



HAL
open science

Protéomique ciblée pour la quantification multiplexée de biomarqueurs : perspectives pour la surveillance environnementale

Anabelle Espeyte, Davide Degli Esposti, Maxime Leprêtre, Hugo Detante, Kévin Sugier, Olivier Geffard, Arnaud Chaumot

► To cite this version:

Anabelle Espeyte, Davide Degli Esposti, Maxime Leprêtre, Hugo Detante, Kévin Sugier, et al.. Protéomique ciblée pour la quantification multiplexée de biomarqueurs : perspectives pour la surveillance environnementale. INRAE; OFB. 2023. hal-04246106

HAL Id: hal-04246106

<https://ofb.hal.science/hal-04246106v1>

Submitted on 17 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Programme 2019/2022 – Action n° 4
Thème *Risques liés à la contamination chimique des milieux aquatiques* -



Protéomique ciblée pour la quantification multiplexée de biomarqueurs: perspectives pour la surveillance environnementale

**Application d'une stratégie généralisable entre
espèces pour l'identification de protéines
d'intérêt en biosurveillance**

Rapport final

**Anabelle Espeyte, Davide Degli Esposti, Maxime
Leprêtre, Hugo Detante, Kévin Sugier, Olivier Geffard,
Arnaud Chaumot**

Février 2023

- **CONTRIBUTEURS**

Laboratoire d'écotoxicologie, UR RiverLy, INRAE Lyon-Villeurbanne

Anabelle Espeyte, ingénieur d'étude (INRAE)
Davide Degli Esposti, chargé de recherche (INRAE)
Maxime Leprêtre, post-doctorant (INRAE)
Kévin Sugier, post-doctorant (INRAE)
Hugo Detante, stagiaire (INRAE-ESTBB)
Olivier Geffard, directeur de recherche (INRAE)
Arnaud Chaumot, directeur de recherche (INRAE)

Laboratoires partenaires impliqués

UMR SEBIO Reims : M Palos Ladeiro, A Geffard, R Peden	modèle dreissène
UMR SEBIO Ineris : A Bado-Nilles, JM Porcher	modèle épinoche
UMR SEBIO Le Havre : B Xuereb	modèles <i>Palaemon</i>
UMR LIEC Université de Lorraine : S Devin, S Pain-Devin	modèle dreissène, quagga
UMR ISA Université Lyon1 : A Salvador, S Ayciriex, J Faugère	protéomique ciblée
CEA Li2D : J Armengaud	protéogénomique
UMR LBBE Lyon 1 : C Lopes	valeurs de référence

- **CORRESPONDANTS**

OFB
Olivier Perceval,
Chef de service
Direction de la recherche et de l'appui scientifique
olivier.perceval@ofb.gouv.fr

INRAE :
Arnaud Chaumot,
Directeur de recherche (INRAE)
UR RiverLy
arnaud.chaumot@inrae.fr

- **RESUME**

Bien qu'une attente grandissante existe vis-à-vis de l'application d'outils biologiques pour l'évaluation de la qualité des milieux, l'utilisation des **outils moléculaires (biomarqueurs)** reste aujourd'hui limitée (particulièrement chez les invertébrés) du fait de l'existence de différents verrous techniques et scientifiques. Comme nous avons pu le démontrer par les travaux INRAE-OFB chez l'espèce sentinelle *Gammarus fossarum*, la mise en place d'une **biosurveillance active** basée sur l'encagement d'organismes calibrés provenant d'une unique population source, permet de contrôler la variabilité biologique des organismes utilisés, facilitant ainsi l'interprétation des marqueurs suivis. Nous avons par ailleurs pu bénéficier chez cette espèce des dernières innovations en termes de biologie moléculaire et notamment pu développer une approche dite de protéogénomique en collaboration avec le CEA (Marcoule - DMTS). Ceci a conduit à la construction de listes de plusieurs centaines de protéines séquencées chez ce crustacé, protéines spécifiques et associées à diverses fonctions physiologiques. Dans le même temps, en collaboration avec l'ISA (CNRS Lyon1), nous avons fait la preuve de concept qu'à l'instar des approches multi-résidus développées pour le dosage des contaminants, les méthodes de spectrométrie de masse disponibles en **protéomique ciblée** permettent aujourd'hui de quantifier sur un même échantillon et en une unique analyse plusieurs dizaines de ces **protéines** via le dosage de peptides spécifiques **biomarqueurs**. L'étude présentée ici a eu pour objectif général de prolonger cette dynamique vers la mise en œuvre **opérationnelle** de ces méthodologies.

Pour cela, deux axes de travail ont été poursuivis. Le premier axe a développé une capacité d'**acquisition haut-débit** et automatisée de données en protéomique ciblée pour permettre notamment la définition de **valeurs de référence** et l'interprétation des variations de concentrations en peptides observées lors de campagnes de biosurveillance (encagements *in situ* sur une cinquantaine de stations du réseau de surveillance RCS). Le deuxième axe a ouvert la réflexion à d'autres **espèces** d'intérêt environnemental déjà proposées comme **sentinelles** de la surveillance des milieux aquatiques d'**eaux douces, de transition et marines**. Associant différents partenaires scientifiques nationaux (SEBIO Reims / Le Havre / Ineris ; LIEC Metz), l'étude a permis de définir et de faire la démonstration de l'application d'une **stratégie commune** de développement de biomarqueurs protéiques mesurés sur organes chez six espèces (gammare, crevette bouquet, crevette blanche, dreissène, moule quagga, épinoche). Cette stratégie passe par une première phase de définition des marqueurs protéiques qui s'appuie sur une démarche de protéogénomique (couplage de séquençage du transcriptome et d'acquisition de données de protéomique massive) permettant de documenter le catalogue de protéines dans les organes d'intérêt chez les différentes espèces. Les peptides rapporteurs de protéines d'intérêt choisies parmi ces catalogues (ici associées à la détoxification, l'osmorégulation ou l'immunité) sont identifiés *in fine* en spectrométrie de masse ciblée.

MOTS CLES : CONTAMINATION CHIMIQUE, TOXICITE, BIOMONITORING, PROTEOMIQUE, BIOMARQUEURS, BIODIVERSITE, ESPECES

Table des matières

1. Introduction	5
1.1. Contexte, état de l'art	5
1.2. Objectifs de l'action et organisation du rapport	7
2. Du développement d'un dosage multiplexé de biomarqueurs protéiques à la définition de valeurs de référence : démonstration chez le gammare	8
2.1. Jeu d'échantillons de l'étude	8
2.2. Développements analytiques pour un dosage multi-marqueur standardisé.....	9
2.3. Définition de valeurs de référence et seuils d'interprétation.....	11
2.4. Bilan	15
3. Exercice multi-espèces : définition d'une stratégie commune et généralisable pour l'identification de biomarqueurs protéiques chez les espèces sentinelles	16
3.1. Choix des cas d'étude	16
3.2. Mobilisation et acquisition de ressources moléculaires chez les 6 espèces d'étude	19
3.2.1. Transcriptomes de référence (RNAseq).....	19
3.2.2. Catalogues protéiques sur organes (protéomique shotgun)	20
3.3. Identification de peptides biomarqueurs par protéomique ciblée (MRM)	23
3.3.1. Procédures analytiques	23
3.3.2. Résultats sur les 3 organes chez <i>G. fossarum</i> , <i>D. polymorpha</i> et <i>P. serratus</i>	26
3.3.3. S'appuyer sur des catalogues protéomiques de la littérature : exemple de l'épinoche	29
3.3.4. S'appuyer sur des catalogues protéomiques acquis sur une autre espèce du même genre : exemple des <i>Palaemon</i> et des dreissènes.....	29
3.4. Bilan	30
4. Conclusion	32
5. Références	33
6. Annexes	36

1. Introduction

1.1. Contexte, état de l'art

Cette action concerne une question nationale relative à la surveillance de la qualité chimique et toxique des milieux aquatiques. L'objectif de l'action qui s'inscrit dans le développement des outils de biosurveillance soutenu depuis une dizaine d'années par le partenariat initialement AFB-Irstea aujourd'hui OFB-INRAE, a été de lever différents verrous scientifiques devant permettre l'utilisation de biomarqueurs moléculaires à large échelle. Cette action a été mise en œuvre en synergie avec les dynamiques actuellement mises en place d'une part autour de l'utilisation du gammare encagé dans le cadre de la surveillance chimique des milieux (par les agences de l'eau, suivi DCE) (Geffard et al 2021), et d'autre part sur la question de l'intégration de la diversité des espèces pour une meilleure surveillance (e.g., projet OFB-GIP Seine Aval Sashimi-Biosurveillance).

L'application de la DCE pour la surveillance de la contamination chimique des eaux de surface consiste en premier lieu à déterminer si les niveaux de contamination des masses d'eau sont conformes vis à vis de normes de qualité environnementale (NQE). Les NQE sont des concentrations de polluants prioritaires dans l'eau ou dans le biote qui ne doivent pas être dépassées, afin de protéger la santé humaine et l'environnement. Pour la conformité aux NQE-biote, les approches passives (échantillonnage d'organismes résidents dans les écosystèmes) ont été les premières à émerger pour les environnements côtiers avec, par exemple, le "Mussel Watch" initié en 1976 (Borja et al 2008). Ces approches passives sont moins développées pour les masses d'eau continentales qui sont plus complexes au regard de la grande diversité de ces hydrosystèmes (imposant l'étude d'un très grand nombre de sites d'échantillonnage) et de la diversité des cortèges d'espèces aquatiques associés à ces différents habitats. La biosurveillance active, basée sur l'utilisation d'organismes transplantés, a récemment été proposée comme approche alternative. Elle a l'avantage d'utiliser une seule espèce sur l'ensemble du territoire, de minimiser la variabilité biologique en utilisant des organismes calibrés (taille, sexe...) provenant de la même population, et enfin de maîtriser le temps d'exposition (Bervoets et al 2005). Dans ce contexte, le laboratoire d'écotoxicologie de Lyon (INRAE) a développé un outil de diagnostic de la contamination chimique des milieux, basé sur l'encagement de l'amphipode *Gammarus fossarum* (Geffard et al 2014, 2021), qui est actuellement appliqué comme un outil de surveillance pour la conformité aux NQE biote¹.

Parallèlement à cette approche « chimique » qui procède substance par substance, l'évaluation de la qualité chimique des milieux aquatiques peut également bénéficier de l'utilisation d'outils biologiques permettant d'intégrer l'effet toxique de l'ensemble des contaminants auxquels sont co-exposés les organismes dans les écosystèmes contaminés. Cette approche peut considérer soit des réponses biologiques génériques face au stress chimique, soit des réponses plus spécifiques de certains modes d'action toxiques. Aujourd'hui, de tels outils écotoxicologiques (bioessais et biomarqueurs) ne sont pas inclus dans la DCE, mais leur intérêt et leur utilisation sont interrogés et encouragés, aussi bien au niveau européen au travers des projets comme SOLUTIONS et DEMEAU, qu'au niveau national au travers du plan Micropolluants (2016-2021), de l'appel à manifestation d'intérêt (AMI 2017) porté par l'Agence Française de la Biodiversité, ou de la mise en place de groupes de travail comme celui porté par INERIS sur le rôle et l'intérêt que doivent avoir les bioessais dans la surveillance des milieux aquatiques. Chez *G. fossarum*, différents marqueurs d'effets individuels (activité alimentaire, fécondité, mue) sont ainsi appliqués aujourd'hui dans le cadre de l'évaluation de la qualité des milieux aquatiques au niveau de stations du réseau de surveillance de différentes agences de l'eau françaises (Chaumot et al 2021), suite à différentes études de validation de ces outils de biosurveillance active menées à l'échelle nationale lors d'actions de recherche INRAE-OFB (Geffard et al 2019).

Concernant les biomarqueurs moléculaires, différentes limites techniques contraignent leur utilisation en routine en biosurveillance, notamment le manque de méthodes de quantification directe permettant d'assurer la répétabilité/reproductibilité dans le temps et la comparaison des résultats entre études avec des unités de mesures qui ne soient pas arbitraires ou dépendantes de chaque protocole ou laboratoire. L'approche multi-biomarqueurs est par ailleurs essentielle pour intégrer une large gamme de typologie de réponses biologiques / physiologiques induites par les différentes classes de contaminants. Certains indices comme l'IBR (Beliaeff et Burgeot, 2002) ont été proposés pour interpréter de manière intégrée les modulations de plusieurs biomarqueurs. Cependant, malgré cette aide à l'interprétation, il reste encore nécessaire de devoir mettre en place une méthode spécifique

¹ EC. Technical Report 2014 – 083. Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance Document N°32 on biota monitoring (the implementation of EQS-Biota) under the Water Framework Directive.

pour chacun des biomarqueurs d'intérêt considéré. Comme discuté par Trapp et al (2014a), la plupart des biomarqueurs moléculaires disponibles aujourd'hui chez les invertébrés reposent sur des méthodes spécifiques à chaque biomarqueur, multipliant le travail de laboratoire, le coût et le temps nécessaire à l'analyse d'un grand nombre d'échantillons, comme cela est nécessaire dans les programmes de surveillance. S'ajoute à cette lourdeur du développement de biomarqueurs « à façon » et de l'application de leur mesure en routine, la nécessité de disposer de marqueurs sur un ensemble d'espèces représentatives de la diversité des communautés biologiques aquatiques. En effet, un des enjeux en écotoxicologie est de pouvoir intégrer la diversité des sensibilités des espèces pour mieux évaluer l'impact des polluants dans les milieux aquatiques (Catteau et al 2022).

Au cours de la dernière décennie, les énormes progrès technologiques réalisés en chimie analytique ont fait émerger des méthodes par spectrométrie de masse hautement performantes en biochimie, notamment le dosage multiplexé de biomarqueurs protéiques dans le domaine du diagnostic médical. Dans le domaine du diagnostic environnemental, l'absence de données génomiques et protéomiques chez les espèces aquatiques sentinelles les plus couramment utilisées en écotoxicologie (notamment invertébrés) limite fortement le développement de biomarqueurs spécifiques et notamment l'application de telles méthodes de dosage de marqueurs protéiques. Pour lever ce verrou, une approche analytique spécifique, la protéogénomique, a été proposée par notre consortium de partenaires comme une approche de rupture en écotoxicologie (Armengaud et al 2014). Elle apparaît comme une alternative au difficile séquençage du génome chez chaque espèce d'intérêt environnemental. Elle consiste à séquencer les ARN messagers (partie codante du génome) en parallèle de l'acquisition de données de protéomique massive (approche shotgun) pour obtenir des informations sur les séquences protéiques spécifiques à l'espèce d'intérêt (Figure 1).

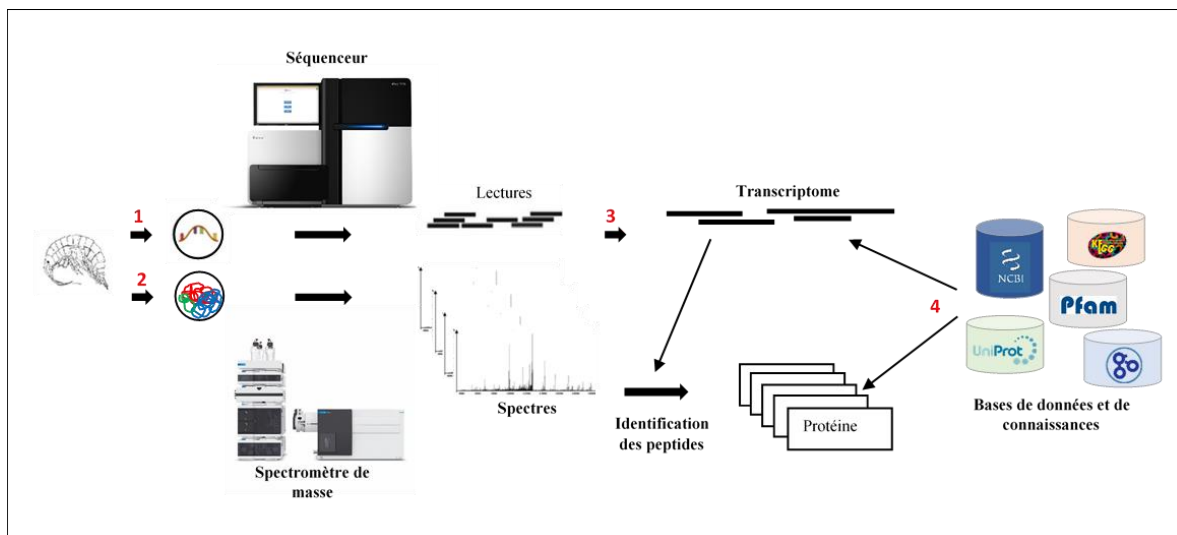


Figure 1 : Illustration de l'approche protéogénomique permettant d'établir un catalogue de protéines présentes chez n'importe quelle espèce sentinelle d'intérêt. Echantillonnage (individuel) pour 1. extraire l'ARN total et séquençage après construction des banques Illumina; et 2. disséquer un organe, et extraire les protéines pour analyse en spectrométrie de masse (shotgun). 3. Les lectures obtenues par le séquenceur sont nettoyées et combinées entre elles par un assembleur (ex : Trinity, Grabherr et al 2011) pour construire le transcriptome de référence de l'espèce (RNA-Seq). Après assemblage, une prédiction des séquences codantes est effectuée en utilisant TransDecoder, et permet ainsi d'obtenir une prédiction du protéome (séquences de protéines prédites). Ces prédictions servent ensuite de référence pour interpréter les spectres protéomiques et les assigner à des séquences protéiques compatibles avec les masses enregistrées (séquences de protéines observées) 4. Dans le but d'identifier les fonctions des protéines prédites ou identifiées, une annotation fonctionnelle peut être réalisée. Pour cela, les séquences protéiques prédites sont alignées contre diverses banques de données composées de séquences connues : par homologie, les caractéristiques (nom, fonction...) des séquences alignées de la base de données sont transférées aux séquences nouvellement construites.

Aujourd'hui, la technologie RNA-Seq permet en effet un séquençage *de novo* profond du transcriptome chez n'importe quelle espèce. Le transcriptome peut ainsi être facilement établi et utilisé pour interpréter les spectres obtenus par protéomique shotgun pour l'identification des protéines. Ceci permet alors d'établir un catalogue de séquences protéiques identifiées par spectrométrie de masse et spécifiques de l'espèce étudiée. Ces dernières années, nous avons ainsi développé en collaboration avec le CEA Marcoule une telle approche chez l'espèce *Gammarus fossarum*. Celle-ci

a permis d'établir le protéome d'organes cibles chez *G. fossarum*, et un total de 1800 protéines a été initialement certifié chez cette espèce (Trapp et al 2014b). Aujourd'hui ce catalogue a été étendu à environ 3000 séquences protéiques expérimentales. En parallèle, nous avons couplé ces approches de découverte aux approches de protéomiques ciblées qui offrent de nouvelles opportunités pour le dosage haut débit multiplexé de protéines par spectrométrie de masse (collaboration ISA CNRS Lyon1) (Gouveia et al 2019). A partir des catalogues protéomiques définis chez *G. fossarum*, une première étude de démonstration a permis de développer une méthode quantitative de dosage multiplexé d'une cinquantaine de peptides biomarqueurs et de l'appliquer sur le terrain (à une échelle régionale) en recourant à l'approche de biomonitoring actif développée chez notre espèce sentinelle (Gouveia et al 2017).

1.2. Objectifs de l'action et organisation du rapport

Le travail proposé dans cette action engagée entre 2019 et 2022 s'est articulé autour de deux axes.

1 – Suite au développement méthodologique qui a abouti à l'identification et la quantification simultanée de plusieurs dizaines de protéines biomarqueurs chez le gammare (Gouveia et al 2017), le premier axe de travail de cette action a été de développer une capacité d'acquisition massive de données en protéomique ciblée pour permettre notamment la définition de valeurs de référence chez le gammare. Ceci est passé par (i) le développement d'une nouvelle méthode multiplexe qui permet de quantifier simultanément la concentration de plusieurs dizaines de peptides rapporteurs de grandes fonctions biologiques chez ce crustacé, (ii) l'automatisation de la préparation des échantillons, et le test de la robustesse du protocole de préparation et d'analyse via l'acquisition des niveaux de ces biomarqueurs sur 325 échantillons collectés à l'échelle nationale (engagements sur le réseau de suivi). Ces développements ont fait l'objet d'un rapport d'étape en 2019 qui détaille les tests de différents protocoles/procédures analytiques qui ont abouti à la validation de la méthode retenue (Espeyte et al 2019). Nous faisons ici la synthèse de ces résultats de développements analytiques, et rapportons également des travaux d'analyse de données qui ont suivi pour aller jusqu'à la définition de valeurs de référence pour ce nouveau type de biomarqueurs protéiques.

2- Le deuxième axe a visé à ouvrir la réflexion « protéomique pour la surveillance » à d'autres espèces d'intérêt environnemental proposées comme sentinelles des milieux aquatiques d'eaux douces, de transition et marines en France. Associant différents partenaires scientifiques (SEBIO Reims, INERIS, LIEC, SEBIO Le Havre), l'objectif a été de définir et faire la démonstration de l'application d'une même stratégie de développement de biomarqueurs protéiques chez 6 espèces : gammare, crevette bouquet, crevette blanche, dreissène, moule quagga, épinoche. Pour accompagner cet exercice de comparaison et de transférabilité entre espèces de la démarche établie chez le gammare qui couple protéomique de découverte (shotgun) et protéomique ciblée, le cadre de réflexion s'est concentré sur l'identification de biomarqueurs par spectrométrie de masse sur organes, en visant 3 types d'organes assurant des fonctions analogues chez les 6 espèces aquatiques retenues (branchies pour l'osmorégulation ; sang-hémolymphe pour l'immunité; foie-hépatopancréas-glande digestive pour la détoxification).

2. Du développement d'un dosage multiplexé de biomarqueurs protéiques à la définition de valeurs de référence : démonstration chez le gammare

Nous synthétisons dans cette partie les avancées accomplies dans le cadre de cette action INRAE-OFB pour asseoir l'opérationnalité du dosage de biomarqueurs protéiques chez le gammare (sur organisme entier) par protéomique ciblée, en ayant comme point de départ la preuve de concept réalisée dans le cadre de la thèse de Duarte Gouveia (Gouveia et al 2017). Les travaux ont consisté tout d'abord à adapter la méthode MSMS sur un nouvel équipement HPLC-MSMS (en mode dynamic MRM, « dMRM »), à optimiser et automatiser la préparation d'échantillons (utilisation d'un robot) afin d'augmenter la capacité d'acquisition de données et de garantir l'absence d'effet entre dates et séries d'analyses (comparabilité des dosages dans le temps et entre études). Le descriptif technique précis de ces travaux est détaillé dans un premier rapport (Espeyte et al 2019), disponible en libre accès. Nous avons mis en œuvre ensuite les mêmes méthodes statistiques déjà appliquées chez le gammare dans le cadre de données de bioaccumulation (Geffard et al 2014) et de toxicité (Geffard et al 2019), pour démontrer ici sur un jeu d'échantillons issus d'encagements sur les stations du réseau de surveillance nationale, la possibilité de définir des valeurs de référence et seuils d'interprétation de ces biomarqueurs d'une nouvelle nature (concentrations peptidiques).

2.1. Jeu d'échantillons de l'étude

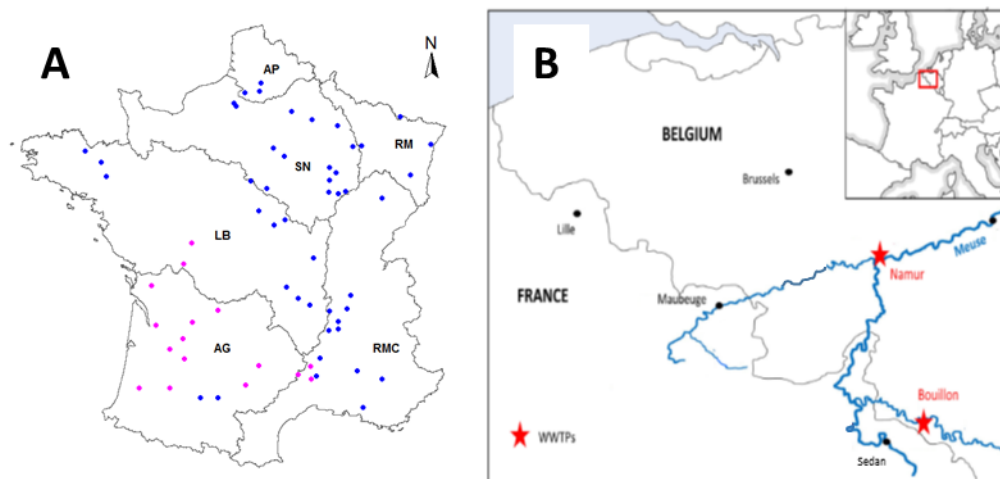


Figure 2 : Localisation des stations de l'étude (encagement de gammarès). A- campagnes AFB48 (Geffard et al 2019). Les points bleus et roses correspondent respectivement aux campagnes 1 et 2. Les délimitations géographiques sur la carte représentent les limites de six agences de bassin. B- localisation des 2 sites d'étude retenus du projet DIADEM (3 stations sur Namur + 2 sur Bouillon réparties sur amont / aval de station d'épuration) (Catteau et al 2022).

Sites d'encagement sur les réseaux de surveillance - Au cours d'une précédente action Irstea-AFB (Geffard et al 2019 ; action 48), deux campagnes d'encagement *in situ* ont été mises en œuvre afin d'exposer pendant 7 jours des gammarès mâles dans les cours d'eau de 64 stations à l'échelle nationale française sur le réseau de surveillance nationale RCS (Figure 2A). L'exposition de gammarès a été réalisée selon le protocole décrit par Besse et al (2013). Il consiste à transplanter sur site des gammarès mâles calibrés en taille et provenant d'une population contrôlée située ici dans une ancienne cressonnière à Saint Maurice de Rémois (Ain 01). Après exposition, cinq gammarès ont été prélevés par site, pesés et congelés individuellement dans de l'azote liquide pour les dosages protéomiques par spectrométrie de masse. Ils ont été conservés dans des congélateurs à -80°C au laboratoire. Durant ces expositions, des événements ponctuels tels que l'ensablement des cages ou un étiage sévère ont eu lieu sur 9 sites et ont parfois conduit à un taux de mortalité supérieur à 50%. Par conséquent, les gammarès prélevés sur ces sites n'ont pas été considérés dans notre étude de modulation des biomarqueurs protéiques. Ainsi, au total cinq gammarès sur 55 sites soit 275 échantillons ont été analysés.

Campagne d'encagements amont /aval sur stations d'épuration (projet DIADEM) - Un autre jeu de déploiements sur site a été ajouté pour élargir le nombre de campagnes et les contextes d'exposition à la contamination. Il s'agit d'expositions réalisées dans le cadre du projet interreg franco-belge DIADEM (<https://www.interregdiadem.eu/>), s'intéressant au développement des approches de biosurveillance active pour caractériser l'impact des rejets urbains sur la qualité des cours d'eau du bassin de la Meuse. Pour ce projet, suivant le même protocole d'encagement, des gammars mâles ont été exposés pendant 7 jours dans les cours d'eau de 5 sites franco-belges, en amont et aval de stations d'épuration (Figure 2B) en octobre 2018 (10 stations au total). Toujours avec des gammars provenant de la cressonnière à Saint Maurice de Rémens et selon le protocole d'exposition décrit par Besse et al (2013), 10 individus par site ont été prélevés, pesés et congelés individuellement. Seules 5 stations ont été retenues pour l'étude, en ciblant des sites où différents effets physiologiques (inhibition alimentaire, perturbation de la mue) ont été enregistrés parallèlement au cours de ces mêmes expositions. Il s'agit des stations Bouillon Amont STEP, Bouillon Aval STEP, Namur Amont STEP et ville, Namur Amont STEP, Namur Aval STEP. Au total, 10 individus sur 5 stations soit 50 échantillons ont été analysés.

2.2. Développements analytiques pour un dosage multi-marqueur standardisé

L'ensemble de la démarche de développement et de validation de ce volet analytique est présenté dans le rapport d'étape Espeyte et al (2019). La méthode pour la préparation et l'analyse des échantillons par chromatographie liquide haute performance couplée à la spectrométrie de masse en tandem (HPLC-MS/MS) a été développée en collaboration avec l'ISA et est schématisée sur la Figure 3.

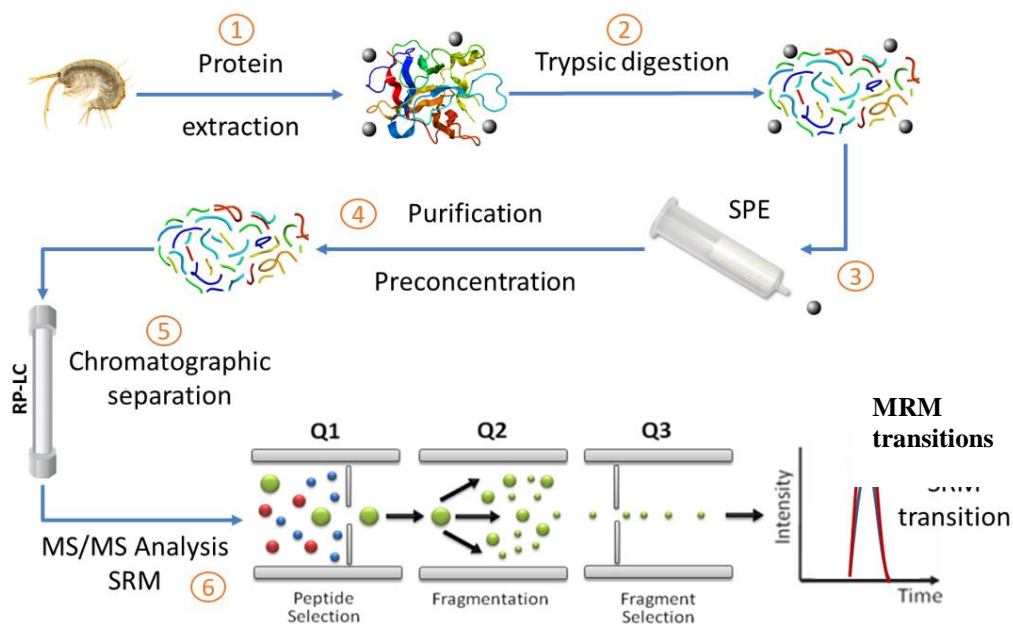


Figure 3 : Schéma de la démarche de recherche et d'identification de potentiels peptides biomarqueurs. Les protéines sont extraites et dosées (1) puis subissent une digestion trypsique (2) afin de récupérer les peptides, ces derniers sont purifiés par SPE (Solid Phase Extraction) (3) et concentrés par évaporation (4). Ils sont alors séparés sur le système chromatographique (5) puis envoyés pour analyse ciblée par MRM sur le spectromètre de masse en tandem (6).

Peptide	Protéine	Fonction	MW (Da)	Groupe fonctionnel	
HIEIFSPITK	64	Vitellogenin	1183.67	Female yolk proteins	
VVPSLSAEDTLSQR	276	Hemolymph clottable protein	1500.79		
HAEFVSNPPLDSTQAVK	276		1838.92		
IYPAEALTIVIEK	276		1571.92		
IFNVLQPIAESK	17046		Clottable protein 2		1357.77
TSEVFLPLTNELYQQTK	17046		2010.04		
ISPLINSPDLPK	39606	Clotting protein precursor	1379.77		
ELTSAAEVVSSLLK	200426	Vitellogenin	1445.81		
AAIETAFVNHK	206469	Clotting protein precursor	1312.72		
SFILSDLTNLQQSQAPYK	194758	Apolipoprotein	2052.06		
FVGLISLIDPPR	32234	Na+K+ ATPase α 1 subunit	1325.78	Osmoregulation	
VIMVTGDHPITAK	110907	Na+K+ ATPase	1380.75		
LGAIVAVTGDGVNDSPALK	110907		1795.98		
LQTNPDGLSTAEAR	209438	Na+K+ ATPase	1572.78		
EVNGDASEAALLK	208857	Na+ K+ ATPase	1315.67		
SGQDGVPIK	208857		1012.56		
LIFDNK	17127	sodium-potassium-activated adenosine triphosphatase alpha subunit D isoform	861.50		
LIIVEGQQR	17127		1086.59		
GIDIIGDAFEADR	2562		Prophenoloxidase		1390.68
APILEGYFSK	2562				1123.60
GIDFGTTQSVR	2562	1179.60			
ATQPSYTVAQLELPGVNITR	2562	2157.15			
ILEDVDFVFN	100255	Cytochrome P450 enzyme, CYP4C39	1365.70	Molt-related	
VPAILESPGR	15561	Prophenoloxidase	1184.66		
AFWGSPLR	144144	JHE-like carboxylesterase 1	1045.58		
ILTTMWADFAR	144144		1323.67		
EFWIADHNEVR	166723	Farnesoic acid methyltransferase	1515.72		
EAFDTVGR	9574	Endochitinase	893.43		
LVLGTATYGR	181833	Chitinase	1049.59		
GTLAVIPQNR	1917	Transglutaminase	1166.68		
GLCLPTDACPR	1975	Hemocytin-like	1258.59		
VLAVDILAK	7169	Transglutaminase	940.60		
FSESAILR	40079	Glutathion-S-transferase	1008.53		
LGSNFLQIPVNCYPYR	45375	Catalase	1776.91	General biomarkers	
ADPALGQAIQER	110912		1267.70		
LADNIAGHVINTQEFIR	110912	Catalase	1910.01		
LSAWLAACK	142711	Glutathion-S-transferase	1018.53		
ELDFADAHR	213317	Cellulase	1219.57		

Table 1 : Liste des 38 peptides suivis par le dosage multiplexé dMRM et quantifiés grâce à l'ajout de peptides lourds. Le code pour chaque protéine correspondante est relatif à l'identifiant des contigs de la base transcriptomique de référence GFOSS (Trapp et al 2014b).

Brièvement, une méthode analytique par protéomique ciblée (de type dMRM) a été développée, optimisée, validée pour le dosage multiplexé de concentrations de peptides rapporteurs de différentes fonctions biologiques chez le gammare. La quarantaine de peptides quantifiés (par ajout de peptides lourds) retenus dans la méthode finale sont présentés dans la Table 1. Dans le cadre de cette action, cette méthode a ainsi permis une acquisition à haut-débit de données biomarqueurs (mesures de 38 biomarqueurs peptidiques sur 325 échantillons individuels de gammare mâles en 2 mois), données correspondant à des concentrations peptidiques spécifiques de 26 protéines ainsi quantifiées de manière directe. L'ensemble des données de concentrations sont disponibles dans la publication en Open Access Leprêtre et al (2022a) ([doi.org/ 10.1186/ s12302-022-00692-2](https://doi.org/10.1186/s12302-022-00692-2) ; Table S3). Le recours à un automate de préparation a permis d'optimiser le traitement d'échantillons pour la digestion trypsique et la purification des peptides par SPE. La validation du protocole de préparation d'échantillons couplée à l'analyse par HPLC-MS/MS a été menée avec succès (comparaison protocole manuel *versus* automatisé). L'ensemble de la procédure analytique retenue et des équipements utilisés est décrit dans le rapport Espeyte et al (2019). Dans ce rapport, l'analyse globale de ces résultats nous ont notamment permis de conclure à l'absence d'effet campagne et d'effet séries d'analyses sur la durée de l'étude (Figure 4).

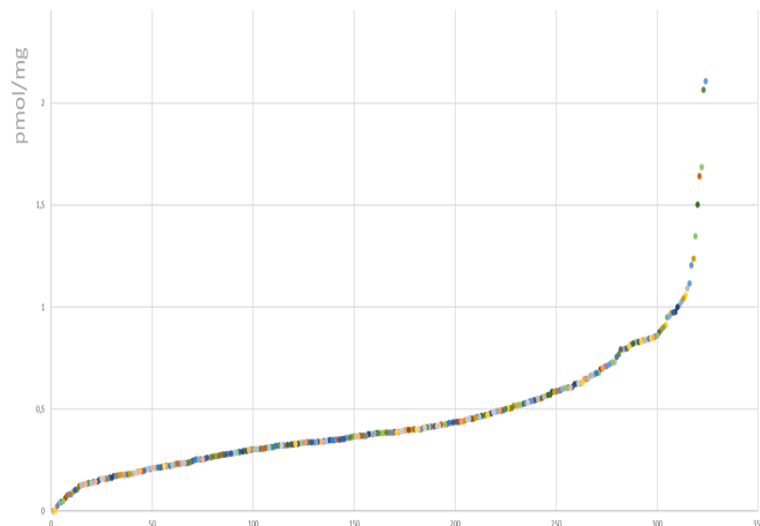


Figure 4 : Exemple de distribution des 325 échantillons individuels AFB et DIADEM classés par ordre croissant et colorés selon les 28 séries d'analyses (2 campagnes AFB48 + déploiement Diadem) avec pour exemple les concentrations du peptide ADPALGQAIQER (relatif à une activité catalase) ; valeurs de concentration en pmol de peptide par gramme sec de gammare. On peut remarquer la répartition aléatoire des différentes séries.

2.3. Définition de valeurs de référence et seuils d'interprétation

Du fait que les profils observés de distributions des concentrations peptidiques étaient très similaires aux distributions obtenues pour les données de bioaccumulation chez le gammare encagé (Geffard et al 2014, 2021) ou de marqueurs de toxicité (Geffard et al 2019, Chaumot et al 2021), nous avons appliqué les méthodes de définition de valeurs de référence mises en œuvre précédemment sur les données de biosurveillance gammare (Besse et al 2013 ; Chaumot et al 2015). L'ensemble de la démarche a fait en partie l'objet d'une publication dans la revue Environmental Sciences Europe (Leprêtre et al 2022a). On voit le plus souvent se dessiner un profil de distribution des concentrations peptidiques où se discriminent deux groupes de données : un premier groupe majoritaire distribué de façon gaussienne et un ensemble minoritaire de valeurs hautes qui se détachent de ce groupe gaussien (Figure 4).

Afin d'établir des distributions de référence, l'exercice de calibration à l'échelle nationale n'a considéré dans un premier temps que les données provenant de la cinquantaine de sites des campagnes sur le RCS (Figure 2A), ce qui garantit une représentativité des conditions de déploiement rencontrées dans nos hydrosystèmes. N'ont été pris en compte parmi les 38 peptides quantifiés à l'aide de standards, que les peptides présentant un taux de quantification de plus de 70% sur l'ensemble de ces échantillons (25 peptides au final). En utilisant des tests de corrélation, nous avons évalué si les variations de ces 25 biomarqueurs n'étaient pas expliquées par les facteurs de confusion température, conductivité, pH et niveaux d'O₂ dissous lors des encagements. Aucune corrélation n'a été décelée. Contrairement aux données de bioaccumulation, la procédure de Besse et al (2013) a dû être adaptée pour pouvoir intégrer à la fois la possibilité d'induction (valeurs anormalement hautes des peptides rapporteurs) tout comme la possibilité d'inhibition (valeurs anormalement basses) (Figure 5), les deux types de perturbations pouvant être attendus pour la plupart des fonctions suivies. L'ensemble des données de calibration de ces valeurs de référence sont produites en Annexe 1 pour les 25 peptides. Un résumé est fourni dans la Table 2.

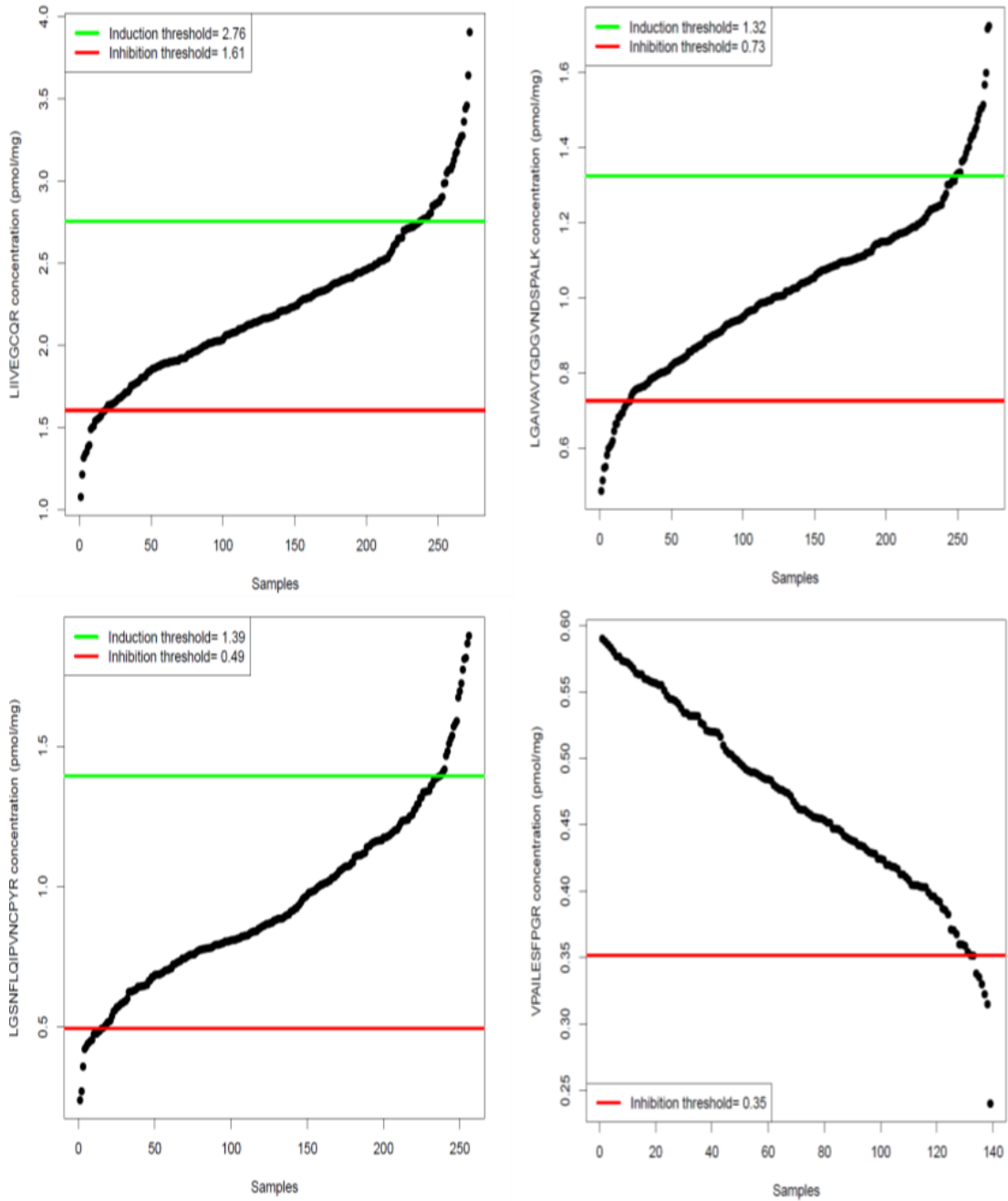


Figure 5 : Définition de valeurs seuils d'induction et/ou d'inhibition sur les données individuelles de concentrations peptidiques mesurées sur les sites du déploiement AFB-48. Exemple de 4 peptides. L'ensemble des profils pour les 25 peptides sont reportés en Annexe 1.

Functional group	Blast annotation	Peptide	Inhibition thresholds		Induction thresholds	
			Threshold values	Data below the threshold (%)	Threshold values	Data above the threshold (%)
Female yolk proteins	Vitellogenin	HIEIFSPITK	-	-	-	-
	Vitellogenin	ELTSAAEVVSSLLK	-	-	-	-
	Clotting protein precursor	AAIETAFAVNHLLK	-	-	-	-
Osmoregulation	Na ⁺ K ⁺ ATPase α1 subunit	FVGLSLIDPPR	1.96	8.45	3.58	7.35
	Na ⁺ K ⁺ ATPase	VIMVTGDHPITAK	3.23	9.55	6.51	11.02
	Na ⁺ K ⁺ ATPase	LGAIAVAVTGDGVNDSPALK	0.73	7.72	1.32	8.82
	Na ⁺ K ⁺ ATPase	LQTNPDITGLSTAEAR	0.13	8.56	0.36	12.06
	Na ⁺ K ⁺ ATPase	EVNGDASEAALIK	3.45	9.19	6.13	15.07
	Na ⁺ K ⁺ ATPase	SGQDGVPLIK	1.05	5.51	2.03	9.55
Molt-related	sodium-potassium-activated adenosine triphosphatase alpha subunit D isoform	LIFDNLK	1.09	4.41	2.05	11.02
	sodium-potassium-activated adenosine triphosphatase alpha subunit D isoform	LIIVEGCQR	1.61	6.61	2.76	12.5
	Prophenoloxidase	GIDIIGDAFEADR	0.59	5.1	2.1	9.69
		GIDFGTTQSVR	0.62	7.01	1.44	11.43
		ATQPSYTVAQLELPGVNITR	0.2	4.91	0.45	11.06
	Cytochrome P450 enzyme, CYP4C39	ILEDVDFVFNRR	-	-	-	-
		Prophenoloxidase	APILEGVFSK	0.32	4.77	0.86
	Prophenoloxidase	VPAILESFPGR	0.35	2.81	-	-
	JHE-like carboxylesterase 1	AFWGSILPLR	-	-	-	-
	JHE-like carboxylesterase 1	ILTTMWADFAR	-	-	-	-
	Farnesoic acid methyltransferase	EFWIATDHNEVR	2.68	8.45	4.84	14.33
	E-ochitinase	EAFDTVGR	1.53	7.35	3.41	10.29
Chitinase	LVLGTATYGR	0.89	4.04	3.86	9.92	
Transglutaminase	GTLAVIPVQNR	0.24	12.86	0.56	9.92	
Hemocytin-like	GLCLPTDACPR	-	-	-	-	
Transglutaminase	VLAVDILAK	0.07	3.69	0.51	15.86	
General biomarkers	Glutathion-S-transferase	FSESSAILR	1.09	5.51	2.92	7.72
	Catalase 1	LGSNFLQIPVNCPPYR	0.49	5.85	1.39	8.2
	Catalase 1	NLPADQAAALASSDPDYAIR	0.72	7.42	1.58	13.28
	Catalase 2	ADPALGQAIQER	0.13	5.18	0.64	21.48
	Catalase 2	LADNIAGHVINTQEFIR	1.14	4.29	8.03	16.4
	Glutathion-S-transferase	LSAWLAACK	0.48	7.11	1.11	10.48
	Cellulase	ELDFDADAHR	0.56	5.51	2.66	10.66

Table 2 : Etablissement de valeurs seuils d'inhibition et/ou d'induction (et taux de déclassement sur les échantillons AFB-48) pour 25 peptides biomarqueurs associés à différentes grandes fonctions biologiques (d'après Leprêtre et al 2022a)

Pour l'ensemble des peptides retenus, un profil conforme avec la présence d'un seuil d'induction et d'un seuil d'inhibition a été validé. La seule exception est un des peptides de la pro-phénoloxdase (pro-PO) pour lequel seul un seuil d'inhibition a été défini (peptide VPAILESFPGR sur la Figure 5). Pour l'ensemble des peptides, on constate sur l'ensemble de ces échantillons individuels des taux de déclassement de l'ordre de 10 à 15% des individus car présentant une induction du biomarqueur, auxquels s'ajoutent 5 à 10% environ pour des inhibitions (Table 2). Ceci souligne la sensibilité des biomarqueurs peptidiques et la possibilité de discriminer des situations de perturbation lors d'exposition sur les sites du réseau RCS, représentatifs des conditions rencontrées sur les cours d'eau nationaux.

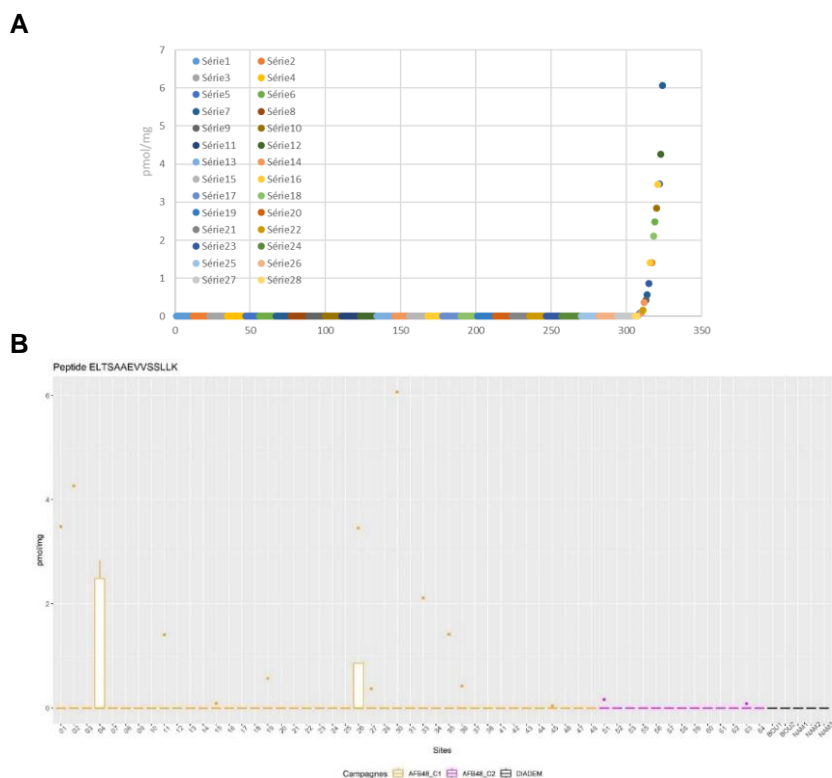


Figure 6 : A- Distribution individuelle des concentrations pour le peptide ELTSAAEVVSSLLK (relatif à une protéine vitellogénine-like) ; B- regroupement par sites d'encagement.

Concernant la douzaine de peptides qui ont été écartés de la procédure de calibration de valeurs de référence du fait de taux de quantification insuffisants (inférieur à 70% des échantillons), l'essentiel de ces peptides sont liés aux protéines de la vitellogénèse, souvent proposées comme biomarqueur de perturbation endocrinienne chez les mâles de différents taxons, ou alors des protéines en lien avec le contrôle hormonal de la mue (Table 2). Chez les mâles exposés *in situ*, ces peptides ont tout de même été détectés de manière ponctuelle chez certains individus. Par exemple, pour le peptide ELTSAAEVVSSLLK, lié à une protéine vitellogénine, 17 individus sur les 325 testés expriment ce peptide (Figure 6A). Le niveau de base étant en théorie très bas voire nul pour ce type de peptides chez le mâle, l'approche valeurs de référence utilisée précédemment ne s'applique pas (toute quantification doit être *a priori* interprétée comme anormale). Si l'on regarde maintenant pour ce même peptide la distribution regroupée par site d'exposition (Figure 6B), les individus « positifs » (*i.e.* où l'on quantifie ce peptide) ne se répartissent probablement pas aléatoirement et un ou deux sites ressortent comme induisant spécifiquement ces protéines. Il est dur d'aller plus loin avec 5 individus par site au regard de la variabilité inter-individu qui apparaît pour ces inductions. Aussi, en cohérence avec nos précédents travaux sur l'expression de la vitellogénine chez le crustacé comme biomarqueur de perturbation endocrinienne chez le mâle (Jubeaux et al 2012, Gouveia et al 2017), une réflexion devrait être menée sur le nombre d'individus analysés, la possibilité de travailler sur des pools d'individus (comme pour les données de bioaccumulation), ou encore la formalisation d'indicateur de perturbation (pourcentage d'individus positifs ?) pour pouvoir exploiter dans un cadre diagnostique ce type de biomarqueurs (notamment de perturbation endocrinienne) qui sont marqués par une variabilité inter-individu importante et un type de réponse plus qualitative (positif/négatif) que quantitative qui en découle.

Les valeurs seuils ont ensuite été appliquées aux données de biomarqueurs acquises à partir des gammars mis en cage en amont et en aval des deux stations d'épuration du projet Diadem. Parmi les 25 biomarqueurs peptidiques quantifiés et disposant de valeurs de référence, un seul peptide a été observé en dehors des valeurs seuils et a été considéré comme impacté. Plus précisément, il s'agit du peptide VPAILESFPGR qui a été trouvé à des concentrations inférieures au seuil d'inhibition chez 50% et 90% des gammars mis en cage en amont de la station de Bouillon et en aval de la station de Namur, respectivement (Figure 7).

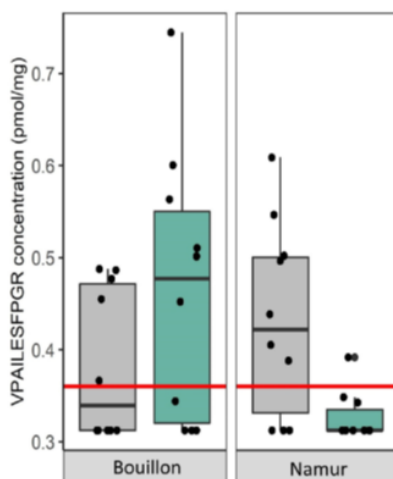


Figure 7 : modulation de la concentration en peptide VPAILESFPGR lors des encagements amont (en gris) et aval (en vert) des deux stations d'épuration du projet Diadem.

Pour aller plus loin, on peut noter que ce peptide modulé VPAILESFPGR est situé au niveau d'un site de clivage de la pro-PO (Leprêtre et al 2022a). Par conséquent, la diminution des concentrations de VPAILESFPGR observée chez les gammars mis en cage en amont de la station d'épuration de Bouillon et en aval de la STEP de Namur suggère une activation de l'enzyme PO dans ces deux sites. Au passage, on illustre ici encore que le recours à des valeurs de référence externes permet de qualifier de façon absolue la qualité des milieux, s'affranchissant de l'approche classique amont/aval, avec ici l'observation que la qualité du cours d'eau sur le site de Bouillon est dégradée dès l'amont du rejet étudié, ce qui conditionne les conclusions en termes d'impact de celui-ci sur le milieu récepteur.

2.4. Bilan

Une méthode analytique par protéomique ciblée a été développée, optimisée, validée pour le dosage multiplexé de concentrations de peptides rapporteurs de différentes fonctions biologiques chez le gammare. Elle permet une acquisition haut-débit de données multi-marqueurs (mesure de 38 biomarqueurs sur 325 échantillons individuels de gammare mâles en 2 mois d'analyse), correspondant à des concentrations de peptides rapporteurs de différentes grandes fonctions biologiques, et quantifiées de manière directe chez cette espèce sentinelle. Nous avons pu conclure à l'absence d'effet campagne et d'effet série, ce qui a été rendu possible à la fois par l'automatisation de la préparation des échantillons, la robustesse de l'analyse par HPLC-MS/MS, en combinaison avec le choix de travailler sur des organismes calibrés issus d'une même population source et exposés selon des protocoles fixés (biosurveillance active). Ceci aboutit à la possibilité de définir des niveaux de référence et des seuils d'induction ou d'inhibition de ces différents biomarqueurs protéiques au regard de la diversité des stations du réseau de surveillance nationale des Agences de l'eau. Les seuils ainsi définis offrent la possibilité de comparer ces données biologiques collectées sur une multitude de sites environnementaux et de dates d'échantillonnage. Les résultats des campagnes menées dans cette étude montrent la possibilité de discriminer une part significative des stations du réseau au regard d'induction et/ou d'inhibition de ces biomarqueurs protéiques. Dans un objectif d'application, les poursuites de ce travail sont double : (i) proposer un indicateur intégré multi-marqueur tenant compte de la déviation de chacun de ces biomarqueurs par rapport à sa gamme de référence (à l'image de l'adaptation de l'approche IBR appliquée dans Catteau et al 2022) ; (ii) travailler sur la définition de signatures spécifiques de différentes typologies d'exposition toxique (dans un but de discrimination de pressions) en bénéficiant de la multi-dimensionnalité offerte par le dosage multi-marqueur en protéomique ciblée.

3. Exercice multi-espèces : définition d'une stratégie commune et généralisable pour l'identification de biomarqueurs protéiques chez les espèces sentinelles

3.1. Choix des cas d'étude

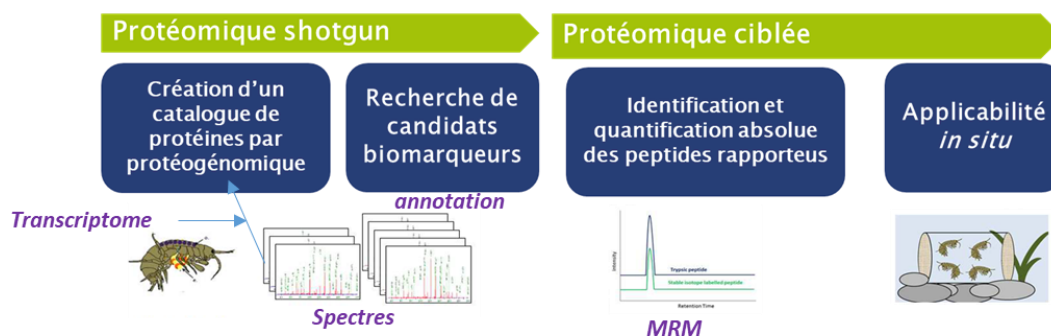


Figure 8 : Démarche générale en deux étapes proposée pour le développement de biomarqueurs protéiques en écotoxicologie; couplage des approches de protéomique massive de découverte (i.e. shotgun) et protéomique ciblée (adapté de Gouveia et al 2019).

Comme présenté dans l'introduction de ce rapport, les développements accomplis chez le gammare en protéomique pour le développement de biomarqueurs s'appuient sur la combinaison d'une première approche de protéomique dite de découverte (protéomique shotgun via le recours à la démarche protéogénomique chez les espèces non séquencées) qui permet d'établir des catalogues de protéines chez l'espèce d'étude (Figure 1), puis dans un deuxième temps, le recours à la protéomique ciblée (de type MRM) pour un dosage quantitatif de peptides rapporteurs de protéines choisies parmi les catalogues protéiques établis en première étape (Figure 8). L'objectif de ce deuxième volet de l'action a été de faire la démonstration qu'une stratégie commune, uniformisée, s'appuyant sur ce principe en 2 étapes, pouvait s'appliquer sur n'importe quelle espèce sentinelle d'intérêt. Cette étude de démonstration a ainsi visé l'identification de peptides rapporteurs de protéines impliquées dans trois grandes fonctions biologiques chez six organismes sentinelles des milieux aquatiques dulçaquicoles, estuariens et marins.







Espèce sentinelle	Gammare	Dreissène	Moule quagga	Crevette bouquet	Crevette blanche	Epinoche
	<i>Gammarus fossarum</i>	<i>Dreissena polymorpha</i>	<i>Dreissena r. bugensis</i>	<i>Palaemon serratus</i>	<i>Palaemon longirostris</i>	<i>Gasterosteus aculeatus</i>
	Crustacés	Bivalves	Bivalves	Crustacés	Crustacés	Poissons
Milieu	Eaux douces	Eaux douces	Eaux douces	Eaux marines	Eaux de transition	Eaux douces ou marines
Population d'étude	Population référence INRAE Lyon (Ain)	Population référence SEBIO Reims (lac du Der)	Population d'étude LIEC (Lorraine - Jouy)	Population d'étude en Cotentin SEBIO Le Havre	Population d'étude en Seine (Tancarville) SEBIO Le Havre	Elevage Ineris
						
Fonction/ Organe étudié	Osmorégulation Détoxification Immunité	Branchies Caeca Hémolymphe	Branchies Glande digestive Hémolymphe	Branchies Glande digestive Hémolymphe	Branchies Hépatopancréas Hémolymphe	Branchies Foie Sang

Table 3 : Espèces sentinelles et organes cibles retenus

Dans l'objectif de couvrir la diversité des espèces des milieux aquatiques, les six espèces sélectionnées appartiennent à trois grands groupes animaux : poissons, bivalves et crustacés (Table 3). Pour aboutir à cette sélection, une réunion a rassemblé en mai 2019 les différents partenaires engagés dans cette proposition de recherche. Dans un esprit de démonstration et d'opérationnalité, il a semblé pertinent de prendre en compte différents milieux (eaux douces, estuaires, eaux littorales) pour lesquels le développement de démarche de monitoring actif (transplantation d'une population de référence) est mené actuellement. C'est ainsi que le gammare (*Gammarus fossarum*), la crevette bouquet (*Palaemon serratus*), la dreissène (moule zébrée *Dreissena polymorpha*) et l'épinoche (*Gasterosteus aculeatus*) ont été retenues. Avec en tête l'évaluation de la transférabilité des biomarqueurs, ont été ajoutées à celles-ci des espèces proches (i.e., de même genre) et d'intérêt en biosurveillance au sein des laboratoires partenaires, à savoir la crevette blanche (estuarienne *Palaemon longirostris*) et la moule quagga (*Dreissena r. bugensis*). Il a été choisi de cibler des protéines potentiellement effectrices de grandes fonctions physiologiques retrouvées chez l'ensemble des groupes animaux, en ciblant le dosage de ces protéines dans trois types d'organes *a priori* impliqués dans la réalisation de ces trois fonctions : les branchies pour l'osmorégulation, la fonction d'immunité pour le sang/hémolymphe, et la détoxication pour le foie/hépatopancréas/caeca/glande digestive (Table 3).

Suivant la stratégie proposée (rappelée sur la Figure 8), la première étape consiste à définir un catalogue de protéines exprimées dans l'organe cible chez l'espèce d'intérêt. Pour pouvoir mettre en place l'approche de protéogénomique qui permet d'établir ces catalogues, deux types d'informations moléculaires sont nécessaires pour chacune des populations expérimentales dont sont originaires les organismes sentinelles utilisés: un transcriptome de référence qui permet d'accéder à la connaissance exacte des séquences protéiques spécifiques à la population d'étude, et une connaissance du sous-ensemble de ces protéines qui s'expriment réellement dans chacun des organes cibles (protéome d'organe). Dans notre étude, la technique de RNA-Seq a été utilisée sur les 6 espèces retenues afin de connaître via le séquençage du transcriptome, les séquences codantes pour les protéines en se basant sur des données acquises spécifiquement sur chaque population étudiée. La connaissance du répertoire de protéines spécifiquement exprimées dans l'organe cible s'est appuyé quant à lui sur l'acquisition de données de protéomique massive sur chacun des organes visés pour le dosage de biomarqueurs. Cette démarche protéogénomique (protéomique shotgun couplée à la connaissance du transcriptome de l'espèce) permet en effet d'établir sans a priori la connaissance des protéines exprimées dans les organes cibles (Figure 1).

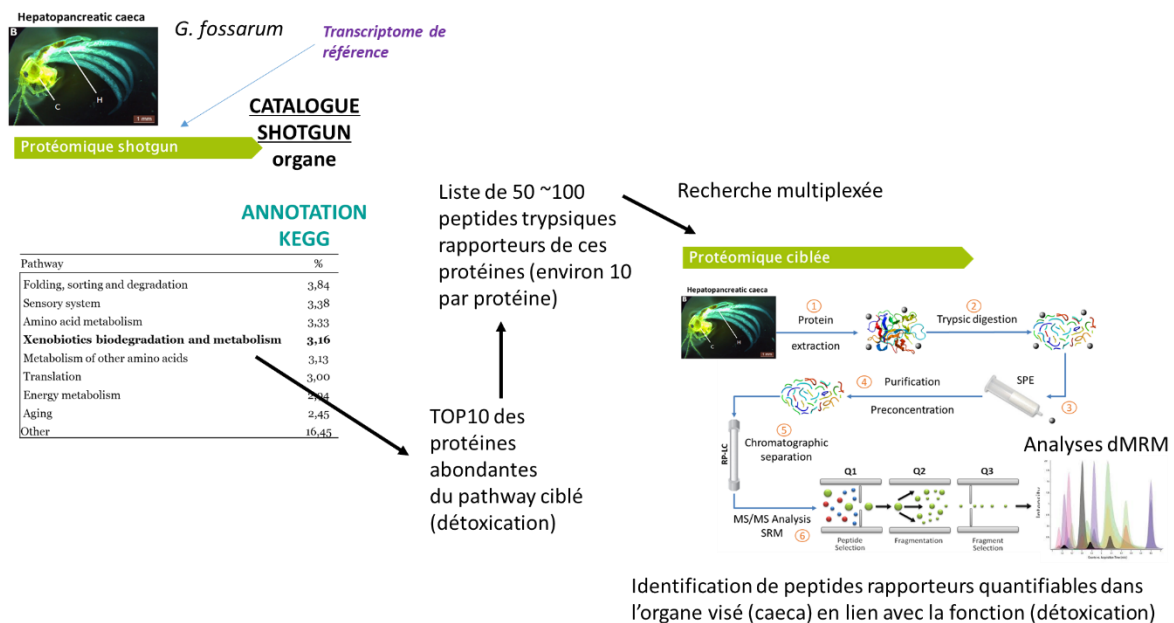


Figure 9 : Démarche suivie pour la recherche de peptides rapporteurs de protéines associées à une fonction physiologique. Stratégie appliquée à partir des catalogues protéomiques établis sur organes (chez *G. fossarum*, *D. polymorpha*, *P. serratus*, *G. aculeatus*) ; l'exemple est pris ici de l'identification de biomarqueurs de la fonction de détoxication analysés dans les caeca chez *Gammarus*.

Grâce à la protéogénomique, de potentiels biomarqueurs ont ainsi été sélectionnés parmi les catalogues protéiques des grandes fonctions d'intérêt sur chaque organe d'étude (Figure 9). Pour chaque organe (exemple « caeca hépatopancréatiques » sur la figure illustrative), l'annotation fonctionnelle des catalogues nous a permis d'identifier une dizaine de protéines abondantes parmi les protéines annotées fonctionnellement comme étant impliquée dans la fonction visée (exemple « détoxification ») (base de données KEGG; Kanehisa et al 2000). Une méthode multiplexée de protéomique ciblée a ensuite été développée pour valider l'identification d'une dizaine de peptides rapporteurs par protéine, donc une centaine de peptides par couple organe-espèce.

La transférabilité des biomarqueurs peptidiques entre espèces a par ailleurs été questionnée en adoptant une stratégie dédiée (Figure 10). Nous avons choisi pour les invertébrés de n'acquérir les données shotgun d'organes que chez une espèce par genre en faisant le pari que le protéome organe spécifique pourra être extrapolé entre espèces du même genre (ici entre dreissènes et entre crevettes palaemons). Notre hypothèse a été qu'une protéine abondante dans un organe comme l'hépatopancréas chez la crevette bouquet par exemple, a toutes les chances d'être également abondante dans l'hépatopancréas chez la crevette blanche (conservation fonctionnelle entre protéines orthologues), et que pour l'identification de cette protéine par protéomique ciblée MRM, on pourra prendre en compte les éventuelles différences de la séquence de cette protéine entre les espèces grâce à la connaissance du transcriptome établi pour toutes les populations d'étude. La démarche suivie est présentée avec l'exemple des dreissènes sur la Figure 10. Pour l'épinoche, poisson très étudié dans la littérature, une stratégie intermédiaire a été suivie en tentant de nous appuyer sur des données de protéomique shotgun sur organes obtenues sur d'autres populations disponibles dans la littérature.

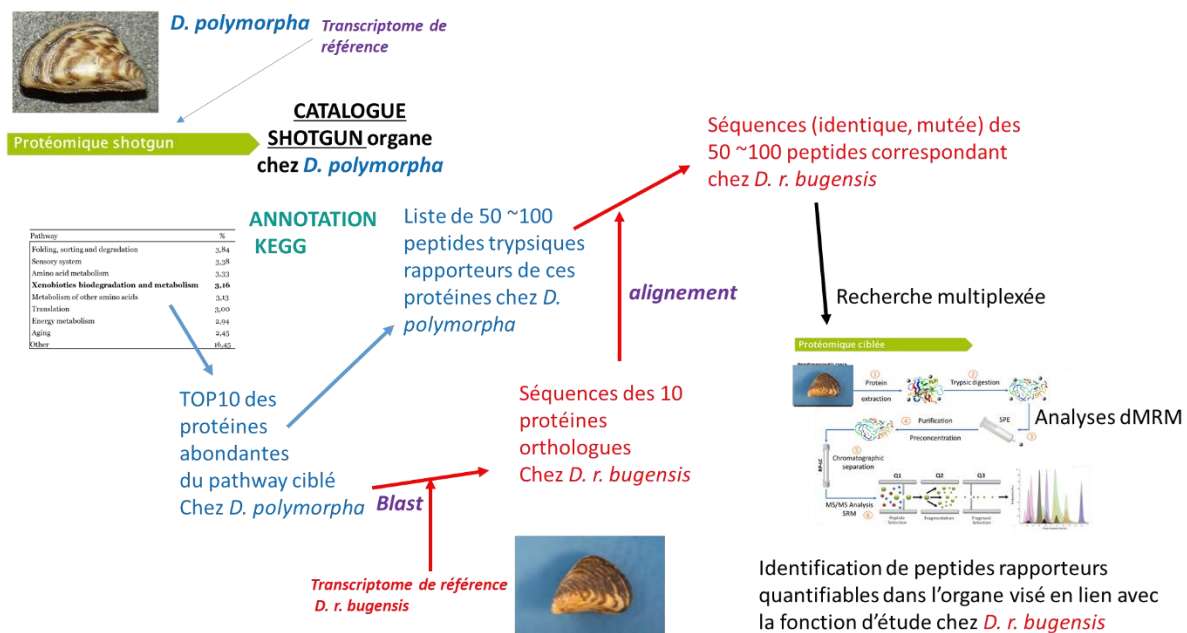


Figure 10 : Adaptation de la démarche suivie pour la recherche de peptides rapporteurs de protéines associées à une fonction physiologique à partir des catalogues protéomiques établis sur organes chez une autre espèce du même genre : l'exemple est pris ici de l'identification de biomarqueurs chez la moule quagga (*D. r. bugensis*) à partir des catalogues protéomiques de la moule zébrée (*D. polymorpha*).

3.2. Mobilisation et acquisition de ressources moléculaires chez les 6 espèces d'étude

L'ensemble des ressources moléculaires mobilisées sur les 6 espèces retenues est présenté dans la Table 4.

	crevette bouquet	crevette blanche	gammare	moule zèbre	moule quagga	épinoche
Transcriptome de référence	PRJNA809144 SRR18114569	PRJNA809144 SRR18095025	INRAE Cogne <i>et al</i> 2019	SEBIO Leprêtre <i>et al</i> 2019	LIEC Péden <i>et al</i> 2019	PRJNA809144 SRR18096867
Protéome « foie/hépatopancréas/caeca/glande digestive »	PXD031698		Dépôt en cours	LIEC		Nuez-Ortin <i>et al</i> 2018
Protéome « sang/hemolymphe »	PXD031698		PXD031777	SEBIO Leprêtre <i>et al</i> 2019 PXD012991		Kultz <i>et al</i> 2015
Protéome branchies	PXD031698		Dépôt en cours	PXD031681		Li <i>et al</i> 2018

Table 4 : Inventaire des ressources moléculaires mobilisées chez les 6 espèces d'étude. En bleu = données acquises au cours de l'étude ; en rouge = ressources préexistantes sur les populations d'étude au sein du consortium ; en vert données de la littérature sur d'autres populations. Les données transcriptomiques acquises au cours du projet ont été déposées sur ncbi (<https://www.ncbi.nlm.nih.gov/>) sous le code d'accès projet (PRJNA809144S), les données de protéomiques shotgun sur le dépôt PRIDE de l'EMBL (<https://www.ebi.ac.uk/pride/>). Les flèches bleues identifient la stratégie d'extrapolation des données de protéomique shotgun entre espèces du même genre (cf Figure 10).

3.2.1. Transcriptomes de référence (RNAseq)

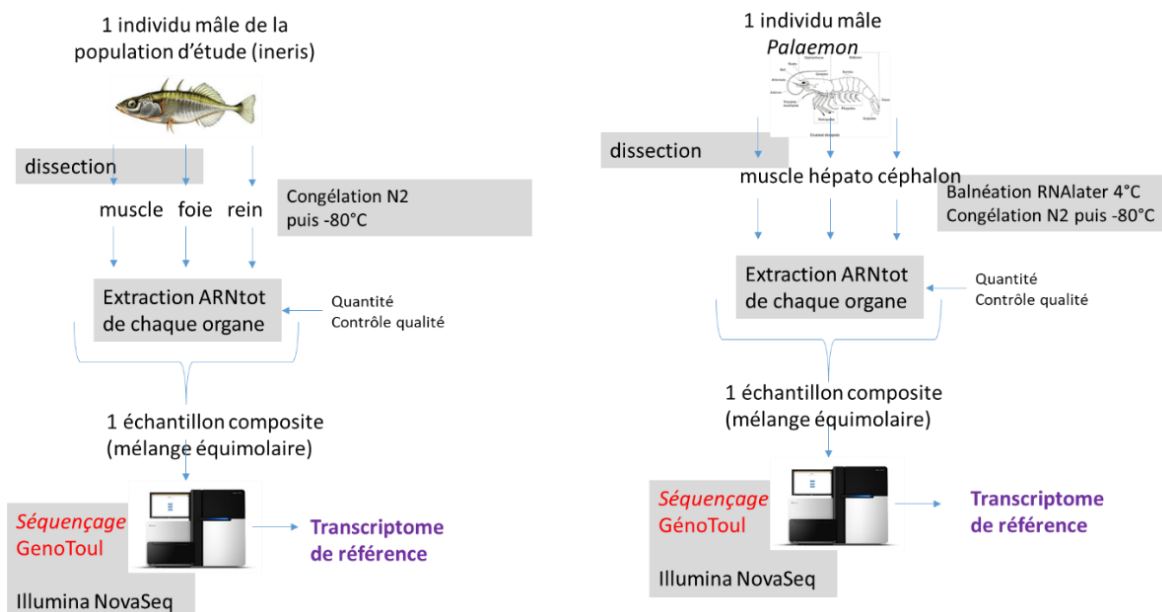


Figure 11 : Protocole suivi pour l'obtention des transcriptomes de référence pour la population d'épinoches de l'étude et pour les 2 espèces (marine et estuarienne) de crevettes Palaemon.

Pour posséder un transcriptome de référence pour chacune des populations d'étude des 6 espèces sentinelles cibles, la construction des 3 transcriptomes de référence manquants (l'épinoche et les deux crevettes *Palaemon*) a été réalisée (Figure 11). Le RNA-Seq de référence n'a pas été nécessairement acquis sur les organes ciblés par le dosage final en protéomique. Au contraire, nous nous sommes appuyés pour chacune des espèces sur un RNA-Seq profond réalisé sur des échantillons composites d'organes où l'extraction de matériel nucléotidique est le plus efficace. Le séquençage de l'ARN en profondeur permet en effet l'identification de la plupart du génome codant le protéome d'un organisme. Le séquençage a été réalisé à partir d'un seul individu de chacune des trois espèces en partant d'un mélange des ARN issus de différents tissus de ces individus (foie, muscle et rein pour l'épinoche ; hépatopancréas, muscle, céphalon pour les deux espèces de crevettes), dans un esprit de couverture large du transcriptome pour les bases de données de référence. Les échantillons ont été séquencés sur la plateforme GenoToul (technologie Illumina NovaSeq).

	épinoche	crevette bouquet	crevette blanche	gammare
	SRR18096867	SRR18114569	SRR18095025	Cogne <i>et al</i> 2019
Nbr bases	272 270 893	202 018 591	212 927 918	263 406 154
Nbr contigs	205 265	170 072	200 458	344 409
GC	49,11%	39,49%	39,14%	43%
Longueur médiane	552	377	372	296
Longueur moyenne	1 326	1 187	1 062	764
N50	2 805	3 466	3 068	1 354
Nbr protéines prédites (TransDecoder)	104 076	52 782	51 767	140 867
Nbr protéines annotées (Pfam & 10-3)	70 629 68%	30 681 58%	28 721 55%	55 977 40%
Analyse BUSCO				
arthropoda_odb10:1 013 BUSCOs	Complet 87,1%	87,8%	97,2%	90,7%
vertebrata_odb10:3 354 BUSCOs	Fragment 6,4%	5,1%	1,2%	5,3%
	Non trouvé 6,5%	7,1%	1,6%	4%

Table 5 : Caractéristiques des transcriptomes de référence obtenus sur épinoche et crevettes palaemons (et comparatif avec le transcriptome déjà décrit chez le gammare).

Par rapport au transcriptome utilisé chez le gammare pour nos précédents développements, l'analyse des caractéristiques des 3 nouveaux transcriptomes assemblés (Table 5) a révélé des assemblages moins fragmentés avec environ 200 000 contigs et des N50 de l'ordre de 3000pb, et présentant un niveau de complétion important (de l'ordre de 90% et plus), ce qui valide leur utilisation comme base d'information solide pour l'identification de protéines par protéogénomique.

3.2.2. Catalogues protéiques sur organes (protéomique shotgun)

Concernant le volet protéomique massive sur organe, des données ont été nouvellement acquises pour le gammare, la dreissène et la crevette bouquet sur les trois organes d'étude (Table 4). Pour l'épinoche, les connaissances protéomiques disponibles dans la littérature nous ont bien permis d'accéder à des protéomes sur plasma et branchies au sein d'autres populations internationales d'épinoches (Kültz *et al* 2015; Li *et al* 2018; Li et Kültz 2020), mais les données disponibles sur foie n'étaient pas suffisamment exhaustives et nous avons été contraints d'avoir recours à des données de catalogues protéiques disponibles chez le saumon (Nuez-Ortín *et al* 2018) en appliquant pour cet organe la stratégie imaginée initialement pour le transfert entre espèces du même genre (Figure 10).

L'acquisition des 7 nouveaux jeux de données shotgun a été réalisée sur la plateforme ProGenoMix du CEA à Marcoule. Les échantillons traités correspondent tous à des organes prélevés sur un unique individu (Figure 12), à l'exception de l'hémolymphe de gammares qui a nécessité de pooler les prélèvements de 25 individus. Les protéines de chaque échantillon sont séparées sur gel SDS-PAGE en 5 fractions différentes selon leur masse moléculaire. Après digestion trypsique, chacune de ces fractions est analysée séparément par spectrométrie de masse en tandem à l'aide d'un Q-Exactive HF (Thermo). Le fractionnement n'a pas été appliqué pour les données branchies et caeca chez le gammare, la plateforme s'étant dotée pour ces analyses d'un équipement de nouvelle génération

(Exploris 480). L'analyse des données de spectrométrie de masse en tandem est réalisée en se rapportant au transcriptome de référence de chaque population pour l'annotation des spectres obtenus et établir les catalogues protéiques de chaque organe.

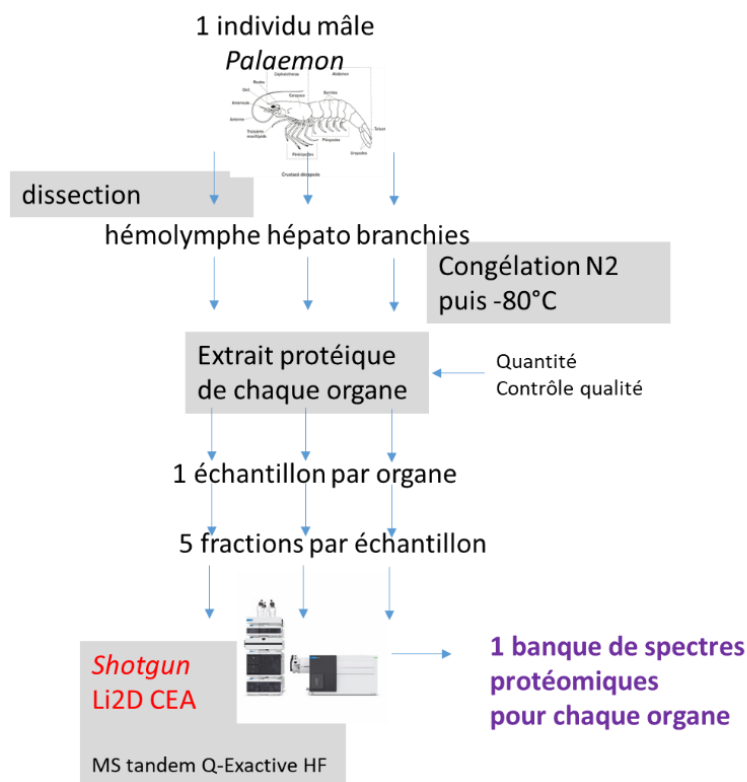


Figure 12 : Protocole mis en place pour l'obtention des 7 nouveaux protéomes d'organes par protéomique shotgun : exemple de *Palaemon serratus*.

Les catalogues obtenus permettent ainsi de disposer de l'information sur la séquence des protéines exprimées dans les différents organes, ainsi qu'une quantification relative en se basant sur les spectral count enregistrés pour les différents peptides appartenant à celles-ci. Une analyse fonctionnelle (base de données KEGG, Kanehisa et al 2000) des catalogues protéiques ainsi obtenus a permis de valider le choix des organes pour les 3 grandes fonctions visées pour le développement de biomarqueurs. Les Tables 6 (A,B,C) présentent les résultats de cette analyse chez les trois invertébrés et soulignent les différences marquées du répertoire fonctionnel qui s'exprime entre organes d'une même espèce, et au contraire une certaine similarité entre espèces pour un même type d'organe (ex pathway « immune system » dans l'hémolymphe des 3 espèces, ou encore « transport & catabolism » ou « signal transduction » dans les branchies ...). Dans la suite des développements, les protéines recherchées dans l'hémolymphe et ciblées sur l'immunité ont été choisies parmi les protéines les plus abondantes du groupe KEGG « immune system », pour l'osmorégulation sur branchies dans « transport & catabolism », et pour la détoxification dans le foie/hépatopancréas/caeca/glande digestive dans le pathway « xenobiotics biodegradation and metabolism » (catégories en jaune dans les Tables 6 et protéines détaillées en Annexe 2).

<i>Gammarus fossarum</i>				
Caeca		Hémolymphe		Branchies
3739 protéines détectées 2214 protéines annotées (59%)		627 protéines détectées 297 protéines annotées (47%)		1285 protéines détectées 770 protéines annotées (60%)
<i>KEGG Pathway</i>	%	<i>KEGG Pathway</i>	%	<i>KEGG Pathway</i>
Transport and catabolism	9,58	Immune system	19,84	Transport and catabolism
Endocrine system	8,07	Signal transduction	11,90	Cellular community - eukaryotes
Signal transduction	7,99	Cellular community - eukaryotes	10,50	Energy metabolism
Carbohydrate metabolism	6,63	Transport and catabolism	8,98	Signal transduction
Cellular community - eukaryotes	5,96	Endocrine system	8,04	Carbohydrate metabolism
Immune system	5,93	Cell growth and death	7,57	Cell growth and death
Lipid metabolism	5,23	Environmental adaptation	6,15	Environmental adaptation
Cell growth and death	4,92	Sensory system	5,88	Amino acid metabolism
Environmental adaptation	4,01	Signaling molecules and interaction	3,54	Nervous system
Folding, sorting and degradation	3,84	Biosynthesis of other secondary metabolites	1,80	Immune system
Sensory system	3,38	Amino acid metabolism	1,75	Excretory system
Amino acid metabolism	3,33	Carbohydrate metabolism	1,64	Endocrine system
Xenobiotics biodegradation and metabolism	3,16	Digestive system	1,19	Metabolism of other amino acids
Metabolism of other amino acids	3,13	Xenobiotics biodegradation and metabolism	1,04	Lipid metabolism
Translation	3,00	Nucleotide metabolism	0,88	Signaling molecules and interaction
Energy metabolism	2,94	Energy metabolism	0,70	Xenobiotics biodegradation and metabolism
Aging	2,45	Folding, sorting and degradation	0,66	Digestive system
Other	16,45	Metabolism of other amino acids	0,62	Other
		Cell motility	0,49	
		Other	2,98	

A

<i>Dreissena polymorpha</i>				
Glande digestive <i>données antérieures: LIEC</i>		Hémolymphe <i>données antérieures: SEBIO Zems</i>		Branchies
4092 protéines détectées 2193 protéines annotées (53%)		3719 protéines détectées 1403 protéines annotées (38%)		7919 protéines détectées 5121 protéines annotées (65%)
<i>KEGG Pathway</i>	%	<i>KEGG Pathway</i>	%	<i>KEGG Pathway</i>
Transport and catabolism	10,41	Immune system	10,27	Signal transduction
Signal transduction	7,74	Signal transduction	10,16	Cellular community - eukaryotes
Cellular community - eukaryotes	6,92	Cellular community - eukaryotes	9,24	Transport and catabolism
Endocrine system	6,59	Transport and catabolism	8,67	Cell growth and death
Carbohydrate metabolism	6,17	Endocrine system	7,14	Endocrine system
Immune system	6,07	Cell growth and death	7,10	Immune system
Cell growth and death	5,68	Carbohydrate metabolism	5,68	Environmental adaptation
Folding, sorting and degradation	5,14	Environmental adaptation	3,64	Cell motility
Translation	5,10	Translation	3,53	Sensory system
Environmental adaptation	4,03	Energy metabolism	3,50	Translation
Amino acid metabolism	3,66	Cell motility	3,41	Carbohydrate metabolism
Energy metabolism	3,06	Folding, sorting and degradation	3,38	Circulatory system
Cell motility	2,85	Nervous system	2,61	Folding, sorting and degradation
Lipid metabolism	2,76	Sensory system	2,34	Energy metabolism
Sensory system	2,15	Digestive system	2,26	Amino acid metabolism
Metabolism of other amino acids	2,10	Other	17,05	Nervous system
Nervous system	2,09			Excretory system
Metabolism of cofactors and vitamins	2,08			Digestive system
Xenobiotics biodegradation and metabolism	1,92			Signaling molecules and interaction
Other	13,47			Transcription
				Lipid metabolism
				Other

B

<i>Palaemon serratus</i>				
Hépatopancréas		Hémolymphe		Branchies
2023 protéines détectées 1261 protéines annotées (62%)		625 protéines détectées 285 protéines annotées (45%)		3717 protéines détectées 2041 protéines annotées (55%)
<i>KEGG Pathway</i>	%	<i>KEGG Pathway</i>	%	<i>KEGG Pathway</i>
Carbohydrate metabolism	19,18	Immune system	22,33	Transport and catabolism
Transport and catabolism	10,91	Endocrine system	13,52	Signal transduction
Endocrine system	6,58	Amino acid metabolism	10,18	Cellular community - eukaryotes
Immune system	6,19	Biosynthesis of other secondary metabolites	9,73	Cell growth and death
Signal transduction	5,73	Cellular community - eukaryotes	9,02	Endocrine system
Cell growth and death	5,49	Signal transduction	7,96	Energy metabolism
Amino acid metabolism	5,27	Transport and catabolism	5,31	Translation
Digestive system	4,97	Cell growth and death	4,08	Carbohydrate metabolism
Metabolism of other amino acids	3,59	Circulatory system	3,59	Immune system
Glycan biosynthesis and metabolism	3,47	Environmental adaptation	2,75	Folding, sorting and degradation
Lipid metabolism	3,33	Sensory system	2,65	Excretory system
Xenobiotics biodegradation and metabolism	3,15	Signaling molecules and interaction	2,60	Amino acid metabolism
Metabolism of cofactors and vitamins	2,7	Carbohydrate metabolism	1,37	Environmental adaptation
Biosynthesis of other secondary metabolites	2,47	Other	7,66	Digestive system
Energy metabolism	2,21			Nervous system
Translation	2,01			Circulatory system
Folding, sorting and degradation	1,86			Transcription
Other	10,9			Sensory system
				Glycan biosynthesis and metabolism
				Other

C

Table 6 : Catalogues protéiques des organes obtenus par protéomique shotgun (Analyse KEGG pathway, Kanehisa et al 2000), en jaune les pathways ciblés pour la sélection de protéines en lien avec la détoxification, l'immunité et l'osmorégulation dans chacun des 3 types d'organes.

3.3. Identification de peptides biomarqueurs par protéomique ciblée (MRM)

L'ensemble des développements pour la recherche de peptides rapporteurs des 10 protéines sélectionnées pour les différents couples organe-espèce, a été réalisé sur des échantillons individuels. La seule exception a été pour l'hémolymphe de gammare où les prélèvements de 10 individus ont été poolés et repris dans 50 µL de tampon HBSS.

3.3.1. Procédures analytiques

Extraction des protéines, digestion et SPE

Le protocole d'extraction des protéines sur organes est adapté de celui utilisé pour les organismes entiers chez le gammare (Simon et al 2010) (Figure 3). Le broyage de l'échantillon est réalisé dans un tampon tris (Tris 50 mM, NaCl 100 mM, EDTA 1 mM, Triton X-100 0.1% v/v) à pH 7,8 contenant 10 µg/mL des inhibiteurs à protéases leupeptine et aprotinine, à l'aide d'un broyeur à billes. Ce tampon permet de solubiliser les protéines. L'homogénat est ensuite centrifugé à 10 000 g pendant 15 min à +4°C afin de récupérer le surnageant dans un nouveau tube. Une solution de délipidation contenant de l'éthanol absolu et du diéthyléther en proportion 1 : 1 (v /v) est ajoutée au surnageant. Le mélange est ensuite vortexé et incubé 10 min dans de la glace puis centrifugé à 10 000 g pendant 10 min à +4°C. Le surnageant contenant la fraction lipidique solubilisée dans le solvant organique est éliminé alors que le culot contenant les protéines précipitées au contact du solvant organique est re-suspendu dans un volume de tampon tris différent selon les organes et espèces étudiées, la quantité de protéines totale étant différente dans chacun d'eux. L'extrait protéique est ensuite dosé par la méthode de Bradford afin de connaître la quantité de protéines totale dans l'extrait. Un spectrophotomètre Tecan Safire Microplate Reader est utilisé. Une fois l'extraction et le dosage des protéines terminés, celles-ci sont digérées par ajout de trypsine. Les peptides tryptiques sont ensuite purifiés par extraction sur phase solide (SPE) en pression positive et concentrés avant analyse. Notre postulat de départ était d'injecter l'équivalent de 50 µg de protéines-peptides afin d'obtenir une réponse la plus intense possible pour la détection de ces potentiels peptides biomarqueurs. De ce fait, la quantité de protéines initiale a été adaptée suivant le nombre d'injections souhaité par échantillon donc le volume d'extrait a varié d'un échantillon à l'autre (fonction de l'organe, de l'espèce, etc.). Les extraits protéiques sont alors complétés à 1,2 mL de bicarbonate d'ammonium (AMBIC) à 50 mM, tampon à pH 8 permettant à la trypsine d'avoir une activité optimale. Après ajout de 130 µL de dithiothréitol (DTT) à 150 mM, les échantillons sont incubés à 60°C pendant 40 min afin de réduire les ponts disulfures et linéariser les protéines. Après retour à température ambiante, les échantillons sont complétés avec 200 µL d'iodoacétamide (IAM) à 150 mM et portés à l'obscurité pendant 40 min à température ambiante afin d'alkyler les cystéines réduites et empêcher la reformation des ponts disulfure. La digestion est ensuite réalisée en ajoutant 150 µL d'une solution de trypsine TPCK à une concentration dépendante de la quantité de protéines initiale selon un rapport 1 : 1 de quantité protéines : trypsine. Les échantillons sont incubés à 37°C pendant 1h. 20 µL d'acide formique (AF) permettent finalement de bloquer la réaction de digestion. La purification des peptides est ensuite réalisée sur une cartouche de phase inverse Oasis HLB 3 cc (60 mg). Les échantillons sont d'abord centrifugés à 10 000 g pendant 5 min à température ambiante. La SPE se déroule alors comme suit : les cartouches sont d'abord conditionnées à l'aide de 1 mL de méthanol pur et 1 mL d'eau 0,5% d'acide formique (AF ; v:v), puis tout le surnageant de l'échantillon est déposé sur la cartouche, un rinçage avec 1 mL d'une solution eau/méthanol (95/5) 0,5% AF (v:v) permet d'éliminer les impuretés restantes et enfin l'élution des peptides est réalisée avec 1 mL de méthanol 0,5% AF (v:v). 100 µL d'une solution de méthanol contenant 10% de glycérol sont ajoutés afin de retenir les peptides au fond du tube lors de l'évaporation sous flux d'azote à sec pour concentration des peptides. Les culots sont re-suspendus dans une solution eau/acétonitrile (ACN) en proportions (90/10) 0,1% AF (v:v) avec un volume variant suivant la quantité de protéines initiale et suivant la quantité protéique d'injection souhaitée. Une dernière centrifugation est réalisée à 15 000 g pendant 5 min à température ambiante afin d'éviter d'injecter les derniers résidus indésirables non éliminés par les étapes précédentes. Tout le surnageant est alors transféré en vial avec insert pour injection en HPLC-MS/MS.

Paramètres HPLC-MS/MS

L'équipe d'écotoxicologie dispose d'une chaîne HPLC Agilent 1290 Infinity II couplée à un spectromètre de masse triple quadropôle Agilent QQQ 6495B équipé d'une source Electrospray

Agilent Jet Stream (AJS). Une colonne Waters Xbridge C18 de dimensions 100mmx2,1mmx3,5µm est utilisée pour la séparation des analytes par HPLC. 20µL d'échantillon sont injectés par analyse. Le four est thermostaté à 40°C. Le débit appliqué est de 300 µL/min sur une phase mobile contenant eau 0,1% AF (v:v) en voie A et ACN 0,1% AF (v:v) en voie B. Le gradient d'élution appliqué est le suivant : palier de 2 min à 5% d'ACN, puis gradient linéaire de 5% à 35% d'ACN en 36 min, palier de 5 min à 100 % d'ACN puis retour à 5% d'ACN pendant 5 min. Les paliers à 5% d'ACN permettent l'équilibrage de la colonne tandis que le palier à 100% d'ACN permet de laver la colonne des éventuelles impuretés restées accrochées en fin de gradient. La durée totale d'injection est de 48 min. Les paramètres du spectromètre de masse ont été réglés de la manière suivante : résolution unitaire des quadrupôles, dwell time de 5 ms, polarité positive, température et débit de gaz azote 210°C et 17 L/min respectivement, pression appliquée sur le nébuliseur 40 psi, température et débit du gaz secondaire azote 250c et 12 L/min respectivement, tension appliquée au capillaire de transfert des ions 4500 V, les 2 tensions appliquées sur l'iFunnel étaient de 145 V et 115 V. L'acquisition des données se fait grâce au logiciel MassHunter Data Acquisition (version B.09.00) et le retraitement se fait via le logiciel Skyline (64-bit version 21.1.0.278).

Sélection des transitions et création des méthodes MRM/dMRM

Afin d'analyser l'ensemble des peptides sélectionnés précédemment, des méthodes MRM et dMRM ont été développées. La première méthode MRM (Multiple Reaction Monitoring) permet l'analyse de 200 transitions tout au long de la méthode, une transition étant la sélection d'un couple ion précurseur-fragment. Si l'on considère qu'un peptide est généralement identifié à l'aide de 3 transitions, il est possible de suivre une soixantaine de peptides au cours d'une méthode MRM. Pour pallier à cette limite numérique, la seconde méthode appelée dMRM (dynamic Multiple Reaction Monitoring) permet de suivre des transitions sur une fenêtre de temps connue et beaucoup plus courte, de l'ordre de la minute, autour du temps de rétention (RT) attendu pour le peptide d'intérêt. In fine, cette dernière étant plus sensible permet l'analyse d'un nombre beaucoup plus élevé de transitions au cours d'une méthode et par conséquent, d'analyser beaucoup plus de peptides en une seule méthode, on parle de multiplexage. Cette méthode a pour avantages un gain de temps et de matériel biologique, il est cependant nécessaire de connaître les RT des peptides d'intérêt.

Le logiciel ProteinProspector (version 6.3.1) permet de prédire toutes les transitions qui seront préférentiellement détectées par (d)MRM, leur ordre d'intensité et le rapport entre les transitions. Ces informations sont lues dans le logiciel Skyline, ce dernier permettant de générer l'ensemble des transitions potentielles pour les peptides d'intérêt selon des critères de sélection choisis par l'utilisateur. Seuls les ions précurseurs di ou tri chargés sont retenus ainsi que des ions fragments mono voire di chargés de type y, b, x, z, a, c. Seuls les ions précurseurs et fragments avec un m/z allant de 250 à 1500 sont sélectionnés. Un peptide est considéré comme détecté et identifié lorsqu'au moins 3 transitions sont superposées. Au final, 6 transitions les plus intenses et répondant à tous ces critères ont été sélectionnées pour chaque peptide afin d'apporter une confiance supplémentaire dans l'identification du peptide lorsque les 6 transitions sont superposées. Dans une recherche optimale de 10 peptides pour chacune des 10 protéines pour chaque couple organe-espèce, il y avait plus de 600 transitions à analyser. Afin d'analyser un nombre aussi conséquent de transitions, une nouvelle stratégie pour prédire les RT de chacun de ces peptides a été mise en place afin de bénéficier de la grande capacité de multiplexage offerte par la dMRM. Pour chaque organe de chaque espèce, une méthode MRM a été initialement créée avec une dizaine de peptides les plus abondants d'après les données en shotgun (spectral count élevé) afin de déterminer le temps de rétention de ces peptides dans la méthode de 48 minutes. Il a été alors possible d'établir une relation linéaire entre les RT observés en shotgun et ceux observés dans notre méthode MRM 48 minutes (Faugère et al 2020), comme le montre la Figure 13. Ceci nous a permis ensuite de prédire les RT de tous les autres peptides ciblés pour l'organe de l'espèce et ainsi de créer directement des méthodes dMRM pour leur analyse. L'avantage de cette stratégie a été de pouvoir chercher un grand nombre de peptides en utilisant le moins de méthodes possibles et donc en un temps le plus court possible avec un besoin de matrice biologique réduit.

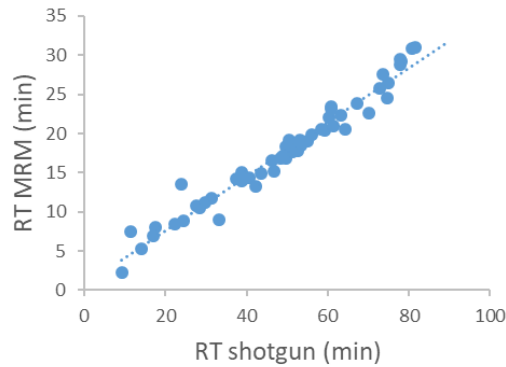


Figure 12 : Droite de corrélation entre les RT obtenus en shotgun versus ceux obtenus par MRM sur le caecum de l'espèce *Gammarus fossarum*.

Critères d'identification des peptides

Une fois l'échantillon analysé avec les méthodes MRM et/ou dMRM, les résultats obtenus sont exploités sur le logiciel Skyline (Figure 14). Pour chaque peptide, Skyline affiche son chromatogramme avec l'ensemble des transitions suivies, si ces dernières sont superposées sur un même pic le logiciel fournit le temps de rétention du peptide. Il affiche également les prédictions issues de ProteinProspector : le spectre de fragmentation prédit, l'ordre d'intensité des transitions prédit et les rapports de transitions prédits. Sur Skyline, ces prédictions sont nommées « library ».

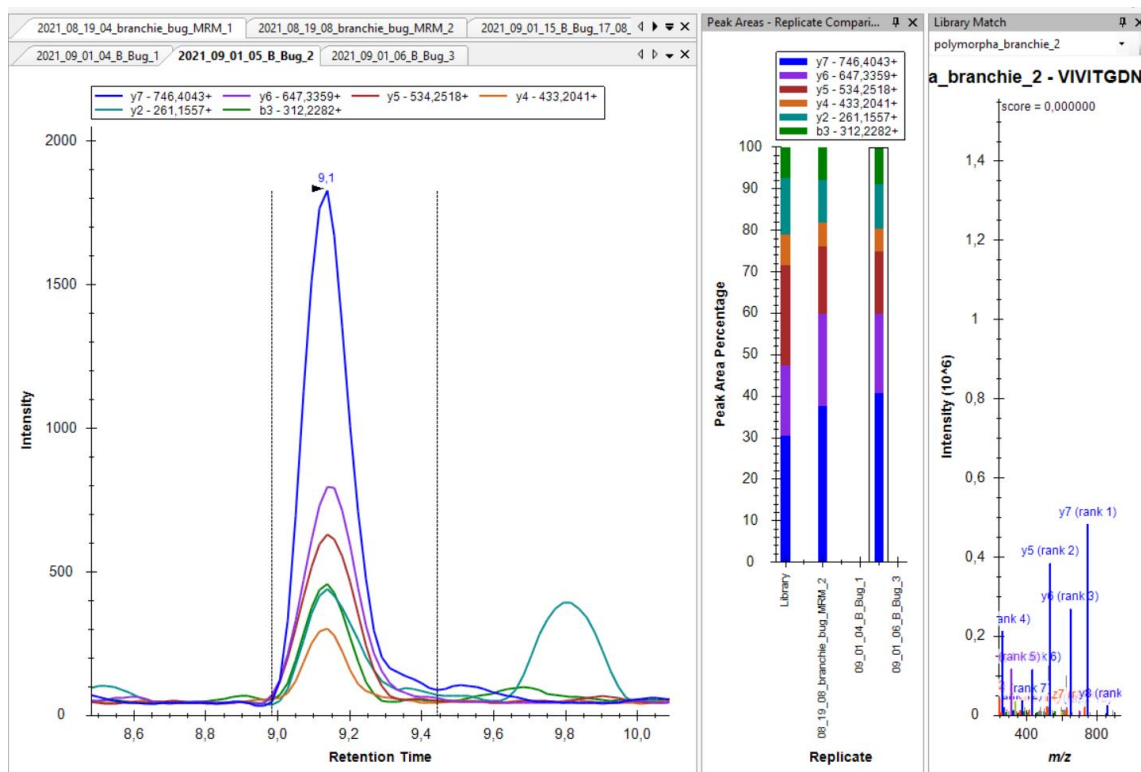


Figure 14 : Exemple de résultat Skyline pour le peptide VIVITGDNK (*Sarcoplasmic/endoplasmic reticulum calcium ATPase-like*, annotation GO *water homeostasis/osmo-régulation*) de la branchie de l'espèce *D. r. bugensis*. Gauche : chromatogramme du peptide, Milieu : rapports de transitions prédits et observés, Droite : spectre de fragmentation prédit.

Les résultats sont compilés dans un tableur Excel (exemple en Annexe 3 pour la branchie de *D. r. bugensis*) où sont renseignés le temps de rétention du pic chromatographique observé, son intensité maximale, le nombre de transitions alignées sur ce pic ainsi que des commentaires indiquant

notamment s'il n'y pas de correspondance avec la library ou si d'autres pics sont éventuellement présents sur le chromatogramme du peptide. Deux essais sont réalisés pour chaque organe afin de confirmer l'observation d'un peptide sur deux individus différents et d'apporter un degré de confiance supplémentaire. Un peptide a été considéré comme validé en termes d'identification lorsque les critères suivants étaient respectés :

- le nombre de transitions superposées et alignées sur un même pic est au minimum de 3.
- l'intensité des pics est supérieure à 250.
- correspondance entre les rapports de transitions obtenus en MRM et ceux prédits par la library, l'ordre d'intensité des transitions devant être similaire aux rangs des transitions prédits par le spectre de fragmentation.

3.3.2. Résultats sur les 3 organes chez *G. fossarum*, *D. polymorpha* et *P. serratus*

Dans un premier temps, nous reportons les résultats pour les 3 espèces pour lesquelles nous sommes appuyés sur des données shotgun acquises spécifiquement sur la population d'étude (stratégie présentée sur le Figure 9). Étant donné le nombre conséquent de résultats obtenus, il a été choisi pour le corps du rapport de présenter tout d'abord les résultats détaillés obtenus pour chaque organe de l'espèce *Gammarus fossarum*, puis dans un second temps les résultats globaux obtenus pour les deux autres espèces (*D. polymorpha* et *P. serratus*). L'ensemble des données détaillées équivalents sur les différentes espèces sont fournies en Annexe 4.

Gammarus fossarum

Les résultats sont présentés ici dans des tableaux faisant l'état pour chaque protéine ciblée du nombre de peptides observés après les deux essais (deux individus indépendants), avec en vert le nombre de peptides détectés avec une grande confiance selon les critères d'identification définis précédemment, en orange le nombre de peptides potentiellement identifiés avec une confiance moindre (RT observés différents entre les 2 essais, ou détection dans un seul essai ou signal de faible intensité), et en rouge le nombre de peptides non détectés.

Annotation KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Chloroalkane and chloroalkene degradation	Aldehyde dehydrogenase-like	GFBM_DN110488_c0_g1_i1: :g.154816::m.154816	4	4	2	10
		GFBM_DN75144_c0_g1_i1: :g.851303::m.851303	5	2	3	10
Caprolactam degradation	Aldose reductase-related protein-like	GFBM_DN113180_c0_g1_i2: :g.521934::m.521934	7	2	1	10
Caprolactam degradation; Aminobenzoate degradation	Trifunctional enzyme subunit alpha, mitochondrial-like	GFBM_DN119980_c5_g1_i1: :g.339948::m.339948	3	5	2	10
Drug metabolism - other enzymes	Xanthine dehydrogenase/oxidase-like	GFBM_DN118844_c0_g1_i2: :g.463279::m.463279	4	5	1	10
	Carboxylesterase	GFBM_DN101596_c0_g1_i2: :g.253580::m.253580	2	4	4	10
	Xanthine dehydrogenase-like	GFBM_DN118151_c0_g2_i1: :g.671259::m.671259	2	5	3	10
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Glutathione S-transferase 1-like	GFBM_DN109777_c0_g1_i2: :g.746927::m.746927	7	2	0	9
	UDP-glucuronosyl/UDP-glucosyltransferase	GFBM_DN117325_c0_g1_i1: :g.9889::m.9889	7	1	2	10
	UDP-glucuronosyl/UDP-glucosyltransferase	GFBM_DN117330_c4_g1_i1: :g.14917::m.14917	7	1	2	10
Total			48	31	20	99

Table 7 : Résultats de la recherche de peptides potentiels biomarqueurs de la « détoxification » sur le caecum de gammare.

Pour le caecum de gammare (Table 7), 99 peptides ont été recherchés par (d)MRM avec un total de 792 transitions. Ces derniers appartiennent à 10 protéines associées à des fonctions de détoxification par l'approche KEGG. Sur les 99 peptides, 48 sont identifiés soit environ la moitié des peptides recherchés, avec au minimum 2 peptides identifiés pour chacune des 10 protéines visées. S'ajoutent

31 peptides qui resteraient à être confirmés avec des analyses complémentaires. Ces résultats très satisfaisants permettent d'affirmer que le caecum est un organe de choix pour étudier les mécanismes de détoxification chez le gammare.

38 peptides ont été sélectionnés pour l'hémolymphe de gammare appartenant à 10 protéines associées *a priori* à l'immunité via l'approche KEGG : les cascades de la coagulation et du complément, l'activation des plaquettes et la migration trans-endothéliale des leucocytes. On peut noter ici tout de même une limite de l'approche KEGG avec la sur-représentation de protéines d'actine, certainement peu spécifiques aux processus d'immunité (Table 8). Au total 324 transitions ont été suivies par 2 méthodes dMRM.

Annotation KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Complement and coagulation cascades	Alpha-2-macroglobulin-like	GFBM_DN120882_c4_g1_i7::g.302215::m.302215	7 dont 2 doublons avec 100087	1	2	10
		GFBM_DN106701_c0_g1_i1::g.516543::m.516543	1	1	0	2
		GFBM_DN100087_c0_g1_i2::g.764399::m.764399	4 dont 2 doublons avec 120882	0	1	5
Platelet activation; Leukocyte transendothelial migration	Actin	GFBM_DN118952_c0_g3_i1::g.476665::m.476665	5 dont 5 doublons	1 dont 1 doublon	0	6
		GFBM_DN112551_c1_g2_i2::g.573473::m.573473	4 dont 3 doublons	1 dont 1 doublon	0	5
		GFBM_DN120284_c2_g3_i2::g.215083::m.215083	7 dont 7 doublons	1 dont 1 doublon	0	8
		GFBM_DN116398_c2_g3_i2::g.46357::m.46357	7 dont 4 doublons	0	1	8
		GFBM_DN118080_c0_g1_i1::g.1971::m.1971	4 dont 4 doublons	1 dont 1 doublon	0	5
		GFBM_DN88039_c0_g1_i1::g.819692::m.819692	5 dont 5 doublons	0	0	5
Platelet activation; Complement and coagulation cascades	Hemocytin-like	GFBM_DN120348_c0_g2_i1::g.39721::m.39721	0	3	5	8
/	Beta-1,3-glucan-binding protein precursor	/	8	1	5	14
Total			31	7	14	52

Table 8 : Résultats de la recherche de peptides potentiels biomarqueurs de l'immunité sur l'hémolymphe de gammare.

Pour les 3 processus liés à l'immunité pré-cités, 9 peptides sur les 38 recherchés n'ont finalement pas été détectés. Les 29 autres peptides ont été identifiés mais avec beaucoup de doublons entre les protéines. Pour l'exercice de démonstration et afin de cibler des protéines plus spécifiquement en lien avec l'immunité, une recherche manuelle dans les annotations du shotgun nous a permis d'ajouter la protéine beta-1,3-glucan-binding protein à la liste des protéines sélectionnées (dernière ligne de la Table 8). Cette dernière est connue pour être impliquée dans le système de défense contre les pathogènes chez les invertébrés (Phupet et al 2018). Sur la séquence identifiée chez le gammare, elle comprend 14 peptides tryptiques dont 8 ont été identifiés avec une forte confiance, 1 serait à confirmer et 5 n'ont pas été détectés. Au total, 60% des peptides recherchés en lien avec l'immunité ont été identifiés dans l'hémolymphe avec une grande confiance. Finalement, seule une protéine sur les 11 (l'hémocytin-like) n'a pas donné de résultats convaincants.

Concernant les branchies de gammare, seulement 25 peptides répondant aux critères de sélection ont été cherchés (Table 9). Ils appartiennent à 7 protéines différentes associées à l'osmorégulation dans les bases KEGG. Un total de 198 transitions a été analysée en MRM. 8 peptides ont été identifiés, 9 n'ont pas été détectés et 8 restent à confirmer.

Annotation KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
water homeostasis; osmoregulation	IF-complex protein 1 subunit gamma-like	GFBM_DN82398_c0_g1::GFBM_DN82398_c0_g1_i1::g.80593::m.80593	6	2	2	10
	IF-complex protein 1 subunit gamma-like	GFBM_DN113594_c4_g2::GFBM_DN113594_c4_g2_i3::g.89565::m.89565	0	1	2	3
	Clathrin light chain-like	GFBM_DN118869_c5_g2::GFBM_DN118869_c5_g2_i1::g.45372::m.45372	0	3	4	7
	Adaptor protein AP-2 complex subunit alpha-like	GFBM_DN117888_c0_g1::GFBM_DN117888_c0_g1_i2::g.153913::m.153913	1	0	0	1
	Protein kinase-like	GFBM_DN120756_c0_g3::GFBM_DN120756_c0_g3_i1::g.43521::m.43521	1	0	0	1
	AMP-dependent protein kinase catalytic subunit 1-like	GFBM_DN82432_c0_g1::GFBM_DN82432_c0_g1_i3::g.102115::m.102115	0	1	1	2
	Serine/Threonine-protein kinase OSR1-like	GFBM_DN113326_c3_g2::GFBM_DN113326_c3_g2_i2::g.224363::m.224363	0	1	0	1
Total			8	8	9	25

Table 9 : Résultats de la recherche de peptides potentiels biomarqueurs de l'osmorégulation sur les branchies de gammare.

Bilan sur les 3 espèces *Gammarus fossarum*, *Dreissena polymorpha* et *Palaemon serratus*

Les résultats d'identification des peptides pour l'ensemble des couples organe-espèce de l'étude sont synthétisés dans la Table 10 ci-après. Le détail des peptides détectés ou non selon les annotations protéiques est donné en Annexe 4.

		Identifiés avec forte confiance	Identification à confirmer	Non détectés	Total de peptides cherchés	
<i>Palaemon serratus</i>	Branchie	15	15	24	54	229
	Hémolymphe	20	27	46	93	
	Hépatopancréas	23	11	48	82	
<i>Palaemon longirostris</i>	Branchie	6	8	35	49	176
	Hémolymphe	1	12	40	53	
	Hépatopancréas	19	9	46	74	
<i>Dreissena polymorpha</i>	Branchie	27	18	17	63	216
	Hémolymphe	13	40	30	83	
	Glande Digestive	33	21	16	70	
<i>Dreissena r. bugensis</i>	Branchie	19	12	27	58	162
	Hémolymphe	8	7	22	37	
	Glande Digestive	34	13	20	67	
<i>Gammarus fossarum</i>	Branchie	8	8	9	25	176
	Hémolymphe	31	7	14	52	
	Caecum	48	31	20	99	
<i>Gasterosteus aculeatus</i>	Branchie	16	11	5	32	70
	Sang	3	5	4	12	
	Foie	8	12	6	26	

Table 10 : Bilan global de l'étude synthétisant les résultats de l'identification dMRM des peptides pour les 18 couples organes-espèces, peptides potentiels biomarqueurs de l'osmorégulation sur branchies, de l'immunité sur hémolymphe-sang, et de la détoxification sur hépatopancréas-caecum-foie-glande digestive.

Au total pour le gammare, sur les 176 peptides tout organe confondu, 87 ont été observés et confirmés dans les deux essais réalisés, ce qui représente 49% du total. 24% n'ont pas du tout été détectés et 26% seraient à confirmer (par de nouveaux essais, ou encore via le recours à des peptides synthétiques). En ce qui concerne *D. polymorpha*, le taux de peptides observés et confirmés dans les deux essais est de 34%, taux qui monte à 71% si on considère les peptides qui resteraient à être confirmés. Une disparité entre organes apparaît avec environ 45% des peptides recherchés qui sont détectés dans les branchies et la glande digestive, alors que le taux d'identification est beaucoup plus faible (15%) dans l'hémolymphe. Pour *P. serratus*, le taux de détection est plus homogène entre organes avec en moyenne 27% des peptides qui sont validés. Toutefois beaucoup de peptides ne

son pas du tout détectés (52%). Malgré ce taux plus faible d'identification des peptides, on peut noter qu'en termes de protéines pour lesquelles au moins un peptide rapporteur est validé ou encore à confirmer, on couvre encore un répertoire important des fonctions visées (Annexe 4), avec des exemples de protéines candidates intéressantes (e.g. GST, cytochrome P450 dans l'hépatopancréas).

De manière générale, l'identification de peptides biomarqueurs s'est avérée encourageante. Le nombre de peptides détectés est au moins égal au tiers du nombre de peptides total. Ce constat est d'autant plus encourageant que les méthodes HPLC-MS/MS n'ont pas été optimisées ici spécifiquement pour ces peptides, les paramètres étant ceux utilisés pour le premier axe du projet (sur gammare entier). Pour la suite, si l'on souhaite valider l'utilisation de ces peptides comme biomarqueurs, il faudra les valider par ajout de peptides synthétiques à la méthode.

3.3.3. S'appuyer sur des catalogues protéomiques de la littérature : exemple de l'épinoche

Pour l'épinoche, *G. aculeatus*, nous nous sommes appuyés sur des données shotgun de la littérature acquises sur d'autres populations pour le plasma et les branchies, et sur le saumon atlantique pour le foie. C'est dans les branchies que le taux de détection de peptides a été le plus important (Table 10). En effet, la moitié des peptides sélectionnés a été identifiée avec une forte confiance. Ensuite, un tiers des peptides sélectionnés dans le foie ont été détectés dans les deux essais et enfin, un quart dans le sang. Il faut cependant noter que pour cette espèce, beaucoup moins de peptides candidats ont pu être définis et ensuite recherchés en (d)MRM. Ceci s'explique par le fait qu'on ait été finalement contraint par l'absence de données shotgun spécifiques de la population d'étude voire de l'espèce, ce qui s'est traduit par le fait d'avoir été tributaire de la qualité des données protéomiques disponibles, ainsi qu'une perte importante de peptides candidats lors de la phase d'alignement de séquences sur le transcriptome de notre population d'étude (Figure 10).

Toutefois, entre 25% et 50% des peptides ont pu être identifiés dans chaque organe ce qui est cohérent avec les résultats sur les 3 premières espèces et n'invalide pas totalement cette stratégie alternative.

3.3.4. S'appuyer sur des catalogues protéomiques acquis sur une autre espèce du même genre : exemple des *Palaemon* et des dreissènes

Dans cette partie sont détaillés les résultats d'identification pour les 2 espèces de crevettes *P. longirostris* et *D. r. bugensis*, pour lesquelles le shotgun d'une espèce du même genre est utilisé comme référence (Figure 10). Ainsi, seul le RNA-Seq a été acquis chez les 2 espèces *P. longirostris* et *D. r. bugensis* afin d'identifier les séquences (identiques ou mutées) d'une liste de peptides orthologues aux peptides recherchés chez les *P. serratus* et *D. polymorpha* respectivement. Assez logiquement, le nombre de peptides recherchés est inférieur du fait principalement d'échec dans la recherche de ces peptides orthologues (Table 10). Cette perte de peptides candidats est relativement modeste si ce n'est pour l'hémolymphe chez les dreissénidés. L'ensemble des peptides sélectionnés et des correspondances entre espèces sont présentés en Annexe 5.

Sur les 54 peptides sélectionnés pour *P. serratus* dans les branchies, 46 étaient identiques chez *P. longirostris*, 5 n'étaient plus présents et 3 présentaient des mutations sur 1 à 2 acides aminés. Ces 3 derniers peptides à mutations ne sont d'ailleurs détectés ni chez *P. longirostris* ni chez *P. serratus* sans mutations. En ce qui concerne les peptides analysés dans l'hémolymphe de ces deux espèces, sur les 94 peptides initiaux de *P. serratus*, 40 peptides sont identiques entre les deux espèces, 48 ne sont pas présents dans le transcriptome de *P. longirostris* et 6 présentent 1 à 2 mutations d'acides aminés par rapport à *P. serratus*. Les 6 peptides à mutations ne sont pas détectés alors que 3/6 le sont chez *P. serratus*. Pour la sélection des peptides de l'hépatopancréas de ces deux espèces, seul un peptide n'est pas présent chez *P. longirostris*, 75 sont identiques entre les deux espèces et 6 ont des modifications sur 1 à 2 acides aminés. 3 peptides à mutations n'ont pas été analysés pour cause d'erreur opérateur. Sur les 3 autres, 1 n'est pas détecté alors qu'il l'était potentiellement chez *P. serratus* et les 2 autres sont bien détectés alors qu'ils ne l'étaient pas chez *P. serratus*. Au final, quasiment autant de peptides sont détectés dans les deux espèces.

Le bilan fait état d'un nombre de peptides conservés très important entre les deux espèces, entre 42 et 92% selon les organes. Très peu de peptides à mutations sont présents, souvent non détectés par rapport au peptide sans modification. Globalement, si le peptide n'est pas détecté chez *P. serratus*, alors il ne l'est pas non plus chez *P. longirostris*, et si inversement un peptide est détecté chez

P. longirostris alors il l'est aussi chez *serratus*. Les peptides avec ou sans mutations sont souvent détectés de la même façon dans l'une ou l'autre des deux espèces. Ceci signifie que la stratégie d'extrapolation des données protéomiques shotgun dans l'optique de généralisation des biomarqueurs au sein d'un même genre est pertinente pour l'analyse des deux espèces de *Palaemon*.

Concernant maintenant les deux espèces de bivalves, 42 peptides recherchés dans les branchies étaient identiques entre les deux espèces, 3 ne sont pas conservés chez *D. r. bugensis* et 18 ont des mutations sur 1 à 4 acides aminés dont 2 mutations apportant un miss cleavage dans la séquence du peptide (lysine ou arginine apparu dans la séquence peptidique) qui n'ont donc pas été analysés. Sur les 16 peptides à mutations restants, 11 n'ont pas été détectés (6 ne l'étaient pas non plus chez *D. polymorpha*, 3 restaient à être à confirmer et 2 étaient identifiés) et 5 demandent une potentielle confirmation (seul 1 n'était pas détecté chez *D. polymorpha*). Au final, 45% des peptides ne sont pas détectés chez *D. r. bugensis* alors que chez *D. polymorpha*, cela ne représentait que 27%.

Sur les 83 peptides sélectionnés dans l'hémolymphe de *D. polymorpha*, 21 sont identiques chez *D. r. bugensis*, 39 ne sont pas retrouvés dans le transcriptome et 23 présentent des mutations de 1 à 6 acides aminés dont 6 mutations apportant un miss cleavage dans la séquence du peptide ou l'absence d'une lysine ou arginine en C-ter du peptide (empêchant le clivage trypsique). Globalement, si le peptide à mutation n'est pas détecté chez *D. r. bugensis*, il ne l'était pas non plus chez *D. polymorpha*. Le nombre de peptides non détectés dans cet organe est assez important puisqu'il représente 60% de l'ensemble des peptides alors que chez *D. polymorpha* il se situait autour de 30%.

Au niveau de la glande digestive, 36 peptides étaient identiques sur 70, 9 ne sont pas présents dans le transcriptome de *D. r. bugensis* et 21 sont modifiés sur 1 à 3 acides aminés par peptide dont deux induisant un miss cleavage. Si le peptide à mutations n'est pas détecté, alors il ne l'était généralement pas non plus chez *D. polymorpha*, et à l'inverse, si le peptide à mutations est détecté, il l'était aussi chez *D. polymorpha*. 23% des peptides totaux de cet organe ne sont pas détectés chez *D. r. bugensis* contre 30% chez *D. polymorpha*.

Ainsi, comme pour les deux crevettes, les peptides trypsiques sont assez bien retrouvés entre les deux espèces de moules : 51% dans les glandes digestives, 23% dans l'hémolymphe et 67% dans les branchies. Il y a cependant beaucoup plus de mutations sur la séquence de ces peptides. Globalement, si le peptide n'est pas détecté chez *D. polymorpha*, alors il ne l'est pas non plus chez *D. r. bugensis*, ceci vaut également pour les peptides à mutations. Les peptides avec ou sans mutations sont souvent de « même couleur » dans les tables de l'Annexe 5 pour les deux espèces donc aussi bien détectés dans l'une ou l'autre des deux espèces.

Pour conclure cette partie de comparaison entre deux espèces d'un même genre qui a adopté la stratégie d'extrapolation du shotgun d'une des deux espèces à l'autre (Figure 10), il y a toujours plus de peptides détectés pour l'espèce référente ayant sa propre analyse de protéomique shotgun. Cette comparaison a également montré que si le peptide n'est pas détecté dans l'espèce référente alors il ne l'est pas non plus dans l'autre espèce même avec mutations. Au contraire, le peptide muté a plus de chance d'être détecté si son homologue non muté dans l'espèce référente est détecté. Il est plus rare de voir un peptide muté détecté alors que son homologue ne l'est pas. Ceci souligne que disposer de jeux de protéomique shotgun sur l'espèce d'étude donne la possibilité de pouvoir rechercher et finalement détecter plus de peptides, mais que la stratégie d'extrapolation des données protéomiques shotgun entre espèces du même genre est tout de même pertinente pour l'analyse puisqu'elle permet tout de même de détecter au moins 40% des peptides qui le sont dans la première espèce. D'un point de vue plus applicatif, si on raisonne à l'échelle des protéines, cette perte de peptides rapporteurs ne se traduit d'ailleurs pas comme une perte drastique du nombre de protéines qui peuvent être suivies dans la deuxième espèce, du fait que la perte des peptides est relativement aléatoire et pas lié à leur appartenance à une même protéine (Annexe 4).

3.4. Bilan

L'exercice de démonstration mené ici sur des espèces de 3 taxons très différents, sur des matrices biologiques de différentes natures (3 types d'organes) nous montre que la stratégie proposée par Gouveia et al (2019) (Figure 8) qui se base sur le couplage de la protéogénomique et de la protéomique ciblée, est clairement applicable à n'importe quelle espèce sentinelle. L'investissement nécessaire dans l'acquisition d'un RNAseq sur la population d'étude est aujourd'hui très faible (de l'ordre de quelques k€) et accessible (en prestation notamment sur des plateformes nationales) à n'importe quel laboratoire. L'exemple de l'épinoche nous a par ailleurs montré que s'appuyer sur

l'acquisition de données de protéomiques shotgun en propre (plutôt que de la littérature) est à privilégier car elle permet d'optimiser la recherche par protéomique ciblée d'un grand nombre de peptides candidats (notamment transfert de paramètres analytiques entre les 2 types de méthodes). Aujourd'hui, l'acquisition de données protéogénomiques est également pratiquée en routine et accessible à n'importe quel laboratoire via différentes plateformes de protéomiques au niveau national (de l'ordre de 1 k€ par échantillon d'organe).

Globalement, si on cible des peptides « bien vus » en shotgun (ici critère sur spectral counts), nous avons constaté un succès de l'ordre de 50% d'identification de chacun de ces peptides dans le suivi multiplexé en protéomique ciblée (dMRM), ce qui garantit de pouvoir suivre la grande majorité des protéines sélectionnées (chacune possédant différents peptides rapporteurs possibles). Pour l'exercice, nous ne nous sommes pas concentrés ici sur un choix fin des protéines à suivre dans chaque situation couple organe-espèce (approche automatique d'annotation via KEGG), mais un examen plus fin et à façon des annotations fonctionnelles au sein des catalogues protéiques shotgun ouvrira clairement la possibilité de se concentrer sur des protéines plus spécifiques des fonctions à suivre.

Enfin, l'exercice de transfert entre espèces d'un même genre réalisé ici chez *les Palaemon* et *Dreissena*, a montré la possibilité de développer des méthodes communes (existence de peptides rapporteurs communs) qui peuvent avoir un intérêt dans le cadre d'études qui visent à évaluer la différence de sensibilité entre espèces ou qui s'appuient sur la complémentarité des espèces (ex : suivi de continuum eaux douces, marines), voire entre populations (ex : approches de biosurveillance passive).

4. Conclusion

Les travaux engagés dans cette action de démonstration de l'approche « protéomique en biosurveillance » ont permis d'établir avec un réseau de laboratoires nationaux, une stratégie partagée et commune qui a été appliquée sur 6 espèces sentinelles issues de différents groupes taxonomiques (poissons, bivalves, crustacés) et d'intérêt actuel pour la surveillance des milieux aquatiques nationaux le long du continuum eaux-douces, eaux de transition et eaux marines.

Cette action a permis de développer une capacité d'acquisition haut-débit et automatisée de données de protéomique ciblée pour l'écotoxicologie. Nous avons ici pu démontrer que l'acquisition de ce nouveau type de données biomarqueurs peut répondre à des exigences de reproductibilité, d'inter-comparabilité des données produites (quantification directe) et permettre notamment la définition de valeurs de référence. Le premier retour sur le déploiement national opéré ici lors d'encagements de gammars sur les réseaux de suivi montrent que ces biomarqueurs peptidiques permettent de décrire des gradients de réponses des organismes à l'exposition (avec des réponses d'induction ou d'inhibition). Le haut degré de multiplexage offert par les technologies de protéomiques actuelles (de l'ordre de la centaine de protéines en (d)MRM ; voir Leprêtre et al 2022b) ouvre la possibilité de décrire des signatures biologiques de l'exposition (caractéristiques de voies d'action toxique, ou encore de l'exposition à différentes typologies de pressions), tout en questionnant le besoin d'indicateurs synthétiques pour le diagnostic de la qualité des milieux. Une telle réflexion sur des indicateurs intégrés peut clairement s'inspirer du travail déjà réalisé sur les indicateurs de contamination biodisponible basée sur les données de bioaccumulation multi-résidus chez le gammare (Alric et al 2019). Dans ce cadre, la démonstration multi-espèces menée dans le deuxième volet de cette action de recherche vient également souligner tout l'intérêt de ce nouveau type de biomarqueurs protéomiques dans notre capacité à développer des indicateurs multi-espèces plus intégrateurs de la pression chimique toxique au sein des écosystèmes (Cateau et al 2022) car basés sur des données comparables entre les espèces choisies comme sentinelles de la biodiversité qu'ils abritent.

5. Références

- Alric B, Geffard O, Chandesris A, Ferréol M, François A, Perceval O, Piffady J, Villeneuve B & Chaumot A (2019) Multisubstance Indicators Based on Caged Gammarus Bioaccumulation Reveal the Influence of Chemical Contamination on Stream Macroinvertebrate Abundances across France. *Environmental Science and Technology*, 53(10):5906-5915.
<https://dx.doi.org/10.1021/acs.est.9b01271>
- Armengaud et al (2014). Non-model organisms, a species endangered by proteogenomics. *J. Proteo*, 105 (2014): 5-18. <https://doi.org/10.1016/j.jprot.2014.01.007>
- Beliaeff & Burgeot (2002). Integrated biomarker response: A useful tool for ecological risk assessment. *Environ Toxicol Chem* 21: 1316-1322.
- Bervoets et al (2005). Use of transplanted Zebra mussels (*Dreissena polymorpha*) to assess the bioavailability of microcontaminants in Flemish surface waters. *Environ Sci Technol* 39: 1492-505.
- Besse JP, Coquery M, Lopes C, Chaumot A, Budzinski H, Labadie P & Geffard O (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 Research*, 47(2):650-660. <https://doi.org/10.1016/j.watres.2012.10.024>
- Borja et al (2008). Overview of integrative tools and methods in assessing ecological integrity in estuarine and coastal systems worldwide. *Mar. Pollut. Bull.* 56: 1519-1537.
- Catteau A, Porcher J-M, Bado-Nilles A, Bonnard I, Bonnard M, Chaumot A, David E, Dedourge-Geffard O, Delahaut L, Delorme N, François A, Garnero L, Lopes C, Nott K, Noury P, Palluel O, Palos-Ladeiro M, Quéau H, Ronkart S, Sossey-Alaoui K, Turiès C, Tychon B, Geffard O & Geffard A (2022) Interest of a multispecies approach in active biomonitoring: Application in the Meuse watershed. *Science Of The Total Environment*, 808:152148.
<https://doi.org/10.1016/j.scitotenv.2021.152148>
- Chaumot A, Geffard O, Armengaud J & Maltby L (2015). Chapter 11 - Gammarids as reference species for freshwater monitoring. In C. Mouneyrac, C. Amiard-Triquet, & J.-C. Amiard (Eds.), *Aquatic Ecotoxicology: Advancing tools for dealing with emerging risks* (pp. 253-280): Academic Press. <https://doi.org/10.1016/B978-0-12-800949-9.00011-5>
- Chaumot A, Perceval O, Recoura-Massaquant R, Degli-Esposti D & Geffard O (2021) Focus – Des bioessais écotoxicologiques in situ pour évaluer les impacts biologiques de la contamination chimique des cours d'eau nationaux : l'expérience du gammare. *Sciences Eaux & Territoires*(37):88-91. <https://doi.org/10.14758/SET-REVUE.2021.4.16>
- Cogne Y, Degli-Esposti D, Pible O, Gouveia D, François A, Bouchez O, Eché C, Ford A, Geffard O, Armengaud J, Chaumot A & Almunia C (2019) De novo transcriptomes of 14 gammarid individuals for proteogenomic analysis of seven taxonomic groups. *Scientific data*, 6(1):184.
<https://doi.org/10.1038/s41597-019-0192-5>.
- El-Gebali, Sara, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, et al (2019) The Pfam protein families database in 2019. *Nucleic Acids Research* 47 (D1): D427-32. <https://doi.org/10.1093/nar/gky995>.
- Espeyte et al (2019). Protéomique ciblée pour la quantification multiplexée de biomarqueurs: perspectives pour la surveillance environnementale - Cas d'étude chez *Gammarus fossarum*. Rapport d'étape, programme OFB-INRAE 2019-2022.
https://professionnels.ofb.fr/sites/default/files/pdf/Biosurveillance_OP/AFB_2019-2021_Action4_biomarqueurs_prot%C3%A9omiques_rapport_%C3%A9tape_2019.pdf
- Faugère et al (2020). High-multiplexed monitoring of protein biomarkers in the sentinel *Gammarus fossarum* by targeted scout-MRM assay, a new vision for ecotoxicoproteomics. *Journal of Proteomics* 226, 103901. <https://doi.org/10.1016/j.jprot.2020.103901>
- Geffard et al (2014). Développement d'une méthodologie pour l'amélioration du suivi chimique des eaux continentales Rapport de synthèse de l'étude pilote - Déploiement de l'outil gammare engagé au niveau national, résultats pour les métaux ciblés. Rapport Onema – programme 2013-2015.
<https://hal.inrae.fr/hal-02599952v1>
- Geffard et al (2019). Lien entre la toxicité, la contamination des milieux aquatiques mesurés chez *Gammarus fossarum* et la perturbation des communautés biologiques - Proposition de valeurs seuils au niveau national pour la mesure de marqueurs de toxicité chez *Gammarus fossarum*. Rapport AFB – programme 2016-2018. <https://hal.inrae.fr/hal-03799794>
- Geffard O, Coquery M, Recoura-Massaquant R, Perceval O, Bolzan D, Bourrain X, Rebillard JP, Navarro L, Nicolai M, Casterot B, Ferreol M, Villeneuve B & Chaumot A (2021) La biosurveillance active pour le suivi de l'état chimique des cours d'eau continentaux. *Sciences Eaux & Territoires*

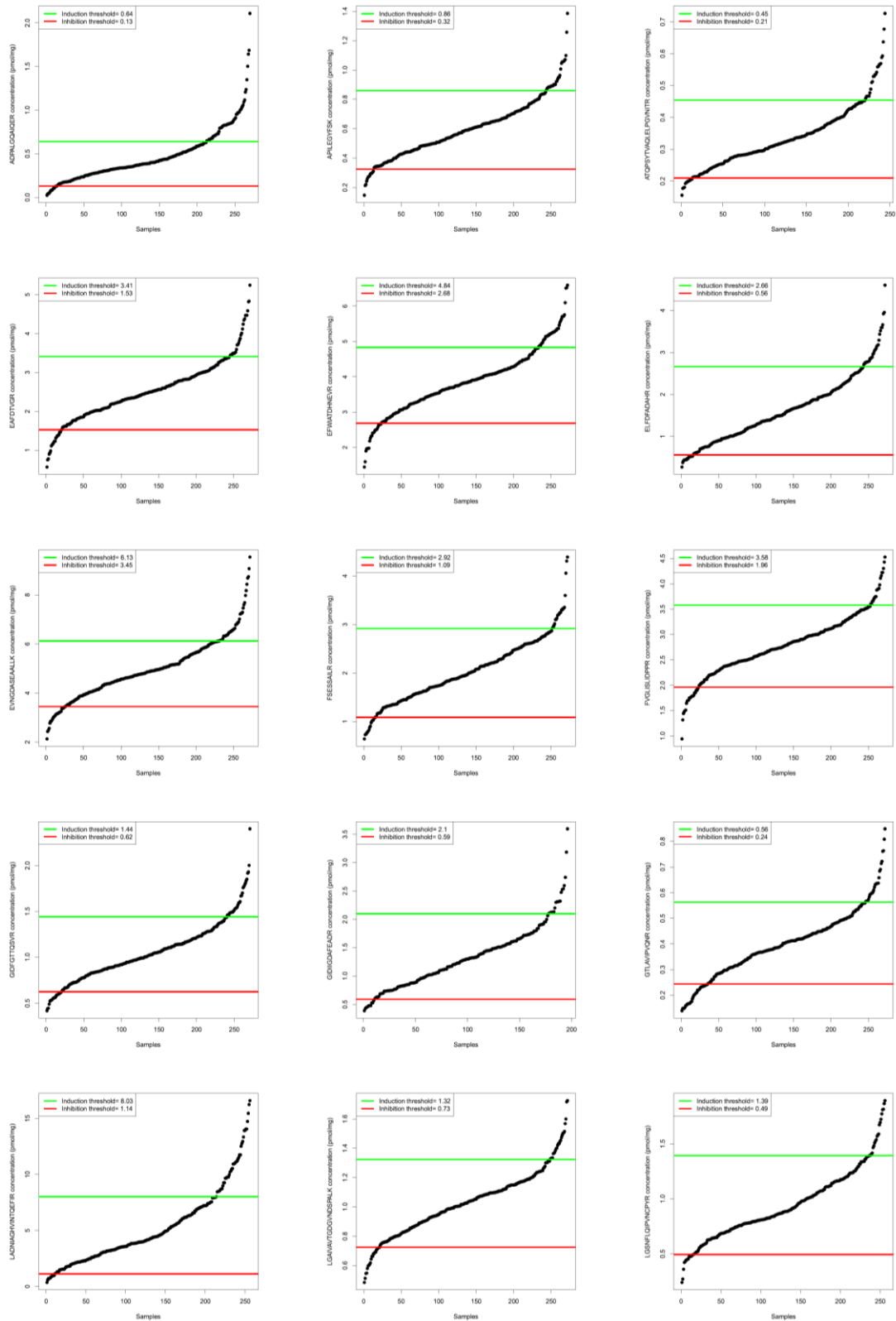
- (37):82-87. <https://doi.org/10.14758/SET-REVUE.2021.4.15>
- Gouveia et al (2017). Ecotoxicoproteomics for aquatic environmental monitoring: first in situ application of a new proteomics-based multibiomarker assay using caged amphipods. *Environ Sci Technol* 51: 13417-13426. <https://doi.org/10.1021/acs.est.7b03736>
- Gouveia D, et al (2019). Ecotoxicoproteomics: A decade of progress in our understanding of anthropogenic impact on the environment. *Journal of Proteomics*, 198:66-77. <https://doi.org/10.1016/j.jprot.2018.12.001>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 15;29(7):644-52.
- Jubeaux G, Simon R, Salvador A, Lopes C, Lacaze E, Quéau H, Chaumot A & Geffard O (2012) Vitellogenin-like protein measurement in caged *Gammarus fossarum* males as a biomarker of endocrine disruptor exposure: Inconclusive experience. *Aquatic Toxicology*, 122-123:9-18. <https://doi.org/10.1016/j.aquatox.2012.05.007>
- Kanehisa, M., et S. Goto. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28 (1): 27-30. <https://doi.org/10.1093/nar/28.1.27>.
- Kültz D, Li J, Zhang X, Villarreal F, Pham T, Paguio D. (2015) Population-specific plasma proteomes of marine and freshwater three-spined sticklebacks (*Gasterosteus aculeatus*). *Proteomics* 15 (23-24): 3980-92. <https://doi.org/10.1002/pmic.201500132>.
- Leprêtre M, Almunia C, Armengaud J, Salvador A, Geffard A & Palos-Ladeiro M (2019) The immune system of the freshwater zebra mussel, *Dreissena polymorpha*, decrypted by proteogenomics of hemocytes and plasma compartments. *Journal of Proteomics*, 202:103366. <https://doi.org/10.1016/j.jprot.2019.04.016>
- Leprêtre M, Geffard A, Palos Ladeiro M, Dedourge-Geffard O, David E, Delahaut L, Bonnard I, Barjhoux I, Nicolai M, Noury P, Espeyte A, Chaumot A, Degli-Esposti D, Geffard O & Lopes C (2022a) Determination of biomarkers threshold values and illustration of their use for the diagnostic in large-scale freshwater biomonitoring surveys. *Environmental Sciences Europe*, 34(1):115. <https://doi.org/10.1186/s12302-022-00692-2>
- Leprêtre M, Geffard O, Espeyte A, Faugère J, Ayciriex S, Salvador A, Delorme N, Chaumot A & Degli-Esposti D (2022b) Multiple reaction monitoring mass spectrometry for the discovery of environmentally modulated proteins in an aquatic invertebrate sentinel species, *Gammarus fossarum*. *Environmental Pollution*:120393. <https://doi.org/10.1016/j.envpol.2022.120393>
- Li, Johnathon, et Dietmar Kültz (2020) Proteomics of osmoregulatory responses in threespine stickleback gills. *Integrative and Comparative Biology* 60 (2): 304-17. <https://doi.org/10.1093/icb/icaa042>
- Li, Johnathon, Bryn Levitan, Silvia Gomez-Jimenez, & Dietmar Kültz (2018) Development of a gill assay library for ecological proteomics of threespine sticklebacks (*Gasterosteus Aculeatus*). *Molecular & Cellular Proteomics: MCP* 17 (11): 2146-63. <https://doi.org/10.1074/mcp.RA118.000973>.
- Nuez-Ortín, W.G., Carter, C.G., Nichols, P.D. et al Liver proteome response of pre-harvest Atlantic salmon following exposure to elevated temperature. *BMC Genomics* 19, 133 (2018). <https://doi.org/10.1186/s12864-018-4517-0>
- Péden, Romain, Pascal Poupin, Bénédicte Sohm, Justine Flayac, Laure Giambérini, Christophe Klopp, Fanny Louis, et al (2019) Environmental transcriptomes of invasive *Dreissena*, a model species in ecotoxicology and invasion biology. *Scientific Data* 6 (1): 234. <https://doi.org/10.1038/s41597-019-0252-x>.
- Phupet et al (2018) Lipopolysaccharide- and β -1,3-glucan-binding protein from *Litopenaeus vannamei*: Purification, cloning and contribution in shrimp defense immunity via phenoloxidase activation. *Developmental and Comparative Immunology*, <https://doi.org/10.1016/j.dci.2017.11.016>
- Seppey, Mathieu, Mosè Manni, et Evgeny M. Zdobnov (2019) BUSCO: Assessing Genome Assembly and Annotation Completeness. In *Gene Prediction: Methods and Protocols*, édité par Martin Kollmar, 227-45. *Methods in Molecular Biology*. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-9173-0_14.
- Simon et al (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*). *Journal of Chromatography A*, 1217(31):5109-5115. <https://doi.org/10.1016/j.chroma.2010.06.015>
- Trapp et al (2014a). Next-generation proteomics: Toward customized biomarkers for environmental

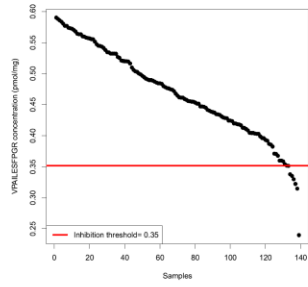
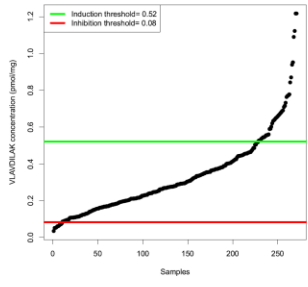
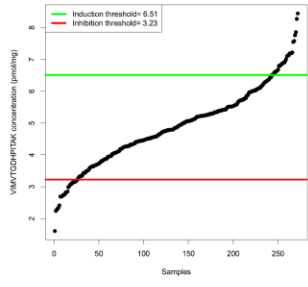
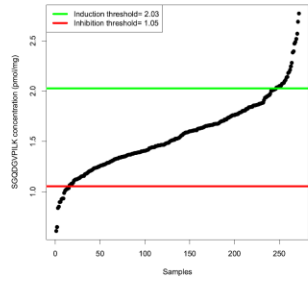
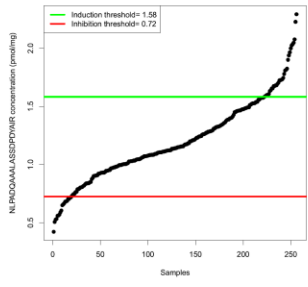
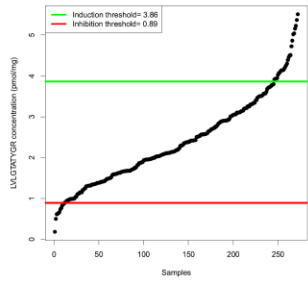
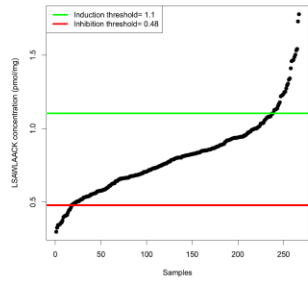
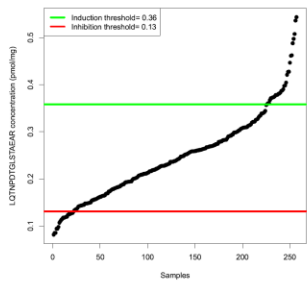
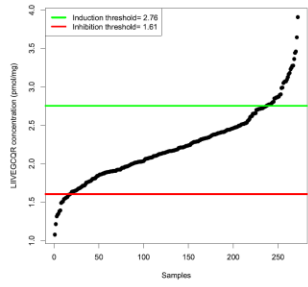
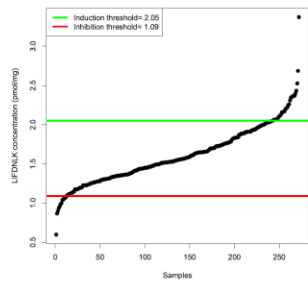
biomonitoring. Environ Sci Technol 48: 13560-13572. <https://doi.org/10.1021/es501673s>
Trapp J, et al (2014b) Proteogenomics of Gammarus fossarum to Document the Reproductive System of Amphipods. Molecular & Cellular Proteomics, 13(12):3612-3625.
<https://doi.org/10.1074/mcp.m114.038851>

6. Annexes

Annexe 1 : données de calibration des valeurs de référence de concentrations peptidiques.

Données individuelles de concentrations enregistrées sur gammarets engagés sur le réseau de surveillance (campagne AFB-48), pour les 25 peptides quantifiés dans plus de 70% des échantillons.





Annexe 2 : Liste et détail des annotations fonctionnelles des protéines recherchées pour chaque espèce d'étude.

Palaemon serratus* et *Palaemon longirostris

Organisme/Organe	ProteineID (<i>P. serratus</i>)	ProteineID (<i>P. longirostris</i>)	Annotation nr/IP	Annotations KEGG Pathway/GO term
Palaemon / Hépatopancréas	Pserratus_DN9025_c1_g3_i1.p1	Plongirostris_DN8009_c0_g3_i1.p1	Aldehyde oxidase/xanthine dehydrogenase-like	Drug metabolism - other enzymes;
	Pserratus_DN7572_c0_g1_i1.p1	Plongirostris_DN7183_c0_g1_i1.p1	Aldehyde oxidase/xanthine dehydrogenase-like	Drug metabolism - cytochrome P450
	Pserratus_DN10656_c0_g1_i1.p1	Plongirostris_DN10031_c1_g1_i1.p1	Aldehyde oxidase/xanthine dehydrogenase-like	
	Pserratus_DN1498_c0_g1_i4.p1	Plongirostris_DN78_c0_g1_i1.p1	Beta-glucuronidase-like	
	Pserratus_DN6358_c1_g1_i1.p1	Plongirostris_DN2815_c0_g1_i1.p1	Beta-glucuronidase-like	Drug metabolism - other enzymes
	Pserratus_DN70045_c0_g1_i1.p2	Plongirostris_DN1377_c0_g1_i2.p2	Nucleoside diphosphate kinase	
	Pserratus_DN8839_c0_g1_i3.p1	Plongirostris_DN2981_c1_g1_i1.p1	Glutathione S-transferase-like	Metabolism of xenobiotics by cytochrome P450; Drug metabolism - other enzymes; Drug metabolism - cytochrome P450
	Pserratus_DN28968_c0_g1_i1.p1	Plongirostris_DN569_c1_g1_i5.p6	hematopoietic prostaglandin D synthase-like	
	Pserratus_DN4311_c0_g1_i1.p1	Plongirostris_DN31724_c0_g1_i1.p1	Aldehyde dehydrogenase, dimeric NADP-preferring-like	Metabolism of xenobiotics by cytochrome P450;
	Pserratus_DN85230_c0_g1_i1.p1	Plongirostris_DN3358_c0_g1_i2.p1	Aldehyde dehydrogenase, dimeric NADP-preferring-like	Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation
Palaemon / Hémolymphhe	Pserratus_DN1513_c0_g1_i1.p1	Plongirostris_DN116115_c0_g1_i1.p1	Macroglobulin-like	
	Pserratus_DN1380_c0_g1_i5.p1	Plongirostris_DN35363_c0_g1_i1.p1	Alpha-2-macroglobulin-like	Complement and coagulation cascades
	Pserratus_DN2620_c0_g1_i1.p1	Plongirostris_DN11621_c0_g1_i1.p1	Macroglobulin-like	
	Pserratus_DN3222_c0_g1_i1.p1	/	gamma-glutamyltransferase-like	
	Pserratus_DN1128_c0_g1_i9.p1	Plongirostris_DN971_c0_g1_i1.p1	Actin	
	Pserratus_DN7793_c0_g1_i1.p1	Plongirostris_DN30538_c0_g1_i1.p1	Actin	
	Pserratus_DN23276_c0_g1_i2.p1	Plongirostris_DN2242_c0_g2_i4.p1	Actin	Platelet activation; Leukocyte transendothelial migration
	Pserratus_DN272_c0_g3_i1.p1	Plongirostris_DN3169_c0_g1_i1.p1	Actin	
	Pserratus_DN35_c0_g1_i1.p1	Plongirostris_DN971_c0_g2_i3.p1	Actin	
	Pserratus_DN2591_c0_g1_i2.p1	Plongirostris_DN4802_c0_g1_i5.p1	Integrin beta-PS-like	
	Pserratus_DN11247_c0_g2_i1.p1	/	actin-like	
	Pserratus_DN309_c0_g1_i11.p1	Plongirostris_DN7554_c1_g1_i2.p1	Hemocytin-like	Platelet activation; IL-17 signaling pathway; Complement and coagulation cascades
	Pserratus_DN1641_c0_g1_i2.p1	/	lectin-like	Hematopoietic cell lineage
Pserratus_DN556_c0_g2_i1.p1	/	alpha-actinin sarcomeric-like	Leukocyte transendothelial migration	
Pserratus_DN7715_c0_g1_i3.p1	/	cofilin/actin-depolymerizing factor homolog	Fc gamma R-mediated phagocytosis	
Palaemon / Branchies	Pserratus_DN911_c0_g3_i1.p1	Plongirostris_DN279_c0_g2_i1.p1	Calcium-transporting ATPase	
	Pserratus_DN437_c0_g1_i4.p1	Plongirostris_DN74_c0_g1_i2.p1	sarcoplasmic/endoplasmic reticulum type-like	
	Pserratus_DN2880_c0_g2_i1.p1	Plongirostris_DN447_c1_g1_i1.p1	Sodium- and chloride-dependent GABA transporter-like	Osmoregulation ; water homeostasis
	Pserratus_DN51473_c0_g1_i1.p1	Plongirostris_DN433_c0_g1_i1.p1	Ubiquitin-protein ligase NEDD4-like	
	Pserratus_DN94_c0_g1_i1.p1	Plongirostris_DN178_c0_g2_i2.p1	Stathmin-like	Osmoregulation ; water homeostasis
	Pserratus_DN4233_c1_g2_i2.p1	Plongirostris_DN1174_c0_g2_i1.p1	Clathrin light chain-like	Osmoregulation ; water homeostasis
	Pserratus_DN11973_c0_g1_i1.p1	Plongirostris_DN102707_c0_g1_i1.p1	Glucosylceramidase-like	Osmoregulation ; water homeostasis
	Pserratus_DN1804_c1_g1_i10.p1	Plongirostris_DN1062_c0_g1_i1.p1	T-complex protein 1 subunit gamma-like	Osmoregulation ; water homeostasis
	Pserratus_DN1338_c0_g1_i4.p1	Plongirostris_DN227_c1_g2_i2.p1	Serine/threonine-protein kinase OSR1-like	Osmoregulation ; water homeostasis
	Pserratus_DN457_c0_g1_i4.p1	/	AMP-dependent protein kinase type II regulatory subunit-like	Osmoregulation ; water homeostasis
		rab11-family interacting protein-like	Osmoregulation ; water homeostasis	

Gammarus fossarum

Organisme/Organe	ProteineID	Annotation nr/IP	Annotations KEGG Pathway/GO term
G.fossB / Caecum	GFBM_DN110488_c0_g1_i1::g.154816::m.154816	Aldehyde dehydrogenase-like	Chloroalkane and chloroalkene degradation
	GFBM_DN75144_c0_g1_i1::g.851303::m.851303		
	GFBM_DN113180_c0_g1_i2::g.521934::m.521934	Aldehyde dehydrogenase-like	Caprolactam degradation
	GFBM_DN119980_c5_g1_i1::g.339948::m.339948	Trifunctional enzyme subunit alpha, mitochondrial-like	Caprolactam degradation; Aminobenzoate degradation
	GFBM_DN118844_c0_g1_i2::g.463279::m.463279	Xanthine dehydrogenase/oxidase-like	
	GFBM_DN101596_c0_g1_i2::g.253580::m.253580	Carboxylesterase	Drug metabolism - other enzymes
	GFBM_DN118151_c0_g2_i1::g.671259::m.671259	Xanthine dehydrogenase-like	
	GFBM_DN109777_c0_g1_i2::g.746927::m.746927	Glutathione S-transferase 1-like	Metabolism of xenobiotics by cytochrome P450;
	GFBM_DN117325_c0_g1_i11::g.9889::m.9889	UDP-glucuronosyl/UDP-glucosyltransferase	Drug metabolism - other enzymes;
	GFBM_DN117330_c4_g1_i1::g.14917::m.14917	UDP-glucuronosyl/UDP-glucosyltransferase	Drug metabolism - cytochrome P450
G.fossB / Hémolymphhe	GFBM_DN120882_c4_g1_i7::g.302215::m.302215	Alpha-2-macroglobulin-like	
	GFBM_DN106701_c0_g1_i1::g.516543::m.516543	Alpha-2-macroglobulin-like	Complement and coagulation cascades
	GFBM_DN100087_c0_g1_i2::g.764399::m.764399	Alpha-2-macroglobulin-like	
	GFBM_DN118952_c0_g3_i1::g.476665::m.476665	Actin	
	GFBM_DN112551_c1_g2_i2::g.573473::m.573473	Actin	
	GFBM_DN120284_c2_g3_i2::g.215083::m.215083	Actin	Platelet activation; Leukocyte transendothelial migration
	GFBM_DN118080_c0_g1_i1::g.1971::m.1971	Actin	
	GFBM_DN116398_c2_g3_i2::g.46357::m.46357	Actin	
	GFBM_DN88039_c0_g1_i1::g.819692::m.819692	Actin	
	GFBM_DN120348_c0_g2_i1::g.39721::m.39721	Hemocytin-like	Platelet activation; Complement and coagulation cascades
G.fossB / Branchies	GFBM_DN121054_c5_g1_i1::g.721459::m.721459	Beta-1,3-glucan-binding protein precursor	
	GFBM_DN82398_c0_g1_i1::g.80593::m.80593	T-complex protein 1 subunit gamma-like	Osmoregulation ; water homeostasis
	GFBM_DN117888_c0_g1_i2::g.153913::m.153913	T-complex protein 1 subunit gamma-like	Osmoregulation ; water homeostasis
	GFBM_DN82432_c0_g1_i3::g.102115::m.102115	Clathrin light chain-like	Osmoregulation ; water homeostasis
	GFBM_DN118869_c5_g2_i1::g.45372::m.45372	Adaptor protein AP-2 complex subunit alpha-like	Osmoregulation ; water homeostasis
	GFBM_DN113326_c3_g2_i2::g.224363::m.224363	Protein kinase-like	Osmoregulation ; water homeostasis
	GFBM_DN113594_c4_g2_i3::g.89565::m.89565	AMP-dependent protein kinase catalytic subunit 1-like	Osmoregulation ; water homeostasis
GFBM_DN120756_c0_g3_i1::g.43521::m.43521	Serine/Threonine-protein kinase OSR1-like		

Dreissena polymorpha et Dreissena r. bugensis

Organisme/Organe	ProteineID (D. polymorpha)	ProteineID (D. r. bugensis)	Annotation nr/IP	Annotations KEGG Pathway/GO term
Dreissena / Glandes dig	Dp_LOC105340590.2.2:g.453987:m.453987	Db_LOC105340590.1.3:g.5430:m.5430	Dihydropyrimidinase-like	Drug metabolism - other enzymes
	Dp_LOC105340590.1.2:g.22680:m.22680	Db_LOC105340590.3.3:g.256819:m.256819	Dihydropyrimidinase-like	Drug metabolism - other enzymes
	Dp_LOC102809579.3.9:g.279328:m.279328	Db_LOC106054059:g.19801:m.19801	Glutathione S-transferase-like	Metabolism of xenobiotics by cytochrome P450;
	Dp_LOC105333880.3.8:g.181063:m.181063	Db_LOC105333880.7.7:g.595487:m.595487	Glutathione S-transferase-like	Drug metabolism - other enzymes; Drug metabolism - cytochrome P450
	Dp_LOC105340488.3.4:g.419498:m.419498	Db_LOC105340488.4.7:g.351317:m.351317	Glutathione S-transferase-like	Drug metabolism - cytochrome P450
	Dp_LOTGIDRAFT_203490:g.363359:m.363359	Db_LOTGIDRAFT_203490:g.417465:m.417465	Hydroxyacyl-CoA dehydrogenase, mitochondrial-like	Caprolactam degradation; Aminobenzoate degradation
	Dp_LOC105342066:g.23160:m.23160	Db_LOC101856538:g.621044:m.621044	Trifunctional enzyme subunit alpha, mitochondrial-like	Caprolactam degradation; Aminobenzoate degradation
	Dp_LOC101869038:g.87598:m.87598	Db_LOC101869038.2.2:g.544162:m.544162	Aldehyde dehydrogenase-like	Metabolism of xenobiotics by cytochrome P450;
	Dp_LOC106156945:g.215850:m.215850	Db_LOC106156945.1.2:g.76385:m.76385	Aldehyde dehydrogenase family 16 member A1-like	Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation
	Dp_LOC106174241.1.2:g.11679:m.11679	Db_LOC106174241.2.2:g.445665:m.445665	Alcohol dehydrogenase class-3-like	Naphthalene degradation; Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation
Dreissena / Hémolymphes	TRINITY_DN215561_c0_g3_i6	Db_LOC106612870:g.588130:m.588130	Macroglobulin-like	Complement and coagulation cascades
	TRINITY_DN208778_c4_g2_i11	Db_LOC105333243.1.2:g.190723:m.190723	Alpha-2-macroglobulin-like	
	TRINITY_DN194686_c0_g4_i1	Db_LOC106612870:g.588130:m.588130	Macroglobulin-like	
	TRINITY_DN220774_c7_g4_i2	/	/	Complement C3-like
	TRINITY_DN205441_c5_g15_i1	Db_LOC105329085.2.3:g.211635:m.211635	High mobility group protein B2-like	innate immune response
	TRINITY_DN212368_c1_g6_i6	/	/	Alpha-2-macroglobulin-like
	TRINITY_DN199594_c10_g9_i1	Db_LOC106051403.4.4:g.354549:m.354549	Concanavalin A-like lectin/glucanase	regulation of innate immune response
	TRINITY_DN192719_c2_g5_i2	Db_LOC105334536.3.3:g.535212:m.535212	Sarcoplasmic calcium-binding protein	positive regulation of inflammatory response
	TRINITY_DN207483_c0_g4_i1	Db_LOC105341865.1.4:g.82954:m.82954	ATP-dependent RNA helicase DDX3Y-like	RIG-I-like receptor signaling pathway
	TRINITY_DN211288_c0_g1_i8	/	/	Complement component C4-like
TRINITY_DN205477_c0_g1_i1	/	/	Alpha-macroglobulin, receptor-binding	regulation of inflammatory response
Dreissena / Branchies	TRINITY_DN207134_c2_g1_i3:g.873451:m.873451	Db_LOC101846151.1.4:g.47625:m.47625	Sarcoplasmic/endoplasmic reticulum calcium ATPase-like	Osmoregulation ; water homeostasis
	TRINITY_DN203049_c2_g3_i3:g.953583:m.953583	Db_LOC106173262.1.5:g.107302:m.107302	Annexin	Osmoregulation ; water homeostasis
	TRINITY_DN214176_c2_g3_i1:g.1091047:m.1091047	Db_LOC105342582:g.481244:m.481244	T-complex protein 1 subunit gamma-like	Osmoregulation ; water homeostasis
	TRINITY_DN211570_c1_g2_i2:g.880414:m.880414	Db_LOTGIDRAFT_203023.1.2:g.32520:m.32520	Adaptor protein AP-2 complex subunit alpha-like	Osmoregulation ; water homeostasis
	TRINITY_DN213857_c0_g2_i2:g.554529:m.554529	Db_1433Z.1.2:g.302511:m.302511	T4-3-3-like protein	Osmoregulation ; water homeostasis
	TRINITY_DN195056_c0_g1_i1:g.126724:m.126724	Db_LOC101855225.1.2:g.186247:m.186247	Wolfamin-like	Osmoregulation ; water homeostasis
	TRINITY_DN181093_c1_g2_i2:g.384799:m.384799	Db_LOC106179046:g.214164:m.214164	Clathrin light chain-like	Osmoregulation ; water homeostasis
	TRINITY_DN218026_c1_g3_i6:g.794724:m.794724	Db_KAPR.1.2:g.486098:m.486098	cAMP-dependent protein kinase regulatory subunit-like	Osmoregulation ; water homeostasis
	TRINITY_DN199632_c0_g1_i1:g.360986:m.360986	Db_LOC106874074.1.2:g.12774:m.12774	cAMP-dependent protein kinase catalytic subunit-like	Osmoregulation ; water homeostasis
	TRINITY_DN206276_c2_g1_i1:g.1110224:m.1110224	Db_LOC105342800:g.349076:m.349076	cAMP-dependent protein kinase type II regulatory subunit-like	Osmoregulation ; water homeostasis

Gasterosteus aculeatus

Organisme/Organe	ProteineID	Annotation nr/IP	Annotations KEGG Pathway/GO term
Gasterosteus / Foie	Gasterosteus_DN13358_c0_g1_i1.p1	Trifunctional enzyme subunit alpha, mitochondrial-like	Caprolactam degradation; Aminobenzoate degradation
	Gasterosteus_DN1722_c0_g2_i1.p1	Glutaryl-CoA dehydrogenase	Xenobiotics biodegradation and metabolism
	Gasterosteus_DN183_c1_g1_i11.p1	Xanthine dehydrogenase/oxidase-like	Drug metabolism - other enzymes
	Gasterosteus_DN2314_c0_g1_i1.p1	Aldehyde dehydrogenase, mitochondrial-like	Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450
	Gasterosteus_DN1345_c0_g1_i10.p2	Cytochrome P450-like	Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450
Gasterosteus / Sang	Gasterosteus_DN14489_c0_g2_i2.p1	Complement C1q-like	Complement and coagulation cascades
	Gasterosteus_DN12002_c0_g1_i1.p1	NACHT, LRR, and PYD domains-containing protein-like	innate immune response
	Gasterosteus_DN18312_c0_g2_i1.p1	Eukaryotic translation initiation factor-like	-
	Gasterosteus_DN2813_c0_g1_i1.p1	Y-box transcription factor-like	-
	Gasterosteus_DN4514_c0_g1_i1.p1	Proteasome subunit alpha type-5	-
Gasterosteus / Branchies	Gasterosteus_DN11408_c0_g1_i1.p1	Keratin, type I	water homeostasis; osmoregulation
	Gasterosteus_DN10771_c0_g1_i1.p1	Keratin, type I	
	Gasterosteus_DN1102_c0_g1_i1.p1	Keratin, type I	
	Gasterosteus_DN7304_c0_g1_i1.p1	T-complex protein 1 subunit gamma-like	water homeostasis; osmoregulation
	Gasterosteus_DN729_c1_g2_i1.p1	arcoplasmic/endoplasmic reticulum calcium ATPase-like	water homeostasis; osmoregulation
	Gasterosteus_DN117_c1_g1_i1.p1	T4-3-3-like protein	water homeostasis; osmoregulation
	Gasterosteus_DN22970_c0_g2_i1.p1	Clathrin light chain A-like	Osmoregulation ; water homeostasis
Gasterosteus_DN561_c0_g2_i10.p1	Clathrin light chain B-like	Osmoregulation; water homeostasis	

Annexe 3 : Exemple de tableau résultats des analyses (d)MRM pour la branchie de *D. r. bugensis*.

19/08/2021						
	Peptide	RT shotgun	RT MRM	Intensité	Nb transition	Commentaires
MRM_1	GAPEGLLDR	45,717	14,4	7000	3	
	IDQSILTGESVSILK	#N/A		/		
	GAGTDEQAIIGIISQR	95,467	22,7	600	11	
	LQTQGPPEAVIDR	48,867	11,7	2200	4	
	GVSDLAQHLYLR	72,817	18,3	4500	11	
	SITGVHQWPYR	46,233		/		
	LLQNYPPPEPAVK	57,45		/		pic à 13,9 mais trop faible
	SGSAVQFQLLHKS	51,7		/		pic à 15,5 mais trop faible
	NLLSVAYK	60,233	15,9	25000	6	
	DSTLIMQLLR	111,283	28,1	23000	8	
	DNLTLLWK	#N/A	17,1	270	4	
	ETASAVLEAALEFASR	#N/A		/		
	IGENISENEFVK	63,117		/		
	QGEPEGEDFFIILEGSAAVLQR	120,083		/		20,9 pic large peut-être
	HNVQQLLK	23,833	8,4	150	5	très faible intensité
	LTVADALEPINFEDGQEIVR	#N/A		/		
	TMAALSK	#N/A		/		
	SILLQLK	67,85	18,4	2200	5	ou 11,7
	WDHPGQNTAGLDDFER	59,2		/		
FSEPHSR	9,417		/			
LAVLDVQAFER	93,533	22,9	500	6		
MNIADALYSK	73,633	18	400	5		
MRM_2	NILFSGTNIASGK	72,633	18,2	450	4	
	VGEATETALSVLVEK	90,35		/		
	EFTLEFSR	78,633	18,9	230	4	
	GIVIGTGLNTEIGK	#N/A		/		
	INFFNTEK	#N/A	16,3	320	3	autre pic à 13,1-330-4
	VTTIHSTTLR	#N/A	6,9	240	5	
	SASEMVLADDNFATIVAAVEEGR	#N/A		/		
	VIVITGDNK	35,267	9,5	1000	6	
	LLLALVGQ	110,25	25,2	300	6	
	GAGTQDSTLIR	30,717	8,3	540	4	
	MGTDESALFTVLAVR	#N/A	27,1	230	5	
	VQIMNLFK	#N/A	22,9	1800	6	
	ATFEEYVK	50,017	12,3	660	3	
	DILSSIDR	74,55	17	480	2	
	EGFMALVMSAK	#N/A	24,8	1200	6	
MALEDLLETLLK	113,65	30,5	900	6		
NVMQEPFLVPGGAAEMALAEQELNEK	#N/A		/			
MRM_3	EINVQHPAAK	17,217	7,8	420	4	
	TAIETAILLR	107,867	27,2	3000	6	
	ALEVIPR	48,083	12,1	3500	5	
	MEEEFIEK	45,35	10,9	444	4	
	EMAEGLSVDIPK	75,083	17,4	270	4	
	GLAVFISDIR	93,05	22,8	1000	5	
	VGGYILGEFGNLIAGDPR	111,95	29,6	600	6	que 2 transitions intenses
	ATSPIDHEQVK	17,75		/		
	LIGFSVSVLEK	95,483	20,3	1500	3	
	LLVSGDTIDSVK	65,417	15,5	550	3	
	NSDVELQQR	#N/A		/		
	YITPAEMDANTFFTR	98,633		/		
	LTNGMPVSIAGSGFQK	75,083		/		
	STAYEAAGWFDK	#N/A		/		
	SENIMEIDDLIK	#N/A		/		
	TLGTGSFGR	36,85	9,2	1800	6	
	GPGDTSNFDDYEEELR	#N/A		/		
	NINEYEEQLVTVFGSK	#N/A	25,4	150	4	très faible intensité
	AAISDLR	29,95	7,5	1200	3	

Annexe 4 : Bilan de la recherche de peptides sur organe chez les espèces de Palaemon, Dreissena et Gasterosteus

Palaemon serratus

<i>P. serratus</i> - Hépatopancréas						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Aldehyde oxidase/xanthine dehydrogenase-like	Pserratus_DN9025_c1_g3_i1.p1	0	1	9	10
	Aldehyde oxidase/xanthine dehydrogenase-like	Pserratus_DN7572_c0_g1_i1.p1	0	0	10	10
	Aldehyde oxidase/xanthine dehydrogenase-like	Pserratus_DN10656_c0_g1_i1.p1	0	2	7	9
Drug metabolism - other enzymes	Beta-glucuronidase-like	Pserratus_DN1498_c0_g1_i4.p1	1	1	7	9
	Beta-glucuronidase-like	Pserratus_DN6358_c1_g1_i1.p1	0	1	6	7
	Nucleoside diphosphate kinase	Pserratus_DN70045_c0_g1_i1.p2	1	0	1	2
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Glutathione S-transferase-like	Pserratus_DN8839_c0_g1_i3.p1	7	0	1	8
	hematopoietic prostaglandin D synthase-like	Pserratus_DN28968_c0_g1_i1.p1	5	2	1	8
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation	Aldehyde dehydrogenase, dimeric NADP-preferring-like	Pserratus_DN4311_c0_g1_i1.p1	0	3	5	8
	Aldehyde dehydrogenase, dimeric NADP-preferring-like	Pserratus_DN85230_c0_g1_i1.p1	9	1	1	11
Total			23	11	48	82

<i>P. serratus</i> - Branchies						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
water homeostasis; osmoregulation	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type-like	Pserratus_DN911_c0_g3_i1.p1	2	4	3	9
	Sodium- and chloride-dependent GABA transporter-like	Pserratus_DN437_c0_g1_i4.p1	0	2	1	3
	E3 ubiquitin-protein ligase NEDD4-like	Pserratus_DN2880_c0_g2_i1.p1	0	2	4	6
	Stathmin-like	Pserratus_DN51473_c0_g1_i1.p1	1	2	3	6
	Clathrin light chain-like	Pserratus_DN94_c0_g1_i1.p1	0	0	1	1
	Glucosylceramidase-like	Pserratus_DN4233_c1_g2_i2.p1	1	1	6	8
	T-complex protein 1 subunit gamma-like	Pserratus_DN111973_c0_g1_i1.p1	6	1	3	10
	cAMP-dependent protein kinase type II regulatory subunit-like	Pserratus_DN1338_c0_g1_i4.p1	3	1	2	6
	Serine/threonine-protein kinase OSR1-like	Pserratus_DN1804_c1_g1_i10.p1	2	2	0	4
rab11-family interacting protein-like	Pserratus_DN457_c0_g1_i4.p1	0	0	1	1	
Total			15	15	24	54

<i>P. serratus</i> - Hémolymphe						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Complement and coagulation cascades	Macroglobulin-like	Pserratus_DN1513_c0_g1_i1.p1	5 dont 1 doublon avec prot 1380 et 2620	1	4	10
	Alpha-2-macroglobulin-like	Pserratus_DN1380_c0_g1_i5.p1	5 dont 1 doublon avec prot 1513 et 2620	2	3	10
	Macroglobulin-like	Pserratus_DN2620_c0_g1_i1.p1	1 dont 1 doublon avec prot 1513 et 1380	4	4	9
	gamma-glutamyltransferase-like	Pserratus_DN3222_c0_g1_i1.p1	0	2	8	10
Platelet activation; IL-17 signaling pathway; Complement and coagulation cascades	Hemoctin-like	Pserratus_DN309_c0_g1_i1.p1	2	3	5	10
Platelet activation; Leukocyte transendothelial migration	Actin	Pserratus_DN1128_c0_g1_i3.p1	4 dont 4 doublons avec prot 35, 23276 et 7793	1 dont 1 doublon avec prot 35 et 23276	5 dont 3 doublons avec prot 11247, 23276, 35, 7793 et 272	10
	Actin	Pserratus_DN7793_c0_g1_i1.p1	2 dont 1 doublon avec prot 1128 et 35	0	3 dont 3 doublons avec prot 11247, 23276, 35 et 1128	5
	Actin	Pserratus_DN23276_c0_g1_i2.p1	2 dont 2 doublons avec prot 1128 et 35	3 dont 1 doublon avec prot 1128 et 35	5 dont 3 doublons avec prot 35, 272, 11247, 7793 et 1128	10
	Actin	Pserratus_DN272_c0_g3_i1.p1	0	1	2 dont 2 doublons avec prot 11247, 23276, 1128 et 35	3
	Actin	Pserratus_DN35_c0_g1_i1.p1	4 dont 4 doublons avec prot 1128, 23276 et 7793	1 dont 1 doublon avec prot 1128 et 23276	5 dont 5 doublons avec prot 1128, 23276, 7793, 11247 et 272	10
	Integrin beta-PS-like	Pserratus_DN2591_c0_g1_i2.p1	2	5	1	8
	actin-like	Pserratus_DN11247_c0_g2_i1.p1	0	0	2 dont 2 doublons avec prot 272 et 23276	2
Fc gamma R-mediated phagocytosis	cofilin/actin-depolymerizing factor homolog	Pserratus_DN7715_c0_g1_i3.p1	1	2	1	4
Leukocyte transendothelial migration	alpha-actinin sarcomeric-like	Pserratus_DN556_c0_g2_i1.p1	2	1	7	10
Hematopoietic cell lineage	Pserratus_DN1641_c0_g1_i2.p1	Pserratus_DN1641_c0_g1_i2.p1	0	3	2	5
Total			20	27	46	93

Palaemon longirostris

<i>P. longirostris</i> - Hépatopancréas						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Aldehyde oxidase/xanthine dehydrogenase-like	Plongirostris_DN8009_c0_g3_i1.p1	0	0	8	8
	Aldehyde oxidase/xanthine dehydrogenase-like	Plongirostris_DN7183_c0_g1_i1.p1	1	0	8	9
Drug metabolism - other enzymes	Aldehyde oxidase/xanthine dehydrogenase-like	Plongirostris_DN10031_c1_g1_i1.p1	0	0	9	9
	Beta-glucuronidase-like	Plongirostris_DN78_c0_g1_i1.p1	0	0	9 dont 1 doublon	9
	Beta-glucuronidase-like	Plongirostris_DN2815_c0_g1_i1.p1	0	0	8 dont 1 doublon	8
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Nucleoside diphosphate kinase	Plongirostris_DN1377_c0_g1_i2.p2	2	0	1	3
	Glutathione S-transferase-like	Plongirostris_DN2981_c1_g1_i1.p1	7	0	0	7
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450	hematopoietic prostaglandin D synthase-like	Plongirostris_DN569_c1_g1_i5.p6	4	1	2	7
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation	Aldehyde dehydrogenase, dimeric NADP-prefering-like	Plongirostris_DN31724_c0_g1_i1.p1	2	4	0	6
	Aldehyde dehydrogenase, dimeric NADP-prefering-like	Plongirostris_DN3358_c0_g1_i2.p1	3	4	1	8
Total			19	9	46	74

<i>P. longirostris</i> - Branchies						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
water homeostasis; osmoregulation	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type-like	Plongirostris_DN279_c0_g2_i1.p1	1	2	6	9
	Sodium- and chloride-dependent GABA transporter-like	Plongirostris_DN74_c0_g1_i2.p1	0	1	2	3
	E3 ubiquitin-protein ligase NEDD4-like	Plongirostris_DN447_c1_g1_i1.p1	0	0	4	4
	Stathmin-like	Plongirostris_DN433_c0_g1_i1.p1	0	1	5	6
	Clathrin light chain-like	Plongirostris_DN178_c0_g1_i2.p1	0	0	1	1
	Glucosylceramidase-like	Plongirostris_DN1174_c0_g2_i1.p1	2	0	6	8
	T-complex protein 1 subunit gamma-like	Plongirostris_DN102707_c0_g1_i1.p1	2	2	5	9
	Serine/threonine-protein kinase OSR1-like	Plongirostris_DN1062_c0_g1_i1.p1	1	0	2	3
cAMP-dependent protein kinase type II regulatory subunit-like	Plongirostris_DN227_c1_g2_i2.p1	0	2	4	6	
Total			6	8	35	49

<i>P. longirostris</i> - Hémolymphhe						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Complement and coagulation cascades	Macroglobulin-like	Plongirostris_DN116115_c0_g1_i1.p1	0	1	7	8
	Alpha-2-macroglobulin-like	Plongirostris_DN33363_c0_g3_i1.p1	0	2	3	5
	Macroglobulin-like	Plongirostris_DN11621_c0_g1_i1.p1	1	1	6	8
Platelet activation; Leukocyte transendothelial migration	Actin	Plongirostris_DN971_c0_g1_i1.p1	0	5 doublons	5 doublons	10
	Actin	Plongirostris_DN30538_c0_g1_i1.p1	0	1 doublon	1	2
	Actin	Plongirostris_DN2242_c0_g2_i4.p1	0	3 doublons	3 dont 1 doublon	6
	Actin	Plongirostris_DN3169_c0_g1_i1.p1	0	1 doublon	1	2
	Actin	Plongirostris_DN971_c0_g2_i3.p1	0	4 doublons	6 dont 5 doublons	10
Platelet activation; IL-17 signaling pathway; Complement and coagulation cascades	Integrin beta-PS-like	Plongirostris_DN4802_c0_g1_i5.p1	0	1	7	8
	Hemocytin-like	Plongirostris_DN7554_c1_g1_i2.p1	0	1	7	8
Total			1	12	40	53

Dreissena polymorpha

<i>D. polymorpha</i> - Glandes digestives						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Drug metabolism - other enzymes	Dihydropyrimidinase-like	Dp_LOC105340590.2.2.:g.453987.:m.453987	3 dont 3 doublons	3	4	10
	Dihydropyrimidinase-like	Dp_LOC105340590.1.2.:g.22680.:m.22680	3 dont 3 doublons	4	3	1
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Glutathione S-transferase-like	Dp_LOC102609579.3.3.:g.279328.:m.279328	6	0	0	6
	Glutathione S-transferase-like	Dp_LOC105333880.3.8.:g.181063.:m.181063	4	4	1	9
	Glutathione S-transferase-like	Dp_LOC105340488.3.4.:g.419498.:m.419498	6	0	0	6
Caprolactam degradation; Aminobenzoate degradation	Hydroxyacyl-CoA dehydrogenase, mitochondrial-like	Dp_LOC106174241.1.2.:g.11679.:m.11679	2	3	5	9
	Trifunctional enzyme subunit alpha, mitochondrial-like	Dp_LOC105342066.:g.23160.:m.23160	4	2	2	8
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation	Aldehyde dehydrogenase-like	Dp_LOC101869038.:g.87598.:m.87598	7	2	1	10
	Aldehyde dehydrogenase family 16 member A1-like	Dp_LOC106156945.:g.215850.:m.215850	1	4	2	7
Naphthalene degradation; Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation	Alcohol dehydrogenase class-3-like	Dp_LOC106174241.1.2.:g.11679.:m.11679	0	2	2	4
Total			33	21	16	70

<i>D. polymorpha</i> - Branchies						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
water homeostasis; osmoregulation	Sarcoplasmic/endoplasmic reticulum calcium ATPase-like	TRINITY_DN207134_c2_g1_i3::g.873451:m.873451	3	3	4	10
	Annexin	TRINITY_DN203049_c2_g3_i3::g.953583:m.953583	7	3	0	10
	T-complex protein 1 subunit gamma-like	TRINITY_DN214176_c2_g3_i1::g.1091047:m.1091047	8	2	0	10
	Adaptor protein AP-2 complex subunit alpha-like	TRINITY_DN211570_c1_g2_i2::g.880414:m.880414	1	6	3	10
	14-3-3-like protein	TRINITY_DN213857_c0_g2_i2::g.554529:m.554529	2	1	0	3
	Wolframin-like	TRINITY_DN195056_c0_g1_i1::g.126724:m.126724	0	0	6	6
	Clathrin light chain-like	TRINITY_DN181093_c1_g2_i2::g.384799:m.384799	1	0	0	1
	cAMP-dependent protein kinase regulatory subunit-like	TRINITY_DN218026_c1_g3_i6::g.794724:m.794724	1	1	2	5
	cAMP-dependent protein kinase catalytic subunit-like	TRINITY_DN199632_c0_g1_i1::g.360986:m.360986	0	2	2	4
cAMP-dependent protein kinase type II regulatory subunit-like	TRINITY_DN206276_c2_g1_i1::g.1110224:m.1110224	4	0	0	4	
Total			27	18	17	63

<i>D. polymorpha</i> - Hémolymphe						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Complement and coagulation cascades	Macroglobulin-like	TRINITY_DN215561_c0_g3_i6	0	7	3	10
	Alpha-2-macroglobulin-like	TRINITY_DN208778_c4_g2_i1	0	6	4	10
	Complement C3-like	TRINITY_DN194686_c0_g4_i1	0	1	4	5
	Complement C3-like	TRINITY_DN220774_c7_g4_i2	2	3	2	7
innate immune response	High mobility group protein B2-like	TRINITY_DN205441_c5_g15_i1	1	3	1	5
	Alpha-2-macroglobulin-like	TRINITY_DN212368_c1_g6_i6	1	3	1	5
regulation of innate immune response	Concanavalin A-like lectin/glucanase	TRINITY_DN193954_c10_g9_i1	6	1	2	9
positive regulation of inflammatory response	Sarcoplasmic calcium-binding protein	TRINITY_DN192719_c2_g5_i2	1	3	4	8
RIG-I-like receptor signaling pathway	ATP-dependent RNA helicase DDX3Y-like	TRINITY_DN207483_c0_g4_i1	1	7	2	10
complement activation	Complement component C4-like	TRINITY_DN211288_c0_g1_i8	1	3	6	10
regulation of inflammatory response	Alpha-macroglobulin, receptor-binding	TRINITY_DN205477_c0_g1_i1	0	3	1	4
Total			13	40	30	63

Dreissena r. bugensis

<i>D. r. bugensis</i> - Glandes digestives						
Annotations KEGG Pathway /GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Drug metabolism - other enzymes	Dihydropyrimidinase-like	Db_LOC105340590.1.3::g.5430::m.5430	5 dont 3 doublons	0	5 dont 2 doublons	10
	Dihydropyrimidinase-like	Db_LOC105340590.3.3::g.256819::m.256819	4 dont 3 doublons	0	5 dont 2 doublons	9
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - other enzymes; Drug metabolism - cytochrome P450	Glutathione S-transferase-like	Db_LOC106054059::g.19801::m.19801	2	2	0	4
	Glutathione S-transferase-like	Db_LOC105333880.7.7::g.595487::m.595487	3	0	3	6
	Glutathione S-transferase-like	Db_LOC105340488.4.7::g.351317::m.351317	6	0	0	6
Caprolactam degradation; Aminobenzoate degradation	Hydroxyacyl-CoA dehydrogenase, mitochondrial-like	Db_LOC107GIDRAFT_203490::g.417465::m.417465	4	3	1	8
	Trifunctional enzyme subunit alpha, mitochondrial-like	Db_LOC101856538::g.621044::m.621044	5	3	0	8
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation	Aldehyde dehydrogenase-like	Db_LOC101869038.2.2::g.544162::m.544162	6	2	1	9
	Aldehyde dehydrogenase family 16 member A1-like	Db_LOC106156945.1.2::g.76385::m.76385	0	1	7	8
Naphthalene degradation; Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450; Chloroalkane and chloroalkene degradation	Alcohol dehydrogenase class-3-like	Db_LOC106174241.2.2::g.445665::m.445665	1	2	0	3
Total			34	13	20	67

<i>D. r. bugensis</i> - Branchies						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
water homeostasis; osmoregulation	Sarcoplasmic/endoplasmic reticulum calcium ATPase-like	Db_LOC101846151.1.4::g.47625::m.47625	3	3	4	10
	Annexin	Db_LOC106173262.1.5::g.107302::m.107302	3	4	2	9
	T-complex protein 1 subunit gamma-like	Db_LOC105342582::g.481244::m.481244	5	1	2	8
	Adaptor protein AP-2 complex subunit alpha-like	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	2	1	7	10
	14-3-3-like protein	Db_1433Z.1.2::g.302511::m.302511	2	1	0	3
	Wolframin-like	Db_LOC101855225.1.2::g.186247::m.186247	0	0	5	5
	Clathrin light chain-like	Db_LOC106179046::g.214164::m.214164	0	1	0	1
	cAMP-dependent protein kinase regulatory subunit-like	Db_KAPR.1.2::g.486098::m.486098	0	1	3	4
	cAMP-dependent protein kinase catalytic subunit-like	Db_LOC106874074.1.2::g.12774::m.12774	1	0	3	4
cAMP-dependent protein kinase type II regulatory subunit-like	Db_LOC105342800::g.349076::m.349076	3	0	1	4	
Total			19	12	27	58

<i>D. r. bugensis</i> - Hémolymphe						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Complement and coagulation cascades	Macroglobulin-like	Db_LOC106612870::g.588130::m.588130	2 dont 1 doublon	1	1	4
	Alpha-2-macroglobulin-like	Db_LOC105333243.1.2::g.190723::m.190723	1	0	1	2
	Macroglobulin-like	Db_LOC106612870::g.588130::m.588130	2 dont 1 doublon	0	2	4
innate immune response	High mobility group protein B2-like	Db_LOC105329085.2.3::g.211635::m.211635	0	1	4	5
regulation of innate immune response	Concanavalin A-like lectin/glucanase	Db_LOC106051403.4.4::g.354549::m.354549	1	4	4	9
positive regulation of inflammatory response	Sarcoplasmic calcium-binding protein	Db_LOC105334536.3.3::g.535212::m.535212	3	0	1	4
RIG-I-like receptor signaling pathway	ATP-dependent RNA helicase DDX3Y-like	Db_LOC105341865.1.4::g.82354::m.82354	0	1	9	10
Total			8	7	22	37

Gasterosteus aculeatus

<i>G. aculeatus</i> - Foie						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Caprolactam degradation; Aminobenzoate degradation	Trifunctional enzyme subunit alpha, mitochondrial-like	Gasterosteus_DN13358_c0_g1_i1.p	2	2	1	5
Xenobiotics biodegradation and metabolism	Glutaryl-CoA dehydrogenase	Gasterosteus_DN1722_c0_g2_i1.p1	1	6	2	9
Drug metabolism - other enzymes	Xanthine dehydrogenase/oxidase-like	Gasterosteus_DN183_c1_g1_i11.p1	1	1	0	2
Metabolism of xenobiotics by cytochrome P450; Drug metabolism - cytochrome P450	Aldehyde dehydrogenase, mitochondrial-like	Gasterosteus_DN2314_c0_g1_i1.p1	4	3	1	8
	Cytochrome P450-like	Gasterosteus_DN1345_c0_g1_i10.p	0	0	2	2
Total			8	12	6	26

<i>G. aculeatus</i> - Branchies							
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés	
water homeostasis; osmoregulation	Keratin, type I	Gasterosteus_DN11408_c0_g1_i1.p1	3	1	2	6	
		Gasterosteus_DN10771_c0_g1_i1.p1	2	2	0	4	
		Gasterosteus_DN1102_c0_g1_i1.p1	0	1	1	2	
		Gasterosteus_DN7304_c0_g1_i1.p1	4	3	2	9	
		Sarcoplasmic/endoplasmic reticulum calcium ATPase-like	Gasterosteus_DN729_c1_g2_i1.p1	5	1	0	6
		14-3-3-like protein	Gasterosteus_DN117_c1_g1_i1.p1	2	1	0	3
		Clathrin light chain A-like	Gasterosteus_DN22970_c0_g2_i1.p1	0	1	0	1
	Clathrin light chain B-like	Gasterosteus_DN561_c0_g2_i10.p1	0	1	0	1	
Total			16	11	5	32	

<i>G. aculeatus</i> - Sang						
Annotations KEGG Pathway/GO term	Annotation nr/IP	Protéine ID	Identifiés avec forte confiance	Identification à confirmer	Non détectés dans les deux essais	Total de peptides cherchés
Complement and coagulation cascades	Complement C1q-like	Gasterosteus_DN14489_c0_g2_i2.p1	0	2	1	3
Innate immune response	NACT, LRR, and PYD domains-containing protein-like	Gasterosteus_DN12002_c0_g1_i1.p1	0	1	1	2
-	Eukaryotic translation initiation factor-like	Gasterosteus_DN18312_c0_g2_i1.p1	1	1	0	2
-	Y-box transcription factor-like	Gasterosteus_DN2813_c0_g1_i1.p1	0	1	1	2
-	Proteasome subunit alpha type-5	Gasterosteus_DN4514_c0_g1_i1.p1	2	0	1	3
Total			3	5	4	12

Annexe 5: comparaison des peptides retrouvés entre espèces d'un même genre : *Paleamon* et *Dreissène*.

Un peptide de couleur verte signifie qu'il a été identifié avec une forte confiance. Un peptide de couleur orange signifie qu'il est identifié avec une moindre confiance et sera à confirmer. Un peptide de couleur rose signifie qu'il n'a pas été détecté. Les cases surlignées jaunes fluo correspondent aux peptides doublons entre protéines. Les mutations dans les peptides sont montrées par une lettre de couleur rouge. Les peptides surlignés en rouge correspondent à des peptides avec mutations entraînant la présence d'au moins un miss cleavage au sein de la séquence par une lysine K ou une arginine R, ou l'absence de K ou R en fin de séquence peptidique.

Annexe 5a : comparaison des peptides dans les branchies de *P. serratus* et *P. longirostris*.

<i>P. serratus</i> - branchie	peptide	<i>P. longirostris</i> - branchie	peptide
Pserratus_DN911_c0_g3_i1.p1	SASEMVLADDNFSSIVAAVEEGR	Plongirostris_DN279_c0_g2_i1.p1	SASEMVLADDNFSSIVAAVEEGR
Pserratus_DN911_c0_g3_i1.p1	EFTLEFSR	Plongirostris_DN279_c0_g2_i1.p1	EFTLEFSR
Pserratus_DN911_c0_g3_i1.p1	ALGIVIGTGLGTAIGK	Plongirostris_DN279_c0_g2_i1.p1	ALGIVIGTGLGTAIGK
Pserratus_DN911_c0_g3_i1.p1	IDQSILTGESVSVIK	Plongirostris_DN279_c0_g2_i1.p1	IDQSILTGESVSVIK
Pserratus_DN911_c0_g3_i1.p1	NILFSGTNVAAGK	Plongirostris_DN279_c0_g2_i1.p1	NILFSGTNVAAGK
Pserratus_DN911_c0_g3_i1.p1	ISFPVILDETFK	Plongirostris_DN279_c0_g2_i1.p1	ISFPVILDETFK
Pserratus_DN911_c0_g3_i1.p1	GAPESILDR	Plongirostris_DN279_c0_g2_i1.p1	GAPESILDR
Pserratus_DN911_c0_g3_i1.p1	EFDELSVAEQR	Plongirostris_DN279_c0_g2_i1.p1	EFDELSVAEQR
Pserratus_DN911_c0_g3_i1.p1	TGTLTTNQMSVSR	Plongirostris_DN279_c0_g2_i1.p1	TGTLTTNQMSVSR
Pserratus_DN111973_c0_g1_i1.p1	GVSDLAQHFLVK	Plongirostris_DN102707_c0_g1_i1.p1	GVSDLAQHFLVK
Pserratus_DN111973_c0_g1_i1.p1	LVLGGGAVEMALSHILQEK	Plongirostris_DN102707_c0_g1_i1.p1	LVLGGGAVEMALSHILQEK
Pserratus_DN111973_c0_g1_i1.p1	TAVETAVLLLR	Plongirostris_DN102707_c0_g1_i1.p1	TAVETAVLLLR
Pserratus_DN111973_c0_g1_i1.p1	MHPVVIIQAYR	Plongirostris_DN102707_c0_g1_i1.p1	MHPVVIIQAYR
Pserratus_DN111973_c0_g1_i1.p1	NLQDALGVAR	Plongirostris_DN102707_c0_g1_i1.p1	NLQDALGVAR
Pserratus_DN111973_c0_g1_i1.p1	QWPYSALGQALEVIPR	Plongirostris_DN102707_c0_g1_i1.p1	QWPYSALGQALEVIPR
Pserratus_DN111973_c0_g1_i1.p1	VQLQININAGK	Plongirostris_DN102707_c0_g1_i1.p1	VQLQININAGK
Pserratus_DN111973_c0_g1_i1.p1	MIADVIR		
Pserratus_DN111973_c0_g1_i1.p1	EITVQHPPAAK	Plongirostris_DN102707_c0_g1_i1.p1	EITVQHPPAAK
Pserratus_DN111973_c0_g1_i1.p1	TQDEEVGDGTTSSVILAGEMLAVAEQFLQHK	Plongirostris_DN102707_c0_g1_i1.p1	TQDEEVGDGTTSSVILAGEMLAVAEQFLQHK
Pserratus_DN51473_c0_g1_i1.p1	LSHEEAIR	Plongirostris_DN433_c0_g1_i1.p1	LSHEEAIR
Pserratus_DN51473_c0_g1_i1.p1	AYLNGLR	Plongirostris_DN433_c0_g1_i1.p1	AYLNGLR
Pserratus_DN51473_c0_g1_i1.p1	SLDEQTLLELR	Plongirostris_DN433_c0_g1_i1.p1	SLDEQTLLELR
Pserratus_DN51473_c0_g1_i1.p1	ITDHLNNVDGVR	Plongirostris_DN433_c0_g1_i1.p1	ITDHLNNVDGVR
Pserratus_DN51473_c0_g1_i1.p1	LASVNER	Plongirostris_DN433_c0_g1_i1.p1	LASVNER
Pserratus_DN51473_c0_g1_i1.p1	LVLAEER	Plongirostris_DN433_c0_g1_i1.p1	LVLAEER
Pserratus_DN1338_c0_g1_i4.p1	LNLADALVPK	Plongirostris_DN227_c1_g2_i2.p1	LNLADALVPK
Pserratus_DN1338_c0_g1_i4.p1	SLDNDQIGEVLDAMFER	Plongirostris_DN227_c1_g2_i2.p1	SLDNDQIGEVLDAMFER
Pserratus_DN1338_c0_g1_i4.p1	AATIQAQVTTGSLWAMDR	Plongirostris_DN227_c1_g2_i2.p1	AATIQAQVTTGSLWAMDR
Pserratus_DN1338_c0_g1_i4.p1	VAFLDVEAFER	Plongirostris_DN227_c1_g2_i2.p1	VAFLDVEAFER
Pserratus_DN1338_c0_g1_i4.p1	ANISDIR	Plongirostris_DN227_c1_g2_i2.p1	ANISDIR
Pserratus_DN1338_c0_g1_i4.p1	VTVGGYFGELALVTK	Plongirostris_DN227_c1_g2_i2.p1	VTVGGYFGELALVTK
Pserratus_DN4233_c1_g2_i2.p1	VVIILNR	Plongirostris_DN1174_c0_g2_i1.p1	VVIILNR
Pserratus_DN4233_c1_g2_i2.p1	SYFSESIGIEYNLGR	Plongirostris_DN1174_c0_g2_i1.p1	TYFSDSGIEYNLGR
Pserratus_DN4233_c1_g2_i2.p1	ASLPIEGK	Plongirostris_DN1174_c0_g2_i1.p1	ASLPIEGK
Pserratus_DN4233_c1_g2_i2.p1	FLDGYAANGVK	Plongirostris_DN1174_c0_g2_i1.p1	FLDGYAANGVK
Pserratus_DN4233_c1_g2_i2.p1	SLQTAVFA	Plongirostris_DN1174_c0_g2_i1.p1	SLQTAVFA
Pserratus_DN4233_c1_g2_i2.p1	SETAVEVR	Plongirostris_DN1174_c0_g2_i1.p1	SETAVEVR
Pserratus_DN4233_c1_g2_i2.p1	SVHDFDAEAGVGR	Plongirostris_DN1174_c0_g2_i1.p1	SVHDFDAEAGVGR
Pserratus_DN4233_c1_g2_i2.p1	LELTSLSDPGQLEATAFR	Plongirostris_DN1174_c0_g2_i1.p1	LELTSLSDHGQLEATAFR
Pserratus_DN437_c0_g1_i4.p1	DSVVFEFWDNK	Plongirostris_DN74_c0_g1_i2.p1	DSVVFEFWDNK
Pserratus_DN437_c0_g1_i4.p1	GVTLPGASEGIR	Plongirostris_DN74_c0_g1_i2.p1	GVTLPGASEGIR
Pserratus_DN437_c0_g1_i4.p1	FDNNFIK	Plongirostris_DN74_c0_g1_i2.p1	FDNNFIK
Pserratus_DN94_c0_g1_i1.p1	ESDLPESSSYPEPAR	Plongirostris_DN178_c0_g1_i2.p1	ESDLPESSSYPEPAR
Pserratus_DN2880_c0_g2_i1.p1	VFFIDHINK	Plongirostris_DN447_c1_g1_i1.p1	VFFIDHINK
Pserratus_DN2880_c0_g2_i1.p1	DDFLGLVQVPLINLPR	Plongirostris_DN447_c1_g1_i1.p1	DDFLGLVQVPLINLPR
Pserratus_DN2880_c0_g2_i1.p1	EFFFLSK	Plongirostris_DN447_c1_g1_i1.p1	EFFFLSK
Pserratus_DN2880_c0_g2_i1.p1	WDEEFIFR	Plongirostris_DN447_c1_g1_i1.p1	WDEEFIFR
Pserratus_DN2880_c0_g2_i1.p2	FHISVDDNEVSR		
Pserratus_DN2880_c0_g2_i1.p3	NNILEDFAFGIISQVNR		
Pserratus_DN457_c0_g1_i4.p1	LPQEIHNK		
Pserratus_DN1804_c1_g1_i10.p1	DTSEGVASELVAAGLVDGR	Plongirostris_DN1062_c0_g1_i1.p1	DTSEGVASELVAAGLVDGR
Pserratus_DN1804_c1_g1_i10.p1	HGVFEFPVIATALR	Plongirostris_DN1062_c0_g1_i1.p1	HGVFEFPVIATALR
Pserratus_DN1804_c1_g1_i10.p1	YLQSVLLGGPISIEER	Plongirostris_DN1062_c0_g1_i1.p1	YLQSVLLGGPISIEER
Pserratus_DN1804_c1_g1_i10.p2	DLIVIAANLDK		

Annexe 5b : comparaison des peptides dans l'hémolymphe de *P. serratus* et *P. longirostris*.

<i>P. serratus</i> - hemolymphe		<i>P. longirostris</i> - hemolymphe	
proteine	peptide	proteine	peptide
P.serratus_DN1513_c0_g1_l1.p1	GEILPVK		
P.serratus_DN1513_c0_g1_l1.p1	NLVSLADTEDASFTANIDR	P.longirostris_DN116115_c0_g1_l1.p1	NLVTLADTNDASFTANIDR
P.serratus_DN1513_c0_g1_l1.p1	IHSVISISAUTEEGTGVEMR	P.longirostris_DN116115_c0_g1_l1.p1	IHSVISISAUTEEGTGVVEIK
P.serratus_DN1513_c0_g1_l1.p1	YNIPPEEASDAFGLTVSAR	P.longirostris_DN116115_c0_g1_l1.p1	YNIPPEEASDAFGLTVSAR
P.serratus_DN1513_c0_g1_l1.p1	TFTPFFLDLTPPSVK	P.longirostris_DN116115_c0_g1_l1.p1	TFTPFFLDLTPPSVK
P.serratus_DN1513_c0_g1_l1.p1	LLQQQVTVPSFPPTIK	P.longirostris_DN116115_c0_g1_l1.p1	LLQQQVTVPSFPPTIK
P.serratus_DN1513_c0_g1_l1.p1	SIEVETAGYAILAMLK	P.longirostris_DN116115_c0_g1_l1.p1	SIEVETAGYAILAMLK
P.serratus_DN1513_c0_g1_l1.p1	HDGFSFAFGNADESGSTWLTAFVLK	P.longirostris_DN116115_c0_g1_l1.p1	HDGFSFAFGNADESGSTWLTAFVLK
P.serratus_DN1513_c0_g1_l1.p1	ILTFSGSMFMDVETK	P.longirostris_DN116115_c0_g1_l1.p1	ILTFSGSMFMDVETK
P.serratus_DN1513_c0_g1_l1.p1	DDSF5FR		
P.serratus_DN1380_c0_g1_l5.p1	GGIVGAGSPVPLTAYVUSLLESSEPSTSK	P.longirostris_DN35363_c0_g3_l1.p1	GGIVGAGSPVPLTAYVUSLLESSEPSTSK
P.serratus_DN1380_c0_g1_l5.p1	GEILPVK		
P.serratus_DN1380_c0_g1_l5.p1	VLTLGEPFLK		
P.serratus_DN1380_c0_g1_l5.p1	IVQWNDVENESGLIHLSFK		
P.serratus_DN1380_c0_g1_l5.p1	ASITFTFTFFDLTPPSVK	P.longirostris_DN35363_c0_g3_l1.p1	ASITFTFTFFDLTPPSVK
P.serratus_DN1380_c0_g1_l5.p1	NQPSAVFTVQVISR		
P.serratus_DN1380_c0_g1_l5.p1	GWVTVAVGDLLGPTLENLGSLSVR	P.longirostris_DN35363_c0_g3_l1.p1	GWVTVAVGDLLGPTLENLGSLSVR
P.serratus_DN1380_c0_g1_l5.p1	IVGYGTGLFK	P.longirostris_DN35363_c0_g3_l1.p1	IVGYGTGLFK
P.serratus_DN1380_c0_g1_l5.p1	AVEPDPDTR		
P.serratus_DN1380_c0_g1_l5.p1	ISVFNYLESPLPVK	P.longirostris_DN35363_c0_g3_l1.p1	ISVFNYLESPLPVK
P.serratus_DN309_c0_g1_l11.p1	VYVDAPAEFR	P.longirostris_DN7554_c1_g1_l2.p1	VYVDAPAEFR
P.serratus_DN309_c0_g1_l11.p1	VSGDDQEVK	P.longirostris_DN7554_c1_g1_l2.p1	VSGDDQEVK
P.serratus_DN309_c0_g1_l11.p1	AFSGPVSASQPLR	P.longirostris_DN7554_c1_g1_l2.p1	AFSGPVSASQPLR
P.serratus_DN309_c0_g1_l11.p1	VGDTLYESK	P.longirostris_DN7554_c1_g1_l2.p1	VGDTLYESK
P.serratus_DN309_c0_g1_l11.p1	VIPSIFYAISLR	P.longirostris_DN7554_c1_g1_l2.p1	VIPSIFYAISLR
P.serratus_DN309_c0_g1_l11.p1	DLVTAVR		
P.serratus_DN309_c0_g1_l11.p1	GQHAFQVIAHMK	P.longirostris_DN7554_c1_g1_l2.p1	GQHAFQVIAHMK
P.serratus_DN309_c0_g1_l11.p1	ISGILTQGR	P.longirostris_DN7554_c1_g1_l2.p1	ISGILTQGR
P.serratus_DN309_c0_g1_l11.p1	SVEEFGFR		
P.serratus_DN309_c0_g1_l11.p1	VHNFIAIEWNR	P.longirostris_DN7554_c1_g1_l2.p1	VHNFIAIEWNR
P.serratus_DN1128_c0_g1_l9.p1	SYELPDGQVITIGNER	P.longirostris_DN871_c0_g1_l1.p1	SYELPDGQVITIGNER
P.serratus_DN1128_c0_g1_l9.p1	AGFAGDDAPR	P.longirostris_DN871_c0_g1_l1.p1	AGFAGDDAPR
P.serratus_DN1128_c0_g1_l9.p1	DLTDYLMK	P.longirostris_DN871_c0_g1_l1.p1	DLTDYLMK
P.serratus_DN1128_c0_g1_l9.p1	VAPPEHPVLLTEAPLNPK	P.longirostris_DN871_c0_g1_l1.p1	VAPPEHPVLLTEAPLNPK
P.serratus_DN1128_c0_g1_l9.p1	IIAPPER	P.longirostris_DN871_c0_g1_l1.p1	IIAPPER
P.serratus_DN1128_c0_g1_l9.p1	IWHHTFYNELR	P.longirostris_DN871_c0_g1_l1.p1	IWHHTFYNELR
P.serratus_DN1128_c0_g1_l9.p1	HQGVMMVGMGQK	P.longirostris_DN871_c0_g1_l1.p1	HQGVMMVGMGQK
P.serratus_DN1128_c0_g1_l9.p1	QYEDSGPSIVHR	P.longirostris_DN871_c0_g1_l1.p1	QYEDSGPSIVHR
P.serratus_DN1128_c0_g1_l9.p1	DSYVGDEAQSK	P.longirostris_DN871_c0_g1_l1.p1	DSYVGDEAQSK
P.serratus_DN1128_c0_g1_l9.p1	GYFTTTAER	P.longirostris_DN871_c0_g1_l1.p1	GYFTTTAER
P.serratus_DN35_c0_g1_l1.p1	SYELPDGQVITIGNER	P.longirostris_DN871_c0_g2_l3.p1	SYELPDGQVITIGNER
P.serratus_DN35_c0_g1_l1.p1	AGFAGDDAPR	P.longirostris_DN871_c0_g2_l3.p1	SGFAGDDAPR
P.serratus_DN35_c0_g1_l1.p1	DLTDYLMK	P.longirostris_DN871_c0_g2_l3.p1	DLTDYLMK
P.serratus_DN35_c0_g1_l1.p1	VAPPEHPVLLTEAPLNPK	P.longirostris_DN871_c0_g2_l3.p1	VAPPEHPVLLTEAPLNPK
P.serratus_DN35_c0_g1_l1.p1	IIAPPER	P.longirostris_DN871_c0_g2_l3.p1	IIAPPER
P.serratus_DN35_c0_g1_l1.p1	IWHHTFYNELR	P.longirostris_DN871_c0_g2_l3.p1	IWHHTFYNELR
P.serratus_DN35_c0_g1_l1.p1	HQGVMMVGMGQK	P.longirostris_DN871_c0_g2_l3.p1	HQGVMMVGMGQK
P.serratus_DN35_c0_g1_l1.p1	TTGIVLDSGDGVSHTVPIYEGYALPHAILR	P.longirostris_DN871_c0_g2_l3.p1	TTGIVLDSGDGVSHTVPIYEGYALPHAILR
P.serratus_DN35_c0_g1_l1.p1	DSYVGDEAQSK	P.longirostris_DN871_c0_g2_l3.p1	DSYVGDEAQSK
P.serratus_DN35_c0_g1_l1.p1	GYFTTTAER	P.longirostris_DN871_c0_g2_l3.p1	GYFTTTAER
P.serratus_DN2591_c0_g1_l2.p1	FTSPFLK	P.longirostris_DN4802_c0_g1_l5.p1	FTSPFLK
P.serratus_DN2591_c0_g1_l2.p1	GEIVGAYVQIPK	P.longirostris_DN4802_c0_g1_l5.p1	GEIVGAYVQIPK
P.serratus_DN2591_c0_g1_l2.p1	SHLDLEGSFGPNAGR	P.longirostris_DN4802_c0_g1_l5.p1	SHLDLEGSFGPNAGR
P.serratus_DN2591_c0_g1_l2.p1	AFDQLPYLEVLSVK	P.longirostris_DN4802_c0_g1_l5.p1	AFDQLPYLEVLSVK
P.serratus_DN2591_c0_g1_l2.p1	LTISSPSEHK	P.longirostris_DN4802_c0_g1_l5.p1	LTISSPSEHK
P.serratus_DN2591_c0_g1_l2.p1	LYISLNEK	P.longirostris_DN4802_c0_g1_l5.p1	LYISLNEK
P.serratus_DN2591_c0_g1_l2.p1	DNPDDEISGR	P.longirostris_DN4802_c0_g1_l5.p1	DNPDDEISGR
P.serratus_DN2591_c0_g1_l2.p1	AIALDSSGR	P.longirostris_DN4802_c0_g1_l5.p1	AIALDSSGR
P.serratus_DN23276_c0_g1_l2.p1	SYELPDGQVITIGNER	P.longirostris_DN2242_c0_g2_l4.p1	SYELPDGQVITIGNER
P.serratus_DN23276_c0_g1_l2.p1	AGFAGDDAPR		
P.serratus_DN23276_c0_g1_l2.p1	IIAPPER	P.longirostris_DN2242_c0_g2_l4.p1	IIAPPER
P.serratus_DN23276_c0_g1_l2.p1	GYSTTTAER	P.longirostris_DN2242_c0_g2_l4.p1	GYSTTTAER
P.serratus_DN23276_c0_g1_l2.p1	HQGVMMVGMGQK		
P.serratus_DN23276_c0_g1_l2.p1	TTGIVLDSGDGVSHTVPIYEGYALPHAILR	P.longirostris_DN2242_c0_g2_l4.p1	TTGIVLDSGDGVSHTVPIYEGYALPHAILR
P.serratus_DN23276_c0_g1_l2.p1	YPIEHGIITNWDOMEK		
P.serratus_DN23276_c0_g1_l2.p1	EITALAPSTIK	P.longirostris_DN2242_c0_g2_l4.p1	EITALAPSTIK
P.serratus_DN23276_c0_g1_l2.p1	DAVVGDEAQSK		
P.serratus_DN23276_c0_g1_l2.p1	DLTAYLMK	P.longirostris_DN2242_c0_g2_l4.p1	DLTAYLMK

Pserratus_DN7793_c0_g1_i1.p1	SYELPDGQVITIGNER	Plongirostris_DN30538_c0_g1_i1.p1	SYELPDGQVITIGNER
Pserratus_DN7793_c0_g1_i1.p1	IIAPPER		
Pserratus_DN7793_c0_g1_i1.p1	DLYANTVLSGGTTMYPGIADR	Plongirostris_DN30538_c0_g1_i1.p1	DLYANTVLSGGTTMYPGIADR
Pserratus_DN7793_c0_g1_i1.p1	QEYDESGPSIVHR		
Pserratus_DN7793_c0_g1_i1.p1	EITALAPSTMK		
Pserratus_DN272_c0_g3_i1.p1	DLTDYLMK	Plongirostris_DN3169_c0_g1_i1.p1	DLTDYLMK
Pserratus_DN272_c0_g3_i1.p1	GYSFTTAER	Plongirostris_DN3169_c0_g1_i1.p1	GYSFTTAER
Pserratus_DN272_c0_g3_i1.p1	SGFAGDDAPR	Plongirostris_DN3169_c0_g1_i1.p1	SGFAGDDAPR
Pserratus_DN11247_c0_g2_i1.p1	SYELPDGQVITIGNER		
Pserratus_DN11247_c0_g2_i1.p1	GYSFTTAER		
Pserratus_DN2620_c0_g1_i1.p1	GEILPVK	Plongirostris_DN11621_c0_g1_i1.p1	GEILPVK
Pserratus_DN2620_c0_g1_i1.p1	SNMAVIEVNLVSGYIPEK	Plongirostris_DN11621_c0_g1_i1.p1	SNMAVIEVNLVSGYIPEK
Pserratus_DN2620_c0_g1_i1.p1	YEVGSK		
Pserratus_DN2620_c0_g1_i1.p1	VSSDPYTLALK	Plongirostris_DN11621_c0_g1_i1.p1	VSSDPYTLALK
Pserratus_DN2620_c0_g1_i1.p1	ISVFNLYK	Plongirostris_DN11621_c0_g1_i1.p1	ISVFNLYK
Pserratus_DN2620_c0_g1_i1.p1	YLDVTQQNTPESA EK	Plongirostris_DN11621_c0_g1_i1.p1	YLDVTQQNTPESA EK
Pserratus_DN2620_c0_g1_i1.p1	VGEHVLP R	Plongirostris_DN11621_c0_g1_i1.p1	VGEHVLP R
Pserratus_DN2620_c0_g1_i1.p1	ADLHSASFSR	Plongirostris_DN11621_c0_g1_i1.p1	ADLHSASFSR
Pserratus_DN2620_c0_g1_i1.p1	LEESSEYEVDEPSEGVK	Plongirostris_DN11621_c0_g1_i1.p1	LEESSEYEVDEPSEGVK
Pserratus_DN7715_c0_g1_i3.p1	QSLEGIGK		
Pserratus_DN7715_c0_g1_i3.p1	MLYSSSFDA LK		
Pserratus_DN7715_c0_g1_i3.p1	VIDVETTGD R		
Pserratus_DN7715_c0_g1_i3.p1	YIIFHIK		
Pserratus_DN556_c0_g2_i1.p1	TINEVENQILTR		
Pserratus_DN556_c0_g2_i1.p1	VGWEQLT SINR		
Pserratus_DN556_c0_g2_i1.p1	SVTDIANAWK		
Pserratus_DN556_c0_g2_i1.p1	GISQEQLTEFR		
Pserratus_DN556_c0_g2_i1.p1	LAPDEFK		
Pserratus_DN556_c0_g2_i1.p1	QQDN TLADLQK		
Pserratus_DN556_c0_g2_i1.p1	LETNFNTLQTK		
Pserratus_DN556_c0_g2_i1.p1	LTSDDLLEWIR		
Pserratus_DN556_c0_g2_i1.p1	ENVSDTAEQVIDSFR		
Pserratus_DN556_c0_g2_i1.p1	SFEDWLLSEMMR		
Pserratus_DN1641_c0_g1_i2.p1	IFLDGLK		
Pserratus_DN1641_c0_g1_i2.p1	EGESHTESGINHIER		
Pserratus_DN1641_c0_g1_i2.p1	IMEEAQEEYER		
Pserratus_DN1641_c0_g1_i2.p1	WTLWVGGR		
Pserratus_DN1641_c0_g1_i2.p1	WLVNDGTTLPNFFWSK		
Pserratus_DN3222_c0_g1_i1.p1	TLTLPVK		
Pserratus_DN3222_c0_g1_i1.p1	TVSALLTASSVYTGAK		
Pserratus_DN3222_c0_g1_i1.p1	SSDIPLLVEHATIK		
Pserratus_DN3222_c0_g1_i1.p1	IVNSNDDNGVVLGR		
Pserratus_DN3222_c0_g1_i1.p1	ITLYSAAR		
Pserratus_DN3222_c0_g1_i1.p1	TVHISNVAAR		
Pserratus_DN3222_c0_g1_i1.p1	WDGNYEDGSPPTK		
Pserratus_DN3222_c0_g1_i1.p1	VVVGSYPTAQGR		
Pserratus_DN3222_c0_g1_i1.p1	SNQYHVGK		
Pserratus_DN3222_c0_g1_i1.p1	SEHLT YILFNPWNK		

Annexe 5c : comparaison des peptides dans l'hépatopancréas de *P. serratus* et *P. longirostris*.

<i>P. serratus</i> - hepato		<i>P. longirostris</i> - hepato	
proteine	peptide	proteine	peptide
Pseratus_DN9025_c1_g3_i1.p1	IILELALPK	Plangirostris_DN8009_c0_g3_i1.p1	IILELALPK
Pseratus_DN9025_c1_g3_i1.p1	VSYYGQPLGLLVTEGR	Plangirostris_DN8009_c0_g3_i1.p1	VSYYGQPLGLLVTEGR
Pseratus_DN9025_c1_g3_i1.p1	IVSIDPSEAIK	Plangirostris_DN8009_c0_g3_i1.p1	IVSIDPSEAIK
Pseratus_DN9025_c1_g1_i1.p1	TILGIIQDVASPDLSGGATNLVRLPSGK	Plangirostris_DN8009_c0_g3_i1.p1	TILGIIQDVASPDLSGGATNLVRLPSGK
Pseratus_DN9025_c1_g3_i1.p1	GPAYGALGAVK	Plangirostris_DN8009_c0_g3_i1.p1	GPAYGALGAVK
Pseratus_DN9025_c1_g3_i1.p1	QVYDENQEIWPLGK	Plangirostris_DN8009_c0_g3_i1.p1	QVYDENQEIWPLGK
Pseratus_DN9025_c1_g3_i1.p1	VTGEPPLVLSFSVVMALR	Plangirostris_DN8009_c0_g3_i1.p1	VTGEPPLVLSFSVVMALR
Pseratus_DN9025_c1_g3_i1.p1	LGGAFGSK		
Pseratus_DN9025_c1_g3_i1.p1	SVPELSTVNQTADK	Plangirostris_DN8009_c0_g3_i1.p1	SVPELSTVNQTADK
Pseratus_DN9025_c1_g3_i1.p1	VTLENSPNPIGVQLSK	Plangirostris_DN8009_c0_g3_i1.p1	VTLENSPNPIGVQLSK
Pseratus_DN7572_c0_g1_i1.p1	LNENETVLEAVR	Plangirostris_DN7183_c0_g1_i1.p1	LNENETVLEAVR
Pseratus_DN7572_c0_g1_i1.p1	QHILSSAAVAAR	Plangirostris_DN7183_c0_g1_i1.p1	QHILSSAAVAAR
Pseratus_DN7572_c0_g1_i1.p1	ALADPGYAHLGEIAK	Plangirostris_DN7183_c0_g1_i1.p1	ALADPGYAHLGEIAK
Pseratus_DN7572_c0_g1_i1.p1	TWVFLGDDQASSK	Plangirostris_DN7183_c0_g1_i1.p1	TWVFLGDDQASSK
Pseratus_DN7572_c0_g1_i1.p1	LLGEEVSPDNHFEDASPEYR	Plangirostris_DN7183_c0_g1_i1.p1	LLGEEVSPDNHFEDASPEYR
Pseratus_DN7572_c0_g1_i1.p1	NVAADDAIPILR	Plangirostris_DN7183_c0_g1_i1.p1	NVAADDAIPILR
Pseratus_DN7572_c0_g1_i1.p1	VAQVAAYALR	Plangirostris_DN7183_c0_g1_i1.p1	VAQVAAYALR
Pseratus_DN7572_c0_g1_i1.p1	HWQYVANVSVR	Plangirostris_DN7183_c0_g1_i1.p1	HWQYVANVSVR
Pseratus_DN7572_c0_g1_i1.p1	ITVSDVHWVSPK	Plangirostris_DN7183_c0_g1_i1.p1	ITVSDVHWVSPK
Pseratus_DN7572_c0_g1_i1.p1	YAGQPVGAIVAENR	Plangirostris_DN7183_c0_g1_i1.p1	YAGQPVGAIVAENR
Pseratus_DN1498_c0_g1_i4.p1	SLDGLWNFR	Plangirostris_DN78_c0_g1_i1.p1	SLDGLWNFR
Pseratus_DN1498_c0_g1_i4.p1	TFIVPTR	Plangirostris_DN78_c0_g1_i1.p1	TFIVPTR
Pseratus_DN1498_c0_g1_i4.p1	WLGANSFR	Plangirostris_DN78_c0_g1_i1.p1	WLGANSFR
Pseratus_DN1498_c0_g1_i4.p1	NGDLAASLDIILVNR	Plangirostris_DN78_c0_g1_i1.p1	NGDLAASLDIILVNR
Pseratus_DN1498_c0_g1_i4.p1	GVDYPLVK	Plangirostris_DN78_c0_g1_i1.p1	GVDYPLVK
Pseratus_DN1498_c0_g1_i4.p1	DHIGWAWYDR	Plangirostris_DN78_c0_g1_i1.p1	DHIGWAWYDR
Pseratus_DN1498_c0_g1_i4.p1	HEDADIR	Plangirostris_DN78_c0_g1_i1.p1	HEDADIR
Pseratus_DN1498_c0_g1_i4.p1	MAGHLMR	Plangirostris_DN78_c0_g1_i1.p1	MAGHLMR
Pseratus_DN1498_c0_g1_i4.p1	VVGWNSTLTINGVPAYLR	Plangirostris_DN78_c0_g1_i1.p1	VVGWNSTLTINGVPAYLR
Pseratus_DN85230_c0_g1_i1.p1	FDYFYTGSTGVGK	Plangirostris_DN3358_c0_g1_i2.p1	FDYFYTGSTGVGK
Pseratus_DN85230_c0_g1_i1.p1	EVPSVVPEDK	Plangirostris_DN3358_c0_g1_i2.p1	EVPSVVPEDK
Pseratus_DN85230_c0_g1_i1.p1	DLDFILSR	Plangirostris_DN3358_c0_g1_i2.p1	DLDFILSR
Pseratus_DN85230_c0_g1_i1.p1	IMQEEIFGPIPLNVK	Plangirostris_DN3358_c0_g1_i2.p1	IMQEEIFGPIPLNVK
Pseratus_DN85230_c0_g1_i1.p1	DVQDTFVQK	Plangirostris_DN3358_c0_g1_i2.p1	DVQDTFVQK
Pseratus_DN85230_c0_g1_i1.p1	MFLEALK	Plangirostris_DN3358_c0_g1_i2.p1	MFLEALK
Pseratus_DN85230_c0_g1_i1.p1	YLPTTLELGGK	Plangirostris_DN3358_c0_g1_i2.p1	YLPTTLELGGK
Pseratus_DN85230_c0_g1_i1.p1	DVINFNER	Plangirostris_DN3358_c0_g1_i2.p1	DVINFNER
Pseratus_DN85230_c0_g1_i1.p1	MYVENEK	Plangirostris_DN3358_c0_g1_i2.p1	MYVENEK
Pseratus_DN85230_c0_g1_i1.p1	VIETLVPK	Plangirostris_DN3358_c0_g1_i2.p1	VIETLVPK
Pseratus_DN6358_c1_g1_i1.p1	SLDGLWNFR	Plangirostris_DN2815_c0_g1_i1.p1	SLDGLWNFR
Pseratus_DN6358_c1_g1_i1.p1	LAAHILR	Plangirostris_DN2815_c0_g1_i1.p1	LAAHILR
Pseratus_DN6358_c1_g1_i1.p1	ITVYGDNLTK	Plangirostris_DN2815_c0_g1_i1.p1	ITVYGDNLTK
Pseratus_DN6358_c1_g1_i1.p1	WLGANVFR	Plangirostris_DN2815_c0_g1_i1.p1	WLGANVFR
Pseratus_DN6358_c1_g1_i1.p1	SPSFLFSSEYQAEELR	Plangirostris_DN2815_c0_g1_i1.p1	SPSFLFSSEYQAEELR
Pseratus_DN6358_c1_g1_i1.p1	HEDADIR	Plangirostris_DN2815_c0_g1_i1.p1	HEDADIR
Pseratus_DN6358_c1_g1_i1.p1	FGSAHYSTR	Plangirostris_DN2815_c0_g1_i1.p1	FGSAHYSTR
Pseratus_DN6358_c1_g1_i1.p1	SSQASADSYFR	Plangirostris_DN2815_c0_g1_i1.p1	SSQASADSYFR
Pseratus_DN6358_c1_g1_i1.p1	GTTGIINYK	Plangirostris_DN2815_c0_g1_i1.p1	GTTGIINYK
Pseratus_DN28968_c0_g1_i1.p1	LIYFNLR	Plangirostris_DN2815_c0_g1_i1.p1	LIYFNLR
Pseratus_DN28968_c0_g1_i1.p1	WIFAYAGIPYTDER	Plangirostris_DN2815_c0_g1_i1.p1	WIFAYAGIPYTDER
Pseratus_DN28968_c0_g1_i1.p1	TLEPALSINK	Plangirostris_DN2815_c0_g1_i1.p1	TLEPALSINK
Pseratus_DN28968_c0_g1_i1.p1	LYAELNOK	Plangirostris_DN2815_c0_g1_i1.p1	LYAELNOK
Pseratus_DN28968_c0_g1_i1.p1	ITWADLALGLVLGGIAK	Plangirostris_DN2815_c0_g1_i1.p1	ITWADLALGLVLGGIAK
Pseratus_DN28968_c0_g1_i1.p1	IQEWIEK	Plangirostris_DN2815_c0_g1_i1.p1	IQEWIEK
Pseratus_DN28968_c0_g1_i1.p1	STIPFGK	Plangirostris_DN2815_c0_g1_i1.p1	STIPFGK
Pseratus_DN28968_c0_g1_i1.p1	LHMMTDEAER	Plangirostris_DN2815_c0_g1_i1.p1	LHMMTDEAER
Pseratus_DN8839_c0_g1_i3.p1	LVLEYVGEQYEEK	Plangirostris_DN2981_c1_g1_i1.p1	LVLEYVGEQYEEK
Pseratus_DN8839_c0_g1_i3.p1	HISNLQPTLK	Plangirostris_DN2981_c1_g1_i1.p1	HISNLQPTLK
Pseratus_DN8839_c0_g1_i3.p1	LLSNILGK	Plangirostris_DN2981_c1_g1_i1.p1	LLSNILGK
Pseratus_DN8839_c0_g1_i3.p1	GLGQYIR	Plangirostris_DN2981_c1_g1_i1.p1	GLGQYIR
Pseratus_DN8839_c0_g1_i3.p1	FEDLPAIK	Plangirostris_DN2981_c1_g1_i1.p1	FEDLPAIK
Pseratus_DN8839_c0_g1_i3.p1	FGPPPEYDR	Plangirostris_DN2981_c1_g1_i1.p1	FGPPPEYDR
Pseratus_DN8839_c0_g1_i3.p1	SUNGPSAAK	Plangirostris_DN2981_c1_g1_i1.p1	SUNGPSAAK

Pserratus_DN10656_c0_g1_i1.p1	TNLPSNTAFR	Plongirostris_DN10031_c1_g1_i1.p1	TNLPSNTAFR
Pserratus_DN10656_c0_g1_i1.p1	VAAFLNVDSYLVR	Plongirostris_DN10031_c1_g1_i1.p1	IAAFLNVDSYLVR
Pserratus_DN10656_c0_g1_i1.p1	MIQVASR	Plongirostris_DN10031_c1_g1_i1.p1	MIQVASR
Pserratus_DN10656_c0_g1_i1.p1	AIGEPPLFLGASVIFSIK	Plongirostris_DN10031_c1_g1_i1.p1	AIGEPPLFLGASVIFSIK
Pserratus_DN10656_c0_g1_i1.p1	VLSLATGFYR	Plongirostris_DN10031_c1_g1_i1.p1	VLSLATGFYR
Pserratus_DN10656_c0_g1_i1.p1	YSPEGVLLTR	Plongirostris_DN10031_c1_g1_i1.p1	YSPEGVLLTR
Pserratus_DN10656_c0_g1_i1.p1	TSIDSFNR	Plongirostris_DN10031_c1_g1_i1.p1	TSIDSFNR
Pserratus_DN10656_c0_g1_i1.p1	VNTDLQTSYLEFHGPR	Plongirostris_DN10031_c1_g1_i1.p1	VNTDLQTSYLEFHGPR
Pserratus_DN10656_c0_g1_i1.p1	LDSPATAER	Plongirostris_DN10031_c1_g1_i1.p1	LDSPATAER
Pserratus_DN4311_c0_g1_i1.p1	WVSPITLVDVSPESK	Plongirostris_DN31724_c0_g1_i1.p1	WISPTILVDVSPESK
Pserratus_DN4311_c0_g1_i1.p1	YQQIVESTR	Plongirostris_DN31724_c0_g1_i1.p1	YQQIVESTR
Pserratus_DN4311_c0_g1_i1.p1	EVFLSAK	Plongirostris_DN31724_c0_g1_i1.p1	EVFLSAK
Pserratus_DN4311_c0_g1_i1.p1	FDYIFYTGSPVVGK	Plongirostris_DN31724_c0_g1_i1.p1	FDYIFYTGSPVVGK
Pserratus_DN4311_c0_g1_i1.p1	HLTPTTLELGGK	Plongirostris_DN31724_c0_g1_i1.p1	HLTPTTLELGGK
Pserratus_DN4311_c0_g1_i1.p1	ASYDTFTHR	Plongirostris_DN31724_c0_g1_i1.p1	ASYDTFTHR
Pserratus_DN4311_c0_g1_i1.p1	VMTEEIFGPILPVINVSSAYEAIR	Plongirostris_DN31724_c0_g1_i1.p1	VMTEEIFGPILPVINVSSAYEAIR
Pserratus_DN70045_c0_g1_i1.p2	TFIAVKPDGVQR	Plongirostris_DN1377_c0_g1_i2.p2	TFIAVKPDGVQR
Pserratus_DN70045_c0_g1_i1.p2	NIIHGSDSIESANK	Plongirostris_DN1377_c0_g1_i2.p2	NIIHGSDSIESANK
Pserratus_DN70045_c0_g1_i1.p2	GLVGEIIK	Plongirostris_DN1377_c0_g1_i2.p2	GLVGEIIK

Annexe 5d : comparaison des peptides dans les branchies de *D. polymorpha* et *D. r. bugensis*.

<i>D. polymorpha</i> - branchie	Peptide	<i>D. r. bugensis</i> - branchie	Peptide
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	GAPEGLDR	Db_LOC101846151.1.4::g.47625::m.47625	GAPEGLDR
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	IDQSILGTSVSVLK	Db_LOC101846151.1.4::g.47625::m.47625	IDQSILGTSVSVLK
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	NILFSGTNIASGK	Db_LOC101846151.1.4::g.47625::m.47625	NILFSGTNIASGK
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	VGEATELTVLVEK	Db_LOC101846151.1.4::g.47625::m.47625	VGEATELTVLVEK
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	EFTLEFSR	Db_LOC101846151.1.4::g.47625::m.47625	EFTLEFSR
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	GVVIGTGLNTEIGK	Db_LOC101846151.1.4::g.47625::m.47625	GVVIGTGLNTEIGK
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	INYYNTEK	Db_LOC101846151.1.4::g.47625::m.47625	INFFNTEK
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	LTIHSTTLR	Db_LOC101846151.1.4::g.47625::m.47625	VTHIHTTLR
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	SASEMVLADDNFATVAAVEEGR	Db_LOC101846151.1.4::g.47625::m.47625	SASEMVLADDNFATVAAVEEGR
TRINITY_DN207134_c2_g1::TRINITY_DN207134_c2_g1_j3::g.873451::m.873451	VIVITGDNK	Db_LOC101846151.1.4::g.47625::m.47625	VIVITGDNK
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	GAGTDEQAIIGISQR	Db_LOC106173262.1.5::g.107302::m.107302	GAGTDEQAIIGISQR
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	LQTQGPPEAVIDR	Db_LOC106173262.1.5::g.107302::m.107302	LQTQGPPEAVIDR
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	LLLALVQK	Db_LOC106173262.1.5::g.107302::m.107302	LLLALVQK
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	GAGTQDSTLIR	Db_LOC106173262.1.5::g.107302::m.107302	GAGTQDSTLIR
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	LGTDSEAFSLVLAIR	Db_LOC106173262.1.5::g.107302::m.107302	MGTDSEAFSLVLAIR
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	VQIMNMFK	Db_LOC106173262.1.5::g.107302::m.107302	VQIMNMFK
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	ATFEEYVK	Db_LOC106173262.1.5::g.107302::m.107302	ATFEEYVK
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	DILSSIDR	Db_LOC106173262.1.5::g.107302::m.107302	DILSSIDR
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	AELTPDQWQR	Db_LOC106173262.1.5::g.107302::m.107302	EL_PDQWQR
TRINITY_DN203049_c2_g3::TRINITY_DN203049_c2_g3_j3::g.953583::m.953583	DGFMALVMSAK	Db_LOC106173262.1.5::g.107302::m.107302	DGFMALVMSAK
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	GVSDLAQHLYLR	Db_LOC105342582.g.481244::m.481244	GVSDLAQHLYLR
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	AGITAIR		
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	SITGVHQWPYR	Db_LOC105342582.g.481244::m.481244	SITGVHQWPYR
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	MALEDLETLK	Db_LOC105342582.g.481244::m.481244	MALEDLETLK
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	NVMQEPFLVPGGAAEMALQALNEK	Db_LOC105342582.g.481244::m.481244	NVMQEPFLVPGGAAEMALQALNEK
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	EINVQHPAAK	Db_LOC105342582.g.481244::m.481244	EINVQHPAAK
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	TAIETAILLR	Db_LOC105342582.g.481244::m.481244	TAIETAILLR
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	ALEVIPR	Db_LOC105342582.g.481244::m.481244	ALEVIPR
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	MEEEFIEK	Db_LOC105342582.g.481244::m.481244	MEEEFIEK
TRINITY_DN214176_c2_g3::TRINITY_DN214176_c2_g3_j1::g.1091047::m.1091047	AIADVIPR		
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	LLQNYPPPEPAVK	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	LLQNYPPPEPAVK
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	SGSAVQFQLLHSHK	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	SGSAVQFQLLHSHK
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	EMAEGSLVDIPK	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	EMAEGSLVDIPK
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	GLAVFISDIR	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	GLAVFISDIR
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	VGGVILGFEFGLIAGDPR	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	VGGVILGFEFGLIAGDPR
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	ATSPIDHEQVK	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	ATSPIDHEQVK
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	LIGFSVSVLEK	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	LIGFSVSVLEK
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	LLVSGDITDSVK	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	LLVSGDITDSVK
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	NADVELQQR	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	NADVELQQR
TRINITY_DN211570_c1_g2::TRINITY_DN211570_c1_g2_j2::g.880414::m.880414	YITPAEMDANFTFTR	Db_LOTGIDRAFT_203023.1.2::g.32520::m.32520	YITPAEMDANFTFTR
TRINITY_DN213857_c0_g2::TRINITY_DN213857_c0_g2_j2::g.554529::m.554529	NLLSVAYK	Db_1433Z.1.2::g.302511::m.302511	NLLSVAYK
TRINITY_DN213857_c0_g2::TRINITY_DN213857_c0_g2_j2::g.554529::m.554529	DSTLMQLLR	Db_1433Z.1.2::g.302511::m.302511	DSTLMQLLR
TRINITY_DN213857_c0_g2::TRINITY_DN213857_c0_g2_j2::g.554529::m.554529	DNLTLLWK	Db_1433Z.1.2::g.302511::m.302511	DNLTLLWK
TRINITY_DN195056_c0_g1::TRINITY_DN195056_c0_g1_j1::g.126724::m.126724	ETASAVGEAALEAFASR	Db_LOC101855225.1.2::g.186247::m.186247	ETASAVGEAALEAFASR
TRINITY_DN195056_c0_g1::TRINITY_DN195056_c0_g1_j1::g.126724::m.126724	IGENISENEFVK	Db_LOC101855225.1.2::g.186247::m.186247	IGENISENEFVK
TRINITY_DN195056_c0_g1::TRINITY_DN195056_c0_g1_j1::g.126724::m.126724	LTNGMPVSIIEAGSGFQK	Db_LOC101855225.1.2::g.186247::m.186247	LTNGMPVSIIEAGSGFQK
TRINITY_DN195056_c0_g1::TRINITY_DN195056_c0_g1_j1::g.126724::m.126724	SAAYEAAGWLEK	Db_LOC101855225.1.2::g.186247::m.186247	STAYEAAGWFDK
TRINITY_DN195056_c0_g1::TRINITY_DN195056_c0_g1_j1::g.126724::m.126724	NMFHTINSAQK	Db_LOC101855225.1.2::g.186247::m.186247	NMFH_INSA_K
TRINITY_DN195056_c0_g1::TRINITY_DN195056_c0_g1_j1::g.126724::m.126724	SGNIMEIDDLIK	Db_LOC101855225.1.2::g.186247::m.186247	SGNIMEIDDLIK
TRINITY_DN218026_c1_g3::TRINITY_DN218026_c1_g3_j6::g.794724::m.794724	QQEPGEDFFIILEGSAVLQQR	Db_KAPR.1.2::g.486098::m.486098	QQEPGEDFFIILEGSAVLQQR
TRINITY_DN218026_c1_g3::TRINITY_DN218026_c1_g3_j6::g.794724::m.794724	AATVIAR		
TRINITY_DN218026_c1_g3::TRINITY_DN218026_c1_g3_j6::g.794724::m.794724	HNVQQLLK	Db_KAPR.1.2::g.486098::m.486098	HNVQQLLK
TRINITY_DN218026_c1_g3::TRINITY_DN218026_c1_g3_j6::g.794724::m.794724	LTVADALEPQVFDGQEIWR	Db_KAPR.1.2::g.486098::m.486098	LTVADALEPQVFDGQEIWR
TRINITY_DN218026_c1_g3::TRINITY_DN218026_c1_g3_j6::g.794724::m.794724	TMAALSK	Db_KAPR.1.2::g.486098::m.486098	TMAALSK
TRINITY_DN181093_c1_g2::TRINITY_DN181093_c1_g2_j2::g.384799::m.384799	SILLQLK	Db_LOC106179046.g.214164::m.214164	SILLQLK
TRINITY_DN199632_c0_g1::TRINITY_DN199632_c0_g1_j1::g.360986::m.360986	WDHPGQNTAGLDDFER	Db_LOC106874074.1.2::g.12774::m.12774	WDHPGQNTAGLDDFER
TRINITY_DN199632_c0_g1::TRINITY_DN199632_c0_g1_j1::g.360986::m.360986	FSEPHSR	Db_LOC106874074.1.2::g.12774::m.12774	FSEPHSR
TRINITY_DN199632_c0_g1::TRINITY_DN199632_c0_g1_j1::g.360986::m.360986	TLGTGSFGR	Db_LOC106874074.1.2::g.12774::m.12774	TLGTGSFGR
TRINITY_DN199632_c0_g1::TRINITY_DN199632_c0_g1_j1::g.360986::m.360986	GPGDTSNFDDYEEPLR	Db_LOC106874074.1.2::g.12774::m.12774	GPGDTSNFDDYEEPLR
TRINITY_DN206276_c2_g1::TRINITY_DN206276_c2_g1_j1::g.1110224::m.1110224	LAVLDVQAFER	Db_LOC105342800.g.349076::m.349076	LAVLDVQAFER
TRINITY_DN206276_c2_g1::TRINITY_DN206276_c2_g1_j1::g.1110224::m.1110224	MNIADALYSK	Db_LOC105342800.g.349076::m.349076	MNIADALYSK
TRINITY_DN206276_c2_g1::TRINITY_DN206276_c2_g1_j1::g.1110224::m.1110224	NINEYEQLVTFGSK	Db_LOC105342800.g.349076::m.349076	NINEYEQLVTFGSK
TRINITY_DN206276_c2_g1::TRINITY_DN206276_c2_g1_j1::g.1110224::m.1110224	AAISDLR	Db_LOC105342800.g.349076::m.349076	AAISDLR

Annexe 5e: comparaison des peptides dans l'hémolymphe de *D. polymorpha* et *D. r. bugensis*.

D. polymorpha - hemolymphe		D. bugensis - hemolymphe	
Protéine	Peptide	Protéine	Peptide
TRINITY_DN215561_c0_g3_i6	GHGGQFPAGALLELSR	Db_LOC106612870:g.588130:m.588130	GFGGPFPAEAVLELSS
TRINITY_DN215561_c0_g3_i6	VTVFNYLQNEQTAR	Db_LOC106612870:g.588130:m.588130	VTVFNYMNMMSMDAR
TRINITY_DN215561_c0_g3_i6	LETYAIPIEVR	Db_LOC106612870:g.588130:m.588130	LETYAIPIEVR
TRINITY_DN215561_c0_g3_i6	DVVVPDAEHPVQVTIK	Db_LOC106612870:g.588130:m.588130	ITDAAHPVEVTK
TRINITY_DN215561_c0_g3_i6	GTVTQQNLQR		
TRINITY_DN215561_c0_g3_i6	LTVTGPSPGMWVGFNVIDK	Db_LOC106612870:g.588130:m.588130	LAVNGPSGMWVGFNVID
TRINITY_DN215561_c0_g3_i6	TNAVQDSIHR		
TRINITY_DN215561_c0_g3_i6	DGASFIEAK		
TRINITY_DN215561_c0_g3_i6	SVFSNQDQYTR	Db_LOC106612870:g.588130:m.588130	SMFSSQDKYT
TRINITY_DN215561_c0_g3_i6	YVLAPNVIR	Db_LOC106612870:g.588130:m.588130	YVLAPNVIR
TRINITY_DN208778_c4_g2_i11	IDMFYNR	Db_LOC105333243.1.2:g.190723:m.19072	IDMFYNR
TRINITY_DN208778_c4_g2_i11	TLAYISK		
TRINITY_DN208778_c4_g2_i11	STQDVTVMQALAAYSAR	Db_LOC105333243.1.2:g.190723:m.19072	TTQDVTVMQALATYSAR
TRINITY_DN208778_c4_g2_i11	THVNSSENSLQOR		
TRINITY_DN208778_c4_g2_i11	LIQDLPLEAGDSK		
TRINITY_DN208778_c4_g2_i11	TPEVQNSINR		
TRINITY_DN208778_c4_g2_i11	FVVTASGR		
TRINITY_DN208778_c4_g2_i11	GIAGMDSLEK		
TRINITY_DN208778_c4_g2_i11	VVESDLK		
TRINITY_DN208778_c4_g2_i11	TMESTHEDSGIK		
TRINITY_DN220774_c7_g4_i2	ALLLNSDK		
TRINITY_DN220774_c7_g4_i2	VDVEGHPFIR		
TRINITY_DN220774_c7_g4_i2	NAGVTLLISSGTLNK		
TRINITY_DN220774_c7_g4_i2	TDTVVADSSSFYIESDQR		
TRINITY_DN220774_c7_g4_i2	IIVYEHR		
TRINITY_DN220774_c7_g4_i2	TNNIVVEK		
TRINITY_DN220774_c7_g4_i2	IVAFVVDK		
TRINITY_DN205441_c5_g15_i1	EMTNYVPSSTGK	Db_LOC105329085.2.3:g.211635:m.21163	EMTNYVPSSTGK
TRINITY_DN205441_c5_g15_i1	HPAENVVFAEFTK	Db_LOC105329085.2.3:g.211635:m.21163	HPAENVVFAEFTK
TRINITY_DN205441_c5_g15_i1	HPPEWSVGDIK	Db_LOC105329085.2.3:g.211635:m.21163	HPPEWSVGDIK
TRINITY_DN205441_c5_g15_i1	FEDMAEQDK	Db_LOC105329085.2.3:g.211635:m.21163	FEDMAEQDK
TRINITY_DN205441_c5_g15_i1	WEAVTDR	Db_LOC105329085.2.3:g.211635:m.21163	WEAVTDR
TRINITY_DN199594_c10_g9_i1	GGFHGHPWVSGSK	Db_LOC106051403.4.4:g.354549:m.35454	GGFHGHPWVSGSK
TRINITY_DN199594_c10_g9_i1	VPVIGSGGFQWK	Db_LOC106051403.4.4:g.354549:m.35454	VPIESGGFWQK
TRINITY_DN199594_c10_g9_i1	HLQYGGWPASGEIDIMESR	Db_LOC106051403.4.4:g.354549:m.35454	HTQYGGWPASGEIDIMESR
TRINITY_DN199594_c10_g9_i1	FGHEFLHGGTMDVDR	Db_LOC106051403.4.4:g.354549:m.35454	FGHEFLHGGTMDVDR
TRINITY_DN199594_c10_g9_i1	NEIGATLHWGPPSQNR	Db_LOC106051403.4.4:g.354549:m.35454	NEIGATLHWGPNPSQNR
TRINITY_DN199594_c10_g9_i1	INGILPPVMSGK	Db_LOC106051403.4.4:g.354549:m.35454	INGILPPVMSGK
TRINITY_DN199594_c10_g9_i1	WELTHAAR	Db_LOC106051403.4.4:g.354549:m.35454	YELTHAAR
TRINITY_DN199594_c10_g9_i1	IPQGDWIIWPAIWMLPR	Db_LOC106051403.4.4:g.354549:m.35454	IPQGDWIIWPAIWMLPR
TRINITY_DN199594_c10_g9_i1	TDGFSFNSFHLVSLDWTVDHIIVSDNQVLR	Db_LOC106051403.4.4:g.354549:m.35454	TDGFSFNSFHLVSLDWTVDHIIVSDNQVLR
TRINITY_DN211288_c0_g1_i8	SVFSNFHDQYER		
TRINITY_DN211288_c0_g1_i8	VFAFYVVK		
TRINITY_DN211288_c0_g1_i8	HFDTSFFFEK		
TRINITY_DN211288_c0_g1_i8	HVDGTPAGETDIAVEIYNDIGR		
TRINITY_DN211288_c0_g1_i8	GYPYMQVYMR		
TRINITY_DN211288_c0_g1_i8	QLLVSEGIQEK		
TRINITY_DN211288_c0_g1_i8	FGDTSFIQASTDIDTSR		
TRINITY_DN211288_c0_g1_i8	TNNDGHLTLVNGFNATAVIR		
TRINITY_DN211288_c0_g1_i8	VEFDLPADLVK		
TRINITY_DN211288_c0_g1_i8	EQSGSDSVTFINAPYK		
TRINITY_DN192719_c2_g5_i2	SISEEFLIAFK	Db_LOC105334536.3.3:g.535212:m.53521	SISEEFLIAFK
TRINITY_DN192719_c2_g5_i2	AYGHENVALDTK	Db_LOC105334536.3.3:g.535212:m.53521	AYGHENVALDTK
TRINITY_DN192719_c2_g5_i2	SLDVNHDGKITFADVEESR	Db_LOC105334536.3.3:g.535212:m.53521	SLDVNHDGKITFADVEESR
TRINITY_DN192719_c2_g5_i2	DGLVPLR	Db_LOC105334536.3.3:g.535212:m.53521	DGLVPLR
TRINITY_DN192719_c2_g5_i2	WWNEFIFR	Db_LOC105334536.3.3:g.535212:m.53521	WWNEFIFR
TRINITY_DN192719_c2_g5_i2	FSELHLDEQK	Db_LOC105334536.3.3:g.535212:m.53521	FSELHLPEEQ
TRINITY_DN192719_c2_g5_i2	QVMDFNEK		
TRINITY_DN192719_c2_g5_i2	TGEISESEFVDMLNK	Db_LOC105334536.3.3:g.535212:m.53521	TGEISESEDFIKMLSK
TRINITY_DN207483_c0_g4_i1	HIEFTDDTFGEIIR	Db_LOC105341865.1.4:g.82954:m.82954	HIDTFDDTFGEIIR
TRINITY_DN207483_c0_g4_i1	HVVNFDLPSDIEEYVHR	Db_LOC105341865.1.4:g.82954:m.82954	HVVNFDLPSDIEEYVHR
TRINITY_DN207483_c0_g4_i1	LYETGPGDAPVSGLPQGGK	Db_LOC105341865.1.4:g.82954:m.82954	LYDTGPGDAPVGLPQGGK
TRINITY_DN207483_c0_g4_i1	HYQSQQNYDVNSR	Db_LOC105341865.1.4:g.82954:m.82954	HYQSQQNYDVNSR
TRINITY_DN207483_c0_g4_i1	VGSTSENITQK	Db_LOC105341865.1.4:g.82954:m.82954	VGSTSENITQK
TRINITY_DN207483_c0_g4_i1	GLDIPHVR	Db_LOC105341865.1.4:g.82954:m.82954	GLDIPHVR
TRINITY_DN207483_c0_g4_i1	ELASQIYDEAR	Db_LOC105341865.1.4:g.82954:m.82954	ELASQIYDEAR
TRINITY_DN207483_c0_g4_i1	FULDEADR	Db_LOC105341865.1.4:g.82954:m.82954	FULDEADR
TRINITY_DN207483_c0_g4_i1	TAAFLVPTLNK	Db_LOC105341865.1.4:g.82954:m.82954	TAAFLVPTLNK
TRINITY_DN207483_c0_g4_i1	QFPALVLAPTR	Db_LOC105341865.1.4:g.82954:m.82954	QFPALVLAPTR
TRINITY_DN212368_c1_g6_i6	IDLFYSR		
TRINITY_DN212368_c1_g6_i6	DSAVAYQAAAFSAR		
TRINITY_DN212368_c1_g6_i6	MLGYLSTR		
TRINITY_DN212368_c1_g6_i6	DGYDNFDLPIPIYSSK		
TRINITY_DN212368_c1_g6_i6	QTGTITEQQK		
TRINITY_DN205477_c0_g1_i1	VVDYNNPEK		
TRINITY_DN205477_c0_g1_i1	QYSINPNK		
TRINITY_DN205477_c0_g1_i1	GFIIFYLSK		
TRINITY_DN205477_c0_g1_i1	MASAVVK		
TRINITY_DN194686_c0_g4_i1	LETYAIPIEVR	Db_LOC106612870:g.588130:m.588130	LETYAIPIEVR
TRINITY_DN194686_c0_g4_i1	LENEETLLVGLGNR	Db_LOC106612870:g.588130:m.588130	LENEETLLVGLGNR
TRINITY_DN194686_c0_g4_i1	VSFEPGSSK		
TRINITY_DN194686_c0_g4_i1	QLAIELISPTNMTVGR	Db_LOC106612870:g.588130:m.588130	QLAIELISPTNMTVGR
TRINITY_DN194686_c0_g4_i1	GQFELPPFDGTWTIK	Db_LOC106612870:g.588130:m.588130	GEFELPPFDGTWTIK

Annexe 5f : comparaison des peptides dans les glandes digestives de *D. polymorpha* et *D. r. bugensis*.

<i>D. polymorpha</i> - glandes digestive		<i>D. r. bugensis</i> - glandes digestive	
Protéine	Peptide	Protéine	Peptide
Dp_LOC105342066::g.23160::m.23160	IGLVDPVVEPIGVGK	Db_LOC101856538::g.621044::m.621044	IGLVDPVVEPIGVGK
Dp_LOC105342066::g.23160::m.23160	ILAPTLSEAVR	Db_LOC101856538::g.621044::m.621044	ILAPTLSEAVR
Dp_LOC105342066::g.23160::m.23160	FGGGNINVLR	Db_LOC101856538::g.621044::m.621044	FGGGNINVLR
Dp_LOC105342066::g.23160::m.23160	LLPFQATPMMMLTGK	Db_LOC101856538::g.621044::m.621044	LLPFATPMMMLTGK
Dp_LOC105342066::g.23160::m.23160	SANENTMEELER	Db_LOC101856538::g.621044::m.621044	SANENTMEELER
Dp_LOC105342066::g.23160::m.23160	YIDMFASSR	Db_LOC101856538::g.621044::m.621044	YIDMFASSR
Dp_LOC105342066::g.23160::m.23160	DMVTAGFLGR	Db_LOC101856538::g.621044::m.621044	DMVTAGFLGR
Dp_LOC105342066::g.23160::m.23160	TSQDTAASAVSVGLK	Db_LOC101856538::g.621044::m.621044	TSQDTAASAVSVGLK
Dp_LOC101869038::g.87598::m.87598	LAEQTPLSGLHIASLFK	Db_LOC101869038.2.2::g.544162::m.544162	LAEQTPLSGLHIASLFK
Dp_LOC101869038::g.87598::m.87598	EAGYPEGVLNVITGYGPTAGAAISR	Db_LOC101869038.2.2::g.544162::m.544162	EAGYPEGVLNVITGYGPTAGAAISK
Dp_LOC101869038::g.87598::m.87598	IYDEFVER	Db_LOC101869038.2.2::g.544162::m.544162	IYDEFVER
Dp_LOC101869038::g.87598::m.87598	ELGEYGMAYTEVK	Db_LOC101869038.2.2::g.544162::m.544162	ELGEYGMAYTEVK
Dp_LOC101869038::g.87598::m.87598	LFINNEFR	Db_LOC101869038.2.2::g.544162::m.544162	LFINNEFR
Dp_LOC101869038::g.87598::m.87598	VTLELGGK	Db_LOC101869038.2.2::g.544162::m.544162	VTLELGGK
Dp_LOC101869038::g.87598::m.87598	YGLAAGVVTNDVSR	Db_LOC101869038.2.2::g.544162::m.544162	YGLAAGVVTNDVSR
Dp_LOC101869038::g.87598::m.87598	VEQGPQVDEQMK	Db_LOC101869038.2.2::g.544162::m.544162	VEQGPQVDEQMK
Dp_LOC101869038::g.87598::m.87598	AINVANSLR	Db_LOC101869038.2.2::g.544162::m.544162	AINVANSLR
Dp_LOC101869038::g.87598::m.87598	TAEFVER		
Dp_LOC102809579.3.9::g.279328::m.279328	TPVPLEVDGR		
Dp_LOC102809579.3.9::g.279328::m.279328	LYYDFDR		
Dp_LOC102809579.3.9::g.279328::m.279328	QITQSMATR	Db_LOC106054059::g.19801::m.19801	QITQSMATR
Dp_LOC102809579.3.9::g.279328::m.279328	TTNAPVYFNLVK	Db_LOC106054059::g.19801::m.19801	TTNAPVYFNLVK
Dp_LOC102809579.3.9::g.279328::m.279328	QAGVDFQDIR		
Dp_LOC102809579.3.9::g.279328::m.279328	SGSGFYVGA	Db_LOC106054059::g.19801::m.19801	SGSGFYVGA
Dp_LOC102809579.3.9::g.279328::m.279328	GLAELPR		
Dp_LOC102809579.3.9::g.279328::m.279328	IYSEQDETK	Db_LOC106054059::g.19801::m.19801	IYSEQDETK
Dp_LOC105340590.2.2::g.453987::m.453987	AALAGGTTMIMDFVLDQK	Db_LOC105340590.1.3::g.5430::m.5430	AALAGGTTMIMDFVLDQK
Dp_LOC105340590.2.2::g.453987::m.453987	QIGNNLVIPPGAR	Db_LOC105340590.1.3::g.5430::m.5430	QIGNNLVIPPGAR
Dp_LOC105340590.2.2::g.453987::m.453987	FVAVTSTNAK	Db_LOC105340590.1.3::g.5430::m.5430	FVAVTSTNAK
Dp_LOC105340590.2.2::g.453987::m.453987	IAVGSDADIVWDPK	Db_LOC105340590.1.3::g.5430::m.5430	IAVGSDADIVWDPK
Dp_LOC105340590.2.2::g.453987::m.453987	TNSIETAPPR	Db_LOC105340590.1.3::g.5430::m.5430	TNSIETAPPR
Dp_LOC105340590.2.2::g.453987::m.453987	LFNIYPQK	Db_LOC105340590.1.3::g.5430::m.5430	LFNIYPQK
Dp_LOC105340590.2.2::g.453987::m.453987	NQQDSSFSFGDQDDGQQR	Db_LOC105340590.1.3::g.5430::m.5430	NQQDSSFSFGDQDDGQQR
Dp_LOC105340590.2.2::g.453987::m.453987	SAADMVTR	Db_LOC105340590.1.3::g.5430::m.5430	SAADMVTR
Dp_LOC105340590.2.2::g.453987::m.453987	VPIHLQSAQSR	Db_LOC105340590.1.3::g.5430::m.5430	VPIHLQSAQSR
Dp_LOC105340590.2.2::g.453987::m.453987	MSVIWEK	Db_LOC105340590.1.3::g.5430::m.5430	MSVIWEK
Dp_LOC105340590.1.2::g.22680::m.22680	AALAGGTTMIMDFVLDQK	Dp_LOC105340590.1.2::g.22680::m.22680	AALAGGTTMIMDFVLDQK
Dp_LOC105340590.1.2::g.22680::m.22680	FVAVTSTNAK	Dp_LOC105340590.1.2::g.22680::m.22680	FVAVTSTNAK
Dp_LOC105340590.1.2::g.22680::m.22680	IAVGSDADIVWDPK	Dp_LOC105340590.1.2::g.22680::m.22680	IAVGSDADIVWDPK
Dp_LOC105340590.1.2::g.22680::m.22680	QIGNNLVIPPGAR	Dp_LOC105340590.1.2::g.22680::m.22680	QIGNNLVIPPGAR
Dp_LOC105340590.1.2::g.22680::m.22680	SAADMVTR		
Dp_LOC105340590.1.2::g.22680::m.22680	TNSIETAPPR	Dp_LOC105340590.1.2::g.22680::m.22680	TNSIETAPPR
Dp_LOC105340590.1.2::g.22680::m.22680	LFNIYPQK	Dp_LOC105340590.1.2::g.22680::m.22680	LFNIYPQK
Dp_LOC105340590.1.2::g.22680::m.22680	MSVIWEK	Dp_LOC105340590.1.2::g.22680::m.22680	MSVIWEK
Dp_LOC105340590.1.2::g.22680::m.22680	VPIHLQSAQSR	Dp_LOC105340590.1.2::g.22680::m.22680	VPIHLQSAQSR
Dp_LOC105340590.1.2::g.22680::m.22680	IPNGVNGVEDR	Dp_LOC105340590.1.2::g.22680::m.22680	IPNGVNGVEDR
Dp_LOC105340590.1.2::g.22680::m.22680	SSDLVIEAIVENLEVK	Db_LOC105340590.1.2::g.22680::m.22680	SSDLVIEAIVENLEVK
Dp_LOC105340590.1.2::g.22680::m.22680	FIIDGWAK	Db_LOC105340590.1.2::g.22680::m.22680	FIIDGWAK
Dp_LOC105340590.1.2::g.22680::m.22680	YPNPFLDPSLLNK	Db_LOC105340590.1.2::g.22680::m.22680	YPNPFLDPSLLNK
Dp_LOC105340590.1.2::g.22680::m.22680	LLVPYMMEA VR	Db_LOC105340590.1.2::g.22680::m.22680	LLVPYMMEA VR
Dp_LOC105340590.1.2::g.22680::m.22680	DTPGFIVNR	Db_LOC105340590.1.2::g.22680::m.22680	DTPGFIVNR
Dp_LOC105340590.1.2::g.22680::m.22680	FGGLHYFNPPVVMK	Db_LOC105340590.1.2::g.22680::m.22680	FGGLHYFNPPVVMK
Dp_LOC105340590.1.2::g.22680::m.22680	LFASLDK		
Dp_LOC105340590.1.2::g.22680::m.22680	GEATPQDVTAMK	Db_LOC105340590.1.2::g.22680::m.22680	GEATPQDVTAMK
Dp_LOC105340590.1.2::g.22680::m.22680	VGEGFYK	Db_LOC105340590.1.2::g.22680::m.22680	VGEGFYK
Dp_LOC10533880.3.8::g.181063::m.181063	VTVADFPYELLQHIR	Db_LOC10533880.7.7::g.595487::m.595487	VTVADFPYELLQHIR
Dp_LOC10533880.3.8::g.181063::m.181063	ITQSNAILR	Db_LOC10533880.7.7::g.595487::m.595487	ITQSNAILR
Dp_LOC10533880.3.8::g.181063::m.181063	NLPINNK	Db_LOC10533880.7.7::g.595487::m.595487	NLPINNK
Dp_LOC10533880.3.8::g.181063::m.181063	GSLELFEK		
Dp_LOC10533880.3.8::g.181063::m.181063	SDLSYSR		
Dp_LOC10533880.3.8::g.181063::m.181063	EYLATDNVK	Db_LOC10533880.7.7::g.595487::m.595487	EYLATDNVK
Dp_LOC10533880.3.8::g.181063::m.181063	HNLDGESVQEK	Db_LOC10533880.7.7::g.595487::m.595487	HNLDGESVQEK
Dp_LOC10533880.3.8::g.181063::m.181063	AEVDMMLDQAMDRLR	Db_LOC10533880.7.7::g.595487::m.595487	AEVDMMLDQAMDRLR
Dp_LOC10533880.3.8::g.181063::m.181063	NWFGGNK	Db_LOC10533880.7.7::g.595487::m.595487	NWFGGNK
Dp_LOC106174241.1.2::g.11679::m.11679	SVDSVPLVDQYMR	Db_LOC106174241.2.2::g.445665::m.445665	SVDSVPLVDQYMR
Dp_LOC106174241.1.2::g.11679::m.11679	GTAFGGWK	Db_LOC106174241.2.2::g.445665::m.445665	GTAFGGWK
Dp_LOC106174241.1.2::g.11679::m.11679	ETQAGVMPNGTAR	Db_LOC106174241.2.2::g.445665::m.445665	ETQAGVMPNGTAR
Dp_LOC106174241.1.2::g.11679::m.11679	AAVAWEAK	Db_LOC106174241.2.2::g.445665::m.445665	AAVAWEAK
Dp_LOC105340488.3.4::g.419498::m.419498	TDVVDNHIPR	Db_LOC105340488.4.7::g.351317::m.351317	TDFIENHGPR
Dp_LOC105340488.3.4::g.419498::m.419498	FPDVAHAK	Db_LOC105340488.4.7::g.351317::m.351317	FPDVAHAK
Dp_LOC105340488.3.4::g.419498::m.419498	LVFSAAK	Db_LOC105340488.4.7::g.351317::m.351317	LVFGAAK
Dp_LOC105340488.3.4::g.419498::m.419498	NVSAYLAK	Db_LOC105340488.4.7::g.351317::m.351317	NVAAYLAK
Dp_LOC105340488.3.4::g.419498::m.419498	LGNYSIK	Db_LOC105340488.4.7::g.351317::m.351317	LGNVYAK
Dp_LOC105340488.3.4::g.419498::m.419498	NGFIIGSK	Db_LOC105340488.4.7::g.351317::m.351317	NGFIIGSK
Dp_LOC106156945::g.215850::m.215850	VVAEEIFGPVLVLPFR	Db_LOC106156945.1.2::g.76385::m.76385	VVAEEIFGPVLVLPFR
Dp_LOC106156945::g.215850::m.215850	TSYGPAPAEADNVAQAWLDDHER	Db_LOC106156945.1.2::g.76385::m.76385	TSYGPAPAEADNVAQAWLDDHER
Dp_LOC106156945::g.215850::m.215850	DADIPVYVR	Db_LOC106156945.1.2::g.76385::m.76385	DADVPLVYR
Dp_LOC106156945::g.215850::m.215850	LLVQESVYGR	Db_LOC106156945.1.2::g.76385::m.76385	LLVQESVYGR
Dp_LOC106156945::g.215850::m.215850	LSLELGGK	Db_LOC106156945.1.2::g.76385::m.76385	LSLELGGK
Dp_LOC106156945::g.215850::m.215850	EVDSLISR	Db_LOC106156945.1.2::g.76385::m.76385	EVDSLISR
Dp_LOC106156945::g.215850::m.215850	NAVEAAHAAAPGWGK	Db_LOC106156945.1.2::g.76385::m.76385	NAVEAAHAAAPGWGK
Dp_LOC106156945::g.215850::m.215850		Db_LOC106156945.1.2::g.76385::m.76385	NPATGQNLLATLQGTEDVDMAVK