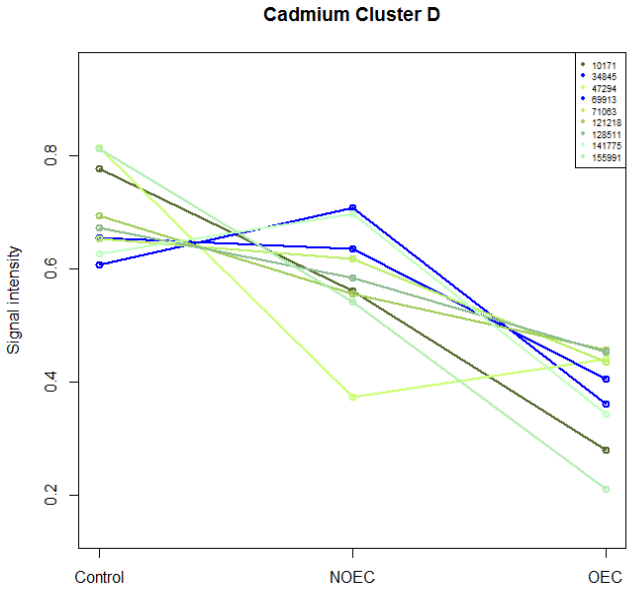
Crease, T. J.; Tang, H.; Lucas, S. M.; Robertson, H. M.; Bork, P.; Koonin, E. V.; Zdobnov, E. M.; Grigoriev, I. V.; Lynch, M.; Boore, J. L., The ecoresponsive genome of Daphnia pulex. *Science* **2011,** 331, (6017), 555-61.
67. Drummond, D. A.; Bloom, J. D.; Adami, C.; Wilke, C. O.; Arnold, F. H., Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* **2005,** 102, (40), 14338-43.
68. Simon, R.; Jubeaux, G.; Chaumot, A.; Lemoine, J.; Geffard, O.; Salvador, A., Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum). *J Chromatogr A* **2010,** 1217, (31), 5109-15.

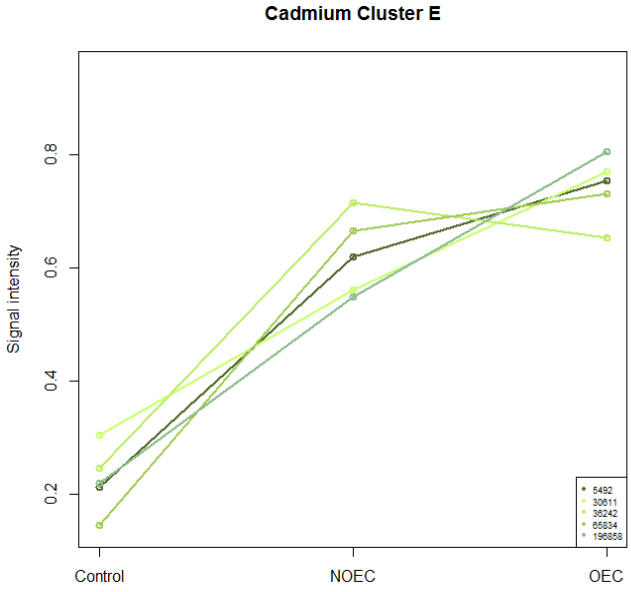
Appendix 2

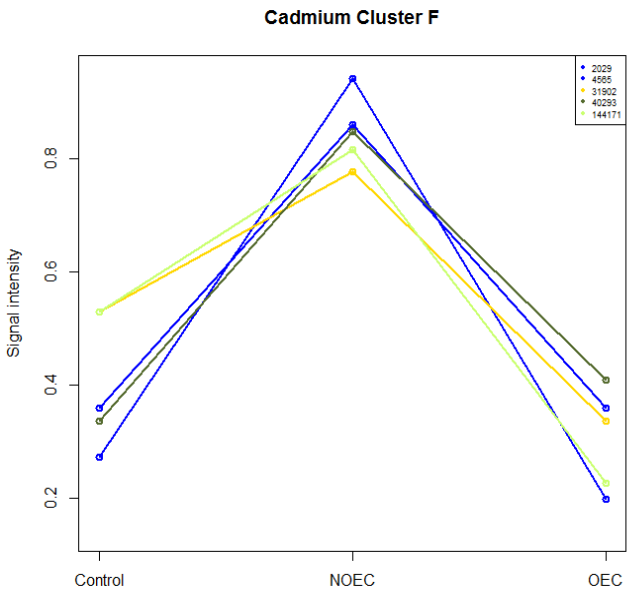

Cadmium Cluster A
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
4227
14404
16972
30247
30375
30921
34079
100349
166827
186201
192695
195102
195483
198421

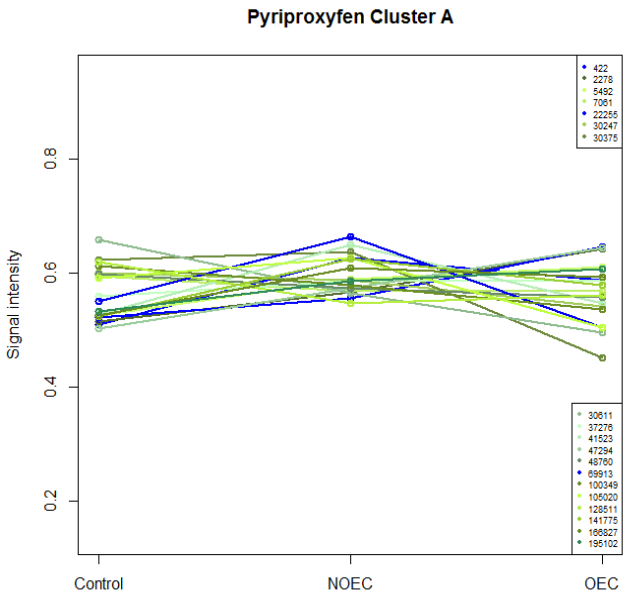

Cadmium Cluster B
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
14
2278
2853
9438
13250
22255
37276
48760
105020
135487

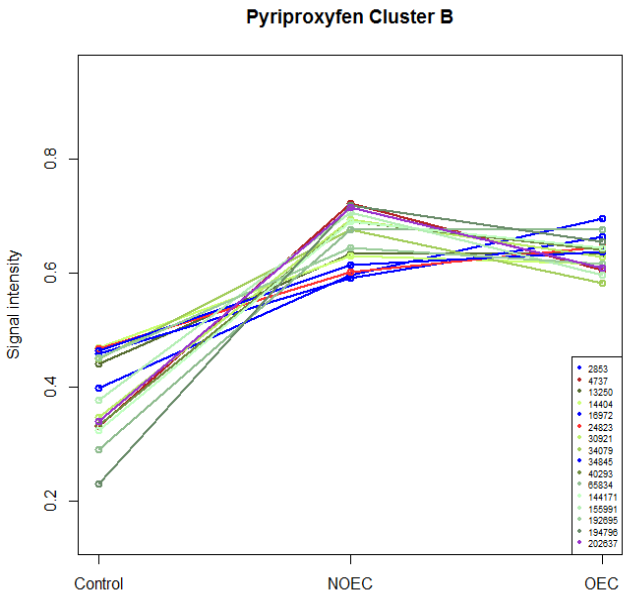

Cadmium Cluster C
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
422
470
4635
7061
41523
182098
194796
196745
200301

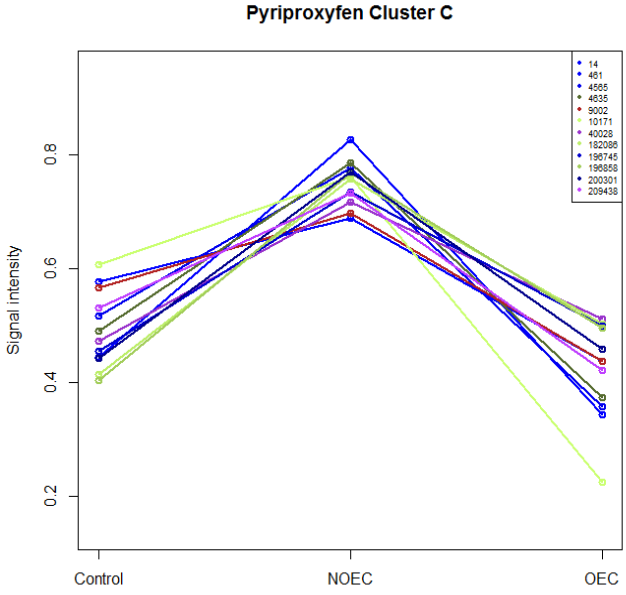

Cadmium Cluster D
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
10171
34845
47294
69913
71063
121218
128511
141775
155991

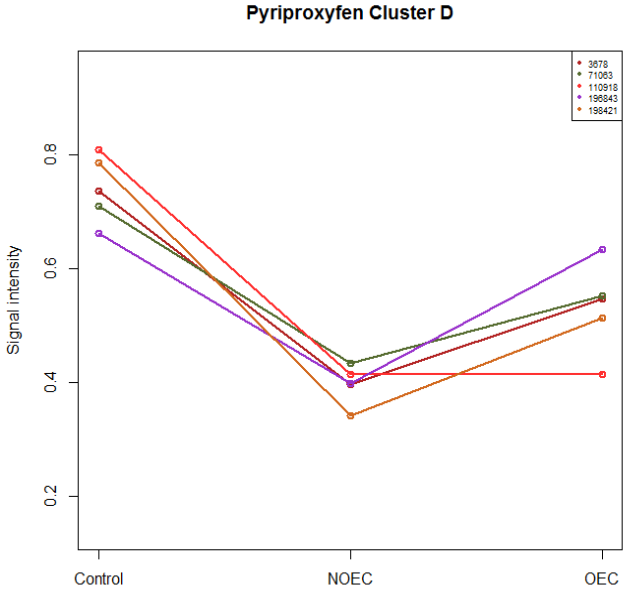

Cadmium Cluster E
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
5492
30611
36242
65834
196858

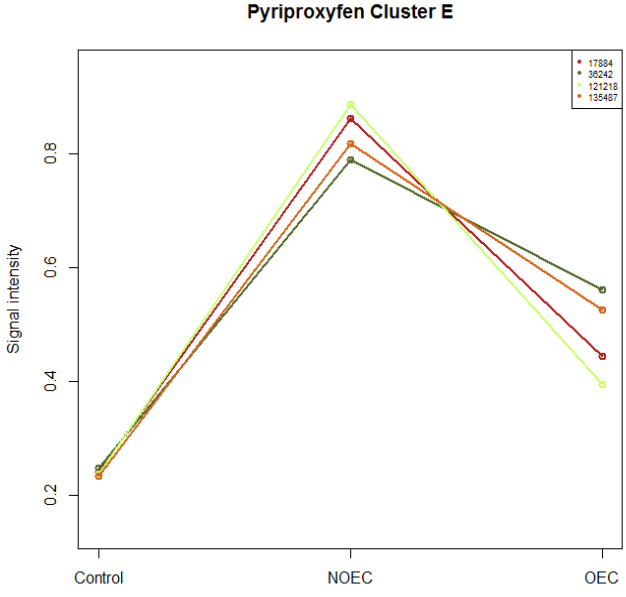

Cadmium Cluster F
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
2029
4566
31902
40293
144171

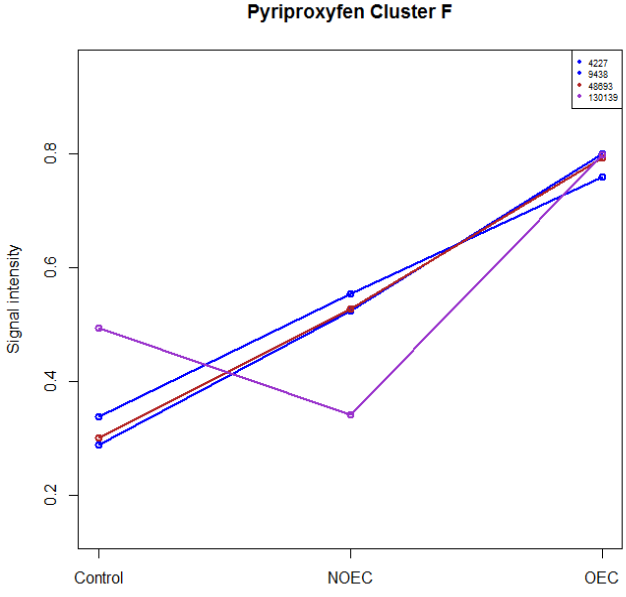

Pyriproxyfen Cluster A
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
422
2278
5492
7061
22255
30247
30375
30611
37276
41523
47294
48760
69913
100349
105020
128511
141775
166827
195102


Pyriproxyfen Cluster B
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
2853
4737
13250
14404
16972
24523
30921
34079
34845
40293
65834
144171
155991
192695
194796
202637

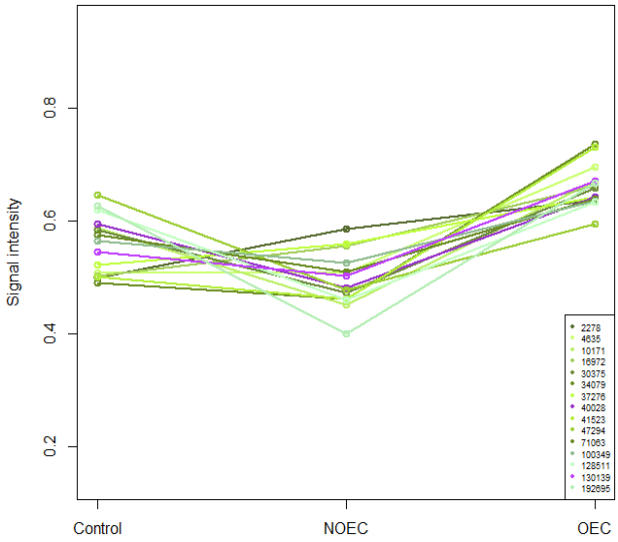

Pyriproxyfen Cluster C
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
14
461
4565
4635
9002
10171
40028
182086
196745
196858
200301
209438


Pyriproxyfen Cluster D
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
3678
71063
110918
196843
198421

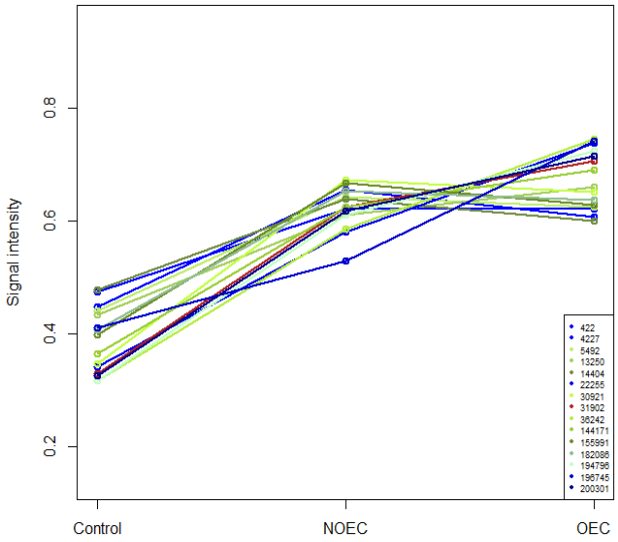

Pyriproxyfen Cluster E
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
17884
36242
121218
135487


Pyriproxyfen Cluster F
Signal intensity
0.2
0.4
0.6
0.8
Control
NOEC
OEC
4227
9438
48693
130139

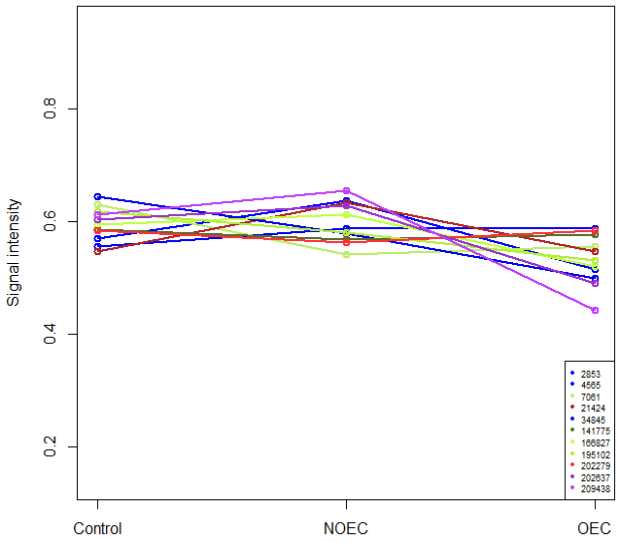
Methoxyfenozide Cluster A



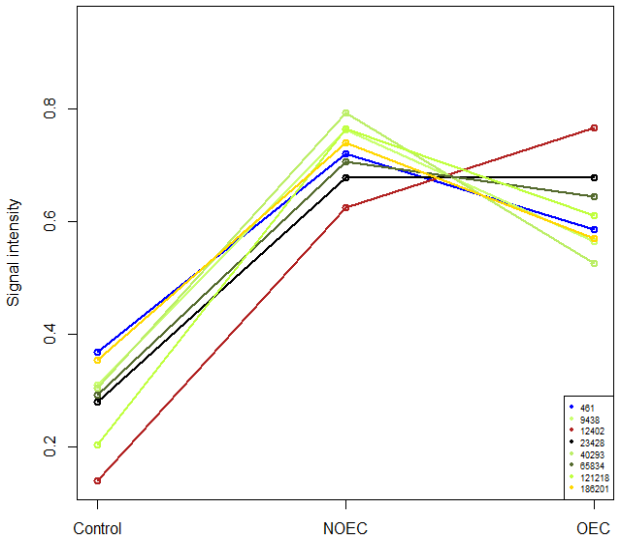
Methoxyfenozide Cluster B



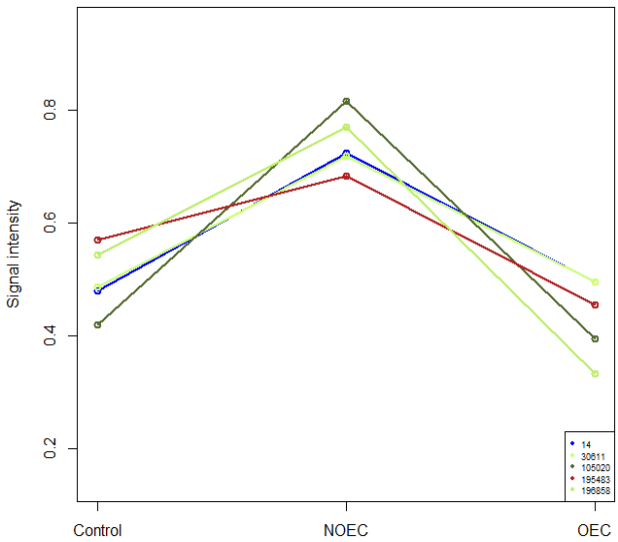
Methoxyfenozide Cluster C



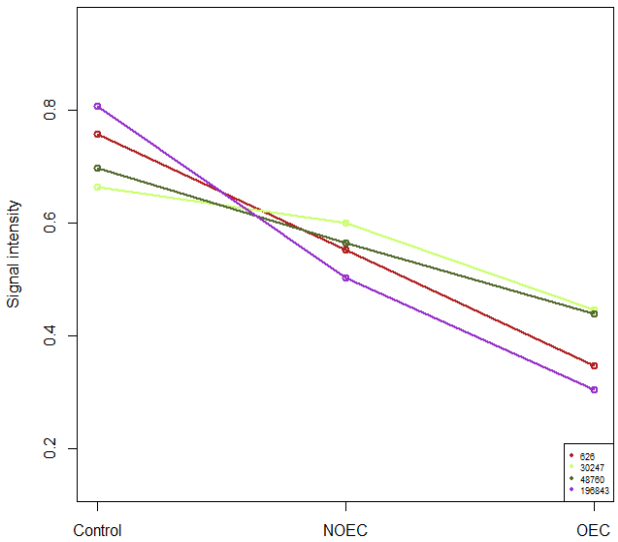
Methoxyfenozide Cluster D



Methoxyfenozide Cluster E



Methoxyfenozide Cluster F



CHAPITRE V : Synthèse et discussion générale

Ce dernier chapitre présente une synthèse de ces travaux de thèse, résumant les principaux résultats détaillés dans les publications précédemment présentées.

La première partie (V-1) expose les différents verrous à l'application de biomarqueurs chez les invertébrés, pour l'évaluation écotoxicologique des milieux et les nouvelles opportunités offertes par les technologies de séquençage nouvelle génération.

La deuxième partie (V-2) présente la stratégie protéogénomique mise en place chez *Gammarus fossarum*, ainsi que les différentes applications pouvant en découler.

Enfin la dernière partie (V-3) se concentre sur le développement d'outils pour l'étude de la reprotoxicité. Dans un premier temps, les protéines identifiées comme étant impliquées dans le processus reproductif sont présentées. Puis, les principaux résultats concernant l'étude des altérations du protéome et du succès reproducteur chez le mâle en réponse à la perturbation endocrine sont exposés.

1. Les biomarqueurs chez les invertébrés : limites et nouvelles opportunités

1.1. Limites à l'utilisation des biomarqueurs

Alors que les biomarqueurs sont considérés comme des outils pertinents pour détecter l'exposition et/ou les effets d'un polluant, ces derniers ne sont pas inclus dans les programmes de surveillance des milieux aquatiques au même titre que les indices biocénotiques. En prenant comme modèle biologique les amphipodes du genre *Gammarus*, nous avons réalisé un bilan actuel de l'utilisation des biomarqueurs protéiques en écotoxicologie aquatique (**Publication 1**). Dans 70 % des études conduites (*n*=29), l'application des biomarqueurs se fait en laboratoire dans une approche prédictive pour l'évaluation du risque. De même, les biomarqueurs mesurés rendent davantage compte d'une perturbation générale en réponse à la pression toxique : défense antioxydative ou induction des HSP que spécifique d'un type de perturbation. Ce constat trouve son origine de limites [1] d'ordre analytique et [2] dans l'interprétation des réponses.

Le test ELISA constitue la méthode de choix pour l'identification et la quantification des protéines. Le développement de ce type de méthode se révèle relativement coûteux, de par la nécessité de purifier la protéine, produire les anticorps et s'assurer de leur spécificité. Chez *Gammarus*, certaines protéines ont été suivies par Western Blot telles que les HSP ou les métallothionéines, en utilisant des anticorps provenant d'organismes vertébrés. Toutefois, l'analyse de la bibliographie a mis en évidence qu'aucune caractérisation biochimique préalable n'avait été réalisée : *i.e*. spécificité de la réaction ou phénomène de réactivité croisée, questionnant la robustesse de l'analyse.

En écotoxicologie, les méthodes analytiques actuellement employées chez les invertébrés se basent majoritairement sur des méthodes indirectes par le suivi de l'activité enzymatique (75 % des biomarqueurs mesurés chez *Gammarus*), limitant l'étude des protéines non enzymatiques. Aussi, pour chaque marqueur enzymatique, un protocole analytique dédié doit être employé (*i.e*. tampon d'homogénéisation, milieu réactionnel) entrainant une démultiplication des analyses dans une démarche de diagnostic multi-biomarqueurs. Cette lourdeur analytique freine les étapes de validation de la réponse du marqueur ainsi que son application dans un contexte de biosurveillance à large échelle.

Aussi, un des constats tirés de notre revue bibliographique est que les biomarqueurs employés chez les invertébrés résultent de la transposition directe de méthodes ayant fait leur preuve chez des organismes vertébrés. Or, en raison de la distance phylogénétique, ces outils peuvent se révéler peu sensibles et discriminants face au stress toxique. Un des cas les plus documentés est celui du suivi de l'activité EROD rendant compte de l'activité de métabolisation du cytochrome P450. Alors qu'il se révèle sensible et pertinent chez les poissons, chez les crustacés aucune induction suivant une relation dose-réponse n'a pu être démontrée (James and Boyle 1998). Chez le gammare, aucune induction n'a pu être mesurée (Leroy *et al.* 2010). Cette absence de développement spécifique chez les invertébrés s'explique par l'absence de connaissances fondamentales sur ces espèces dites « non-modèles » pour les biologistes moléculaires.

1.2. Intérêt des nouvelles techniques omiques

Alors qu'en biochimie des protéines, le Western Blot était la méthode de routine pour détecter et quantifier les protéines, la spectrométrie de masse est en phase de devenir la méthode prévalente dans ce domaine (Aebersold *et al.* 2013). L'emploi de méthodologie ciblée hautement spécifique SRM/MRN (Selected/ Multiple Reaction Monitoring) pour la quantification de peptides devient de plus en plus routinier en bioanalyse médicale. Dans la **Publication 1**, nous avons plaidé pour l'utilisation de ce type d'approche dans l'évaluation écotoxicologique, qui permettrait de lever différents verrous méthodologiques en termes de :

[1] spécificité, les techniques SRM/MRN ciblant les peptides spécifiques de la protéine d'intérêt.

[2] exhaustivité, en ouvrant la possibilité de quantifier des protéines non enzymatiques.

[3] rendement analytique, ces techniques permettent la quantification de plusieurs centaines de protéines sur un même échantillon biologique (*i.e*. multiplexage) (Percy *et al.* 2014), facilitant ainsi la validation et l'application des biomarqueurs.

L'absence de connaissance fondamentale chez les invertébrés employés dans l'évaluation écotoxicologique reste un verrou fort pour l'utilisation de ces techniques, les protéines et leurs séquences associées étant peu caractérisées chez ces organismes. Les avancées dans le domaine du séquençage de l'information en acides nucléiques et en acides aminés permettent à présent la découverte rapide d'un vaste ensemble de gènes et de protéines chez les espèces non-modèles.

Comme présenté dans la **Publication 2**, ces approches se sont principalement démocratisées pour le séquençage de l'acide nucléique. L'approche la plus couramment employée est celle du séquençage du transcriptome, technique désormais appliquée chez une large gamme d'organismes. En revanche, le séquençage des protéines se révèle techniquement plus exigeant en raison de l'importante gamme dynamique du protéome et des multiples combinaisons possibles en acides aminés. Par conséquent, l'information en acides aminés se révèle moins étudiée que celle en acides nucléiques alors que le protéome est au plus proche du phénotype des organismes. En écotoxicologie aquatique, nous avons montré que l'étude du protéome est portée par une stratégie d'identification des protéines par homologie et non de découverte, en raison de l'absence de BD de séquences protéiques spécifiques pour l'interprétation des spectres MS. Cette approche par homologie permet l'identification des protéines les plus conservées et les plus ubiquitaires au sein du vivant : ATP synthase, protéine du cytosquelette, arginine kinase. Or, il a été démontré chez la daphnie que les gènes spécifiques à l'organisme (gènes orphelins) sont ceux qui répondent le plus aux stress environnementaux (Colbourne *et al.* 2011). C'est pourquoi afin d'appréhender les mécanismes moléculaires mis en place par l'organisme pour faire face aux changements de son environnement, il est nécessaire d'adopter une démarche de découverte des protéines.

1.3. La protéogénomique pour la découverte de protéines chez les espèces non modèles

Une stratégie efficace pour la découverte de protéines chez les eucaryotes est le séquençage en parallèle du transcriptome et du protéome sur un même échantillon biologique, afin de générer à partir du séquençage des ARNm, une BD protéique spécifique pour l'interprétation des spectres MS/MS (Wang *et al.* 2012). Comme présenté en **Publication 2**, l'utilisation de ces deux techniques combinées, qualifiée d'approche « protéogénomique », se révèle prometteuse pour la découverte de protéines chez des espèces non modèles et est appelée à se démocratiser ces prochaines années. Chez notre modèle d'étude *Gammarus fossarum*, nous avons proposé dans la **Publication 1**, une stratégie pour le développement et l'application de nouveaux biomarqueurs pour l'évaluation de la reprotoxicité, en combinant les approches de protéogénomique et de quantification ciblée MRN.

En revanche, la découverte de protéines par protéogénomique ne permet pas de renseigner sur la fonction biologique qu'elles occupent. A cette fin, nous avons proposé deux approches

complémentaires basées sur [1] le traitement bio-informatique des séquences et [2] sur l'observation empirique.

La première stratégie repose sur la comparaison des séquences de notre modèle d'étude avec les séquences présentes sur les bases de données publiques. Si une séquence homologue est rencontrée et que celle-ci présente une annotation fonctionnelle, alors la même fonction est inférée à notre modèle d'étude. Toutefois, cette prédiction peut être erronée car bien que les séquences soient proches, la protéine de notre modèle d'étude a pu acquérir de nouvelles fonctions au cours de l'évolution. C'est pourquoi, pour certaines familles de protéines dont l'histoire évolutive a été reconstruite, la protéine de notre modèle d'étude peut être intégrée au sein d'arbre phylogénétique permettant d'affiner l'annotation fonctionnelle.

La seconde stratégie s'inspire directement de travaux en physiologie animale et repose sur l'étude empirique de l'évolution de l'abondance des protéines au cours de différents processus physiologiques, démarche rendue possible chez le gammare grâce à la connaissance de sa physiologie reproductive.

2. Mise en place d'une approche protéogénomique chez *Gammarus fossarum*

2.1. Création d'un catalogue de protéines

Afin d'optimiser la découverte de protéines d'intérêt dans la reproduction, le séquençage en parallèle du transcriptome et protéome a été réalisé sur trois tissus impliqués dans le processus reproductif: les tissus reproducteurs mâles et femelles et le céphalon. Les *cæca* hépatopancréatiques ont également été utilisés, ces derniers étant le centre métabolique de l'organisme et concentrant différentes réponses d'intérêt en écotoxicologie (**Publication 3**) (Figure V-1).

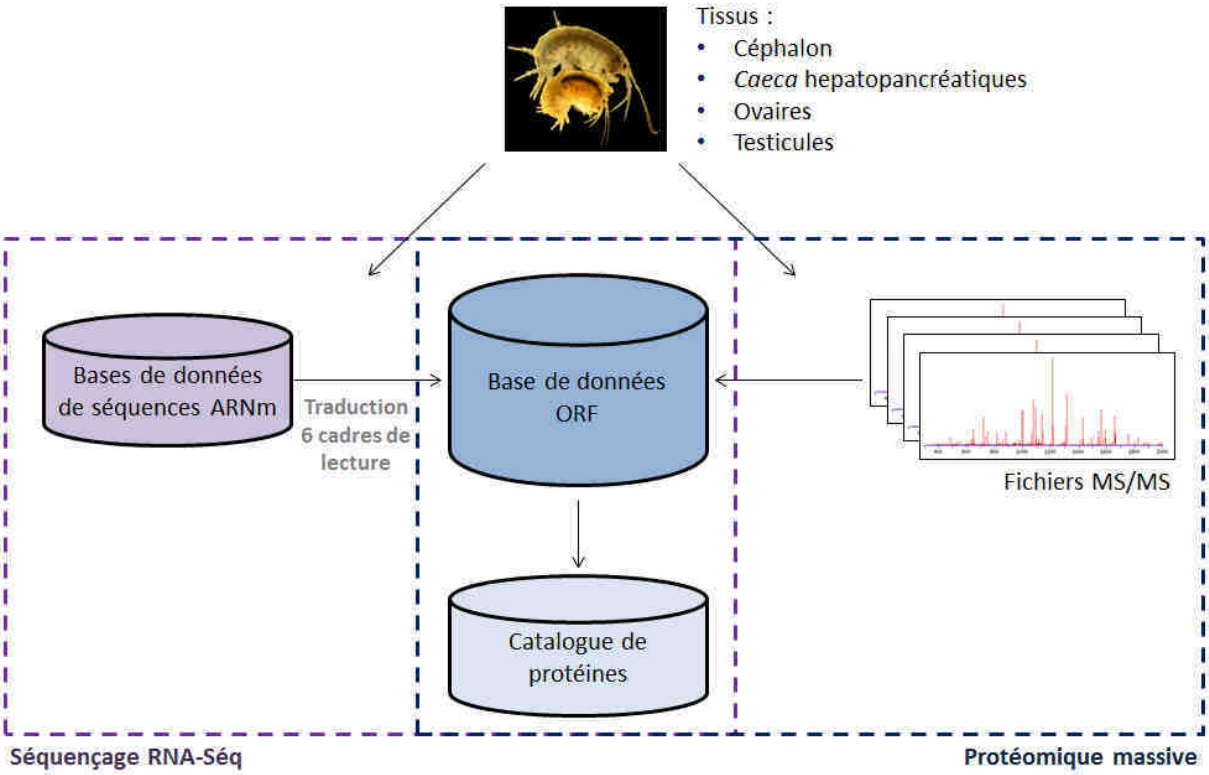


Figure V-1 : Stratégie protéogénomique adoptée pour la découverte de protéines chez *Gammarus fossarum*

Afin de disposer d'une BD générique, les organismes choisis pour générer le transcriptome sont issus du complexe *fossarum-pulex* en échantillonnant des organismes de tailles hétérogènes au sein de la population de la Tour du Pin. Au contraire, pour les expériences protéomiques, seuls des individus de classe de taille homogène et ayant été identifiés comme étant *fossarum* sur des critères morphologiques, ont été employés.

Le séquençage du transcriptome a permis d'identifier 218 574 contigs dont 58 % ne possèdent pas d'homologues sur les bases de données publiques, ainsi que 386 238 singletons. Seuls les contigs ont été employés pour créer la base de données ORF, ces derniers étant issus de l'assemblage de séquences multiples alors que les singletons sont principalement artéfactuels ou fragmentaires car provenant de transcrits de faible abondance (Meyer *et al.* 2009).

Les différentes expérimentations protéomiques ont permis de confirmer l'existence d'un total de 2257 protéines par nanoLC/MS-MS (Annexe I), dont 278 protéines orphelines (12 %) (Figure V-2). L'organisme présentant le plus grand nombre d'homologues est *Daphnia pulex*, seul crustacé ayant son génome séquencé et complétement assemblé (Colbourne *et al.* 2011).

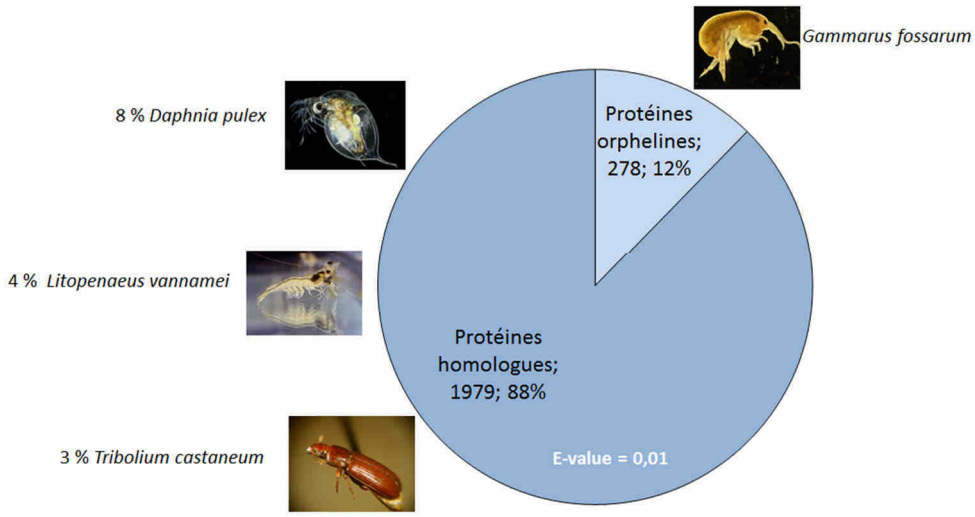


Figure V-2 : Homologie des protéines identifiées chez *G. fossarum* par l'interrogation de la base de données NCBI (Evalue = 0,01)

2.2. La base de données ORF : fondement pour une bonne annotation structurale

2.2.1. Détection d'erreurs d'assemblage du transcriptome

Bien que dans un souci de qualité, seuls les contigs ont été utilisés pour créer la BD ORF, les expérimentations protéomiques ont mis en évidence que certains contigs étaient lus selon différents ORF, ces changements de cadre de lecture s'opérant toujours dans le même sens (sens ou anti-sens). Au cours de la création du catalogue de protéines (**Publication 3**), la présence de

plusieurs ORF a été détectée pour 64 contigs. Ces observations peuvent être expliquées au niveau [1] biologique par des processus post-traductionnelles d'épissage et [2] technique par l'existence d'erreurs de séquençage *i.e.* addition ou délétion d'un nucléotide et d'erreurs au cours de l'assemblage *de novo* du transcriptome. Les proportions attribuables aux processus biologiques ou aux artéfacts analytiques ne peuvent être définies. Toutefois, l'absence de génome de référence affecte certainement la qualité d'assemblage du transcriptome.

La qualité de la BD ORF impacte fortement l'annotation structurale des protéines. Certaines ne peuvent être découvertes, les peptides séquencés ne pouvant être interprétés car présents dans des zones de changement de cadre de lecture ou de codons STOP sur la séquence prédite.

De même, comme présenté en Figure V-3, il est envisageable que certaines protéines soient chimériques, le contig réassemblé appartenant en réalité à deux brins d'ADNc distincts (Hypothèse 2) ou au contraire que deux protéines distinctes validées par spectrométrie de masse qui n'en forme en réalité qu'une (Hypothèse 1).

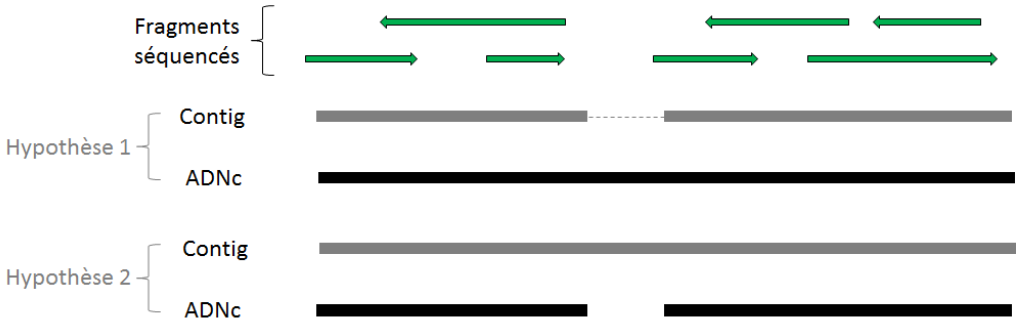


Figure V-3 : Origine des protéines chimériques prédites par le séquençage du transcriptome

Un meilleur assemblage du transcriptome permettrait d'améliorer l'annotation structurale et d'augmenter le taux de découverte des protéines.

2.2.2. Stratégies pour une meilleure annotation structurale des protéines

Par rapport aux études transcriptomiques chez les espèces dépourvues de génome de référence, notre démarche protéogénomique permet d'améliorer l'annotation structurale de la protéine par la cartographie des peptides séquencés sur la séquence prédite (**Note 2**). Toutefois, ce travail a été réalisé pour certaines protéines d'intérêt présentant une bonne couverture de

séquençage par nanoLC-MS/MS. Afin d'améliorer de manière globale l'annotation structurale ainsi que la découverte de protéines, un travail en amont sur la qualité de la BD ORF ainsi que sur son caractère générique doit être entrepris. Deux stratégies sont ici proposées, chacune s'adressant à deux objectifs distincts : [1] le développement de biomarqueurs robustes pour le suivi de la reprotoxicité en utilisant des organismes de la population de référence de la Tour du Pin et [2] la possibilité de disposer d'une BD générique ouvrant des perspectives en terme de caractérisation du protéome issu d'autres tissus ou d'autres populations au sein du complexe *fossarum-pulex*.

La première stratégie repose sur l'augmentation de la redondance de couverture du transcriptome par la réalisation d'un nouveau séquençage RNA-Seq par Illumina ou une technique NGS alternative. Dans le cadre de ces travaux de thèse, la technique Illumina/Solexa a été choisie. Cette dernière domine le marché des technologies NGS en raison de sa rapidité et de son faible coût. Elle est caractérisée par le séquençage de fragments courts (~300 bp) et par une importante profondeur de lecture permettant la détection de phénomène de polymorphisme. A l'opposé, la technique de pyroséquençage Roche 454 et son successeur PacBio permettent le séquençage de fragments longs (454 : 1 kb et PacBio 20 kb), mais avec une faible profondeur de lecture, générant davantage de séquences consensus. Elles se révèlent plus appropriées pour l'assemblage *de novo* mais présentent un coût plus prohibitif (résumé par van Dijk *et al.* 2014). He *et al.* (2012) ont réalisé le séquençage du transcriptome du roseau commun *Pharagmites australis* en utilisant la technique Illumina et 454 puis ont produit trois assemblages : individuel ou hydride. Individuellement, la technique Illumina permet d'apporter une meilleure couverture et profondeur de lecture du transcriptome. Toutefois, la méthode hydride produit des données de meilleure qualité, la taille des contigs étant augmentée et un meilleure assemblage, connectant des contigs non chevauchants au niveau individuel. En revanche, les auteurs n'ont pu déterminer si la combinaison intrinsèque des deux techniques ou si l'augmentation de la redondance du séquençage étaient à l'origine de cette amélioration.

Alors que cette première stratégie est pertinente pour l'étude du protéome chez des gammares standardisés originaires de la population référence de la Tour du Pin, elle montre ses limites dans une démarche de caractérisation spécifique du protéome au sein de la diversité du complexe *fossarum/pulex*. Bien que le transcriptome ait été réalisé à partir d'un assemblage des deux espèces, le taux d'attribution des spectres pour le protéome des ovaires des femelles *G. fossarum* de la Tour du Pin était de 20 %, il tombe à 8,5 % lors du séquençage des mêmes tissus

chez une population de *G. pulex* de la Tanche (Jura) (Trapp *et al.* 2015), ces deux populations étant issues de lignées phylogénétiques différentes (Vigneron *et al.* 2014). Aussi, notre BD a été construite à partir de tissus impliqués dans la fonction reproductrice, limitant l'étude d'autres tissus d'intérêt en écotoxicologie tels que les branchies, surface d'échange avec le milieu extérieur en première ligne face à la contamination ou la chaine de ganglion nerveux, cible de certains pesticides neurotoxiques ou médicaments psychotropes.

Alors que les techniques NGS, deviennent de plus en plus répandues et abordables, les études utilisant des BD dérivées du transcriptome pour la découverte de nouvelles protéines se multiplient chez les organismes modèles et sont même appelées à devenir routinières (Wang *et al.* 2012, Menschaert *et al.* 2013, Omasits *et al.* 2013, Chocu *et al.* 2014). Dans une perspective de baisse continue de coût de séquençage NGS, il est donc envisageable de créer une BD personnalisée pour les lignées phylogénétiques *fossarum*/*pulex* ou les tissus considérés. Toutefois pour les organismes modèles, cette approche protéogénomique systématique est facilitée par l'existence d'un génome de référence permettant un assemblage rapide et de qualité du transcriptome. Par conséquent, à terme, une stratégie plus ambitieuse serait d'initier un projet de séquençage du génome de notre modèle d'étude. Motivée par la baisse des coûts de séquençage, l'initiative i5k (5000 arthropod genome sequencing initiative) a vu le jour et intègre en son sein deux génomes d'amphipodes, celui de *Parhyale hawaiensis* un modèle en Evo-Devo et l'alter-ego à notre modèle au niveau nord-américain, *Hyalella azteca*. Ce dernier étant également issu d'un complexe d'espèces cryptiques (Major *et al.* 2013), la stratégie choisie a été de créer une «lignée consanguine» à partir d'organismes élevés en laboratoire afin d'abaisser la diversité génétique. Le matériel génétique a été séquencé dans un premier temps par la technique 454 afin de créer des séquences consensus (Poynton *et al.* 2013). Outre les perspectives scientifiques offertes par l'obtention du génome, d'un point de vue de l'annotation structurale, une stratégie similaire de création de génome consensus sur notre population de référence permettrait [1] de disposer d'une ossature pour améliorer l'assemblage des reads obtenus en transcriptomique et [2] réaliser des études protéogénomiques au sens originel par la réannotation du génome.

2.3. Intérêt de la ressource protéogénomique

Jusqu'à récemment, les ressources moléculaires étaient rares chez les pancrustacés et leur émergence vont permettre de clarifier les relations phylogénétiques au sein de ce taxon (von Reumont *et al.* 2012). Les amphipodes sont placés au sein des malacostracés et donc proches des décapodes. Toutefois, les Vtg identifiées chez *G. fossarum* se situent de manière intermédiaire auprès de celles des insectes et branchiopodes et auprès des décapodes (**Publication 4**). De même, au cours de la création du catalogue de protéines (**Publication 3**), différentes protéines impliquées dans le métabolisme des hormones juvénoïdes ont été séquencées dont celles du catabolisme, annotées sous le terme d'hormone juvénile. Alors que le terme en lui-même peut se référer de manière générique aux hormones terpenoïdes des arthropodes, une de ces protéines est une époxyde hydrolase, dégradant spécifiquement le groupement époxyde, groupement supposé être une nouveauté évolutive unique aux insectes (Figure V-4) (De Loof *et al.* 2013). Au cours de ces travaux, il a été mis en évidence que certaines prédictions fonctionnelles peuvent être mal avisées. Toutefois cette observation pose la question de la nature chimique des hormones juvénoïdes chez les amphipodes et par conséquent de leur place au sein des pancrustacés.

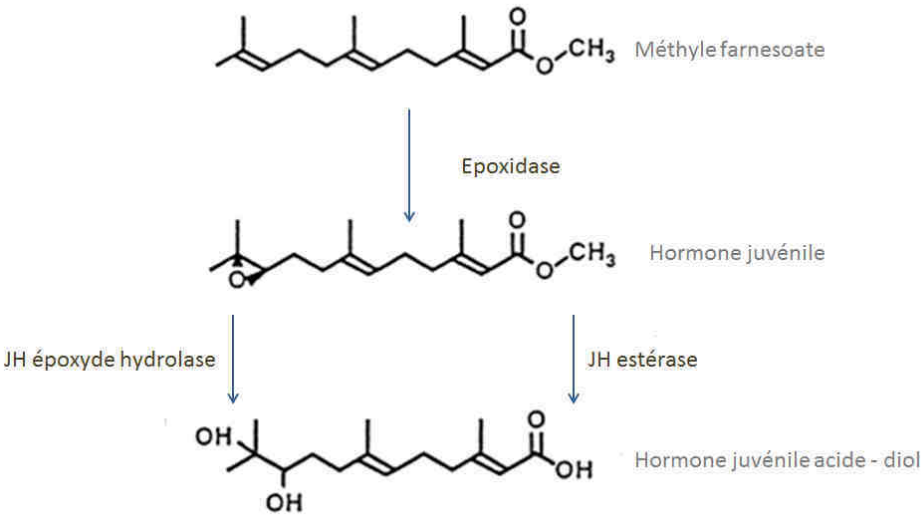


Figure V-4 : Synthèse et dégradation de l'hormone juvénile chez les insectes

Le complexe *fossarum/pulex* est également un modèle d'étude pour comprendre les mécanismes liés aux phénomènes de cryptisme (Lagrue *et al.* 2014). Dans la **Publication 3**, une proportion plus importante de protéines orphelines au sein des tissus reproducteurs a été mise en évidence face au

céphalon. De cette divergence, des phénomènes de spéciation entre espèces proches peuvent émerger (Clark *et al.* 2006). Les protéines orphelines identifiées de manière spécifique dans les tissus reproducteurs sont des candidats intéressants pour étudier ces processus au sein du complexe *fossarum/pulex.*

Parmi les 2257 protéines séquencées au cours de ces travaux, la recherche d'homologues par BLAST sur NCBI a permis de prédire la fonction de 1453 (64 %) d'entre-elles. Concernant une utilisation en écotoxicologie, plusieurs protéines d'intérêt ont été directement identifiées sur la base de leur annotation fonctionnelle, permettant d'alimenter une démarche de développement de méthodologie de quantification multiplexée (Annexe II). Différentes protéines impliquées dans la défense antioxydative ont été séquencées : catalase, peroxydase et superoxyde dismutase, la glutathion-S-transférase impliquée dans la détoxification des xénobiotiques, différentes protéines ayant un rôle dans la digestion : cellulase, chitinase, cellobiohydrolase et amylase, des protéines de l'immunité : peptide antimicrobien, protéine de reconnaissance des pathogènes, hémocytine, hémolectine, prophénoloxydase et transglutaminase, le transporteur ionique Na^+/K^+ ATPase et enfin la dopamine beta hydrolase, impliquée dans la synthèse des neurotransmetteurs. Certaines de ces protéines ont des fonctions pléiotropiques, la chitinase et la cellulase étant également impliquées dans la défense des pathogènes. Une analyse de la phylogénie moléculaire des différentes formes séquencées pourrait éclairer sur leur spécificité. Il est toutefois nécessaire, en amont d'une application directe en biosurveillance, de valider la pertinence de ces candidats biomarqueurs en laboratoire. En effet, certaines protéines d'intérêt sont des biomarqueurs développés chez les vertébrés et ne sont donc pas implicitement sensibles à la perturbation chez le gammare. De même, certaines prédictions fonctionnelles peuvent être erronées.

3. Développement d'outils pour l'évaluation de la reprotoxicité

3.1. Une première identification par annotation fonctionnelle automatique

Une première approche, basée sur l'identification de protéines d'intérêt par prédiction fonctionnelle, a permis d'obtenir une liste de quatorze candidats biomarqueurs, représentant différents processus clés dans la reproduction tels que le métabolisme hormonal ou la mue. Le Tableau V-1 présente ces différentes protéines d'intérêt mais n'inclue pas celles de type Vtg-like, ces dernière faisant l'objet d'une identification spécifique en partie V-3.2.2.

Tableau V-1 : Première identification de candidats biomarqueurs impliqués dans la reproduction sur la base des prédictions fonctionnelles

Fonction prédite	Identité de la protéine	Tissus / Expérimentation
Acide farnesoïque O-méthyltransférase	134275_fr5	Céphalon, testicules
Acide farnesoïque O-méthyltransférase	166723_fr2	Testicules
Hormone juvénile carboxyle-estérase 1	144144_fr2	Spermatogénèse, reprotoxicité
Hormone juvénile époxyde hydrolase	191918_fr1	Testicules
17-bêta-hydroxystéroid déshydrogénase	141191_fr5	Testicules, spermatogénèse
17-bêta-hydroxystéroid déshydrogénase	121329_fr5	Testicules
17-bêta-hydroxystéroid déshydrogénase	101072_fr3	Testicules
Hydroxystéroid déshydrogénase	147007_fr3	Spermatogénèse
Cytochrome P450 CYP12A2	122081_fr5	Ovaires
Cytochrome P450 enzyme, CYP4C39	100255_fr4	Testicules
Chitinase 2	12415_fr3	Céphalon
Epididymal sperm-binding protein 1	14545_fr1	Spermatogénèse, reprotoxicité
Epididymal sperm-binding protein 1	186168_fr5	Spermatogénèse, reprotoxicité
Epididymal sperm-binding protein 1	18609_fr3	Spermatogénèse

Plusieurs protéines impliquées dans le métabolisme des hormones terpenoïdes ont été séquencées (Figure V-4), majoritairement dans le tissu reproducteur masculin, renforçant l'hypothèse que l'hormone juvénile joue un rôle de déterminant sexuel masculin (Nagaraju 2007). Deux formes de CYP450 ont été identifiées dans les tissus reproducteurs mâles et femelles, suggérant une possible implication dans le métabolisme des ecdystéroïdes. Plus surprenant, dans le tissu reproducteur mâle, trois protéines annotées comme étant des 17-bêta-hydroxystéroid déshydrogénase ont été séquencées. Ces dernières sont supposées catalyser l'interconversion de l'estrone en œstradiol ainsi que de l'androstenedione et de la testostérone (Mindnich *et al.* 2004). Alors que les arthropodes n'ont pas d'hormones stéroïdiennes sexuelles vertébrés, cette identification peut être

expliquée comme [1] un mécanisme de détoxification face à ces molécules ubiquitaires dans l'environnement (Scott 2012) ou [2] ces protéines ont acquis une nouvelle fonction dans le métabolisme des hormones stéroïdiennes d'arthropodes et sont paralogues à celles des vertébrés.

Plusieurs protéines annotées comme étant des chitinases ont été séquencées au cours de ces travaux de thèse, leur analyse phylogénétique (Huang *et al.* 2010) ainsi que leur localisation tissulaire ont mis en évidence que la chitinase 2 est impliquée dans le processus de mue, qui par conséquent, est un bon candidat pour suivre les altérations de ce processus au niveau moléculaire.

Enfin, plusieurs protéines annotées comme epididymal sperm-binding protein ont été identifiées, ces dernières se lient aux spermatozoïdes sénescents afin d'initier leur élimination (D'Amours *et al.* 2012)

3.2. Identification de protéines impliquées dans la physiologie reproductrice

La démarche d'identification de nouvelles protéines d'intérêt dans la reproduction se décline selon deux approches : [1] l'identification de protéines présentant un dimorphisme sexuel en comparant l'abondance des protéines des tissus reproducteurs mâles et femelles et [2] le suivi de la dynamique du protéome des ovaires, des embryons et des testicules au cours respectivement de l'ovogénèse, de l'embryogénèse et de la spermatogénèse.

3.2.1. Protéomique comparative des tissus reproducteurs mâles et femelles

L'étape de protéomique comparative sur les données obtenues au cours de la création du catalogue protéique (**Publication 3**), a permis d'identifier respectivement 129 et 75 protéines spécifiques du tissu reproducteur mâle et femelle (Figure V-5) (Annexe III). Basé sur les valeurs de fold change, le dimorphisme sexuel se révèle plus prononcé pour les protéines spécifiques de la femelle, en raison du rôle des ovaires comme tissu de réserve pour le développement embryonnaire, accumulant certaines protéines en très forte abondance.

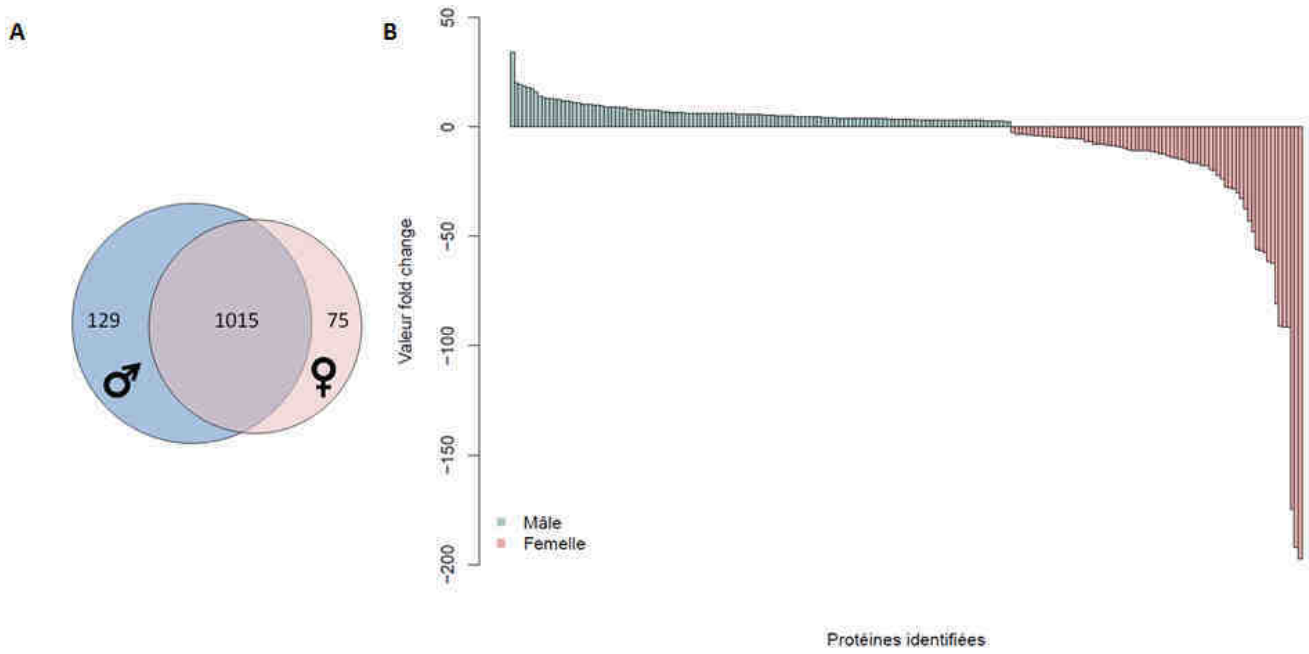


Figure V-5 : Analyse comparative de l'abondance des protéines issues des tissus reproducteurs mâle et femelle

(A) Nombre de protéines présentant un dimorphisme sexuel significatif (fold change> 2 et p-value<0,05) (B) Distribution de ces protéines selon la valeur de fold change.

Chez les femelles, 40 % des protéines sont orphelines et 33 % sont membres de la famille des lipoprotéines LLTP. Alors que chez les mâles, 41 % des protéines sont dépourvues d'annotation fonctionnelle, incluant 8 % d'orphelines. Les autres protéines spécifiques du mâle incluent majoritairement des protéines ubiquitaires de maintenance mais plusieurs protéines sont impliquées dans la maintenance de l'homéostasie calcique ou sous dépendance du calcium, comme la protéine Copine-8. Cette dernière est exprimée de manière prédominante dans la prostate et les testicules chez les humains, suggérant un rôle dans le développement et le maintien des caractères sexuels masculins (Maitra *et al.* 2003). On retrouve également plusieurs protéines potentiellement impliquées dans la réaction de l'acrosome : une protéine membranaire la flotilin-1 et des glycohydrolases (Tulsiani *et al.* 1998, Wu *et al.* 2012).

3.2.2. Protéines impliquées dans l'ovogénèse

L'étude de la dynamique du protéome des ovaires (**Publication 4**) a mis en évidence la modulation de 27 protéines au cours de l'ovogénèse, distribuées en trois profils distincts (A, B et C). Le cluster A (Figure V-6) est constitué de deux protéines de la famille des LLTP et une protéine orpheline. Deux fonctions possibles ont été proposées pour ces protéines : [1] le transport des ecdystéroïdes pour la coordination de l'ovogénèse ou [2] le transport du cholestérol pour une synthèse *de novo* d'ecdystéroïdes au sein des ovaires.

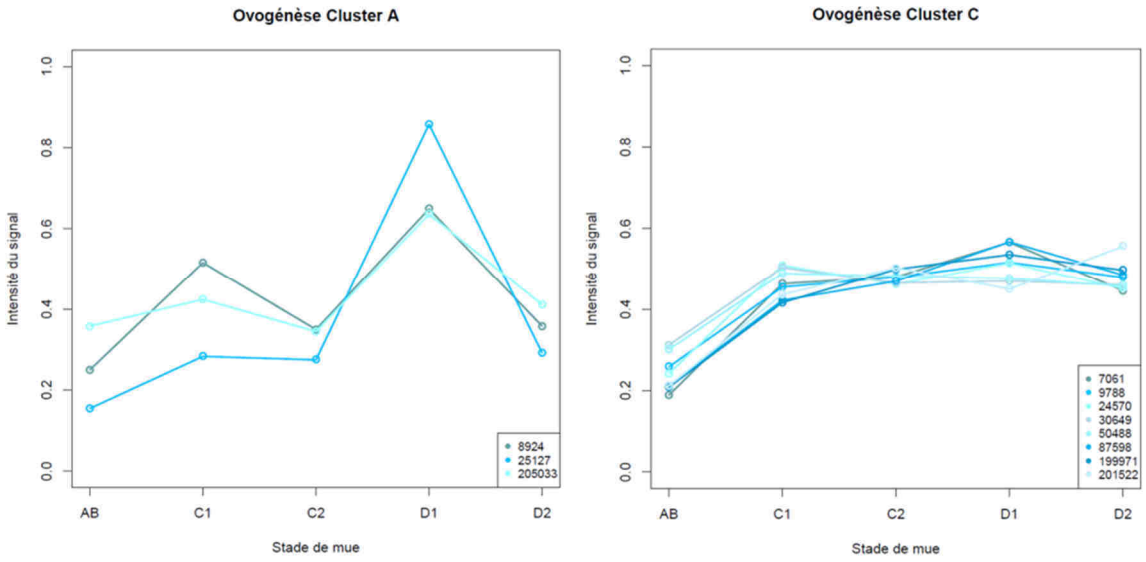


Figure V-6 : Protéines modulées au cours de l'ovogénèse selon deux profils distincts

Les différents profils ont été obtenus à partir du spectral count en utilisant le module TrendQuest du programme PatternLab. La santé du cluster a été fixée à 0.9 et le nombre minimum de protéines par cluster à 3

Les protéines incluses dans le cluster C semblent impliquées dans le processus de vitellogénèse primaire, leur abondance n'augmentant qu'en phase de post-mue alors que les protéines incluses dans le cluster B sont davantage impliquées dans la vitellogénèse secondaire, car elles présentent un profil d'accumulation constant au cours de l'ovogenèse (Figure V-7).

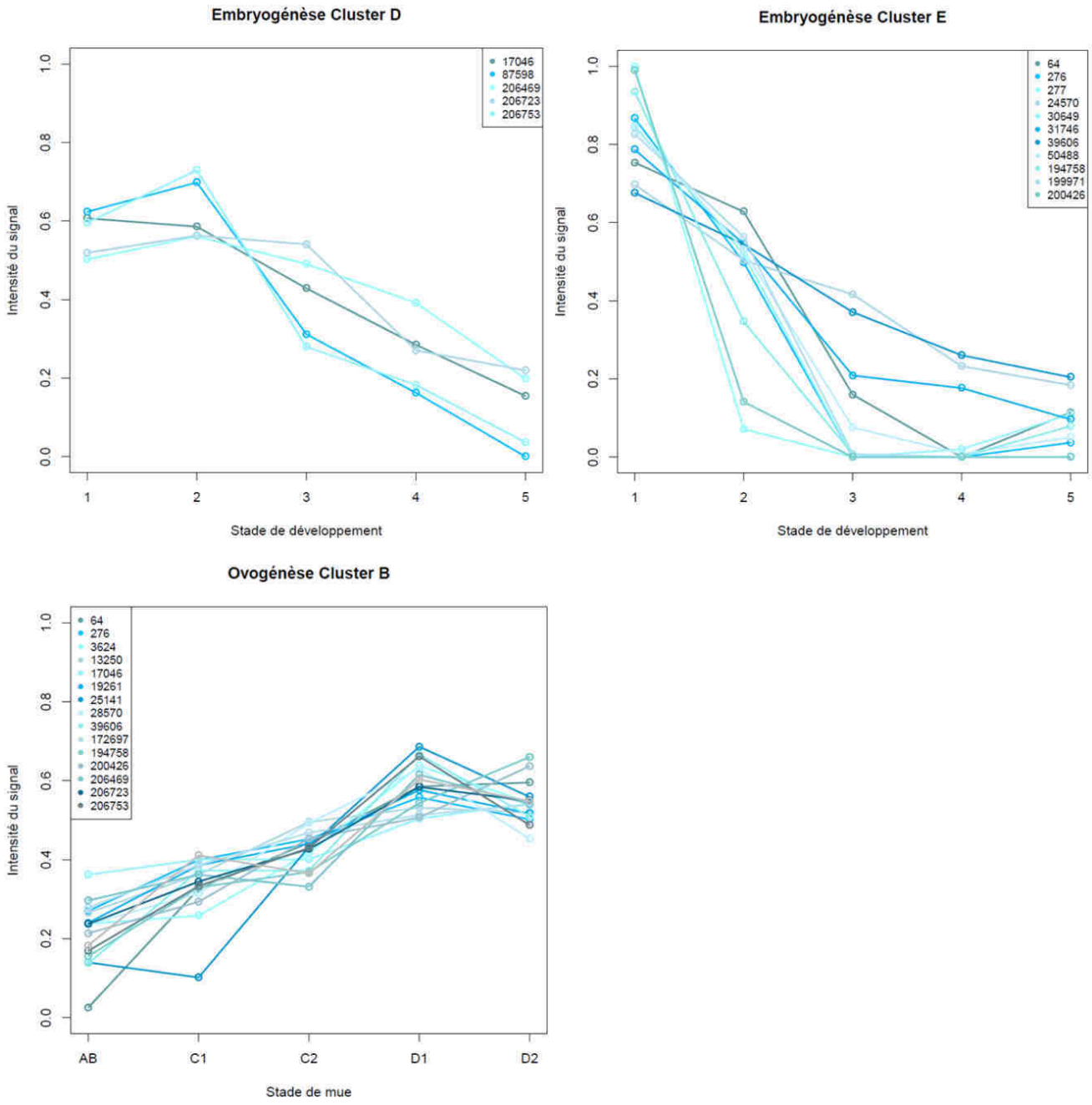


Figure V-7 : Protéines présentant un profil d'évolution au cours de l'ovogénèse et de l'embryogénèse cohérant avec un rôle de réserve du développement embryonnaire

Les différents profils ont été obtenus à partir du spectral count en utilisant le module TrendQuest du programme PatternLab. La santé du cluster a été fixée à 0.9 et le nombre minimum de protéines par cluster à 3

Afin d'identifier les protéines occupant la fonction de vitellogénine chez *G. fossarum*, la dynamique du protéome au cours de l'embryogénèse a également été étudiée. Deux clusters, D et E, présentent un profil de variation cohérant avec une utilisation de la ressource vitelline (Figure V-7). Une identification croisée entre les protéines identifiées dans le cluster B, D et E et celles classées comme étant spécifiques du tissu reproducteur femelle a permis l'identification de huit protéines occupant la fonction de vitellogénine (Tableau V-2).

Tableau V-2 : Protéines identifiées comme vitellogénines chez *G. fossarum*

Fonction prédite	Identité de la protéine
Protéine de coagulation	17046_fr6
Protéine de coagulation	206469_fr3
Protéine de coagulation	39606_fr3
Protéine de coagulation	276_fr4
Apolipocrustacéine	194758_fr3
Vitellogénine	277_fr1
Vitellogénine	200426_fr2
Vitellogénine	64_fr5

L'analyse phylogénétique de ces protéines met en évidence que l'annotation fonctionnelle automatique reflète davantage la phylogénie de ces protéines que leur fonction, remettant en doute l'annotation par homologie qui s'est généralisée lors de l'arrivée sur le marché de la nouvelle génération de séquenceurs d'ADN. La seule recherche d'homologues pour l'annotation fonctionnelle peut conduire à établir de fausses hypothèses mécanistiques. Il apparait donc nécessaire d'adopter une démarche d'annotation fonctionnelle spécifique pour l'organisme étudié.

Enfin, chez le gammare, ses Vtg se partagent entre les Vtg des insectes, vraies orthologues de la Vtg et celles des décapodes, issues de la famille paralogue à la Vtg, des apolipoprotéines (Hayward *et al.* 2010, Wu *et al.* 2013).

3.2.3. Protéines impliquées dans la spermatogénèse

L'analyse du protéome des tissus mâles à sept stades différents de la spermatogénèse a mis en évidence la modulation de 17 protéines au cours de ce processus (Tableau V-3), dont la majorité est dépourvue d'annotation fonctionnelle.

Tableau V-3 : Protéines modulées au cours de la spermatogénèse

Fonction prédite	Identité de la protéine	Cluster
Hémocytine	461_fr1	A (spécifique du mâle)
Pas de prédiction	1162_fr1	A
Pas de prédiction	521_fr6	A
Heat shock protéine 70 kDa	155_fr4	B
Pas de prédiction	1344_fr4	B
Aldéhyde oxydase	2906_fr1	B
Pas de prédiction	1366_fr1	C
Pas de prédiction	2450_fr1	C
Pas de prédiction	439_fr3	C
Pas de prédiction	108_fr5	D
Pas de prédiction	202_fr5	D
Cellobiohydrolase	1736_fr5	D
Pas de prédiction	608_fr6	E (spécifique du mâle)
Hémocytine	1094_fr3	E
Déshydrogénase à chaîne courte	1277_fr6	E
Pompe calcique ATPase (SERCA)	4227_fr2	E (spécifique du mâle)
Calréticuline	5310_fr1	E

Ces protéines se distribuent sur six profils différents, démontrant la complexité de ce processus au niveau moléculaire (Figure V-8). En comparaison aux différents profils établis au cours de l'ovogénèse, ceux de la spermatogénèse sont plus difficilement interprétables.

Parmi les protéines présentant une annotation fonctionnelle, on retrouve notamment une pompe calcique, la réticulum sarcoplasmique / endoplasmique Ca^{2+} ATPase. Cette dernière est suspectée de jouer un rôle dans la séquestration de Calcium au sein de l'acrosome. Le contrôle de l'homéostasie calcique est crucial pour les spermatozoïdes, pour leur maturation, pour acquérir leur potentiel de fertilisation ainsi que pour leurs interactions avec les ovocytes (Jimenez-Gonzalez *et al.* 2006).

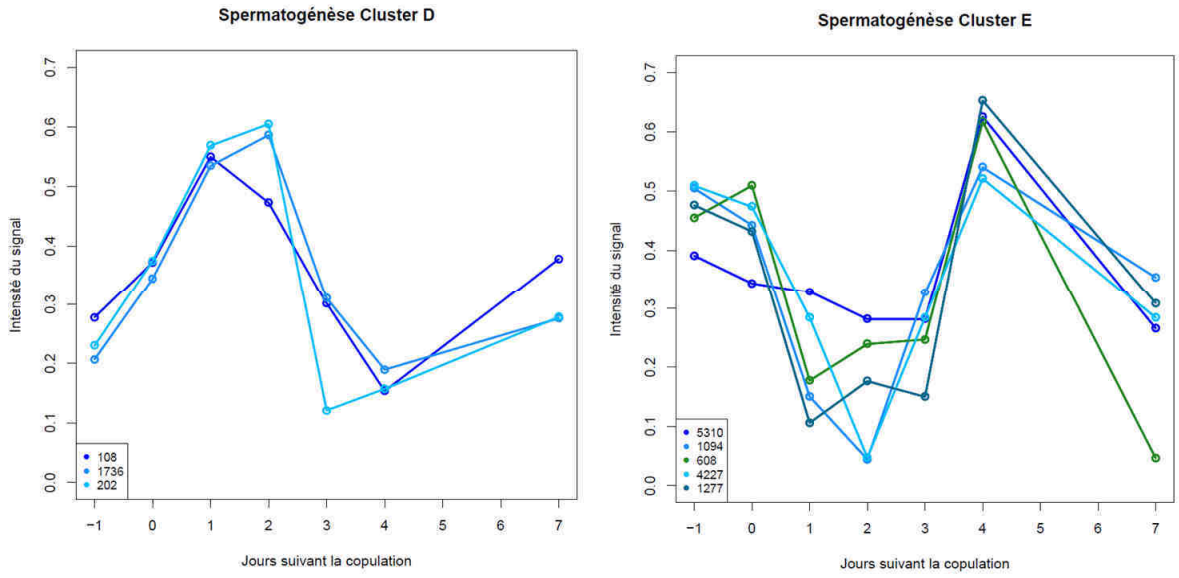


Figure V-8 : Exemples de profils d'évolution des protéines au cours de la spermatogénèse

Les différents profils ont été obtenus à partir du spectral count en utilisant le module TrendQuest du programme PatternLab. La santé du cluster a été fixée à 0.9 et le nombre minimum de protéines par cluster à 3. Les protéines apparaissant en vert sont celles ayant été préalablement classifiée comme spécifique du mâle.

Alors que le catalogue de protéines a été réalisé sur quatre tissus, la démarche d'annotation fonctionnelle a été uniquement effectuée sur les organes reproducteurs. Les protéines séquencées au sein de ces tissus se situent davantage en fin de cascade neuroendocrine en tant qu'effectrices finales des processus physiologiques, *ex* : identification de la vitellogénine. En revanche, au sein du céphalon, différentes glandes de régulation neuroendocrine sont présentes : organe X/glande du sinus, organe Y et organe mandibulaire, sécrétant respectivement les neuropeptides, les hormones ecdystéroïdiennes et les hormones terpenoïdes. Par conséquent, afin d'identifier les biomolécules en amont de la cascade neuroendocrine, c.-à-d. les neuropeptides initiant la synthèse hormonale et les enzymes la contrôlant, l'étude de leur dynamique au cours du cycle de mue chez la femelle permettrait d'identifier sans *a priori* de nouveaux biomarqueurs d'intérêt. Toutefois, alors que la BD ORF permet théoriquement d'identifier les spectres MS/MS générés par ces biomolécules, la démarche analytique shotgun adoptée au cours de ces travaux semble inadaptée pour caractériser les neuropeptides. Chez les décapodes, des travaux portant sur la caractérisation du neuropeptidome par MS reposent sur une double approche MALDI-MS/MS directement sur coupe tissulaire ou ESI-MS/MS sur extrait tissulaire, sans digestion enzymatique (résumé par Yu *et al.* 2014). Ces approches chez les décapodes sont facilitées par la taille des glandes neuroendocrines,

permettant de travailler spécifiquement sur un type de glande alors que pour *G. fossarum*, la difficulté réside dans l'isolement des différentes glandes. Toutefois, de récents développements dans le couplage des méthodes de microdissection assistée par laser et de protéomique (Xu 2010) rendent ces approches envisageables pour notre modèle d'étude.

3.3. Impact de la perturbation endocrine sur la reproduction chez le mâle

Alors que l'infertilité chez les invertébrés en réponse à la contamination est un phénomène peu étudié, son étude chez les organismes mâles est d'autant plus rare. Chez le gammare, les travaux portant sur le développement de marqueurs de génotoxicité, ont mis en évidence que les spermatozoïdes présentent plus de cassures à l'ADN que les ovocytes et les hémocytes (Lacaze *et al.* 2010). Ces derniers sont supposés être plus sensibles à la contamination car dépourvus de système de défense et de réparation (Aitken and Baker 2006). C'est pourquoi, au cours de ce travail doctoral, l'accent a été mis sur le développement d'outils pour le suivi de la reprotoxicité masculine. Pour cela, les organismes mâles ont été exposés au cours de deux spermatogénèses successives (15 jours) à trois différents xénobiotiques, le cadmium et deux pesticides, potentiels perturbateurs endocriniens le methoxyfenozide, ciblant la voie des ecdystéroïdes et le pyriproxyfen ciblant la voie des juvénoïdes. Au même titre que pour l'étape d'annotation fonctionnelle spécifique, l'approche adoptée couple l'étude du protéome des gonades et de la physiologie reproductive, par le biais de marqueurs de qualité spermatique. Cette démarche a permis d'orienter l'étude protéomique en choisissant pour chaque xénobiotique, d'étudier une concentration pour laquelle aucun effet n'est observé (NOEC ; *No Observed Effect Concentration*) et une concentration entrainant des effets néfastes au niveau physiologique (EC ; *Effective Concentration*).

3.3.1. Suivi de la qualité spermatique pour l'évaluation de l'infertilité masculine

Le suivi de la viabilité spermatique ne semble pas être un paramètre sensible chez *G. fossarum* (Lacaze *et al.* 2011), le choix s'est donc porté sur l'étude de la production et de la maturation des gamètes par la numération des spermatides et des spermatozoïdes (**Publication 5**). Tandis que la production de spermatides n'a pas été affectée par l'exposition aux polluants, la

production de spermatozoïdes s'est révélée être sensible au stress toxique, selon une relation dose-réponse (Figure V-9). Par conséquent, le choix des conditions étudiées en protéomique s'est fait sur la base de ce paramètre.

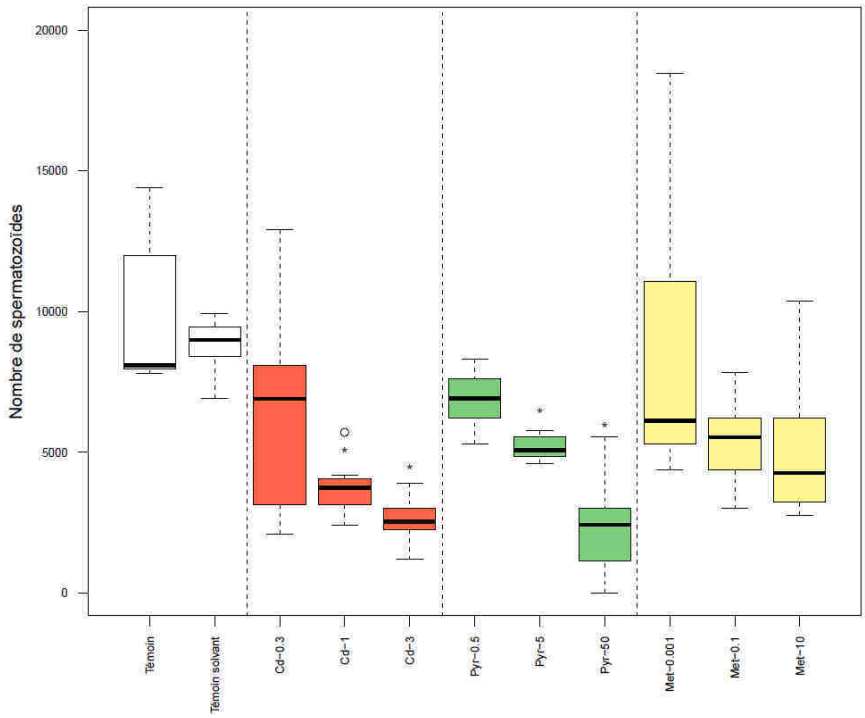


Figure V-9 : Nombre de spermatozoïdes dans les gonades après 15 jours d'exposition aux polluants

Les témoins solvant et sans solvant sont représentés en blanc ; Cd (0.3, 1, and 3 µg.L^{-1}) en rouge; Pyr (0.5, 5 and 50 µg.L^{-1}) en vert; Met (0.001, 0.1, and 10 µg.L^{-1}) en jaune. Les boites s'étendent du 1er au 3ème quartile avec une ligne en gras présentant la valeur médiane, les ronds représentent les points les plus extrêmes. *= différence significative face à la condition témoin associée (p<0,05). *n*= 7-10.

Tandis que chez les vertébrés, les effets toxiques interfèrent avec la maturation des gamètes au cours de la phase de méiose *i.e.* spermatocytes en cours de division (Evans and Ganjam 2011), chez les crustacés, la phase sensible semble être celle de différenciation. Pour seulement une seule condition, à la plus forte concentration en Pyr, des effets significatifs apparaissent sur la production de spermatides, suggérant une désorganisation au niveau tissulaire plutôt qu'un mécanisme spécifique sur la maturation cellulaire (Revathi *et al.* 2014).

Les marqueurs de qualité spermatique se révèlent donc sensibles et pertinents pour évaluer l'infertilité masculine chez notre modèle d'étude. L'intégration de ces marqueurs au sein de la batterie de biotests déjà existants permettrait de proposer une évaluation complète de la reprotoxicité chez *G. fossarum*. Toutefois, leurs futures applications dans une démarche prédictive en laboratoire ou en biosurveillance sont actuellement limitées en raison de [1] la nécessitée de travailler sur tissus frais et [2] la faible cadence analytique de numération. Ces deux paramètres limitent le nombre d'organismes pouvant être suivis dans une même journée et par conséquent conditionnent la taille des expérimentations. L'application à large échelle de ces marqueurs passe donc nécessairement par des développements analytiques.

Dans un premier temps, un travail sur la conservation de la suspension cellulaire doit être effectué afin de pouvoir traiter les échantillons biologiques *a posteriori* de l'expérimentation. Les techniques de cryoconservation des gamètes ont fait l'objet de développements poussés chez les mammifères, relatifs aux techniques d'insémination artificielle. Ces techniques requièrent l'utilisation de tampons cryopréservateurs, l'optimisation des temps de congélation et de décongélation et permet de conserver des gamètes fonctionnels sur plusieurs décennies (Yoshida 2000).

Le deuxième point méthodologique à améliorer est la méthode de numération cellulaire. La cytométrie de flux constitue la méthode analytique de choix, automatisable et à haut rendement. Elle permet de distinguer les différentes populations cellulaires, et de par l'utilisation de sondes fluorescentes, d'approfondir l'étude de la physiologie cellulaire *i.e*. phénomène d'apoptose, de génotoxicité, réaction de l'acrosome. D'autres alternatives sont également possibles, nécessitant un équipement analytique moins onéreux tel que la mesure de la turbidité cellulaire par spectrophotométrie ou la microscopie optique couplée à la prise de vue. Cette dernière alternative permet un traitement *a posteriori* des images, traitement pouvant être assisté par un logiciel, certaines solutions informatiques étant déjà développées et disponibles en libre-service, *ex* : Image J (http://rsb.info.nih.gov/ij/index.html).

3.3.2. Modifications du protéome face à la pression toxique

L'analyse du protéome des gonades mâles en réponse aux différents xénobiotiques a mis en évidence la modulation de 70 protéines (Annexe IV). Seuls les groupes de protéines présentant une variation de leur abondance entre la condition témoin et la NOEC ont été considérées, comme présenté en Figure V-10.

L'exposition des organismes mâles aux différents xénobiotiques a provoqué une surexpression des transporteurs calciques. De nombreux polluants sont capables d'inhiber ces cibles et par conséquent, cette surexpression semble être un mécanisme compensatoire pour limiter l'apparition de phénomènes d'apoptose déclenchés par une dérégulation de l'homéostasie calcique (Kirk *et al.* 2003). De même, de manière globale, la défense immunitaire a été impactée. Or, en cas d'immunodépression, l'installation de parasites féminisants est facilitée, les phénomènes d'intersexualité sur populations naturelles étant une conséquence indirecte de la contamination (Ford 2012) (voir I-2.2.2).

Seule une protéine définie par protéomique comparative des gonades mâles et femelles (V-3.2.1) comme ayant un fort dimorphisme sexuel envers la femelle a été induite en réponse à la pression toxique, une HSP 70. Cette protéine chaperon est impliquée dans une large gamme de processus biologiques et son accumulation dans les ovaires en conditions physiologiques normales sert vraisemblablement au développement embryonnaire alors que son induction chez le mâle exposé à un stress toxique favorise l'homéostasie et la survie cellulaire (Rupik *et al.* 2011). Alors que chez les vertébrés, l'exposition des mâles aux PE est accompagnée de phénomènes de féminisation avec l'induction de protéines femelles, chez les crustacés, c'est plutôt une modulation des protéines spécifiques du mâle qui est observée. A part des protéines ubiquitaires, deux protéines dépourvues d'annotation fonctionnelle (ID 9438 et 69913) semblent être de bons candidats en tant que biomarqueurs d'exposition aux PE.

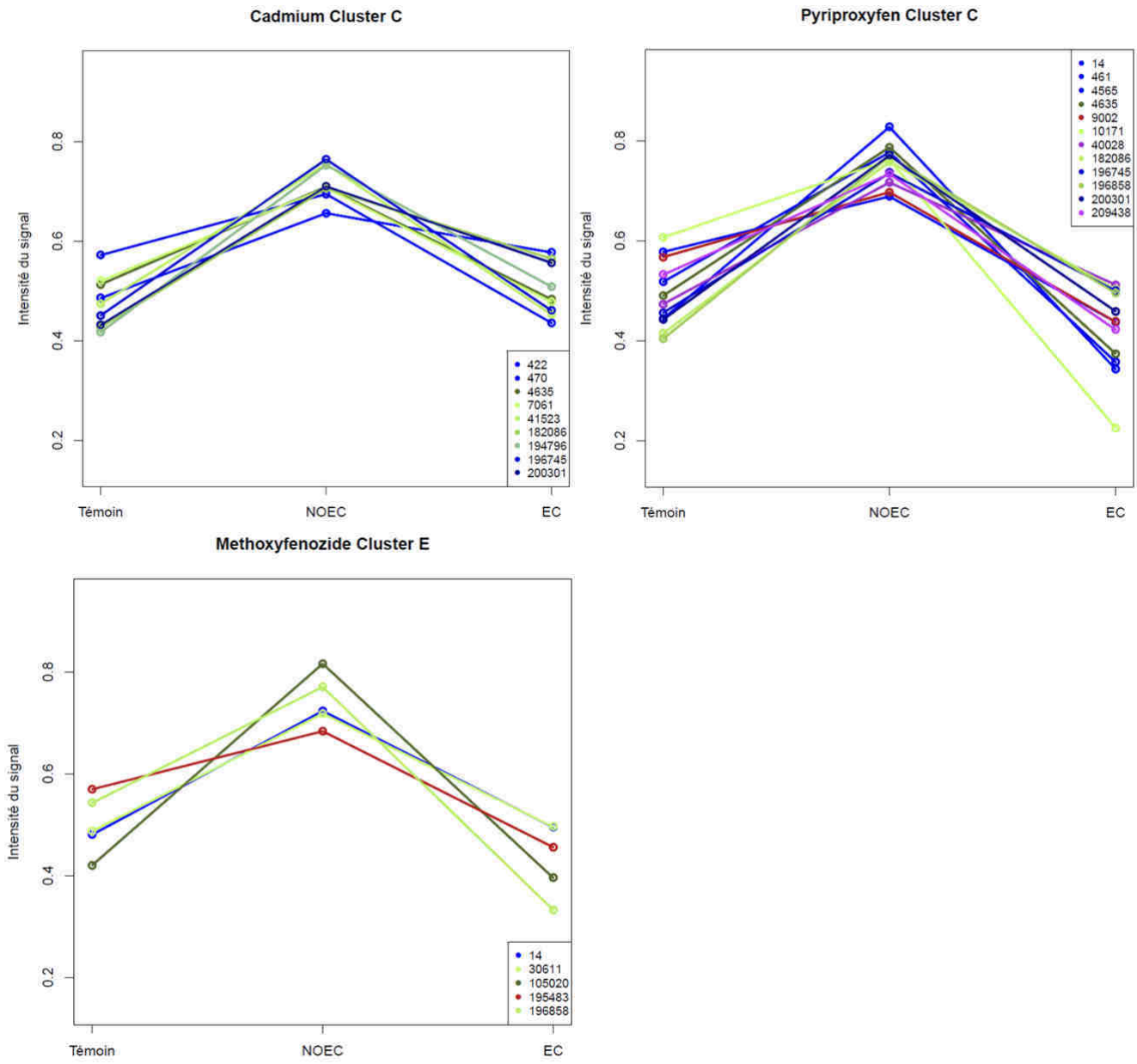


Figure V-10 : Exemples de profil d'évolution de l'abondance des protéines chez les gammares mâles exposés aux polluants

Les différents profils ont été obtenus à partir du spectral count en utilisant le module TrendQuest du programme PatternLab. Le nombre minimum de protéines par cluster est fixé à 3. Pour Pyr et Met, le témoin est le témoin solvant. La NOEC a été fixée à 0.3 µg.L^{-1}, 0.5 µg.L^{-1}, et 0.001 µg.L^{-1} pour Cd, Pyr et Met respectivement. La EC a été fixée à 3 µg.L^{-1}, 50 µg.L^{-1} et 0.1 µg.L^{-1} pour Cd, Pyr et Met respectivement. Légende : le bleu pour les protéines spécifiques du mâle, le vert pour celles présentant des profils de variation temporelle pour tous les xénobiotiques, le jaune pour celles présentant des profils de variation temporelle pour Cd et Met, le violet pour celles présentant des profils de variation temporelle pour Pyr et Met et rouge pour celle uniquement modulée par le xénobiotique considéré.

Au cours de cette dernière expérimentation en condition toxique, 20 % des protéines modulées étaient dépourvues d'annotation fonctionnelle. Ces dernières constituent un tiers des protéines sur le protéome total dont 8 % de protéines orphelines, nombreuses étant nouvellement séquencées (24 protéines homologues à la Daphnie et 18 orphelines). Les protéines issues des organes reproducteurs sont parmi les plus diversifiées au sein de l'arbre du vivant, participant aux processus d'isolement reproductif et donc de spéciation (Marshall *et al.* 2011). Ces protéines hautement divergentes peuvent donc être potentiellement impliquées dans les interactions spermatozoïdes-ovocytes. Or toutes anomalies sur ces protéines abaissent le potentiel reproducteur de l'individu (du Plessis *et al.* 2011). De même, comme mis en évidence chez la Daphnie, les gènes orphelins se révèlent être les plus sensibles face à un stress écophysiologique ou toxique (Colbourne *et al.* 2011). Toutefois, nombre de ces protéines hautement divergentes sont présentes en faible abondance dans le protéome et les techniques shotgun permettent leur découverte mais pas le suivi de leur modulation, nécessitant des techniques plus ciblées.

La recherche de candidats biomarqueurs pour détecter l'apparition d'effets reprotoxiques se sont concentrées sur le mâle, une approche identique est facilement envisageable chez la femelle, par l'utilisation du test de reprotoxicité préalablement développé (Geffard *et al.* 2010). Une telle expérimentation permettrait de compléter la liste des candidats biomarqueurs pour une évaluation complète de l'infertilité, aussi bien masculine que féminine.

CONCLUSION & PERSPECTIVES

Face à la nécessité de préserver et restaurer les masses d'eau, il est nécessaire de disposer d'outils permettant d'évaluer la qualité de l'eau et la santé des écosystèmes. Dans le cas de pollution chimique, les biomarqueurs réagissent de manière précoce et sont capables de faire le lien entre la présence de xénobiotiques et l'apparition d'effets toxiques associés. Alors que ces outils ont montré leur pertinence chez les organismes vertébrés *i.e*. poisson, ils se révèlent peu sensibles et discriminants chez les invertébrés. En effet, ces derniers n'ont pu bénéficier de développement spécifique en raison du manque de connaissance fondamentale de leur régulation au niveau moléculaire, notamment au niveau de la reproduction. Les molécules impliquées dans la fonction reproductrice se sont fortement diversifiées au cours de l'évolution, conduisant à des systèmes endocriniens très divers et différents de ceux connus ou caractérisés chez les vertébrés. Or les perturbateurs endocriniens font partie des substances à risque, menaçant le maintien de l'espèce dans son environnement, et rendent nécessaire le développement de biomarqueurs capables d'évaluer l'exposition à ces substances. Dans ce contexte, ces travaux de thèse portent sur la découverte de protéines impliquées dans la reproduction du crustacé d'eau douce *Gammarus fossarum* afin de mieux comprendre la régulation de cette fonction ainsi que les mécanismes de la perturbation endocrinienne chez les crustacés. Ces travaux s'appuient sur les avancées technologiques les plus récentes dans le domaine du séquençage de l'information génétique et protéique.

Le premier volet de ma thèse s'est intéressé à une problématique méthodologique visant à proposer une approche novatrice pour la découverte de protéines chez une espèce « non-modèle » par une approche protéogénomique, en combinant le séquençage en parallèle du transcriptome et du protéome. Cette démarche vise à générer une BD protéique spécifique à partir du séquençage des séquences d'ARNm, pour l'interprétation des spectres MS/MS obtenus par protéomique shotgun. Afin d'optimiser la découverte de protéines d'intérêt dans la reproduction, quatre tissus impliqués dans le processus reproductif, les tissus reproducteurs mâles et femelles, le céphalon et les *cæca* hépatopancréatiques ont été caractérisés par protéomique et transcriptomique.

Le séquençage du transcriptome a permis d'identifier 218 574 contigs et les différentes expérimentations protéomiques menées au cours de ces travaux ont permis de confirmer l'existence d'un total de 2257 protéines dont 278 orphelines. Les approches protéogénomiques permettent de lever certains verrous pour l'étude d'espèces « non-modèle » au niveau protéique et par conséquent, sont probablement amenées à se généraliser au cours de ces prochaines années.

La deuxième étape visait à identifier de nouvelles protéines d'intérêt dans la reproduction. En alliant l'analyse bio-informatique des séquences des protéines ainsi que leur localisation tissulaire, différentes protéines impliquées dans la mue, dans le métabolisme hormonal des ecdystéroïdes et terpenoïdes et dans la réaction de l'acrosome ont été identifiées. L'étude de la dynamique du protéome des ovaires au cours de l'ovogénèse a permis d'isoler des protéines potentiellement impliquées dans la coordination de l'ovogénèse ou dans la biogénèse des ecdystéroïdes. Aussi, en y combinant l'étude de la dynamique du protéome des embryons au cours de l'embryogénèse, huit protéines occupant la fonction de Vtg ont été isolées, révélant une diversité insoupçonnée de ces protéines au sein d'un même organisme. La physiologie reproductive masculine par le suivi de la dynamique du protéome des testicules au cours de la spermatogénèse se révèle plus complexe à étudier, mettant en évidence une importante portion de protéines orphelines.

Enfin, la dernière étape de ces travaux porte sur l'étude de l'infertilité masculine, en réponse à l'exposition à des perturbateurs endocriniens. Cette étude couple l'étude du protéome des gonades et de la physiologie reproductive par le biais de marqueurs de qualité spermatique. La numération spermatique se montre être un paramètre sensible pour l'évaluation de la reprotoxicité masculine et pertinent pour une implantation au sein de la batterie de biotests préexistants chez *G. fossarum*. L'étude du protéome a révélé une perturbation de l'homéostasie calcique et ionique ainsi qu'une immunosuppression. En revanche, contrairement aux vertébrés, aucun phénomène de féminisation n'a été détecté *i.e.* induction Vtg, mais plutôt une modulation des protéines spécifiques du mâle. Enfin, une importante proportion de protéines orphelines s'est montrée sensible au stress toxique, justifiant la mise en place d'approche de découverte spécifique au modèle d'étude.

La protéomique shotgun permet une étude exhaustive du protéome et se révèle être un outil puissant pour la découverte de nouveaux biomarqueurs. Dans le cas de l'identification de la Vtg chez le gammare, l'utilisation d'une approche ciblée par Jubeaux (2012) validait son rôle en tant que protéine de réserve du développement embryonnaire. Toutefois, notre caractérisation

exhaustive a permis d'isoler des protéines ayant un rôle plus prépondérant dans ce processus. Cependant, le spectromètre de masse employé au cours de ces travaux permet une meilleure caractérisation des peptides abondants que des peptides de faible abondance. Aussi, certaines protéines de faible abondance, notamment les protéines orphelines, ont pu être détectées sur l'ensemble de l'expérimentation mais leur modulation n'a pas pu être suivie. Il est à noter que la récente mise sur le marché d'une nouvelle génération de spectromètre de masse de type Q-exactive HF permet désormais de caractériser encore plus exhaustivement les échantillons (Arrey *et al.* 2014). Il est en tout cas envisagé d'adopter une démarche de quantification ciblée des candidats biomarqueurs découverts au cours de ces travaux de thèse afin de valider leur pertinence.

Les outils de protéomique massive ne sont également pas adaptés dans une optique de suivi en routine de biomarqueurs, la validation de ces derniers passe par l'utilisation de techniques ciblées de type SRM/MRN avec des instruments plus abordables (Gallien *et al.* 2011, Muraoka *et al.* 2012). La Figure i présente la démarche de développement de nouveaux biomarqueurs chez *G. fossarum*, allant d'une démarche de découverte jusqu'à une application en biosurveillance.

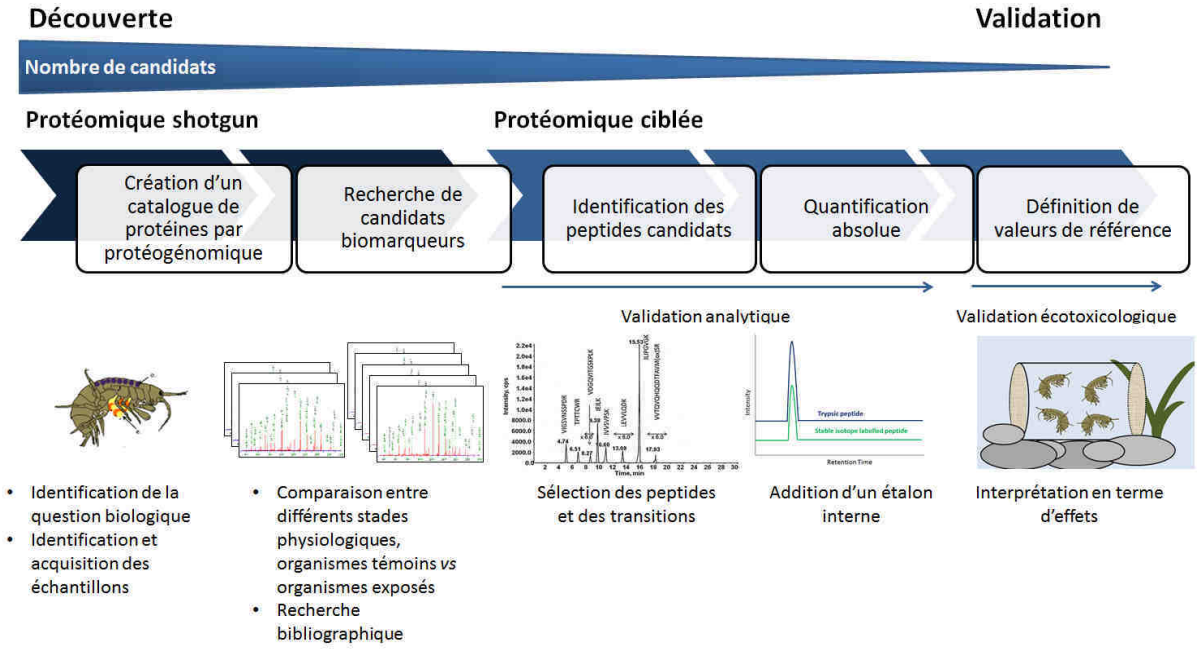


Figure i : Stratégie pour le développement de nouveaux biomarqueurs chez *Gammarus fossarum* par des approches protéomiques

En amont d'une utilisation en biosurveillance, il est nécessaire de caractériser les facteurs biotiques influant sur la réponse des paramètres biologiques mesurés afin de définir des organismes calibrés. En raison de son synchronisme avec l'ovogénèse et l'embryogénèse, le cycle de mue de la femelle a été finement caractérisé. Cette démarche a permis de calibrer les organismes afin de constituer un groupe d'organismes homogènes, limitant la variabilité inter-individuelle. Chez les mâles, le cycle de mue n'a pas été décrit, ce dernier étant supposé avoir peu d'influence sur la reproduction. Toutefois, au cours de ces travaux de thèse, certaines observations contredisent cette hypothèse telle que l'importante variabilité interindividuelle dans la production spermatique en réponse au Met, analogue de la 20-HE. Dans un objectif de développement d'outils pour évaluer l'exposition aux PE, la caractérisation du cycle de mue du mâle doit être réalisée, permettant par la suite d'étudier l'effet du stade de mue sur les variables biologiques suivies (biomarqueurs et marqueurs de qualité spermatique). Aussi, bien que les biomarqueurs protéiques ont été suivis au cours de différents processus physiologiques par protéomique shotgun, il est nécessaire de les suivre à nouveau par MRM afin [1] d'observer si les protéines présentent le même profil de variation temporelle selon les deux méthodes, validant ainsi leur rôle dans ces processus et [2] de décrire l'évolution des protéines de faible abondance dont la modulation n'a pu être suivie.

Le protocole d'exposition pour le suivi de l'infertilité chez le mâle a été développé afin de maximiser l'apparition d'effets en exposant les organismes sur deux cycles de spermatogénèse. Cette approche semble difficilement applicable en contexte naturel. L'évaluation *in situ* de la reprotoxicité chez la femelle bénéficie d'un protocole robuste, appliqué de manière routinière, se basant sur l'encagement pendant 21 jours, de précopulât dont la femelle est en stade de pré-mue (Lacaze *et al.* 2011, Coulaud *et al.* 2014). Son application pour l'étude des phénomènes reprotoxiques chez le mâle constitue une première piste à explorer. Une fois le protocole d'exposition validé, la définition des niveaux de base des différents paramètres biologiques pourra être réalisée par le suivi *in situ* de la variabilité spatiotemporelle.

Enfin, dans un contexte d'évaluation de la qualité environnementale, les biomarqueurs doivent avoir une visée diagnostique et/ou prédictive (*i.e*. apparition d'effets aux niveaux d'organisation biologique supérieurs) (Lagadic 2002). Pour le mâle, comme pour la femelle, les bioessais disponibles permettent de réaliser en laboratoire, une caractérisation fine de la relation dose-réponse pour une bonne interprétation de la réponse du biomarqueur.

BIBLIOGRAPHIE

Adams, S. M., A. M. Brown and R. W. Goede (1993). "A Quantitative Health Assessment Index for Rapid Evaluation of Fish Condition in the Field." Transactions of the American Fisheries Society **122**(1): 63-73.

Aebersold, R., A. L. Burlingame and R. A. Bradshaw (2013). "Western blots versus selected reaction monitoring assays: Time to turn the tables?" Molecular and Cellular Proteomics **12**(9): 2381-2382.

Agilent Technologies (2009). Agilent 3100 OFFGEL Fractionator Kit : Quick Start Guide. USA, Agilent Technologies, Inc.

Aitken, R. J. and M. A. Baker (2006). "Oxidative stress, sperm survival and fertility control." Mol Cell Endocrinol **250**(1-2): 66-69.

Allain, F. (2014). Nouvelle méthode d'interprétation de données de spectrométrie de masse en tandem pour l'identification de microorganismes dans un échantillon complexe. PhD, L'Institut National des Sciences Appliquées de Rouen.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**(3): 403-410.

Amelina, H., I. Apraiz, W. Sun and S. Cristobal (2007). "Proteomics-based method for the assessment of marine pollution using liquid chromatography coupled with two-dimensional electrophoresis." Journal of Proteome Research **6**(6): 2094-2104.

Amiard, J. C., C. Amiard-Triquet, S. Barka, J. Pellerin and P. S. Rainbow (2006). "Metallothioneins in aquatic invertebrates: Their role in metal detoxification and their use as biomarkers." Aquatic Toxicology **76**(2): 160-202.

Ansong, C., S. O. Purvine, J. N. Adkins, M. S. Lipton and R. D. Smith (2008). "Proteogenomics: needs and roles to be filled by proteomics in genome annotation." Brief Funct Genomic Proteomic **7**(1): 50-62.

Apraiz, I., J. Mi and S. Cristobal (2006). "Identification of proteomic signatures of exposure to marine pollutants in mussels (Mytilus edulis)." Mol Cell Proteomics **5**(7): 1274-1285.

Armengaud, J. (2009). "A perfect genome annotation is within reach with the proteomics and genomics alliance." Current Opinion in Microbiology **12**(3): 292-300.

Armengaud, J. (2013). "Microbiology and proteomics, getting the best of both worlds!" Environmental Microbiology **15**(1): 12-23.

Armengaud, J. and C. Bland (2014). Proteogenomics: A New Integrative Approach for a Better Description of Protein Diversity Found in Soil Microflora. Omics in Soil Science. P. Nannipieri, G. Pietramellara and G. Renella, Caister Acadamic Press.

Arrey, T., E. Damoc, K. Scheffler, Y. Xuan, B. Maciej, T. Moehring and M. Kellmann. (2014). "Improvements in Shotgun Proteomics Using a Benchtop Quadrupole Ultra-High-Field Orbitrap." from http://www.thermoscientific.com/en/product/q-exactive-hf-hybrid-quadrupole-orbitrap.html.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: Tool for the unification of biology." Nature Genetics **25**(1): 25-29.

Attrill, M. J. and M. H. Depledge (1997). "Community and population indicators of ecosystem health: Targeting links between levels of biological organisation." Aquatic Toxicology **38**(1-3): 183-197.

Barnard, J. L. and C. M. Barnard (1983). Freshwater Amphipoda of the world- I. Evolutionary patterns.

Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall and E. L. L. Sonnhammer (2002). "The pfam protein families database." Nucleic Acids Research **30**(1): 276-280.

Bedulina, D. S., M. A. Timofeyev, M. Zimmer, E. Zwirnmann, R. Menzel and C. E. Steinberg (2010). "Different natural organic matter isolates cause similar stress response patterns in the freshwater amphipod, Gammarus pulex." Environ Sci Pollut Res Int **17**(2): 261-269.

Bedulina, D. S., M. Zimmer and M. A. Timofeyev (2010). "Sub-littoral and supra-littoral amphipods respond differently to acute thermal stress." Comp Biochem Physiol B Biochem Mol Biol **155**(4): 413-418.
Besse, J. P., M. Coquery, C. Lopes, A. Chaumot, H. Budzinski, P. Labadie and O. Geffard (2013). "Caged Gammarus fossarum (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: towards the determination of threshold values." Water Res **47**(2): 650-660.
Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." Biomed Environ Mass Spectrom **16**(1-12): 99-111.
Blanchet-Tournier, M. F. (1980). Mue et vitellogenèse chez le crustacé amphipode Orchestia gammarellus (Pallas) : contrôle endocrine et interactions. Docteur, Université Pierre et Marie Curie.
Bland, C., L. Bellanger and J. Armengaud (2014). "Magnetic immunoaffinity enrichment for selective capture and MS/MS analysis of N-terminal-TMPP-labeled peptides." Journal of Proteome Research **13**(2): 668-680.
Bodzon-Kulakowska, A., A. Bierczynska-Krzysik, T. Dylag, A. Drabik, P. Suder, M. Noga, J. Jarzebinska and J. Silberring (2007). "Methods for samples preparation in proteomic research." Journal of Chromatography B **849**(1–2): 1-31.
Bollache, L. and F. Cézilly (2004). "Sexual selection on male body size and assortative pairing in Gammarus pulex (Crustacea: Amphipoda): field surveys and laboratory experiments." Journal of Zoology **264**(2): 135-141.
Borowsky, R. and M. M. Guarna (1989). "Excess Amylase in Gammarus-Palustris (Crustacea, Amphipoda) - Its Release into and Possible Roles in the Environment." Marine Biology **101**(4): 529-534.
Borst, D. W., H. Laufer, M. Landau, E. S. Chang, W. A. Hertz, F. C. Baker and D. A. Schooley (1987). "Methyl farnesoate and its role in crustacean reproduction and development." Insect Biochemistry **17**(7): 1123-1127.
Bouetard, A., C. Noirot, A. L. Besnard, O. Bouchez, D. Choisne, E. Robe, C. Klopp, L. Lagadic and M. A. Coutellec (2012). "Pyrosequencing-based transcriptomic resources in the pond snail Lymnaea stagnalis, with a focus on genes involved in molecular response to diquat-induced stress." Ecotoxicology **21**(8): 2222-2234.
Breitholtz, M. and B. E. Bengtsson (2001). "Oestrogens have no hormonal effect on the development and reproduction of the harpacticoid copepod Nitocra spinipes." Marine Pollution Bulletin **42**(10): 879-886.
Brooks, S. J. and C. Lloyd Mills (2006). "Gill Na(+), K(+)-ATPase in a series of hyper-regulating gammarid amphipods: enzyme characterisation and the effects of salinity acclimation." Comp Biochem Physiol A Mol Integr Physiol **144**(1): 24-32.
Brousseau, P., S. Pillet, H. Frouin, M. Auffret, F. Gagné and M. Fournier (2012). Linking immunotoxicity and ecotoxicological effects at higher biological levels. Ecological Biomarkers - Indicators of Ecotoxicological Effects. C. Amiard-Triquet, J. C. Amiard and P. S. Rainbow. Boca Raton, CRC Press**:** 131-154.
Cakouros, D., T. J. Daish and S. Kumar (2004). "Ecdysone receptor directly binds the promoter of the Drosophila caspase dronc, regulating its expression in specific tissues." The Journal of Cell Biology **165**(5): 631-640.
Cañas, B., C. Piñeiro, E. Calvo, D. López-Ferrer and J. M. Gallardo (2007). "Trends in sample preparation for classical and second generation proteomics." Journal of Chromatography A **1153**(1–2): 235-258.
Carlson, G. R., T. S. Dhadialla, R. Hunter, R. K. Jansson, C. S. Jany, Z. Lidert and R. A. Slawecki (2001). "The chemical and biological properties of methoxyfenozide, a new insecticidal ecdysteroid agonist." Pest Manag Sci **57**(2): 115-119.
Carvalho, P. C., J. S. Fischer, E. I. Chen, J. R. Yates, 3rd and V. C. Barbosa (2008). "PatternLab for proteomics: a tool for differential shotgun proteomics." BMC Bioinformatics **9**: 316.
Castellana, N. E., S. H. Payne, Z. Shen, M. Stanke, V. Bafna and S. P. Briggs (2008). "Discovery and revision of Arabidopsis genes by proteogenomics." Proceedings of the National Academy of Sciences **105**(52): 21034-21038.
Chang, E. S. and J. D. O'Connor (1978). "In vitro secretion and hydroxylation of α-ecdysone as a function of the crustacean molt cycle." General and Comparative Endocrinology **36**(1): 151-160.
Charniaux-Cotton, H. (1954). "Discovery in, an amphipod crustacean (Orchestia gammarella) of an endocrine gland responsible for the differentiation of primary and secondary male sex characteristics." Découverte chez un crustacé amphipode (Orchestia gammarella) d'une glande endocrine responsable de la différenciation de caractères sexuels primaires et secondaires mâles. **239**(13): 780-782.

Charniaux-Cotton, H. (1965). Hormonal control of sex differentiation in invertebrates. Organogenesis. R. L. Dehaan and H. Ursprung. New York, Holt, Rinehart, and Winston**:** 701-740.
Charniaux-Cotton, H. (1973). "Description et contrôle de l'ovogenèse chez les crustacés supérieurs." Ann. Biol. anim. Bioch. Biophys. **13**(Hors-série): 21-37.
Charniaux-Cotton, H. and G. Payen (1988). Crustacean reproduction. Endocrinology of Selected Invertebrate Types. H. Laufer and R. Downer. New York, Alan R. Liss,**:** 279–303.
Charron, L., O. Geffard, A. Chaumot, R. Coulaud, A. Jaffal, V. Gaillet, O. Dedourge-Geffard and A. Geffard (2014). "Influence of molting and starvation on digestive enzyme activities and energy storage in Gammarus fossarum." PLoS ONE **9**(4).
Charron, L., O. Geffard, A. Chaumot, R. Coulaud, H. Queau, A. Geffard and O. Dedourge-Geffard (2013). "Effect of water quality and confounding factors on digestive enzyme activities in Gammarus fossarum." Environ Sci Pollut Res Int **20**(12): 9044-9056.
Chenau, J., S. Michelland, J. Sidibe and M. Seve (2008). "Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling." Proteome Science **6**(1): 9.
Chevreux, E. and L. Fage, Eds. (1970). Faune de France, Tome 9 : Amphipodes. Nendeln (Lichtenstein).
Chocu, S., B. Evrard, R. Lavigne, A. D. Rolland, F. Aubry, B. Jegou, F. Chalmel and C. Pineau (2014). "Forty-Four Novel Protein-Coding Loci Discovered Using a Proteomics Informed by Transcriptomics (PIT) Approach in Rat Male Germ Cells." Biol Reprod.
Chora, S., M. Starita-Geribaldi, J. M. Guigonis, M. Samson, M. Romeo and M. J. Bebianno (2009). "Effect of cadmium in the clam Ruditapes decussatus assessed by proteomic analysis." Aquat Toxicol **94**(4): 300-308.
Christie, A. E. (2011). "Crustacean neuroendocrine systems and their signaling agents." Cell and Tissue Research **345**(1): 41-67.
Clare, A. S., D. Rittschof and J. D. Costlow (1992). "Effects of the nonsteroidal ecdysone mimic RH 5849 on larval crustaceans." Journal of Experimental Zoology **262**(4): 436-440.
Clark, N. L., J. E. Aagaard and W. J. Swanson (2006). "Evolution of reproductive proteins from animals and plants." Reproduction **131**(1): 11-22.
Clements, W. H. (2000). "Integrating effects of contaminants across levels of biological organization: An overview." Journal of Aquatic Ecosystem Stress and Recovery **7**(2): 113-116.
Coen, W. M. D., C. R. Janssen, W. M. De Coen and C. R. Janssen (1997). "The use of biomarkers in Daphnia magna toxicity testing. IV. Cellular Energy Allocation: a new methodology to assess the energy budget of toxicant-stressed Daphnia populations." Journal of Aquatic Ecosystem Stress and Recovery **6**(1): 43-55.
Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Caceres, L. Carmel, C. Casola, J. H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Frohlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V. Koonin, E. M. Zdobnov, I. V. Grigoriev, M. Lynch and J. L. Boore (2011). "The ecoresponsive genome of Daphnia pulex." Science **331**(6017): 555-561.
Correia, A. D., F. O. Costa, T. Neuparth, M. E. Diniz and M. H. Costa (2001). "Sub-lethal Effects of Copper-Spiked Sediments on the Marine Amphipod *Gammarus locusta* : Evidence of Hormesis ? ." Ecotoxicology and Environmental Restoration(2): 32-38.
Correia, A. D., M. H. Costa, O. J. Luis and D. R. Livingstone (2003). "Age-related changes in antioxidant enzyme activities, fatty acid composition and lipid peroxidation in whole body Gammarus locusta (Crustacea : Amphipoda)." Journal of Experimental Marine Biology and Ecology **289**(1): 83-101.
Cottrell, J. S. (2011). "Protein identification using MS/MS data." J Proteomics **74**(10): 1842-1851.
Coulaud, R., O. Geffard, A. Coquillat, H. Queau, S. Charles and A. Chaumot (2014). "Ecological modeling for the extrapolation of ecotoxicological effects measured during in situ assays in Gammarus." Environ Sci Technol **48**(11): 6428-6436.

Coulaud, R., O. Geffard, B. Xuereb, E. Lacaze, H. Queau, J. Garric, S. Charles and A. Chaumot (2011). "In situ feeding assay with Gammarus fossarum (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring." Water Res **45**(19): 6417-6429.
D'Amours, O., G. Frenette, L. J. Bordeleau, N. Allard, P. Leclerc, P. Blondin and R. Sullivan (2012). "Epididymosomes transfer epididymal sperm binding protein 1 (ELSPBP1) to dead spermatozoa during epididymal transit in bovine." Biology of Reproduction **87**(4).
Dangles, O. and F. Guérold (2001). "Linking shredders and leaf litter processing: Insights from an acidic stream study." Int Rev Hydrobiol **86**(4-5): 395-406.
de Hoffmann, E. and V. Stroobant (2005). Spectrométrie de masse - Cours et exercices corrigés. Paris, Dunod.
De Loof, A., B. Boerjan, U. R. Ernst and L. Schoofs (2013). "The mode of action of juvenile hormone and ecdysone: towards an epi-endocrinological paradigm?" Gen Comp Endocrinol **188**(0): 35-45.
Dedourge-Geffard, O., L. Charron, C. Hofbauer, V. Gaillet, F. Palais, E. Lacaze, A. Geffard and O. Geffard (2013). "Temporal patterns of digestive enzyme activities and feeding rate in gammarids (Gammarus fossarum) exposed to inland polluted waters." Ecotoxicol Environ Saf **97**: 139-146.
Dedourge-Geffard, O., F. Palais, S. Biagianti-Risbourg, O. Geffard and A. Geffard (2009). "Effects of metals on feeding rate and digestive enzymes in Gammarus fossarum: an in situ experiment." Chemosphere **77**(11): 1569-1576.
Dedourge-Geffard, O., F. Palais, A. Geffard and C. Amiard-Triquet (2012). Origin of Energy Metabolism Impairments. Ecological Biomarkers - Indicators of Ecotoxicological Effects. C. Amiard-Triquet, J. C. Amiard and P. S. Rainbow. Boca Raton, CRC Press**:** 279-306.
Depledge, M. H. and M. C. Fossi (1994). "The role of biomarkers in environmental assessment (2). Invertebrates." Ecotoxicology **3**(3): 161-172.
Depledge, M. H. and T. S. Galloway (2005). "Healthy animals, healthy ecosystems." Frontiers in Ecology and the Environment **3**(5): 251-258.
Dhainaut, A., J. Bonaly, J. P. Barque, C. Minier and T. Caquet (1997). Protéines de choc thermique et résistance multixénobiotique. Biomarqueurs en Ecotoxicologie - Aspects fondamentaux. L. Lagadic, T. Caquet, J. C. Amiard and F. Ramade. Paris, Masson**:** 67-87.
du Plessis, S. S., A. H. Kashou, D. J. Benjamin, S. P. Yadav and A. Agarwal (2011). "Proteomics: A subcellular look at spermatozoa." Reprod Biol Endocrinol. **9**.
Duchet, C., M. Mitie Inafuku, T. Caquet, M. Larroque, E. Franquet, C. Lagneau and L. Lagadic (2011). "Chitobiase activity as an indicator of altered survival, growth and reproduction in Daphnia pulex and Daphnia magna (Crustacea: Cladocera) exposed to spinosad and diflubenzuron." Ecotoxicology and Environmental Safety **74**(4): 800-810.
Dupierris, V., C. Masselon, M. Court, S. Kieffer-Jaquinod and C. Bruley (2009). "A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa." Bioinformatics **25**(15): 1980-1981.
Durica, D. S., A. C.-K. Chung and P. M. Hopkins (1999). "Characterization of EcR and RXR Gene Homologs and Receptor Expression During the Molt Cycle in the Crab, Uca pugilator." American Zoologist **39**(4): 758-773.
Durica, D. S., X. Wu, G. Anilkumar, P. M. Hopkins and A. C. Chung (2002). "Characterization of crab EcR and RXR homologs and expression during limb regeneration and oocyte maturation." Mol Cell Endocrinol **189**(1-2): 59-76.
Durighello, E., J. A. Christie-Oleza and J. Armengaud (2014). "Assessing the exoproteome of marine bacteria, lesson from a RTX-toxin abundantly secreted by Phaeobacter Strain DSM 17395." PLoS ONE **9**(2).
Ekblom, R. and J. Galindo (2011). "Applications of next generation sequencing in molecular ecology of non-model organisms." Heredity (Edinb) **107**(1): 1-15.
El Haj, A. J., P. Harrison and E. S. Chang (1994). "Localization of ecdysteroid receptor immunoreactivity in eyestalk and muscle tissue of the American lobster, Homarus americanus." Journal of Experimental Zoology **270**(4): 343-349.

Eng, J. K., A. L. McCormack and J. R. Yates Iii (1994). "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database." Journal of the American Society for Mass Spectrometry **5**(11): 976-989.
Ernoult, E., E. Gamelin and C. Guette (2008). "Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation." Proteome Science **6**(1): 27.
European Commission (1999). Community Strategy for Endocrine Disruptors - a range of substances suspected of interfering with the hormone systems of humans and wildlife. COMMUNICATION FROM THE COMMISSION TO THE COUNCIL AND THE EUROPEAN PARLIAMENT. Brussels.
Evans, T. J. and V. K. Ganjam (2011). Reproductive Anatomy and Physiology. Reproductive and Developmental Toxicology**:** 7-32.
Falkenstein, E., H. C. Tillmann, M. Christ, M. Feuring and M. Wehling (2000). "Multiple actions of steroid hormones--a focus on rapid, nongenomic effects." Pharmacol Rev **52**(4): 513-556.
Felten, V. (2003). Effets de l'acidification des ruisseaux vosgiens sur la biologie, l'écologie et l'écophysiologie de *Gammarus fossarum* Koch, 1835 (*Crustacea Amphipoda*) : Approche intégrée à différents niveaux d'organisation. P.h.D., Université de Metz.
Felten, V., G. Charmantier, M. Charmantier-Daures, F. Aujoulat, J. Garric and O. Geffard (2008). "Physiological and behavioural responses of Gammarus pulex exposed to acid stress." Comp Biochem Physiol C Toxicol Pharmacol **147**(2): 189-197.
Felten, V., G. Charmantier, R. Mons, A. Geffard, P. Rousselle, M. Coquery, J. Garric and O. Geffard (2008). "Physiological and behavioural responses of Gammarus pulex (Crustacea: Amphipoda) exposed to cadmium." Aquat Toxicol **86**(3): 413-425.
Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse (1989). "Electrospray ionization for mass spectrometry of large biomolecules." Science **246**(4926): 64-71.
Fields, S. (2001). "Proteomics. Proteomics in genomeland." Science **291**(5507): 1221-1224.
Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick and et al. (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." Science **269**(5223): 496-512.
Ford, A. T. (2008). "Can you feminise a crustacean?" Aquatic Toxicology **88**(4): 316-321.
Ford, A. T. (2012). "Intersexuality in Crustacea: an environmental issue?" Aquat Toxicol **108**: 125-129.
Ford, A. T., T. F. Fernandes, S. A. Rider, P. A. Read, C. D. Robinson and I. M. Davies (2003). "Measuring sublethal impacts of pollution on reproductive output of marine Crustacea." Marine Ecology Progress Series **265**: 303-309.
Frank, S. N., S. Godehardt, M. Nachev, A. Trubiroha, W. Kloas and B. Sures (2013). "Influence of the cestode Ligula intestinalis and the acanthocephalan Polymorphus minutus on levels of heat shock proteins (HSP70) and metallothioneins in their fish and crustacean intermediate hosts." Environmental Pollution **180**: 173-179.
Gagné, F. and C. Blaise (2004). "Shell protein characteristics and vitellogenin-like proteins in brine shrimp Artemia franciscana exposed to municipal effluent and 20-hydroxyecdysone." Comparative Biochemistry and Physiology - C Toxicology and Pharmacology **138**(4): 515-522.
Gagné, F., C. Blaise and J. Pellerin (2005). "Altered exoskeleton composition and vitellogenesis in the crustacean Gammarus sp. collected at polluted sites in the Saguenay Fjord, Quebec, Canada." Environmental Research **98**(1): 89-99.
Gallien, S., E. Duriez and B. Domon (2011). "Selected reaction monitoring applied to proteomics." J Mass Spectrom **46**(3): 298-312.
Galloway, T. S., R. J. Brown, M. A. Browne, A. Dissanayake, D. Lowe, M. B. Jones and M. H. Depledge (2004). "A multibiomarker approach to environmental assessment." Environ Sci Technol **38**(6): 1723-1731.
Gaspari, M. and G. Cuda (2011). "Nano LC-MS/MS: a robust setup for proteomic analysis." Methods Mol Biol **790**: 115-126.
Geffard, A., H. Queau, O. Dedourge, S. Biagianti-Risboug and O. Geffard (2007). "Influence of biotic and abiotic factors on metallothionein level in Gammarus pulex." Comp Biochem Physiol C Toxicol Pharmacol **145**(4): 632-640.

Geffard, O., B. Xuereb, A. Chaumot, A. Geffard, S. Biagianti, C. Noel, K. Abbaci, J. Garric, G. Charmantier and M. Charmantier-Daures (2010). "Ovarian cycle and embryonic development in Gammarus fossarum: application for reproductive toxicity assessment." Environ Toxicol Chem **29**(10): 2249-2259.
Ghekiere, A., M. Fenske, T. Verslycke, C. Tyler and C. Janssen (2005). "Development of a quantitative enzyme-linked immunosorbent assay for vitellin in the mysid Neomysis integer (Crustacea: Mysidacea)." Comp Biochem Physiol A Mol Integr Physiol **142**(1): 43-49.
Ghekiere, A., T. Verslycke and C. Janssen (2006). "Effects of methoprene, nonylphenol, and estrone on the vitellogenesis of the mysid Neomysis integer." Gen Comp Endocrinol **147**(2): 190-195.
Gismondi, E., T. Rigaud, J. N. Beisel and C. Cossu-Leguille (2012). "Microsporidia parasites disrupt the responses to cadmium exposure in a gammarid." Environ Pollut **160**(1): 17-23.
Gore, A. (2009). La naissance du mouvement écologiste. Printemps silencieux. R. Carson. France, Wildproject.
Gouy, M., S. Guindon and O. Gascuel (2010). "Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building." Molecular Biology and Evolution **27**(2): 221-224.
Haine, E. R., E. Brondani, K. D. Hume, M.-J. Perrot-Minnot, M. Gaillard and T. Rigaud (2004). "Coexistence of three microsporidia parasites in populations of the freshwater amphipod Gammarus roeseli: evidence for vertical transmission and positive effect on reproduction." International Journal for Parasitology **34**(10): 1137-1146.
Halanych, K. M. (2004). "The new view of animal phylogeny." Annual Review of Ecology Evolution and Systematics **35**: 229-256.
Hannas, B. R., Y. H. Wang, S. Thomson, G. Kwon, H. Li and G. A. LeBlanc (2011). "Regulation and dysregulation of vitellogenin mRNA accumulation in daphnids (Daphnia magna)." Aquatic Toxicology **101**(2): 351-357.
Hanson, N., L. Forlin and A. Larsson (2010). "Spatial and annual variation to define the normal range of biological endpoints: an example with biomarkers in perch." Environ Toxicol Chem **29**(11): 2616-2624.
Hartmann, E. M. and J. Armengaud (2014). "Shotgun proteomics suggests involvement of additional enzymes in dioxin degradation by Sphingomonas wittichiiRW1." Environ Microbiol **16**(1): 162-176.
Hawkins, T., M. Chitale and D. Kihara (2008). "New paradigm in protein function prediction for large scale omics analysis." Mol Biosyst **4**(3): 223-231.
Hayward, A., T. Takahashi, W. G. Bendena, S. S. Tobe and J. H. Hui (2010). "Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans." FEBS Lett **584**(6): 1273-1278.
He, R., M. J. Kim, W. Nelson, T. S. Balbuena, R. Kim, R. Kramer, J. A. Crow, G. D. May, J. J. Thelen, C. A. Soderlund and D. R. Gang (2012). "Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, Phragmites australis (Poaceae), reveals genes involved in invasiveness and rhizome specificity." American Journal of Botany **99**(2): 232-247.
Henzel, W. J., T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley and C. Watanabe (1993). "Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases." Proc Natl Acad Sci U S A **90**(11): 5011-5015.
Holt, M. S. (2000). "Sources of chemical contaminants and routes into the freshwater environment." Food Chem Toxicol **38**(1 Suppl): S21-27.
Hook, S. E., N. A. Twine, S. L. Simpson, D. A. Spadaro, P. Moncuquet and M. R. Wilkins (2014). "454 pyrosequencing-based analysis of gene expression profiles in the amphipod Melita plumulosa: Transcriptome assembly and toxicant induced changes." Aquatic Toxicology **153**: 73-88.
Howe, K., M. D. Clark, C. F. Torroja, J. Torrance, C. Berthelot, M. Muffato, J. E. Collins, S. Humphray, K. McLaren, L. Matthews, S. McLaren, I. Sealy, M. Caccamo, C. Churcher, C. Scott, J. C. Barrett, R. Koch, G. J. Rauch, S. White, W. Chow, B. Kilian, L. T. Quintais, J. A. Guerra-Assunção, Y. Zhou, Y. Gu, J. Yen, J. H. Vogel, T. Eyre, S. Redmond, R. Banerjee, J. Chi, B. Fu, E. Langley, S. F. Maguire, G. K. Laird, D. Lloyd, E. Kenyon, S. Donaldson, H. Sehra, J. Almeida-King, J. Loveland, S. Trevanion, M. Jones, M. Quail, D. Willey, A. Hunt, J. Burton, S. Sims, K. McLay, B. Plumb, J. Davis, C. Clee, K. Oliver, R. Clark, C. Riddle, D. Eliott, G. Threadgold, G. Harden, D. Ware, B. Mortimer, G. Kerry, P. Heath, B. Phillimore, A. Tracey, N. Corby, M. Dunn, C. Johnson, J.

Wood, S. Clark, S. Pelan, G. Griffiths, M. Smith, R. Glithero, P. Howden, N. Barker, C. Stevens, J. Harley, K. Holt, G. Panagiotidis, J. Lovell, H. Beasley, C. Henderson, D. Gordon, K. Auger, D. Wright, J. Collins, C. Raisen, L. Dyer, K. Leung, L. Robertson, K. Ambridge, D. Leongamornlert, S. McGuire, R. Gilderthorp, C. Griffiths, D. Manthravadi, S. Nichol, G. Barker, S. Whitehead, M. Kay, J. Brown, C. Murnane, E. Gray, M. Humphries, N. Sycamore, D. Barker, D. Saunders, J. Wallis, A. Babbage, S. Hammond, M. Mashreghi-Mohammadi, L. Barr, S. Martin, P. Wray, A. Ellington, N. Matthews, M. Ellwood, R. Woodmansey, G. Clark, J. Cooper, A. Tromans, D. Grafham, C. Skuce, R. Pandian, R. Andrews, E. Harrison, A. Kimberley, J. Garnett, N. Fosker, R. Hall, P. Garner, D. Kelly, C. Bird, S. Palmer, I. Gehring, A. Berger, C. M. Dooley, Z. Ersan-Ürün, C. Eser, H. Geiger, M. Geisler, L. Karotki, A. Kirn, J. Konantz, M. Konantz, M. Oberländer, S. Rudolph-Geiger, M. Teucke, K. Osoegawa, B. Zhu, A. Rapp, S. Widaa, C. Langford, F. Yang, N. P. Carter, J. Harrow, Z. Ning, J. Herrero, S. M. J. Searle, A. Enright, R. Geisler, R. H. A. Plasterk, C. Lee, M. Westerfield, P. J. De Jong, L. I. Zon, J. H. Postlethwait, C. Nüsslein-Volhard, T. J. P. Hubbard, H. R. Crollius, J. Rogers and D. L. Stemple (2013). "The zebrafish reference genome sequence and its relationship to the human genome." Nature **496**(7446): 498-503.

Huang, Q.-S., J.-H. Yan, J.-Y. Tang, Y.-M. Tao, X.-L. Xie, Y. Wang, X.-Q. Wei, Q.-H. Yan and Q.-X. Chen (2010). "Cloning and tissue expressions of seven chitinase family genes in Litopenaeus vannamei." Fish & Shellfish Immunology **29**(1): 75-81.

Hughes, P. J., H. McLellan, D. A. Lowes, S. Z. Kahn, J. G. Bilmen, S. C. Tovey, R. E. Godfrey, R. H. Michell, C. J. Kirk and F. Michelangeli (2000). "Estrogenic alkylphenols induce cell death by inhibiting testis endoplasmic reticulum Ca(2+) pumps." Biochem Biophys Res Commun **277**(3): 568-574.

Hui, J. H. L., A. Hayward, W. G. Bendena, T. Takahashi and S. S. Tobe (2010). "Evolution and functional divergence of enzymes involved in sesquiterpenoid hormone biosynthesis in crustaceans and insects." Peptides **31**(3): 451-455.

Hyne, R. V. (2011). "Review of the reproductive biology of amphipods and their endocrine regulation: identification of mechanistic pathways for reproductive toxicants." Environ Toxicol Chem **30**(12): 2647-2657.

Hyne, R. V., S. A. Gale and C. K. King (2005). "Laboratory culture and life-cycle experiments with the benthic amphipod Melita plumulosa (Zeidler)." Environ Toxicol Chem **24**(8): 2065-2073.

Issartel, J., V. Boulo, S. Wallon, O. Geffard and G. Charmantier (2010). "Cellular and molecular osmoregulatory responses to cadmium exposure in Gammarus fossarum (Crustacea, Amphipoda)." Chemosphere **81**(6): 701-710.

Jaffe, J. D., H. C. Berg and G. M. Church (2004). "Proteogenomic mapping as a complementary method to perform genome annotation." Proteomics **4**(1): 59-77.

James, M. O. and S. M. Boyle (1998). "Cytochromes P450 in crustacea." Comp Biochem Physiol C Pharmacol Toxicol Endocrinol **121**(1-3): 157-172.

Jeuniaux, C. (1963). Chitine et chitinolyse. Un chapitre de la biologie moléculaire. Thèse d'agrégation de l'Enseignement Supérieur.

Jimenez-Gonzalez, C., F. Michelangeli, C. V. Harper, C. L. Barratt and S. J. Publicover (2006). "Calcium signalling in human spermatozoa: a specialized 'toolkit' of channels, transporters and stores." Hum Reprod Update **12**(3): 253-267.

Jobling, S., M. Nolan, C. R. Tyler, G. Brighty and J. P. Sumpter (1998). "Widespread sexual disruption in wild fish." Environ Sci Technol **32**(17): 2498-2506.

Jubeaux, G., R. Simon, A. Salvador, C. Lopes, E. Lacaze, H. Queau, A. Chaumot and O. Geffard (2012). "Vitellogenin-like protein measurement in caged Gammarus fossarum males as a biomarker of endocrine disruptor exposure: inconclusive experience." Aquat Toxicol **122-123**: 9-18.

Jubeaux, G., R. Simon, A. Salvador, H. Queau, A. Chaumot and O. Geffard (2012). "Vitellogenin-like proteins in the freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization throughout reproductive process, potential for use as an indicator of oocyte quality and endocrine disruption biomarker in males." Aquat Toxicol **112-113**(0): 72-82.

Jungmann, D., V. Ladewig, K. U. Ludwichowski, P. Petzsch and R. Nagel (2004). "Intersexuality in Gammarus fossarum Koch - A common inducible phenomenon?" Arch Hydrobio **159**(4): 511-529.

Kanehisa, M. and S. Goto (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes." Nucleic Acids Research **28**(1): 27-30.

Karaman, G. S. (1977). "Freshwater Gammarus species from Europe, North Africa and adjacent regions of Asia (Crustacea-Amphipoda). Part I. Gammarus pulex-group and related species " Bijdragen Tot de Dierkunde **47**.

Karaouzas, I., E. Cotou, T. A. Albanis, A. Kamarianos, N. T. Skoulikidis and U. Giannakou (2011). "Bioassays and biochemical biomarkers for assessing olive mill and citrus processing wastewater toxicity." Environ Toxicol **26**(6): 669-676.

Karas, M. and F. Hillenkamp (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons [1]." Analytical Chemistry **60**(20): 2299-2301.

Kato, Y., K. Kobayashi, H. Watanabe and T. Iguchi (2011). "Environmental sex determination in the branchiopod crustacean Daphnia magna: Deep conservation of a Doublesex gene in the sex-determining pathway." Plos Genetics **7**(3).

Kato, Y., S. Tokishita, T. Ohta and H. Yamagata (2004). "A vitellogenin chain containing a superoxide dismutase-like domain is the major component of yolk proteins in cladoceran crustacean Daphnia magna." Gene **334**(0): 157-165.

Kenny, N. J., S. Quah, P. W. H. Holland, S. S. Tobe and J. H. L. Hui (2013). "How are comparative genomics and the study of microRNAs changing our views on arthropod endocrinology and adaptations to the environment?" General and Comparative Endocrinology **188**(0): 16-22.

Kim, M.-S., S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.-C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda and A. Pandey (2014). "A draft map of the human proteome." Nature **509**(7502): 575-581.

Kime, D. E., J. P. Nash and A. P. Scott (1999). "Vitellogenesis as a biomarker of reproductive disruption by xenobiotics." Aquaculture **177**(1-4): 345-352.

King-Jones, K. and C. Thummel (2005). "Nuclear receptors - A perspective from *Drosophila*." Nature Reviews **6**.

Kirk, C. J., L. Bottomley, N. Minican, H. Carpenter, S. Shaw, N. Kohli, M. Winter, E. W. Taylor, R. H. Waring, F. Michelangeli and R. M. Harris (2003). "Environmental endocrine disrupters dysregulate estrogen metabolism and Ca2+ homeostasis in fish and mammals via receptor-independent mechanisms." Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology **135**(1): 1-8.

Knotz, S., M. Boersma and R. Saborowski (2006). "Microassays for a set of enzymes in individual small marine copepods." Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology **145**(3): 406-411.

Koenig, S., P. Fernandez and M. Sole (2012). "Differences in cytochrome P450 enzyme activities between fish and crustacea: Relationship with the bioaccumulation patterns of polychlorobiphenyls (PCBs)." Aquatic Toxicology **108**: 11-17.

Köhler, H. R., W. Kloas, M. Schirling, I. Lutz, A. L. Reye, J. S. Langen, R. Triebskorn, R. Nagel and G. Schönfelder (2007). "Sex steroid receptor evolution and signalling in aquatic invertebrates." Ecotoxicology **16**(1): 131-143.

Kunz, P. Y., C. Kienle and A. Gerhardt (2010). *Gammarus* spp. in Aquatic Ecotoxicology and Water Quality Assessment: Toward Integrated Multilevel Tests. Reviews of Environmental Contamination and Toxicology. D. M. Whitacre, Springer New York. **205:** 1-76.

Lacaze, E. (2011). Un biomarqueur de génotoxicité chez *Gammarus fossarum* : développement, signification fonctionelle et application au milieu naturel. PhD, Université de Metz.

Lacaze, E., A. Devaux, G. Jubeaux, R. Mons, M. Gardette, S. Bony, J. Garric and O. Geffard (2011). "DNA damage in Gammarus fossarum sperm as a biomarker of genotoxic pressure: intrinsic variability and reference level." Sci Total Environ **409**(17): 3230-3236.
Lacaze, E., A. Devaux, R. Mons, S. Bony, J. Garric, A. Geffard and O. Geffard (2011). "DNA damage in caged Gammarus fossarum amphipods: a tool for freshwater genotoxicity assessment." Environ Pollut **159**(6): 1682-1691.
Lacaze, E., O. Geffard, S. Bony and A. Devaux (2010). "Genotoxicity assessment in the amphipod Gammarus fossarum by use of the alkaline Comet assay." Mutat Res-Gen Tox En **700**(1–2): 32-38.
Lacaze, E., O. Geffard, D. Goyet, S. Bony and A. Devaux (2011). "Linking genotoxic responses in Gammarus fossarum germ cells with reproduction impairment, using the Comet assay." Environmental Research **111**(5): 626-634.
Lafont, R. (2000). "The endocrinology of invertebrates." Ecotoxicology **9**(1-2): 41-57.
Lagadic, L. (2002). "Biomarkers: Useful tools for the monitoring of aquatic environments." Revue de Medecine Veterinaire **153**(8-9): 581-588.
Lagadic, L. and T. Caquet (1998). "Invertebrates in testing of environmental chemicals: Are they alternatives?" Environmental Health Perspectives **106**(SUPPL. 2): 593-611.
Lagadic, L., T. Caquet and J.-C. Amiard (2008). Biomarqeurs en écotoxicologie : principes et définitions. Les biomarqueurs dans l'évaluation de l'état écologique des milieux aquatiques. T. Doc. Paris, Lavoisier**:** 1-9.
Lagrue, C., R. Wattier, M. Galipaud, Z. Gauthey, J.-P. Rullmann, C. Dubreuil, T. Rigaud and L. Bollache (2014). "Confrontation of cryptic diversity and mate discrimination within Gammarus pulex and Gammarus fossarum species complexes." Freshwater Biology: n/a-n/a.
Lampert, W., K. P. Lampert and P. Larsson (2012). "Induction of male production in clones of Daphnia pulex by the juvenoid hormone methyl farnesoate under short photoperiod." Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **156**(2): 130-133.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. Fitzhugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. Levine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, M. L. Hong, J. Dubois, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. De La Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. R. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. A. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.

R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, A. Patrinos and M. J. Morgan (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lawrence, A. J. and C. Poulter (1998). "Development of a sub-lethal pollution bioassay using the estuarine amphipod *Gammarus duebeni*." Water Research **32**(3): 569-578.

Le Roux, M.-L. (1933). Recherches sur la sexualité des gammariens, Faculté des Sciences de l'Université de Paris.

LeBlanc, G. A. (2007). "Crustacean endocrine toxicology: a review." Ecotoxicology **16**(1): 61-81.

Lebrun, J. D., M. Perret, A. Geffard and C. Gourlay-France (2012). "Modelling copper bioaccumulation in Gammarus pulex and alterations of digestive metabolism." Ecotoxicology **21**(7): 2022-2030.

Lee, R. F. and T. Noone (1995). "Effect of reproductive toxicants on lipovitellin in female blue crabs, Callinectes sapidus." Marine Environmental Research **39**(1-4): 151-154.

Leroy, D., E. Haubruge, E. De Pauw, J. P. Thome and F. Francis (2010). "Development of ecotoxicoproteomics on the freshwater amphipod Gammarus pulex: identification of PCB biomarkers in glycolysis and glutamate pathways." Ecotoxicol Environ Saf **73**(3): 343-352.

Leung, P. T., Y. Wang, S. S. Mak, W. C. Ng and K. M. Leung (2011). "Differential proteomic responses in hepatopancreas and adductor muscles of the green-lipped mussel Perna viridis to stresses induced by cadmium and hydrogen peroxide." Aquat Toxicol **105**(1-2): 49-61.

Li, Z., Z. Zhang, P. Yan, S. Huang, Z. Fei and K. Lin (2011). "RNA-Seq improves annotation of protein-coding genes in the cucumber genome." BMC Genomics **12**(1): 540.

Liu, F., D. Z. Wang and W. X. Wang (2012). "Cadmium-induced changes in trace element bioaccumulation and proteomics perspective in four marine bivalves." Environ Toxicol Chem **31**(6): 1292-1300.

Liu, F. J. and W. X. Wang (2012). "Proteome pattern in oysters as a diagnostic tool for metal pollution." Journal of Hazardous Materials **239**: 241-248.

Luconi, M. (2006). "Rapid responses to estrogens." Medecine Therapeutique Medecine de la Reproduction **8**(2): 119-127.

Macneil, C., J. T. A. Dick and R. W. Elwood (1997). "The trophic ecology of freshwater Gammarus spp. (Crustacea: Amphipoda): Problems and perspectives concerning the functional feeding group concept." Biological Reviews of the Cambridge Philosophical Society **72**(3): 349-364.

Macneil, C., J. T. A. Dick and R. W. Elwood (1999). "The dynamics of predation on Gammarus spp. (Crustacea: Amphipoda)." Biological Reviews of the Cambridge Philosophical Society **74**(4): 375-395.

Maitra, R., D. N. Grigoryev, T. Kumar Bera, I. H. Pastan and B. Lee (2003). "Cloning, molecular characterization, and expression analysis of Copine 8." Biochemical and Biophysical Research Communications **303**(3): 842-847.

Major, K., D. J. Soucek, R. Giordano, M. J. Wetzel and F. Soto-Adames (2013). "The common ecotoxicology laboratory strain of Hyalella azteca is genetically distinct from most wild strains sampled in Eastern North America." Environ Toxicol Chem **32**(11): 2637-2647.

Maltby, L., S. A. Clayton, R. M. Wood and N. McLoughlin (2002). "Evaluation of the Gammarus pulex in situ feeding assay as a biomonitor of water quality: Robustness, responsiveness, and relevance." Environmental Toxicology and Chemistry **21**(2): 361-368.

Maltby, L. and L. Hills (2008). "Spray drift of pesticides and stream macroinvertebrates: experimental evidence of impacts and effectiveness of mitigation measures." Environ Pollut **156**(3): 1112-1120.

Manduzio, H., P. Cosette, L. Gricourt, T. Jouenne, C. Lenz, O. K. Andersen, F. Leboulenger and B. Rocher (2005). "Proteome modifications of blue mussel (Mytilus edulis L.) gills as an effect of water pollution." Proteomics **5**(18): 4958-4963.

Marinković, M., W. C. de Leeuw, M. de Jong, M. H. S. Kraak, W. Admiraal, T. M. Breit and M. J. Jonker (2012). "Combining Next-Generation Sequencing and Microarray Technology into a Transcriptomics Approach for the Non-Model Organism Chironomus riparius." PLoS ONE **7**(10).

Marshall, J. L., D. L. Huestis, C. Garcia, Y. Hiromasa, S. Wheeler, S. Noh, J. M. Tomich and D. J. Howard (2011). "Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes." Mol Biol Evol **28**(1): 423-435.
Martin-Diaz, M. L., A. Villena-Lincoln, S. Bamber, J. Blasco and T. A. DelValls (2005). "An integrated approach using bioaccumulation and biomarker measurements in female shore crab, Carcinus maenas." Chemosphere **58**(5): 615-626.
Martin, J. W. and G. E. Davis (2001). An updated classification of the recent crustacea. Natural History Museum of Los Angeles County. Los Angeles.
Martyniuk, C. J., S. Alvarez and N. D. Denslow (2012). "DIGE and iTRAQ as biomarker discovery tools in aquatic toxicology." Ecotoxicology and Environmental Safety **76**(1): 3-10.
Matozzo, V., F. Gagne, M. G. Marin, F. Ricciardi and C. Blaise (2008). "Vitellogenin as a biomarker of exposure to estrogenic compounds in aquatic invertebrates: a review." Environ Int **34**(4): 531-545.
Matthiessen, P. (2003). "Endocrine disruption in marine fish." Pure and Applied Chemistry **75**(11-12): 2249-2261.
Mazurová, E., K. Hilscherová, T. Šídlová-Štěpánková, H. Köhler, R. Triebskorn, D. Jungmann, J. P. Giesy and L. Bláha (2010). "Chronic toxicity of contaminated sediments on reproduction and histopathology of the crustacean Gammarus fossarum and relationship with the chemical contamination and in vitro effects." Journal of Soils and Sediments **10**(3): 423-433.
Mazurová, E., K. Hilscherová, R. Triebskorn, H.-R. Köhler, B. Maršálek and L. Bláha (2008). "Endocrine regulation of the reproduction in crustaceans: Identification of potential targets for toxicants and environmental contaminants." Biologia **63**(2): 139-150.
McKusick, V. A. and F. H. Ruddle (1987). "A new discipline, a new name, a new journal." Genomics **1**(1): 1-2.
McLoughlin, N., D. Q. Yin, L. Maltby, R. M. Wood and H. X. Yu (2000). "Evaluation of sensitivity and specificity of two crustacean biochemical biomarkers." Environmental Toxicology and Chemistry **19**(8): 2085-2092.
Menschaert, G., W. Van Criekinge, T. Notelaers, A. Koch, J. Crappe, K. Gevaert and P. Van Damme (2013). "Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events." Mol Cell Proteomics **12**(7): 1780-1790.
Mesquita, S. R., L. Guilhermino and L. Guimarães (2011). "Biochemical and locomotor responses of Carcinus maenas exposed to the serotonin reuptake inhibitor fluoxetine." Chemosphere **85**(6): 967-976.
Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nature reviews. Genetics **11**(1): 31-46.
Meyer, E., G. V. Aglyamova, S. Wang, J. Buchanan-Carter, D. Abrego, J. K. Colbourne, B. L. Willis and M. V. Matz (2009). "Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx." BMC Genomics **10**: 219.
Miller, J. R., S. Koren and G. Sutton (2010). "Assembly algorithms for next-generation sequencing data." Genomics **95**(6): 315-327.
Mindnich, R., G. Möller and J. Adamski (2004). "The role of 17 beta-hydroxysteroid dehydrogenases." Molecular and Cellular Endocrinology **218**(1–2): 7-20.
Muraoka, S., H. Kume, S. Watanabe, J. Adachi, M. Kuwano, M. Sato, N. Kawasaki, Y. Kodera, M. Ishitobi, H. Inaji, Y. Miyamoto, K. Kato and T. Tomonaga (2012). "Strategy for SRM-based verification of biomarker candidates discovered by iTRAQ method in limited breast cancer tissue samples." J Proteome Res **11**(8): 4201-4210.
Mykles, D. (2011). "Ecdysteroid metabolism in crustaceans." The Journal of Steroid Biochemistry and Molecular Biology **127**(3–5): 196-203.
Nagaraju, G. P. C. (2007). "Is methyl farnesoate a crustacean hormone?" Aquaculture **272**(1-4): 39-54.
Nagaraju, G. P. C. (2011). "Reproductive regulators in decapod crustaceans: an overview." The Journal of Experimental Biology **214**(1): 3-16.
Nagaraju, G. P. C. and D. W. Borst (2008). "Methyl farnesoate couples environmental changes to testicular development in the crab Carcinus maenas." Journal of Experimental Biology **211**(17): 2773-2778.

Nagele, E., M. Vollmer, P. Horth and C. Vad (2004). "2D-LC/MS techniques for the identification of proteins in highly complex mixtures." Expert Review of Proteomics **1**(1): 37-46.
Nakagawa, Y. and V. C. Henrich (2009). "Arthropod nuclear receptors and their role in molting." FEBS Journal **276**(21): 6128-6157.
Nakatsuji, T., C.-Y. Lee and R. D. Watson (2009). "Crustacean molt-inhibiting hormone: Structure, function, and cellular mode of action." Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology **152**(2): 139-148.
Nakatsuji, T., C. Y. Lee and R. D. Watson (2009). "Crustacean molt-inhibiting hormone: Structure, function, and cellular mode of action." Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology **152**(2): 139-148.
Nakazawa, T., M. Yamaguchi, T.-a. Okamura, E. Ando, O. Nishimura and S. Tsunasawa (2008). "Terminal proteomics: N- and C-terminal analyses for high-fidelity identification of proteins using MS." PROTEOMICS **8**(4): 673-685.
Narbonne, J.-F. (1998). Historique-Fondements biologiques de l'utilisation des biomarqeurs en écotoxicology Utilisation de biomarqueurs pour la surveillence de la qualité de l'environnement. L. Lagadic, T. Caquet, J. C. Amiard and F. Ramade. Paris, Lavoisier**:** 1-6.
Neave, M. J., C. Streten-Joyce, A. S. Nouwens, C. J. Glasby, K. A. McGuinness, D. L. Parry and K. S. Gibb (2012). "The transcriptome and proteome are altered in marine polychaetes (Annelida) exposed to elevated metal levels." J Proteomics **75**(9): 2721-2735.
Neuparth, T., A. D. Correia, F. O. Costa, G. Lima and M. H. Costa (2005). "Multi-level assessment of chronic toxicity of estuarine sediments with the amphipod Gammarus locusta: I. Biochemical endpoints." Mar Environ Res **60**(1): 69-91.
Okumura, T. (1992). "Changes in hemolymph vitellogenin and ecdysteroid levels during the reproductive and nonreproductive molt cycles in the freshwater prawn Macrobrachium nipponense." Zool. Sci. **9**: 37-45.
Okumura, T. and M. Hara (2004). "Androgenic gland cell structure and spermatogenesis during the molt cycle and correlation to morphotypic differentiation in the giant freshwater prawn, Macrobrachium rosenbergii." Zoological Science **21**(6): 621-628.
Okuno, A., Y. Hasegawa, T. Ohira, Y. Katakura and H. Nagasawa (1999). "Characterization and cDNA cloning of androgenic gland hormone of the terrestrial isopod Armadillidium vulgare." Biochemical and Biophysical Research Communications **264**(2): 419-423.
Olmstead, A. W. and G. A. Leblanc (2002). "Juvenoid hormone methyl farnesoate is a sex determinant in the crustacean Daphnia magna." Journal of Experimental Zoology **293**(7): 736-739.
Olmstead, A. W. and G. A. LeBlanc (2003). "Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean Daphnia magna." Environ Health Perspect **111**(7): 919-924.
Omasits, U., M. Quebatte, D. J. Stekhoven, C. Fortes, B. Roschitzki, M. D. Robinson, C. Dehio and C. H. Ahrens (2013). "Directed shotgun proteomics guided by saturated RNA-seq identifies a complete expressed prokaryotic proteome." Genome Res **23**(11): 1916-1927.
Otto, A., J. Bernhardt, M. Hecker and D. Becher (2012). "Global relative and absolute quantitation in microbial proteomics." Current Opinion in Microbiology **15**(3): 364-372.
Pain, S., S. Biaganti-Risbourg and M. Parant (2005). Relevance of the multixenobiotic defence mechanism (MXDM) for the biological monitoring of freshwaters - example of its use in zebra mussels. Trends in water pollution research. J. V. Livingston. New York, Nova Science Publishers**:** 203-220.
Parnes, S., S. Raviv, A. Shechter and A. Sagi (2006). "Males also have their time of the month! Cyclic disposal of old spermatophores, timed by the molt cycle, in a marine shrimp." J Exp Biol **209**(Pt 24): 4974-4983.
Patterson, S. D. and R. H. Aebersold (2003). "Proteomics: the first decade and beyond." Nat Genet **33 Suppl**: 311-323.
Percy, A. J., R. Simon, A. G. Chambers and C. H. Borchers (2014). "Enhanced sensitivity and multiplexing with 2D LC/MRM-MS and labeled standards for deeper and more comprehensive protein quantitation." Journal of Proteomics **106**(0): 113-124.
Petryk, A., J. T. Warren, G. Marques, M. P. Jarcho, L. I. Gilbert, J. Kahler, J. P. Parvy, Y. Li, C. Dauphin-Villemant and M. B. O'Connor (2003). "Shade is the Drosophila P450 enzyme that mediates the

hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone." Proc Natl Acad Sci U S A **100**(24): 13773-13778.
Poynton, H., D. Weston, G. Wellborn, M. Lydy, B. Blalock, K. Major and J. Lazorchak (2013). Genome Sequencing of Hyalella azteca: a model for evolutionary toxicology and ecological exposure. SETAC 2013. S. o. E. T. a. C. A. Meeting. Nashville.
R Development Core Team (2008). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.
Rabilloud, T., M. Chevallet, S. Luche and C. Lelong (2010). "Two-dimensional gel electrophoresis in proteomics: Past, present and future." Journal of Proteomics **73**(11): 2064-2077.
Ralston-Hooper, K. J., B. C. Sanchez, J. Adamec and M. S. Sepulveda (2011). "Proteomics in aquatic amphipods: can it be used to determine mechanisms of toxicity and interspecies responses after exposure to atrazine?" Environ Toxicol Chem **30**(5): 1197-1203.
Reeves, G. A., D. Talavera and J. M. Thornton (2009). "Genome and proteome annotation: organization, interpretation and integration." J R Soc Interface **6**(31): 129-147.
Reger, J. (1966). "A comparative study on the fine structure of developing spermatozoa in the isopod, Oniscus asellus, and the amphipod, Orchestoidea sp." Zeitschrift für Zellforschung und Mikroskopische Anatomie **75**(3): 579-590.
Regier, N., L. Baerlocher, M. Münsterkötter, L. Farinelli and C. Cosio (2013). "Analysis of the Elodea nuttallii transcriptome in response to mercury and cadmium pollution: Development of sensitive tools for rapid ecotoxicological testing." Environmental Science and Technology **47**(15): 8825-8834.
Revathi, P., P. Iyapparaj, L. A. Vasanthi, N. Munuswamy and M. Krishnan (2014). "Ultrastructural changes during spermatogenesis, biochemical and hormonal evidences of testicular toxicity caused by TBT in freshwater prawn Macrobrachium rosenbergii (De Man, 1879)." Environ Toxicol **29**(10): 1171-1181.
Riva, C., A. Binelli, F. Rusconi, G. Colombo, A. Pedriali, R. Zippel and A. Provini (2011). "A proteomic study using zebra mussels (D. polymorpha) exposed to benzo(α)pyrene: The role of gender and exposure concentrations." Aquatic Toxicology **104**(1–2): 14-22.
Riva, C., S. Cristoni and A. Binelli (2012). "Effects of triclosan in the freshwater mussel Dreissena polymorpha: a proteomic investigation." Aquat Toxicol **118-119**(0): 62-71.
Rodgers-Gray, T. P., J. E. Smith, A. E. Ashcroft, R. E. Isaac and A. M. Dunn (2004). "Mechanisms of parasite-induced sex reversal in Gammarus duebeni." International Journal for Parasitology **34**(6): 747-753.
Rodriguez-Ortega, M. J., B. E. Grosvik, A. Rodriguez-Ariza, A. Goksoyr and J. Lopez-Barea (2003). "Changes in protein expression profiles in bivalve molluscs (Chamaelea gallina) exposed to four model environmental pollutants." Proteomics **3**(8): 1535-1543.
Rodríguez, E. M., D. A. Medesani and M. Fingerman (2007). "Endocrine disruption in crustaceans due to pollutants: A review." Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology **146**(4): 661-671.
Roméo, M. and L. Giambérini (2012). History of Biomakers. Ecological Biomarkers - Indicators of Ecotoxicological Effects. C. Amiard-Triquet, J. C. Amiard and P. S. Rainbow. Boca Raton, CRC Press**:** 15-43.
Romero-Ruiz, A., M. Carrascal, J. Alhama, J. L. Gomez-Ariza, J. Abian and J. Lopez-Barea (2006). "Utility of proteomics to assess pollutant response of clams from the Donana bank of Guadalquivir Estuary (SW Spain)." Proteomics **6 Suppl 1**: S245-255.
Roux, L. A. (1970). "Les gammares du groupe pulex." Archives de Zoologie Expérimentale et Générale **313-356**(111).
Rubiano-Labrador, C., C. Bland, G. Miotello, P. Guerin, O. Pible, S. Baena and J. Armengaud (2014). "Proteogenomic insights into salt tolerance by a halotolerant alpha-proteobacterium isolated from an Andean saline spring." J Proteomics **97**(0): 36-47.
Rupik, W., K. Jasik, J. Bembenek and W. Widłak (2011). "The expression patterns of heat shock genes and proteins and their role during vertebrate's development." Comp Biochem Physiol A Mol Integr Physiol **159**(4): 349-366.
Sánchez, L. (2008). "Sex-determining mechanisms in insects." International Journal of Developmental Biology **52**(7): 837-856.

Sanders, M. B., Z. Billinghurst, M. H. Depledge and A. S. Clare (2005). "Larval Development and Vitellin-like Protein Expression in Palaemon elegans Larvae Following Xeno-oestrogen Exposure1." Integrative and Comparative Biology **45**(1): 51-60.
Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.
Scheil, V., R. Triebskorn and H. R. Kohler (2008). "Cellular and stress protein responses to the UV filter 3-benzylidene camphor in the amphipod crustacean Gammarus fossarum (Koch 1835)." Arch Environ Contam Toxicol **54**(4): 684-689.
Schill, R. O., H. Gorlitz and H. R. Kohler (2003). "Laboratory simulation of a mining accident: acute toxicity, hsc/hsp70 response, and recovery from stress in Gammarus fossarum (Crustacea, Amphipoda) exposed to a pulse of cadmium." BioMetals **16**(3): 391-401.
Schirling, M., D. Jungmann, V. Ladewig, K. U. Ludwichowski, R. Nagel, H. R. Kohler and R. Triebskorn (2006). "Bisphenol A in artificial indoor streams: II. Stress response and gonad histology in Gammarus fossarum (Amphipoda)." Ecotoxicology **15**(2): 143-156.
Schirling, M., D. Jungmann, V. Ladewig, R. Nagel, R. Triebskorn and H. R. Kohler (2005). "Endocrine effects in Gammarus fossarum (Amphipoda): influence of wastewater effluents, temporal variability, and spatial aspects on natural populations." Arch Environ Contam Toxicol **49**(1): 53-61.
Schliesky, S., U. Gowik, A. P. Weber and A. Brautigam (2012). "RNA-Seq Assembly - Are We There Yet?" Front Plant Sci **3**: 220.
Scott, A. P. (2012). "Do mollusks use vertebrate sex steroids as reproductive hormones? Part I: Critical appraisal of the evidence for the presence, biosynthesis and uptake of steroids." Steroids **77**(13): 1450-1468.
Seehuus, S. C., K. Norberg, U. Gimsa, T. Krekling and G. V. Amdam (2006). "Reproductive protein protects functionally sterile honey bee workers from oxidative stress." Proceedings of the National Academy of Sciences of the United States of America **103**(4): 962-967.
Semenza, J. C., P. E. Tolbert, C. H. Rubin, Guillette L.J, Jr. and R. J. Jackson (1997). "Reproductive toxins and alligator abnormalities at Lake Apopka, Florida." Environmental Health Perspectives **105**(10): 1030-1032.
Serafini, M. (2001). Pyriproxyfen - Registration of a Major Change of Label for Esteem® 3/01. D. o. S. H. Materials. New York, New York State Department of Environmental Conservation.
Shatilina, Z., T. Pobezhimova, O. Grabel'nykh, D. Bedulina, M. Protopopova, V. Pavlichenko and M. Timofeev (2010). "Heat shock proteins in the mechanisms of stress adaptation in Baikal amphipods and palearctic *Gammarus lacustris* Sars. I. HSP70 family." Contemporary Problems of Ecology **3**(1): 41-49.
Short, S., G. Yang, P. Kille and A. T. Ford (2014). "Vitellogenin is not an appropriate biomarker of feminisation in a crustacean." Aquat Toxicol **153**(0): 89-97.
Silvestre, F., V. Gillardin and J. Dorts (2012). "Proteomics to assess the role of phenotypic plasticity in aquatic organisms exposed to pollution and global warming." Integr Comp Biol **52**(5): 681-694.
Simon, R., G. Jubeaux, A. Chaumot, J. Lemoine, O. Geffard and A. Salvador (2010). "Mass spectrometry assay as an alternative to the enzyme-linked immunosorbent assay test for biomarker quantitation in ecotoxicology: application to vitellogenin in Crustacea (Gammarus fossarum)." J Chromatogr A **1217**(31): 5109-5115.
Soetaert, A., L. N. Moens, K. Van der Ven, K. Van Leemput, B. Naudts, R. Blust and W. M. De Coen (2006). "Molecular impact of propiconazole on Daphnia magna using a reproduction-related cDNA array." Comp Biochem Physiol C Toxicol Pharmacol **142**(1-2): 66-76.
Song, M. Y., J. D. Stark and J. J. Brown (1997). "Comparative toxicity of four insecticides, including imidacloprid and tebufenozide, to four aquatic arthropods." Environmental Toxicology and Chemistry **16**(12): 2494-2500.
Sousa, A. C. A., M. R. Pastorinho, S. Takahashi and S. Tanabe (2014). "History on organotin compounds, from snails to humans." Environmental Chemistry Letters **12**(1): 117-137.
Spindler-Barth, M., A. Wormhoudt and K. Spindler (1990). "Chitinolytic enzymes in the integument and midgut-gland of the shrimp<i>Palaemon serratus</i> during the moulting cycle." Marine Biology **106**(1): 49-52.

Sroda, S. and C. Cossu-Leguille (2011). "Effects of sublethal copper exposure on two gammarid species: which is the best competitor?" Ecotoxicology **20**(1): 264-273.
Sroda, S. and C. Cossu-Leguille (2011). "Seasonal variability of antioxidant biomarkers and energy reserves in the freshwater gammarid Gammarus roeseli." Chemosphere **83**(4): 538-544.
Stranneheim, H. and J. Lundeberg (2012). "Stepping stones in DNA sequencing." Biotechnol J **7**(9): 1063-1073.
Stuenkel, E. L. (1983). "Biosynthesis and axonal transport of proteins and identified peptide hormones in the X-organ sinus gland neurosecretory system." Journal of Comparative Physiology ■ B **153**(2): 191-205.
Styrishave, B., T. Lund and O. Andersen (2008). "Ecdysteroids in female shore crabs Carcinus maenas during the moulting cycle and oocyte development." J Mar Biol Assoc UK **88**(3): 575-581.
Subramoniam, T. (2000). "Crustacean ecdysteriods in reproduction and embryogenesis." Comp Biochem Physiol C Toxicol Pharmacol **125**(2): 135-156.
Tachet, H., M. Bournaud, P. Richoux and P. Usseglio-Polaterra (2000). Invertébrés d'eau douce : Systématiques, biologie, écologie. Paris, CNRS édition.
Takizawa, Y. (2000). Understanding minamata disease and strategies to prevent further environmental contamination by methylmercury. **42:** 139-146.
Thompson, E. L., D. A. Taylor, S. V. Nair, G. Birch, P. A. Haynes and D. A. Raftos (2011). "A proteomic analysis of the effects of metal contamination on Sydney Rock Oyster (Saccostrea glomerata) haemolymph." Aquat Toxicol **103**(3-4): 241-249.
Timofeyev, M. A., Z. M. Shatilina, M. V. Protopopova, D. S. Bedulina, V. V. Pavlichenko, A. V. Kolesnichenko and C. E. W. Steinberg (2009). "Thermal stress defense in freshwater amphipods from contrasting habitats with emphasis on small heat shock proteins (sHSPs)." Journal of Thermal Biology **34**(6): 281-285.
Tokishita, S., Y. Kato, T. Kobayashi, S. Nakamura, T. Ohta and H. Yamagata (2006). "Organization and repression by juvenile hormone of a vitellogenin gene cluster in the crustacean, Daphnia magna." Biochem Biophys Res Commun **345**(1): 362-370.
Transboundary Freshwater Dispute Database (2003). Atlas of International Freshwater Agreements. Oregon State University. http://www.transboundarywaters.orst.edu/publications/atlas/, United Nations Environment Programme

Food and Agriculture Organization of the United Nations.
Trapp, J., C. Almunia, J. C. Gaillard, O. Pible, A. Chaumot, O. Geffard and J. Armengaud (2015). "Proteogenomic insights into the core-proteome of female reproductive tissues from crustacean amphipods." J Proteomics.
Triebskorn, R., S. Adam, H. Casper, W. Honnen, M. Pawert, M. Schramm, J. Schwaiger and H. R. Kohler (2002). "Biomarkers as diagnostic tools for evaluating effects of unknown past water quality conditions on stream organisms." Ecotoxicology **11**(6): 451-465.
Tuberty, S. R. and C. L. McKenney Jr (2005). "Ecdysteroid responses of estuarine crustaceans exposed through complete larval development to juvenile hormone agonist insecticides." Integrative and Comparative Biology **45**(1): 106-117.
Tulsiani, D. R. P., A. Abou-Haila, C. R. Loeser and B. M. J. Pereira (1998). "The biological and functional significance of the sperm acrosome and acrosomal enzymes in mammalian fertilization." Experimental Cell Research **240**(2): 151-164.
Turja, R., L. Guimaraes, A. Nevala, H. Kankaanpaa, S. Korpinen and K. K. Lehtonen (2014). "Cumulative effects of exposure to cyanobacteria bloom extracts and benzo[a]pyrene on antioxidant defence biomarkers in Gammarus oceanicus (Crustacea: Amphipoda)." Toxicon **78**: 68-77.
van der Oost, R., J. Beyer and N. P. Vermeulen (2003). "Fish bioaccumulation and biomarkers in environmental risk assessment: a review." Environ Toxicol Pharmacol **13**(2): 57-149.
van Dijk, E. L., H. Auger, Y. Jaszczyszyn and C. Thermes (2014). "Ten years of next-generation sequencing technology." Trends Genet **30**(9): 418-426.
Vandenbergh, G. F., D. Adriaens, T. Verslycke and C. R. Janssen (2003). "Effects of 17α-ethinylestradiol on sexual development of the amphipod Hyalella azteca." Ecotoxicology and Environmental Safety **54**(2): 216-222.

Vandenbrouck, T., A. Soetaert, K. van der Ven, R. Blust and W. De Coen (2009). "Nickel and binary metal mixture responses in Daphnia magna: molecular fingerprints and (sub)organismal effects." Aquat Toxicol **92**(1): 18-29.
Ventura, T., O. Rosen and A. Sagi (2011). "From the discovery of the crustacean androgenic gland to the insulin-like hormone in six decades." General and Comparative Endocrinology **173**(3): 381-388.
Vigneron, A., O. Geffard, H. Pellerin, R. Recoura-Massaquant, A. François, H. Quéau, T. Rigaud, R. Wattier and A. Chaumot (2014). Variability of Cd-sensitivity and phylogenetic diversity of field populations throughout the Gammarus fossarum/pulex species complex. SETAC Europe 24th Annual Meeting. Basel.
Volz, D. C., T. Kawaguchi and G. T. Chandler (2002). "Purification and characterization of the common yolk protein, vitellin, from the estuarine amphipod Leptocheirus plumulosus." Prep Biochem Biotechnol **32**(2): 103-116.
von Reumont, B. M., R. A. Jenner, M. A. Wills, E. Dell'ampio, G. Pass, I. Ebersberger, B. Meyer, S. Koenemann, T. M. Iliffe, A. Stamatakis, O. Niehuis, K. Meusemann and B. Misof (2012). "Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda." Molecular Biology and Evolution **29**(3): 1031-1045.
Wainwright, G., S. G. Webster, M. C. Wilkinson, J. S. Chung and H. H. Rees (1996). "Structure and Significance of Mandibular Organ-inhibiting Hormone in the Crab, Cancer pagurus: INVOLVEMENT IN MULTIHORMONAL REGULATION OF GROWTH AND REPRODUCTION." Journal of Biological Chemistry **271**(22): 12749-12754.
Wang, X., R. J. C. Slebos, D. Wang, P. J. Halvey, D. L. Tabb, D. C. Liebler and B. Zhang (2012). "Protein identification using customized protein sequence databases derived from RNA-seq data." Journal of Proteome Research **11**(2): 1009-1017.
Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: A revolutionary tool for transcriptomics." Nature Reviews Genetics **10**(1): 57-63.
Washburn, M. P., D. Wolters and J. R. Yates, 3rd (2001). "Large-scale analysis of the yeast proteome by multidimensional protein identification technology." Nat Biotechnol **19**(3): 242-247.
Watts, M. M., D. Pascoe and K. Carroll (2002). "Population responses of the freshwater amphipod Gammarus pulex (L.) to an environmental estrogen, 17α-ethinylestradiol." Environmental Toxicology and Chemistry **21**(2): 445-450.
Webster, S. G., R. Keller and H. Dircksen (2012). "The CHH-superfamily of multifunctional peptide hormones controlling crustacean metabolism, osmoregulation, moulting, and reproduction." General and Comparative Endocrinology **175**(2): 217-233.
Winkler, H. (1920). Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche by Winkler. Jena, Germany, Fischer, G. .
Wu, L. T., J. H. Hui and K. H. Chu (2013). "Origin and evolution of yolk proteins: expansion and functional diversification of large lipid transfer protein superfamily." Biol Reprod **88**(4): 102.
Wu, Y., X. Chen, S. Wang, M. Jiang, B. Zheng, Q. Zhou, Y. Bi, Z. Zhou, X. Huang and J. Sha (2012). "Flotillin-2 is an acrosome-related protein involved in mouse spermiogenesis." Journal of Biomedical Research **26**(4): 278-287.
Xu, B. J. (2010). "Combining laser capture microdissection and proteomics: methodologies and clinical applications." Proteomics Clin Appl **4**(2): 116-123.
Xuereb, B. (2009). Développement de marqueurs de neurotoxicité et de perturbations endocrines chez l'amphipode d'eau douce *Gammarus fossarum*. P.h.D, Paul Verlaine.
Xuereb, B., L. Bezin, A. Chaumot, H. Budzinski, S. Augagneur, R. Tutundjian, J. Garric and O. Geffard (2011). "Vitellogenin-like gene expression in freshwater amphipod Gammarus fossarum (Koch, 1835): functional characterization in females and potential for use as an endocrine disruption biomarker in males." Ecotoxicology **20**(6): 1286-1299.
Xuereb, B., A. Chaumot, R. Mons, J. Garric and O. Geffard (2009). "Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda) Intrinsic variability, reference levels, and a reliable tool for field surveys." Aquat Toxicol **93**(4): 225-233.
Xuereb, B., E. Lefevre, J. Garric and O. Geffard (2009). "Acetylcholinesterase activity in Gammarus fossarum (Crustacea Amphipoda): linking AChE inhibition and behavioural alteration." Aquat Toxicol **94**(2): 114-122.

Yang, G., P. Kille and A. T. Ford (2008). "Infertility in a marine crustacean: Have we been ignoring pollution impacts on male invertebrates?" Aquatic Toxicology **88**(1): 81-87.
Yates, J. R., J. K. Eng and A. L. McCormack (1995). "Mining Genomes: Correlating Tandem Mass Spectra of Modified and Unmodified Peptides to Sequences in Nucleotide Databases." Analytical Chemistry **67**(18): 3202-3210.
Yoshida, M. (2000). "Conservation of sperms: current status and new trends." Animal Reproduction Science **60–61**(0): 349-355.
Yu, Q., C. OuYang, Z. Liang and L. Li (2014). "Mass spectrometric characterization of the crustacean neuropeptidome." EuPA Open Proteomics **3**(0): 152-170.
Zhou, J., Z. H. Cai, L. Li, Y. F. Gao and T. H. Hutchinson (2010). "A proteomics based approach to assessing the toxicity of bisphenol A and diallyl phthalate to the abalone (Haliotis diversicolor supertexta)." Chemosphere **79**(5): 595-604.
Zhou, J., W.-N. Wang, A.-L. Wang, W.-Y. He, Q.-T. Zhou, Y. Liu and J. Xu (2009). "Glutathione S-transferase in the white shrimp Litopenaeus vannamei: Characterization and regulation under pH stress." Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **150**(2): 224-230.
Zivanovic, Y., J. Armengaud, A. Lagorce, C. Leplat, P. Guerin, M. Dutertre, V. Anthouard, P. Forterre, P. Wincker and F. Confalonieri (2009). "Genome analysis and genome-wide proteomics of Thermococcus gammatolerans, the most radioresistant organism known amongst the Archaea." Genome Biology **10**(6): 76.
Zou, E. and M. Fingerman (1999). "Chitobiase activity in the epidermis and hepatopancreas of the fiddler crab Uca pugilator during the molting cycle." Marine Biology **133**(1): 97-101.

ANNEXES

Annexe I : Protéines séquencées par nanoLC/MS-MS au cours de l'ensemble des expérimentations

Annexe II : Protéines identifiées comme étant d'intérêt en écotoxicologie sur la base de leur annotation fonctionnelle

Annexe III : Protéines définies comme présentant un dimorphisme sexuel d'après l'analyse comparative du protéome des gonades mâles et femelles

Annexe IV : Protéines modulées dans les gonades mâles en réponse à la pression toxique

Annexe I : Protéines séquencées par nanoLC/MS-MS au cours de l'ensemble des expérimentations

Contig_ID	ORF	pI	Mass	NBCI_ID	Fonction prédite	Origine du plus proche homologue	Blast score	Blast Evalue	Céphalon	Caecums	Ovaires	Testicules	Ovogénèse	Embryogénèse	Spermatogénèse	Expérimentation toxique
5	fr4	5,3	47606,3	gb|AAL39886.1|	No predicted function	*Drosophila melanogaster*	1440	0,0E+00	X							
8	fr4	8,2	63382,2	ref|XP_795649.3|	d-3-phosphoglycerate dehydrogenase-like	*Strongylocentrotus purpuratus*	1174	6,0E-151	X		X	X			X	X
14	fr3	9,1	46360,4	gb|AAP59794.1|	slow muscle myosin S1 heavy chain	*Homarus americanus*	970	2,0E-123	X		X	X	X	X	X	X
18	fr4	10,5	111884,3	gb|AET06086.1|	envelope glycoprotein, partial	*Human immunodeficiency virus 1*	104	5,3E-02			X		X			
47	fr4	9,5	57210,0	sp|Q05808.1|	Vitellogenin	*Anthonomus grandis*	110	1,7E-02			X	X	X	X		
51	fr5	7,9	66502,9	gb|EMC83869.1|	Methylmalonate-semialdehyde dehydrogenase	*Columba livia*	1781	0,0E+00	X		X	X				
55	fr1	8,9	100910,8	ref|NP_001071208.1|	methylcrotonoyl-CoA carboxylase subunit alpha, mitochondria	*Danio rerio*	1871	0,0E+00	X							
64	fr6	10,3	119422,5	XP_001607388	vitellogenin	*Nasonia vitripennis*	122	3,1E-03			X	X	X	X		X
95	fr1	8,6	119466,6	dbj|BAM19130.1|	limpet protein	*Papilio polytes*	926	3,0E-113				X			X	
102	fr5	10,2	60667,6	ESO11450.1	No predicted function	*Helobdella robusta*	657	3,0E-75					X			
108	fr5	9,5	77512,8	ref|XP_002424350.1|	No predicted function	*Pediculus humanus corporis*	538	2,0E-58							X	X
118	fr4	9,6	94638,9	ref|XP_975384.1|	No predicted function	*Tribolium castaneum*	342	6,0E-34			X	X				
124	fr5	5,5	82935,3	ref|XP_966352.1|	ubiquitin-activating enzyme E1	*Tribolium castaneum*	1586	0,0E+00			X	X		X	X	X
129	fr5	9,1	80707,2	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	2074	0,0E+00	X			X			X	
143	fr6	6,4	66728,1	gb|EFX86896.1|	No predicted function	*Daphnia pulex*	182	8,0E-12							X	X
145	fr4	9,2	177775,8	ref|NP_001036035.1|	chitinase 7 precursor	*Tribolium castaneum*	1739	0,0E+00			X					
151	fr6	9,2	69238,2		No detected homolog				X		X	X	X	X	X	
153	fr2	7,6	51178,0	gb|AAM47019.1|	dopamine beta hydroxylase	*Homarus americanus*	594	6,0E-68			X					
155	fr4	8,5	145135,7	ref|XP_003691374.1|	heat shock 70 kDa protein 4L-like	*Apis florea*	1645	0,0E+00			X	X			X	X
156	fr4	7,8	203122,5	ref|XP_002430009.1|	Spectrin alpha chain,	*Pediculus humanus corporis*	4298	0,0E+00	X		X	X			X	X
184	fr3	5,4	38884,4	gb|EKC23074.1|	Putative aminopeptidase	*Crassostrea gigas*	968	1,0E-123		X						
192	fr3	5,6	195797,9	gb|EFX86426.1|	No predicted function	*Daphnia pulex*	5426	0,0E+00	X			X				
202	fr5	9,0	151818,7	gb|EFX77111.1|	No predicted function	*Daphnia pulex*	1768	0,0E+00							X	X
213	fr1	8,4	152253,2	ref|NP_001025406.1|	echinoderm microtubule-associated protein	*Danio rerio*	1479	0,0E+00	X		X	X			X	
276	fr4	9,6	130902,0	gb|ABI95361.1|	hemolymph clottable protein	*Litopenaeus vannamei*	129	3,0E-04			X		X	X		
277	fr1	9,5	61856,9	gb|ACU51164.1|	vitellogenin	*Pandalopsis japonica*	189	9,0E-12			X		X	X		
286	fr2	5,6	84630,3		No detected homolog										X	
327	fr3	9,0	45326,7	BAD11098.1	vitellogenin	*Pandalus hypsinotus*	280	2,9E-23					X	X		
336	fr1	9,9	129064,9	gb|EFX73618.1|	No predicted function	*Daphnia pulex*	393	2,0E-40				X				
343	fr5	10,5	121579,3	CAA39363	clathrin heavy chain	*Homo sapiens*	542	2,7E-58								X
345	fr2	10,4	90851,6	ACZ58333.1	elongation factor-1 alpha	*Cyamus erraticus*	712	6,8E-82						X		X
374	fr6	6,0	126489,4	ref|XP_002194844.2|	gem-associated protein 5	*Taeniopygia guttata*	487	2,0E-46			X					
379	fr6		219170,0	ref|XP_002054824.1|	No predicted function	*Drosophila virilis*	723	3,0E-78	X							
417	fr5	10,4	42414,8	gb|ACO14779.1|	60S acidic ribosomal protein P0	*Caligus clemensi*	929	2,0E-120	X			X		X		
422	fr4	5,3	40443,8	gb|ACQ72825.1|	myosin heavy chain	*Bombyx mandarina*	582	3,0E-64	X		X	X	X	X	X	X
439	fr2	8,9	140345,0	gb|EFX89582.1|	No predicted function	*Daphnia pulex*	223	2,0E-15				X			X	X
441	fr2	7,9	49658,2	ref|XP_001986651.1|	No predicted function	*Drosophila grimshawi*	554	2,0E-61			X	X	X	X	X	X
451	fr4	5,3	34820,0	gb|ACO36738.1|	ubiquitin carboxyl-terminal esterase L3	*Scylla paramamosain*	602	2,0E-73			X	X				
452	fr5	8,4	60984,3	gb|EFX79712.1|	No predicted function	*Daphnia pulex*	1581	0,0E+00			X		X			
461	fr1	8,9	110793,5	ref|XP_003694231.1|	hemocytin-like	*Apis florea*	260	1,0E-19				X			X	X
470	fr4	10,5	263691,2	gb|EFX67006.1|	No predicted function	*Daphnia pulex*	4512	0,0E+00	X		X	X		X	X	X
521	fr6	8,5	130458,9	gb|EFX89414.1|	No predicted function	*Daphnia pulex*	1053	4,0E-128	X		X	X			X	X
525	fr3	8,2	57084,7	ref|XP_001607937.1|	glycerol-3-phosphate dehydrogenase	*Nasonia vitripennis*	1278	4,0E-170			X	X				X
559	fr2	9,2	92139,0	XP_005187617.1	delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondria	*Musca domestica*	671	7,9E-74			X		X			
562	fr1	9,2	86049,2	gb|AFE19188.1|	plasma membrane calcium ATPase	*Callinectes sapidus*	1852	0,0E+00	X							
572	fr1	5,9	53013,7	gb|AEV89752.1|	heat shock protein 60	*Schistocerca gregaria*	1466	0,0E+00			X					
581	fr4	9,8	219781,2	gb|EFA07696.1|	No predicted function	*Tribolium castaneum*	155	6,0E-07	X			X				
584	fr2	6,1	120518,6	gb|EFX86468.1|	No predicted function	*Daphnia pulex*	3429	0,0E+00				X				
608	fr6	5,0	65246,6	ERL94345	No predicted function	*Dendroctonus ponderosae*	1155	7,5E-146	X			X			X	X
626	fr4	9,2	66588,3	gb|AFN66646.1|	glucan pattern-recognition lipoprotein	*Fenneropenaeus chinensis*	446	3,0E-43	X		X	X	X	X	X	X
654	fr3	9,7	138776,4	ref|XP_966465.1|	alanyl-tRNA synthetase	*Tribolium castaneum*	2014	0,0E+00				X				
657	fr4	9,9	58824,0	gb|EGI64907.1|	Phosphoglycerate kinase	*Acromyrmex echinatior*	954	2,0E-120	X		X	X			X	X
679	fr2	9,8	154942,7	gb|EFA03240.1|	No predicted function	*Tribolium castaneum*	484	6,0E-47				X			X	X
681	fr6	6,5	35017,2		No detected homolog						X		X			
697	fr6	8,6	64735,6	gb|EFX83224.1|	No predicted function	*Daphnia pulex*	1161	1,0E-142							X	X
724	fr1	9,1	87774,5	gb|EFX86357.1|	No predicted function	*Daphnia pulex*	769	6,0E-84		X						
737	fr2	9,1	74512,8	ref|XP_001607215.1|	1-acyl-sn-glycerol-3-phosphate acyltransferase gamma-like	*Nasonia vitripennis*	957	8,0E-120							X	X
774	fr1	8,3	127790,6		No detected homolog						X					
806	fr5	5,0	26328,5		No detected homolog						X					
813	fr4	9,4	71547,2	AGG09855	glutamate dehydrogenase 1, partial	*Nilaparvata lugens*	794	1,5E-95								X
823	fr5	9,0	150932,2	gb|EFX71199.1|	lamin-like protein	*Daphnia pulex*	483	1,0E-47			X	X				X
880	fr4	9,5	92568,4	ref|XP_003704912.1|	insulin-like growth factor 2 mRNA-binding protein 1-like	*Megachile rotundata*	731	4,0E-82			X					
933	fr1	9,3	184324,4	gb|EFX88967.1|	No predicted function	*Daphnia pulex*	1028	2,0E-125			X	X				
1019	fr4	9,2	159401,3	gb|EFX83852.1|	No predicted function	*Daphnia pulex*	1197	2,0E-143	X			X			X	X
1021	fr2	9,8	100276,1	ref|XP_001603180.2|	threonyl-tRNA synthetase, cytoplasmic-like	*Nasonia vitripennis*	1675	0,0E+00				X				
1024	fr3	7,4	167318,2	ref|XP_001017770.1|	No predicted function	*Tetrahymena thermophila*	176	9,0E-10	X			X			X	
1034	fr3	9,0	68890,8	gb|EHJ69744.1|	serine protease inhibitor 6	*Danaus plexippus*	448	1,0E-45				X				
1038	fr1	5,4	111373,3	ELU16163	No predicted function	*Capitella teleta*	2279	0,0E+00								X
1065	fr4	8,7	26989,3	gb|ACM91676.1|	F1F0-ATP synthase beta subunit	*Marsupenaeus japonicus*	947	3,0E-122	X		X	X	X	X	X	X
1071	fr3	9,4	59021,4	gb|EFX80375.1|	No predicted function	*Daphnia pulex*	1558	0,0E+00	X							
1075	fr6	9,5	93072,8	ref|XP_004080918.1|	enoyl-CoA delta isomerase 2, mitochondrial-like	*Oryzias latipes*	776	8,0E-91			X					
1094	fr3	5,2	60728,1	ref|XP_001952865.2|	hemocytin-like	*Acyrthosiphon pisum*	251	3,0E-19							X	X
1098	fr4	9,4	51973,9	gb|EFX72448.1|	No predicted function	*Daphnia pulex*	1133	8,0E-150			X					
1134	fr1	9,5	125446,6	ref|XP_003737268.1|	transitional endoplasmic reticulum ATPase TER94-like isoform	*Metaseiulus occidentalis*	1447	0,0E+00						X	X	X
1139	fr2	9,6	57024,5	ADQ43366	No predicted function	*Litopenaeus vannamei*	515	3,6E-58								X
1148	fr3	10,5	60609,2	EFX67764	No predicted function	*Daphnia pulex*	240	1,7E-18								X
1157	fr3	11,1	29933,8	XP_005072715.1	carboxypeptidase Q	*Mesocricetus auratus*	341	2,9E-33					X	X		
1162	fr1	9,1	93938,8	gb|EFX79772.1|	No predicted function	*Daphnia pulex*	929	7,0E-113				X			X	X
1169	fr5	8,3	102489,5	gb|EFZ17862.1|	No predicted function	*Solenopsis invicta*	2111	0,0E+00				X			X	
1170	fr4	9,4	41044,6	gb|ACY66388.1|	chaperonin 10	*Scylla paramamosain*	348	7,0E-37			X					
1172	fr4	10,2	53295,2	ref|XP_003400694.1|	60S ribosomal protein L3-like	*Bombus terrestris*	1516	0,0E+00			X	X		X	X	
1186	fr5	5,7	111349,4	gb|EFN89572.1|	No predicted function	*Harpegnathos saltator*	195	6,0E-13							X	
1216	fr3	7,0	33082,5	EHJ72859.1	rap55	*Danaus plexippus*	374	9,6E-37					X			
1226	fr5		72556,0	ref|NP_001086341.1|	enoyl CoA hydratase 1, peroxisomal	*Xenopus laevis*	744	1,0E-89	X		X				X	
1256	fr1	8,0	74080,8	AGD79959.1	heat shock protein 60	*Marsupenaeus japonicus*	1023	1,4E-126	X				X	X	X	X
1262	fr5	10,5	71796,7		No detected homolog							X				
1263	fr2	6,3	74103,8	gb|EFX84177.1|	No predicted function	*Daphnia pulex*	818	3,0E-98	X							
1270	fr2	9,2	172897,8	gb|ADQ64537.1|	No predicted function	*Bactrocera oleae*	231	1,0E-17			X	X				
1277	fr6	9,4	68868,3	ref|XP_001657588.1|	short-chain dehydrogenase	*Aedes aegypti*	491	3,0E-53							X	X
1282	fr3	10,0	49025,9	EAY79196	No predicted function	*Oryza sativa Indica Group*	545	7,2E-60								X
1295	fr5	6,8	50556,8	gb|ADJ94951.2|	voltage-dependent anion-selective channel	*Eriocheir sinensis*	1256	7,0E-169			X				X	X
1304	fr2	8,8	43064,0	ref|XP_002594275.1|	No predicted function	*Branchiostoma floridae*	834	3,0E-105			X					
1344	fr4	8,7	111446,9	gb|EFX74372.1|	No predicted function	*Daphnia pulex*	1380	1,0E-176							X	X
1366	fr1		125601,7	gb|EFZ14493.1|	No predicted function	*Solenopsis invicta*	2143	0,0E+00	X			X			X	X
1425	fr6	11,0	37970,8	gb|AAW34065.1|	heat shock protein 90	*Homarus americanus*	768	8,0E-97				X			X	
1524	fr3	9,0	31938,2	ref|XP_001656296.1|	snare protein sec22	*Aedes aegypti*	695	5,0E-88			X	X				
1538	fr1	5,2	29559,5	gb|EFX80396.1|	No predicted function	*Daphnia pulex*	380	3,0E-38				X			X	
1562	fr4	9,0	68281,5	ref|XP_002413056.1|	phosphoglucomutase,	*Ixodes scapularis*	915	3,0E-111				X				
1574	fr5	11,7	135597,7	XP_002423351.1	Filamin-C, putative	*Pediculus humanus corporis*	1200	1,1E-143	X			X		X	X	X
1580	fr3	9,3	45374,4		No detected homolog						X		X			
1596	fr2	4,9	28387,5	gb|EFX80810.1|	No predicted function	*Daphnia pulex*	284	3,0E-26			X					
1601	fr2	9,7	70828,2	gb|ELU14716.1|	No predicted function	*Capitella teleta*	1179	2,0E-152							X	X
1605	fr6	8,4	84496,9	ref|XP_318078.2|	No predicted function	*Anopheles gambiae str. PEST*	2183	0,0E+00			X					
1628	fr1		45724,9	gb|EFX71334.1|	cytosolic malate dehydrogenase	*Daphnia pulex*	1175	1,0E-156			X	X			X	X
1641	fr2	8,8	81035,0	gb|ENN70365.1|	No predicted function	*Dendroctonus ponderosae*	1295	2,0E-169	X							
1683	fr4	8,4	78016,3		No detected homolog											X
1703	fr5	7,2	80232,9		No detected homolog						X					
1718	fr3	9,3	76094,1	gb|ENN77096.1|	No predicted function	*Dendroctonus ponderosae*	981	5,0E-118			X	X				
1719	fr3	5,4	54189,8		No detected homolog										X	X
1736	fr5	10,7	67035,9	gb|AEM31042.1|	cellobiohydrolase I, partial	*Uncharacterized*	376	7,0E-39							X	
1737	fr2	8,3	99504,6	XP_001850333	6-phosphofructokinase	*Culex quinquefasciatus*	2521	0,0E+00								X
1742	fr4	10,5	38502,6	ref|XP_002156937.1|	electron transfer flavoprotein subunit beta-like	*Hydra magnipapillata*	485	4,0E-55	X			X				
1784	fr6	8,0	114133,3	gb|AFC36877.1|	14-3-3-like protein	*Litopenaeus vannamei*	1067	4,0E-134	X			X		X		X
1803	fr2	5,9	72437,3	gb|EFX72588.1|	No predicted function	*Daphnia pulex*	917	2,0E-110							X	X
1811	fr1	9,3	170351,6	EKC21093	Ribosome-binding protein 1	*Crassostrea gigas*	203	1,2E-12								X
1826	fr5	9,5	94736,4	gb|EFX80375.1|	No predicted function	*Daphnia pulex*	1764	0,0E+00	X		X	X				
1836	fr6	5,1	118396,1	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	2653	0,0E+00	X			X				X
1917	fr2	6,1	46277,4	ref|XP_972752.1|	transglutaminase	*Tribolium castaneum*	997	6,0E-125	X			X			X	X
1966	fr2	9,8	111904,9		No detected homolog										X	X
1975	fr6	8,5	91671,8	ref|XP_003694231.1|	hemocytin-like	*Apis florea*	247	3,0E-18				X			X	X
1998	fr2	9,8	171959,3	dbj|BAG56705.1|	No predicted function	*Homo sapiens*	637	2,0E-67	X							
2029	fr2	6,4	311727,3	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	10929	0,0E+00	X			X		X	X	X
2041	fr6	7,3	45487,3	ref|XP_976164.2|	No predicted function	*Tribolium castaneum*	1208	1,0E-159	X			X				
2042	fr5	9,0	151526,2	gb|ADE34166.1|	vitellogenin receptor	*Nilaparvata lugens*	353	2,0E-30			X					
2051	fr3	9,4	78024,6	gb|EFX78146.1|	No predicted function	*Daphnia pulex*	1560	0,0E+00				X				
2058	fr6	8,0	43764,8		No detected homolog				X			X				
2100	fr3	5,7	31372,6	ref|XP_004441663.1|	periaxin-like	*Ceratotherium simum simum*	158	5,0E-09	X			X				
2101	fr3	9,6	70187,2	gb|ACY66386.2|	histone H2AV	*Scylla paramamosain*	668	7,0E-81						X	X	X

2151	fr2	6,5	32588,3	dbj|BAJ23879.1|	glycogen phosphorylase	*Marsupenaeus japonicus*	1068	7,0E-136								X		X
2179	fr2	9,7	48927,5	dbj|BAH57291.1|	ovarian lipoprotein receptor	*Marsupenaeus japonicus*	150	2,0E-07				X						
2216	fr2	8,7	131693,0	gb|ELU03765.1|	No predicted function	*Capitella teleta*	3203	0,0E+00	X				X					
2235	fr4	9,2	124720,2	CDJ15074	filamin	*Hymenolepis microstoma*	1577	0,0E+00	X			X	X				X	X
2278	fr6	8,9	45041,7	XP_005182615.1	fructose-bisphosphate aldolaseisoform X1	*Musca domestica*	1094	6,6E-144				X	X		X		X	X
2292	fr2	9,8	35196,7	gb|EKC32468.1|	40S ribosomal protein S4, X isoform	*Crassostrea gigas*	824	2,0E-106				X	X		X		X	X
2308	fr6	9,3	101678,9	gb|EKC41795.1|	Copine-8	*Crassostrea gigas*	810	3,0E-93				X		X	X		X	X
2320	fr6	5,4	84070,0	gb|EFX77072.1|	No predicted function	*Daphnia pulex*	1689	0,0E+00	X				X					
2338	fr5	9,2	72388,5	ref|XP_001964418.1|	No predicted function	*Drosophila ananassae*	907	5,0E-112					X					
2365	fr2	4,6	98342,1	EFX89249	No predicted function	*Daphnia pulex*	1603	0,0E+00										X
2415	fr3	8,2	130078,5	ref|XP_972752.1|	transglutaminase	*Tribolium castaneum*	1047	1,0E-123	X									
2450	fr1	9,8	52772,6	gb|EFX70812.1|	No predicted function	*Daphnia pulex*	1344	3,0E-178				X					X	X
2459	fr1	7,7	69384,2	ref|XP_001847482.1|	utp-glucose-1-phosphate uridylyltransferase 2	*Culex quinquefasciatus*	1881	0,0E+00	X				X				X	X
2460	fr1	6,7	65864,0	emb|CAD54729.1|	beta-1,4-endoglucanase	*Mastotermes darwiniensis*	812	5,0E-98									X	
2491	fr4	9,0	82857,3	XP_004922008	myosin heavy chain, non-muscle-like	*Bombyx mori*	2446	0,0E+00										X
2493	fr4	10,2	40469,3	gb|AAR01298.1|	elongation factor-2	*Libinia emarginata*	975	3,0E-122	X			X	X		X		X	X
2523	fr5	7,1	43373,6	gb|EFX74120.1|	No predicted function	*Daphnia pulex*	950	1,0E-123				X						
2550	fr6	9,9	66870,5	ref|XP_002004439.1|	No predicted function	*Drosophila mojavensis*	904	3,0E-111				X						
2552	fr6		47992,4	gb|EFZ19648.1|	No predicted function	*Solenopsis invicta*	484	5,0E-51					X					
2558	fr4	9,5	73086,5	ref|XP_003962989.1|	dolichyl-diphosphooligosaccharide--protein glycosyltransferas	*Takifugu rubripes*	694	8,0E-80	X				X					
2562	fr2	9,5	80738,3	gb|AAX48011.1|	prophenoloxidase	*Macrobrachium rosenbergii*	1248	5,0E-161	X				X				X	X
2598	fr4	4,7	28769,5	prf||1405340A	protein 40kD	*Uncharacterized*	727	6,0E-92							X		X	X
2625	fr2	8,4	86315,0	XP_396106	probable V-type proton ATPase 116 kDa subunit a-like	*Apis mellifera*	535	3,7E-60										X
2645	fr4	9,1	38360,9	emb|CAD39002.1|	No predicted function	*Homo sapiens*	298	8,0E-29	X									
2663	fr3	9,2	54987,9	ref|NP_001090100.1|	5-aminoimidazole-4-carboxamide ribonucleotide formyltransfi	*Xenopus laevis*	988	1,0E-123				X		X				
2687	fr5	7,6	28467,2	ref|XP_970004.1|	NADH:ubiquinone oxidoreductase ESSS subunit	*Tribolium castaneum*	279	2,0E-27					X					
2701	fr4	9,9	79948,8	gb|EFX89528.1|	No predicted function	*Daphnia pulex*	443	9,0E-43					X					
2737	fr5	9,2	78394,9	AAL27405.1	chaperonin subunit 1	*Artemia franciscana*	710	1,3E-80							X			
2745	fr2	7,3	35692,0		No detected homolog							X		X				
2763	fr2	10,9	54809,2	gb|ABB55269.2|	peroxinectin	*Fenneropenaeus chinensis*	530	1,0E-55	X									
2775	fr2	9,0	127044,8	ref|XP_002430266.1|	Alanine aminotransferase,	*Pediculus humanus corporis*	1667	0,0E+00	X			X	X				X	X
2799	fr2	10,1	30687,8	XP_002116274	No predicted function	*Trichoplax adhaerens*	605	2,1E-75										X
2801	fr6	8,9	121470,9	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	2162	0,0E+00	X				X					
2839	fr5	8,5	121456,0	gb|EFX63600.1|	No predicted function	*Daphnia pulex*	1108	3,0E-137					X					X
2843	fr6	5,0	67471,1	gb|ADO00931.1|	calnexin	*Penaeus monodon*	1152	9,0E-147					X					
2853	fr3	9,3	58697,3	gb|AEG79567.1|	sarcoplasmic calcium-binding protein variant a	*Procambarus clarkii*	657	2,0E-79	X			X	X	X	X		X	X
2906	fr1	9,3	44335,1	ref|XP_001603962.1|	aldehyde oxidase 2-like	*Nasonia vitripennis*	323	6,0E-29									X	X
2916	fr1	8,2	98359,1	gb|EFX87607.1|	No predicted function	*Daphnia pulex*	2448	0,0E+00	X				X				X	X
2920	fr6	9,9	71441,9	gb|ABY64694.1|	masquerade-like protein, partial	*Armadillidium vulgare*	443	1,0E-47	X									
2997	fr4		43883,4	ref|XP_001657918.1|	Acetyl-CoA acetyltransferase, mitochondrial	*Aedes aegypti*	1166	2,0E-154									X	X
3010	fr5	8,6	80209,9	ref|XP_003700352.1|	F-actin-capping protein subunit alpha-like	*Megachile rotundata*	1088	3,0E-140				X	X					
3013	fr1	4,9	58231,2	gb|AGC54636.1|	heat shock protein 90	*Scylla paramamosain*	1528	0,0E+00				X	X					
3029	fr4	5,2	55069,6	ref|XP_001855895.1|	transaldolase	*Culex quinquefasciatus*	1240	4,0E-165	X			X	X					
3057	fr5	9,3	124223,8	gb|EFX88735.1|	No predicted function	*Daphnia pulex*	381	1,0E-34	X			X	X					
3082	fr3	10,2	36897,0	XP_004529974	extensin-3-like	*Ceratitis capitata*	131	5,6E-06										X
3093	fr3	9,1	40520,6		No detected homolog												X	X
3100	fr1	6,2	61705,7	ref|XP_003703000.1|	No predicted function	*Megachile rotundata*	2347	0,0E+00					X				X	
3110	fr2	10,9	24765,3	ABM55418.1	40S ribosomal protein S3	*Xenopsylla cheopis*	691	8,9E-88							X			X
3149	fr2	9,4	57449,9	gb|AAP59794.1|	slow muscle myosin S1 heavy chain	*Homarus americanus*	714	7,0E-84				X	X	X	X		X	X
3152	fr3	9,5	107008,9	XP_006030852	ectonucleotide pyrophosphatase/phosphodiesterase family m	*Alligator sinensis*	235	7,3E-17										X
3157	fr1	6,2	183945,6	gb|EFX80619.1|	No predicted function	*Daphnia pulex*	2694	0,0E+00	X				X		X		X	X
3203	fr6	5,7	70090,2	ref|XP_002101751.1|	Hex-A	*Drosophila yakuba*	1546	0,0E+00	X				X					
3212	fr4	8,9	93652,0	gb|EHJ71918.1|	No predicted function	*Danaus plexippus*	567	1,0E-64					X					
3241	fr5	9,6	85160,8	ref|XP_003487165.1|	trifunctional enzyme subunit alpha, mitochondrial-like	*Bombus impatiens*	1871	0,0E+00	X			X	X		X		X	X
3252	fr4	5,3	126990,8	ref|XP_001809722.1|	Muscle-specific protein 300 CG33715-PD	*Tribolium castaneum*	776	4,0E-81	X									
3319	fr5	9,3	82040,5	gb|ABW82155.1|	catalase	*Fenneropenaeus chinensis*	469	2,0E-47				X						
3334	fr4	5,9	56877,6	gb|AAC64661.1|	pacifastin light chain precursor	*Pacifastacus leniusculus*	462	2,0E-48					X					
3409	fr2	7,7	66665,5	dbj|BAM17968.1|	glycine hydroxylmethyltransferase	*Papilio xuthus*	1615	0,0E+00				X						
3452	fr6	4,5	25270,9	ref|NP_001037168.1|	ovarian serine protease	*Bombyx mori*	234	1,0E-18				X						
3461	fr4	9,3	75560,2	gb|EFX67149.1|	No predicted function	*Daphnia pulex*	1255	3,0E-165				X					X	
3473	fr6	9,1	62372,5	gb|AAG41119.1|AF115331_1	26S protease regulatory subunit	*Amblyomma americanum*	725	1,0E-88					X					
3503	fr4	9,8	80814,4	gb|EFX76980.1|	No predicted function	*Daphnia pulex*	541	3,0E-59					X					
3508	fr6	9,2	64938,3	EKC26290	No predicted function	*Crassostrea gigas*	268	3,9E-23										X
3514	fr3	9,5	108641,6	ref|XP_003486675.1|	lysosomal-trafficking regulator-like	*Bombus impatiens*	113	1,9E-02				X						
3535	fr3	7,2	34860,8	ref|XP_002072914.1|	No predicted function	*Drosophila willistoni*	1048	5,0E-136					X					
3601	fr6	8,4	38580,4		No detected homolog												X	X
3624	fr5	9,6	54073,9		No detected homolog							X	X	X	X			X
3645	fr6	6,3	124676,7	ref|XP_003707843.1|	glycogen debranching enzyme	*Megachile rotundata*	2797	0,0E+00					X					
3678	fr1	9,1	33740,1	ref|XP_001943087.1|	No predicted function	*Acyrthosiphon pisum*	147	8,0E-08				X	X				X	X
3706	fr1	10,2	83760,3	ADO51769.1	elongation factor 2	*Litopenaeus vannamei*	1351	7,0E-171					X		X			
3755	fr4	4,6	75520,1	XP_001658230	glutamyl aminopeptidase	*Aedes aegypti*	1519	0,0E+00										X
3764	fr3	9,8	95491,6	ref|XP_001603962.1|	aldehyde oxidase 2-like	*Nasonia vitripennis*	412	4,0E-38									X	X
3774	fr3	9,1	32209,5	gb|AET34917.1|	apoptosis inhibitor	*Macrobrachium rosenbergii*	572	9,0E-66				X	X		X		X	X
3799	fr6	9,3	66582,3	ref|NP_001017244.1|	apoptosis-inducing factor, mitochondrion-associated, 1	*Xenopus (Silurana) tropicalis*	1401	0,0E+00				X	X					
3833	fr1	8,9	84836,2	XP_001601449	aspartate aminotransferase, cytoplasmic-like	*Nasonia vitripennis*	998	3,8E-124										X
3846	fr2	5,3	66932,3	gb|AAK51547.1|	laminin beta chain	*Schistocerca gregaria*	1083	2,0E-130					X					X
3913	fr1	10,0	80215,2		No detected homolog							X		X				
3935	fr4	9,3	97880,4		No detected homolog							X						
3948	fr4	8,6	88471,3	dbj|BAH98114.1|	GTP binding protein alpha subunit Go	*Marsupenaeus japonicus*	1778	0,0E+00	X				X				X	
3967	fr2	9,8	77803,5	gb|ABG82044.1|	ALG-2 interacting protein x	*Penaeus monodon*	638	3,0E-68				X						
4015	fr6	9,5	36078,9	gb|ACD13589.1|	prohibitin 2	*Penaeus monodon*	1051	7,0E-140				X						
4020	fr4	9,5	117633,6	gb|ENN80698.1|	No predicted function	*Dendroctonus ponderosae*	1921	0,0E+00				X	X				X	X
4071	fr2	9,4	217976,7	gb|ADB28896.1|	No predicted function	*Macrobrachium nipponense*	2146	0,0E+00	X			X	X	X	X		X	X
4109	fr3	8,8	101902,5	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	2918	0,0E+00	X				X					
4123	fr3	9,1	101949,6	gb|EFX87974.1|	No predicted function	*Daphnia pulex*	1843	0,0E+00	X			X	X	X	X		X	X
4137	fr4	8,7	66161,4	ref|XP_002430488.1|	Hrp65 protein	*Pediculus humanus corporis*	455	4,0E-46				X	X					
4152	fr4	5,5	101348,5	gb|ADE60732.1|	heat shock protein 90	*Eriocheir sinensis*	2217	0,0E+00	X			X	X	X	X		X	X
4166	fr3	9,3	153932,4	EFX69050	No predicted function	*Daphnia pulex*	4031	0,0E+00										X
4212	fr2	6,5	91306,3	ref|XP_003493785.1|	dynamin-like isoform 1	*Bombus impatiens*	3412	0,0E+00	X								X	
4222	fr1	9,9	78488,0	gb|EFX75184.1|	No predicted function	*Daphnia pulex*	1276	3,0E-164									X	
4227	fr2	7,3	61098,8	ref|XP_001943129.2|	calcium-transporting ATPase sarcoplasmic/endoplasmic reticu	*Acyrthosiphon pisum*	1910	0,0E+00	X			X	X		X		X	X
4242	fr1	6,1	37345,1	BAL60587	cellulase	*Neomysis intermedia*	523	2,1E-58										X
4252	fr3	9,0	97824,4	YP_006416755	Acetyl-CoA acetyltransferase, mitochondrial	*Aequorivita sublithincola DSM 14*	113	2,9E-03										X
4277	fr6	6,4	35824,9	ref|YP_002333271.1|	nucleoprotein N	*Wongabel virus*	248	9,0E-21									X	
4287	fr3	5,7	46692,3	XP_003689685	xaa-Pro aminopeptidase 1-like	*Apis florea*	633	4,5E-71										X
4315	fr5	8,9	31780,1	AGU01545	antimicrobial peptide type 2 precursor IIc	*Pandalopsis japonica*	285	8,4E-28	X			X	X				X	X
4316	fr5	6,7	56335,0	gb|EFX83182.1|	No predicted function	*Daphnia pulex*	1156	3,0E-149	X									
4384	fr5	9,3	86838,7	AEK49107.1	Tudor staphylococcal nuclease	*Penaeus monodon*	2412	0,0E+00							X			
4401	fr4	9,6	51068,6	gb|ENN78676.1|	No predicted function	*Dendroctonus ponderosae*	1169	4,0E-151				X			X		X	X
4456	fr5	9,4	210637,3	gb|EFX88672.1|	No predicted function	*Daphnia pulex*	2024	0,0E+00	X				X				X	X
4531	fr5	4,2	53492,3	ref|XP_002670944.1|	No predicted function	*Naegleria gruberi*	130	7,0E-05									X	
4553	fr3	9,6	76028,6		No detected homolog										X			
4565	fr1	9,0	80261,1	gb|AEK11985.1|	calmin-like protein	*Tigriopus californicus*	245	5,0E-21	X			X	X	X	X		X	X
4581	fr2	8,7	71093,6	ref|XP_971164.2|	widerborst CG5643-PC	*Tribolium castaneum*	2127	0,0E+00	X				X					
4614	fr3	6,7	89901,9	ref|XP_557743.2|	No predicted function	*Anopheles gambiae str. PEST*	2292	0,0E+00				X						
4631	fr3	9,1	53955,7	CAA56703	beta-1,3-D-glucan binding protein	*Pacifastacus leniusculus*	274	2,0E-22										X
4635	fr5	8,7	91259,6	ref|XP_002423552.1|	No predicted function	*Pediculus humanus corporis*	1270	7,0E-148	X			X	X	X			X	X
4646	fr4	9,9	33049,4	CAX65684	beta-1,3-glucan-binding protein	*Pontastacus leptodactylus*	400	9,9E-40										X
4647	fr6	8,1	193182,8	gb|EFN78084.1|	Ankyrin-2	*Harpegnathos saltator*	2575	0,0E+00	X				X		X		X	X
4674	fr2	10,6	41262,9	XP_318679.4	No predicted function	*Anopheles gambiae str. PEST*	242	3,8E-19							X			
4696	fr1	9,3	138971,5	gb|EFN89532.1|	Myosin heavy chain, non-muscle	*Harpegnathos saltator*	2089	0,0E+00				X	X	X	X		X	X
4728	fr4	9,6	112969,0	ref|XP_002603203.1|	No predicted function	*Branchiostoma floridae*	1730	0,0E+00	X									
4729	fr4	9,3	113660,7	ENN74946.1	No predicted function	*Dendroctonus ponderosae*	832	1,7E-97							X			
4737	fr3	8,2	37012,8	gb|ABY66598.1|	pyruvate kinase 3	*Litopenaeus vannamei*	1144	2,0E-149				X	X	X			X	X
4741	fr1	8,8	143320,0	gb|AEC48728.1|	alpha-tubulin	*Eriocheir sinensis*	2115	0,0E+00									X	
4746	fr4	10,0	57380,3		No detected homolog							X	X		X		X	X
4753	fr3	9,3	88567,8	ref|XP_002814519.1|	D-beta-hydroxybutyrate dehydrogenase, mitochondrial-like	*Pongo abelii*	609	1,0E-69					X					
4757	fr2	10,4	43250,0	gb|AAR19062.1|	alpha-amylase	*Litopenaeus vannamei*	306	2,0E-29									X	X
4816	fr4	10,1	119332,2		No detected homolog							X	X					
4825	fr6	7,1	86258,5	ref|XP_002165626.2|	No predicted function	*Hydra magnipapillata*	217	4,0E-15									X	
4833	fr1	9,2	37557,0	gb|EHJ67850.1|	No predicted function	*Danaus plexippus*	437	3,0E-49	X									
4834	fr6	5,3	45630,8	gb|ACT88125.1|	No predicted function	*Drosophila melanogaster*	720	6,0E-91	X									
4872	fr6	8,3	119491,5	EFR27785	No predicted function	*Anopheles darlingi*	695	9,2E-72										X
4879	fr4	9,6	129550,8	gb|EFX83224.1|	No predicted function	*Daphnia pulex*	2695	0,0E+00	X			X	X				X	X
4935	fr3	5,1	130960,8	gb|EFX76950.1|	No predicted function	*Daphnia pulex*	3002	0,0E+00							X		X	X
5003	fr4	11,4	49019,8	AAD16454.1	clotting protein precursor	*Pacifastacus leniusculus*	111	1,4E-02						X				
5059	fr2	8,6	103602,1	ref|XP_002430367.1|	Flotillin-1	*Pediculus humanus corporis*	1782	0,0E+00					X					
5094	fr6	9,7	62946,6	ref|XP_004054746.1|	propionyl-CoA carboxylase alpha chain, mitochondrial isoform	*Gorilla gorilla gorilla*	1058	1,0E-131									X	X
5115	fr4	8,6	101110,9	gb|EMC82662.1|	Thimet oligopeptidase, partial	*Columba livia*	1689	0,0E+00					X					
5137	fr4	4,7	37086,4		No detected homolog							X	X					

5150	fr4	10,9	123397,3	ref|XP_003696500.1|	papilin-like	*Apis florea*	1690	0,0E+00				X				X	
5155	fr2	8,8	67361,7	gb|ABO33174.1|	serine protease-like protein 3	*Penaeus monodon*	750	2,0E-89			X						
5204	fr4	6,2	79194,5	gb|AET34915.1|	heat shock protein 21	*Macrobrachium rosenbergii*	235	3,0E-18			X						
5269	fr4	9,3	30571,6	sp|O65353.1|RL5_HELAN	60S ribosomal protein L5	*Helianthus annuus*	759	1,0E-96				X					X
5310	fr1	5,7	68528,5	gb|AEC50079.1|	calreticulin	*Pacifastacus leniusculus*	1037	5,0E-132						X		X	X
5366	fr5	9,6	52964,0	XP_004950477	neuroblast differentiation-associated protein AHNAK-like, part	*Gallus gallus*	221	6,8E-16	X			X					X
5392	fr6	9,0	19892,0	gb|AAS88999.1|	Actin	*Sitobion avenae*	615	1,0E-78				X				X	X
5407	fr1	9,6	32824,9	ref|XP_004365566.1|	peptidylprolyl isomerase	*Capsaspora owczarzaki ATCC 308*	759	3,0E-98					X	X		X	X
5492	fr3	10,3	114102,2	ref|XP_002423351.1|	Filamin-C	*Pediculus humanus corporis*	862	8,0E-97	X		X	X	X	X		X	X
5557	fr4	8,6	82534,6	XP_005175755.1	protein arginine N-methyltransferase 1-A	*Musca domestica*	862	4,8E-105						X			
5580	fr6	9,8	84444,8	gb|EFX81838.1|	No predicted function	*Daphnia pulex*	728	1,0E-83			X						
5589	fr4	8,6	82234,1	ERL85336	No predicted function	*Dendroctonus ponderosae*	170	3,0E-09				X				X	X
5631	fr5	9,5	93422,7	EFZ09701	No predicted function	*Solenopsis invicta*	497	7,9E-49									X
5739	fr5	6,2	63163,7	ref|XP_972494.1|	fascin	*Tribolium castaneum*	1750	0,0E+00	X								
5753	fr2	6,6	46255,6	gb|EFN61404.1|	Muscle M-line assembly protein unc-89	*Camponotus floridanus*	1110	3,0E-132	X			X					
5763	fr2	4,7	122538,8	gb|EFX77181.1|	No predicted function	*Daphnia pulex*	2893	0,0E+00	X			X					X
5764	fr6	6,8	73561,8	gb|EFZ20618.1|	No predicted function	*Solenopsis invicta*	1916	0,0E+00	X			X					
5777	fr5	10,1	99700,5	gb|ABX76969.1|	Vasa	*Parhyale hawaiensis*	1640	0,0E+00								X	
5839	fr3	9,9	53152,7	dbj|BAG30740.1|	muscle myosin heavy chain	*Papilio xuthus*	272	2,0E-22				X		X		X	X
5861	fr5	4,8	60753,7	gb|EFA06228.1|	No predicted function	*Tribolium castaneum*	1223	3,0E-145	X								
5872	fr6	9,1	58052,5	gb|AFD61666.1|	cytosolic manganese superoxide dismutase	*Cherax quadricarinatus*	885	3,0E-112	X		X	X					
5884	fr2	5,2	35643,1	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	501	2,7E-53									X
5902	fr1	6,8	55584,8	ref|XP_972994.2|	sodium bicarbonate cotransporter	*Tribolium castaneum*	1068	1,0E-130	X								
5917	fr5	10,6	104816,1	ABU41117	metalloproteinase	*Lepeophtheirus salmonis*	125	1,3E-04									X
5953	fr4	9,6	59149,7		No detected homolog				X								
5991	fr4	11,9	34749,4	ETE60281	60S ribosomal protein L7a	*Ophiophagus hannah*	470	7,2E-53								X	X
6005	fr3	4,9	46390,8	ref|XP_001812981.1|	ADP-ribosyltransferase 5	*Tribolium castaneum*	135	4,0E-06				X					
6039	fr2	10,5	51666,3		No detected homolog									X			
6287	fr4	6,0	61958,8	gb|EFX87974.1|	No predicted function	*Daphnia pulex*	1894	0,0E+00	X		X	X	X	X		X	X
6306	fr2	9,1	68711,7	ref|XP_001654511.1|	aldehyde oxidase	*Aedes aegypti*	938	2,0E-108	X			X				X	X
6350	fr2	4,8	90413,2	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	1801	0,0E+00									X
6358	fr4	8,6	46492,3	AEC22817	No predicted function	*Macrobrachium nipponense*	159	1,9E-08									X
6366	fr2	7,0	65405,1	ref|YP_002333271.1|	nucleoprotein N	*Wongabel virus*	488	2,0E-51								X	
6394	fr3	7,5	35132,4	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	737	1,0E-90	X		X	X				X	
6432	fr4	9,6	59984,0	XP_002598196	No predicted function	*Branchiostoma floridae*	314	7,8E-28									X
6435	fr2	8,6	39818,2	gb|ABR21044.1|	venom insulin-like growth factor binding protein-1	*Mesobuthus eupeus*	156	1,0E-09			X	X					
6440	fr2	9,8	80767,8	ref|XP_001601998.1|	partner of Y14 and mago-like	*Nasonia vitripennis*	250	1,0E-20				X					
6444	fr2	5,5	87246,6	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	3465	0,0E+00	X			X				X	X
6495	fr1	8,8	69538,1	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	1944	0,0E+00	X			X				X	X
6538	fr6	10,5	14619,9	gb|ADV40146.1|	ADP/ATP translocase	*Latrodectus hesperus*	466	6,0E-55			X	X	X	X		X	X
6689	fr5	9,8	104353,9	gb|AAT77811.1|	calpain B	*Gecarcinus lateralis*	824	4,0E-93	X			X				X	X
6749	fr5	8,4	48036,5	ref|XP_003700392.1|	No predicted function	*Megachile rotundata*	950	9,0E-111	X			X					
6766	fr3	8,7	43207,4		No detected homolog				X		X	X				X	X
6807	fr2	6,6	90131,3	XP_005023902.1	lysosomal-trafficking regulator isoform X1	*Anas platyrhynchos*	128	4,0E-04			X		X				
6815	fr6	9,3	62757,4	ref|XP_974030.1|	Annexin IX CG5730-PC	*Tribolium castaneum*	573	7,0E-65			X						
6885	fr5	10,1	118384,1	ref|NP_001098066.2|	aubergine protein	*Bombyx mori*	480	2,0E-46			X	X					X
6933	fr5	10,5	99772,3	sp|Q9U572.1|CLOT_PENMO	Hemolymph clottable protein	*Penaeus monodon*	146	2,0E-06			X		X	X			
6945	fr6	7,7	100235,3	gb|AFD61667.1|	prophenoloxidase	*Cherax quadricarinatus*	761	2,0E-85						X		X	X
7061	fr2	6,2	29114,5	ref|XP_002432927.1|	No predicted function	*Pediculus humanus corporis*	157	3,0E-09			X		X	X		X	X
7089	fr4	5,9	35061,5	gb|AET34915.1|	heat shock protein 21	*Macrobrachium rosenbergii*	185	1,0E-12	X								
7108	fr3	8,1	31720,9	gb|ACI66366.1|	Nattectin precursor	*Salmo salar*	111	1,0E-03				X					
7151	fr1	5,5	50489,8	gb|ELU06714.1|	No predicted function	*Capitella teleta*	1666	0,0E+00	X								
7169	fr4	5,2	58816,6	gb|AAK69205.1|AF336805_1	transglutaminase	*Pacifastacus leniusculus*	1644	0,0E+00								X	X
7221	fr1	4,8	24542,8	ref|NP_668873.1|	No predicted function	*Yersinia pestis KIM10+*	108	2,0E-03				X					
7367	fr6	9,9	82824,6	EOB00010	Alpha-mannosidase 2C1, partial	*Anas platyrhynchos*	475	2,0E-46									X
7403	fr5	9,5	122933,5	gb|EFZ17862.1|	No predicted function	*Solenopsis invicta*	954	9,0E-111				X					
7432	fr3	8,6	49606,1		No detected homolog									X		X	X
7525	fr6	7,0	60323,7	gb|EFX69841.1|	No predicted function	*Daphnia pulex*	1264	2,0E-167	X			X					
7587	fr4	9,0	45284,4	dbj|BAD17952.1|	phosphoglycerate kinase	*Lethenteron reissneri*	504	1,0E-55	X			X					
7625	fr2	8,8	66803,0	ref|XP_003703974.1|	3-oxoacyl-acyl-carrier-protein synthase	*Megachile rotundata*	1324	1,0E-174				X	X			X	X
7714	fr1	7,4	56297,0	ref|NP_001039250.1|	long-chain acyl-CoA dehydrogenase	*Xenopus (Silurana) tropicalis*	1585	0,0E+00	X			X					
7767	fr5	6,7	91602,1		No detected homolog				X								
7880	fr2	9,9	83026,0		No detected homolog									X			
7908	fr4	8,1	54876,2	dbj|BAM18606.1|	hemolectin	*Papilio xuthus*	129	2,0E-05								X	X
7942	fr6	9,1	86949,0	ref|XP_309663.5|	No predicted function	*Anopheles gambiae str. PEST*	546	5,0E-57			X	X	X	X		X	X
7953	fr6	7,3	72095,5	gb|EFX88799.1|	No predicted function	*Daphnia pulex*	2007	0,0E+00				X				X	X
7983	fr6	10,0	135800,4	ref|XP_002427775.1|	myosin-9	*Pediculus humanus corporis*	952	2,0E-103	X		X	X				X	X
8013	fr1	8,3	57789,6	XP_001846253.1	bicoid stability factor	*Culex quinquefasciatus*	893	2,8E-103						X			
8055	fr2	10,5	46740,9	ref|XP_002116186.1|	No predicted function	*Trichoplax adhaerens*	591	1,0E-70			X						
8064	fr3	10,0	89122,2	gb|AFN42969.1|	glucose-6-phosphate isomerase	*Tigriopus californicus*	1070	3,0E-132	X							X	X
8112	fr6	9,3	128224,8	NP_001022662	Protein EMB-9, isoform a	*Caenorhabditis elegans*	1017	1,2E-111									X
8132	fr6	4,7	64550,1	XP_003693732	laminin subunit alpha-like	*Apis florea*	686	1,4E-73									X
8186	fr2	5,1	78186,3	emb|CBY24127.1|	No predicted function	*Oikopleura dioica*	227	5,0E-16				X				X	X
8190	fr1	9,7	57596,4	gb|ACH69158.1|	myosin heavy chain	*Bombyx mori*	994	2,0E-124					X	X		X	X
8288	fr5	10,2	40245,5	ref|XP_002424248.1|	Troponin T	*Pediculus humanus corporis*	657	2,0E-79								X	X
8483	fr6	8,5	48670,5	gb|EFX74525.1|	No predicted function	*Daphnia pulex*	746	1,0E-93			X	X				X	X
8504	fr2	9,7	63559,4	emb|CAX83018.1|	Stress protein DDR48 (DNA damage-responsive protein 48)	*Schistosoma japonicum*	130	6,0E-05	X								
8708	fr1	9,3	77726,4	gb|ADY90107.1|	protein disulfide isomerase A6	*Penaeus monodon*	748	9,0E-88	X			X				X	X
8722	fr3	9,4	87061,6	gb|EFX79712.1|	No predicted function	*Daphnia pulex*	1208	4,0E-153			X	X	X			X	X
8832	fr2	8,9	94804,5	gb|ABY66597.1|	pyruvate kinase 2	*Litopenaeus vannamei*	1456	0,0E+00			X	X				X	X
8873	fr3	9,1	59709,6	gb|AFW99837.1|	troponin I	*Litopenaeus vannamei*	192	2,0E-13	X							X	
8888	fr3	6,7	50177,5	ref|XP_002734509.1|	aldehyde dehydrogenase 2-like	*Saccoglossus kowalevskii*	1109	2,0E-144	X								
8924	fr2	9,5	43521,8	gb|ADF81052.1|	vitellogenin-like protein precursor	*Gammarus fossarum*	747	2,0E-94			X		X	X			
9002	fr1	5,5	66376,1	ref|XP_002609252.1|	No predicted function	*Branchiostoma floridae*	1455	0,0E+00					X	X		X	X
9053	fr2	5,8	30487,3	gb|ABZ80828.1|	peroxiredoxin	*Penaeus monodon*	1051	1,0E-141				X					
9227	fr3	5,7	54136,8	ref|XP_003692078.1|	protein hu-li tai shao-like	*Apis florea*	1190	1,0E-152				X					
9264	fr5	7,1	85114,9	ref|XP_003706217.1|	tubulin beta-1 chain-like	*Megachile rotundata*	1433	0,0E+00								X	X
9333	fr6	9,8	65054,5	AEE36486.1	protein disulfide isomerase 2	*Fenneropenaeus chinensis*	605	3,3E-67								X	
9416	fr2	10,1	73839,0		No detected homolog						X		X	X			
9438	fr2	9,9	118622,0	gb|ADQ64493.1|	No predicted function	*Bactrocera oleae*	967	1,0E-118	X		X	X		X		X	X
9612	fr5	8,8	69318,2		No detected homolog						X		X				
9667	fr6	7,8	44585,6	emb|CAE47485.1|	beta-1,3-glucan binding protein	*Homarus gammarus*	1022	2,0E-133	X		X	X				X	X
9730	fr4	9,5	21002,0	gb|ABK59925.1|	clottable protein	*Marsupenaeus japonicus*	122	4,0E-05								X	X
9731	fr2	9,6	38104,9	gb|AFD28290.1|	14-3-3-like protein	*Eriocheir sinensis*	654	1,0E-80			X			X		X	
9788	fr4	9,8	27504,1	gb|ABU41135.1|	vitellogenin 2	*Lepeophtheirus salmonis*	106	1,3E-02			X		X	X			
9798	fr4	10,3	92219,1	ref|XP_002423394.1|	Apoptosis inhibitor	*Pediculus humanus corporis*	165	6,0E-09				X					
9819	fr1	10,4	33056,6	AET34917.1	apoptosis inhibitor	*Macrobrachium rosenbergii*	325	9,4E-31						X			
9962	fr1	9,1	76354,9	gb|ACM66845.1|	nuclear autoantigenic sperm protein	*Penaeus monodon*	309	2,0E-26			X						
10004	fr1	8,8	12955,6	gb|ACX46120.1|	catalase	*Scylla paramamosain*	255	2,0E-23								X	
10101	fr5	7,0	62351,0	gb|EFN65250.1|	Nesprin-1	*Camponotus floridanus*	221	2,0E-15	X			X				X	X
10107	fr2	8,9	22769,0	gb|ABB29594.1|	ribosomal protein P0 large subunit	*Platynereis dumerilii*	524	1,0E-62	X		X	X					
10109	fr4	9,9	57081,9	ref|ZP_08782073.1|	Glucosylceramidase	*Methylobacter tundripaludum SV*	427	2,0E-43								X	X
10148	fr4	9,3	77419,1	gb|ACY66440.1|	myosin regulatory light chain 2 smooth muscle	*Scylla paramamosain*	621	7,0E-73			X	X					
10149	fr3	8,8	43490,5	gb|EFX64695.1|	No predicted function	*Daphnia pulex*	383	3,0E-38			X		X	X		X	X
10171	fr5	5,8	44429,6	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	1544	0,0E+00	X		X	X	X	X		X	X
10189	fr3	9,1	34049,8	gb|EFR27027.1|	No predicted function	*Anopheles darlingi*	351	3,0E-34				X				X	X
10262	fr1	6,3	75121,3	gb|ELT97789.1|	No predicted function	*Capitella teleta*	232	2,0E-18				X				X	X
10319	fr3	8,9	47814,4	ref|XP_003974830.1|	heterogeneous nuclear ribonucleoprotein K-like	*Takifugu rubripes*	611	2,0E-70			X						
10389	fr3		21796,4	gb|AEG80154.1|	Tcp-1-beta	*Litopenaeus vannamei*	665	7,0E-81						X		X	
10397	fr1	4,9	57520,0	gb|AGC54636.1|	heat shock protein 90	*Scylla paramamosain*	1468	0,0E+00	X		X	X	X	X		X	X
10406	fr2	11,2	34967,5	gb|ABO33174.1|	serine protease-like protein 3	*Penaeus monodon*	367	6,0E-37	X		X	X					
10413	fr5	10,3	77229,0	ref|XP_001648302.1|	glucose dehydrogenase	*Aedes aegypti*	728	9,0E-83			X						
10436	fr1	10,5	103991,7	dbj|BAE42637.1|	No predicted function	*Mus musculus*	1719	0,0E+00	X		X	X					
10466	fr4	8,3	58238,3	NP_001267545	calcium activated chloride channel precursor	*Xenopus laevis*	252	2,0E-19									X
10498	fr6	9,2	48071,1	ref|XP_002432182.1|	Heterogeneous nuclear ribonucleoprotein A1,	*Pediculus humanus corporis*	871	9,0E-109			X	X		X			
10515	fr1	9,8	56397,3	ref|XP_002423394.1|	Apoptosis inhibitor	*Pediculus humanus corporis*	198	3,0E-13						X		X	X
10580	fr5	5,3	34153,3	gb|ABG74710.1|	tubulin beta-1 chain, partial	*Diaphorina citri*	1156	2,0E-156					X	X		X	
10582	fr1	10,2	32633,4	gb|AEI88118.1|	calsequestrin-1 precursor	*Scylla paramamosain*	350	2,0E-37	X								X
10587	fr4	9,2	22021,6	gb|ACY66555.1|	ribosomal protein L19e	*Scylla paramamosain*	205	1,0E-17			X			X		X	
10596	fr1	8,8	49083,8	ref|XP_001607760.1|	GTP-binding protein CG1354-like	*Nasonia vitripennis*	1179	3,0E-156				X				X	
10601	fr5	9,4	23514,1		No detected homolog									X			
10630	fr3	10,1	13430,3	ref|XP_002492949.1|	ubiquitin-40S ribosomal protein S31 fusion protein	*Komagataella pastoris GS115*	377	2,0E-43								X	X
10635	fr3	9,0	59507,8	ref|XP_003692078.1|	protein hu-li tai shao-like	*Apis florea*	981	7,0E-121			X		X	X		X	X
10648	fr2	11,1	17248,3	ref|XP_003241023.1|	histone H2A-like	*Acyrthosiphon pisum*	512	2,0E-63				X					
10740	fr1	8,7	37198,7	XP_004534773.1	cofilin/actin-depolymerizing factor homolog	*Ceratitis capitata*	341	2,6E-35	X			X		X		X	X
10743	fr4	8,1	13745,3	gb|ACY66388.1|	chaperonin 10	*Scylla paramamosain*	271	4,0E-28			X	X					
10750	fr4	5,7	44004,2		No detected homolog						X						
10761	fr5	9,1	37615,1	XP_003705965.1	proclotting enzyme	*Megachile rotundata*	182	6,5E-12					X				
10767	fr1	9,6	26012,8	pdb|2HF5|A	No predicted function	*Uncharacterized*	346	3,0E-38	X								
10775	fr2	6,5	17271,3	gb|ADE58569.1|	GH5 family protein GH5E	*Limnoria quadripunctata*	373	6,0E-40								X	

10825	fr5	6,3	57478,9	ref|NP_001262104.1|	eukaryotic release factor 1, isoform H	*Drosophila melanogaster*	1553	0,0E+00	X								
10848	fr6	10,0	44723,2	ref|XP_002585525.1|	No predicted function	*Branchiostoma floridae*	757	2,0E-93				X					
10884	fr5	8,1	35411,5	ref|XP_318716.4|	No predicted function	*Anopheles gambiae str. PEST*	193	9,0E-14			X						
10905	fr1	4,8	73523,0	ELU16365	No predicted function	*Capitella teleta*	882	3,2E-108									X
10958	fr4	9,4	31514,2	ref|XP_003396556.1|	alpha-L-fucosidase-like	*Bombus terrestris*	508	6,0E-57			X					X	
11145	fr5	6,1	22813,8	XP_971239	Hemolectin	*Tribolium castaneum*	147	7,7E-08									X
11178	fr5	9,0	118453,2	ref|XP_971500.1|	midline fasciclin	*Tribolium castaneum*	507	5,0E-50								X	
11190	fr1	5,9	25907,7	XP_001608247	myosin regulatory light chain 2-like isoform 1	*Nasonia vitripennis*	367	3,4E-39									X
11199	fr3	8,4	18920,3	gb|EFX79782.1|	No predicted function	*Daphnia pulex*	110	1,0E-03			X					X	
11259	fr1	8,5	68253,9	gb|EKC36313.1|	S-formylglutathione hydrolase	*Crassostrea gigas*	912	4,0E-115			X	X					
11310	fr4	9,2	113336,7	gb|ABG67961.1|	eukaryotic initiation factor 4A	*Callinectes sapidus*	1024	3,0E-125			X	X	X	X	X	X	
11425	fr6	9,9	18356,2	ADD52189.1	ribosomal protein L8	*Pandalus platyceros*	196	2,4E-16						X			
11505	fr1	9,6	121816,2	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	1904	0,0E+00				X				X	X
11506	fr3	9,4	120467,8	BAC66140	projectin	*Procambarus clarkii*	1049	6,7E-116									X
11557	fr1	6,0	19548,7	ref|XP_001608251.2|	No predicted function	*Nasonia vitripennis*	314	2,0E-31			X	X					
11585	fr5	10,7	43930,5	ref|NP_001011291.1|	fucosidase, alpha-L- 2, plasma precursor	*Xenopus (Silurana) tropicalis*	468	8,0E-50	X								
11594	fr5	9,0	12630,5	gb|AEB54622.1|	Actin	*Procambarus clarkii*	400	1,0E-47				X					
11635	fr5	5,5	39009,9	ref|XP_002026190.1|	No predicted function	*Drosophila persimilis*	978	2,0E-123			X	X					
11648	fr5	9,2	23896,7	gb|ABQ96198.1|	crustin isoform 2	*Litopenaeus schmitti*	175	1,0E-13			X	X					
11651	fr4	9,6	19723,0		No detected homolog				X					X			
11694	fr6	6,8	19224,4	gb|ADM64456.1|	fatty acid binding protein	*Eriocheir sinensis*	409	1,0E-47	X		X	X				X	X
11700	fr6	9,4	30769,8	XP_005178098.1	V-type proton ATPase subunit E	*Musca domestica*	330	2,5E-33						X		X	
11705	fr4	6,3	55910,1	gb|ADO00927.1|	calreticulin	*Penaeus monodon*	1095	2,0E-142			X	X					X
11710	fr5	6,7	62249,7	ref|XP_002409147.1|	beta-glucocerebrosidase	*Ixodes scapularis*	632	1,0E-71								X	X
11717	fr4	7,6	44587,7	gb|ABC59662.1|	beta tubulin	*Hericium erinaceum*	1109	3,0E-146					X	X		X	X
11723	fr3	8,9	39198,1	dbj|BAE42496.1|	No predicted function	*Mus musculus*	1081	1,0E-143			X	X	X	X		X	
11772	fr1	5,8	12145,0	gb|ACR78689.1|	cuticle protein	*Rimicaris exoculata*	153	9,0E-11								X	
11839	fr4	10,1	36193,4	AEZ68611.1	adenine nucleotide translocase	*Litopenaeus vannamei*	296	1,0E-27			X		X	X		X	
11860	fr1	8,3	63695,9	gb|ENN75033.1|	No predicted function	*Dendroctonus ponderosae*	740	1,0E-88				X					
11975	fr6	9,4	29234,1	ref|XP_001376605.2|	epoxide hydrolase 1-like	*Monodelphis domestica*	522	1,0E-59								X	
11988	fr3	6,0	80747,6	ref|XP_392215.4|	No predicted function	*Apis mellifera*	890	5,0E-99	X		X	X					
12002	fr2	10,5	63978,3	ref|XP_971371.1|	No predicted function	*Tribolium castaneum*	527	1,0E-56			X		X	X			
12042	fr4	9,9	38467,3	ref|XP_001948247.2|	dihydrolipoyl dehydrogenase, mitochondrial-like	*Acyrthosiphon pisum*	1064	2,0E-138								X	X
12109	fr2	9,5	39443,9	ref|XP_976197.1|	Bmsqd-2 isoform 3	*Tribolium castaneum*	484	4,0E-54				X					
12127	fr1	4,3	49264,8	gb|EFX84260.1|	No predicted function	*Daphnia pulex*	864	8,0E-107			X	X					
12147	fr6	6,7	44972,2	ACS36538.1	elongation factor 2	*Homarus americanus*	1238	2,6E-159						X			
12172	fr5	9,6	29271,2		No detected homolog						X		X				
12316	fr3	5,3	35647,0	gb|AEC50083.1|	alpha-2-macroglobulin 2 isoform 3	*Pacifastacus leniusculus*	363	2,0E-34						X		X	X
12339	fr3	9,0	63991,3	gb|ENN71944.1|	No predicted function	*Dendroctonus ponderosae*	288	6,0E-24				X					
12395	fr2	9,0	19152,8	gb|AFD61666.1|	cytosolic manganese superoxide dismutase	*Cherax quadricarinatus*	441	1,0E-50			X						
12402	fr1	7,7	47401,6	gb|ACY66376.1|	HSP90	*Scylla paramamosain*	984	4,0E-126					X	X		X	X
12415	fr3	6,0	50795,6	gb|ADG22164.1|	chitinase 2	*Penaeus monodon*	1168	6,0E-154	X								
12473	fr4	9,9	66599,7		No detected homolog						X	X	X	X			
12497	fr5	9,7	105243,2	gb|EFR19083.1|	No predicted function	*Anopheles darlingi*	127	5,0E-04				X					X
12555	fr1	8,5	43582,5	gb|ABM92447.1|	glucose-regulated protein 78	*Fenneropenaeus chinensis*	758	2,0E-90					X	X		X	
12562	fr1	10,1	79316,8	gb|ACY66498.1|	T-complex protein 1 subunit alpha (TCP-1-alpha) (CCT-alpha) is	*Scylla paramamosain*	1011	5,0E-129			X	X					
12571	fr5	8,9	58633,5	XP_004924796	four and a half LIM domains protein 2-like isoform X3	*Bombyx mori*	1334	8,8E-179	X							X	X
12596	fr4	5,3	23070,2	ref|NP_001014354.1|	lysosomal alpha-mannosidase precursor	*Danio rerio*	534	1,0E-59			X	X				X	
12612	fr4	9,9	18835,9	gb|ADD70028.1|	myosin light chain	*Marsupenaeus japonicus*	280	2,0E-28	X		X	X					X
12649	fr1	9,0	46585,2	gb|EAW79627.1|	ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A, isofo	*Homo sapiens*	1416	0,0E+00						X		X	X
12679	fr4	9,2	31680,9	emb|CAD91421.1|	ribosomal protein L5	*Crassostrea gigas*	898	6,0E-118				X					
12681	fr4	11,0	45957,2	ref|XP_003424098.1|	myosin heavy chain, muscle isoform 3	*Nasonia vitripennis*	577	1,0E-61	X		X	X				X	X
12683	fr4	8,8	39958,8	ref|XP_973414.1|	No predicted function	*Tribolium castaneum*	1039	3,0E-130			X	X					
12812	fr2	9,3	41224,2	EFN88457	Myosin heavy chain, muscle	*Harpegnathos saltator*	497	3,2E-51									X
12830	fr2		48220,0	ref|XP_001238828.3|	No predicted function	*Anopheles gambiae str. PEST*	904	3,0E-117								X	
12849	fr2	7,2	73832,7	gb|AAT01139.1|	vitellogenin	*Metapenaeus ensis*	124	5,0E-04			X		X	X			
12885	fr4	9,9	44125,3	gb|ABX89617.1|	vitellogenin	*Callinectes sapidus*	231	2,0E-18			X		X	X			
12901	fr5	5,5	42809,7	gb|EFX67104.1|	No predicted function	*Daphnia pulex*	451	6,0E-46								X	X
12912	fr4	9,4	105393,2	gb|EFX68668.1|	No predicted function	*Daphnia pulex*	846	2,0E-98				X				X	
12917	fr4	9,4	35428,9		No detected homolog								X				
12941	fr2	9,7	93949,0	emb|CBN81210.1|	Aconitate hydratase	*Dicentrarchus labrax*	1439	0,0E+00						X		X	X
12995	fr2	9,8	78363,4	gb|EFX80396.1|	No predicted function	*Daphnia pulex*	350	3,0E-31	X								
12996	fr4	9,3	44533,5	ref|XP_002430232.1|	muscle lim protein	*Pediculus humanus corporis*	573	1,0E-64	X			X					
13018	fr6	9,6	58137,4	gb|ELT99118.1|	No predicted function	*Capitella teleta*	349	2,0E-33				X					
13026	fr5	9,7	36538,6	gb|EFX72553.1|	No predicted function	*Daphnia pulex*	774	2,0E-96	X			X				X	
13039	fr2	8,6	57697,7	ref|XP_001660597.1|	protein disulfide-isomerase A6 precursor	*Aedes aegypti*	792	3,0E-96								X	X
13076	fr3	5,1	20011,6	ref|XP_003793694.1|	tubulin alpha chain-like	*Otolemur garnettii*	613	3,0E-78				X					
13125	fr5	5,0	18725,3	ref|XP_395092.2|	No predicted function	*Apis mellifera*	101	5,0E-03								X	
13141	fr2	6,9	19627,0	gb|AFN42990.1|	glucose-6-phosphate isomerase	*Tigriopus californicus*	566	2,0E-66								X	X
13142	fr5	10,1	81127,0	gb|AAR00498.1|	Hsp90-like protein	*Sphoeroides annulatus*	619	2,0E-72			X		X	X		X	X
13154	fr3	8,5	100349,6	gb|ABM55628.1|	No predicted function	*Maconellicoccus hirsutus*	426	4,0E-44	X		X	X					
13177	fr2	9,5	24878,4	ref|XP_001954532.1|	No predicted function	*Drosophila ananassae*	101	4,2E-02	X								
13187	fr6	6,2	72637,9	gb|EFX89582.1|	No predicted function	*Daphnia pulex*	226	3,0E-16				X				X	X
13242	fr1	9,4	37455,7	dbj|BAH98115.1|	GTP binding protein alpha subunit Gq	*Marsupenaeus japonicus*	954	2,0E-124				X					
13250	fr6	10,5	15546,7	sp|Q6PV61.3|H2A_PENVA	Histone H2A	*Uncharacterized*	622	2,0E-80					X	X		X	X
13269	fr1	5,4	60883,3	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	1507	0,0E+00	X			X					
13503	fr3	9,8	59568,4	EFX84286	No predicted function	*Daphnia pulex*	193	3,0E-13									X
13767	fr6	9,1	67712,9	XP_004519700	fructose-bisphosphate aldolase-like	*Ceratitis capitata*	661	2,2E-76									X
13775	fr4	9,1	31624,1	gb|EFX72553.1|	No predicted function	*Daphnia pulex*	797	1,0E-100			X					X	
13781	fr3	11,7	35569,3	gb|ADV40154.1|	ribosomal protein L4	*Latrodectus hesperus*	758	2,0E-95						X		X	X
13812	fr6	7,7	25393,6	gb|ADJ94951.2|	voltage-dependent anion-selective channel	*Eriocheir sinensis*	655	1,0E-81	X		X	X					
13825	fr3	10,8	57954,9	AFI61881.1	serine proteinaseprotein	*Macrobrachium nipponense*	111	1,7E-02					X				
13838	fr4	7,8	58351,3	gb|ACY66440.1|	myosin regulatory light chain 2 smooth muscle	*Scylla paramamosain*	424	9,0E-46	X								
13849	fr2	5,9	39856,5	ref|XP_002424552.1|	collagen alpha-1, IV, chain precursor	*Pediculus humanus corporis*	999	3,0E-119								X	X
13857	fr2		46239,1	ref|XP_002642905.1|	No predicted function	*Caenorhabditis briggsae*	448	2,0E-48				X					
13886	fr2	10,4	42487,0		No detected homolog				X								
13912	fr2	9,7	20472,7	gb|AEE62547.1|	No predicted function	*Dendroctonus ponderosae*	583	7,0E-74	X								
13974	fr6	10,0	17034,0	gb|ABW23229.1|	ribosomal protein rpl30	*Eurythoe complanata*	504	1,0E-62	X		X	X		X		X	X
14118	fr1	9,4	111324,4	XP_004930078.1	lamin-C	*Bombyx mori*	794	1,1E-89	X		X	X		X		X	X
14181	fr4	9,3	13489,7	AGN98123	14-3-3-like protein	*Penaeus monodon*	270	1,6E-26									X
14190	fr2	9,3	46777,2	EFX89528	No predicted function	*Daphnia pulex*	519	9,5E-55									X
14232	fr3	4,6	29495,7	gb|EFX73391.1|	No predicted function	*Daphnia pulex*	117	1,0E-04						X		X	X
14235	fr6	10,4	62095,6	gb|EFX71611.1|	No predicted function	*Daphnia pulex*	1192	4,0E-154								X	X
14258	fr1	8,3	72228,3	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	755	2,0E-84									X
14260	fr3	9,2	23047,0	ref|XP_002742190.1|	Patterned Expression Site family member (pes-9)-like, partial	*Saccoglossus kowalevskii*	501	6,0E-60	X								
14268	fr4	7,9	31036,4		No detected homolog						X						
14309	fr2	7,6	50381,2		No detected homolog						X						
14348	fr5	6,2	50578,3	gb|AAC47524.1|	alpha-3-tubulin	*Gecarcinus lateralis*	2111	0,0E+00	X			X				X	
14368	fr4	10,3	56965,6		No detected homolog							X					X
14388	fr6	9,2	42659,2	gb|AFT92034.1|	triose-phosphate isomerase	*Litopenaeus vannamei*	892	1,0E-115						X		X	X
14404	fr1	7,0	48683,0	gb|ACV96855.1|	arginine kinase	*Scylla serrata*	1031	3,0E-134	X		X	X	X	X		X	X
14446	fr6	9,2	72985,6	ESP02281	No predicted function	*Lottia gigantea*	847	3,2E-101			X						X
14451	fr4	4,5	45513,4	ESP00722	No predicted function	*Lottia gigantea*	758	1,5E-91									X
14454	fr5	6,9	35907,8	EFX72698	No predicted function	*Daphnia pulex*	491	1,1E-52									X
14471	fr1	4,7	29958,9	ref|XP_002592994.1|	No predicted function	*Branchiostoma floridae*	810	3,0E-102	X		X	X					
14545	fr1	8,9	60475,2	ref|XP_004400971.1|	epididymal sperm-binding protein 1	*Odobenus rosmarus divergens*	126	6,0E-05								X	X
14564	fr4	8,8	19218,7	ref|XP_001962562.1|	No predicted function	*Drosophila ananassae*	251	3,0E-24				X					
14580	fr6	8,6	39478,9	ref|XP_003393706.1|	transketolase-like protein 2-like isoform 2	*Bombus terrestris*	808	2,0E-98			X		X	X		X	X
14600	fr1	7,7	35646,1	AAF63714	beta-1,4-glucanase 3	*Polyphaga aegyptiaca*	389	7,6E-41									X
14720	fr4	9,9	86684,6		No detected homolog						X		X	X			
14887	fr5	9,2	57212,8	gb|EFN69572.1|	Serine/threonine-protein phosphatase alpha-1 isoform	*Camponotus floridanus*	1196	2,0E-158	X								
14923	fr5	7,1	45191,9	gb|EFX79090.1|	No predicted function	*Daphnia pulex*	938	3,0E-121								X	
14943	fr4	5,8	46920,1	gb|EFX88723.1|	No predicted function	*Daphnia pulex*	768	7,0E-88			X						
15109	fr6	5,8	26430,1	gb|ABS57456.1|	alpha-tubulin	*Heliconius melpomene*	880	2,0E-116				X					
15119	fr5	10,5	35884,6	gb|EFN64477.1|	Solute carrier family 2, facilitated glucose transporter member	*Camponotus floridanus*	461	4,0E-50	X								
15128	fr2	10,2	71138,8	gb|EHJ71027.1|	No predicted function	*Danaus plexippus*	246	1,0E-19				X					
15140	fr5	9,3	26301,0	BAN20954.1	igf2 mRNA binding protein, putative	*Riptortus pedestris*	387	7,6E-40				X	X				
15282	fr2	6,2	30766,3	gb|ACH69158.1|	myosin heavy chain	*Bombyx mori*	516	2,0E-57	X		X	X	X			X	X
15433	fr2	9,4	17939,2	gb|ABR23401.1|	40S ribosomal protein S10	*Ornithodoros parkeri*	411	6,0E-48			X						
15435	fr3	10,1	30907,7		No detected homolog											X	
15457	fr5	8,5	38310,7		No detected homolog							X					
15501	fr2	9,8	70161,2	ref|XP_002612557.1|	3-hydroxybutyrate dehydrogenase	*Branchiostoma floridae*	647	8,0E-75								X	
15561	fr1	10,5	61903,1	gb|ADR50356.1|	prophenoloxidase	*Oratosquilla oratoria*	620	6,0E-68								X	X
15750	fr4	10,0	30770,6	P56252.1	Enolase	*Homarus gammarus*	547	4,5E-63						X			
15759	fr3	9,5	31285,1		No detected homolog								X				
15922	fr4	8,2	60052,0	gb|EGI59374.1|	Serine/threonine-protein phosphatase alpha-1 isoform	*Acromyrmex echinatior*	1526	0,0E+00				X					
16028	fr6	8,0	46334,8	ref|XP_967788.1|	26S protease regulatory subunit S10b	*Tribolium castaneum*	1558	0,0E+00	X		X						
16110	fr4	9,7	78108,1	AGN96208.1	vitellogenin	*Scylla paramamosain*	194	4,4E-12			X		X	X			
16143	fr4	9,3	31420,6	gb|ADM64456.1|	fatty acid binding protein	*Eriocheir sinensis*	409	3,0E-46	X		X	X				X	X
16167	fr2	10,1	73712,4	gb|AAC78141.1|	phosphopyruvate hydratase	*Penaeus monodon*	1091	7,0E-139			X	X	X	X		X	X

16199	fr5	9,4	45149,6	ref|XP_001946246.1|	No predicted function	*Acyrthosiphon pisum*	195	2,0E-13									X	X
16488	fr1	10,0	39226,0	ADY16617.1	ribosomal protein S3a	*Fenneropenaeus merguiensis*	687	4,8E-85								X			
16655	fr5	10,1	94875,8	ref|XP_001603180.2|	threonyl-tRNA synthetase, cytoplasmic-like	*Nasonia vitripennis*	720	1,0E-79	X										
16696	fr1	8,2	57594,9	ref|XP_001657107.1|	laminin beta-2 chain	*Aedes aegypti*	255	6,0E-20					X			X	X	X	
16739	fr5	10,7	53606,9	gb|EKC28803.1|	D-3-phosphoglycerate dehydrogenase, partial	*Crassostrea gigas*	349	8,0E-33										X	
16861	fr6	9,2	124031,1	emb|CAJ83333.1|	aldehyde dehydrogenase 1 family, member L1	*Xenopus (Silurana) tropicalis*	1759	0,0E+00	X				X						
16901	fr2	6,6	45546,7	gb|AFK24473.1|	protein phosphatase 2A regulatory subunit B	*Scylla paramamosain*	1730	0,0E+00					X						
16919	fr2	8,4	53280,3	gb|ACS91344.1|	peroxiredoxin	*Fenneropenaeus indicus*	724	5,0E-90				X						X	
16972	fr2	9,9	62910,5	gb|ADC55251.1|	ATP synthase subunit alpha precursor	*Litopenaeus vannamei*	1257	2,0E-163	X			X	X		X		X	X	X
16976	fr2	9,3	73776,3	ref|XP_001688503.1|	No predicted function	*Anopheles gambiae str. PEST*	716	1,0E-82					X						
17029	fr2	10,7	42582,4	gb|ELU14393.1|	No predicted function	*Capitella teleta*	461	4,0E-52	X			X	X			X	X	X	X
17046	fr6	9,5	133858,8	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	260	1,0E-19	X	X		X	X		X	X			
17091	fr1	9,6	80377,2	gb|AEV89752.1|	heat shock protein 60	*Schistocerca gregaria*	1192	9,0E-152				X	X			X	X	X	X
17127	fr3	10,3	73501,4	emb|CAB56173.1|	Na/K-ATPase a1 subunit	*Artemia franciscana*	943	2,0E-117				X				X	X	X	X
17183	fr2	5,3	28669,0	gb|AAC78141.1|	phosphopyruvate hydratase	*Penaeus monodon*	815	2,0E-103	X				X						
17207	fr1	11,1	49232,2	gb|AAA17371.1|	fast myosin heavy chain	*Homarus americanus*	598	9,0E-69	X						X	X	X	X	
17244	fr3	10,5	40265,9	XP_002043030.1	No predicted function	*Drosophila sechellia*	399	3,4E-41								X			
17385	fr4	6,6	27192,4	AAC78141.1	phosphopyruvate hydratase	*Penaeus monodon*	825	5,7E-105							X	X			
17423	fr2	8,5	29605,1	EFX65940.1	No predicted function	*Daphnia pulex*	226	2,0E-17								X			X
17452	fr5	9,7	36852,6	gb|ACO36738.1|	ubiquitin carboxyl-terminal esterase L3	*Scylla paramamosain*	230	3,0E-19				X							
17502	fr5	9,3	61365,7	AGM75775.1	vitellogenin	*Eriocheir sinensis*	283	4,2E-23						X		X			
17585	fr4	10,3	118876,4	XP_005177435.1	apolipophorins	*Musca domestica*	126	8,8E-04				X		X		X			
17708	fr3	4,5	12552,0		No detected homolog													X	
17829	fr5	10,3	31421,0	gb|ELW47039.1|	ADP/ATP translocase 1	*Tupaia chinensis*	680	1,0E-84	X										
17853	fr6	10,0	28973,9	ref|XP_001654511.1|	aldehyde oxidase	*Aedes aegypti*	301	4,0E-27				X							
17884	fr3	9,5	91439,8	ref|XP_001955651.1|	No predicted function	*Drosophila ananassae*	1058	6,0E-128								X	X	X	
17932	fr2	4,8	24390,0	gb|EFX90170.1|	No predicted function	*Daphnia pulex*	327	2,0E-32				X							
17967	fr6	6,3	46737,5	gb|ACO54865.1|	aconitase 2	*Gillichthys seta*	1475	0,0E+00	X									X	X
18033	fr4	9,7	26901,0	gb|EFZ13851.1|	No predicted function	*Solenopsis invicta*	564	3,0E-64	X				X						
18162	fr1	9,6	23683,7	ref|XP_001807949.1|	oxidase/peroxidase	*Tribolium castaneum*	144	1,0E-07				X							
18185	fr6	7,0	42838,3	gb|EFX73844.1|	No predicted function	*Daphnia pulex*	326	4,0E-33				X							
18200	fr5	4,4	25795,0	EFX67764	No predicted function	*Daphnia pulex*	409	2,0E-43											X
18209	fr6	10,0	27983,9	gb|ABM55456.1|	S10e ribosomal protein	*Xenopsylla cheopis*	405	1,0E-45					X						
18211	fr2	7,7	28747,5	ADO00930	endoplasmic reticulum protein 57	*Penaeus monodon*	666	2,0E-80											X
18324	fr6	9,6	44832,4	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	171	4,0E-10	X			X	X		X	X			
18386	fr3	9,8	64655,1	gb|EFN70601.1|	Acetyl-CoA acetyltransferase, mitochondrial	*Camponotus floridanus*	908	2,0E-111	X									X	
18406	fr4	4,9	38913,7	gb|ACO36738.1|	ubiquitin carboxyl-terminal esterase L3	*Scylla paramamosain*	465	2,0E-52	X										
18437	fr1	5,0	30664,4		No detected homolog														X
18473	fr2	9,8	92589,4	gb|EKC41795.1|	Copine-8	*Crassostrea gigas*	885	7,0E-105					X		X	X	X	X	X
18500	fr3	9,6	37037,9	sp|Q9U572.1|CLOT_PENMO	Hemolymph clottable protein	*Penaeus monodon*	204	1,0E-14		X					X	X			X
18513	fr2	9,2	33184,9	BAG07411.1	beta-tubulin	*Riptortus pedestris*	905	1,2E-119							X				
18548	fr5	10,3	24971,8	gb|EFX69015.1|	No predicted function	*Daphnia pulex*	223	4,0E-18					X						
18564	fr4	10,6	34665,5	ADB44901.1	DEAD(Asp-Glu-Ala-Asp) box polypeptide 39	*Macrobrachium nipponense*	667	1,4E-80								X			
18581	fr5	5,9	38463,8	gb|EFX80034.1|	No predicted function	*Daphnia pulex*	1579	0,0E+00	X				X						
18609	fr3	7,6	28439,7	ref|XP_003406806.1|	epididymal sperm-binding protein 1-like	*Loxodonta africana*	138	4,0E-07										X	
18677	fr2	10,6	56122,0	ref|YP_005064173.1|	chaperone protein DnaK	*Mycoplasma haemocanis str. Illin*	106	4,8E-02				X							
18893	fr6	9,1	41079,4	ref|XP_001862481.1|	No predicted function	*Culex quinquefasciatus*	274	3,0E-26				X							
18951	fr5	5,0	41562,6		No detected homolog					X			X						
18968	fr2	10,7	47843,2		No detected homolog							X	X	X					
18979	fr5	10,2	65769,2	ref|XP_003392894.1|	methenyltetrahydrofolate synthase domain-containing protei	*Bombus terrestris*	176	1,0E-10				X	X						
18981	fr1	8,7	84646,5	ref|XP_002425619.1|	T-complex protein 1 subunit theta,	*Pediculus humanus corporis*	578	8,0E-62	X			X	X						
19019	fr5	11,3	106879,1	gb|EFX89062.1|	fascilin 2-like protein	*Daphnia pulex*	761	1,0E-84					X					X	
19112	fr3	9,6	65155,6	gb|EFR28681.1|	No predicted function	*Anopheles darlingi*	684	2,0E-79				X							
19128	fr1	6,5	34124,3		No detected homolog							X		X		X			
19138	fr4	4,4	11275,5	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	194	6,4E-15											X
19169	fr6	9,5	70007,2	gb|ABG67961.1|	eukaryotic initiation factor 4A	*Callinectes sapidus*	694	9,0E-81	X			X	X						
19196	fr3	9,3	72863,3	ref|XP_003691238.1|	phosphoglycerate kinase-like	*Apis florea*	1004	3,0E-125				X	X			X	X	X	X
19261	fr6	10,0	70035,2		No detected homolog							X	X	X		X			
19453	fr3	9,0	38068,0	gb|ADE60733.1|	myosin essential light chain	*Eriocheir sinensis*	429	2,0E-48				X	X					X	
19465	fr3	7,5	44940,0	gb|AGC10822.1|	Rab-1	*Eriocheir sinensis*	924	6,0E-121				X	X					X	
19552	fr2	9,7	40711,8	AEC12821	lactate dehydrogenase	*Litopenaeus vannamei*	825	2,0E-104											X
19608	fr6	9,2	63137,3	gb|ACL77780.1|	ATP synthase b	*Spodoptera exigua*	596	3,0E-69	X				X						
19615	fr1	7,0	22120,5		No detected homolog									X					
19623	fr4	10,2	36846,2	gb|ADY39535.1|	60S Ribosomal protein L7-like	*Hottentotta judaicus*	269	8,0E-26					X						
19666	fr3	8,7	49281,2	ADO51769.1	elongation factor 2	*Litopenaeus vannamei*	1042	1,4E-129								X			
19701	fr2	9,0	72251,9		No detected homolog														X
19703	fr4	4,5	31447,7	EFN87052	Aminopeptidase N	*Harpegnathos saltator*	651	1,9E-74											X
19757	fr6	7,9	83065,3	XP_002026190	No predicted function	*Drosophila persimilis*	1203	3,0E-151											X
19763	fr6	5,4	51833,5	emb|CAQ60115.1|	glyceraldehyde-3-phosphate dehydrogenase	*Gammarus locusta*	895	2,0E-113				X							X
19775	fr1	7,0	21726,0	gb|ACY66376.1|	HSP90	*Scylla paramamosain*	807	4,0E-103				X			X	X	X	X	
19886	fr2	7,6	104661,8	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	3673	0,0E+00	X				X						
19930	fr1	8,4	63645,5	gb|EFX84177.1|	No predicted function	*Daphnia pulex*	728	5,0E-86	X										
20034	fr2	8,7	79284,9	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	278	3,0E-22				X			X	X			
20187	fr5	10,0	33780,3	gb|EFX83224.1|	No predicted function	*Daphnia pulex*	966	2,0E-118								X	X	X	
20335	fr3	9,1	50933,8	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	230	4,0E-17				X			X	X	X	X	
20449	fr1	8,4	48197,9	ref|XP_003932292.1|	aldehyde dehydrogenase, mitochondrial isoform 2	*Saimiri boliviensis boliviensis*	806	8,0E-99				X							
20488	fr4	11,6	17028,3	gb|AEE62456.1|	No predicted function	*Dendroctonus ponderosae*	439	7,0E-52					X						
20536	fr3	9,5	156676,7	ref|XP_002035131.1|	No predicted function	*Drosophila sechellia*	1380	4,0E-166				X							
20596	fr2	5,9	21656,1	AAM81357.1	hemocyanin	*Pacifastacus leniusculus*	648	4,1E-77								X			X
20628	fr1	9,3	42264,7	ACU51164.1	vitellogenin	*Pandalopsis japonica*	315	1,0E-27						X		X			
20801	fr4	11,2	96158,5	gb|EEZ99233.1|	alpha spectrin	*Tribolium castaneum*	385	1,0E-34										X	
20828	fr2	6,9	55910,1	ref|XP_001971666.1|	No predicted function	*Drosophila erecta*	1853	0,0E+00	X										
20915	fr6	9,3	98442,1	ref|XP_001601449.1|	aspartate aminotransferase, cytoplasmic-like	*Nasonia vitripennis*	1030	1,0E-127	X			X	X					X	X
20931	fr2	7,6	65260,8	gb|EFX87450.1|	No predicted function	*Daphnia pulex*	1695	0,0E+00	X				X					X	
20974	fr4	9,1	99259,0	ref|XP_001359375.1|	No predicted function	*Drosophila pseudoobscura pseud*	1609	0,0E+00										X	X
20975	fr4	6,3	15297,9	gb|ADD70028.1|	myosin light chain	*Marsupenaeus japonicus*	405	2,0E-47								X		X	X
20978	fr3	9,9	84489,8	ref|XP_003700438.1|	phosphoglycerate kinase-like	*Megachile rotundata*	930	4,0E-114										X	X
21030	fr6	8,8	51568,5	gb|EEZ99519.1|	No predicted function	*Tribolium castaneum*	760	4,0E-93	X			X	X			X		X	X
21077	fr2	9,1	38386,4	ref|XP_002052909.1|	No predicted function	*Drosophila virilis*	841	4,0E-108	X										
21166	fr3	7,9	72160,5	gb|ABW79798.1|	vitellogenin receptor	*Penaeus monodon*	183	6,0E-11				X							
21198	fr3	5,1	36992,5	ref|XP_001991062.1|	No predicted function	*Drosophila grimshawi*	904	4,0E-117				X							
21251	fr1	9,5	70450,6	gb|ADW10720.1|	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	306	7,0E-26								X		X	X
21264	fr5	6,7	65214,9	ref|XP_002737508.1|	No predicted function	*Saccoglossus kowalevskii*	1150	5,0E-148										X	X
21301	fr2	10,6	35368,1	gb|EMP24627.1|	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase	*Chelonia mydas*	332	2,0E-31					X						
21326	fr3	8,1	80028,4	ref|XP_001602512.1|	T-complex protein 1 subunit alpha-like	*Nasonia vitripennis*	1828	0,0E+00				X				X			
21346	fr2	8,5	62944,3	ref|XP_001633389.1|	No predicted function	*Nematostella vectensis*	1383	0,0E+00	X									X	X
21366	fr6	5,2	22675,1	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	884	3,0E-105	X				X						
21392	fr2	7,9	64299,0	ref|XP_972752.1|	transglutaminase	*Tribolium castaneum*	563	5,0E-60	X										
21424	fr5	8,2	46127,9	gb|ADC54203.1|	No predicted function	*Drosophila melanogaster*	1104	2,0E-145								X		X	X
21529	fr6	9,5	76218,8	gb|EKC20273.1|	Spliceosome RNA helicase BAT1	*Crassostrea gigas*	1084	5,0E-137				X	X		X	X		X	X
21548	fr2	9,9	54172,5	ref|XP_001864008.1|	No predicted function	*Culex quinquefasciatus*	887	3,0E-113					X						
21564	fr6	9,3	36538,7	gb|EFN65250.1|	Nesprin-1	*Camponotus floridanus*	112	5,0E-03					X					X	X
21574	fr2	10,2	44911,2	gb|ACC95446.1|	glutamate dehydrogenase	*Litopenaeus vannamei*	862	1,0E-106				X	X					X	
21692	fr3	9,1	44703,8	dbj|BAH57291.1|	ovarian lipoprotein receptor	*Marsupenaeus japonicus*	147	4,0E-07				X							
21694	fr5	10,3	47313,8	ref|XP_003427484.1|	muscle M-line assembly protein unc-89-like	*Nasonia vitripennis*	492	3,0E-50					X						
21803	fr4	10,0	41840,1	gb|EFX72780.1|	No predicted function	*Daphnia pulex*	621	2,0E-72					X					X	
21834	fr4	9,7	45584,0	gb|EFA04932.1|	No predicted function	*Tribolium castaneum*	844	3,0E-106					X			X		X	X
21971	fr5	9,6	34527,1		No detected homolog											X			
22001	fr2	10,2	42779,8	gb|ABK59925.1|	clottable protein [Marsupenaeus japonicus]	*Marsupenaeus japonicus*	150	1,0E-07		X						X			
22004	fr1	9,7	37322,8	dbj|BAM17830.1|	heat shock protein cognate 3	*Papilio xuthus*	653	7,0E-76				X		X				X	X
22010	fr2	9,5	84920,3	ref|XP_002055667.1|	No predicted function	*Drosophila virilis*	1724	0,0E+00				X							
22040	fr1	9,0	29234,8	ref|XP_001999705.1|	No predicted function	*Drosophila mojavensis*	365	5,0E-39										X	
22085	fr4	8,4	44695,9	ref|XP_003492254.1|	venom protease-like	*Bombus impatiens*	305	2,0E-28				X							
22180	fr5	10,0	57282,4	XP_004922073.1	methenyltetrahydrofolate synthase domain-containing protei	*Bombyx mori*	616	9,9E-70				X				X			
22232	fr3	9,0	55587,7	ref|XP_001655738.1|	citrate synthase	*Aedes aegypti*	1178	7,0E-154				X	X					X	X
22255	fr2	9,1	53284,7	gb|AAB34531.2|	myosin heavy chain	*Drosophila melanogaster*	841	9,0E-105	X			X	X	X	X	X		X	X
22261	fr5	8,8	62057,6	ref|XP_002063871.1|	No predicted function	*Drosophila willistoni*	1081	2,0E-135					X	X	X	X		X	X
22266	fr2	5,5	23270,1	gb|AAB34531.2|	myosin heavy chain	*Drosophila melanogaster*	436	2,0E-48					X			X		X	X
22276	fr1	11,3	59539,1	gb|ACK86592.1|	elongation factor 2	*Coutierella tonkinensis*	244	6,0E-21				X						X	
22376	fr4	5,3	54477,9	EFX65940.1	No predicted function	*Daphnia pulex*	881	2,6E-103								X			
22428	fr2	9,6	66454,9	ref|XP_003737757.1|	peroxidasin homolog	*Metaseiulus occidentalis*	382	8,0E-36				X							
22459	fr2	9,2	33064,0	ref|NP_001088200.1|	ubiquitin carboxyl-terminal esterase L3	*Xenopus laevis*	313	8,0E-31	X				X						
22568	fr6	10,7	67494,0	gb|EFA01539.1|	No predicted function	*Tribolium castaneum*	262	5,0E-21				X							
22677	fr4	9,4	36154,5	gb|AAC52237.1|	smooth muscle gamma-actin	*Mus musculus*	577	2,0E-67				X		X					
22690	fr6	9,9	83257,6	gb|AAH96507.1|	No predicted function	*Xenopus (Silurana) tropicalis*	1202	4,0E-153										X	X
22698	fr5	10,5	37803,1	ref|XP_002428431.1|	protein disulfide-isomerase A3 precursor	*Pediculus humanus corporis*	349	7,0E-34	X										
22707	fr2	9,2	36157,1	gb|AEO72077.1|	glutamate dehydrogenase	*Eriocheir sinensis*	977	6,0E-125										X	
22840	fr6	9,3	47217,4	ABI95361.1	hemolymph clottable protein	*Litopenaeus vannamei*	121	1,0E-03						X		X			
22990	fr6	6,5	58035,4	ACO12709.1	Tubulin alpha-1 chain	*Lepeophtheirus salmonis*	1143	3,7E-148						X					
23000	fr2	9,3	65639,0	gb|EFN69012.1|	ATP-citrate synthase	*Camponotus floridanus*	1736	0,0E+00	X										

23124	fr4	10,8	37446,9		No detected homolog								X						
23125	fr3	6,4	66771,1	gb|EFN65250.1|	Nesprin-1	*Camponotus floridanus*	196	2,0E-12	X			X	X				X	X	
23131	fr6	9,2	34359,9	gb|EFN60812.1|	No predicted function	*Camponotus floridanus*	465	5,0E-53	X										
23132	fr4	9,9	51292,1	gb|AEO72077.1|	glutamate dehydrogenase	*Eriocheir sinensis*	728	7,0E-86									X	X	
23165	fr2	11,4	28989,4	gb|EFA00566.1|	No predicted function	*Tribolium castaneum*	223	5,0E-18					X						
23201	fr6	10,1	141336,4	ref|XP_002430009.1|	Spectrin alpha chain,	*Pediculus humanus corporis*	1421	5,0E-164	X				X						
23268	fr3	9,3	139541,0	ref|XP_003705078.1|	plasma membrane calcium-transporting ATPase 3 isoform 2	*Megachile rotundata*	1667	0,0E+00	X				X						
23284	fr3	9,0	109673,9	gb|EKC35027.1|	26S proteasome non-ATPase regulatory subunit 2	*Crassostrea gigas*	1229	9,0E-150				X							
23321	fr2	9,7	41005,5	ref|XP_001952092.2|	myosin heavy chain, muscle isoform 1	*Acyrthosiphon pisum*	375	1,0E-35								X	X	X	
23340	fr1	10,3	58174,9	BAN20944	spectrin	*Riptortus pedestris*	630	1,9E-67										X	
23367	fr2	9,6	40703,9	ref|XP_794208.3|	sorbitol dehydrogenase-like	*Strongylocentrotus purpuratus*	326	5,0E-31				X							
23428	fr4	8,6	43587,9	gb|ABF18258.1|	heat shock cognate 70	*Aedes aegypti*	1097	9,0E-141				X	X	X		X	X	X	
23487	fr5	5,7	33488,0		No detected homolog											X			
23545	fr5	7,5	27700,4	gb|AFI61881.1|	serine proteinase-like protein	*Macrobrachium nipponense*	205	3,0E-15				X							
23556	fr2	5,9	35600,7	gb|ACN89260.1|	protein disulfide isomerase	*Litopenaeus vannamei*	766	2,0E-94									X		
23593	fr2	8,7	49971,0	gb|AFI60316.1|	heat shock protein 70	*Eulimnogammarus verrucosus*	1106	2,0E-141					X			X	X	X	
23819	fr5	9,4	83308,0	EEZ99360	aminopeptidase N-like protein	*Tribolium castaneum*	606	2,9E-63										X	
23850	fr4	9,7	39836,7	ref|XP_002414143.1|	ribosomal DEAD box protein,	*Ixodes scapularis*	837	7,0E-107				X		X					
23928	fr5	9,6	113257,1	ref|XP_002588697.1|	No predicted function	*Branchiostoma floridae*	239	2,0E-17					X				X	X	
23966	fr5	9,1	45265,8	gb|AFI61881.1|	serine proteinase-like protein	*Macrobrachium nipponense*	188	3,0E-12				X							
23970	fr5	8,3	59485,7	gb|ABY66598.1|	pyruvate kinase 3	*Litopenaeus vannamei*	1421	0,0E+00	X				X						
24066	fr3	10,1	49117,8	AAY56092	14-3-3-like protein	*Penaeus monodon*	607	3,7E-72				X	X	X	X	X	X	X	
24069	fr6	9,2	44239,0	gb|EFZ15216.1|	No predicted function	*Solenopsis invicta*	129	5,0E-05				X							
24240	fr3	10,2	46307,9	ACN89260.1	protein disulfide isomerase	*Litopenaeus vannamei*	632	4,3E-73								X			
24256	fr1	9,5	95803,2		No detected homolog											X			
24285	fr5	9,2	33556,0	CAF96433.1	No predicted function	*Tetraodon nigroviridis*	606	2,3E-72								X		X	
24361	fr5	9,8	33391,5	ref|XP_003241254.1|	proclotting enzyme-like isoform 2	*Acyrthosiphon pisum*	211	3,0E-16				X							
24391	fr3	10,7	58201,0	XP_004519624.1	THO complex subunit 4	*Ceratitis capitata*	321	2,6E-30								X	X		
24461	fr4	10,2	16905,9	ref|XP_001657158.1|	histone H2B	*Aedes aegypti*	486	9,0E-60				X				X	X		
24570	fr5	10,3	87427,8	gb|ADF81052.1|	vitellogenin-like protein precursor [Gammarus fossarum]	*Gammarus fossarum*	519	1,0E-57		X		X		X		X		X	
24629	fr4	9,1	58396,6	ref|XP_002404837.1|	GMP reductase,	*Ixodes scapularis*	1470	0,0E+00				X							
24672	fr2	10,2	131005,0	XP_002199494	clathrin heavy chain 1-like, partial	*Taeniopygia guttata*	1044	3,9E-123										X	
24682	fr1	10,0	55497,8	ref|XP_003701628.1|	transketolase-like protein 2-like isoform 4	*Megachile rotundata*	693	3,0E-80				X					X		
24687	fr4	10,2	47413,6	XP_001845218	78 kDa glucose-regulated protein	*Culex quinquefasciatus*	695	1,3E-80										X	
24754	fr2	10,4	55964,9	ref|XP_003741973.1|	No predicted function	*Metaseiulus occidentalis*	487	1,0E-52	X				X			X	X	X	
24823	fr3	9,7	45571,7	ref|XP_001606463.1|	heat shock 70 kDa protein cognate 3-like	*Nasonia vitripennis*	1144	8,0E-148				X	X	X	X	X	X	X	
24835	fr4	9,6	36791,6		No detected homolog											X			
24870	fr6	9,4	76182,3	ADW10720.1	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	318	4,9E-27	X			X	X	X	X	X	X	X	
24872	fr5	10,8	94276,3	gb|EFX66153.1|	No predicted function	*Daphnia pulex*	685	5,0E-78				X							
24883	fr1	4,3	16912,8	ref|XP_001120217.1|	No predicted function	*Apis mellifera*	333	2,0E-35				X							
24884	fr5	8,0	38396,0	BAA34422.1	TAR-binding protein	*Drosophila melanogaster*	636	2,0E-76								X			
24960	fr1	9,4	40889,6	AGZ01980	valosin-containing protein	*Penaeus monodon*	1265	3,7E-164										X	
25005	fr3	9,3	50313,4	ref|ZP_08260757.1|	No predicted function	*Gemella sanguinis M325*	162	8,0E-09	X				X						
25106	fr1	9,8	56573,0	gb|ABM55532.1|	small heat shock protein	*Maconellicoccus hirsutus*	213	2,0E-16	X				X						
25127	fr4	10,1	38944,1	gb|ABK59925.1|	clottable protein	*Marsupenaeus japonicus*	172	2,0E-10				X		X		X			
25141	fr3	9,8	127332,2	sp|Q9U572.1|CLOT_PENMO	Hemolymph clottable protein	*Penaeus monodon*	121	3,0E-03				X		X		X			
25178	fr2	6,5	40345,1	gb|EKC34306.1|	Cytosolic non-specific dipeptidase	*Crassostrea gigas*	899	3,0E-113	X				X						
25193	fr4	6,3	63358,7	emb|CCI71879.1|	gelsolin	*Homarus americanus*	1533	0,0E+00	X										
25201	fr1	10,6	47448,3	ref|XP_003456401.1|	cytochrome b-c1 complex subunit 2, mitochondrial-like	*Oreochromis niloticus*	201	7,0E-14									X		
25209	fr4	6,1	29263,1	gb|AFD28274.1|	14-3-3-like protein	*Scylla paramamosain*	789	4,0E-102	X					X		X	X	X	
25448	fr6	9,4	44702,8	gb|EGI60184.1|	DEAD-box helicase Dbp80	*Acromyrmex echinatior*	997	6,0E-128					X						
25564	fr1	5,6	21887,4	ref|XP_002431395.1|	No predicted function	*Pediculus humanus corporis*	312	8,0E-30					X						
25777	fr6	10,2	42083,2	gb|EFX82704.1|	No predicted function	*Daphnia pulex*	541	2,0E-61					X						
25794	fr6	9,0	28562,9	gb|EFX74164.1|	No predicted function	*Daphnia pulex*	749	5,0E-94	X				X						
25797	fr3	6,9	23692,5		No detected homolog												X		
25891	fr5	9,7	38311,2	ref|XP_001950201.2|	glycogen	*Acyrthosiphon pisum*	666	2,0E-77	X								X		
26367	fr2	9,2	44300,1	ref|XP_975329.1|	bicoid stability factor CG10302-PA	*Tribolium castaneum*	554	4,0E-59	X										
26473	fr4	9,2	25161,8		No detected homolog												X		
26732	fr2	6,1	7782,0	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	315	2,0E-31					X						
26896	fr2	6,4	60890,2	ref|XP_002165626.2|	No predicted function	*Hydra magnipapillata*	152	2,0E-07									X		
27649	fr3	7,3	50584,7	gb|AFM75821.1|	myosin light chain 2	*Procambarus clarkii*	518	6,0E-60	X				X						
27742	fr3	4,4	46191,5		No detected homolog								X						
28073	fr6	5,5	13773,7	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	602	6,0E-69	X										
28350	fr5	6,7	50280,2	ref|XP_002587540.1|	No predicted function	*Branchiostoma floridae*	1283	6,0E-172	X										
28400	fr1	9,1	39251,2		No detected homolog												X	X	
28434	fr5	4,9	25036,8	gb|ENN80241.1|	No predicted function	*Dendroctonus ponderosae*	362	2,0E-35	X								X		
28559	fr5	5,5	53503,3	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	2087	0,0E+00	X				X				X	X	
28570	fr3	7,2	14505,4	AAD16454.1	clotting protein precursor	*Pacifastacus leniusculus*	104	4,5E-03						X		X			
28578	fr2	6,8	50294,4	ref|XP_003705284.1|	alpha-aminoadipic semialdehyde synthase, mitochondrial-like	*Megachile rotundata*	964	4,0E-117					X						
28713	fr5	8,9	14436,4		No detected homolog												X		
28819	fr4	5,5	63254,6	ref|XP_003395464.1|	No predicted function	*Bombus terrestris*	1242	2,0E-147	X									X	
28841	fr1	9,9	29777,4		No detected homolog				X										
29064	fr6	8,7	42636,6	ref|XP_002600835.1|	No predicted function	*Branchiostoma floridae*	347	1,0E-35					X						
29145	fr6	4,9	28437,6		No detected homolog							X		X					
29225	fr1	8,8	42215,8	ref|NP_001171772.1|	succinyl-CoA:3-ketoacid-coenzyme A transferase 1, mitochond	*Saccoglossus kowalevskii*	1191	7,0E-157					X				X		
29283	fr3	6,6	18763,2	prf||1010303J	crystallin alphaA	*Uncharacterized*	204	5,0E-17	X										
29315	fr3	5,9	28434,4	gb|EFX65172.1|	No predicted function	*Daphnia pulex*	995	3,0E-134	X				X						
29320	fr2	4,6	49962,4	ref|XP_002431590.1|	laminin A chain	*Pediculus humanus corporis*	421	1,0E-40	X				X				X	X	
29567	fr1	6,3	44035,8	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	1551	0,0E+00	X				X						
29593	fr4	5,7	57051,0	gb|ACH96582.1|	NDAE-B	*Aedes aegypti*	1293	8,0E-163								X	X	X	
29624	fr6	5,4	53724,3	gb|EFX83462.1|	No predicted function	*Daphnia pulex*	571	7,0E-60	X				X					X	
29874	fr6	5,5	40659,5	ref|XP_001945240.2|	basement membrane-specific heparan sulfate proteoglycan cc	*Acyrthosiphon pisum*	647	4,0E-71				X	X	X			X	X	
29901	fr4	8,4	30715,0	gb|EFX84988.1|	No predicted function	*Daphnia pulex*	1114	8,0E-137					X						
29954	fr5	4,8	36761,4	ref|XP_973346.1|	phosphoacetylglucosamine mutase	*Tribolium castaneum*	926	3,0E-117				X	X						
30105	fr6	6,0	63626,0		No detected homolog													X	
30183	fr3	4,4	18014,2	gb|AEC22811.1|	cathepsin L	*Macrobrachium nipponense*	519	1,0E-61	X										
30189	fr3	9,8	65698,5	gb|ACY72380.1|	cellulose hydrolase	*Bacillus licheniformis*	275	2,0E-22									X	X	
30226	fr1	9,0	52490,2	ref|XP_971540.1|	Glycoprotein 93 CG5520-PA	*Tribolium castaneum*	1002	5,0E-124								X	X	X	
30229	fr4	5,5	50171,5	ref|XP_002739409.1|	No predicted function	*Saccoglossus kowalevskii*	1387	0,0E+00									X		
30230	fr1	5,2	48125,2	gb|EFX88976.1|	myosin regulatory light chain	*Daphnia pulex*	401	1,0E-42					X						
30247	fr5	10,3	46974,4	gb|ABB59715.1|	hemocyanin	*Cyamus scammoni*	507	7,0E-54	X			X	X	X		X	X	X	
30280	fr4	7,2	56256,8	gb|AET34921.1|	chaperonin	*Macrobrachium rosenbergii*	1257	1,0E-164				X							
30292	fr1	9,0	32848,9	ACB11340.1	heat shock protein 70	*Daphnia magna*	954	4,5E-121								X			
30310	fr2	9,4	42399,6	ref|XP_971851.1|	NADH-ubiquinone oxidoreductase 42 kda subunit	*Tribolium castaneum*	667	5,0E-80					X						
30342	fr3	6,6	56272,5	emb|CAB85963.1|	alpha glucosidase	*Litopenaeus vannamei*	794	1,0E-91									X	X	
30353	fr4	5,7	40404,3	AHA11058	vacuolar H+-ATPase	*Exopalaemon carinicauda*	1262	3,9E-168				X					X	X	
30375	fr6	9,4	90105,8	gb|AAN77377.1|	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	1070	3,0E-127	X				X			X	X	X	
30378	fr1	9,2	68996,0	gb|ABY64692.1|	alpha-2-macroglobulin, partial	*Armadillidium vulgare*	792	3,0E-93	X								X	X	
30412	fr5	9,3	42575,8	gb|ACG63697.1|	myosin heavy chain	*Gammarus duebeni*	1203	2,0E-159								X	X	X	
30416	fr4	9,4	48312,6	gb|ENN78269.1|	No predicted function	*Dendroctonus ponderosae*	320	8,0E-31				X							
30449	fr6	7,7	51448,8	gb|EFR27027.1|	No predicted function	*Anopheles darlingi*	525	1,0E-56	X			X	X			X	X	X	
30461	fr2	10,0	208633,4	gb|AAA48986.1|	myosin heavy chain	*Gallus gallus*	1453	8,0E-165	X			X	X			X	X	X	
30486	fr4	9,6	100689,6	AAN77377.1	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	779	5,0E-85	X			X	X			X	X		
30488	fr4	10,0	16286,1		No detected homolog							X							
30490	fr1	9,5	22343,2	gb|ADF81052.1|	vitellogenin-like protein precursor	*Gammarus fossarum*	214	6,0E-18				X		X		X			
30505	fr4	9,3	111238,3	ref|XP_003489788.1|	myosin heavy chain, non-muscle-like	*Bombus impatiens*	770	1,0E-81									X	X	
30514	fr6	9,0	97609,6	ref|XP_003494094.1|	citrate synthase 2, mitochondrial-like	*Bombus impatiens*	1071	6,0E-133	X										
30561	fr3	6,0	46687,1	gb|AAN77377.1|	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	1371	5,0E-177					X						
30611	fr4	10,5	87454,6	gb|ACR78686.1|	paramyosin	*Rimicaris exoculata*	575	1,0E-64				X	X	X		X	X	X	
30649	fr2	9,3	55712,8	gb|ADF81052.1|	vitellogenin-like protein precursor	*Gammarus fossarum*	274	1,0E-24	X			X	X	X		X			
30701	fr6	11,4	70838,1	EFN70346	Filamin-B	*Camponotus floridanus*	706	1,1E-75										X	
30755	fr3	4,7	11075,3	gb|EFZ16730.1|	No predicted function	*Solenopsis invicta*	163	6,0E-12									X	X	
30770	fr5	9,7	18248,4	AGJ90347	Heat shock protein 90	*Ornithodoros moubata*	385	1,4E-40									X	X	
30813	fr2	10,8	56669,4	gb|AEC50079.1|	calreticulin	*Pacifastacus leniusculus*	329	1,0E-30								X	X		
30838	fr5	8,9	26985,2	ref|NP_001119609.1|	sex-lethal	*Acyrthosiphon pisum*	414	2,0E-46				X							
30841	fr6	10,2	66404,6	ABF20530.1	heat shock protein 70	*Metapenaeus ensis*	1085	6,4E-136					X			X	X		
30894	fr1	10,5	42461,0	gb|AAN72821.1|	beta-I tubulin	*Callinectes sapidus*	660	3,0E-81						X		X	X	X	
30899	fr3	6,2	33866,0	ref|XP_974030.1|	Annexin IX CG5730-PC	*Tribolium castaneum*	592	1,0E-70				X							
30921	fr3	9,5	46040,1	AGI96988.1	troponin T	*Nilaparvata lugens*	483	1,2E-52								X	X	X	
30941	fr4	5,9	36112,1	AET34917.1	apoptosis inhibitor	*Macrobrachium rosenbergii*	507	1,2E-55								X		X	
30942	fr6	10,1	86534,4		No detected homolog												X	X	
31248	fr3	9,1	80338,1	gb|EFX84778.1|	No predicted function	*Daphnia pulex*	1521	0,0E+00									X	X	
31263	fr6	8,2	68442,3	XP_971371	No predicted function	*Tribolium castaneum*	529	2,5E-56										X	
31408	fr2	5,6	39849,8	emb|CAB85963.1|	alpha glucosidase	*Litopenaeus vannamei*	812	2,0E-96									X	X	
31415	fr2	9,0	108904,9	ref|XP_002051741.1|	No predicted function	*Drosophila virilis*	1051	2,0E-124									X		
31580	fr4	10,0	50007,4	ref|ZP_09616004.1|	carboxyl transferase	*Mucilaginibacter paludis DSM 18*	854	3,0E-105									X		
31746	fr5	9,4	49203,0		No detected homolog							X	X	X	X				
31805	fr5	8,6	36693,0	sp|Q9U572.1|CLOT_PENMO	Hemolymph clottable protein	*Penaeus monodon*	130	3,0E-05	X								X	X	
31902	fr5	9,0	149839,1	ref|XP_003693213.1|	alpha-actinin, sarcomeric-like	*Apis florea*	1762	0,0E+00								X	X	X	
31948	fr3	9,3	54597,5	ref|XP_001968775.1|	No predicted function	*Drosophila erecta*	180	7,0E-11	X								X	X	

31949	fr6	8,6	82166,8	gb|EFN65250.1|	Nesprin-1	*Camponotus floridanus*	117	5,0E-03				X				X	X
32053	fr2	9,1	58312,8	sp|Q8WQ47.2|TBA_LEPDS	Tubulin alpha chain	*Lepidoglyphus destructor*	1245	7,0E-164				X					
32111	fr5	9,1	133382,9	CAA56703	beta-1,3-D-glucan binding protein	*Pacifastacus leniusculus*	500	4,7E-48									X
32195	fr3	9,7	44253,3	gb|ACY66451.1|	alpha-I tubulin	*Scylla paramamosain*	1122	7,0E-150								X	
32234	fr1	11,2	60625,7	emb|CAB56173.1|	Na/K-ATPase a1 subunit	*Artemia franciscana*	516	3,0E-56							X	X	X
32254	fr1	12,3	23478,1	XP_002742335.1	histone H1-II-1, putative	*Saccoglossus kowalevskii*	119	1,5E-04							X		
32330	fr6	9,0	64177,6	ref|XP_002613480.1|	No predicted function	*Branchiostoma floridae*	1190	2,0E-153			X						
32480	fr1	9,7	47572,8	gb|AAD16454.1|	clotting protein precursor	*Pacifastacus leniusculus*	130	5,0E-05				X					
32503	fr6	7,6	57846,6	gb|EFX87834.1|	No predicted function	*Daphnia pulex*	1684	0,0E+00	X			X					
32837	fr1	9,8	53324,6	XP_002400188.1	peroxidase	*Ixodes scapularis*	395	2,6E-37					X				
32843	fr5	11,2	75011,4	gb|AEC50080.1|	alpha-2-macroglobulin	*Pacifastacus leniusculus*	273	1,0E-21				X					
32861	fr4	10,3	36733,1	ref|XP_002744826.2|	histone H3.3-like, partial	*Callithrix jacchus*	623	2,0E-77							X	X	X
32875	fr1	9,6	34866,6	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	231	6,0E-18	X		X						
33340	fr1	8,4	54119,0	gb|ACS36538.1|	elongation factor 2	*Homarus americanus*	1202	1,0E-152			X	X				X	
33506	fr2	6,8	30953,2		No detected homolog				X								
33679	fr3	8,4	60888,8		No detected homolog				X		X	X				X	X
33719	fr1	9,2	51260,1	ref|XP_002021840.1|	No predicted function	*Drosophila persimilis*	313	5,0E-28				X					
33730	fr4		24542,8	NP_001003349.1	Actin	*Canis lupus familiaris*	615	7,0E-75					X				
33807	fr3	8,3	37253,0	gb|AFK76478.1|	cold shock protein	*Enterobacter cloacae*	291	3,0E-29			X	X					
33874	fr1	8,9	43882,0	XP_002432182	Heterogeneous nuclear ribonucleoprotein A1	*Pediculus humanus corporis*	577	1,0E-65									X
33940	fr6	5,1	33867,4	ref|XP_002831364.1|	acrosin	*Pongo abelii*	195	5,0E-14			X						
33972	fr4	9,3	20112,9	gb|ACU82845.1|	fatty acids binding protein	*Fenneropenaeus chinensis*	136	6,0E-08				X					
34079	fr5	6,1	20331,3	gb|AAB34531.2|	myosin heavy chain	*Drosophila melanogaster*	544	1,0E-64	X		X	X	X		X	X	X
34103	fr4	9,4	49852,7	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	711	1,4E-80									X
34157	fr4	9,5	72080,4	XP_004932596.1	chorion peroxidase	*Bombyx mori*	493	3,8E-49					X				
34262	fr3	10,1	27130,4	ACU52718	ribosomal protein L10a	*Fenneropenaeus merguiensis*	511	9,5E-61									X
34272	fr5	9,8	52950,6	ref|XP_801745.2|	Aconitate hydratase	*Strongylocentrotus purpuratus*	754	6,0E-89				X					
34339	fr5	11,6	85190,9	ref|XP_002423351.1|	Filamin-C	*Pediculus humanus corporis*	403	1,0E-37	X				X		X	X	X
34340	fr6		16949,9		No detected homolog								X				
34536	fr6	10,2	34147,0	XP_001602027.2	apoptosis inhibitor 5-B	*Nasonia vitripennis*	200	3,3E-14							X		
34610	fr6	9,0	77113,5		No detected homolog										X		
34701	fr5		30559,9	ref|XP_001867082.1|	cytochrome B5	*Culex quinquefasciatus*	325	6,0E-34				X					
34824	fr2		52990,0	gb|AFJ22310.1|	heat shock protein cognate 5, partial	*Alicorhagia sp. AD1316*	198	2,0E-13	X								
34845	fr5	9,1	85478,8	gb|EKC41795.1|	Copine-8	*Crassostrea gigas*	765	4,0E-88	X			X	X			X	X
34912	fr1	9,6	47038,4	emb|CAA56703.1|	beta-1,3-D-glucan binding protein	*Pacifastacus leniusculus*	482	3,0E-49				X				X	X
35354	fr1	10,0	61250,6	gb|AEE61406.1|	No predicted function	*Dendroctonus ponderosae*	822	7,0E-100	X			X				X	
35422	fr5	10,1	70880,0	dbj|BAM99026.1|	valosin containing protein, partial	*Struthio camelus*	1024	8,0E-131					X		X	X	X
35613	fr5	9,4	38037,3	ref|XP_004431246.1|	carboxypeptidase Q	*Ceratotherium simum simum*	631	4,0E-74			X		X				
35684	fr2	9,8	42380,4	gb|EHJ72859.1|	rap55	*Danaus plexippus*	375	2,0E-36			X						
35708	fr5	7,4	89614,7	EFX81905	No predicted function	*Daphnia pulex*	789	5,2E-88									X
35752	fr6	10,1	39003,2	gb|ADQ43366.1|	No predicted function	*Litopenaeus vannamei*	287	6,0E-27				X					
35858	fr6	9,9	49810,4	ref|XP_002423351.1|	Filamin-C	*Pediculus humanus corporis*	1120	4,0E-141							X	X	X
36051	fr3	5,3	37020,9	ref|NP_998717.1|	6-phosphogluconate dehydrogenase, decarboxylating isoform	*Danio rerio*	926	3,0E-118				X					
36113	fr3	6,8	40684,6	gb|ADK66338.1|	C-type lectin	*Eriocheir sinensis*	163	2,0E-10	X								
36189	fr1	8,2	27753,8	gb|AFP57558.1|	heat shock cognate protein 70, partial	*Latrodectus hesperus*	863	4,0E-113			X		X				
36242	fr3	5,5	54584,1	gb|AAW57890.1|	hemocyanin subunit 2	*Metacarcinus magister*	1032	1,0E-129			X	X			X	X	X
36277	fr4	8,8	27780,6	XP_005107606	spliceosome RNA helicase Ddx39b-like	*Aplysia californica*	530	2,5E-62									X
36278	fr6	9,2	39333,4	EFX64849.1	No predicted function	*Daphnia pulex*	940	9,1E-121							X		
36330	fr4	8,9	52406,2	gb|ACA83751.1|	ribophorin I	*Penaeus monodon*	827	4,0E-100				X					
36606	fr4	8,9	27751,6	gb|ACM91676.1|	F1F0-ATP synthase beta subunit	*Marsupenaeus japonicus*	711	9,0E-87	X								
36666	fr4	8,8	65043,0	gb|EFN84507.1|	Elongation factor Tu, mitochondrial	*Harpegnathos saltator*	1233	8,0E-162	X			X					
37031	fr3	8,5	61633,7	ref|XP_002409147.1|	beta-glucocerebrosidase	*Ixodes scapularis*	579	6,0E-64	X							X	X
37243	fr4	6,6	45019,0		No detected homolog								X		X		
37248	fr5	4,4	30882,4	sp|Q9GV46.1|LUCB_OPLGR	Oplophorus-luciferin 2-monooxygenase non-catalytic subunit	*Oplophorus gracilirostris*	170	8,0E-11				X					
37255	fr2	8,9	38914,6	sp|P09645.1|TBA8_CHICK	Tubulin alpha-8 chain	*Gallus gallus*	1235	4,0E-167			X	X			X	X	X
37276	fr5	9,6	92006,6	EFX87105.1	myosin heavy chain isoform 2	*Daphnia pulex*	1198	6,0E-140	X		X	X	X		X	X	X
37308	fr2	10,7	65000,8	gb|ACM17460.1|	miniparamyosin	*Bombyx mori*	592	3,0E-65			X		X		X	X	X
37328	fr4	9,5	60670,3	ref|XP_002008153.1|	No predicted function	*Drosophila mojavensis*	451	9,0E-48	X								
37338	fr3	8,8	105339,8	ref|NP_651601.1|	glycoprotein 93	*Drosophila melanogaster*	1379	8,0E-174			X	X	X			X	X
37524	fr1	9,7	35496,0	gb|EFX88544.1|	carbonyl reductase	*Daphnia pulex*	412	1,0E-44								X	
37597	fr3	9,9	39716,3	dbj|BAD17915.1|	triose phosphate isomerase	*Amia calva*	442	5,0E-49								X	
37598	fr5	6,1	26394,9		No detected homolog						X						
37605	fr1	10,3	29401,4	XP_005189831.1	THO complex subunit 5	*Musca domestica*	316	3,7E-31							X	X	
37606	fr2	7,5	48530,5	gb|ABQ10738.1|	cathepsin D	*Penaeus monodon*	1079	6,0E-141			X	X					
37636	fr5	8,8	30436,7		No detected homolog						X						
37670	fr4	9,8	67652,8	gb|ACS27517.1|	phosphoglucose isomerase	*Colias eurytheme*	678	7,0E-77			X	X			X	X	X
37683	fr2	9,9	28567,2	gb|ABB59715.1|	hemocyanin	*Cyamus scammoni*	455	7,0E-49								X	
37698	fr1	8,5	51941,7	gb|ADG85663.1|	lectin A isoform 4	*Marsupenaeus japonicus*	217	1,0E-17	X							X	
37795	fr5	9,5	38191,0	gb|ADF81052.1|	vitellogenin-like protein precursor [Gammarus fossarum]	*Gammarus fossarum*	634	4,0E-78		X	X		X		X		
37799	fr3	9,5	31647,6	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	663	4,0E-80	X								
37820	fr2	9,2	44936,7		No detected homolog						X						
37843	fr2	8,8	45156,5	gb|EFZ17063.1|	No predicted function	*Solenopsis invicta*	651	1,0E-76			X						
37901	fr4	10,6	46660,8		No detected homolog							X					
38027	fr1	8,1	43443,9	dbj|BAH72354.1|	No predicted function	*Acyrthosiphon pisum*	563	9,0E-66				X					
38047	fr5	10,1	27499,6	ref|XP_002155286.2|	annexin-B12-like	*Hydra magnipapillata*	332	2,0E-33				X					
38110	fr6	8,4	34206,5	ref|XP_002754345.1|	heat shock protein HSP 90-alpha-like	*Callithrix jacchus*	878	6,0E-115			X	X	X		X	X	X
38155	fr2	9,9	28956,5	gb|ACP19740.1|	thymosin-repeated protein 1	*Eriocheir sinensis*	188	1,0E-14			X						
38187	fr4	10,3	74642,3	gb|EAW97907.1|	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2, isofc	*Homo sapiens*	711	8,0E-82			X	X				X	
38209	fr1	10,6	36064,0		No detected homolog										X		
38283	fr1	8,3	76311,5	emb|CAB85963.1|	alpha glucosidase	*Litopenaeus vannamei*	877	1,0E-101	X		X	X				X	X
38421	fr6	9,0	95714,1	emb|CCD75479.1|	clathrin heavy chain	*Schistosoma mansoni*	2690	0,0E+00	X			X					X
38577	fr5	9,3	67252,0	ref|XP_002542385.1|	heat shock protein 82	*Uncinocarpus reesii 1704*	764	2,0E-88			X	X	X		X	X	X
38609	fr3	5,9	33886,3	gb|EEZ99519.1|	No predicted function	*Tribolium castaneum*	656	1,0E-79								X	
38697	fr2	8,3	69686,6	ref|XP_002049382.1|	No predicted function	*Drosophila virilis*	488	7,0E-55			X	X					
38765	fr2	5,7	35926,5	ref|XP_001956880.1|	No predicted function	*Drosophila ananassae*	478	9,0E-51								X	
38846	fr3	9,7	24882,9	AEX65804.1	heat shock protein 70	*Gammarus lacustris*	959	7,5E-123							X		
39038	fr2	9,2	63241,8		No detected homolog						X				X	X	
39065	fr4	10,1	50384,9	gb|ACO50110.1|	elongation factor 1 alpha	*Euglena gracilis*	869	2,0E-108				X					
39107	fr3	9,1	80274,4	AFK65746.1	ubiquitin-activating enzyme E1	*Eriocheir sinensis*	976	2,8E-114	X						X		
39162	fr2	10,3	81992,5	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	888	6,0E-108							X	X	X
39262	fr4	9,8	45837,2		No detected homolog						X		X		X		
39357	fr1	9,1	97615,5	ref|XP_003741973.1|	No predicted function	*Metaseiulus occidentalis*	564	1,0E-61								X	
39462	fr2	9,2	70404,7	ref|XP_002594232.1|	No predicted function	*Branchiostoma floridae*	788	3,0E-94			X						
39518	fr3	10,2	61357,5	gb|ABY66597.1|	pyruvate kinase 2	*Litopenaeus vannamei*	1130	5,0E-145								X	
39606	fr3	9,3	64021,1	gb|AAD16454.1|	clotting protein precursor	*Pacifastacus leniusculus*	127	2,0E-04			X	X	X		X		
39612	fr1	9,8	63848,2		No detected homolog								X				
39714	fr6	5,0	27407,3	gb|ACY66440.1|	myosin regulatory light chain 2 smooth muscle	*Scylla paramamosain*	548	4,0E-67	X		X	X					
39721	fr5	4,8	41409,3	ref|XP_971540.1|	Glycoprotein 93 CG5520-PA	*Tribolium castaneum*	985	7,0E-123								X	
39789	fr1	10,1	45693,9	gb|EFN88457.1|	Myosin heavy chain, muscle	*Harpegnathos saltator*	499	2,0E-51	X								
39890	fr4	9,9	79112,7	ref|XP_971568.1|	fucosidase, alpha-L- 2, plasma	*Tribolium castaneum*	608	2,0E-67				X					
40004	fr6	9,6	59136,5	ref|XP_001948977.1|	t-complex protein 1 subunit gamma-like	*Acyrthosiphon pisum*	585	2,0E-64				X					
40028	fr4	9,2	123292,7	ref|XP_001601953.2|	glycogen phosphorylase-like	*Nasonia vitripennis*	1017	2,0E-116	X			X			X	X	X
40034	fr4	9,1	79755,5	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*	886	6,0E-105					X		X	X	X
40044	fr3	8,8	76725,8	ACV97157.1	collagenolytic serine protease	*Fenneropenaeus chinensis*	166	2,2E-09					X				
40069	fr2	7,2	24578,8	Q9GV46	No predicted function	*0*	154	5,1E-09									X
40079	fr3	6,9	24037,9	gb|AAQ74441.1|	glutathione S-transferase	*Haemaphysalis longicornis*	389	5,0E-43			X	X					
40095	fr4	10,0	46062,4		No detected homolog												X
40105	fr6	5,1	17319,4	gb|ACR78689.1|	cuticle protein	*Rimicaris exoculata*	305	2,0E-32	X								
40137	fr6	9,2	34533,9	emb|CAQ60115.1|	glyceraldehyde-3-phosphate dehydrogenase	*Gammarus locusta*	905	2,0E-117								X	
40218	fr2	9,4	87247,1	ref|XP_002738122.1|	trifunctional protein, alpha subunit-like, partial	*Saccoglossus kowalevskii*	627	1,0E-68							X	X	X
40293	fr4	9,1	51903,3	ref|XP_001956880.1|	No predicted function	*Drosophila ananassae*	773	6,0E-90	X			X			X	X	X
40603	fr6	7,3	57413,8	ref|XP_002025550.1|	No predicted function	*Drosophila persimilis*	1313	7,0E-173	X			X					
40632	fr5	9,2	31649,2		No detected homolog							X					
40709	fr6	8,6	81394,5	XP_005141044.1	tubulin alpha-1C chain	*Melopsittacus undulatus*	1273	5,1E-165	X				X		X	X	
40712	fr5	10,2	50141,3	ref|XP_001628246.1|	No predicted function	*Nematostella vectensis*	834	5,0E-103			X						
40895	fr6	9,1	54423,6	ref|XP_003974830.1|	heterogeneous nuclear ribonucleoprotein K-like	*Takifugu rubripes*	606	8,0E-69			X	X					
41027	fr6	4,8	32017,5	emb|CAB85963.1|	alpha glucosidase	*Litopenaeus vannamei*	744	6,0E-88	X			X				X	X
41135	fr1	5,8	29654,9	BAM65720.1	myosin heavy chain type 2	*Penaeus monodon*	639	9,5E-72					X		X		
41268	fr3	3,9	29158,5	AGZ95687	C-type lectin 3	*Exopalaemon modestus*	248	2,8E-21									X
41306	fr4	11,0	72038,2	XP_004525349	paramyosin	*Ceratitis capitata*	637	7,1E-73				X			X	X	X
41523	fr3	10,0	43274,8		No detected homolog								X		X	X	X
41589	fr6	9,9	20390,5	BAJ23879	glycogen phosphorylase	*Marsupenaeus japonicus*	421	1,6E-44									X
41820	fr1	9,3	49620,5		No detected homolog						X	X				X	
41838	fr3	8,8	29660,6	gb|AGG22607.1|	cytochrome c oxidase subunit IV	*Locusta migratoria*	539	2,0E-65	X		X	X					
41923	fr2		59950,2	gb|EFX84493.1|	No predicted function	*Daphnia pulex*	252	4,0E-22								X	
41933	fr1	8,8	70325,6	gb|AAN77377.1|	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	1026	4,0E-123	X			X			X	X	X
41969	fr5	10,0	41887,5	gb|EFN75173.1|	paramyosin	*Harpegnathos saltator*	514	3,0E-56	X		X	X			X	X	X
42134	fr4	9,8	64873,1	ABO21408	pyruvate kinase	*Litopenaeus vannamei*	1286	8,4E-167									X
42143	fr3	10,2	49225,5	WP_016318454.1	No predicted function	*Firmicutes bacterium M10-2*	108	2,7E-02							X	X	X
42149	fr5	9,1	41659,2	gb|AFC60662.1|	chitinase	*Pandalopsis japonica*	364	3,0E-35	X								X

42289	fr3	10,1	42943,0		No detected homolog							X							
42297	fr5	9,2	30011,5	AFD61666	cytosolic manganese superoxide dismutase	*Cherax quadricarinatus*	543	3,9E-64											X
42379	fr2	10,5	92133,4	ref|XP_003737657.1|	peptidyl-prolyl cis-trans isomerase FKBP4-like	*Metaseiulus occidentalis*	210	1,0E-14					X						
42448	fr5	10,4	60798,9	gb|EFR29638.1|	No predicted function	*Anopheles darlingi*	362	8,0E-36				X						X	
42461	fr4	9,2	137406,9	gb|EFX77111.1|	No predicted function	*Daphnia pulex*	1773	0,0E+00				X							
42502	fr1	9,7	26122,1	emb|CAA56703.1|	beta-1,3-D-glucan binding protein	*Pacifastacus leniusculus*	183	2,0E-12	X										
42505	fr2	8,9	54796,9	gb|EMC81649.1|	Tubulin alpha-3 chain, partial	*Columba livia*	1392	0,0E+00				X			X			X	
42523	fr4	9,0	62280,2	XP_002415272.1	peptidase	*Ixodes scapularis*	102	6,3E-02							X				
42757	fr6	7,2	14379,0	ref|XP_001950806.2|	No predicted function	*Acyrthosiphon pisum*	122	2,0E-05										X	
42833	fr6	9,3	25447,7	ref|XP_968221.1|	No predicted function	*Tribolium castaneum*	347	8,0E-38	X				X						
42839	fr2	9,1	63245,2	gb|AEG80154.1|	Tcp-1-beta	*Litopenaeus vannamei*	1298	8,0E-170				X	X						
42924	fr4	11,3	35777,8	gb|ACO12528.1|	60S ribosomal protein L4	*Lepeophtheirus salmonis*	539	6,0E-62									X	X	X
42986	fr5	4,3	21740,4	gb|EFX89905.1|	No predicted function	*Daphnia pulex*	248	7,0E-23										X	X
43139	fr5	8,7	63062,8	gb|AAM81357.1|AF522504_:	hemocyanin	*Pacifastacus leniusculus*	1159	1,0E-147									X	X	X
43172	fr6	9,0	43220,7	ref|XP_001657588.1|	short-chain dehydrogenase	*Aedes aegypti*	406	9,0E-43										X	X
43196	fr6	5,9	45780,6	dbj|BAJ23879.1|	glycogen phosphorylase	*Marsupenaeus japonicus*	1681	0,0E+00										X	X
43493	fr2	8,8	53329,3	ref|XP_004076822.1|	tubulin alpha-1C chain-like	*Oryzias latipes*	815	1,0E-102										X	
43527	fr5	10,0	22756,5	ACG63697.1	myosin heavy chain	*Gammarus duebeni*	389	8,6E-41							X		X		X
43633	fr6	10,4	51774,3	AGZ95685	C-type lectin 1	*Exopalaemon modestus*	151	9,7E-08											X
43739	fr3	11,2	101238,5	ref|XP_001953208.1|	No predicted function	*Drosophila ananassae*	473	1,0E-47					X					X	X
43810	fr6	9,4	57663,2	ref|XP_003435825.1|	No predicted function	*Anopheles gambiae str. PEST*	1247	4,0E-163	X				X						
43912	fr3	9,9	33580,5		No detected homolog								X					X	X
43914	fr2	9,7	96952,8	ref|XP_002738122.1|	trifunctional protein, alpha subunit-like, partial	*Saccoglossus kowalevskii*	961	5,0E-116					X					X	X
43964	fr4	9,2	50439,8	ref|NP_954688.1|	adenosylhomocysteinase	*Danio rerio*	963	1,0E-122				X							
43974	fr2	9,8	97899,9	P81182	No predicted function	*0*	337	6,5E-29											X
44002	fr5	6,7	40557,5	gb|AFM75821.1|	myosin light chain 2	*Procambarus clarkii*	468	2,0E-53				X							
44073	fr1	9,6	69412,7		No detected homolog												X		
44126	fr2	9,3	34891,9	gb|AGE44125.1|	tropomyosin	*Portunus pelagicus*	799	3,0E-102	X			X	X				X	X	X
44177	fr5	10,4	69124,7	gb|AEG80154.1|	Tcp-1-beta	*Litopenaeus vannamei*	1161	1,0E-148									X	X	X
44186	fr4	8,2	32089,6		No detected homolog				X			X	X					X	X
44277	fr2	9,2	69674,4	emb|CAM73269.1|	heat shock protein 70	*Oniscus asellus*	528	1,0E-59	X						X		X	X	X
44380	fr1	7,4	48326,0	EFX74629	No predicted function	*Daphnia pulex*	432	7,1E-44											X
44466	fr3	9,8	49527,7		No detected homolog								X				X	X	X
44520	fr6	9,7	64520,9		No detected homolog													X	X
44616	fr3	9,2	91844,4	EKC20363.1	No predicted function	*Crassostrea gigas*	503	4,7E-49									X		
44678	fr6	6,9	73994,3	ref|XP_002435481.1|	angiotensin-I converting enzyme,	*Ixodes scapularis*	654	1,0E-73					X						
44693	fr2	9,4	30998,2		No detected homolog													X	
44819	fr3	10,3	21812,0	gb|ADY80039.1|	thymosin	*Procambarus clarkii*	145	3,0E-09					X						
44875	fr2	5,6	26508,3		No detected homolog							X							
44983	fr2	9,9	70214,9	gb|EFX69070.1|	No predicted function	*Daphnia pulex*	464	6,0E-47				X						X	
45051	fr1	10,3	34807,0	gb|ADL14588.1|	catalase	*Hyriopsis cumingii*	704	1,0E-85										X	
45104	fr3	10,3	29563,8	gb|ACG63696.1|	myosin heavy chain	*Gammarus duebeni*	332	2,0E-32										X	X
45221	fr1	9,5	54782,9	Q94571	No predicted function	*0*	813	2,0E-99											X
45258	fr3		91730,0	gb|AEC50083.1|	alpha-2-macroglobulin 2 isoform 3	*Pacifastacus leniusculus*	802	2,0E-87										X	X
45296	fr3	8,8	27576,8	gb|ADL14588.1|	catalase	*Hyriopsis cumingii*	732	7,0E-91										X	
45331	fr1	9,5	26946,7	ADD82543	catalase	*Eriocheir sinensis*	690	6,8E-84											X
45375	fr4	10,7	33728,9	AGZ63441	catalase	*Octopus vulgaris*	530	1,9E-59	X				X					X	X
45519	fr4	9,4	43080,0		No detected homolog								X					X	X
45660	fr2	6,4	50850,5	AGM37865	family 7 cellobiohydrolase	*Chelura terebrans*	844	2,7E-104				X	X					X	X
45864	fr1	9,7	38752,1	AAT06185.1	elongation factor 1 alpha	*Metridium senile*	437	2,6E-46							X		X		X
45867	fr6	10,6	34147,5		No detected homolog								X						
46083	fr3	9,2	52098,5	gb|EFX87104.1|	myosin heavy chain isoform 1	*Daphnia pulex*	1627	0,0E+00							X		X	X	X
46143	fr1	9,9	118952,0		No detected homolog												X		
46175	fr6	8,2	33104,0	gb|ENN81074.1|	No predicted function	*Dendroctonus ponderosae*	201	1,0E-15										X	
46249	fr5	10,0	62001,8	ref|XP_001951517.1|	fructose-bisphosphate aldolase-like isoform 1	*Acyrthosiphon pisum*	516	4,0E-56				X	X					X	X
46252	fr2	8,7	58794,5		No detected homolog													X	X
46313	fr4	10,0	33625,0		No detected homolog								X						
46326	fr4	6,3	27598,9	ref|XP_003145362.1|	Actin	*Loa loa*	855	2,0E-113				X	X		X		X	X	X
46501	fr6	5,0	40966,2	gb|AFN20597.1|	lectin 1	*Macrobrachium rosenbergii*	186	1,0E-12				X	X					X	X
46567	fr4	8,8	37081,0	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	686	1,0E-80				X	X		X		X		
46702	fr1	9,5	124074,9	gb|ELT92693.1|	No predicted function	*Capitella teleta*	1353	1,0E-162				X			X				
46762	fr6	6,6	26334,7	ref|XP_001351127.1|	No predicted function	*Plasmodium falciparum 3D7*	114	1,0E-03					X					X	
46841	fr4	9,3	37882,0	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	613	5,0E-70										X	
46903	fr1	9,0	27408,0	gb|AFQ60003.1|	polyubiquitin, partial	*Solen grandis*	668	1,0E-83				X	X						
46934	fr2	10,0	40837,3		No detected homolog								X						X
47272	fr2	8,7	71108,9	CAA71673	Chymotrypsin BII	*Litopenaeus vannamei*	354	1,8E-34											X
47294	fr1	9,3	35401,8	XP_004748769.1	Actin	*Mustela putorius furo*	941	1,8E-121	X			X	X		X		X	X	X
47342	fr2	9,1	34512,0	BAG65416.1	No predicted function	*Homo sapiens*	616	8,9E-74									X		X
47389	fr4	7,3	33617,5	ref|XP_001351525.1|	erythrocyte membrane-associated antigen,	*Plasmodium falciparum 3D7*	132	1,0E-05					X					X	X
47426	fr4	10,8	23301,8	ref|XP_002166291.1|	No predicted function	*Hydra magnipapillata*	102	9,0E-03										X	
47446	fr5	7,9	21635,8	ref|XP_001351127.1|	No predicted function	*Plasmodium falciparum 3D7*	109	3,0E-03										X	
47479	fr5	8,7	58729,0	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	1138	8,0E-145				X	X		X		X	X	X
47577	fr5	10,5	33706,9	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	504	7,0E-55				X	X		X		X		
47660	fr5	9,3	13625,6	ref|XP_001844560.1|	zinc metalloproteinase nas-12 [Culex quinquefasciatus]gb|EDS	*Culex quinquefasciatus*	110	2,0E-04			X									
47680	fr4	8,5	63759,9	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	825	8,0E-98									X	X	X
47690	fr4	9,3	40525,0	sp|P45521.1|ACT_PROCL	Actin	*Uncharacterized*	608	2,0E-72									X	X	X
47791	fr2	7,9	47295,6	ref|XP_003773573.1|	immunoglobulin superfamily member 22	*Sarcophilus harrisii*	128	8,0E-05					X						
47886	fr2	9,4	33248,3	XP_004748769	Actin	*Mustela putorius furo*	518	1,4E-58											X
47943	fr4	9,8	38779,6	ref|XP_002013173.1|	No predicted function	*Drosophila persimilis*	490	2,0E-54									X	X	X
47948	fr6	8,8	67527,8	emb|CAX46431.1|	hemocyanin subunit 4 precursor	*Odontodactylus scyllarus*	882	6,0E-106	X			X	X		X		X	X	X
47954	fr3	8,9	32250,0	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	347	1,0E-33										X	
47963	fr2	10,0	25236,7	ABG27014.1	Actin	*Artemia franciscana*	323	3,2E-33									X		X
47965	fr6	9,7	31905,9	emb|CAX46429.1|	hemocyanin subunit 2 precursor	*Odontodactylus scyllarus*	442	6,0E-47									X	X	X
47988	fr6	6,6	46101,1	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	945	1,0E-117				X	X						X
48065	fr2	5,6	16586,9		No detected homolog				X										
48103	fr2	6,6	68245,4	gb|AAM81357.1|AF522504_:	hemocyanin	*Pacifastacus leniusculus*	892	3,0E-107									X	X	X
48119	fr1	9,2	30977,9	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	490	3,0E-53	X			X	X		X		X	X	
48164	fr2	8,6	29845,0		No detected homolog							X							
48166	fr4	9,1	50497,4	gb|ABB59715.1|	hemocyanin	*Cyamus scammoni*	608	8,0E-68							X		X	X	X
48222	fr4	6,2	16269,0	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	305	1,0E-29	X				X					X	X
48266	fr3	9,8	49998,3	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	413	4,0E-41							X		X	X	X
48351	fr4	9,2	52334,7	CAI78901	hemocyanin subunit 1	*Gammarus roeselii*	912	1,0E-111											X
48395	fr6	8,8	31443,1	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*	726	4,3E-87							X		X		
48484	fr2	8,8	58703,6	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	432	5,0E-43							X		X	X	X
48631	fr4	9,5	35971,5	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	447	6,0E-47										X	
48693	fr3	8,7	22397,8	gb|EJW72694.1|	Actin	*Wuchereria bancrofti*	552	1,0E-68				X			X		X	X	X
48696	fr6	8,4	14069,7	gb|AEB54622.1|	Actin	*Procambarus clarkii*	346	2,0E-39	X										
48717	fr1	8,8	50930,5	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	692	9,0E-80										X	
48739	fr4	9,8	36319,3	gb|AAC78681.1|	Actin	*Penaeus monodon*	699	8,0E-86					X				X	X	
48741	fr3	9,2	65320,4	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	686	1,0E-77	X			X	X		X			X	X
48760	fr1	5,1	36213,2	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	794	1,0E-96				X			X		X	X	X
48769	fr1	8,7	29299,1	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	449	5,0E-48									X	X	
48770	fr4	6,5	25924,4	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*	566	1,9E-64							X				X
48779	fr4	8,7	41447,4	AGR44824	Actin	*Bombyx mori*	1020	3,4E-133					X		X		X	X	X
48828	fr2	9,0	42091,2	gb|AEL29868.1|	Actin	*Charybdis japonica*	544	7,0E-64							X		X	X	X
48854	fr2	9,2	36529,1	gb|ACD39389.1|	beta actin	*Bufo gargarizans*	493	6,0E-58					X						
48864	fr1	10,5	16008,3	ref|XP_002838338.1|	No predicted function	*Tuber melanosporum Mel28*	167	1,0E-11										X	
48868	fr5	9,0	54414,1	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	754	1,0E-88	X			X	X		X		X	X	X
48876	fr6	8,2	42498,8	CAI78901	hemocyanin subunit 1	*Gammarus roeselii*	950	1,2E-118											X
48880	fr4	6,2	27537,8	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*	522	4,2E-58									X		
48901	fr2	7,7	34734,8	gb|AFI60316.1|	heat shock protein 70	*Eulimnogammarus verrucosus*	1093	1,0E-141	X			X	X		X		X	X	X
49014	fr4	9,8	48512,2		No detected homolog							X							
49094	fr6	4,9	11152,3		No detected homolog							X			X		X		
49115	fr1	9,9	38463,4	gb|ADV40137.1|	Actin	*Latrodectus hesperus*	682	2,0E-83				X						X	
49117	fr6	9,5	14959,4	emb|CAA38618.1|	Actin	*Drosophila melanogaster*	362	1,0E-41	X										
49119	fr4	8,6	36922,4	gb|ACZ60616.1|	beta-actin	*Panulirus argus*	687	9,0E-86				X							
49126	fr2	9,3	39631,4	gb|AAL78164.1|	Actin	*Neohelice granulata*	645	1,0E-78									X	X	X
49153	fr2	9,4	34905,2	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	554	7,0E-62					X					X	
49159	fr2	8,4	20783,1	ref|XP_002031056.1|	No predicted function	*Drosophila sechellia*	638	1,0E-81	X				X						
49212	fr6	9,0	40483,1	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	385	4,0E-38					X						
49231	fr6	6,4	43209,3	ABB59715.1	hemocyanin	*Cyamus scammoni*	936	1,4E-116									X		X
49244	fr3	9,8	39216,9	dbj|BAG65416.1|	No predicted function	*Homo sapiens*	1013	4,0E-133				X	X		X		X	X	X
49262	fr1	11,0	17072,7	emb|CAG10388.1|	No predicted function	*Tetraodon nigroviridis*	216	4,0E-18				X	X						
49275	fr3	9,2	42178,4	EMC90781.1	heat shock protein 70	*Columba livia*	297	1,7E-29							X				X
49276	fr4	9,1	50613,5	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	644	5,0E-73							X		X	X	X
49285	fr6	11,1	34648,8	gb|AAF22646.1|	skeletal alpha-actin	*Sparus aurata*	150	4,0E-08										X	
49289	fr4	9,1	40888,5	CAI78901	hemocyanin subunit 1	*Gammarus roeselii*	723	2,3E-85											X
49678	fr2	5,6	52780,0	AGM37865	family 7 cellobiohydrolase	*Chelura terebrans*	623	4,2E-71	X									X	X
49680	fr4	5,5	31081,8	AGM37865	family 7 cellobiohydrolase	*Chelura terebrans*	309	6,1E-29	X			X	X					X	X
49809	fr1	6,7	39666,7	AGM37865	family 7 cellobiohydrolase	*Chelura terebrans*	545	3,4E-61										X	X

50044	fr5	9,1	22095,3	gb|ADI59751.1|	early cuticle protein 3	*Callinectes sapidus*	170	3,0E-12	X								
50392	fr3	9,8	100268,8	gb|EHJ67857.1|	No predicted function	*Danaus plexippus*	1289	2,0E-151							X	X	X
50412	fr1	8,9	51459,8	gb|ACU31810.1|	alpha2 macroglobulin isoform 2	*Fenneropenaeus chinensis*	683	2,0E-75			X	X				X	X
50440	fr1	6,3	48358,7	emb|CAX46431.1|	hemocyanin subunit 4 precursor	*Odontodactylus scyllarus*	984	2,0E-123	X		X	X	X	X		X	X
50478	fr6	9,2	58058,2	gb|ELT97926.1|	No predicted function	*Capitella teleta*	851	8,0E-105			X		X				
50488	fr6	9,2	52320,3	gb|EFX76382.1|	No predicted function	*Daphnia pulex*	154	1,0E-07	X				X	X			
50513	fr3	10,3	40828,8	gb|ADK77876.1|	lipopolysaccharide and beta-1,3-glucan binding protein	*Portunus trituberculatus*	211	8,0E-16	X		X	X					
50566	fr5	8,3	46454,6		No detected homolog												X	X
50586	fr4	8,1	65439,0	gb|EKC26157.1|	Nose resistant to fluoxetine protein 6	*Crassostrea gigas*	182	4,0E-11			X						
50659	fr2	8,8	20973,9	ref|XP_312372.4|	No predicted function	*Anopheles gambiae str. PEST*	412	2,0E-45	X			X					
50669	fr2	8,8	36642,1	gb|ACN89260.1|	protein disulfide isomerase	*Litopenaeus vannamei*	691	5,0E-83							X	X	
50707	fr5	10,6	67305,6	ADC55251.1	ATP synthase subunit alpha precursor	*Litopenaeus vannamei*	747	3,0E-87								X	X
50753	fr5	8,8	38264,8	ref|XP_392401.1|	sorbitol dehydrogenase-like isoform 2	*Apis mellifera*	818	1,0E-103			X	X				X	
50923	fr1	10,7	46612,8	Q07233	No predicted function	*0*	223	2,9E-18									X
51033	fr5	9,3	68789,7	ref|XP_003398522.1|	2-oxoglutarate dehydrogenase, mitochondrial-like	*Bombus terrestris*	1814	0,0E+00	X			X					
51071	fr6	5,2	56188,6	sp|Q9GV46.1|LUCB_OPLGR	Oplophorus-luciferin 2-monooxygenase non-catalytic subunit	*Oplophorus gracilirostris*	262	5,0E-22								X	X
51145	fr6	9,7	58129,1	ref|XP_002080867.1|	No predicted function	*Drosophila simulans*	360	7,0E-34	X								
51195	fr5	9,1	35062,4	gb|EFX87975.1|	No predicted function	*Daphnia pulex*	197	1,0E-14			X						
51327	fr5	4,7	25410,4	ADK77876	lipopolysaccharide and beta-1,3-glucan binding protein	*Portunus trituberculatus*	456	4,5E-51									X
51377	fr1	10,3	53208,7	ref|XP_967224.1|	aspartate aminotransferase	*Tribolium castaneum*	319	3,0E-29				X				X	X
51397	fr1	9,1	43761,3		No detected homolog							X					
51491	fr2	9,6	83253,0	ref|NP_001017852.1|	trans-2-enoyl-CoA reductase, mitochondrial precursor	*Danio rerio*	619	2,0E-69	X		X	X				X	X
51502	fr1	9,0	63666,8	gb|EFR27142.1|	No predicted function	*Anopheles darlingi*	1240	2,0E-163				X					
51551	fr2	9,9	30059,0	gb|AEZ68611.1|	adenine nucleotide translocase	*Litopenaeus vannamei*	243	1,0E-20	X							X	
51714	fr3	9,2	48906,2	gb|ABF18489.1|	No predicted function	*Aedes aegypti*	214	2,0E-16	X								
51847	fr1	8,4	32294,0	ref|XP_624694.1|	GMP reductase 2-like isoform 1	*Apis mellifera*	823	4,0E-105			X						
52227	fr6	9,3	37109,4	ref|XP_002432772.1|	Soluble NSF attachment protein,	*Pediculus humanus corporis*	716	2,0E-89			X						
52368	fr5		57880,7	gb|EFX89414.1|	No predicted function	*Daphnia pulex*	764	4,0E-92				X					
52433	fr4	6,0	57518,1	gb|EFX67825.1|	No predicted function	*Daphnia pulex*	628	1,0E-73	X								
52575	fr4	6,6	55997,0	ref|XP_003704920.1|	heterogeneous nuclear ribonucleoprotein H-like	*Megachile rotundata*	878	2,0E-109				X					
52804	fr1	4,7	61444,4		No detected homolog				X								
53019	fr4	10,2	53604,4	XP_002423352	No predicted function	*Pediculus humanus corporis*	749	4,5E-84									X
53462	fr5	10,5	14656,6	ref|XP_001200730.1|	40S ribosomal protein S28-like	*Strongylocentrotus purpuratus*	319	1,0E-35				X					
55690	fr6	6,3	43002,9	gb|EFX88799.1|	No predicted function	*Daphnia pulex*	1264	3,0E-165								X	
57175	fr1	9,0	16941,0	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	251	3,0E-22	X		X	X	X	X		X	
57964	fr5	9,6	41482,0		No detected homolog				X							X	
58024	fr5	5,1	22952,1		No detected homolog						X		X				
58284	fr2	4,7	19001,6	gb|EFX78004.1|	No predicted function	*Daphnia pulex*	203	2,0E-15	X			X					
58650	fr5	4,7	33054,8	gb|EFX84793.1|	No predicted function	*Daphnia pulex*	656	4,0E-74				X					
58988	fr5	10,4	23156,2	dbj|BAF98646.1|	prophenoloxidase	*Fenneropenaeus chinensis*	350	6,0E-35								X	
61104	fr4	6,3	33121,8	dbj|BAB64298.1|	Kettin	*Procambarus clarkii*	1070	6,0E-129	X			X					
61308	fr1	10,0	21143,6	gb|EKC20565.1|	Glucosylceramidase	*Crassostrea gigas*	134	5,0E-07	X			X					
61558	fr6	6,8	15652,1	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	486	4,0E-53				X					
62638	fr5	9,1	32792,6	EFX74525	No predicted function	*Daphnia pulex*	611	4,2E-75									X
62747	fr2	7,4	34868,3	gb|EFN78871.1|	No predicted function	*Harpegnathos saltator*	396	6,0E-39	X								
63604	fr6	4,6	16714,5	EFX77181	No predicted function	*Daphnia pulex*	231	4,1E-19									X
65174	fr2	9,1	19867,3	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	641	3,0E-73				X					
65834	fr4	11,4	16456,1	ref|XP_004065634.1|	histone H3.2-like	*Oryzias latipes*	714	4,0E-94			X	X	X	X		X	X
65907	fr3	8,4	25687,0	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	832	8,0E-98	X			X					
65991	fr5	5,4	12386,6	gb|EFN64011.1|	L-xylulose reductase	*Camponotus floridanus*	314	4,0E-33								X	
66440	fr3	9,1	19507,0	ref|XP_002133684.1|	No predicted function	*Drosophila pseudoobscura pseud*	523	7,0E-60	X			X					
67263	fr2	10,0	11551,9		No detected homolog								X		X		
68130	fr4	8,9	16610,8	gb|EEZ99980.1|	No predicted function	*Tribolium castaneum*	224	4,0E-18	X		X	X					
68651	fr2	8,8	18457,0	ref|XP_002408828.1|	No predicted function	*Ixodes scapularis*	179	2,0E-12				X					
69698	fr2	5,8	20990,6	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	775	7,0E-91	X								X
69913	fr6	4,6	20738,9	gb|EFX83489.1|	No predicted function	*Daphnia pulex*	722	4,0E-84	X		X	X	X			X	X
70282	fr6	5,3	11719,7	dbj|BAJ23880.1|	glycogen synthase	*Marsupenaeus japonicus*	359	5,0E-39	X		X	X					
70287	fr5	5,0	9124,4	dbj|BAB64298.1|	Kettin	*Procambarus clarkii*	337	3,0E-34	X			X					
70317	fr1	6,1	10205,0	gb|EFN78871.1|	No predicted function	*Harpegnathos saltator*	169	9,0E-12	X			X					
70483	fr5	7,0	18660,7	ref|XP_001978294.1|	No predicted function	*Drosophila erecta*	465	2,0E-51								X	
71063	fr4	6,3	42492,5	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	584	3,0E-65	X		X	X	X	X		X	X
72484	fr6	10,2	16598,6	gb|ABX44790.1|	40S ribosomal protein RPS17	*Flustra foliacea*	388	5,0E-45			X	X		X			X
73530	fr3	4,8	9212,5	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	378	1,0E-39				X					
73635	fr4	4,9	7740,5	ref|XP_003694238.1|	basement membrane-specific heparan sulfate proteoglycan cc	*Apis florea*	207	4,0E-17								X	
81896	fr4	9,3	9407,0	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	318	1,0E-31	X								
82142	fr1	8,8	19280,2	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	677	4,0E-78	X			X					
82548	fr5	8,4	18547,2	gb|EJW74021.1|	No predicted function	*Wuchereria bancrofti*	548	2,0E-68				X					
84749	fr6	8,0	7107,9		No detected homolog						X		X		X		
85980	fr2	4,1	5970,4	ref|XP_794373.3|	basement membrane-specific heparan sulfate proteoglycan cc	*Strongylocentrotus purpuratus*	171	1,0E-12								X	X
86662	fr5	5,4	38679,6		No detected homolog							X					
87598	fr2	9,4	82412,9	gb|ADF81052.1|	vitellogenin-like protein precursor	*Gammarus fossarum*	179	1,0E-11			X	X	X		X		
87602	fr1	6,0	51889,5	gb|AEC48728.1|	alpha-tubulin	*Eriocheir sinensis*	1021	5,0E-131			X	X					
87607	fr3	10,2	52334,0	AAC05737	myosin heavy chain	*Drosophila virilis*	651	5,1E-77									X
87610	fr6	5,0	16803,9		No detected homolog							X					
87621	fr5	9,0	75566,4	gb|AAN77377.1|	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	1398	3,0E-177				X					
87676	fr5	8,9	35698,3	ref|ZP_02716404.1|	zinc metalloprotease ZmpB	*Streptococcus pneumoniae CDC0.*	116	2,0E-03				X					
87677	fr3	9,7	74893,4	gb|ACR78686.1|	paramyosin	*Rimicaris exoculata*	484	2,0E-52	X			X				X	X
87679	fr6	9,0	73974,2		No detected homolog						X						
87717	fr4	8,0	58870,4	ref|NP_001093090.1|	NADPH-specific isocitrate dehydrogenase	*Bombyx mori*	951	5,0E-120	X		X	X				X	X
87769	fr2	9,4	81277,2	gb|ABU41117.1|	metalloproteinase	*Lepeophtheirus salmonis*	185	2,0E-12				X				X	X
87813	fr4	11,1	53934,4	gb|ABF19573.1|	catecholamine-regulated protein 40	*Homo sapiens*	691	1,0E-82				X					
87821	fr6	9,2	24965,5	EGW10424	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate syntheta	*Cricetulus griseus*	719	1,8E-92									X
88060	fr4	10,5	79184,4		No detected homolog							X					
88066	fr3	9,0	40432,5	gb|ADR30117.1|	glutathione S-transferase delta	*Locusta migratoria*	473	1,0E-53								X	
88225	fr5	10,2	68618,6	ref|XP_001950262.1|	2-oxoglutarate dehydrogenase, mitochondrial-like isoform 1	*Acyrthosiphon pisum*	1095	5,0E-133				X					
88280	fr1	9,3	57154,1	gb|ACL77780.1|	ATP synthase b	*Spodoptera exigua*	591	3,0E-69	X		X	X					
88301	fr2	9,0	22113,3	gb|EFR23045.1|	No predicted function	*Anopheles darlingi*	148	1,0E-08								X	
88325	fr3	10,2	34828,6	ref|XP_001602648.2|	phosphoglycerate mutase 2-like	*Nasonia vitripennis*	501	6,0E-58	X								
88538	fr5	5,7	14023,4	ref|ZP_01160790.1|	cold shock-like protein cspG	*Photobacterium sp. SKA34*	156	2,0E-11	X		X	X					
88549	fr3	7,6	52935,6	gb|EGF83039.1|	No predicted function	*Batrachochytrium dendrobatidis.*	979	3,0E-125				X					
88566	fr1	9,3	46791,1	ref|XP_002593173.1|	No predicted function	*Branchiostoma floridae*	438	5,0E-48	X			X					
88709	fr6	7,3	42559,7	ref|NP_001025928.1|	hsc70-interacting protein	*Gallus gallus*	541	9,0E-62				X					
88904	fr6	5,7	22329,7	ref|XP_001950806.2|	No predicted function	*Acyrthosiphon pisum*	195	1,0E-14	X								
88959	fr1	10,0	25829,2	gb|ADY16617.1|	ribosomal protein S3a	*Fenneropenaeus merguiensis*	458	2,0E-52			X	X					
89179	fr2	5,2	77723,5	gb|EFX89400.1|	No predicted function	*Daphnia pulex*	795	3,0E-91				X					
89378	fr3	8,9	41713,6		No detected homolog				X								
90243	fr4	7,3	44604,2	gb|ENN77623.1|	No predicted function	*Dendroctonus ponderosae*	931	9,0E-117	X		X	X					
90255	fr3	9,7	14091,3	emb|CAF87569.1|	No predicted function	*Tetraodon nigroviridis*	353	2,0E-39				X					
90711	fr1	9,6	29248,8	gb|ABI52648.1|	ubiquitin/40S ribosomal protein S27a fusion protein	*Argas monolakensis*	638	1,0E-79								X	X
92200	fr4	9,3	45981,2	ref|XP_001867188.1|	bifunctional aminoacyl-tRNA synthetase	*Culex quinquefasciatus*	515	1,0E-53				X					
94514	fr5	4,9	27632,6	XP_002026190	No predicted function	*Drosophila persimilis*	693	1,3E-82									X
94725	fr5	8,9	30425,3	ref|XP_001688736.1|	No predicted function	*Anopheles gambiae str. PEST*	507	4,0E-60								X	
97936	fr6	9,6	41452,6	XP_005309289.1	Actin	*Chrysemys picta bellii*	435	1,7E-48							X		X
98010	fr6		45204,2	gb|EFX87104.1|	myosin heavy chain isoform 1	*Daphnia pulex*	971	2,0E-114			X	X	X		X	X	X
98012	fr3	9,8	58863,1	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	148	6,0E-07				X				X	X
98015	fr5	9,7	43050,8	emb|CAB85963.1|	alpha glucosidase	*Litopenaeus vannamei*	259	3,0E-21				X					
98023	fr1	7,6	49476,2	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	1327	2,0E-177			X	X	X		X	X	X
98110	fr5	9,7	163924,5	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	2063	0,0E+00				X				X	X
98117	fr2	9,5	43619,9	ref|XP_002023000.1|	No predicted function	*Drosophila persimilis*	821	1,0E-104	X								
98134	fr4	9,3	50613,1	gb|AAQ72746.1|	sodium/potassium ATPase alpha subunit	*Caprella sp. FEA-2003*	708	5,0E-85	X							X	
98145	fr6	8,3	71256,7	gb|EHJ72120.1|	collagen alpha-2 precursor	*Danaus plexippus*	645	1,0E-67	X			X				X	
98152	fr5	8,8	22278,1	gb|ACS36538.1|	elongation factor 2	*Homarus americanus*	652	2,0E-76							X	X	X
98163	fr6	9,7	44377,1	ref|XP_002003905.1|	No predicted function	*Drosophila mojavensis*	555	1,0E-59			X						
98188	fr3	5,8	50683,2	gb|EAW79627.1|	ATPase, H+ transporting, lysosomal 70kDa, V1 subunit A, isofo	*Homo sapiens*	1303	3,0E-173			X					X	X
98247	fr5	9,4	78548,8	XP_005181522.1	protein disulfide-isomerase	*Musca domestica*	1029	2,5E-128	X		X	X			X	X	X
98260	fr1	6,0	95174,6	gb|EFX71627.1|	No predicted function	*Daphnia pulex*	837	4,0E-96								X	X
98464	fr4	7,8	36433,5	gb|ABD77231.1|	ATP synthase, H+ transporting F1 complex beta subunit	*Dasypus novemcinctus*	1073	3,0E-143			X	X	X		X	X	X
98545	fr1	9,6	66103,2	sp|Q0MQI4.1|NDUV1_PONP	NADH dehydrogenase	*Pongo pygmaeus*	1307	7,0E-172	X		X						
98571	fr3	10,2	68351,2	gb|ADI58834.1|	medium-chain acyl-CoA dehydrogenase	*Chiloscyllium plagiosum*	456	6,0E-49			X						
98589	fr1	8,7	56849,3	gb|AEB54655.1|	triosephosphate isomerase	*Procambarus clarkii*	692	4,0E-84	X			X					
98633	fr1	10,2	46421,0		No detected homolog						X						
98641	fr2	9,2	33532,9	gb|EFX73635.1|	No predicted function	*Daphnia pulex*	439	8,0E-49				X					
98669	fr5	10,3	100820,7	ref|XP_001652300.1|	phosphofructokinase	*Aedes aegypti*	668	2,0E-71				X					
98980	fr4	9,2	31697,2	gb|AEF33837.1|	prohibtin protein	*Cherax quadricarinatus*	444	1,0E-49				X					
99109	fr3	9,3	48221,5		No detected homolog					X							
99411	fr2	9,7	41841,1		No detected homolog				X		X	X				X	X
99449	fr2	5,0	45708,9	ref|XP_312645.4|	No predicted function	*Anopheles gambiae str. PEST*	521	1,0E-57				X					
99610	fr3	8,4	23507,9	ref|XP_970004.1|	NADH:ubiquinone oxidoreductase ESSS subunit	*Tribolium castaneum*	298	1,0E-30				X					
100102	fr1	9,4	19303,6	gb|AAF22646.1|	skeletal alpha-actin	*Sparus aurata*	266	9,0E-25				X	X			X	
100208	fr1	5,0	48198,6		No detected homolog							X					

100255	fr4	7,8	50169,0	pir||JC8026	cytochrome P450 enzyme, CYP4C39 enzyme	*Carcinus maenas*	998	5,0E-127					X						
100307	fr3	10,1	16075,5	ABG81986.1	ubiquitin/ribosomal fusion protein	*Diaphorina citri*	368	5,2E-42								X			
100349	fr3	11,0	16836,0	XP_001339600.2	histone H4	*Danio rerio*	517	2,4E-64	X			X	X		X	X	X	X	X
100405	fr3	11,7	17837,4	ref|NP_001119698.1|	ribosomal protein L24	*Acyrthosiphon pisum*	134	9,0E-08								X	X		
100645	fr3	7,7	52729,8	gb|AEY81221.1|	plasma membrane calcium ATPase	*Spodoptera littoralis*	905	1,0E-106	X				X						
100990	fr1	10,0	22554,6	ref|XP_001200730.1|	40S ribosomal protein S28-like	*Strongylocentrotus purpuratus*	325	1,0E-35				X	X						
101072	fr3	8,9	27043,0	ref|XP_002733503.1|	Estradiol 17-beta-dehydrogenase 12-B-like	*Saccoglossus kowalevskii*	423	1,0E-46					X						
101079	fr6	9,1	80369,9	ESO99517	No predicted function	*Lottia gigantea*	701	2,8E-74					X					X	X
101957	fr1	9,1	24826,0	AGM37865.1	family 7 cellobiohydrolase	*Chelura terebrans*	412	5,6E-44								X		X	X
102024	fr2		16358,0	gb|ACR78689.1|	cuticle protein	*Rimicaris exoculata*	149	9,0E-10										X	
102192	fr5	8,3	21914,3	NP_001191331	ribose-phosphate pyrophosphokinase 1 isoform 2	*Homo sapiens*	376	2,4E-42											X
102235	fr2	9,9	40684,4	ref|XP_001951601.1|	fumarate hydratase, mitochondrial-like	*Acyrthosiphon pisum*	647	7,0E-76	X										
102540	fr6	6,1	52780,9	EKC22950	Bifunctional aminoacyl-tRNA synthetase	*Crassostrea gigas*	1444	0,0E+00											X
103101	fr6	5,0	26074,6		No detected homolog				X										
103291	fr4	9,6	33634,9	gb|EFA11221.1|	No predicted function	*Tribolium castaneum*	597	1,0E-65										X	
103394	fr3	10,0	30551,7	sp|Q05187.1|TGMH_TACTR	Hemocyte protein-glutamine gamma-glutamyltransferase	*Tachypleus tridentatus*	414	9,0E-43										X	X
103910	fr2	7,8	29253,6		No detected homolog						X								
104321	fr3	9,5	33816,9	gb|EFX76852.1|	No predicted function	*Daphnia pulex*	354	3,0E-34			X								
104993	fr3	5,6	13518,8		No detected homolog				X										
105002	fr6	10,1	19941,4	gb|AAX48874.1|	L40-ubiquitin	*Suberites domuncula*	607	1,0E-77										X	
105020	fr5	5,5	41816,6	ref|XP_001607303.2|	myosin heavy chain, muscle isoform 1	*Nasonia vitripennis*	850	8,0E-99	X			X	X	X	X			X	X
105024	fr6	9,7	48914,6	ABU41134.1	vitellogenin 1	*Lepeophtheirus salmonis*	122	8,4E-04							X				
105032	fr3	9,6	78987,9	gb|AAN72821.1|	beta-I tubulin	*Callinectes sapidus*	842	3,0E-104										X	X
105035	fr3	9,1	29543,5	AFC60662	chitinase	*Pandalopsis japonica*	251	6,0E-21											X
105044	fr3	8,4	40746,9	gb|AAW84264.1|	heterogeneous nuclear ribonucleoprotein L, partial	*Aplysia californica*	350	3,0E-35			X	X		X			X	X
105105	fr5	9,6	54324,0	gb|ABB55269.2|	peroxinectin	*Fenneropenaeus chinensis*	592	3,0E-64	X		X								
105132	fr2	10,6	53183,5	emb|CAB56173.1|	Na/K-ATPase a1 subunit	*Artemia franciscana*	726	1,0E-87							X		X	X	X
105143	fr3	9,7	67090,0	gb|ENN79035.1|	No predicted function	*Dendroctonus ponderosae*	357	2,0E-32							X		X		
105148	fr3	5,0	13254,6	ABR01223.1	elongation factor 2	*Penaeus monodon*	494	1,1E-55							X				
105194	fr6	9,4	36342,5	ref|XP_002031864.1|	No predicted function	*Drosophila sechellia*	549	4,0E-63						X	X		X		X
105195	fr5	8,9	71530,6		No detected homolog						X								
105224	fr3	9,1	60873,2	ref|XP_003705126.1|	neurogenic locus notch protein homolog	*Megachile rotundata*	489	6,0E-52			X							X	
105272	fr4	8,2	46026,8	ref|XP_003703049.1|	peroxiredoxin-5, mitochondrial-like isoform 1	*Megachile rotundata*	338	3,0E-34			X							X	
105274	fr3	8,6	33730,7	gb|EFX87450.1|	No predicted function	*Daphnia pulex*	712	3,0E-87	X										
105311	fr3	9,5	51142,3	gb|EFX68668.1|	No predicted function	*Daphnia pulex*	1091	2,0E-140					X					X	
105312	fr1	9,1	60998,6	gb|EFX74207.1|	Cct5-prov protein	*Daphnia pulex*	1008	1,0E-126	X			X	X	X					
105322	fr2	8,5	52466,2	ADO00929	endoplasmic reticulum protein 57	*Penaeus monodon*	909	1,8E-113											X
105371	fr5	9,6	45410,3	ref|XP_001660027.1|	transaldolase	*Aedes aegypti*	534	1,0E-60										X	
105389	fr4	8,8	19207,3	BAN20387	calponin/transgelin	*Riptortus pedestris*	517	5,9E-63					X					X	X
105395	fr6	9,2	36471,7	gb|ACS44711.1|	hemocyanin	*Porcellio scaber*	462	1,0E-49				X	X	X					
105399	fr4	8,9	33382,6		No detected homolog						X			X	X				
105406	fr4	10,2	17873,0	EPQ64670.1	No predicted function	*Blumeria graminis f. sp. tritici 96.*	92	6,7E-02							X				
105465	fr4	9,9	41183,6	gb|ACA83751.1|	ribophorin I	*Penaeus monodon*	306	1,0E-27			X							X	
105472	fr5	7,2	36908,6	gb|EFA00566.1|	No predicted function	*Tribolium castaneum*	650	6,0E-78										X	X
105505	fr5	7,3	25917,0	gb|EFN78711.1|	Glucosamine-6-phosphate isomerase	*Harpegnathos saltator*	792	1,0E-102	X		X		X					X	
105524	fr5	9,8	41479,5	ref|XP_001949122.1|	kelch-like protein diablo-like	*Acyrthosiphon pisum*	186	3,0E-12					X						
105618	fr2	9,9	68508,3	XP_002432811	2-oxoglutarate dehydrogenase, putative	*Pediculus humanus corporis*	1151	3,3E-141											X
105625	fr2	10,9	33092,6	XP_004519624.1	THO complex subunit 6	*Ceratitis capitata*	325	4,1E-32								X			
105628	fr5	8,9	61708,3	ref|XP_001633389.1|	No predicted function	*Nematostella vectensis*	985	5,0E-125										X	X
105692	fr1	9,1	51084,0	ref|XP_001602648.2|	phosphoglycerate mutase 2-like	*Nasonia vitripennis*	529	2,0E-60			X		X					X	
105712	fr5	8,2	21526,9	ref|XP_003435825.1|	No predicted function	*Anopheles gambiae str. PEST*	643	1,0E-77								X		X	X
105721	fr1	6,7	38662,5	ref|XP_309462.4|	No predicted function	*Anopheles gambiae str. PEST*	177	6,0E-12	X										
105765	fr4	9,6	45401,5	ref|XP_002116301.1|	No predicted function	*Trichoplax adhaerens*	913	9,0E-117			X							X	X
105797	fr6	8,9	46027,7	ref|XP_004208927.1|	superoxide dismutase	*Cu-Zn*	275	6,0E-26	X										
105830	fr2	6,0	51062,8	ref|NP_001153808.1|	parvin	*Acyrthosiphon pisum*	630	7,0E-74	X				X						
105848	fr3	6,0	24141,0	ref|XP_966588.1|	No predicted function	*Tribolium castaneum*	676	9,0E-86	X				X						
105865	fr6	9,3	26479,4	ref|XP_002429108.1|	inositol monophosphatase family domain containing protein	*Pediculus humanus corporis*	485	6,0E-56			X								
105866	fr2	9,8	39624,1	gb|EFZ17063.1|	No predicted function	*Solenopsis invicta*	439	3,0E-46			X		X						
105868	fr4	9,5	34835,5	gb|AEB54629.1|	aspartate aminotransferase	*Procambarus clarkii*	566	1,0E-67	X					X		X			
105895	fr1	8,9	34482,1		No detected homolog						X								
105949	fr3	9,5	51986,1	ref|NP_001162016.1|	No predicted function	*Nasonia vitripennis*	584	5,0E-69	X										
106001	fr4	11,4	29333,1	gb|ACR78696.1|	ribosomal protein S3	*Rimicaris exoculata*	248	6,0E-23								X		X	X
106017	fr1	8,4	20102,1	AGM32607	inorganic pyrophosphatase, partial	*Coptotermes formosanus*	378	8,9E-41					X						X
106155	fr4	4,8	24833,6	gb|AAH43613.1|	COL5A2 protein	*Homo sapiens*	149	3,0E-08										X	
106201	fr2	9,3	18843,0	gb|ACX30746.1|	thioredoxin	*Fenneropenaeus chinensis*	271	2,0E-27	X									X	
106346	fr1	9,1	68046,9	ref|XP_001633483.1|	No predicted function	*Nematostella vectensis*	150	5,0E-08					X						
106378	fr4	10,5	23520,4	gb|ACV72062.1|	QM protein	*Penaeus monodon*	399	2,0E-44			X		X			X		X	
106445	fr4	9,0	24986,1	gb|AFW19996.1|	cytochrome c oxidase subunit Va	*Litopenaeus vannamei*	264	2,0E-25	X										
106464	fr4	9,9	23890,9	ref|XP_001850941.1|	NADH dehydrogenase	*Culex quinquefasciatus*	365	9,0E-41	X										
106506	fr2	10,7	38071,8	gb|AER42704.1|	dihydrolipoamide S-succinyltransferase (DLST, E2 component	*Epinephelus coioides*	551	4,0E-64	X										
106531	fr4	8,9	31706,8	gb|AFE19188.1|	plasma membrane calcium ATPase	*Callinectes sapidus*	895	6,0E-108										X	
106533	fr2	8,7	52417,3	gb|EKC36313.1|	S-formylglutathione hydrolase	*Crassostrea gigas*	839	5,0E-106				X	X						
106571	fr2	5,4	10778,5		No detected homolog						X								
106629	fr1	11,3	32853,3	gb|AAX62486.1|	ribosomal protein L7 isoform B	*Lysiphlebus testaceipes*	437	7,0E-49								X		X	
106636	fr2	9,2	22944,3	gb|EFX81723.1|	No predicted function	*Daphnia pulex*	187	2,0E-14			X		X						
106723	fr6	9,7	52961,2	gb|EFX89414.1|	No predicted function	*Daphnia pulex*	398	1,0E-39			X							X	
106795	fr3	5,9	54459,0	ref|XP_002173166.1|	No predicted function	*Schizosaccharomyces japonicus y*	144	1,0E-06			X		X					X	X
106932	fr5	9,0	34920,7	ref|XP_001607466.1|	single-stranded DNA-binding protein, mitochondrial-like	*Nasonia vitripennis*	301	5,0E-30		X									
106948	fr3	5,8	15566,7	ref|XP_002755726.1|	60S acidic ribosomal protein P2-like	*Callithrix jacchus*	137	2,0E-08	X				X						
107198	fr2	4,8	25642,4	emb|CCI71880.1|	gelsolin	*Homarus americanus*	605	8,0E-70	X				X						
107252	fr3	10,6	16986,6	ref|XP_001618731.1|	No predicted function	*Nematostella vectensis*	214	9,0E-20					X						
107301	fr2	10,3	47632,1	gb|ABG82044.1|	ALG-2 interacting protein x	*Penaeus monodon*	787	4,0E-92				X	X					X	
107314	fr5	8,5	30315,5	XP_004923968	phosphoglucomutase-like isoform X2	*Bombyx mori*	516	9,5E-58										X	X
107407	fr1	4,2	15836,0		No detected homolog				X										
107487	fr1	9,8	64995,5	ref|XP_392215.4|	No predicted function	*Apis mellifera*	556	5,0E-57					X						
107688	fr4	9,1	65569,2	dbj|BAH70887.1|	No predicted function	*Acyrthosiphon pisum*	518	2,0E-58			X								
107776	fr5	9,1	32788,2	tpg|DAA34329.1|	TPA_inf: splicing factor hnRNP-F	*Amblyomma variegatum*	548	1,0E-65			X								
108041	fr1	9,3	19705,6	gb|ACO11894.1|	Cytochrome c	*Lepeophtheirus salmonis*	313	1,0E-33	X				X					X	
108163	fr3	8,8	31831,6	ref|XP_002596725.1|	No predicted function	*Branchiostoma floridae*	440	1,0E-48					X						
108376	fr1	8,6	70974,1	gb|EFN78871.1|	No predicted function	*Harpegnathos saltator*	779	2,0E-85	X				X						
108701	fr3		38023,7	dbj|BAJ22990.1|	cytochrome c	*Marsupenaeus japonicus*	398	2,0E-44										X	X
109189	fr5	10,6	17868,3	ref|XP_966448.1|	ribosomal protein L24e	*Tribolium castaneum*	335	2,0E-36								X		X	X
109413	fr2	11,6	38181,9	gb|EFX84473.1|	No predicted function	*Daphnia pulex*	104	4,4E-02										X	X
109695	fr1	5,2	22937,6	ABF72887	hemolectin-like	*Belgica antarctica*	182	2,4E-13											X
110907	fr2	8,8	54006,7	gb|AFM54541.1|	Na+/K+_ATPase	*Exopalaemon carinicauda*	1584	0,0E+00	X			X	X	X	X			X	X
110912	fr3	9,2	39569,8	gb|AEX58670.1|	catalase	*Meretrix meretrix*	390	4,0E-39										X	X
110918	fr2	8,9	31420,9		No detected homolog						X		X	X	X			X	X
110946	fr1	8,5	48166,7	ref|XP_796773.3|	alpha-L-fucosidase-like	*Strongylocentrotus purpuratus*	584	2,0E-66	X				X					X	
110964	fr1	7,1	51404,4	ref|XP_003486248.1|	glycerol-3-phosphate dehydrogenase	*Bombus impatiens*	1165	4,0E-154	X				X					X	X
110976	fr3	9,9	45628,0	XP_005181947	muscle LIM protein Mlp84B-like	*Musca domestica*	466	7,3E-49	X									X	X
110978	fr6	9,8	28829,9	gb|ACZ58291.1|	elongation factor-1 alpha	*Cyamus ovalis*	899	7,0E-116						X		X		X	X
111022	fr3	8,8	43456,9	ACC76803	Lit v 3 allergen myosin light chain	*Litopenaeus vannamei*	247	1,4E-21											X
111094	fr4	8,9	78345,6	XP_001952392.2	delta-1-pyrroline-5-carboxylate dehydrogenase, mitochondria	*Acyrthosiphon pisum*	584	5,2E-63						X					
111171	fr2	9,1	60655,8	XP_001660338	xaa-pro aminopeptidase	*Aedes aegypti*	506	1,2E-52											X
111172	fr3	8,3	40929,7	sp|Q9GV46.1|LUCB_OPLGR	Oplophorus-luciferin 2-monooxygenase non-catalytic subunit	*Oplophorus gracilirostris*	204	7,0E-15					X						
111201	fr2	9,1	43647,8	ref|XP_971371.1|	No predicted function	*Tribolium castaneum*	259	1,0E-21				X		X	X				
111225	fr2	9,2	40845,1	gb|ADY90107.1|	protein disulfide isomerase A6	*Penaeus monodon*	835	8,0E-105				X	X		X		X	X	X
111317	fr6	5,6	43852,4	ref|XP_002425619.1|	T-complex protein 1 subunit theta,	*Pediculus humanus corporis*	927	2,0E-116				X	X		X				
111343	fr5	7,1	23977,9		No detected homolog							X							
111356	fr6	7,4	57827,6	gb|ACA83751.1|	ribophorin I	*Penaeus monodon*	1046	4,0E-132				X	X						
111362	fr2	9,2	30150,3	dbj|BAM19324.1|	No predicted function	*Papilio polytes*	392	3,0E-43					X						
111380	fr3	8,9	86464,4		No detected homolog							X							
111401	fr5	9,1	50644,1	ref|XP_001654149.1|	NADH dehydrogenase	*Aedes aegypti*	352	4,0E-36					X						
111405	fr2	9,8	27056,5	ref|XP_003708272.1|	28 kDa heat- and acid-stable phosphoprotein-like	*Megachile rotundata*	265	7,0E-25					X						
111417	fr1	9,5	38664,0	gb|ADK63101.1|	thrombospondin protein	*Penaeus monodon*	163	3,0E-09	X										
111453	fr5	9,2	58510,2	ref|ZP_17361430.1|	40-residue YVTN family beta-propeller	*Bacillus cereus BAG1X1-2*	106	4,9E-02										X	X
111464	fr2	8,9	29928,3	gb|EDL21529.1|	mCG10343, isoform CRA_d	*Mus musculus*	394	4,0E-42	X				X						
111486	fr3	5,5	28554,9	ref|ZP_07641372.1|	surface exclusion protein PrgA	*Streptococcus mitis SK597*	107	9,0E-03				X							
111526	fr5	9,5	57650,8	ref|XP_003387079.1|	xylose isomerase-like	*Amphimedon queenslandica*	865	9,0E-107										X	X
111529	fr5		33998,5	gb|AAO42752.1|	cytosolic manganese superoxide dismutase	*Farfantepenaeus aztecus*	227	1,0E-19										X	
111565	fr5	9,1	38373,4	XP_003689940	heterogeneous nuclear ribonucleoprotein Q-like	*Apis florea*	645	2,8E-74											X
111573	fr3	6,1	21891,3	gb|ACZ06791.1|	eukaryotic translation initiation factor 5A	*Scylla paramamosain*	458	2,0E-54	X									X	
111581	fr5	9,5	46633,5		No detected homolog						X			X					
111621	fr2	9,0	43490,7	ref|XP_004377525.1|	cytochrome c-like	*Trichechus manatus latirostris*	485	1,0E-56			X		X						
111671	fr6	5,2	46387,1		No detected homolog				X							X			
111697	fr2	6,1	55262,2	XP_004922073.1	methenyltetrahydrofolate synthase domain-containing protei	*Bombyx mori*	534	4,7E-58								X			
111710	fr5	9,8	30246,3	ref|XP_001655159.1|	aspartate aminotransferase	*Aedes aegypti*	630	2,0E-75			X		X					X	X
111711	fr6		33296,8	ref|XP_001655005.1|	ATP synthase delta chain	*Aedes aegypti*	243	3,0E-21										X	X
111749	fr2	9,7	42033,1	emb|CAQ60115.1|	glyceraldehyde-3-phosphate dehydrogenase	*Gammarus locusta*	629	4,0E-75	X			X	X	X	X			X	X
111802	fr3	5,2	29140,5	gb|ACS36544.1|	troponin C isoform 4'	*Homarus americanus*	348	4,0E-37	X				X						

111806	fr3	9,8	49911,9	gb|ABO26669.1|	psmc6 protein	*Haliotis discus discus*	1018	2,0E-133			X	X					
111876	fr3	8,3	43925,6		No detected homolog											X	X
111894	fr4	9,1	42882,3		No detected homolog				X								
111922	fr5	9,7	23225,8	ref|XP_002611703.1|	No predicted function	*Branchiostoma floridae*	504	4,0E-60						X	X		
111956	fr4	9,8	43851,7	ref|XP_003459489.1|	ribosome maturation protein SBDS-like isoform 1	*Oreochromis niloticus*	360	1,0E-36				X					
111972	fr3	5,7	52094,6	ref|XP_002423620.1|	Dolichyl-diphosphooligosaccharide--protein glycosyltransferas	*Pediculus humanus corporis*	1773	0,0E+00				X					
112001	fr6		37871,4	gb|EFX82640.1|	No predicted function	*Daphnia pulex*	479	2,0E-51			X						
112216	fr4	9,9	23828,6	ref|XP_003488683.1|	60S ribosomal protein L10-like	*Bombus impatiens*	818	9,0E-108								X	
112239	fr2	9,0	17420,4	BAN20115.1	ribosomal protein LP0	*Riptortus pedestris*	553	4,8E-67						X			
112257	fr4	10,4	23911,3	ref|XP_001604111.1|	40S ribosomal protein S8-like	*Nasonia vitripennis*	410	3,0E-46			X	X		X	X		
112353	fr4	6,7	46529,6	ref|XP_001960926.1|	No predicted function	*Drosophila ananassae*	1149	1,0E-152				X					
112455	fr3	5,6	31388,2	ref|XP_002432772.1|	Soluble NSF attachment protein,	*Pediculus humanus corporis*	829	3,0E-107	X			X				X	
112559	fr6	5,9	60271,1	gb|AFK49794.1|	Actin	*Cryptocercus punctulatus*	1815	0,0E+00								X	X
112586	fr4	9,2	30583,5	gb|ENN79515.1|	No predicted function	*Dendroctonus ponderosae*	112	2,0E-03	X								
112588	fr5	5,4	17028,5		No detected homolog						X						
112617	fr5	5,1	42992,5	ref|XP_002416560.1|	coiled-coil domain-containing protein,	*Ixodes scapularis*	562	3,0E-65	X								
112758	fr2	11,4	28046,1	gb|ABQ96198.1|	crustin isoform 2	*Litopenaeus schmitti*	162	8,0E-12	X								
113852	fr1	9,7	33617,6	gb|ABO26679.1|	hypoxanthine phosphoribosyltransferase 1	*Haliotis discus discus*	558	4,0E-67			X						
114115	fr5	6,9	74807,8		No detected homolog											X	
114513	fr5	5,0	29184,8	gb|EFX80115.1|	No predicted function	*Daphnia pulex*	474	8,0E-55	X								
114728	fr3	6,0	12523,2	CAI78901	hemocyanin subunit 1	*Gammarus roeselii*	309	1,9E-30									X
115747	fr4	9,3	37250,2	gb|ABC67745.1|	ribosomal protein L12	*Anthopleura elegantissima*	651	3,0E-81			X	X		X			
116225	fr4	11,8	35441,1	gb|EFX85184.1|	No predicted function	*Daphnia pulex*	565	2,0E-68				X		X		X	
116241	fr4	6,6	21971,8	ref|XP_002424248.1|	Troponin T	*Pediculus humanus corporis*	417	5,0E-46			X	X		X			
116243	fr4	9,9	60847,7	gb|AFV15454.1|	acyl-CoA delta-9 desaturase	*Eriocheir sinensis*	461	8,0E-49								X	X
116254	fr2	5,4	38805,3	ENN78731.1	No predicted function	*Dendroctonus ponderosae*	1051	3,1E-132					X				X
116334	fr5	5,6	39099,1	gb|EFX88976.1|	myosin regulatory light chain	*Daphnia pulex*	126	2,0E-05				X					
116336	fr1	11,0	40028,7	ref|XP_002196279.1|	aldehyde dehydrogenase, mitochondrial	*Taeniopygia guttata*	826	2,0E-102						X		X	
116338	fr2	10,4	28340,4	ref|XP_002131110.2|	malate dehydrogenase, cytoplasmic-like	*Ciona intestinalis*	221	7,0E-18				X					
116357	fr4	9,6	48596,8	ref|XP_001657588.1|	short-chain dehydrogenase	*Aedes aegypti*	470	1,0E-51								X	X
116383	fr1	6,4	30284,7	dbj|BAE90429.1|	No predicted function	*Macaca fascicularis*	438	6,0E-51					X	X		X	X
116386	fr6	10,2	61332,1	gb|ENN75717.1|	No predicted function	*Dendroctonus ponderosae*	476	1,0E-48								X	X
116392	fr2	5,4	47853,1	gb|AAC98687.1|	ATP sulfurylase/APS kinase 2	*Mus musculus*	1114	3,0E-143	X			X				X	
116403	fr1	10,1	75420,3	ref|XP_001809722.1|	Muscle-specific protein 300 CG33715-PD	*Tribolium castaneum*	127	3,0E-04			X	X		X		X	X
116420	fr4	10,5	32745,8	gb|ABQ52471.1|	mortalin splice variant	*Mya arenaria*	185	2,0E-12	X					X			
116463	fr3	9,9	54972,1	gb|AAT77811.1|	calpain B	*Gecarcinus lateralis*	790	1,0E-92				X					
116469	fr4	10,9	78309,0	gb|ADM67602.1|	V-type H+-ATPase subunit A	*Sepia officinalis*	105	2,6E-02	X								
116490	fr6	10,4	56434,0	ref|NP_649898.1|	CG8149	*Drosophila melanogaster*	136	5,0E-06	X			X					
116496	fr5	9,9	85539,4	ref|XP_002609968.1|	No predicted function	*Branchiostoma floridae*	228	3,0E-17			X						
116501	fr6	4,9	15851,6	gb|EFX72698.1|	No predicted function	*Daphnia pulex*	174	5,0E-12				X					X
116507	fr5	9,8	31423,4	gb|ADC55251.1|	ATP synthase subunit alpha precursor	*Litopenaeus vannamei*	736	1,0E-89						X		X	X
116538	fr6	10,1	51196,2	XP_780013.2	T-complex protein 1 subunit gammaisoform 1	*Strongylocentrotus purpuratus*	667	4,6E-77						X			
116584	fr3	8,8	64887,4	ref|NP_001040108.1|	chaperonin subunit 6a zeta	*Bombyx mori*	1078	1,0E-136			X	X				X	X
116691	fr6	9,1	55108,8	ref|XP_001654149.1|	NADH dehydrogenase	*Aedes aegypti*	543	2,0E-63	X								
116711	fr6	9,7	41514,2	ref|NP_001008171.1|	cystathionine-beta-synthase	*Xenopus (Silurana) tropicalis*	368	6,0E-36				X					
116761	fr2	10,2	27248,7	gb|EHJ69335.1|	acyl-CoA dehydrogenase	*Danaus plexippus*	516	8,0E-59			X						
116790	fr1	10,8	76161,2	ref|XP_001846164.1|	Gelsolin	*Culex quinquefasciatus*	757	6,0E-86	X							X	
116794	fr4	9,5	69779,4	gb|EFX86357.1|	No predicted function	*Daphnia pulex*	545	9,0E-56			X	X					X
116844	fr2	10,5	54992,2	gb|EFN79054.1|	Chromobox protein-like protein 5	*Harpegnathos saltator*	204	1,0E-14				X					
116889	fr5	8,8	25060,1	gb|AAY57407.1|	cytosolic MnSOD	*Litopenaeus vannamei*	411	2,0E-45								X	X
116996	fr2	10,9	24809,3	ref|XP_001867355.1|	small nuclear ribonucleoprotein Sm D1	*Culex quinquefasciatus*	256	2,0E-24			X	X	X	X		X	
117014	fr2	10,1	22709,1	gb|ACO15580.1|	ATP synthase subunit O, mitochondrial precursor	*Caligus clemensi*	272	5,0E-26								X	
117083	fr4	10,9	33678,0	ETE60281.1	60S ribosomal protein L7a	*Ophiophagus hannah*	229	9,2E-19						X		X	X
117102	fr1	9,8	20553,6	ref|XP_003396620.1|	ADP,ATP carrier protein 2-like	*Bombus terrestris*	479	7,0E-56				X	X	X		X	X
117114	fr5	9,0	43771,8	gb|ADO00929.1|	endoplasmic reticulum protein 57	*Penaeus monodon*	930	5,0E-118						X		X	X
117129	fr5	7,8	22158,0		No detected homolog				X								
117162	fr6	9,6	31908,9		No detected homolog				X								
117331	fr3	7,6	12616,1	gb|ACX30746.1|	thioredoxin	*Fenneropenaeus chinensis*	271	3,0E-28				X					
117464	fr5	9,2	17393,9	ACJ22896	60S acidic ribosomal protein P0	*Rana dybowskii*	382	8,2E-44									X
117577	fr2	9,5	16191,1	gb|ACT31598.1|	cytochrome c oxidase subunit Va protein-like protein	*Artemia franciscana*	267	9,0E-27	X								
117630	fr3	11,0	16554,3		No detected homolog											X	
117813	fr6	8,4	51855,5	XP_001658446.1	leucine rich protein, putative	*Aedes aegypti*	205	7,4E-14						X			X
117983	fr5	10,5	33219,9	sp|Q95V34.1|RS4_SPOFR	40S ribosomal protein S4	*Spodoptera frugiperda*	467	6,0E-53			X	X				X	
118014	fr3	4,8	35857,4	gb|EFX67825.1|	No predicted function	*Daphnia pulex*	715	5,0E-89				X					
119169	fr3	9,3	31652,9	gb|ENN79515.1|	No predicted function	*Dendroctonus ponderosae*	207	1,0E-15	X								
119204	fr3	9,6	31676,4	gb|EGI61842.1|	Adenylate kinase 2, mitochondrial	*Acromyrmex echinatior*	489	1,0E-56			X						
120326	fr3	5,5	25970,1	gb|ACR61397.1|	Gasp protein	*Choristoneura fumiferana*	593	6,0E-73	X								
120922	fr5	5,1	8384,8	ref|XP_003983099.1|	carboxypeptidase A2	*Felis catus*	176	2,0E-13	X		X	X					
121034	fr3	10,5	28403,2	XP_005702235.1	tubulin alpha-4A chain-like	*Capra hircus*	375	1,9E-40					X				
121043	fr1	5,6	29856,1	ref|XP_004015742.1|	periaxin	*Ovis aries*	154	1,0E-08	X		X						
121081	fr5	9,8	48837,3		No detected homolog									X		X	X
121109	fr1	5,8	30379,9		No detected homolog						X		X				
121125	fr3	6,6	21046,1	sp|P04069.1|CBPB_ASTFL	Carboxypeptidase B	*Uncharacterized*	265	2,0E-24			X	X		X			X
121141	fr1	9,2	51715,6	gb|EFX88976.1|	myosin regulatory light chain	*Daphnia pulex*	182	2,0E-12			X	X	X	X			
121166	fr4	9,5	36534,5	ref|XP_003386442.1|	No predicted function	*Amphimedon queenslandica*	101	9,7E-02				X				X	
121195	fr5	8,6	23415,8	ACY66376.1	heat shock protein 90	*Scylla paramamosain*	652	1,4E-79						X			
121213	fr3	4,8	51513,1	EFN78057	Glutamyl aminopeptidase	*Harpegnathos saltator*	487	4,5E-50									X
121218	fr5	8,4	35359,8	gb|ADB44901.1|	DEAD(Asp-Glu-Ala-Asp) box polypeptide 39	*Macrobrachium nipponense*	852	3,0E-108			X		X	X		X	X
121249	fr2	9,6	54432,8	XP_001600771.2	No predicted function	*Nasonia vitripennis*	170	8,0E-10						X			
121267	fr4	9,7	41956,8		No detected homolog							X				X	X
121276	fr1	8,7	32768,2	gb|EFA04932.1|	No predicted function	*Tribolium castaneum*	534	1,0E-61	X			X		X		X	X
121300	fr5	9,0	34498,5	gb|EFX69032.1|	malate dehydrogenase	*Daphnia pulex*	584	3,0E-69						X		X	X
121308	fr3	9,4	31658,8	gb|EFX70812.1|	No predicted function	*Daphnia pulex*	672	3,0E-80			X	X	X	X			
121324	fr1	10,6	49136,3	gb|AAT77811.1|	calpain B	*Gecarcinus lateralis*	464	1,0E-47								X	
121329	fr5	9,8	97115,4	gb|AGC26171.1|	17beta-HSD4	*Azumapecten farreri*	239	1,0E-17				X					
121351	fr5	9,9	25931,0	ref|XP_003701427.1|	upstream activation factor subunit spp27-like	*Megachile rotundata*	323	9,0E-33	X		X			X		X	X
121384	fr3	8,9	54855,6	gb|ADB44900.1|	DEAD(Asp-Glu-Ala-Asp) box polypeptide 5	*Macrobrachium nipponense*	1050	1,0E-133			X						
121407	fr1	9,0	36024,3	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	818	3,0E-100					X	X		X	X
121429	fr2	5,9	72265,7	NP_620276	echinoderm microtubule-associated protein-like 2	*Rattus norvegicus*	806	6,5E-94									X
121669	fr2	6,5	20787,4	emb|CAR85673.1|	cytoplasmic manganese superoxide dismutase	*Perisesarma bidens*	381	2,0E-41				X					
121674	fr1	8,0	18316,7	gb|AFJ11397.2|	prostaglandin F synthase	*Penaeus monodon*	302	4,0E-30			X	X					
121767	fr6	9,2	30122,8	ref|XP_313649.4|	No predicted function	*Anopheles gambiae str. PEST*	1011	7,0E-131			X		X				
121829	fr5	9,3	37052,5	AGM32184	ATP synthase oligomycin sensitivity conferral protein	*Coptotermes formosanus*	360	2,4E-37								X	X
121870	fr2	9,2	34377,1	emb|CAQ60115.1|	glyceraldehyde-3-phosphate dehydrogenase	*Gammarus locusta*	852	2,0E-109								X	
121871	fr3	9,8	36783,5	gb|ACU46011.1|	glutathione S-transferase	*Eriocheir sinensis*	390	6,0E-42			X					X	
121888	fr5	9,0	52215,2	ref|XP_002424350.1|	No predicted function	*Pediculus humanus corporis*	652	8,0E-77								X	X
121919	fr4	8,4	27062,4	ref|XP_002400043.1|	beta-N-acetylhexosaminidase	*Ixodes scapularis*	495	3,0E-55				X					
121924	fr5	8,6	40422,3	gb|AEE62005.1|	No predicted function	*Dendroctonus ponderosae*	427	1,0E-47								X	
121988	fr3	11,1	30669,4	ref|XP_003700261.1|	40S ribosomal protein S6-like	*Megachile rotundata*	582	2,0E-70			X	X		X			
121996	fr2	9,6	38380,1	gb|EMC82106.1|	Hsc70-interacting protein, partial	*Columba livia*	421	8,0E-46				X				X	
122081	fr5	9,0	38054,1	ref|XP_001604810.1|	cytochrome P450 CYP12A2	*Nasonia vitripennis*	227	1,0E-17			X						
122099	fr5	9,6	77264,4	ref|XP_003697765.1|	long-chain-fatty-acid--CoA ligase 3-like	*Apis florea*	881	8,0E-104				X					
122312	fr4	9,9	10043,3	XP_001371286.2	40S ribosomal protein S5	*Monodelphis domestica*	321	1,3E-35						X			
122435	fr1	9,1	15855,2	ref|XP_001943087.1|	No predicted function	*Acyrthosiphon pisum*	114	1,0E-04			X	X				X	
122649	fr4	11,3	22818,1	gb|ELT99918.1|	No predicted function	*Capitella teleta*	342	1,0E-36				X					
122754	fr5	9,9	20630,6	AEB54647	ribosomal protein S18	*Procambarus clarkii*	439	9,9E-52									X
123314	fr2	10,0	70928,2	gb|EFX75492.1|	No predicted function	*Daphnia pulex*	295	3,0E-25								X	
123927	fr1	5,0	36847,1	ref|XP_003468979.1|	gastrokine-1-like	*Cavia porcellus*	99	5,0E-02	X								
124058	fr4	8,5	22070,9	ref|XP_001652300.1|	phosphofructokinase	*Aedes aegypti*	676	3,0E-80				X					
124382	fr4	11,3	22167,0		No detected homolog											X	
125202	fr5	9,9	42936,7		No detected homolog						X						
125254	fr6	9,7	22623,1	gb|ELU01136.1|	No predicted function	*Capitella teleta*	250	7,0E-24			X						
125466	fr2	5,5	21892,0	ref|XP_967085.1|	No predicted function	*Tribolium castaneum*	757	8,0E-95	X			X					
125545	fr3	11,7	44351,9	ref|XP_001990119.1|	No predicted function	*Drosophila grimshawi*	345	5,0E-32					X	X		X	X
125555	fr4	9,8	32688,2	EFR27027	No predicted function	*Anopheles darlingi*	332	1,3E-31									X
125600	fr3	6,1	39911,0	gb|EFX86066.1|	No predicted function	*Daphnia pulex*	211	2,0E-16			X						
125607	fr6	10,4	56074,1	AAR01304.1	elongation factor-2	*Neogonodactylus oerstedii*	571	2,4E-61						X			
125640	fr6	9,6	28724,7	ref|XP_002196279.1|	aldehyde dehydrogenase, mitochondrial	*Taeniopygia guttata*	636	1,0E-75			X						
125644	fr6	9,5	55083,3	gb|ABG67961.1|	eukaryotic initiation factor 4A	*Callinectes sapidus*	1121	4,0E-146						X		X	X
125653	fr4	9,8	27365,4	ABV44280.1	heat shock protein 90	*Nilaparvata lugens*	332	1,3E-34					X				
125702	fr2	7,3	37305,5	gb|ACN89260.1|	protein disulfide isomerase	*Litopenaeus vannamei*	512	1,0E-56			X	X				X	
125722	fr2	5,2	31305,0	gb|ACP43727.1|	C-type lectin galactose-binding protein, partial	*Anopheles stephensi*	235	3,0E-20			X						
125730	fr1	8,8	69650,3	ref|XP_002019799.1|	No predicted function	*Drosophila persimilis*	510	5,0E-57	X		X	X				X	X
125741	fr1	9,5	66938,4	XP_001605758	T-complex protein 1 subunit epsilon-like	*Nasonia vitripennis*	1009	4,5E-126									X
125751	fr6	9,0	84194,8	dbj|BAH72354.1|	No predicted function	*Acyrthosiphon pisum*	961	1,0E-120	X			X		X			
125764	fr5	8,9	37684,2	gb|ADI40504.1|	H+-transporting ATP synthase F1 complex alpha subunit 1	*Scotophilus kuhlii*	1082	2,0E-144						X		X	X
125789	fr1	9,7	64660,2	gb|AET34917.1|	apoptosis inhibitor	*Macrobrachium rosenbergii*	520	5,0E-55						X		X	X
125804	fr4	9,0	56415,2		No detected homolog						X						
125832	fr3	9,3	36512,6	ref|XP_002164858.2|	No predicted function	*Hydra magnipapillata*	160	1,0E-10	X								
125877	fr1	8,5	41419,2	gb|EFN66295.1|	Glucosamine-6-phosphate isomerase	*Camponotus floridanus*	714	1,0E-88			X						

Annexe I

125896	fr3	8,9	37983,3	ref|XP_780172.1|	adenosylhomocysteinase A-like	*Strongylocentrotus purpuratus*	642	2,0E-76	X										
125911	fr1	9,8	104019,9		No detected homolog				X			X							
125919	fr3	8,5	38143,9	gb|EFA12023.1|	No predicted function	*Tribolium castaneum*	1359	0,0E+00	X			X							
125938	fr2	9,2	41984,1	ref|XP_003453266.1|	tryptophanyl-tRNA synthetase, cytoplasmic-like	*Oreochromis niloticus*	860	7,0E-108									X		
125959	fr3	8,4	60781,5	gb|EKC25746.1|	28 kDa heat- and acid-stable phosphoprotein	*Crassostrea gigas*	319	6,0E-30	X										
125989	fr3	9,4	54587,8	ref|XP_002038169.1|	No predicted function	*Drosophila sechellia*	308	7,0E-28				X		X					
126060	fr4	8,4	44093,9	ref|XP_003387079.1|	xylose isomerase-like	*Amphimedon queenslandica*	1161	5,0E-153						X			X		
126175	fr5	6,1	43302,3	gb|AFM75821.1|	myosin light chain 2	*Procambarus clarkii*	462	3,0E-52	X			X		X			X		
126199	fr1	7,6	47876,5	gb|AFN66645.1|	C-type lectin-like protein	*Fenneropenaeus chinensis*	128	2,0E-05						X					
126438	fr2	9,0	49563,8	ENN81916	No predicted function	*Dendroctonus ponderosae*	164	5,2E-09											X
126541	fr1	9,7	61157,5	EFX74372	No predicted function	*Daphnia pulex*	730	1,5E-85											X
126600	fr3	9,8	33107,4	ABE57139	heat shock protein Hsp21.7	*Liriomyza sativae*	189	5,0E-14											X
126792	fr4	9,0	44256,4	gb|EGI68534.1|	Glucose-6-phosphate 1-dehydrogenase	*Acromyrmex echinatior*	1141	3,0E-149	X										
126801	fr3	9,9	23205,0		No detected homolog				X										
127549	fr1	10,4	42532,2	gb|EFX61261.1|	No predicted function	*Daphnia pulex*	371	8,0E-36				X		X					
128499	fr4	8,4	31266,1	gb|EFX75492.1|	No predicted function	*Daphnia pulex*	965	1,0E-124									X		X
128511	fr6	9,6	13502,9	gb|EFX89003.1|	No predicted function	*Daphnia pulex*	342	2,0E-38	X			X		X		X		X	X
129051	fr2	4,9	34489,6	ref|XP_003247860.1|	collagen alpha-1(IV) chain-like	*Acyrthosiphon pisum*	741	6,0E-91									X		X
129661	fr6	5,1	7108,7	gb|ACY71124.1|	60S acidic ribosomal protein P2-like protein	*Scylla paramamosain*	145	4,0E-10	X			X		X					
129687	fr3	9,7	33464,7	gb|ACQ72823.1|	myosin heavy chain	*Bombyx mandarina*	458	6,0E-49								X	X		X
129747	fr4	8,9	108102,1	XP_001600229	serine/threonine-protein phosphatase 2A 65 kDa regulatory su	*Nasonia vitripennis*	677	6,8E-74											X
129814	fr2	6,8	67352,9	XP_002428942	collagen alpha-2 precursor, putative	*Pediculus humanus corporis*	856	4,0E-96	X								X		X
129831	fr3	11,2	14266,8	ref|XP_001627564.1|	No predicted function	*Nematostella vectensis*	95	5,0E-03								X	X		
129857	fr4	8,3	34113,9	EFN73041	No predicted function	*Camponotus floridanus*	108	2,0E-02											X
129865	fr5	9,4	32075,9	EFX89905.1	No predicted function	*Daphnia pulex*	201	1,3E-15							X				X
129900	fr3	9,9	38425,3	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	715	6,0E-87				X							
129908	fr3	4,7	40531,0	gb|EFZ14123.1|	No predicted function	*Solenopsis invicta*	234	9,0E-20				X			X	X			
129916	fr5	5,4	23369,0	gb|ACY43127.1|	arg methyltransferase	*Libinia emarginata*	814	9,0E-107				X		X		X			
129929	fr3	9,3	59565,1	ref|XP_002938634.1|	methylmalonate-semialdehyde dehydrogenase	*Xenopus (Silurana) tropicalis*	1803	0,0E+00				X		X			X		
129960	fr4	7,7	32814,6	gb|EFX72698.1|	No predicted function	*Daphnia pulex*	261	4,0E-22									X		X
130021	fr5	5,1	25729,7	CCE46010.1	DEAD-box ATP-dependent RNA helicase	*Cancer pagurus*	375	3,1E-38								X			
130034	fr5	5,9	34886,9	gb|ACM91675.1|	F1F0-ATP synthase beta subunit	*Fenneropenaeus chinensis*	1130	1,0E-148				X		X	X		X		X
130036	fr1	9,4	46680,9	ref|XP_003706377.1|	hrp65 protein-like	*Megachile rotundata*	507	1,0E-54						X		X	X		X
130062	fr5	9,8	80117,3	XP_004530495.1	T-complex protein 1 subunit zeta	*Ceratitis capitata*	918	4,5E-111				X				X	X		
130066	fr3	6,4	44245,1	ref|XP_780172.1|	adenosylhomocysteinase A-like	*Strongylocentrotus purpuratus*	1431	0,0E+00	X								X		X
130094	fr1	9,2	37328,7	XP_004931117.1	bifunctional purine biosynthesis protein PURH	*Bombyx mori*	760	6,4E-92							X				
130139	fr1	9,2	46242,2	gb|AFT92034.1|	triose-phosphate isomerase	*Litopenaeus vannamei*	697	9,0E-86	X					X		X	X		X
130148	fr4	9,3	32977,6	ref|XP_001506076.2|	phosphate carrier protein, mitochondrial-like	*Ornithorhynchus anatinus*	1012	5,0E-134	X					X			X		X
130214	fr4	10,3	39687,6	XP_004519624.1	THO complex subunit 7	*Ceratitis capitata*	314	5,2E-30				X		X	X	X	X		X
130227	fr1	8,6	36673,8	gb|ADM64456.1|	fatty acid binding protein	*Eriocheir sinensis*	254	2,0E-23	X			X		X	X				
130236	fr5	10,5	53478,2	ref|XP_002739772.1|	No predicted function	*Saccoglossus kowalevskii*	264	1,0E-23	X			X		X			X		X
130260	fr6	8,6	41348,3	ref|XP_002098233.1|	No predicted function	*Drosophila yakuba*	434	7,0E-46	X					X					
130273	fr6	6,3	44525,7		No detected homolog							X							
130339	fr3	8,7	81354,8	dbj|BAB13738.1|	DD5	*Marsupenaeus japonicus*	158	6,0E-08	X						X				
130435	fr3	9,9	58633,6	ref|NP_995755.1|	serpin 42Da, isoform J	*Drosophila melanogaster*	442	6,0E-46				X		X			X		
130712	fr5	11,1	41915,3	gb|EFX69073.1|	No predicted function	*Daphnia pulex*	254	4,0E-22						X					
130876	fr1	11,7	21694,1	gb|AFK93891.1|	adenine nucleotide translocase	*Penaeus monodon*	191	2,0E-14	X								X		
130914	fr4	9,0	27200,0		No detected homolog				X										
130943	fr5	5,6	27695,7	gb|EFA10819.1|	No predicted function	*Tribolium castaneum*	210	6,0E-18	X										
131115	fr6	9,4	24268,7	ref|XP_001950806.2|	No predicted function	*Acyrthosiphon pisum*	190	7,0E-14	X										
131219	fr4	7,5	23511,9	gb|EFX79782.1|	No predicted function	*Daphnia pulex*	98	8,5E-02						X					
131743	fr5	6,0	46929,3	dbj|BAB64297.1|	I-connectin	*Procambarus clarkii*	1993	0,0E+00	X					X					
131878	fr5	9,8	31785,8		No detected homolog														X
131949	fr2		27783,3	ref|XP_001862481.1|	No predicted function	*Culex quinquefasciatus*	190	4,0E-15				X		X					
132388	fr1	9,9	11419,0		No detected homolog												X		
133271	fr3		30377,2	ref|XP_001655422.1|	metalloproteinase, putative [Aedes aegypti]gb|EAT36357.1| A	*Aedes aegypti*	205	3,0E-16		X									
133468	fr5	11,5	24386,0	gb|AEB54658.1|	H+ transporting ATP synthase O subunit	*Procambarus clarkii*	239	9,0E-22									X		
133529	fr2	11,6	21081,3	CAJ17239	ribosomal protein L7e	*Scarabaeus laticollis*	391	4,5E-43											X
133650	fr5	9,5	34892,5	gb|EEZ99519.1|	No predicted function	*Tribolium castaneum*	464	3,0E-51				X			X	X	X		
133670	fr2	10,0	64315,5	ref|XP_003701627.1|	transketolase-like protein 2-like isoform 3	*Megachile rotundata*	695	1,0E-79	X			X		X	X	X	X		X
133675	fr4	5,6	42637,8	gb|ACB05778.1|	cuticular protein	*Artemia franciscana*	355	2,0E-37	X										
133685	fr4	10,7	19666,3	gb|AEX20134.1|	elongation factor 1-alpha, partial	*Synisoma lancifer*	398	6,0E-46									X		X
133722	fr1	5,5	36389,2	ref|XP_002746650.1|	ectonucleotide pyrophosphatase/phosphodiesterase family m	*Callithrix jacchus*	180	1,0E-11				X							
133742	fr3	9,0	35005,5	gb|ABB91677.1|	arthrodial cuticle protein AMP16.5	*Callinectes sapidus*	189	3,0E-14	X					X					
133756	fr4	9,7	44078,7	EFX88544	carbonyl reductase	*Daphnia pulex*	414	4,7E-44											X
133838	fr6	9,6	12649,4	gb|EFX90170.1|	No predicted function	*Daphnia pulex*	202	1,0E-16						X			X		
133881	fr1	8,3	40314,5	gb|AAT92148.1|	salivary secreted peptide	*Ixodes pacificus*	203	2,0E-16	X								X		
133896	fr4	9,0	58687,6	ref|XP_002017383.1|	No predicted function	*Drosophila persimilis*	693	7,0E-81				X		X					
133946	fr2	6,7	53572,5	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*	492	4,0E-51				X		X	X	X	X		X
134112	fr4	9,1	37377,0	pdb|1Q7F|A	No predicted function	*Uncharacterized*	208	6,0E-16				X							
134128	fr2	9,7	58758,2	gb|EGI65421.1|	Insulin-like growth factor 2 mRNA-binding protein 1	*Acromyrmex echinatior*	921	3,0E-113				X			X				
134170	fr1	9,7	93352,7	gb|EFN62410.1|	Hemocytin	*Camponotus floridanus*	124	9,0E-04						X			X		X
134172	fr1	9,8	60679,4	ABO33174	serine protease-like protein 3	*Penaeus monodon*	308	1,6E-27											X
134178	fr5	8,3	48382,9	ref|NP_001086341.1|	enoyl CoA hydratase 1, peroxisomal	*Xenopus laevis*	775	1,0E-96						X					
134275	fr5	10,1	27054,2	gb|ACX30003.1|	farnesoic acid O-methyltransferase	*Eriocheir sinensis*	345	2,0E-35	X					X					
134287	fr6	11,8	28631,9	gb|ADF47162.1|	ribosomal protein L6	*Hemiscyllium ocellatum*	270	2,0E-25									X		X
134338	fr3	8,1	40203,8	XP_319705	No predicted function	*Anopheles gambiae str. PEST*	209	3,7E-15											X
134359	fr5	4,7	17680,6	ref|XP_001963598.1|	GF20477 [Drosophila ananassae]gb|EDV44674.1| GF20477 [D	*Drosophila ananassae*	305	1,0E-29		X									
134381	fr4	8,6	29005,6	EFX69044	No predicted function	*Daphnia pulex*	835	9,1E-107											X
134392	fr3	9,3	15101,8	ref|XP_001988955.1|	No predicted function	*Drosophila grimshawi*	342	1,0E-36				X							
134405	fr2	7,8	25690,2	AAX98287	hepatopancreas trypsin, partial	*Pontastacus leptodactylus*	330	1,4E-33											X
134418	fr1	9,2	44354,8	gb|ACD44938.1|	protein-disulfide isomerase	*Scylla paramamosain*	469	7,0E-50									X		
134489	fr4	9,3	47414,8	ref|XP_317609.3|	No predicted function	*Anopheles gambiae str. PEST*	147	9,0E-08				X							
134553	fr6	9,3	57172,6	gb|EFA12511.1|	No predicted function	*Tribolium castaneum*	306	2,0E-29	X										
134627	fr5	8,9	49176,7	ref|XP_002710718.1|	heat shock 70kD protein binding protein	*Oryctolagus cuniculus*	381	5,0E-38				X							
135001	fr5	8,9	29374,4	gb|AEP84099.1|	acyl-CoA-binding protein	*Penaeus monodon*	344	2,0E-37				X							
135140	fr6	9,2	38368,1	gb|EFX89414.1|	No predicted function	*Daphnia pulex*	540	3,0E-61									X		
135282	fr4	4,6	11860,6	AFI61881.1	serine proteinaseprotein	*Macrobrachium nipponense*	108	1,1E-03							X				
135405	fr3	4,1	22043,3	gb|AET36897.1|	troponin C2	*Litopenaeus vannamei*	598	1,0E-75	X								X		
135487	fr5	5,5	40447,5	gb|EFX75492.1|	No predicted function	*Daphnia pulex*	640	1,0E-74							X		X		X
135671	fr6	9,3	51358,5	gb|EFX85065.1|	No predicted function	*Daphnia pulex*	489	3,0E-52	X										
135775	fr5	10,1	15215,1	gb|ADV40146.1|	ADP/ATP translocase	*Latrodectus hesperus*	316	1,0E-32				X							
135923	fr4	9,6	20340,2		No detected homolog				X										
136881	fr5	9,3	25518,1	ref|XP_001943087.1|	No predicted function	*Acyrthosiphon pisum*	98	7,0E-02	X						X		X		X
137322	fr5	10,3	38637,6	gb|ACG63696.1|	myosin heavy chain	*Gammarus duebeni*	919	4,0E-117	X					X	X				
137335	fr1	9,1	94446,4	ref|XP_002411773.1|	multifunctional chaperone	*Ixodes scapularis*	398	6,0E-40								X	X		X
137345	fr3	9,9	86737,3		No detected homolog										X	X			
137352	fr5	6,7	18929,6	ref|XP_001603580.1|	alpha-L-fucosidase-like	*Nasonia vitripennis*	424	1,0E-46	X					X					
137367	fr3	8,5	23687,6	ref|XP_002430579.1|	transketolase,	*Pediculus humanus corporis*	381	3,0E-39									X		
137406	fr3	9,7	65206,3	dbj|BAD12009.1|	endo-beta-1,4-glucanase OfEG2	*Odontotermes formosanus*	391	6,0E-39	X								X		X
137447	fr5	9,6	91763,1	NP_001083317	heat shock 70kDa protein 4	*Xenopus laevis*	515	1,3E-51											X
137448	fr4	9,4	72392,5	ref|XP_002413321.1|	LIM domain-binding protein,	*Ixodes scapularis*	842	2,0E-100						X					
137567	fr1	9,6	43617,8		No detected homolog							X					X		X
137569	fr2	8,8	63417,0	EFN85021	Xaa-Pro aminopeptidase 1	*Harpegnathos saltator*	355	7,1E-33											X
137592	fr1	5,8	37244,0	ref|XP_001661866.1|	succinyl-coa synthetase beta chain	*Aedes aegypti*	695	4,0E-84									X		
137627	fr1	10,5	76844,9	XP_784737	coatomer subunit gamma-2-like	*Strongylocentrotus purpuratus*	954	6,7E-113											X
137660	fr4	10,0	39695,7	ref|XP_002082643.1|	No predicted function	*Drosophila simulans*	601	5,0E-73									X		
137702	fr5	4,5	27190,6	gb|EFX79463.1|	No predicted function	*Daphnia pulex*	570	6,0E-67						X					
137731	fr4	8,9	55122,1	ref|XP_001963854.1|	No predicted function	*Drosophila ananassae*	918	3,0E-115						X					
137871	fr4	8,5	28054,0	emb|CAF90687.1|	No predicted function	*Tetraodon nigroviridis*	551	4,0E-63								X	X		X
138096	fr2	8,9	29832,9	gb|ACO10726.1|	disulfide-isomerase A3 precursor	*Caligus rogercresseyi*	530	2,0E-60									X		
138186	fr3	10,8	81905,7	XP_001999455	No predicted function	*Drosophila mojavensis*	562	5,5E-60											X
138262	fr6	9,8	37202,7	gb|EFX82640.1|	No predicted function	*Daphnia pulex*	383	3,0E-38				X							
138719	fr2	7,7	30323,8	gb|AET34915.1|	heat shock protein 21	*Macrobrachium rosenbergii*	199	7,0E-15									X		
138924	fr2	10,0	44408,2	ref|XP_001951931.1|	protein ROP-like isoform 2	*Acyrthosiphon pisum*	644	2,0E-74									X		
139091	fr5	10,3	35148,8		No detected homolog												X		X
140303	fr6	9,8	24916,7	gb|AAT67403.1|	RNA binding motif protein 8B	*Equus caballus*	438	2,0E-51						X					
140441	fr2	8,9	11990,3		No detected homolog														X
140868	fr2	5,3	16635,2	ref|XP_002411773.1|	multifunctional chaperone	*Ixodes scapularis*	467	4,0E-55								X	X		X
140931	fr3	6,7	46140,6	ref|XP_001600229.1|	serine/threonine-protein phosphatase 2A 65 kDa regulatory su	*Nasonia vitripennis*	1038	2,0E-132									X		
141052	fr6	9,6	72431,6	gb|EFX81514.1|	No predicted function	*Daphnia pulex*	738	3,0E-84	X										X
141086	fr5	10,4	41660,0	gb|ELU08533.1|	No predicted function	*Capitella teleta*	780	9,0E-95								X	X		X
141093	fr6	9,5	74984,7	dbj|BAH11992.1|	No predicted function	*Homo sapiens*	999	1,0E-126				X							
141153	fr2	6,9	23451,5	AGM32416	calponin-likey domain containing protein	*Coptotermes formosanus*	397	1,6E-44				X							X
141191	fr5	9,6	75132,9	gb|AGC26171.1|	17beta-HSD4	*Azumapecten farreri*	821	7,0E-95						X			X		
141220	fr4	9,7	45156,3	gb|ACT78699.1|	delta glutathione S-transferase GST	*Eriocheir sinensis*	310	8,0E-30									X		X
141228	fr1	10,2	86541,2	gb|EFX89062.1|	fascilin 2-like protein	*Daphnia pulex*	287	2,0E-23				X		X			X		X
141251	fr6	8,8	25830,2	ref|XP_001946818.1|	t-complex protein 1 subunit theta-like isoform 1	*Acyrthosiphon pisum*	464	9,0E-51				X							
141294	fr5	9,6	82875,2	ref|XP_002025869.1|	No predicted function	*Drosophila persimilis*	752	4,0E-88				X					X		
141321	fr3	8,8	58794,3	ref|XP_001656100.1|	heterogeneous nuclear ribonucleoprotein r	*Aedes aegypti*	1009	1,0E-127				X				X			

141565	fr2	9,3	18305,1	EFN76741.1	No predicted function	*Harpegnathos saltator*	323	7,4E-34							X			
141582	fr1	7,4	35313,9	dbj|BAB13738.1|	DD5	*Marsupenaeus japonicus*	226	2,0E-17	X									
141610	fr1	9,9	48702,4	gb|ACD44938.1|	protein-disulfide isomerase	*Scylla paramamosain*	601	2,0E-68			X	X			X	X		
141642	fr4	9,7	26387,0	gb|AFM75821.1|	myosin light chain 2	*Procambarus clarkii*	234	1,0E-20							X	X	X	X
141775	fr4	7,6	36553,9	dbj|BAC77082.1|	glyceraldehyde-3-phosphate dehydrogenase	*Procambarus clarkii*	1053	1,0E-139				X	X		X	X	X	X
142013	fr6	10,2	29994,7	gb|ACO11744.1|	60S ribosomal protein L23a	*Caligus rogercresseyi*	246	4,0E-23	X			X						
142049	fr4	11,0	17948,2		No detected homolog				X									
142293	fr1	9,1	24798,2		No detected homolog				X									
142711	fr4	8,7	20982,5	ref|XP_001654620.1|	glutathione-s-transferase theta, gst	*Aedes aegypti*	167	1,0E-11									X	
143364	fr4	10,4	25975,2	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	330	7,0E-32					X		X	X	X	
143688	fr1	10,9	21043,1	gb|ACR24952.1|	ribosomal protein L28	*Lepidochitona cinerea*	165	7,0E-12			X							
143811	fr2	10,4	17233,1	XP_004699157	ADP/ATP translocase 3-like, partial	*Echinops telfairi*	559	1,6E-69									X	X
144144	fr2	4,9	53675,6	gb|ADZ96217.1|	JHE-like carboxylesterase 1	*Pandalopsis japonica*	330	3,0E-30									X	X
144152	fr3	8,8	46334,7	ref|XP_004085772.1|	IgGFc-binding protein-like, partial	*Oryzias latipes*	138	6,0E-06			X	X					X	X
144157	fr4	4,5	41994,1	gb|ACR23314.1|	chitinase 4 precursor	*Litopenaeus vannamei*	1084	1,0E-139									X	
144165	fr4	8,2	29097,7	gb|ACU31810.1|	alpha2 macroglobulin isoform 2	*Fenneropenaeus chinensis*	439	2,0E-45	X		X	X						
144171	fr6	10,2	191658,5	ref|XP_002423351.1|	Filamin-C	*Pediculus humanus corporis*	1327	5,0E-158					X		X		X	X
144184	fr5	9,7	22094,3	EGI66180.1	Transketolaseprotein 2	*Acromyrmex echinatior*	640	6,2E-76					X					
144210	fr3	8,8	52490,5		No detected homolog						X							
144272	fr1	9,5	41765,2	EFA03240	No predicted function	*Tribolium castaneum*	760	1,9E-89										X
144281	fr6	9,9	49376,0	dbj|BAJ23879.1|	glycogen phosphorylase	*Marsupenaeus japonicus*	1000	2,0E-123				X						X
144320	fr1	10,0	53308,8	XP_005182614.1	fructose-bisphosphate aldolaseisoform X1	*Musca domestica*	485	3,0E-52	X			X	X				X	
144332	fr2	9,0	30255,0	gb|ACV04936.1|	pancreatic amylase mRNA	*Ctenopharyngodon idella*	605	1,0E-73	X		X	X					X	X
144378	fr3	10,0	50960,4	BAE87341.1	No predicted function	*Macaca fascicularis*	633	1,1E-73							X			
144425	fr5	9,6	45861,6	ref|XP_003452166.1|	plasma glutamate carboxypeptidase-like	*Oreochromis niloticus*	380	1,0E-37			X							
144441	fr6	9,8	57741,0	EFX60715	No predicted function	*Daphnia pulex*	177	3,5E-11										X
144517	fr2	5,2	42241,1	gb|EFX79463.1|	No predicted function	*Daphnia pulex*	801	3,0E-99	X			X						
144790	fr4	5,7	34988,6	ref|XP_002427522.1|	adenosylhomocysteinase,	*Pediculus humanus corporis*	1227	7,0E-163									X	
145005	fr5	8,5	15777,5	gb|EFZ21522.1|	No predicted function	*Solenopsis invicta*	334	1,0E-35			X							
145050	fr1	8,8	46232,5	gb|ADM64319.1|	prohibitin	*Eriocheir sinensis*	1154	2,0E-154			X							
145070	fr4	7,8	18608,1	emb|CAQ60115.1|	glyceraldehyde-3-phosphate dehydrogenase	*Gammarus locusta*	622	3,0E-77			X		X		X		X	X
145074	fr6	10,9	21124,2	gb|ABV60391.1|	histone H2A	*Artemia franciscana*	567	4,0E-71			X							
145138	fr3	10,4	76840,8	gb|EFN66634.1|	D-3-phosphoglycerate dehydrogenase	*Camponotus floridanus*	214	6,0E-15									X	
145634	fr1	9,0	28835,7	ref|XP_001865898.1|	antioxidant enzyme	*Culex quinquefasciatus*	159	4,0E-11			X	X						
145692	fr6	5,3	31986,8	XP_002606848	No predicted function	*Branchiostoma floridae*	104	4,7E-02										X
146288	fr6	6,2	27031,4	ref|XP_970004.1|	NADH:ubiquinone oxidoreductase ESSS subunit	*Tribolium castaneum*	269	5,0E-26	X									
147007	fr3	9,3	15280,7	ref|XP_002120177.1|	hydroxysteroid dehydrogenase like 2-like	*Ciona intestinalis*	516	5,0E-61									X	
147296	fr5	9,8	85633,1	AGT95756	carboxylesterase A5	*Locusta migratoria*	302	7,9E-26										X
147298	fr2	9,3	44001,0	XP_004622462.1	tubulin alpha chain	*Sorex araneus*	684	2,4E-83					X					
147363	fr5	9,2	63623,7	AGI96992.1	paramyosin	*Nilaparvata lugens*	549	1,7E-57	X		X	X	X		X		X	X
147377	fr2	9,2	53428,5	gb|EFA11828.1|	No predicted function	*Tribolium castaneum*	140	1,0E-06	X									
147388	fr2	9,3	51961,7	gb|EFX68208.1|	No predicted function	*Daphnia pulex*	310	6,0E-29			X							
147423	fr4	9,6	45700,6	ref|XP_317541.3|	No predicted function	*Anopheles gambiae str. PEST*	1056	3,0E-132									X	
147444	fr5	9,5	59520,4	gb|EFX72994.1|	No predicted function	*Daphnia pulex*	483	8,0E-55			X							
147451	fr3	8,1	41804,6	ref|XP_002734509.1|	aldehyde dehydrogenase 2-like	*Saccoglossus kowalevskii*	789	2,0E-97				X						
147469	fr2	9,1	35816,0	gb|EKC20273.1|	Spliceosome RNA helicase BAT1	*Crassostrea gigas*	490	9,0E-54				X						
147563	fr2	9,9	49138,0	gb|EKC38123.1|	SAP domain-containing ribonucleoprotein	*Crassostrea gigas*	127	4,0E-05				X						
147623	fr5	6,5	23338,4	gb|ABU97480.1|	allergen Tyr p 13	*Tyrophagus putrescentiae*	151	8,0E-10			X	X						
147727	fr5	8,6	78397,3	ref|XP_001963854.1|	No predicted function	*Drosophila ananassae*	1091	2,0E-138				X						
147818	fr2	9,9	27841,0	gb|AAS98886.1|	allergen Pen m 2	*Fenneropenaeus chinensis*	102	2,0E-03				X						
147839	fr2	8,9	26812,8	ref|XP_001607760.1|	GTP-binding protein CG1354-like	*Nasonia vitripennis*	656	2,0E-80	X									
147876	fr3	10,8	28034,1	XP_005504901.1	60S ribosomal protein L13	*Columba livia*	379	2,6E-41			X				X		X	
147887	fr5	10,1	58476,0	XP_001842491.1	No predicted function	*Culex quinquefasciatus*	114	4,3E-03					X					
148010	fr5	10,7	21730,4	ref|XP_004024378.1|	ribosomal protein,	*Ichthyophthirius multifiliis*	163	2,0E-11				X						
148355	fr4	9,1	66723,4	gb|EFX76404.1|	No predicted function	*Daphnia pulex*	635	7,0E-74									X	X
148587	fr2	10,6	39802,3	gb|AAT77811.1|	calpain B	*Gecarcinus lateralis*	446	6,0E-46	X		X						X	
149155	fr3	9,5	33460,8	gb|EGI64669.1|	No predicted function	*Acromyrmex echinatior*	106	1,1E-02				X						
150178	fr1	6,1	8217,0	gb|AEZ68611.1|	adenine nucleotide translocase	*Litopenaeus vannamei*	174	1,0E-13									X	X
150268	fr2	8,8	30610,3	AEB54656	glutathione S transferase D1	*Procambarus clarkii*	275	1,6E-25										X
150286	fr1	8,2	39866,2	ref|XP_002061081.1|	No predicted function	*Drosophila willistoni*	298	4,0E-29				X						
150370	fr2	8,3	41981,4	AAR01304.1	elongation factor-2	*Neogonodactylus oerstedii*	528	6,5E-57							X			
150376	fr5	9,3	46849,6		No detected homolog													X
150378	fr1	10,0	57709,4	BAJ23879	glycogen phosphorylase	*Marsupenaeus japonicus*	1310	2,3E-168										X
150386	fr2	11,2	48247,3	ref|NP_001261009.1|	Z band alternatively spliced PDZ-motif protein 52, isoform P	*Drosophila melanogaster*	421	4,0E-45	X									
150407	fr4	6,5	44009,1		No detected homolog				X									
150411	fr1	5,9	25630,7	ref|XP_002734509.1|	aldehyde dehydrogenase 2-like	*Saccoglossus kowalevskii*	734	2,0E-91	X									
150708	fr6	9,4	21415,6		No detected homolog							X						
150938	fr4	11,9	36635,5	ref|XP_004024375.1|	ribosomal protein,	*Ichthyophthirius multifiliis*	152	4,0E-09	X		X	X						
151196	fr1	10,3	21069,7	ref|XP_004318392.1|	60S ribosomal protein L23	*Tursiops truncatus*	192	3,0E-15			X	X						
151500	fr2	5,9	51227,9	gb|ACO15256.1|	Zinc metalloproteinase nas-12 precursor	*Caligus clemensi*	112	4,0E-03				X					X	
151543	fr6	9,1	30689,1		No detected homolog						X							
151692	fr5	11,3	33477,3	EFW41215	XDH protein	*Capsaspora owczarzaki ATCC 308*	255	9,1E-21										X
151887	fr2	8,8	42745,1	ref|XP_001641575.1|	No predicted function	*Nematostella vectensis*	242	6,0E-19	X									
151942	fr2	9,4	40951,4		No detected homolog							X					X	X
153026	fr1	9,8	49310,4	gb|EFX75803.1|	No predicted function	*Daphnia pulex*	832	3,0E-104				X						
153034	fr3	7,2	33209,5	gb|ADG85668.1|	lectin B isoform 1	*Marsupenaeus japonicus*	122	3,0E-05	X				X				X	X
153040	fr3	9,7	56543,9	XP_004929847.1	failed axon connectionsisoform X2	*Bombyx mori*	403	1,9E-40							X			
153076	fr1	9,9	74252,2		No detected homolog								X					
153091	fr5	9,5	37027,5	XP_785580.3	isocitrate dehydrogenase	*NADP*	656	3,0E-78							X			
153114	fr1	9,7	48244,8	sp|P04069.1|CBPB_ASTFL	Carboxypeptidase B	*Uncharacterized*	427	2,0E-45	X									
153121	fr6	9,7	31914,3	gb|EFX88544.1|	carbonyl reductase	*Daphnia pulex*	548	4,0E-65									X	X
153160	fr6	10,3	20752,0	ref|XP_003748179.1|	histone H1.0-A-like	*Metaseiulus occidentalis*	262	3,0E-25							X		X	
153193	fr2	5,8	56003,2	gb|ACN89260.1|	protein disulfide isomerase	*Litopenaeus vannamei*	1345	1,0E-178			X	X	X		X		X	X
153198	fr5	8,6	45512,3	CAA54524	preamylase 1	*Litopenaeus vannamei*	336	1,9E-31										X
153235	fr3	10,9	47384,8		No detected homolog													X
153280	fr4	9,4	26569,7	gb|AET34917.1|	apoptosis inhibitor	*Macrobrachium rosenbergii*	391	1,0E-40	X			X			X			
153496	fr1	9,9	69858,7	ref|XP_001633389.1|	No predicted function	*Nematostella vectensis*	937	5,0E-117									X	X
153529	fr1	10,2	37944,6	gb|ENN80417.1|	No predicted function	*Dendroctonus ponderosae*	547	6,0E-61			X							
153630	fr6	11,3	18610,8	gb|EFN76243.1|	L-lactate dehydrogenase	*Harpegnathos saltator*	450	1,0E-51									X	X
153779	fr4	10,8	30966,9	prf||671058A	dehydrogenase,glyceraldehydephosphate	*Uncharacterized*	307	2,0E-29	X			X						
153870	fr6	9,3	61473,4	gb|EKC39191.1|	Thioredoxin-like protein 1	*Crassostrea gigas*	783	2,0E-96			X							
154258	fr2	9,2	73477,5	gb|EFX76404.1|	No predicted function	*Daphnia pulex*	569	9,0E-64									X	X
155076	fr6	8,5	41189,7	gb|AEA95820.1|	prohibitin	*Charybdis japonica*	774	1,0E-97	X			X						
155245	fr1	9,5	43816,9	ref|XP_002591929.1|	No predicted function	*Branchiostoma floridae*	502	8,0E-54									X	
155311	fr4	9,1	47342,9	XP_004551026	trans-2-enoyl-CoA reductase, mitochondrial-like	*Maylandia zebra*	438	6,0E-46									X	X
155615	fr4	9,7	44608,3	dbj|BAE87441.1|	No predicted function	*Macaca fascicularis*	1051	1,0E-139			X		X		X			
155620	fr2	9,4	41306,7	gb|ACQ91270.1|	myosin heavy chain variant B	*Bombyx mandarina*	1077	1,0E-136	X		X	X	X		X		X	X
155642	fr5	8,0	54392,3	ERL93980.1	No predicted function	*Dendroctonus ponderosae*	1337	3,5E-172				X			X		X	X
155689	fr5	10,8	23929,8	XP_002409673.1	NADP-dependent isocitrate dehydrogenase	*Ixodes scapularis*	313	2,2E-30							X			X
155751	fr6	10,1	89195,3	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	791	5,2E-88										X
155762	fr2		68659,5		No detected homolog												X	
155798	fr4	9,7	79179,3	gb|EFX68668.1|	No predicted function	*Daphnia pulex*	1257	1,0E-161									X	
155803	fr2	9,3	49086,8	ESOQ2590.1	No predicted function	*Helobdella robusta*	234	1,2E-18							X			
155828	fr6	11,8	79725,4	ref|XP_001942763.2|	collagen alpha-2(I) chain-like	*Acyrthosiphon pisum*	152	3,0E-07				X						
155837	fr3	8,7	60496,7	gb|EFX69032.1|	malate dehydrogenase	*Daphnia pulex*	1185	6,0E-156	X		X	X			X		X	X
155889	fr2	9,8	43092,8	XP_003782607.1	carboxypeptidase Q	*Otolemur garnettii*	317	3,3E-29					X					
155991	fr2	9,7	46958,7	gb|ABN12066.1|	ATP synthase beta subunit	*Maconellicoccus hirsutus*	1189	4,0E-158	X		X	X	X		X		X	X
156197	fr3	9,6	22112,0	ref|XP_310955.4|	No predicted function	*Anopheles gambiae str. PEST*	356	1,0E-36			X							
156236	fr1	8,5	21656,6		No detected homolog												X	X
156252	fr6	9,6	34579,6	gb|AEC12821.1|	lactate dehydrogenase	*Litopenaeus vannamei*	746	2,0E-93									X	
156342	fr4	8,8	34870,1	gb|EFX67986.1|	No predicted function	*Daphnia pulex*	343	1,0E-33	X									
156582	fr6	10,1	65946,8	gb|EFX86358.1|	No predicted function	*Daphnia pulex*	244	2,0E-18			X	X						
156597	fr3	8,8	16953,3	gb|ABU53620.1|	sorbitol dehydrogenase	*Pyrrhocoris apterus*	319	2,0E-32				X						
156898	fr6		30796,4	ref|XP_003241023.1|	histone H2A-like	*Acyrthosiphon pisum*	329	2,0E-34			X							
157736	fr6	10,9	34700,4	gb|ACO12528.1|	60S ribosomal protein L4	*Lepeophtheirus salmonis*	542	1,0E-62				X			X		X	X
157852	fr2	8,9	27188,4	gb|EFX66153.1|	No predicted function	*Daphnia pulex*	444	6,0E-49			X	X						
157947	fr6	5,6	24836,2	pdb|3PRY|A	No predicted function	*Uncharacterized*	559	9,0E-68							X		X	
158015	fr5	9,3	37836,2	ref|XP_974058.1|	annexin B13a	*Tribolium castaneum*	660	1,0E-78									X	X
158027	fr2	10,3	35512,5	gb|ACH69159.1|	myosin heavy chain	*Bombyx mori*	854	2,0E-104							X		X	X
158043	fr6	10,5	21938,0	ref|XP_002435347.1|	glutathione S-transferase,	*Ixodes scapularis*	212	8,0E-19				X	X				X	
158083	fr3	9,2	131489,4	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	1230	2,0E-138									X	X
158099	fr6	9,4	60926,7	EFZ13522	No predicted function	*Solenopsis invicta*	595	2,7E-65										X
158168	fr2	8,7	31011,0	gb|AAF08281.1|AF189779_1	vacuolar ATP synthase subunit B K form	*Carcinus maenas*	856	2,0E-108			X							
158172	fr4	8,3	98387,1	gb|AEK49107.1|	Tudor staphylococcal nuclease	*Penaeus monodon*	2008	0,0E+00	X		X							
158231	fr1	9,5	59831,0	ref|XP_001658309.1|	threonyl-tRNA synthetase	*Aedes aegypti*	1141	2,0E-144				X						
158265	fr5	10,8	51104,2	EHJ66818	phosphoglycerate kinase	*Danaus plexippus*	625	2,2E-72										X
158299	fr2	10,8	24983,0	dbj|BAE87341.1|	No predicted function	*Macaca fascicularis*	166	1,0E-10			X	X						
158389	fr1	10,0	23348,0	gb|ACY66547.1|	ribosomal protein L5	*Scylla paramamosain*	420	2,0E-48				X						
158419	fr1	9,4	31176,8	gb|ENN73642.1|	No predicted function	*Dendroctonus ponderosae*	144	3,0E-07			X		X		X			
158439	fr4	8,8	34091,2	gb|EFZ21595.1|	No predicted function	*Solenopsis invicta*	634	8,0E-78				X						

158464	fr6	4,5	26888,9	ref|YP_001213298.1|	No predicted function	*Pelotomaculum thermopropionic*	197	8,0E-15				X						
158593	fr3	6,8	21527,9	XP_001607772.1	multifunctional protein ADE2	*Nasonia vitripennis*	368	2,9E-38							X			
158681	fr4	10,8	15368,8		No detected homolog								X					
158809	fr6	8,0	22501,8	gb|AEE61966.1|	No predicted function	*Dendroctonus ponderosae*	311	7,0E-33				X							
159182	fr2	9,6	47240,0	XP_005094236	indole-3-acetaldehyde oxidase-like	*Aplysia californica*	241	1,1E-18											X
159387	fr2	9,7	32988,7	gb|AEE63197.1|	No predicted function	*Dendroctonus ponderosae*	401	1,0E-41					X						X
159799	fr1	8,9	14041,2		No detected homolog							X							
160313	fr5	9,7	26649,7	gb|ABQ52471.1|	mortalin splice variant	*Mya arenaria*	488	9,0E-54					X						
160359	fr6	9,0	71691,5	gb|EHJ67857.1|	No predicted function	*Danaus plexippus*	1522	0,0E+00								X		X	
160363	fr1	9,6	74380,9	AAN77377.1	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	1272	1,9E-158							X		X		
160369	fr4	5,2	44982,0	gb|ACI04663.1|	tropomyosin	*Jasus lalandii*	736	3,0E-92	X				X					X	
160381	fr4	8,2	60279,2	XP_005188828	hemocytin-like	*Musca domestica*	234	5,5E-17										X	X
160448	fr2	8,9	36206,0	ref|XP_002731568.1|	cytosolic malate dehydrogenase-like	*Saccoglossus kowalevskii*	638	4,0E-77	X				X						
160483	fr4	7,9	54447,6	gb|ACO36738.1|	ubiquitin carboxyl-terminal esterase L3	*Scylla paramamosain*	387	5,0E-40				X	X		X			X	
160484	fr4	7,5	56845,3	AHA11058	vacuolar H+-ATPase	*Exopalaemon carinicauda*	1127	1,6E-145									X	X	X
160507	fr3	9,9	63115,8	gb|ADK66280.1|	intracellular fatty acid-binding protein	*Litopenaeus vannamei*	146	6,0E-08				X							
160555	fr4	9,1	78656,7	ref|XP_003931000.1|	propionyl-CoA carboxylase beta chain, mitochondrial isoform	*Saimiri boliviensis boliviensis*	1006	1,0E-125					X					X	X
160605	fr4	10,4	43480,0	XP_003741655	phosphoglycerate kinase-like	*Metaseiulus occidentalis*	584	4,7E-67											X
160777	fr2	9,4	63569,7	ref|XP_003977210.1|	alcohol dehydrogenase class-3 chain L-like	*Takifugu rubripes*	721	8,0E-86				X	X						
160935	fr3	9,4	39010,5	ref|NP_001161324.1|	ATP synthase subunit b, mitochondrial	*Nasonia vitripennis*	425	2,0E-46										X	
161076	fr4	4,7	16560,8	gb|ACC76803.1|	Lit v 3 allergen myosin light chain	*Litopenaeus vannamei*	422	2,0E-49	X				X						
161448	fr1	10,3	52200,5		No detected homolog								X						
161609	fr2	8,7	44680,7		No detected homolog								X						
161699	fr3	10,0	41422,0	gb|EGT55281.1|	No predicted function	*Caenorhabditis brenneri*	555	5,0E-64				X				X			
162106	fr1	6,4	24053,6	ref|XP_003741655.1|	phosphoglycerate kinase-like	*Metaseiulus occidentalis*	668	3,0E-82				X	X					X	X
162401	fr4	8,2	21334,1	gb|AEE62370.1|	No predicted function	*Dendroctonus ponderosae*	613	5,0E-73				X							
162469	fr3	9,8	95738,4	ref|XP_001605732.2|	phosphoribosylformylglycinamidine synthase-like	*Nasonia vitripennis*	739	3,0E-79				X							
162505	fr6	8,6	88542,4	ref|XP_001120260.1|	sorting nexin-2-like isoform 2	*Apis mellifera*	725	1,0E-82				X							
162506	fr3		64843,6		No detected homolog													X	
162527	fr4	5,7	26542,1	gb|ACC76803.1|	Lit v 3 allergen myosin light chain	*Litopenaeus vannamei*	241	1,0E-21				X	X						
162551	fr3	8,4	32917,7	gb|AAC32663.1|	glycerol-3-phosphate dehydrogenase, partial	*Ceratitis capitata*	651	2,0E-80										X	
162564	fr1	9,2	30183,7	gb|AEI88031.1|	ATP synthase subunit b	*Scylla paramamosain*	205	6,0E-17								X		X	
162578	fr5	9,5	23472,9	emb|CAG32523.1|	No predicted function	*Gallus gallus*	495	7,0E-59						X		X		X	X
162582	fr2	9,0	87933,4	gb|ADB44901.1|	DEAD(Asp-Glu-Ala-Asp) box polypeptide 39	*Macrobrachium nipponense*	1120	1,0E-141				X	X	X		X		X	X
162612	fr3	9,0	37124,0	ref|XP_003697978.1|	mitochondrial-processing peptidase subunit beta-like	*Apis florea*	557	2,0E-63				X	X						
162616	fr2	5,6	29202,3	AGZ89665.1	eukaryotic initiation factor 4A	*Litopenaeus vannamei*	644	1,3E-77	X								X		
162638	fr3	8,1	10592,6		No detected homolog								X						
162710	fr5	9,8	40425,6	ref|XP_975021.1|	stathmin	*Tribolium castaneum*	318	5,0E-30				X							
162718	fr2	8,3	32249,1	gb|AET34921.1|	chaperonin	*Macrobrachium rosenbergii*	848	2,0E-106	X										
162837	fr5	9,8	61169,9	gb|EFA12793.1|	No predicted function	*Tribolium castaneum*	290	1,0E-25					X						
163123	fr3	7,9	24228,1	gb|EEZ97818.1|	No predicted function	*Tribolium castaneum*	330	4,0E-32	X			X	X						
163286	fr3	8,5	32391,6	gb|ADM64456.1|	fatty acid binding protein	*Eriocheir sinensis*	291	1,0E-28				X							
163642	fr3		15284,6	gb|ENN75662.1|	No predicted function	*Dendroctonus ponderosae*	288	5,0E-29									X	X	X
164345	fr6	5,4	11775,6		No detected homolog							X							
164353	fr6	9,4	39119,1	XP_002427775.1	myosin-9, putative	*Pediculus humanus corporis*	485	5,2E-50									X		
164360	fr5	9,0	30688,7	ref|NP_001040164.1|	14-3-3-like protein	*Bombyx mori*	455	9,0E-52					X		X		X	X	
164403	fr3	9,6	39457,9	ref|XP_003703942.1|	proliferation-associated protein 2G4-like	*Megachile rotundata*	507	8,0E-57				X	X						
164562	fr1	4,8	21253,7	AAR01291.1	elongation factor-2	*Forficula auricularia*	681	3,0E-87									X		
164608	fr1	9,2	43770,9	gb|AET34921.1|	chaperonin	*Macrobrachium rosenbergii*	992	2,0E-126				X							
164711	fr4	9,4	43298,6	ref|XP_003705965.1|	proclotting enzyme-like	*Megachile rotundata*	108	1,3E-02				X							
164713	fr5	6,8	67246,1	ref|XP_002130881.1|	glucosamine-6-phosphate isomerase 1-like	*Ciona intestinalis*	1008	1,0E-129					X					X	
165164	fr3	9,5	22926,9	gb|EFN73927.1|	pyruvate dehydrogenase E1 component subunit alpha, mitoch	*Camponotus floridanus*	217	1,0E-17					X						
165204	fr1	4,2	19885,6	gb|ADI59751.1|	early cuticle protein 3	*Callinectes sapidus*	220	2,0E-19	X				X						
165315	fr4	8,5	16625,2	gb|ABU53620.1|	sorbitol dehydrogenase	*Pyrrhocoris apterus*	376	1,0E-40	X										
166043	fr4	4,4	12900,2	ref|XP_001660597.1|	protein disulfide-isomerase A6 precursor	*Aedes aegypti*	372	4,0E-40										X	X
166183	fr3	9,8	53201,5	gb|AGE44125.1|	tropomyosin	*Portunus pelagicus*	521	1,0E-58										X	
166184	fr3	10,0	96960,0	CCW43201	alpha-2-macroglobulin	*Scylla paramamosain*	948	3,2E-106	X			X	X					X	X
166195	fr2	11,6	60852,8	XP_004923608	filamin-A-like	*Bombyx mori*	522	1,4E-55											X
166198	fr4	8,0	46357,6	gb|AEA72216.1|	alpha-amylase	*Crassostrea gigas*	887	3,0E-110										X	X
166287	fr4	8,2	43441,4	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	592	7,2E-65											X
166361	fr5	8,9	71493,0	dbj|BAM18606.1|	hemolectin	*Papilio xuthus*	141	8,0E-07										X	
166374	fr5	8,8	23312,7		No detected homolog													X	X
166399	fr4	10,7	34771,5	BAD17952	phosphoglycerate kinase	*Lethenteron reissneri*	546	8,5E-63											X
166437	fr1	5,6	19630,9	gb|ABV60327.1|	fructose-bisphosphate aldolase	*Lutzomyia longipalpis*	558	4,0E-68										X	
166697	fr2	9,1	51006,8	ref|XP_002430488.1|	Hrp65 protein,	*Pediculus humanus corporis*	294	9,0E-26	X			X	X						
166723	fr2	10,4	30663,3	gb|ACX30003.1|	farnesoic acid O-methyltransferase	*Eriocheir sinensis*	222	3,0E-18					X						
166827	fr1	10,1	24869,8	gb|ABB59715.1|	hemocyanin	*Cyamus scammoni*	554	4,0E-63				X	X	X		X		X	X
166887	fr1	8,9	30456,6		No detected homolog							X							
166905	fr1	8,7	49448,5		No detected homolog								X						
167012	fr6	9,7	49633,3		No detected homolog							X							
167020	fr1	9,6	34011,4	gb|AAR32721.1|	ubiquinol-cytochrome c oxidoreductase subunit 9 and Rieske i	*Pithecia irrorata*	262	7,0E-24	X				X						
167195	fr1	9,1	74482,1	EFX61932.1	No predicted function	*Daphnia pulex*	823	6,6E-98									X		
167919	fr6	9,8	62646,2	dbj|BAK61430.1|	myosin heavy chain type b	*Marsupenaeus japonicus*	692	4,0E-75				X		X				X	
167946	fr1	9,1	17914,0	XP_003704083.1	filamin-A	*Megachile rotundata*	407	4,6E-42						X					
167958	fr3	6,4	40353,8	gb|EFX77111.1|	No predicted function	*Daphnia pulex*	969	1,0E-115								X		X	
167959	fr3	11,5	47866,6	AAW84264.1	heterogeneous nuclear ribonucleoprotein L	*Aplysia californica*	297	1,2E-27								X			
168006	fr3	9,3	43154,4	XP_001601449	aspartate aminotransferase, cytoplasmic-like	*Nasonia vitripennis*	536	3,6E-60											X
168010	fr2	6,1	39776,2	gb|AGC10822.1|	Rab-1	*Eriocheir sinensis*	613	1,0E-74				X							
168076	fr2	9,3	21261,6	ref|XP_396106.4|	V-type proton ATPase 116 kDa subunit a-like	*Apis mellifera*	241	2,0E-22										X	
168116	fr2	9,4	39859,6	ref|XP_002072914.1|	No predicted function	*Drosophila willistoni*	509	1,0E-55	X										
168131	fr3	9,1	35910,0	sp|Q25009.1|TBB1_HOMAM	Tubulin beta-1 chain	*Homarus americanus*	979	6,0E-127					X				X	X	X
168248	fr3	5,9	26029,9	gb|AEG79567.1|	sarcoplasmic calcium-binding protein variant a	*Procambarus clarkii*	593	9,0E-74				X	X		X		X	X	X
168270	fr4	9,4	39839,0	gb|AFT92034.1|	triose-phosphate isomerase	*Litopenaeus vannamei*	689	2,0E-85				X						X	X
168287	fr4	4,8	29142,4	gb|AFJ59951.1|	globular head complement component C1q receptor	*Fenneropenaeus chinensis*	604	5,0E-74				X	X						
168378	fr6	8,8	48872,3	gb|EHJ75654.1|	transgelin	*Danaus plexippus*	664	1,0E-81	X				X						
168505	fr1	9,1	35492,7	ref|XP_003453266.1|	tryptophanyl-tRNA synthetase, cytoplasmic-like	*Oreochromis niloticus*	839	1,0E-105										X	
168713	fr2	8,5	41817,5	ref|XP_001951792.1|	heat shock 70 kDa protein 4L-like	*Acyrthosiphon pisum*	1071	1,0E-135				X							
168929	fr1	9,6	21309,7	gb|AEC48730.1|	ubiquitin-conjugating enzyme E2 b	*Eriocheir sinensis*	467	6,0E-56				X							
169584	fr4	10,4	15624,9		No detected homolog							X		X					
169586	fr1	10,0	73410,9		No detected homolog									X		X			X
169603	fr2	10,0	57024,9	gb|ABY66597.1|	pyruvate kinase 2	*Litopenaeus vannamei*	921	2,0E-114				X		X		X		X	X
169622	fr1	10,1	71974,4	gb|EFZ13892.1|	No predicted function	*Solenopsis invicta*	614	1,0E-69					X					X	X
169623	fr4	5,5	21049,5	EGI70840	Aminopeptidase N	*Acromyrmex echinatior*	353	5,5E-35											X
169648	fr6	6,3	69357,1	dbj|BAM20313.1|	paramyosin	*Papilio polytes*	917	2,0E-112	X				X				X	X	X
169669	fr4	9,0	34411,9		No detected homolog													X	X
169690	fr2	10,9	44472,4	gb|AEC50079.1|	calreticulin	*Pacifastacus leniusculus*	216	4,0E-16				X	X	X					
169703	fr3	10,2	28370,0	gb|ACR56806.1|	lipopolysaccharide and beta-1,3-glucan binding protein	*Fenneropenaeus merguiensis*	401	1,0E-42										X	X
169712	fr3	9,8	24280,7	ref|XP_001844560.1|	zinc metalloproteinase nas-12 [Culex quinquefasciatus]gb|EDS	*Culex quinquefasciatus*	109	2,0E-03			X								
169716	fr5	6,0	41418,9	gb|EFN65250.1|	Nesprin-1	*Camponotus floridanus*	219	3,0E-16	X										
169737	fr2	9,1	82681,0	gb|EFX72698.1|	No predicted function	*Daphnia pulex*	234	4,0E-17					X						
169742	fr1	6,2	30538,9	sp|P04069.1|CBPB_ASTFL	Carboxypeptidase B	*Uncharacterized*	642	5,0E-79				X	X				X	X	X
169778	fr4	8,2	46222,2		No detected homolog								X					X	
169790	fr3	9,6	35143,9	gb|AET34917.1|	apoptosis inhibitor	*Macrobrachium rosenbergii*	265	9,0E-23	X										
169820	fr2	5,0	20827,3		No detected homolog														X
169890	fr3	5,2	35494,8	ref|XP_003393148.1|	No predicted function	*Bombus terrestris*	105	3,1E-02										X	
170306	fr3	11,1	13755,7	gb|ADV40154.1|	ribosomal protein L4	*Latrodectus hesperus*	378	1,0E-41	X										
170352	fr4	7,3	39430,8	gb|AEG80154.1|	Tcp-1-beta	*Litopenaeus vannamei*	1186	3,0E-156	X			X	X				X	X	
170702	fr2	6,7	29812,7	gb|AEG79567.1|	sarcoplasmic calcium-binding protein variant a	*Procambarus clarkii*	627	3,0E-78										X	X
171051	fr5	6,5	8291,1		No detected homolog				X										
171058	fr5	10,9	19142,9	gb|AFA51707.1|	elongation factor 1 alpha, partial	*Melita plumulosa*	354	3,0E-39							X		X	X	X
171126	fr4	8,7	19193,3	gb|ACO10210.1|	Venom allergen 5.02 precursor	*Caligus rogercresseyi*	139	5,0E-08				X							
171189	fr3	6,6	29939,7	gb|ABS57456.1|	alpha-tubulin	*Heliconius melpomene*	832	8,0E-109				X	X						
171207	fr2	9,5	43393,2	gb|ABY66598.1|	pyruvate kinase 3	*Litopenaeus vannamei*	1042	2,0E-133									X	X	X
171219	fr5	5,3	21586,9	gb|ABM74408.1|	hemocyanin	*Portunus pelagicus*	585	6,0E-72									X	X	
171230	fr3	9,9	69422,3	gb|EHJ72366.1|	No predicted function	*Danaus plexippus*	1044	2,0E-131					X						
171303	fr3	9,1	83506,3	ref|XP_002415244.1|	No predicted function	*Ixodes scapularis*	126	3,0E-05										X	
171315	fr5	8,5	32745,0	AEK11985	calmin-like protein	*Tigriopus californicus*	136	3,3E-07											X
171435	fr4	9,0	66805,5	ref|XP_002666686.1|	v-type proton ATPase catalytic subunit A	*Danio rerio*	1563	0,0E+00				X					X	X	
171530	fr6	9,3	30848,6	gb|AFH75400.1|	elongation factor 2	*Eriocheir sinensis*	582	2,0E-65									X	X	X
172234	fr2	5,5	43358,6	CAB43347	lamin	*Priapulus caudatus*	516	5,6E-56											X
172600	fr1	6,7	33346,7	gb|AET34921.1|	chaperonin	*Macrobrachium rosenbergii*	758	6,0E-93					X						
172695	fr2	8,3	32419,3	XP_002411179	No predicted function	*Ixodes scapularis*	252	2,8E-20											X
172697	fr5	10,4	51360,9		No detected homolog									X		X			
172699	fr5	10,0	59332,9	gb|AFX62797.1|	alpha tubulin 84B, partial	*Drosophila simulans*	1184	4,0E-155					X					X	X
172788	fr6	10,2	17430,6		No detected homolog														X
172832	fr5	8,7	55370,7	gb|AEK11985.1|	calmin-like protein	*Tigriopus californicus*	167	7,0E-11										X	X
173108	fr6	8,7	38971,2	gb|EFX88967.1|	No predicted function	*Daphnia pulex*	713	1,0E-89				X							
173956	fr6	10,3	27451,8		No detected homolog							X							
173980	fr5	10,1	16414,4	gb|ACP50252.1|	arginine kinase	*Tychiodes sp. DDM-2009*	395	1,0E-45					X						
174103	fr1	9,8	13086,8	gb|ABE77334.1|	histone variant H2A.Z	*Hydroides elegans*	381	3,0E-44				X	X						

174113	fr3	9,5	40474,9	gb|ADL14588.1|	catalase	*Hyriopsis cumingii*	960	3,0E-123	X			X	X	X	X	X
174148	fr5	10,0	33440,6	gb|ADB10837.1|	lectin	*Eriocheir sinensis*	169	2,0E-11							X	X
174162	fr1	9,7	50552,6	gb|EFX79712.1|	No predicted function	*Daphnia pulex*	814	7,0E-99				X			X	
174223	fr4	9,6	30433,9	ref|XP_001688561.1|	No predicted function	*Anopheles gambiae str. PEST*	297	1,0E-27	X			X				
174259	fr5	10,1	29186,3		No detected homolog												X
174271	fr4	5,8	47664,0	gb|EHJ66180.1|	No predicted function	*Danaus plexippus*	912	1,0E-115	X						X	
174278	fr6	9,1	55802,7	ref|XP_001601449.1|	aspartate aminotransferase, cytoplasmic-like	*Nasonia vitripennis*	586	2,0E-66				X			X	X
174360	fr1	10,1	74883,2	ref|XP_001989177.1|	No predicted function	*Drosophila grimshawi*	533	1,0E-56			X					
174472	fr2	4,8	20387,8	EFX89582	No predicted function	*Daphnia pulex*	141	3,0E-07								X
174517	fr1	4,4	22254,8		No detected homolog						X	X				
175002	fr1	8,9	24690,7	ref|XP_003388020.1|	sorbitol dehydrogenase-like	*Amphimedon queenslandica*	489	9,0E-57							X	X
175053	fr2	8,1	47992,7	ref|NP_001040544.1|	transaldolase	*Bombyx mori*	799	1,0E-99			X	X				
175139	fr3	4,9	16478,8	gb|ACB05778.1|	cuticular protein	*Artemia franciscana*	215	4,0E-19	X							
175507	fr6	7,9	50493,1	emb|CAO79603.1|	alpha-tubulin	*Fasciola hepatica*	1315	3,0E-175	X		X	X	X	X	X	X
175575	fr3	9,6	41807,8	dbj|BAH28836.1|	elongation factor 1-alpha	*Marsupenaeus japonicus*	782	2,0E-96			X	X				X
175700	fr1	10,4	38819,9	gb|AET34917.1|	apoptosis inhibitor	*Macrobrachium rosenbergii*	277	4,0E-24						X	X	X
175708	fr4	11,6	35391,0	ref|XP_395092.2|	No predicted function	*Apis mellifera*	122	3,0E-05				X			X	
176048	fr6	8,6	26854,7	ref|XP_002737020.1|	superoxide dismutase 1, soluble-like, partial	*Saccoglossus kowalevskii*	296	6,0E-30				X				
176301	fr2	5,8	37098,2		No detected homolog				X			X			X	X
176791	fr1	9,4	39568,7	gb|EFX76969.1|	No predicted function	*Daphnia pulex*	544	2,0E-62							X	X
176997	fr2	9,3	66869,8	AEV89752.1	heat shock protein 60	*Schistocerca gregaria*	1034	1,4E-129					X			
177084	fr4	4,8	29729,5	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	1050	2,0E-138	X			X				
177102	fr5	7,9	22064,0	XP_005566910.1	acrosin	*Macaca fascicularis*	122	5,6E-05					X			
177368	fr1	10,3	31501,7	gb|AEK11985.1|	calmin-like protein	*Tigriopus californicus*	130	1,0E-06			X				X	X
177383	fr6	9,5	30835,1	gb|ELU08533.1|	No predicted function	*Capitella teleta*	1064	6,0E-139					X	X	X	X
177485	fr5	9,2	35236,2	ref|XP_003394081.1|	heat shock 70 kDa protein 4L-like	*Bombus terrestris*	739	4,0E-87				X				
177844	fr6	5,1	34428,3	YP_003289787	No predicted function	*Rhodothermus marinus DSM 425*	208	3,0E-15								X
177973	fr2	9,8	66249,4	gb|AAM73986.1|AF465679_	alpha-tubulin 1	*Homarus gammarus*	790	2,0E-95						X	X	X
177978	fr4	9,6	21910,7	gb|ACG63696.1|	myosin heavy chain	*Gammarus duebeni*	529	2,0E-61	X		X	X	X		X	
178011	fr1	8,9	40733,4	gb|EHJ72366.1|	No predicted function	*Danaus plexippus*	746	1,0E-90			X					
178091	fr2	9,5	51388,8	dbj|BAJ23879.1|	glycogen phosphorylase	*Marsupenaeus japonicus*	1150	2,0E-145			X	X			X	X
178103	fr6	5,4	26230,7		No detected homolog							X			X	X
178109	fr4	11,2	69776,1	sp|Q27319.1|GELS_HOMAM	Gelsolin, cytoplasmic	*Homarus americanus*	564	2,0E-59	X		X	X				
178132	fr5	5,4	35339,6	ref|XP_001602600.1|	venom protease	*Nasonia vitripennis*	126	4,0E-05			X		X			
178152	fr5	9,4	56377,5	gb|AEE62370.1|	No predicted function	*Dendroctonus ponderosae*	792	6,0E-95			X					
178171	fr6	9,4	38436,5	gb|AAV41060.1|	Hsp90alpha, partial	*Xenopus laevis*	358	2,0E-38			X					
178200	fr2	8,8	42001,4	gb|ACK44247.1|	vitelline membrane outer layer 1-like protein	*Scylla paramamosain*	363	3,0E-37			X					
178303	fr6	10,8	34164,1	XP_504936.1	No predicted function	*Yarrowia lipolytica*	765	8,8E-94						X		X
178309	fr3	5,2	23071,7	gb|AEE61406.1|	No predicted function	*Dendroctonus ponderosae*	564	2,0E-66							X	
178360	fr2	6,2	20247,1	AAR01298.1	elongation factor-2	*Libinia emarginata*	555	2,0E-63						X		
178443	fr5	5,1	32902,7	ref|XP_003397965.1|	fatty acid-binding protein-like	*Bombus terrestris*	180	2,0E-13	X							
179005	fr5	9,2	44204,0	gb|ACN11335.1|	Peroxisomal 3,2-trans-enoyl-CoA isomerase	*Salmo salar*	579	9,0E-67			X	X				
179168	fr1	9,5	22276,9	gb|AEC50079.1|	calreticulin	*Pacifastacus leniusculus*	467	6,0E-53	X		X	X				X
179175	fr5	10,1	64296,9	ref|XP_973301.1|	DJ-1	*Tribolium castaneum*	104	2,9E-02			X					
179252	fr2	10,5	33789,3	ref|XP_002399327.1|	histone H1,	*Ixodes scapularis*	195	5,0E-15							X	X
179301	fr4	9,0	37354,3	gb|EFX88976.1|	myosin regulatory light chain	*Daphnia pulex*	357	4,0E-37							X	X
179325	fr3	10,0	62150,5	emb|CAA57964.1|	serine proteinase inhibitor	*Pacifastacus leniusculus*	297	4,0E-26				X			X	
179328	fr4	10,1	46098,2	gb|AFJ22314.1|	heat shock protein cognate 5, partial	*Brachychthoniidae gen. sp. 3 AD1*	689	6,0E-81			X					
179570	fr4	9,0	34255,6	XP_005311798	No predicted function	*Chrysemys picta bellii*	195	1,2E-13								X
179995	fr3	9,8	22926,8	gb|AAH52753.1|	Rtn1b protein	*Danio rerio*	197	2,0E-15							X	
179998	fr5	4,6	29122,5	XP_971371	No predicted function	*Tribolium castaneum*	485	1,0E-53								X
180145	fr6	10,2	62653,1		No detected homolog										X	X
180211	fr3		41160,3		No detected homolog							X				X
180286	fr4	9,2	29883,8		No detected homolog							X				
180452	fr3	9,1	35171,4		No detected homolog						X	X				
180732	fr3	10,1	18820,5	gb|EFX72698.1|	No predicted function	*Daphnia pulex*	176	4,0E-12	X		X					
180766	fr1	10,4	13366,6		No detected homolog				X						X	
181254	fr1	6,7	38548,3	ref|XP_001601449.1|	aspartate aminotransferase, cytoplasmic-like	*Nasonia vitripennis*	947	3,0E-122							X	X
181255	fr1	8,7	37187,5	ref|XP_002734509.1|	aldehyde dehydrogenase 2-like	*Saccoglossus kowalevskii*	1009	4,0E-131			X	X	X		X	
181267	fr5	8,6	32695,4	ref|XP_003973218.1|	protein transport protein Sec61 subunit alpha-like 1-like	*Takifugu rubripes*	1174	2,0E-156							X	X
181349	fr3	5,0	36236,4	gb|EFX88851.1|	No predicted function	*Daphnia pulex*	141	9,0E-07		X						
181483	fr2	10,9	22414,3	ref|XP_003748179.1|	histone H1.0-A-like	*Metaseiulus occidentalis*	267	1,0E-25						X	X	X
181526	fr4	11,1	31159,7	gb|AEE36486.1|	protein disulfide isomerase 2	*Fenneropenaeus chinensis*	284	1,0E-25						X	X	
181833	fr2	6,5	13303,4	gb|AFC60660.1|	chitinase	*Pandalopsis japonica*	258	6,0E-24							X	X
182086	fr1	9,5	61435,7	gb|ACQ72825.1|	myosin heavy chain	*Bombyx mandarina*	567	3,0E-60	X			X		X	X	X
182089	fr2	9,2	83018,8	XP_005183384.1	myosin heavy chain, non-muscleisoform X1	*Musca domestica*	644	6,3E-67						X	X	
182115	fr3	9,1	42363,7	ref|XP_801745.2|	Aconitate hydratase	*Strongylocentrotus purpuratus*	878	8,0E-109			X			X	X	
182203	fr2	9,4	49646,5	EKC36585	Angiotensin-converting enzyme	*Crassostrea gigas*	519	1,9E-53								X
182207	fr2	9,4	53648,5	gb|ACN51891.1|	glutamine-oxaloacetic transaminase	*Daphnia magna*	1008	2,0E-129	X							
182248	fr5	5,5	61674,6	emb|CCI71879.1|	gelsolin	*Homarus americanus*	1654	0,0E+00							X	
182450	fr5	8,9	39568,1	ref|XP_002424317.1|	ubiquinol-cytochrome C reductase, core protein 1	*Pediculus humanus corporis*	446	5,0E-47							X	X
182456	fr2	9,5	41259,1	ref|XP_001870703.1|	T-complex protein 1 subunit gamma	*Culex quinquefasciatus*	740	8,0E-93				X				
182640	fr2	5,7	22102,1	XP_005180435	zinc metalloproteinase nas-4-like	*Musca domestica*	316	3,3E-31							X	X
182791	fr2	5,4	29812,2	EFX72588	No predicted function	*Daphnia pulex*	427	1,4E-44								X
182902	fr1	10,4	40860,6	gb|ACR20474.1|	lipopolysaccharide and beta-1,3-glucan binding protein	*Procambarus clarkii*	419	5,0E-44							X	X
183005	fr5	8,6	33149,9	sp|P06754.2|TPM1_DROME	Tropomyosin-1, isoforms 9A/A/B	*Uncharacterized*	630	2,0E-76	X		X	X				
183015	fr6	10,0	77115,4	gb|ADA61117.1|	glutamate dehydrogenase	*Scylla serrata*	621	7,0E-73	X							
183076	fr1	5,5	28074,3	XP_004529383	glycerol-3-phosphate dehydrogenase	*NAD(+)*	841	2,6E-108							X	X
183084	fr4	5,1	31945,6	gb|AFC36877.1|	14-3-3-like protein	*Litopenaeus vannamei*	952	3,0E-126				X			X	X
183126	fr3	8,7	21402,1		No detected homolog						X					
183169	fr5	9,5	25095,4	XP_002005892.1	No predicted function	*Drosophila mojavensis*	506	6,9E-57						X		
183446	fr4	4,8	16336,0	dbj|BAJ23879.1|	glycogen phosphorylase	*Marsupenaeus japonicus*	340	5,0E-34	X							
183562	fr1	6,0	21328,9		No detected homolog				X							
183721	fr4	8,2	27231,0	ref|XP_002199514.2|	phosphoglycerate kinase-like	*Taeniopygia guttata*	720	9,0E-92	X		X	X				
183844	fr1	10,1	97237,7		No detected homolog											X
183846	fr1	9,3	39190,9	gb|AFD28290.1|	14-3-3-like protein	*Eriocheir sinensis*	749	1,0E-94			X			X	X	X
183890	fr1	10,3	93992,3	dbj|BAC66140.1|	projectin	*Procambarus clarkii*	1001	4,0E-112				X				
183919	fr2	8,3	54621,4	gb|ACD68589.1|	peroxiredoxin	*Eurypanopeus depressus*	703	1,0E-86			X	X				
183925	fr1	8,5	48769,7	AFH58810	chitinase, partial	*Macrobrachium nipponense*	287	1,2E-27								X
183941	fr1	9,5	30976,9	gb|ABF18334.1|	ER protein reticulon	*Aedes aegypti*	310	2,0E-30							X	
183988	fr1	9,4	52388,4	gb|AGC10822.1|	Rab-1	*Eriocheir sinensis*	732	4,0E-91	X			X				
184339	fr4	9,2	55127,2	gb|ACR43475.1|	sarcoplasmic calcium-binding protein	*Crangon crangon*	569	7,0E-67	X			X				
184449	fr2	9,4	10317,3		No detected homolog										X	
184643	fr2	8,5	21090,8	EOA92874.1	Tubulin alpha-1C chain	*Anas platyrhynchos*	931	3,7E-124					X			
184668	fr6	10,2	72756,8	gb|EAT44728.2|	No predicted function	*Aedes aegypti*	413	6,0E-39			X		X			
184825	fr5	4,7	30491,6	XP_001608247	myosin regulatory light chain 2-like isoform 1	*Nasonia vitripennis*	246	1,7E-21								X
184962	fr2	8,3	31239,7	ref|XP_003703722.1|	mitochondrial-processing peptidase subunit beta	*Megachile rotundata*	689	1,0E-83	X		X	X				
185409	fr6	8,2	40609,7	gb|ACO36738.1|	ubiquitin carboxyl-terminal esterase L3	*Scylla paramamosain*	561	1,0E-66			X	X				
185442	fr4	9,2	22014,5	gb|AEC50079.1|	calreticulin	*Pacifastacus leniusculus*	400	4,0E-43						X	X	
185484	fr5	9,0	48336,4	gb|EFX88723.1|	No predicted function	*Daphnia pulex*	314	1,0E-27			X		X			
185487	fr3	9,8	31740,2	XP_004923608	filamin-A-like	*Bombyx mori*	727	7,7E-89								X
185498	fr2	10,5	56333,8	gb|ABK59872.1|	ME2	*Tigriopus californicus*	993	2,0E-124	X		X	X		X	X	X
185506	fr1	9,1	36203,6	gb|EFN60616.1|	Heterogeneous nuclear ribonucleoprotein L	*Camponotus floridanus*	503	8,0E-58			X		X	X	X	X
185516	fr6	10,0	47277,4	ref|XP_001655070.1|	paramyosin	*Aedes aegypti*	272	2,0E-22	X							
185626	fr6	9,0	43264,8		No detected homolog						X					
185800	fr3	7,1	37064,1	sp|P04069.1|CBPB_ASTFL	Carboxypeptidase B	*Uncharacterized*	496	3,0E-56						X	X	X
186112	fr4	4,9	23920,8		No detected homolog								X			
186168	fr5	8,3	29978,9	ref|XP_003224567.1|	epididymal sperm-binding protein 1-like	*Anolis carolinensis*	113	2,0E-03							X	X
186195	fr5	10,3	43502,0	gb|ABW77320.1|	clottable protein 2	*Penaeus monodon*	156	3,0E-08							X	X
186201	fr3	6,5	81829,2	emb|CAX46430.1|	hemocyanin subunit 3 precursor	*Odontodactylus scyllarus*	1200	2,0E-151	X		X	X	X	X	X	X
186210	fr5	10,0	42623,6	XP_002423351.1	Filamin-C, putative	*Pediculus humanus corporis*	602	1,7E-66						X		
186217	fr5	5,5	38126,3	XP_001662326	protease m1 zinc metalloprotease	*Aedes aegypti*	805	1,4E-95								X
186222	fr3	9,3	33270,5		No detected homolog				X							
186237	fr2	10,8	69665,4	ref|XP_001983883.1|	No predicted function	*Drosophila grimshawi*	263	1,0E-20	X			X			X	
186255	fr6	9,2	52725,1	ref|XP_004070089.1|	peptidyl-prolyl cis-trans isomerase FKBP4-like	*Oryzias latipes*	295	3,0E-26			X					
186257	fr2	9,6	46320,1	XP_308397	No predicted function	*Anopheles gambiae str. PEST*	546	1,3E-62								X
186286	fr6	9,1	48528,6	gb|EFX80638.1|	No predicted function	*Daphnia pulex*	480	2,0E-53			X					
186340	fr4	6,5	40199,2		No detected homolog								X			
186427	fr6	6,3	56083,4	gb|EFN65250.1|	Nesprin-1	*Camponotus floridanus*	218	2,0E-15	X							
186515	fr4	7,8	38922,5	gb|EFX88976.1|	myosin regulatory light chain	*Daphnia pulex*	357	7,0E-37	X							
186872	fr3	9,5	44241,4	gb|ABO21408.1|	pyruvate kinase	*Litopenaeus vannamei*	974	5,0E-123			X				X	
187499	fr2	9,6	55527,9	ref|XP_003741973.1|	No predicted function	*Metaseiulus occidentalis*	528	1,0E-58						X	X	X
187507	fr1	10,6	91750,9	EFX76382.1	No predicted function	*Daphnia pulex*	140	1,5E-05					X			
187519	fr2	9,2	26461,3	gb|AEO72077.1|	glutamate dehydrogenase	*Eriocheir sinensis*	819	8,0E-103				X			X	
187544	fr2	8,9	38450,7	gb|EFX85885.1|	No predicted function	*Daphnia pulex*	221	2,0E-17			X	X				
187564	fr1	7,9	27584,1		No detected homolog							X			X	X
187640	fr4	10,0	59580,3	dbj|BAJ23879.1|	glycogen phosphorylase	*Marsupenaeus japonicus*	1071	1,0E-132							X	
188118	fr4	11,3	112049,4	ERL95534	No predicted function	*Dendroctonus ponderosae*	509	4,7E-49								X
188129	fr1	10,1	94347,5	gb|AEO72077.1|	glutamate dehydrogenase	*Eriocheir sinensis*	835	2,0E-97	X		X	X			X	X
188130	fr2	9,6	50744,9	gb|AEB40302.1|	ATP synthase gamma subunit precursor	*Litopenaeus vannamei*	439	6,0E-47							X	X

188208	fr1	8,4	58389,3	XP_004529383	glycerol-3-phosphate dehydrogenase	*NAD(+)*	504	8,1E-55							X	X
188212	fr5	9,9	35713,3		No detected homolog										X	X
188305	fr3	6,1	35810,2	gb|AFC36877.1|	14-3-3-like protein	*Litopenaeus vannamei*	1126	4,0E-152	X		X				X	X
188661	fr3	9,4	28321,6		No detected homolog								X	X		
188696	fr1	9,8	69163,8	gb|AEO72077.1|	glutamate dehydrogenase	*Eriocheir sinensis*	1737	0,0E+00	X		X	X			X	X
188731	fr4	8,2	28974,5	ACR78686.1	paramyosin	*Rimicaris exoculata*	378	3,0E-40						X		
188740	fr6	9,5	26965,5	gb|EFR30275.1|	No predicted function	*Anopheles darlingi*	483	4,0E-57						X	X	X
188898	fr1	9,2	54075,0		No detected homolog						X					
189036	fr5	9,7	28637,8		No detected homolog										X	X
189210	fr4	6,4	39758,6	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	731	1,0E-84								X
189277	fr3	7,9	36921,7	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	389	6,2E-38								X
189299	fr5	9,7	47877,5	ref|XP_001657588.1|	short-chain dehydrogenase	*Aedes aegypti*	309	5,0E-29							X	X
189441	fr6	9,2	29497,1	ref|XP_002593173.1|	No predicted function	*Branchiostoma floridae*	458	8,0E-53							X	X
189460	fr4	9,7	36771,7	ref|NP_001093090.1|	NADPH-specific isocitrate dehydrogenase	*Bombyx mori*	1100	3,0E-145							X	X
189695	fr2	9,0	70099,0	gb|AEC50080.1|	alpha-2-macroglobulin	*Pacifastacus leniusculus*	506	2,0E-50			X	X			X	X
189707	fr2	9,4	33983,0	XP_002425566.1	No predicted function	*Pediculus humanus corporis*	338	2,3E-31					X			
189952	fr5	9,3	60519,9	BAC66140	projectin	*Procambarus clarkii*	760	1,1E-83								X
190020	fr5	5,5	20179,3	ref|XP_001948420.2|	paramyosin	*Acyrthosiphon pisum*	246	6,0E-21	X			X			X	
190195	fr1	9,8	97762,7		No detected homolog						X					
190594	fr3	9,6	34668,0		No detected homolog										X	
190624	fr5	8,9	63440,0	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	1002	1,0E-126			X	X	X	X	X	X
191065	fr3		22999,7	gb|ABY66598.1|	pyruvate kinase 3	*Litopenaeus vannamei*	357	3,0E-36						X	X	X
191315	fr3	8,3	41923,5	gb|AFN20599.1|	lectin 3	*Macrobrachium rosenbergii*	157	8,0E-09				X				
191476	fr2	9,4	52643,5	gb|AFQ00932.1|	alpha-tubulin, partial	*Locusta migratoria*	1202	3,0E-159			X		X	X	X	X
191489	fr4	8,9	115411,6	gb|EFN79764.1|	Solute carrier family 2, facilitated glucose transporter member	*Harpegnathos saltator*	657	3,0E-73	X							
191492	fr6	10,5	34955,4	gb|ABY66597.1|	pyruvate kinase 2	*Litopenaeus vannamei*	679	3,0E-81							X	
191493	fr6	8,4	50953,6	gb|AFD28290.1|	14-3-3-like protein	*Eriocheir sinensis*	634	4,0E-76	X			X		X	X	X
191505	fr1	9,5	72710,3		No detected homolog						X		X			
191570	fr6	4,2	22334,1		No detected homolog							X				
191880	fr1	9,1	79179,4	AEE63290.1	No predicted function	*Dendroctonus ponderosae*	145	8,2E-07						X		
191891	fr5	6,8	18242,2	AAW57890.1	hemocyanin subunit 2	*Metacarcinus magister*	310	6,7E-30					X	X		
191918	fr1	9,3	71913,1	ref|NP_001171331.1|	juvenile hormone epoxide hydrolase-like protein 5	*Bombyx mori*	338	5,0E-31				X				
191946	fr3	8,9	33441,8	XP_002593173.1	No predicted function	*Branchiostoma floridae*	478	3,3E-55					X			
192148	fr6	8,8	17117,5	gb|ACH81781.1|	fructose 1,6-bisphosphate aldolase	*Artemia franciscana*	247	7,0E-23				X		X	X	X
192301	fr5	9,7	29130,2	ACZ58291	elongation factor-1 alpha	*Cyamus ovalis*	834	8,8E-106								X
192326	fr6	11,2	41168,5	gb|AFZ93881.1|	troponin T isoform F	*Spodoptera frugiperda*	495	3,0E-55	X		X	X	X	X	X	X
192342	fr4	8,6	35214,5	gb|ACS91344.1|	peroxiredoxin	*Fenneropenaeus indicus*	713	1,0E-90				X				
192443	fr6	9,8	36826,5	ref|XP_002412933.1|	alpha-L-fucosidase,	*Ixodes scapularis*	624	1,0E-73	X							
192652	fr5	9,1	42923,9	emb|CAB85963.1|	alpha glucosidase	*Litopenaeus vannamei*	486	7,0E-51							X	X
192686	fr4		37170,9	AAC03145.1	elongation factor-1 alpha	*Armadillidium vulgare*	524	1,2E-59					X			
192695	fr3	7,1	50086,7	gb|ABH06849.1|	enolase	*Drosophila simulans*	987	3,0E-126	X		X	X	X	X	X	X
192741	fr2	9,6	47538,6	EJW75092	endoplasmic reticulum	*ER*	677	2,9E-82								X
192788	fr6	9,9	50933,8	gb|EMC79730.1|	No predicted function	*Columba livia*	121	3,0E-04							X	X
192906	fr4	8,9	44269,0	ref|XP_003425981.1|	transketolase-like isoform 2	*Nasonia vitripennis*	798	3,0E-96							X	
193227	fr4	9,3	45359,9	gb|ACR78686.1|	paramyosin	*Rimicaris exoculata*	440	5,0E-48			X					
193410	fr3	9,0	50430,7		No detected homolog						X		X			
193455	fr6	8,3	12899,8	dbj|BAL60587.1|	cellulase	*Neomysis intermedia*	162	6,0E-11							X	
193533	fr1	6,0	40644,8	ref|XP_966352.1|	ubiquitin-activating enzyme E1	*Tribolium castaneum*	1201	2,0E-152				X				
193782	fr4	9,7	73648,3	XP_004922008.1	myosin heavy chain, non-muscle	*Bombyx mori*	1910	0,0E+00					X	X		X
193819	fr5	9,4	36482,8	EFX70513.1	No predicted function	*Daphnia pulex*	475	1,0E-51						X		X
193864	fr2	5,4	16537,3	CAB85963	alpha glucosidase	*Litopenaeus vannamei*	249	1,3E-21								X
194188	fr6	10,8	45080,0	gb|ACC95446.1|	glutamate dehydrogenase	*Litopenaeus vannamei*	328	2,0E-30	X							
194190	fr3	8,3	36009,8	gb|ACS91344.1|	peroxiredoxin	*Fenneropenaeus indicus*	747	2,0E-95							X	
194447	fr6	11,1	20331,3	ref|XP_001651219.1|	myosin heavy chain, nonmuscle or smooth muscle	*Aedes aegypti*	184	9,0E-13	X							
194448	fr3	5,7	31695,5	AAW57890.1	hemocyanin subunit 2	*Metacarcinus magister*	790	1,3E-96						X		X
194450	fr6	9,8	64472,7	AAR87379	tropomyosin	*Squilla aculeata*	582	1,6E-66								X
194469	fr4	8,3	21438,2	ref|XP_001351525.1|	erythrocyte membrane-associated antigen,	*Plasmodium falciparum 3D7*	109	3,0E-03							X	
194500	fr4	10,1	42659,6	gb|AEO72077.1|	glutamate dehydrogenase	*Eriocheir sinensis*	702	4,0E-83							X	X
194528	fr4	5,9	56511,8	ref|XP_002026190.1|	No predicted function	*Drosophila persimilis*	1316	7,0E-172	X		X					X
194720	fr2	9,6	52379,7	emb|CCF55382.1|	hemocyanin beta subunit 1	*Atyopsis moluccensis*	390	8,0E-38						X	X	X
194758	fr3	9,7	148658,8	gb|AAL12620.3|	apolipocrustacein	*Penaeus semisulcatus*	263	9,0E-20			X		X	X		
194768	fr2	5,7	23943,5	BAG30740.1	muscle myosin heavy chain	*Papilio xuthus*	433	7,6E-45						X		X
194792	fr1	5,9	28684,0	gb|ABY64692.1|	alpha-2-macroglobulin, partial	*Armadillidium vulgare*	668	8,0E-80							X	X
194796	fr2	6,6	34973,2	gb|AAA17371.1|	fast myosin heavy chain	*Homarus americanus*	649	6,0E-78	X		X	X	X	X	X	X
194802	fr6	10,0	35411,8	ref|XP_002423351.1|	Filamin-C	*Pediculus humanus corporis*	579	2,0E-64				X		X		X
194806	fr6	5,2	28328,4	gb|ACQ72825.1|	myosin heavy chain	*Bombyx mandarina*	752	4,0E-90			X	X		X	X	X
194910	fr5	5,0	32712,7		No detected homolog						X					
195071	fr5	7,0	35473,4	gb|AAN77377.1|	smooth endoplasmic reticulum calcium ATPase	*Porcellio scaber*	738	5,0E-86	X							
195088	fr4	10,1	22462,6		No detected homolog											X
195102	fr4	9,6	49119,9	dbj|BAM65719.1|	myosin heavy chain type 1	*Penaeus monodon*	546	4,0E-57			X	X	X	X	X	X
195111	fr3	6,4	33812,9	gb|EHJ67857.1|	No predicted function	*Danaus plexippus*	933	5,0E-111							X	X
195190	fr6	5,4	38772,0	gb|ABH06849.1|	enolase	*Drosophila simulans*	1011	2,0E-131							X	
195234	fr1	6,9	22849,4	gb|ACS91344.1|	peroxiredoxin	*Fenneropenaeus indicus*	640	3,0E-81				X				
195360	fr5	9,6	45260,5	gb|ABW82155.1|	catalase	*Fenneropenaeus chinensis*	990	2,0E-126	X			X				
195405	fr4	9,6	33538,2	gb|ACZ58321.1|	elongation factor-1 alpha	*Cyamus gracilis*	873	3,0E-111						X	X	
195409	fr2	8,5	20965,2	ACO14623	Cofilin/actin-depolymerizing factor homolog	*Caligus clemensi*	228	1,9E-20								X
195483	fr1	9,0	33012,6	gb|AAC78141.1|	phosphopyruvate hydratase	*Penaeus monodon*	1057	3,0E-139			X	X	X	X	X	X
195541	fr5	9,9	21982,2	ref|XP_002424248.1|	Troponin T	*Pediculus humanus corporis*	277	4,0E-26	X							
195622	fr2	10,3	60509,5	gb|ADZ76135.1|	catalase 2	*Crassostrea gigas*	472	1,0E-51			X		X	X	X	X
195626	fr6	9,5	38089,1		No detected homolog						X					
195667	fr3	10,7	35289,6	AGY36381.1	elongation factor 1-alpha L1	*Alpheus armatus*	278	9,0E-26						X	X	X
195772	fr1	8,4	11272,3	ref|XP_001653750.1|	enolase	*Aedes aegypti*	106	8,0E-04			X	X				
195874	fr1	9,7	57183,5	Q94571.1	Tubulin beta-1 chain	*Homarus americanus*	1009	2,8E-128					X			
195946	fr2	9,6	29045,1	ref|XP_001604099.2|	heterogeneous nuclear ribonucleoprotein L-like	*Nasonia vitripennis*	460	7,0E-50	X							
196124	fr3	9,5	67250,9	gb|ABI95361.1|	hemolymph clottable protein	*Litopenaeus vannamei*	131	8,0E-05			X		X	X		
196145	fr5	10,9	59512,4	ref|XP_001607090.2|	filamin-C-like	*Nasonia vitripennis*	509	3,0E-51							X	
196197	fr5	10,3	65052,0	ABV58636	hemocyanin-like protein	*Metapenaeus ensis*	396	1,0E-40								X
196397	fr2	5,5	19542,8	gb|ABM74408.1|	hemocyanin	*Portunus pelagicus*	374	1,0E-40					X	X	X	X
196580	fr2	9,7	97480,8	gb|EFX83224.1|	No predicted function	*Daphnia pulex*	975	3,0E-112				X			X	X
196609	fr3	6,0	18273,4	XP_003746355	myosin heavy chain, non-muscle-like	*Metaseiulus occidentalis*	226	3,2E-18								X
196615	fr5	10,1	37654,4	ref|XP_001351525.1|	erythrocyte membrane-associated antigen,	*Plasmodium falciparum 3D7*	139	3,0E-06				X			X	X
196745	fr1	8,6	40557,0	BAN17316.1	myosin heavy chain type 4	*Marsupenaeus japonicus*	465	1,3E-51	X		X	X	X	X	X	X
196843	fr3	7,6	58483,9	emb|CAX46431.1|	hemocyanin subunit 4 precursor	*Odontodactylus scyllarus*	621	7,0E-69	X		X	X	X	X	X	X
196858	fr4	9,9	66032,7		No detected homolog								X	X	X	X
196884	fr4	9,1	26640,2	gb|AAB34531.2|	myosin heavy chain	*Drosophila melanogaster*	447	2,0E-49				X				X
197005	fr6	9,8	55409,0	ref|XP_002423351.1|	Filamin-C	*Pediculus humanus corporis*	1201	2,0E-152	X		X	X	X	X	X	X
197099	fr1	9,8	58321,0		No detected homolog								X	X	X	X
197161	fr2	10,6	30373,7	gb|AAT91089.1|	elongation factor 1-alpha	*Pimephales promelas*	407	1,0E-42						X	X	X
197163	fr5	8,9	31726,5	gb|AAD21849.1|	elongation factor 1-alpha	*Heteromysis formosa*	408	2,0E-43				X				
197192	fr5	10,8	21244,3		No detected homolog										X	X
197269	fr2	10,1	29334,0		No detected homolog						X					
197504	fr2	11,8	27612,0	gb|ACR54115.1|	slow muscle myosin S1 heavy chain	*Palaemonetes varians*	107	4,0E-03				X				
197521	fr3	8,5	26226,3	Q9U572.1	Hemolymph clottable protein	*Penaeus monodon*	154	1,3E-08						X		
197578	fr3	9,7	32913,3		No detected homolog									X		
197627	fr4	9,2	69288,8		No detected homolog						X					
197640	fr5	11,6	45140,6	ref|XP_003385684.1|	NAD(P) transhydrogenase, mitochondrial-like	*Amphimedon queenslandica*	537	7,0E-57						X	X	X
197797	fr1	9,2	71974,6	emb|CAX65684.2|	beta-1,3-glucan-binding protein	*Pontastacus leptodactylus*	298	7,0E-25			X	X	X	X	X	X
197838	fr2	5,5	44302,5	ref|XP_002428965.1|	Aminopeptidase N precursor,	*Pediculus humanus corporis*	639	3,0E-71							X	X
197987	fr2	9,9	50613,8		No detected homolog											X
197999	fr3	7,9	31769,6	ACI04663	tropomyosin	*Jasus lalandii*	381	8,2E-41								X
198177	fr2	9,5	91476,7	dbj|BAF91417.1|	vitellogenin	*Upogebia major*	347	3,0E-30			X		X	X		
198233	fr5	10,1	47739,1	gb|ABK59925.1|	clottable protein	*Marsupenaeus japonicus*	120	1,0E-03						X	X	X
198352	fr6	8,9	40961,2	EKC20356.1	Vitellogenin	*Crassostrea gigas*	245	9,4E-20						X		
198368	fr2	5,6	59310,6	gb|ADO00931.1|	calnexin	*Penaeus monodon*	790	4,0E-94				X				X
198421	fr3	9,5	52212,3	gb|ABU41134.1|	vitellogenin 1	*Lepeophtheirus salmonis*	131	5,0E-05	X		X	X	X	X	X	X
198446	fr5	10,1	50104,2	gb|ABQ95505.1|	14-3-3-like protein	*Litopenaeus vannamei*	446	2,0E-50						X	X	
198451	fr2	9,2	51041,0	ref|XP_003741973.1|	No predicted function	*Metaseiulus occidentalis*	495	4,0E-54				X				X
198506	fr4	8,4	48602,5	CAX46431.1	hemocyanin subunit 4 precursor	*Odontodactylus scyllarus*	544	5,0E-59						X		X
198532	fr3	6,2	31735,5	WP_019668393	No predicted function	*Eudoraea adriatica*	358	2,4E-35								X
198702	fr4	8,9	53237,1	ref|XP_315409.2|	No predicted function	*Anopheles gambiae str. PEST*	844	1,0E-105			X					
198854	fr5	7,7	34516,3	gb|AAY56092.1|	14-3-3-like protein	*Penaeus monodon*	330	6,0E-33					X	X	X	X
198862	fr2	8,9	35901,3	ABO09863.1	vitellogenin	*Homarus americanus*	241	6,6E-19					X	X		
198903	fr6	10,0	16339,4		No detected homolog						X					
198995	fr4	9,4	44674,5	gb|ABK59925.1|	clottable protein	*Marsupenaeus japonicus*	133	2,0E-05						X	X	X
199136	fr3	7,6	20451,7	emb|CCQ42788.1|	catalase	*Leucosolenia complicata*	471	3,0E-54				X				
199312	fr3	9,6	36352,2		No detected homolog									X		
199470	fr6	8,8	17772,0	dbj|BAF95206.1|	tropomyosin	*Oratosquilla oratoria*	485	2,0E-57							X	
199575	fr4	10,1	104911,0		No detected homolog								X			
199594	fr2	10,1	65764,2	gb|EFX84778.1|	No predicted function	*Daphnia pulex*	1107	1,0E-136				X			X	X
199600	fr1	8,3	37587,7	dbj|BAF76431.1|	tropomyosin	*Euphausia pacifica*	595	4,0E-71	X			X				

199608	fr3	8,2	50793,8	dbj|BAF76431.1|	tropomyosin	*Euphausia pacifica*	744	8,0E-92	X			X	X			X	X
199772	fr1		20347,7	gb|EHB00855.1|	Calmodulin	*Heterocephalus glaber*	528	9,0E-65				X	X				
199807	fr1	8,5	28571,2	dbj|BAF95206.1|	tropomyosin	*Oratosquilla oratoria*	665	5,0E-83				X	X		X	X	X
199971	fr3	9,9	102777,4	gb|AAD16454.1|	clotting protein precursor	*Pacifastacus leniusculus*	231	3,0E-16	X			X	X	X	X		
200124	fr5	6,7	23741,1	XP_002738641.1	No predicted function	*Saccoglossus kowalevskii*	410	2,0E-47						X			
200269	fr2	10,0	29290,7	XP_006039481.1	tubulin, alpha 1a	*Alligator sinensis*	683	1,4E-83							X	X	X
200301	fr2	9,5	35093,0	gb|ACG63697.1|	myosin heavy chain	*Gammarus duebeni*	820	9,0E-103	X			X	X	X	X	X	X
200386	fr3	9,0	22314,2	ref|XP_002115153.1|	No predicted function	*Trichoplax adhaerens*	886	6,0E-115					X				
200426	fr2	9,0	81251,0	gb|AAT01139.1|	vitellogenin	*Metapenaeus ensis*	115	8,0E-03				X		X	X		
200434	fr5	9,1	33282,5		No detected homolog									X				
200435	fr2	10,5	51935,2	gb|EFA05963.1|	No predicted function	*Tribolium castaneum*	152	9,0E-08					X				
200559	fr3	9,1	26364,5		No detected homolog							X			X	X		
200648	fr2	9,7	39394,6	ABI95361.1	hemolymph clottable protein	*Litopenaeus vannamei*	235	4,3E-18							X			
200701	fr2	11,5	71814,0	gb|ABS57456.1|	alpha-tubulin	*Heliconius melpomene*	830	3,0E-103	X			X	X	X	X	X	X
200739	fr2	11,1	16077,9		No detected homolog							X						
200782	fr1	10,6	36772,3	ref|XP_002066324.1|	No predicted function	*Drosophila willistoni*	699	5,0E-79				X	X	X	X	X	X
201021	fr5	11,4	13928,1		No detected homolog					X		X	X					
201173	fr6	9,9	20382,4	gb|ELT97844.1|	No predicted function	*Capitella teleta*	251	2,0E-22				X					
201238	fr2	9,5	48477,3	ref|XP_802011.1|	tubulin alpha chain-like isoform 5	*Strongylocentrotus purpuratus*	1196	1,0E-157				X				X	
201362	fr4	10,7	79720,9		No detected homolog							X		X	X			
201379	fr2	10,0	26561,6	dbj|BAM65721.1|	myosin heavy chain type 1	*Litopenaeus vannamei*	617	3,0E-69							X	X	X
201400	fr3	11,5	27003,6	gb|AAU20853.2|	endogenous cellulase	*Reticulitermes flavipes*	198	1,0E-14	X							X	X
201522	fr3	9,7	62106,7	gb|AAT01139.1|	vitellogenin	*Metapenaeus ensis*	108	3,6E-02				X		X	X		
201527	fr6	7,1	30777,4		No detected homolog										X			
201587	fr3	9,8	59582,6	BAD11098.1	vitellogenin	*Pandalus hypsinotus*	321	6,8E-28						X	X			
201667	fr3	8,9	41893,2	gb|ABX89617.1|	vitellogenin	*Callinectes sapidus*	419	5,0E-44				X		X	X			
201744	fr2	8,5	42733,9		No detected homolog									X	X			
201834	fr4	10,6	45773,4		No detected homolog										X			
201844	fr3	10,0	75208,4	pdb|1E32|A	No predicted function	*Uncharacterized*	656	2,0E-74							X	X	X	
201865	fr3	6,5	26820,7	AGM75775.1	vitellogenin	*Eriocheir sinensis*	334	2,6E-31				X		X	X			
201978	fr3	9,8	28714,6	AGM75775.1	vitellogenin	*Eriocheir sinensis*	136	3,5E-06							X			
202109	fr3	10,1	63707,9	gb|ACD63049.1|	heat shock cognate 70 protein	*Loxostege sticticalis*	572	3,0E-63				X	X	X		X	X
202279	fr1	6,6	82952,3	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	1729	0,0E+00	X			X	X	X	X	X	X
202300	fr6	9,9	62474,1	CAI78901.1	hemocyanin subunit 1	*Gammarus roeselii*	306	3,3E-26						X				
202517	fr3	5,5	22688,2	XP_005814809	neural cell adhesion molecule L1.1-like, partial	*Xiphophorus maculatus*	103	2,8E-02									X	
202609	fr5	10,6	32281,1	dbj|BAA12287.1|	chitinase	*Marsupenaeus japonicus*	334	3,0E-32								X		
202637	fr1	8,0	89595,4	emb|CAI78901.1|	hemocyanin subunit 1	*Gammarus roeselii*	1668	0,0E+00	X			X	X	X	X	X	X
202705	fr3	8,2	57172,0	ref|XP_001897491.1|	Tubulin alpha chain	*Brugia malayi*	988	2,0E-126							X	X	X
202894	fr3	6,8	61235,5	EFX72698	No predicted function	*Daphnia pulex*	489	4,5E-50									X
203291	fr2	9,7	77274,6		No detected homolog									X				
203298	fr1	9,7	22837,4	XP_003110714.1	CRE-TBB-1 protein	*Caenorhabditis remanei*	728	7,3E-91						X				
203444	fr6	10,1	12329,2	ref|XP_003994105.1|	annexin A7 isoform 4	*Felis catus*	217	6,0E-19	X							X	X
203606	fr2	4,7	19921,7		No detected homolog				X							X		
203611	fr1	9,7	105046,8	XP_002199483	clathrin heavy chain 1-like, partial	*Taeniopygia guttata*	713	2,3E-82									X	
203627	fr2	6,0	21891,6	gb|ADE60733.1|	myosin essential light chain	*Eriocheir sinensis*	429	3,0E-50				X	X					
203689	fr2	6,8	49905,1		No detected homolog							X						
203698	fr1	9,6	51464,9	gb|ADF81052.1|	vitellogenin-like protein precursor [Gammarus fossarum]	*Gammarus fossarum*	209	5,0E-16			X							
203707	fr1	10,5	16774,0		No detected homolog							X						
203833	fr2	6,8	41533,8	ref|NP_001093319.1|	heterogeneous nuclear ribonucleoprotein A1	*Bombyx mori*	521	3,0E-59				X	X	X	X	X	X
203938	fr3	8,7	34075,9		No detected homolog									X	X			
204366	fr5	9,6	23464,6	AAN72821.1	beta-I tubulin	*Callinectes sapidus*	824	2,6E-108						X				
204380	fr6	9,8	57786,4	gb|ACY44665.1|	phosphogluconate dehydrogenase	*Leiobunum verrucosum*	362	3,0E-36								X	X	
204388	fr1	9,5	41884,2	gb|ELT97182.1|	No predicted function	*Capitella teleta*	181	2,0E-11					X					
204527	fr2	9,1	92601,4	BAN21261	moesin/ezrin/radixin homolog 1	*Riptortus pedestris*	1154	2,5E-144								X	X	
204630	fr1		96924,0	tpg|DAA34772.1|	TPA_inf: dihydrolipoamide dehydrogenase	*Amblyomma variegatum*	609	5,0E-69								X		
204711	fr1	8,9	23793,2	gb|AFP57558.1|	heat shock cognate protein 70, partial	*Latrodectus hesperus*	781	3,0E-101				X		X				
204825	fr6	9,3	44804,7	gb|ABI95361.1|	hemolymph clottable protein	*Litopenaeus vannamei*	172	4,0E-10				X		X	X			
204886	fr1	10,4	49794,8	gb|AFI60312.1|	heat shock protein 70, partial	*Eulimnogammarus cruentus*	525	4,0E-61								X		
204978	fr4	10,7	22444,4		No detected homolog							X		X	X			
204994	fr3	7,5	41853,8	AAN72821.1	beta-I tubulin	*Callinectes sapidus*	592	7,2E-71						X	X		X	
205010	fr4	8,6	30335,6	ref|XP_004365566.1|	peptidylprolyl isomerase	*Capsaspora owczarzaki ATCC 308*	713	2,0E-91	X			X	X	X	X	X	X	
205033	fr1	11,4	77936,6		No detected homolog									X	X		X	
205041	fr4	10,1	13443,9		No detected homolog							X						
205082	fr4	8,8	36811,8	sp|Q94571.1|TBB2_HOMAM	Tubulin beta-2 chain	*Homarus americanus*	1070	2,0E-140								X	X	
205133	fr6	8,7	31739,9		No detected homolog							X		X	X			
205245	fr3	9,5	20438,4		No detected homolog							X						
205250	fr5	10,5	34460,9		No detected homolog							X						
205368	fr4	8,7	37389,4		No detected homolog										X			
205425	fr5	9,2	41564,8		No detected homolog									X				
205532	fr1	8,1	19837,2		No detected homolog								X					
205588	fr2	5,9	16234,2	gb|EFX71611.1|	No predicted function	*Daphnia pulex*	425	2,0E-46								X		
205688	fr2	9,2	14974,4		No detected homolog							X						
205707	fr3	9,5	29068,5	dbj|BAG65398.1|	No predicted function	*Homo sapiens*	556	3,0E-66								X		
205875	fr1	9,3	28491,8	ref|XP_002612643.1|	No predicted function	*Branchiostoma floridae*	600	4,0E-70								X	X	
205967	fr4	9,7	37685,6	AHA44483.1	heat shock protein 70	*Coreius guichenoti*	896	1,4E-111						X				
205972	fr6	8,2	55581,0	EFX87974.1	No predicted function	*Daphnia pulex*	1074	2,1E-138						X				
206033	fr1	10,3	17625,6	gb|EFX71611.1|	No predicted function	*Daphnia pulex*	481	2,0E-54					X					
206392	fr5	11,5	26851,5	gb|AEY75235.1|	Hsp8, partial	*Bactrocera dorsalis*	157	1,0E-10	X							X		
206469	fr3	9,3	73042,8	gb|AAD16454.1|	clotting protein precursor	*Pacifastacus leniusculus*	133	5,0E-05				X		X	X			
206496	fr4	6,6	23765,6		No detected homolog				X				X	X	X			
206518	fr4	11,2	32131,5		No detected homolog							X		X	X			
206576	fr4	9,4	18965,6		No detected homolog							X		X	X			
206681	fr4	10,1	108766,4	gb|EFN61569.1|	Vigilin	*Camponotus floridanus*	895	3,0E-99	X									
206684	fr4	8,3	22934,1		No detected homolog										X			
206723	fr1	9,8	39677,2		No detected homolog							X		X	X			X
206753	fr4	7,2	42154,3		No detected homolog				X			X	X	X	X			
206816	fr3		23093,4		No detected homolog									X	X			
206819	fr1	9,7	20820,3		No detected homolog							X						
206828	fr3	10,9	21337,8		No detected homolog									X				
206860	fr2	8,9	36634,6	ref|XP_003708652.1|	heat shock 70 kDa protein cognate 4-like	*Megachile rotundata*	1236	6,0E-163				X						
206895	fr3	9,4	43580,4	gb|AAQ63611.1|	70kD heat shock-like protein	*Procambarus clarkii*	829	1,0E-104				X	X	X	X	X	X
206906	fr6	5,2	9770,9	gb|AGC96524.1|	70kD heat shock protein, partial	*Scylla paramamosain*	218	7,0E-20	X									
206956	fr2	10,0	24062,3	gb|AEX65804.1|	heat shock protein 70	*Gammarus lacustris*	250	2,0E-21				X						
207005	fr1	9,1	29587,2		No detected homolog													X
207048	fr3	9,0	23066,4	gb|ABM92447.1|	glucose-regulated protein 78	*Fenneropenaeus chinensis*	496	4,0E-55						X	X	X		
207064	fr4	9,5	44155,1		No detected homolog							X						
207095	fr1		16551,5	ref|ZP_09465633.1|	ABC transporter-like protein	*Acetivibrio cellulolyticus CD2*	92	1,0E-01				X						
207097	fr1	10,0	35893,0	gb|ADF81052.1|	vitellogenin-like protein precursor	*Gammarus fossarum*	313	7,0E-31				X		X	X			
207130	fr2	10,7	75555,6		No detected homolog							X		X				
207131	fr4	9,5	33375,1		No detected homolog								X					
207133	fr2	10,0	27469,1		No detected homolog							X						
207135	fr3	6,5	30732,8		No detected homolog							X						
207330	fr5	9,4	34782,3		No detected homolog							X						
207414	fr5	9,1	31834,5		No detected homolog							X						
207425	fr2	8,6	25333,9		No detected homolog									X				
207567	fr4	9,7	87752,4	ref|XP_003487165.1|	trifunctional enzyme subunit alpha, mitochondrial-like	*Bombus impatiens*	1178	1,0E-145								X	X	
207591	fr6	8,2	30505,1	ref|XP_001867082.1|	cytochrome B5	*Culex quinquefasciatus*	321	2,0E-33					X					
207783	fr6	6,8	30905,7	ref|XP_001867082.1|	cytochrome B5	*Culex quinquefasciatus*	316	1,0E-32				X						
207844	fr2	8,3	91989,2	AGI74967.1	heat shock protein 60	*Scylla paramamosain*	1257	2,7E-159	X			X	X	X	X	X	X	
207937	fr5	10,1	81396,7	gb|AFA36427.1|	heat shock protein 60	*Portunus trituberculatus*	1002	8,0E-123				X	X		X	X	X
207957	fr5	9,9	35288,4	ACL00842	heat shock protein 60	*Biomphalaria glabrata*	615	3,7E-71										X
208021	fr3	10,0	55762,5	AGD79959.1	heat shock protein 60	*Marsupenaeus japonicus*	1219	3,1E-158						X	X	X	X	
208036	fr5	10,5	27841,3	gb|ACL00842.1|	heat shock protein 60	*Biomphalaria glabrata*	686	1,0E-82				X		X	X	X	X	
208049	fr4	8,9	26973,7	AFM73647.1	heat shock protein 60	*Bicyclus anynana*	697	1,1E-88						X				
208079	fr3	5,3	40506,1	emb|CCA94937.1|	glycoside hydrolase family 7	*Uncharacterized*	508	5,0E-56					X					
208131	fr2	9,7	31528,0	gb|ABU49887.1|	receptor for activated protein kinase c1	*Penaeus monodon*	1011	2,0E-134					X		X			
208354	fr2	11,2	37436,2	sp|P41569.1|RL8_AEDAL	60S ribosomal protein L8	*Aedes albopictus*	494	2,0E-56	X			X	X		X		X	
208477	fr1	9,7	23969,5	gb|AFW99839.1|	troponin I	*Litopenaeus vannamei*	301	3,0E-30	X				X	X	X	X		
208646	fr4	9,7	58843,5	gb|AFE88209.1|	protaglandin reductase 1	*Penaeus monodon*	650	3,0E-76									X	
208855	fr6	10,5	44807,3	ref|XP_001949866.1|	protein singed-like	*Acyrthosiphon pisum*	345	8,0E-33	X									
208953	fr6	9,4	64907,7	ref|XP_002428695.1|	citrate synthase,	*Pediculus humanus corporis*	719	2,0E-84	X								X	
209438	fr3	8,5	78604,6	gb|AFM54541.1|	Na+/K+_ATPase	*Exopalaemon carinicauda*	2159	0,0E+00	X			X	X	X	X	X	X	
209622	fr3	9,9	54259,7	gb|AAT48984.1|	cathepsin B-like proteinase	*Triatoma sordida*	341	1,0E-32				X						
209919	fr3	10,0	34377,7	sp|P05547.1|TNNI_PONLE	Troponin I	*Uncharacterized*	491	3,0E-57									X	X
209926	fr1	7,0	43578,0	gb|AFW99839.1|	troponin I	*Litopenaeus vannamei*	485	5,0E-55	X			X	X				X	X
209971	fr2	5,9	10733,4	ref|XP_001688628.1|	No predicted function	*Anopheles gambiae str. PEST*	282	3,0E-29							X	X	X	
209980	fr2	8,4	17414,0	ref|XP_003738598.1|	troponin I 4-like	*Metaseiulus occidentalis*	252	9,0E-24					X					
210231	fr4	9,2	29039,9	gb|EHJ79048.1|	No predicted function	*Danaus plexippus*	144	1,0E-07				X		X				
210246	fr2	10,6	36754,1		No detected homolog							X		X				
210317	fr1	10,0	15677,9		No detected homolog							X		X				

210459	fr6	10,0	36346,8		No detected homolog								X					
210517	fr6	9,8	26075,4		No detected homolog								X					
210532	fr6	8,9	44629,8	gb|AAQ72746.1|	sodium/potassium ATPase alpha subunit	*Caprella sp. FEA-2003*	1011	9,0E-131	X						X	X	X
210555	fr5	8,8	61962,1	pdb|1E32|A	No predicted function	*Uncharacterized*	1359	2,0E-180							X	X	X
210630	fr5	8,0	40130,2	dbj|BAM99026.1|	valosin containing protein, partial	*Struthio camelus*	1056	1,0E-139	X			X	X			X	
210664	fr5	6,1	48478,9	ref|XP_002425193.1|	No predicted function	*Pediculus humanus corporis*	1348	8,0E-176	X			X	X				
210668	fr5	5,4	30556,1	AGZ01980.1	valosin-containing protein	*Penaeus monodon*	767	1,2E-91				X	X		X		
210817	fr6	9,7	56907,8	gb|ADW10720.1|	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	511	5,0E-52				X	X	X		X	
210818	fr4	9,8	55225,5	gb|ADW10720.1|	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	394	4,0E-37				X	X				
210873	fr1	9,8	43470,8	ref|XP_002084551.1|	No predicted function	*Drosophila simulans*	644	2,0E-77								X	
210885	fr3	8,9	51355,3	gb|AFM54541.1|	Na+/K+_ATPase	*Exopalaemon carinicauda*	1412	0,0E+00	X					X	X	X	X
211002	fr6	6,1	24105,3	gb|AAQ72746.1|	sodium/potassium ATPase alpha subunit	*Caprella sp. FEA-2003*	871	7,0E-113							X	X	
211177	fr4	9,0	80539,9	EFX65940	No predicted function	*Daphnia pulex*	801	3,1E-89									X
211227	fr4	8,5	51335,3	EKC20356	Vitellogenin	*Crassostrea gigas*	419	3,7E-42									X
211426	fr6	6,4	37823,1	XP_002731917.1	putative gut cathepsin Daspartic protease	*Saccoglossus kowalevskii*	586	2,6E-68						X			
212218	fr6	9,3	37829,4	gb|ACN51928.1|	phosphoglucose isomerase	*Daphnia magna*	1013	7,0E-132					X				
212926	fr5	9,2	46276,7	ref|XP_974030.1|	Annexin IX CG5730-PC	*Tribolium castaneum*	260	3,0E-22					X				
212929	fr6	7,6	16277,9	ref|XP_003473838.1|	annexin A7-like isoform 2	*Cavia porcellus*	255	4,0E-23								X	X
212968	fr1	10,2	40289,5	ACG60512	chitinase 1 precursor	*Scylla serrata*	179	3,4E-11									X
213136	fr3	7,2	40693,3	CAD54729	beta-1,4-endoglucanase	*Mastotermes darwiniensis*	556	7,3E-63									X
213184	fr3	9,2	62087,5	AAK12339	cellulase	*Coptotermes acinaciformis*	702	4,6E-82									X
213294	fr5	10,1	20186,1	gb|ACY70393.1|	endoglucanase	*Porcellio scaber*	223	6,0E-19					X				
213317	fr5	5,1	20384,7	gb|AAO61672.2|	cellulase GHF9	*Cherax quadricarinatus*	491	5,0E-56	X				X			X	
213376	fr3	10,1	33101,7	gb|ADB12483.1|	endo-beta-1,4-glucanase	*Coptotermes formosanus*	314	1,0E-29								X	X
213827	fr4	11,1	38358,2	ref|XP_001950997.1|	annexin-B11-like	*Acyrthosiphon pisum*	366	5,0E-36								X	X
214554	fr4	9,4	36527,0	gb|ACP19740.1|	thymosin-repeated protein 1	*Eriocheir sinensis*	285	9,0E-28	X								
214902	fr3	9,8	34768,6	gb|EFA01066.1|	No predicted function	*Tribolium castaneum*	487	1,0E-53				X					
215120	fr5	8,5	24525,3	gb|ADY80039.1|	thymosin	*Procambarus clarkii*	215	1,0E-18	X				X				
215408	fr6	11,7	19972,7	EFR26430	No predicted function	*Anopheles darlingi*	355	1,4E-38									X
216324	fr4	6,0	23276,6	gb|AEE25938.1|	sodium potassium-transporting ATPase subunit beta	*Litopenaeus vannamei*	545	3,0E-65					X				
216470	fr4	10,1	89950,3	gb|ACG60513.1|	chitinase 1 precursor	*Litopenaeus vannamei*	368	1,0E-33								X	
216753	fr2	9,9	40208,6	XP_002738122	trifunctional protein, alpha subunit-like, partial	*Saccoglossus kowalevskii*	766	2,6E-93									X
216880	fr6	9,7	72169,3	gb|ADW10720.1|	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	479	5,0E-47				X	X		X	X	X
216940	fr5	9,7	19689,4	gb|EHJ74477.1|	hydroxyacyl-coenzyme A dehydrogenase	*Danaus plexippus*	267	7,0E-24								X	X
216982	fr4	9,7	68940,5		No detected homolog										X		
217140	fr3	9,8	48785,6		No detected homolog										X		
217166	fr2	9,8	21751,3	gb|AFN66646.1|	glucan pattern-recognition lipoprotein	*Fenneropenaeus chinensis*	303	4,0E-28	X							X	
217264	fr1	10,0	27276,2	gb|AFN66646.1|	glucan pattern-recognition lipoprotein	*Fenneropenaeus chinensis*	324	2,0E-30				X					
217521	fr5	9,1	30455,1	CAX65684	beta-1,3-glucan-binding protein	*Pontastacus leptodactylus*	413	1,2E-41									X
217581	fr6	9,7	33271,0	sp|P81182.2|BGBP_LITVA	Beta-1,3-glucan-binding protein	*Litopenaeus vannamei*	371	6,0E-36				X	X			X	
217606	fr6	8,8	33630,8		No detected homolog										X		
217647	fr4	9,7	21161,0	emb|CAA56703.1|	beta-1,3-D-glucan binding protein	*Pacifastacus leniusculus*	186	6,0E-13					X			X	X
217669	fr1		46444,0	gb|ADW10720.1|	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	508	1,0E-52					X				
217725	fr5	8,8	20746,2	gb|ADW10720.1|	beta-1,3-glucan-binding protein	*Fenneropenaeus chinensis*	247	8,0E-21								X	X
217864	fr5	10,5	40413,3		No detected homolog									X	X		
217964	fr3	8,9	29476,0	gb|ADD70028.1|	myosin light chain	*Marsupenaeus japonicus*	526	1,0E-63	X								
218127	fr6	8,9	47021,3	gb|EFX82342.1|	No predicted function	*Daphnia pulex*	131	2,0E-05				X					
218527	fr2	5,8	37707,6	ACR15996.1	serine protease 46	*Mamestra configurata*	143	2,0E-07						X	X		
218528	fr5	10,1	44011,9		No detected homolog							X					

Annexe II : Protéines identifiées comme étant d'intérêt en écotoxicologie sur la base de leur annotation fonctionnelle

Candidat biomarqueur	Contig_ID
Catalase	45331_fr1; 45375_fr4; 110912_fr3; 174113_fr3; 10004_fr1; 45051_fr1; 45296_fr3; 195360_fr5; 3319_fr5; 199136_fr3; 195622_fr2
Peroxidase	32837_fr1
Superoxyde dismutase	105797_fr6; 176048_fr6
Glutathion-S-transférase	105797_fr6; 176048_fr6; 105797_fr6; 176048_fr6; 150268_fr2; 121871_fr3; 40079_fr3; 88066_fr3; 158043_fr6; 142711_fr4
Cellulase	213136_fr3; 2460_fr1; 14600_fr1; 4242_fr1; 213184_fr3; 193455_fr6; 213317_fr5; 30189_fr3
Chitinase	105035_fr3; 42149_fr5; 181833_fr2; 202609_fr5; 212968_fr1; 216470_fr4; 144157_fr4; 145_fr4; 183925_fr1
Cellobiohydrolase (famille 7)	45660_fr2; 49678_fr2; 49680_fr4; 49809_fr1; 101957_fr1
Amylase	153198_fr5
Peptide antimicrobien	4315_fr5
Protéine de reconnaissance du 1,3-bêta-glucan	4631_fr3; 32111_fr5; 34912_fr1; 217647_fr4; 42502_fr1; 9667_fr6; 217581_fr6; 4646_fr4; 24870_fr6; 217521_fr5; 21251_fr1; 197797_fr1; 216880_fr6; 217725_fr5; 210817_fr6; 210818_fr4; 217669_fr1; 626_fr4; 217166_fr2; 217264_fr1
Hémocytine	160381_fr4461_fr11094_fr31975_fr6
Hémolectine	11145_fr5; 7908_fr4; 166361_fr5; 109695_fr1
Prophénoloxidase	2562_fr2; 6945_fr6; 15561_fr1; 58988_fr5
Transglutaminase	1917_fr2; 7169_fr4; 2415_fr3; 21392_fr2
Na^+/K^+ ATP ase	17127_fr3; 32234_fr1; 105132_fr2; 110907_fr2; 209438_fr3; 210885_fr3; 105132_fr2; 110907_fr2; 209438_fr3; 210885_fr3
Dopamine beta hydroxylase	153_fr2

Annexe III : Protéines définies comme présentant un dysmorphisme sexuel d'après l'analyse comparative du protéome des gonades mâles et femelles

Contig_ID	Fonction prédite	Fold Change	pValue
2029_fr2	projectin	34,3	0,04
6444_fr2	I-connectin	20,3	0,01
3157_fr1	unknown function	19,3	0,00
22255_fr2	myosin heavy chain	18,7	0,02
608_fr6	unknown function	18,0	0,03
169648_fr6	paramyosin, partial	17,4	0,01
169622_fr1	unknown function	16,0	0,00
23125_fr3	Nesprin-1	14,0	0,03
69913_fr6	unknown function	13,2	0,00
2562_fr2	prophenoloxidase	13,0	0,02
470_fr6	unknown function	12,8	0,00
1019_fr4	unknown function	12,7	0,00
12497_fr5	unknown function	12,7	0,00
45519_fr4	No detected homolog	12,0	0,00
41933_fr1	smooth endoplasmic reticulum calcium ATPase	12,0	0,04
461_fr1	hemocytin-like	11,3	0,01
34845_fr5	Copine-8	11,0	0,00
151942_fr2	No detected homolog	11,0	0,04
51491_fr2	trans-2-enoyl-CoA reductase, mitochondrial precursor	10,4	0,00
161076_fr4	Lit v 3 allergen myosin light chain	10,3	0,00
24754_fr2	unknown function	10,3	0,03
100102_fr1	skeletal alpha-actin	10,0	0,00
6689_fr5	calpain B	10,0	0,00
49159_fr2	unknown function	9,5	0,00
153779_fr4	dehydrogenase,glyceraldehydephosphate	9,3	0,00
11594_fr5	actin1	9,3	0,00
4227_fr2	calcium-transporting ATPase sarcoplasmic/endoplasmi	9,3	0,01
190020_fr5	paramyosin, long form-like	9,0	0,00
1975_fr6	hemocytin-like	8,7	0,00
29320_fr2	laminin A chain	8,7	0,01
180286_fr4	No detected homolog	8,0	0,00
4456_fr5	unknown function	8,0	0,02
2459_fr1	utp-glucose-1-phosphate uridylyltransferase 2	8,0	0,01
38421_fr6	clathrin heavy chain	8,0	0,00
25794_fr6	unknown function	7,7	0,00
200386_fr3	conserved hypothetical protein	7,7	0,00
4565_fr1	calmin-like protein	7,6	0,04
14118_fr1	unknown function	7,5	0,00
44126_fr2	tropomyosin	7,4	0,01
88566_fr1	unknown function	7,0	0,02
18473_fr2	Copine-8	7,0	0,00
14348_fr5	alpha-3-tubulin	6,7	0,02
176301_fr2	No detected homolog	6,7	0,00
186237_fr2	unknown function	6,7	0,00
157736_fr6	60S ribosomal protein L4	6,7	0,00
25178_fr2	Cytosolic non-specific dipeptidase	6,3	0,01
1742_fr4	electron transfer flavoprotein subunit beta-like	6,3	0,03
22266_fr2	myosin heavy chain	6,3	0,00
29315_fr3	unknown function	6,3	0,00
7953_fr6	unknown function	6,3	0,01
156_fr5	Spectrin alpha chain,	6,2	0,01
28559_fr5	I-connectin	6,0	0,01
209926_fr1	troponin I	6,0	0,02
10101_fr5	Nesprin-1	6,0	0,02
30230_fr1	myosin regulatory light chain	6,0	0,02
42833_fr6	unknown function	6,0	0,00
6766_fr3	No detected homolog	6,0	0,00
46249_fr5	fructose-bisphosphate aldolase-like isoform 1	6,0	0,03
201021_fr5	No detected homolog	5,8	0,03
2853_fr3	sarcoplasmic calcium-binding protein variant a	5,8	0,02
18951_fr5	No detected homolog	5,7	0,03
2839_fr5	unknown function	5,7	0,00
61104_fr4	Kettin	5,7	0,00
13269_fr1	I-connectin	5,7	0,02
210630_fr5	valosin containing protein, partial	5,7	0,00
68130_fr4	unknown function	5,4	0,00
199594_fr2	unknown function	5,3	0,03
105848_fr3	unknown function	5,3	0,01
422_fr4	myosin heavy chain	5,2	0,01

2051_fr3	unknown function	5,0	0,01
45867_fr6	No detected homolog	5,0	0,01
121141_fr1	myosin regulatory light chain	5,0	0,02
14_fr3	slow muscle myosin S1 heavy chain	4,9	0,00
46326_fr4	actin 2	4,8	0,00
30449_fr6	unknown function	4,8	0,01
12681_fr4	myosin heavy chain, muscle isoform 3	4,7	0,01
30561_fr3	smooth endoplasmic reticulum calcium ATPase	4,7	0,00
200701_fr2	alpha-tubulin	4,7	0,00
105830_fr2	parvin	4,7	0,00
13026_fr5	unknown function	4,7	0,02
187564_fr1	No detected homolog	4,3	0,03
1574_fr6	unknown function	4,3	0,00
2775_fr2	Alanine aminotransferase,	4,2	0,02
126175_fr5	myosin light chain 2	4,2	0,01
8483_fr6	unknown function	4,0	0,04
1024_fr3	unknown function	4,0	0,00
584_fr2	unknown function	4,0	0,00
5763_fr2	unknown function	4,0	0,00
7525_fr6	unknown function	4,0	0,02
110946_fr1	alpha-L-fucosidase-like	4,0	0,02
5764_fr6	unknown function	4,0	0,00
9438_fr2	unknown function	3,9	0,03
4696_fr1	Myosin heavy chain, non-muscle	3,9	0,03
2235_fr4	filamin	3,8	0,03
108041_fr1	Cytochrome c	3,8	0,00
106948_fr3	60S acidic ribosomal protein P2-like	3,8	0,02
1917_fr2	transglutaminase	3,8	0,02
19019_fr5	fascilin 2-like protein	3,7	0,00
213317_fr5	cellulase GHF9	3,7	0,00
18581_fr5	unknown function	3,7	0,00
3948_fr4	GTP binding protein alpha subunit Go	3,7	0,02
21366_fr6	I-connectin	3,7	0,01
8186_fr2	unknown function	3,7	0,01
20915_fr6	aspartate aminotransferase, cytoplasmic-like	3,4	0,03
68651_fr2	unknown function	3,3	0,03
27649_fr3	myosin light chain 2	3,3	0,01
144332_fr2	pancreatic amylase mRNA	3,3	0,04
204388_fr1	unknown function	3,3	0,00
191918_fr1	juvenile hormone epoxide hydrolase-like protein 5	3,3	0,00
111405_fr2	28 kDa heat- and acid-stable phosphoprotein-like	3,3	0,03
32503_fr6	unknown function	3,3	0,01
3203_fr6	Hex-A	3,3	0,00
4581_fr2	widerborst CG5643-PC	3,3	0,00
15457_fr5	No detected homolog	3,3	0,00
173980_fr5	arginine kinase	3,3	0,00
108163_fr3	unknown function	3,3	0,00
87602_fr1	alpha-tubulin	3,2	0,00
11635_fr5	unknown function	3,1	0,02
130227_fr1	fatty acid binding protein	3,1	0,02
5059_fr2	Flotillin-1	3,0	0,00
52575_fr4	heterogeneous nuclear ribonucleoprotein H-like	3,0	0,00
49262_fr1	unknown function	3,0	0,00
158299_fr2	unknown function	2,8	0,03
144152_fr3	IgGFc-binding protein-like, partial	2,7	0,03
49680_fr4	GH7 family protein	2,7	0,03
213_fr1	echinoderm microtubule-associated protein-like 2	2,6	0,03
196745_fr1	slow-tonic S2 myosin heavy chain	2,6	0,00
58650_fr5	unknown function	2,5	0,01
200301_fr2	myosin heavy chain	2,2	0,00
206956_fr2	heat shock protein 70	-2,7	0,00
178152_fr5	unknown function	-3,3	0,03
17932_fr2	unknown function	-3,3	0,00
119204_fr3	Adenylate kinase 2, mitochondrial	-3,3	0,00
3452_fr6	ovarian serine protease	-3,7	0,01
173956_fr6	No detected homolog	-3,7	0,01
147876_fr3	60S ribosomal protein L13-like	-4,0	0,00
129900_fr3	Tubulin beta-2 chain	-4,0	0,00
5204_fr4	heat shock protein 21	-4,3	0,00
98633_fr1	No detected homolog	-4,3	0,03

200739_fr2	No detected homolog	-4,3	0,03
10750_fr4	No detected homolog	-4,7	0,00
158168_fr2	vacuolar ATP synthase subunit B K form	-4,7	0,02
203707_fr1	No detected homolog	-4,7	0,01
210517_fr6	No detected homolog	-5,0	0,00
806_fr5	No detected homolog	-5,0	0,03
2523_fr5	unknown function	-5,0	0,03
121767_fr6	unknown function	-5,7	0,03
121351_fr5	upstream activation factor subunit spp27-like	-5,7	0,00
17585_fr4	unknown function	-6,7	0,00
23428_fr4	heat shock cognate 70	-6,8	0,00
24872_fr5	unknown function	-7,7	0,02
190195_fr1	No detected homolog	-8,0	0,01
125804_fr4	No detected homolog	-8,0	0,00
193410_fr3	No detected homolog	-8,3	0,01
12473_fr4	No detected homolog	-8,6	0,02
774_fr1	No detected homolog	-8,7	0,04
12885_fr4	vitellogenin	-9,0	0,02
32875_fr1	clottable protein 2	-9,3	0,00
559_fr2	delta-1-pyrroline-5-carboxylate dehydrogenase, mitocł	-9,7	0,04
207095_fr1	ABC transporter-like protein	-10,3	0,00
205688_fr2	No detected homolog	-10,7	0,00
164345_fr6	No detected homolog	-10,7	0,02
29145_fr6	No detected homolog	-10,7	0,00
210317_fr1	No detected homolog	-11,0	0,00
25141_fr3	Hemolymph clottable protein	-11,0	0,00
49094_fr6	No detected homolog	-11,3	0,00
207064_fr4	No detected homolog	-11,7	0,01
47_fr4	Vitellogenin	-12,2	0,01
201865_fr3	vitellogenin	-12,3	0,03
39262_fr4	No detected homolog	-13,3	0,01
201362_fr4	No detected homolog	-14,0	0,02
681_fr6	No detected homolog	-14,3	0,00
19128_fr1	No detected homolog	-14,7	0,00
122081_fr5	cytochrome P450 CYP12A2	-15,0	0,00
23124_fr4	No detected homolog	-15,7	0,00
200559_fr3	No detected homolog	-16,7	0,03
30490_fr1	vitellogenin-like protein precursor	-16,7	0,00
196124_fr3	hemolymph clottable protein	-17,0	0,00
30488_fr4	No detected homolog	-17,7	0,00
205041_fr4	No detected homolog	-17,7	0,00
9788_fr4	vitellogenin 2	-19,7	0,01
20034_fr2	clottable protein 2	-20,0	0,01
3624_fr5	No detected homolog	-22,3	0,03
39606_fr3	clotting protein precursor	-23,6	0,00
178132_fr5	venom protease	-27,7	0,00
206469_fr3	clotting protein precursor	-28,0	0,00
30649_fr2	vitellogenin-like protein precursor	-28,4	0,00
12849_fr2	vitellogenin	-30,0	0,00
199971_fr3	clotting protein precursor	-32,8	0,00
201667_fr3	vitellogenin	-37,3	0,00
204978_fr4	No detected homolog	-43,0	0,04
17046_fr6	clottable protein 2	-48,0	0,00
64_fr5	vitellogenin	-55,8	0,01
87598_fr2	vitellogenin-like protein precursor	-56,7	0,00
24570_fr5	vitellogenin-like protein precursor	-57,3	0,01
31746_fr5	No detected homolog	-61,7	0,03
37795_fr5	vitellogenin-like protein precursor	-62,3	0,00
200426_fr2	vitellogenin	-81,0	0,00
206576_fr4	No detected homolog	-91,0	0,00
276_fr4	hemolymph clottable protein	-91,3	0,00
194758_fr3	apolipocrustacein	-91,3	0,01
19261_fr6	No detected homolog	-174,8	0,04
277_fr1	vitellogenin	-192,3	0,01
25127_fr4	clottable protein	-197,3	0,00

Annexe IV : Protéines modulées dans les gonades mâles en réponse à la pression toxique

Probable function	Contig_ID	Cadmium	Pyriproxyfen	Methoxyfenozide
calcium-transporting ATPase sarcoplasmic/endoplasmi	4227_fr2	A	F	B
arginine kinase	14404_fr1	A	B	B
mitochondrial ATP synthase subunit alpha precursor	16972_fr2	A	B	A
hemocyanin	30247_fr5	A	A	F
smooth endoplasmic reticulum calcium ATPase	30375_fr6	A	A	A
troponin T	30921_fr3	A	B	B
myosin heavy chain	34079_fr5	A	B	A
histone H4-like	100349_fr3	A	A	A
hemocyanin	166827_fr1	A	A	C
enolase	192695_fr3	A	B	A
myosin heavy chain type 1	195102_fr4	A	A	C
slow muscle myosin S1 heavy chain	14_fr3	B	C	E
fructose-bisphosphate aldolase-like isoform X2	2278_fr6	B	A	A
sarcoplasmic calcium-binding protein variant a	2853_fr3	B	B	C
No predicted function	9438_fr2	B	F	D
Histone H2A	13250_fr6	B	B	B
myosin heavy chain	22255_fr2	B	A	B
myosin heavy chain isoform 2	37276_fr6	B	A	A
hemocyanin subunit 1	48760_fr1	B	A	F
myosin heavy chain, muscle isoform 1	105020_fr5	B	A	E
myosin heavy chain	422_fr4	C	A	B
No predicted function	4635_fr5	C	C	A
No predicted function	7061_fr2	C	A	C
No detected homolog	41523_fr3	C	A	A
myosin heavy chain	182086_fr1	C	C	B
fast myosin heavy chain	194796_fr2	C	B	B
myosin heavy chain type 4, partial	196745_fr1	C	C	B
myosin heavy chain	200301_fr2	C	C	B
Tubulin beta-2 chain	10171_fr5	D	C	A
Copine-8	34845_fr5	D	B	C
actin, cytoplasmic 2 isoform X1	47294_fr1	D	A	A
hemocyanin subunit 1	71063_fr4	D	D	A
DEAD(Asp-Glu-Ala-Asp) box polypeptide 39	121218_fr5	D	E	D
No predicted function	128511_fr6	D	A	A
glyceraldehyde-3-phosphate dehydrogenase	141775_fr4	D	A	C
ATP synthase beta subunit	155991_fr2	D	B	B
No predicted function	5492_fr3	E	A	B
paramyosin, partial	30611_fr4	E	A	E
hemocyanin subunit 2	36242_fr3	E	E	B
histone H3.2-like	65834_fr4	E	B	D
No detected homolog	196858_fr4	E	C	E
calmin-like protein	4565_fr1	F	C	C
No predicted function	40293_fr4	F	B	D
Filamin-C, putative	144171_fr6	F	B	B
hemocyanin subunit 3 precursor	186201_fr3	A		D
phosphopyruvate hydratase	195483_fr1	A		E
vitellogenin 1	198421_fr3	A	D	
No predicted function	135487_fr5	B	E	
No predicted function	69913_fr6	D	A	
alpha-actinin, sarcomeric-like	31902_fr5	F		B
hemocytin-like	461_fr1		C	D
glycogen phosphorylase-like	40028_fr4		C	A
triose-phosphate isomerase	130139_fr1		F	A
hemocyanin subunit 4 precursor	196843_fr3		D	F
hemocyanin subunit 1	202637_fr1		B	C
Na+/K+_ATPase	209438_fr3		C	C
spectrin beta chain-like	470_fr4	C		
projectin	2029_fr2	F		
glucan pattern-recognition lipoprotein	626_fr4			F
No predicted function	3678_fr1		D	
pyruvate kinase 3	4737_fr3		B	
alpha tubulin	9002_fr1		C	
Heat shock protein 90	12402_fr1			D
No predicted function	17884_fr3		E	
No predicted function	21424_fr5			C
heat shock cognate 70	23428_fr4			D
heat shock 70 kDa protein cognate 3-like	24823_fr3		B	
actin-2, partial	48693_fr3		F	
No detected homolog	110918_fr2		D	
hemocyanin subunit 1	202279_fr1			C

Résumé

Parmi les outils existants pour l'évaluation de la qualité des milieux aquatiques, les biomarqueurs permettent de détecter précocement l'exposition et/ou les effets d'une contamination. Toutefois, en raison du manque de connaissance fondamentale de leur régulation au niveau moléculaire, peu de biomarqueurs ont été spécifiquement développés chez les invertébrés, alors que ces derniers représentent 95% de la biodiversité animale. Grâce aux récentes avancées technologiques dans le domaine du séquençage de l'information génétique et protéique, la découverte de nouvelles protéines chez ces organismes est à présent possible.

Centré sur l'utilisation d'une espèce d'intérêt en écotoxicologique, l'amphipode d'eau douce *Gammarus fossarum*, et sur la problématique de la perturbation endocrine, ce travail doctoral a pour objectif de découvrir de nouvelles protéines impliquées dans la reproduction du gammare et de proposer des biomarqueurs spécifiques de reprotoxicité. Par une première approche protéogénomique, un important catalogue de protéines a été généré. Puis, afin d'identifier de nouvelles protéines impliquées dans la reproduction, une comparaison entre les protéomes des tissus reproducteurs mâle et femelle a été réalisée, suivi de l'étude de la dynamique du protéome au cours de différents processus physiologiques : ovogénèse, embryogénèse et spermatogénèse. Enfin, une dernière expérience sur des organismes mâles exposés à différents perturbateurs endocriniens a permis d'identifier différents candidats biomarqueurs de reprotoxicité. Cette étude ouvre la voie à des développements rapides de biomarqueurs spécifiques d'une espèce animale d'intérêt en écotoxicologie.

Mots-clés : biomarqueurs, protéomique, reproduction, perturbateur endocrinien, *Gammarus fossarum*

Abstract

Among the tools for assessing biologic quality of aquatic ecosystems, biomarkers are ideally suited for the early detection of contaminant exposure and/or effects. Yet, because of the lack of fundamental knowledge of their molecular regulation, few specific developments have been carried out on invertebrates even though they account for more than 95% of animal biodiversity. Thanks to recent technological advances in nucleic and proteic information sequencing, discovery of new proteins is now possible for these organisms.

Focussed on the use of an ecotoxicologically relevant species, the freshwater amphipod *Gammarus fossarum* and the problem of endocrine disruption, this doctoral thesis aims to discover new proteins involved in gammarid reproductive function and to propose specific biomarkers of endocrine disruption. With a first proteogenomic approach, an important protein catalogue was generated. Next, to identify proteins involved in gammarid reproduction, male and female reproductive tissue proteomes were compared, followed by the study of proteome dynamics in several physiological processes: oogenesis, embryogenesis and spermatogenesis. Finally, a last experiment on male gammarids challenged with different endocrine disrupter chemicals identified several candidate biomarkers of reprotoxicity. This study paves the way for quick developments of specific biomarkers for organisms of interest in ecotoxicology.

Keywords: biomarker, proteomic, reproduction, endocrine disruption, *Gammarus fossarum*