UNIVERSITÉ DU QUÉBEC À RIMOUSKI

Génomique fonctionnelle séquentielle pour la recherche de marqueurs génétiques de la croissance et de la résistance larvaire chez la moule bleue *Mytilus edulis*

Thèse présentée

dans le cadre du doctorat en océanographie

en vue de l'obtention du grade de *Philosophiæ Doctor es* océanographie

PAR

© SLEIMAN BASSIM

Juin 2014

Composition du jury :

[Céline Audet], présidente du jury, [Université de Québec à Rimouski]

[Réjean Tremblay], directeur de recherche, [Université de Québec à Rimouski]

[Dario Moraga], codirecteur de recherche, [Université de Bretagne Occidentale]

[Bassem Allam], examinateur externe, [Stony Brook University]

Dépôt initial le [9 juin 2014] Dépôt final le [3 septembre 2014]

Pour maman Joji, Marie et Zayzaf

Start by doing what is necessary then do what is possible,
and you will soon find yourself doing the impossible.
— Saint Francis of Assisi

REMERCIEMENTS

Tout d'abord je remercie les membres du jury pour leur enthousiasme, leur patience et leur disponibilité. Je remercie également toutes les personnes qui ont contribué de quelque façon que ce soit au bon déroulement de ma thèse que ce soit technique, analytique ou expérimental. Un remerciement spécial pour les directeurs Réjean Tremblay et Dario Moraga, pour votre disponibilité, vos encouragements dans les situations difficiles, votre soutien dans toutes mes initiatives et votre confiance dans mes jugements à mener ce travail vers une fin désirée. Je ne voudrais pas non plus oublier les enseignants, les encadrants et les amis de recherche depuis mon départ du Liban en 2006, passant par la France à Nancy, Montpellier, Paris et Brest : Lionel, Nathalie, Pierre, Philippe, Florence, Gérard, Souad, Arnaud, Sophie, Michel et Martine.

Je tiens également à remercier Louis qui au cours de mes années à Brest m'a fait partager ses passions. Pour les bons moments loin des bases de données et des ordinateurs, je remercie les amis de Rimouski et de Brest qui ont partagé avec moi les moments d'exaltation et qui m'ont fait la preuve de leur amitié. Phill (king of T.), Valérie et nos discussions sans fin, Alexandra, Ossan, Manue, Souad, Anne-Sophie et nos soirées films, Isabelle, Carole, Shanie et Manu. Bien d'autres ont contribué d'une façon ou d'une autre dans l'accomplissement de cette thèse, que Dieu vous bénisse. Je remercie enfin mes sœurs Jocy, Nicole et Coco et en particulier ma mère Joji, pour ses prières et sa patience. Tu as fait développer chez moi les valeurs de la persévérance et de la détermination. Je m'en charge d'ici là.

AVANT PROPOS

Ce texte pourrait être utilisé comme une ressource pour les chercheurs en écologie, génétique et biologie moléculaire. Pour cette raison, les hypothèses de la thèse ont été explorées d'une manière à faire ressortir des liens écoévolutifs basés sur la génétique moléculaire afin d'expliquer l'effet nutritionnel sur la réponse génomique pendant l'ontogenèse larvaire. L'écologie constitue la recherche des effets contemporains des facteurs biotiques et abiotiques de l'environnement. La génétique se base sur le rôle direct de l'ADN dans la fonction cellulaire. La biologie moléculaire établit un lien interactif entre les produits fonctionnels des gènes, la réponse cellulaire et l'adaptation phénotypique. Ainsi, cette approche multidisciplinaire permettra d'explorer les effets de l'adaptation locale. Toutefois, des études complémentaires de ce genre peuvent devenir surchargées d'interprétations complexes. Pour cette raison, la première phase est facilitée par une étude à l'échelle transcriptomique des larves. Ceci ouvre la possibilité d'utiliser des techniques modernes qui peuvent être modulées pour produire une variété d'inférences descriptives de l'étendue des composants génomiques. En effet, les diverses applications incluses ici sont établies pour tester des hypothèses ainsi que pour documenter les observations. De ce fait, les résultats retenues peuvent contenir des informations potentiellement erronées. Or, l'objectif principal est de pénétrer les mystères biologiques, la raison pour laquelle j'encourage la vérification de la crédibilité de nos conclusions. Des idées ont été testées. Certaines idées qui n'ont pas passé les tests ont été donc rejetées. D'autres ont permis de choisir des observations que j'ai utilisé ensuite pour développer mes positions. Les modèles statistiques, la génétique quantitative (analyses multivariées et les profils d'expression des gènes) et les calculs intensifs pour la modélisation moléculaire employés nous ont permis d'examiner explicitement les effets écophysiologiques et l'appa-

rition de possible adaptation locale. Comme il est indiqué dans les chapitres de cette thèse, nous avons combiné une collection de méthodologies classique et de nouvelle génération pour tester ces hypothèses.

Les chapitres ont été écrits de façon à minimiser la redondance du contenu. Un des défis rencontrés repose sur la longueur du texte et l'étendue des approches qui peuvent être entreprises pour l'explication des hypothèses. Les études en génomique basées sur des analyses descriptives des bases de données transcriptomiques génèrent habituellement un grand nombre d'informations. Bien que les résultats de la thèse sont divisés sous trois chapitres, le plan expérimental de leurs différentes parties a été développé pour faire ressortir la complémentarité des hypothèses et des observations. Il se peut que certaines notions de génétique ou de génomique fonctionnelle aient été omises dans l'introduction générale de la thèse, par exemple, la régulation épigénétique, les interactions gène-protéine, l'effet dose des espèces réactives d'oxygène et l'activation de la défense immunitaire, les marqueurs génétiques du stress oxydatif, l'effet direct des acides gras sur le phénotype et le système immunitaire. Effectivement, certaines de ces notions ont été étudiées avec plus de consistance dans d'autres projets dévoués entièrement à leurs effets, or les analyses spécifiques à ce travail de thèse ont été spécialement développées pour caractériser de nouveaux patrons génétiques. Ces gènes candidats sont souvent définis dans le texte d'après le rôle de leur fonction protéique dans ces mêmes voies de résistance et du maintien de l'homéostasie. Le texte a néanmoins été réduit en taille afin d'essayer de couvrir le maximum d'informations relatives aux processus du développement. Malgré la restriction que je me suis imposée, chaque chapitre servira d'introduction pour d'autres revues, livres et travaux publiés concernant ces axes de recherche. D'ailleurs, en raison de l'efficacité des méthodologies utilisées ici pour le traitement des données, de nouveaux travaux publiés peuvent être considérés comme plus appropriés que ceux cités. J'encourage le lecteur à se lancer dans ces approches nouvellement développées puisque même si les avenues des traitements se ressemblent elles ne sont pas identiques. L'approche analytique basée sur le *High Performance Computing* a été développé

récemment, et son efficacité et son interprétabilité sont souvent améliorées dans ces nouvelles publications.

Les figures conceptuelles illustrent les différentes observations des résultats de la thèse. Les propositions faites ne peuvent pas couvrir la totalité des systèmes géniques étudiés, puisque les mécanismes moléculaires sont complexes et nombreux. Pour cette raison, des informations supplémentaires retirées de revues, livres ou publications ont été utilisées pour créer une meilleure structure des résultats. L'objectif de cette approche est de présenter l'information sans réduire sa valeur, et surtout, d'argumenter plausiblement l'efficacité du plan expérimental. Il se peut que certaines informations aient été retirées du texte pour faciliter l'interprétation des figures et tableaux, mais aussi éliminer toute redondance avec les travaux cités qui comportent parfois plus d'informations que nous ne pouvons résumer dans un seul chapitre. Les notes en bas de page et leurs définitions comprennent des notions utilisées dans tous les chapitres. Certaines de ces définitions ont été reprises de travaux publiés, mais la plupart correspondent aux résumés d'observations personnelles qui forment la base de la présente thèse. Pour un aperçu global des différentes notions incluses dans ce travail de thèse, une liste comprenant les indices relatifs a été compilée à la fin de la thèse. Les codes R pour les analyses statistiques et certaines figures (non conceptuelles) sont partagés publiquement sur GitHub. Les codes d'accès se trouvent dans la troisième partie de l'annexe. Une liste des principaux logiciels utilisés pour le développement des pipelines analytiques, les analyses statistiques, la rédaction de la thèse et le système de contrôle des versions (VCS) est introduite également dans la troisième partie de l'annexe. Finalement, j'exprime ma gratitude à toute personne qui a participé dans la lecture et a contribué dans l'édition de ce texte. La rédaction d'un ouvrage basé sur des techniques multidisciplinaires qui traite plusieurs problématiques n'est pas une tâche facile. Effectivement, le compromis entre l'usage d'un jargon technique pour décrire les fonctions moléculaires ou les modèles *Machine Learning* pour la prédiction génique et la simplification des discussions pour une interprétation facile a relevé un défi personnel important. Dans la mesure où nous avons réussi et dans le but

de démontrer la valeur de ma thèse, je dois remercier tous mes éditeurs et les coauteurs qui ont contribué à ce travail.

Sleiman Bassim

Mysteries don't lose their poetry because they are solved,
the solution often is more beautiful than the puzzle,
uncovering deeper mysteries.
— *Richard Dawkins [Dawkins, 2000]*

RÉSUMÉ

La dissertation est divisée en deux parties suivies par une discussion générale pour conclure sur des prospectifs dans les disciplines de transcriptomique, génomique et physiologique. La première partie introduit la sélection naturelle de la larve et son évolution qui varie en fonction de son adaptation locale gène-dépendante et la valeur sélective modulée par les pressions sélectives environnementales. La deuxième partie est divisée en trois chapitres qui traitent en détail de nouvelles découvertes et les expériences effectuées en biologie quantitative. Les analyses des réseaux de gènes ont contribué au développement de décisions théoriques bayésiennes et à la classification supervisée de modèles prédictifs. Enfin je discute l'importance et les avantages d'une approche multidisciplinaire dans l'étude de la biologie évolutive et je propose des précautions à entreprendre pour le traitement des données fonctionnelles en génétique.

L'objectif principal est de caractériser chez la moule bleue *Mytilus edulis* les effets moléculaires d'une carence nutritionnelle en acides gras pro-inflammatoires, appelés les précurseurs des eicosanoïdes. Les eicosanoïdes sont des hormones impliquées dans des mécanismes moléculaires peu connus chez les larves pélago-benthiques. Ces acides gras ont un rôle fonctionnel important dans la structure et la fluidité des membranes cellulaires et sont à l'origine de la valeur sélective des populations. Une approche nutrigénomique nous a permis d'examiner les modulateurs de l'adaptation locale et les effets des facteurs écologiques

sur l'ontogénie des espèces marines. L'histoire de vie des moules est influencée fortement par la qualité et la disponibilité des microalgues qui constituent la nourriture principale pendant la phase larvaire. Effectivement, cette dépendance est centrale dans le développement larvaire chez les invertébrés marins et spécialement chez les bivalves. Plusieurs notions relatives sont introduites dans la première partie de la thèse. D'ailleurs nous décrivons les interactions écophysiologiques et la prédisposition génétique sur la valeur inclusive des espèces. En effet, le taux des réserves énergétiques des larves varie en fonction de la nature de la nourriture. Elle est fondamentale pour leur adaptation locale, la survie, le succès métamorphique puis l'acquisition d'une maturité sexuelle. Ces hypothèses ont été testées par des méthodes modernes en génomique développementale. Nous avons donc réussi à identifier un lien entre la dynamique du génome et la valeur sélective du phénotype à transition rapide.

La biométrie larvaire a été déterminée au début de la période du développement ontogénique. Nous avons ainsi enregistré les différentes transitions d'un organisme jeune, peu développé et susceptible à de fortes mortalités, en un organisme plus résistant. Ensuite, le transcriptome a été séquencé par RNA-seq afin de caractériser les régulations différentielles des larves. Par conséquent, le premier chapitre de la deuxième partie traite l'importance du séquençage à l'échelle d'un génome entier. Notre stratégie a permis d'incorporer toutes les interactions et les variantes géniques dans une étude physiogénétique plus fiable. Par exemple, les mécanismes qui s'activent pendant une défense immunitaire en réponse à une infection bactérienne peuvent aussi être impliqués dans la croissance cellulaire et le développement sans la présence de perturbations exogènes. Par ailleurs, l'interaction entre les régulateurs transcriptionnels et leurs gènes cibles a été étudiée pendant le développement. Ceci nous a aidé à définir la coordination du réseautage génique et les changements physiologiques essentiels pour l'assemblage morphologique des larves. Des modèles de prédiction (Machine Learning) pour la recherche supervisée des signatures de l'expression génique ont permis l'exploration du lien dynamique entre les acides gras nutritionnels et la réponse gé-

nomique. D'ailleurs, nous avons caractérisé de nouveaux biomarqueurs par les différents modèles statistiques employés sur tout le transcriptome de *M. edulis*. Par exemple, nous avons découvert de nouveaux gènes différentiellement exprimés pendant le développement larvaire et dépendamment du régime nutritionnel par classification supervisée. Ces marqueurs sont relatifs à la tolérance aux stress et la réparation de l'ADN. L'exploitation des données physiologiques et génomiques a permis de confirmer l'existence d'un impact environnemental sur la dynamique du génome larvaire de *M. edulis*. En plus, les acides gras pro-inflammatoires sont capables d'activer plusieurs mécanismes de défense et du métabolisme énergétique. Ces mécanismes sont basés sur des connexions géniques complexes dont l'activation varie en fonction de l'âge des larves. Enfin, l'ensemble de ces résultats confirme que l'adaptation locale est dépendante de la valeur sélective de plusieurs réseaux de gènes interconnectés. Pour cette raison nous avons réussi à caractériser 29 biomarqueurs génomiques capables d'associer les effets des précurseurs des eicosanoïdes à la croissance et la mortalité des larves de *M. edulis*. Nous avons présenté dans cette thèse une approche intégrative qui inclue le transcriptome et les données comportementales. De cette manière nous apportons une compréhension globale de la biologie des systèmes. Ceci permettra la mise en place d'une relation entre l'évaluation environnementale et le développement fonctionnel chez les bivalves marins.

Mots clés : *Mytilus edulis*, larve, ontogenèse, précurseurs des eicosanoïdes, RNA-seq, transcriptome, data mining, Machine Learning, croissance, mortalité

ABSTRACT

The dissertation is divided into two parts followed by a general discussion and concluding remarks to future goals in transcriptomics, genomics and physiology. The first part introduces the influencing factors of natural selection and evolution of larvae that are dependent on their local adaptation and their genes' inclusive fitness which is modulated by selective environmental pressures. The second part is divided into three chapters, where each addresses in detail new discoveries and all the biological experiments, expression analysis, and pipeline designs. Analysis of gene networks contributed to the development of Bayesian decision theories and supervised classification models. The discussion assesses the importance of a multidisciplinary approach in studying evolutionary biology and the precautions in the assessment of functional genomic data.

The main objective here was to characterize the impact of a deficient diet in fatty acids in the blue mussel *Mytilus edulis*. These fatty acids are pro-inflammatory eicosanoid precursors responsible of the molecular mechanisms behind larval fitness during early development. They confer further structural fluidity to cell membranes. A nutrigenomic approach provided us with essential information about modulators in marine adaptation and thorough knowledge of the ecological players during the progress of larval ontogenesis. Life cycle history of adult bivalves is strongly influenced by the quality and availability of food during the larval phase. Indeed, this dependency found amongst invertebrates and specifically in bivalves is central to the growing larvae. Several related concepts have been introduced in the first part of the thesis. In addition, we included detailed examples of the effect of eco-physiological interactions and genetic predisposition on the inclusive fitness of species. Energy reserves can vary greatly depending on the microalgal fatty acid composition. Microalgae constitute the main

food source for the feeding larvae and are fundamental to their local adaptation. Moreover, they are nutrient-rich, which can contribute to the survival of larvae, their metamorphic success, and the duration of their transitional period to sexually mature adults. All these physiological hypotheses were tested using modern methods in developmental genomics. We thus identified a connection between the dynamical function of the genome and the fast-changing phenotype's inclusive fitness.

Furthermore, biometrics were measured at early larval development. We recorded the various transitions of a young, underdeveloped body and likely susceptible of high mortality, into a more resistant form. Next, larval transcriptome was RNA-sequenced. Consequently, the first chapter of the second part addresses the importance of whole-genome sequencing. Our strategy helped integrate various constituents of the genome for a more reliable physiogenetic assessment. For example, mechanisms known to be specifically activated during an immune defence in response to infections, can also be associated with cellular growth and development in the absence of exogenous perturbations. In addition, interactions between transcriptional regulators and their target genes were thoroughly investigated. This helped us define coordinated expressions between genes and the physiological changes underlying larval body-plan transitions. Supervised Machine Learning models were used to identify gene signatures from expression profiles during larval development, provided that the genomic response generated was assessed relatively to the impact of dietary fatty acids. Statistical modeling enabled us to map the whole transcriptome of *M. edulis* and to identify new genetic biomarkers. For example, we discovered new differentially regulated genes implicated in stress resistance and DNA repair using specialized supervised classifications. Therefore, data mining of the newly built transcriptome of *M. edulis* and notably the large-scale examination of gene expressions provided new insights on the existence of a direct environmental impact on the dynamic of larval transcriptome. Furthermore, pro-inflammatory fatty acids can activate depending on the stage of development, a newly discovered set of interrelated genes that promote defence and metabolic regulation. Lastly, these results confirm that

the inclusive fitness of the identified genetic regulatory networks (GRNs) can affect the local adaptation of larvae. For this reason, we were able to characterize 29 genetic markers that connect the effect of eicosanoid precursors with *M. edulis* larval growth and mortality. We have presented in this thesis an integrative approach of transcriptomic and behavioural data that brings a holistic understanding of systems biology. This will help to establish links between environmental assessment and functional development in marine bivalves.

Keywords : *Mytilus edulis*, larvae, ontogenesis, eicosanoid precursors, RNA-seq, transcriptome, data mining, Machine Learning, growth, mortality.

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

ACRONYMES

4-HB-CoAT 4-hydroxybutyrate coenzyme A transferase
ABCF3 ATP-binding cassette sub-family F member 3
ACTL actin-like protein
AIC Akaike's information criterion
AIF-1 allograft inflammatory factor-1
AMP antimicrobial peptide
ANAPC7 anaphase-promoting complex subunit 7
ANN artificial neural networks
ANOVA one-way analysis of variance
APOD apolipoprotein D
ArA arachidonic acid
ARF4 ADP-ribosylation factor 4
ARF6 ADP-ribosylation factor 6
ATRX transcriptional regulator atrx
B9D1 B9 domain-containing protein 1
bagging boostrap aggregating
BP biological process
CALM2 calmodulin 2
CAT Catalase
CBX3 chromobox protein 3
CC cellular component
CDC42 cell division control protein 42
ChIP immunoprécipitation de la chromatine
CGI hypothetical protein CGI
CnVAS1 vasa-like protein
Co cocktail diet
CO_2 dioxyde de carbone
CRD carbohydrate recognition domain
CSNK2B casein kinase ii subunit beta isoform 1

CTNNB beta-catenin
CV cross-validation
DI dissoconche I
DII dissoconche II
DERL2 derlin-2
DET differentially expressed transcript
DHA docosahexaenoic acid
DNMT1 DNA (cytosine-5)-methyltransferase 1
DPF day post-fertilization
dsRNA double stranded ribonucleic acid
E egg
EBDBN empirical Bayes dynamic Bayesian network
ECM extracellular matrix
EFA essential fatty acid
EPA eicosapentaenoic acid
EL ensemble learning
ERK extracellular signal-regulated kinase
EST expressed sequence tags
FA fatty acid
FDR false discovery rate
FFA free fatty acid
FCM fuzzy c-means
GAM generalized additive models
GATAD1 gata zinc finger domain-containing protein 1
GEO gene expression omnibus
GLM generalized linear models
GM2A ganglioside gm2 activator
GNBP gram-negative binding proteins
GO gene ontology
GPX glutathione peroxidase
GRN genetic regulatory network
GST glutathione-S-transferase
GTPase guanosine triphosphatase
GWAS genome-wide association study
HISTH2A histone H2A
HSP heat shock protein

Hsp70 heat-shock protein 70
IRAK IL-1 receptor-associated kinase
ISMER Institut des Sciences de la mer de Rimouski
J juvenile
JAK janus kinase
JAK/STAT janus kinase/signal transducer and activator of transcription
JC cocktail-reared juvenile
JNK c-jun amino-terminal kinase
JPF jours postfertilisation
jpf jours postfécondation
JT T-Iso-reared juvenile
IkB inhibitory protein kB
KEGG Kyoto encyclopedia of genes and genomes
KDP kazal domain-containing peptide
lasso least absolution shrinkage and selection operator
LBP/BPI LPS-binding/bactericidal-permeability increasing
LDL low density lipoprotein
LEMAR Laboratoire des sciences de l'Environnement MARin
LGBP LPS- and B-1,3-glucan
LITAF lipopolysaccharide-induced tumor necrosis factor-alpha factor-like protein
LNA linolenic acid
LOA linoleic acid
LPS lipopolysaccharide
LRR leucine-rich repeat
MACP membrane attack complex and perforin
MANOVA multiple analyses of variance
MAPK mitogen-activated protein kinase
MBL mannose-binding protein
Med6 mediator of RNA polymerase II transcription subunit 6
MERP1 mammalian ependymin-related protein 1
MF molecular function
MG-RAST Metagenomics RAST
ML Machine Learning
MRMR minimum redundancy maximum relevance
MMC modulated modularity clustering

MPSS massive parallel signature sequencing
MUFA monounsaturated fatty acid
MyD88 myeloid differentiation primary response protein 88
N népioconche
NA non annotated
NCBI National Center for Biotechnology Information
ncRNA non-coding RNA
NGS next generation sequencing
NFkB nuclear factor-kB
NLR NOD-like receptor
NRARP notch-regulated ankyrin repeat-containing protein
NO nitric oxide
NOAH-2 nompA-homolog
ORF open reading frame
PAMP pathogen-associated molecular pattern
P pediveliger
PC cocktail-reared pediveliger
PCA principal components analysis
PGN peptidoglycan
PCR principal components regression
PGRP peptidoglycan recognition protein
PIP5K1A phosphatidylinositol 4-phosphate 5-kinase type-1 alpha
PIR Protein Information Resource
PL phospholipid
PRDX peroxiredoxin
PRR pattern recognition receptor
PT T-Iso-reared pediveliger
PUFA polyunsaturated fatty acid
RDA redundancy analysis
RF random forest
Rho1 RAS-like GTP-binding protein Rho1-like
RMSE root-mean-square error
ROI reactive oxygen intermediate
ROPN1L ropporin-1-like protein
ROS reactive oxygen species

RRF regularized random forest
PACRG parkin coregulated gene protein
PBX1 pre-B-cell leukemia transcription factor 1 isoform 2
PDI prodissoconche I
PDII prodissoconche II
PLS partial least squares
PTPN13 tyrosine-protein phosphatase non-receptor type 13
PUFA polyunsaturated fatty acid
RLR rig-like receptor
RNAi acide ribonucléique interférence
RT-q(PCR) reverse-transcription quantitative polymerase chain reaction
SeGPX selenium-dependent glutathione peroxidase
SAGE serial analysis of gene expression
SAM significance analysis of microarray
SCP sarcoplasmic calcium-binding protein
SD standard deviation
SEM standard error of the mean
SFA saturated fatty acid
SMARCD1 swi/snf-related matrix-associated actin-dependent regulator of chromatin subfamily d member 1
SNP single nucleotide polymorphism
SOD superoxide dismutase
SPE substance polymérique extracellulaire
SR scavenger receptor
SSH suppression subtractive hybridization
ST sterol
STAT signal transducer and activator of transcription
SVM support vector machines
SWI/SNF swItch/sucrose non fermentable
T trochophore
TAG triacylglycerol
TEP thioester-containing protein
TF family transcriptional regulator
TGF transforming growth factor
Ti T-Iso diet
TIMP tissue inhibitor of metalloproteinase

TIR toll/interleukin 1 receptor
TIRAP adapter-like protein
TLR toll-like receptor
TNF tumor necrosis factor
TNKS2 a chain human tankyrase 2 - catalytic parp domain
TR family transcriptional regulator
TRAF tumor necrosis factor receptor-associated factor
TSS transcription-start site
Tukey's HSD Tukey post hoc honestly significance difference
UBN1 ubinuclein-1
UniProt Universal Protein Resource
UV UltraViolet
V veliger
VC cocktail-reared veliger
VT T-Iso-reared trochophore
WNT16 protein wnt-16
YWHAZ 14-3-3 protein zeta

Première partie

INTRODUCTION GÉNÉRALE

1

INTRODUCTION

Ici, il faut courir aussi vite que tu peux pour garder la même place. Si tu veux te déplacer, tu dois courir au moins deux fois plus vite. Cette déclaration par la Reine Rouge à Alice dans le conte *Through the Looking-Glass* de Lewis Carroll illustre le principe de la *sélection naturelle*[1]. Un être vivant doit s'adapter pour survivre aux forces de la contre-sélection. Depuis l'éclosion jusqu'à la phase de précompétence métamorphique, les larves ont développé un système morphogénétique dynamique qui s'adapte aux forces variables de l'environnement afin de se métamorphoser en adultes capables de transmettre leurs gènes. L'évolution de la larve est discutée ici compte tenu de l'effet des forces et signaux environnementaux sur la variation fonctionnelle de la réponse génomique. L'interaction entre environnement et gènes est à la base de la plasticité phénotypique des organismes, associée à une manifestation différentielle anatomique, morphologique et comportementale. Ainsi, l'adaptation locale[2] des larves est variable en fonction du flux interpopulationnel des gènes, sachant que la sélection naturelle contribue à la rétention de la structure fonctionnelle des génomes avec une plus grande *valeur sélective*[3]. Les effets sélectifs de l'environnement et la diversité des structures génétiques pourront donc servir à une évolution génomique intraspécifique qui représente l'axe central des problématiques de cette thèse chez la moule bleue, *Mytilus edulis*. L'éco-

1 *La sélection naturelle* : la thèse de Charles Darwin (1809-1882) explique l'évolution en terme de conflit continu entre les espèces où seulement un nombre limité survivra à cause de leur *fitness* ou le succès de leur valeur sélective pour se reproduire [Darwin, 1872].

2 *Adaptation locale* : la dynamique du génome fonctionnel à amener l'organisme à extraire la meilleure partie des conditions locales.

3 *Valeur sélective* : en anglais *fitness*, c'est la mesure de l'avantage sélectif d'un individu ou un groupe apparenté, par comparaison des vitesses de diffusion de deux gènes (via la reproduction) au niveau du groupe. Les mesures prennent en compte des composantes de la valeur sélective, telles que la fécondité, la longévité, la taille, le taux d'acquisition et d'allocation des ressources énergétiques et le succès reproducteur [Newkirk, 1980].

logie évolutive tente de regrouper les différentes explications pour la question *pourquoi et comment les organismes évoluent-ils de façon à maximiser leur succès reproducteur* ?

Cette question pourrait être discutée sur plusieurs thématiques, telles que l'écophysiologie, le comportement, la génétique, la biologie moléculaire ou par analyse des traits de l'histoire de vie. Les principales notions argumentées dans ce texte sont donc résumées dans le Tableau 1.1 page 6. L'effet des facteurs trophiques joue un rôle important dans l'adaptation locale des larves de la moule bleue. Par exemple, les acides gras essentiels provenant de la nourriture peuvent améliorer la croissance et le taux de survie des larves de *M. edulis*, quoique, les mécanismes moléculaires qui sont à la base de cette plasticité phénotypique sont peu connus. D'ailleurs, la dynamique du génome qui est à l'origine de la régulation de ces mécanismes n'a jamais été étudiée pendant l'ontogenèse larvaire de *M. edulis* à l'échelle de tout le transcriptome. Bien que nous ne connaissons pas comment les facteurs environnementaux trophiques influencent la plasticité génomique, des hypothèses qui étudient la cause d'une telle relation ont montré la complexité de la nature du phénotype et du génotype. Pour cette raison, je présente les effets significatifs environnementaux sur le phénotype larvaire et je propose des avantages et des désavantages sélectifs de la *morphogenèse fonctionnelle*[4] chez les bivalves marins. En effet, les propriétés du génome seront ainsi définies dans le cadre de la régulation dynamique des gènes et de leurs effets irréversibles sur l'organisme. Les interactions géniques sont donc sélectivement corrélées aux effets environnementaux, seulement afin d'améliorer leur valeur sélective inclusive et par conséquent contribuer à l'adaptation locale. De ce fait, ce premier chapitre introduit l'importance des facteurs environnementaux trophiques sur la valeur sélective des gènes pendant l'ontogenèse larvaire.

4 *Morphogenèse fonctionnelle* : regroupe les formes physiologiques successives des larves qui sont préférentiellement conservées parce qu'elles favorisent l'adaptation locale en réponse à des pressions sélectives différentes.

1.1 ÉTAT DES CONNAISSANCES SUR LE DÉVELOPPEMENT DES INVERTÉBRÉS MARINS

Les *modèles de reproduction* peuvent être classés en trois différents types de développement : planctotrophique, lécithotrophique et direct[5]. Ces modèles peuvent être quantifiés par une estimation de l'abondance relative des descendants et de leurs distributions latitudinales. Des mesures ont été développées pour la caractérisation de la dispersion larvaire (pélagique *vs* benthique), les modes trophiques (larve planctotrophe *vs* lecithotrophe) et le soin parental (incubation *vs* libération à grande échelle des gamètes mâle et femelle dans la colonne d'eau) [Vance, 1973]. Chacun de ces modèles de développement larvaire est caractérisé par des compromis sélectifs.

Les *larves planctoniques* (pélagiques) réfèrent à un investissement parental relativement faible et en vue de leur nombre élevé favorisent la dispersion de l'espèce [Levin, 2006]. Cependant, elles subissent souvent une forte mortalité en réponse à une prédation intense (temporaire dans le plancton) et une dilution alimentaire dans la colonne d'eau [Rumrill, 1990]. D'autre part, les larves issues d'un *œuf lécithotrophe*, sont caractérisées par une indépendance trophique non influencée par les ressources nutritives fluctuantes dans le plancton [Hines, 1986], quoique, la taille de l'œuf lécithotrophe est corrélée avec une allocation élevée en énergie et une baisse de la fécondité parentale [Ramirez Llodra, 2002; Marshall and Bolton, 2007]. La prévalence de gros œufs est la conséquence de l'adaptation parentale aux courtes périodes de reproduction (par exemple dans les zones polaires) pour protéger la descendance. Il en existe une mortalité réduite corrélée à un nombre limité de descendants [Thorson, 1950].

5 *Développement direct* : sans phase intermédiaire larvaire.

Tableau 1.1: **Résumé des notions discutées dans le texte qui justifient le choix des hypothèses de la thèse et la démarche entreprise pour les tester.**

Paramètres	Description
Ontogenèse	**Quels sont s'il y a lieu, le ou les réseaux géniques impliqués dans le développement chez les larves ?** (i) La croissance est la progression du cycle cellulaire une fois la cellule différenciée. (ii) La différenciation cellulaire est partie intégrante du processus de développement ontogénique. (iii) Le développement chez les larves est caractérisé par une apparition d'organes larvaires fonctionnels, puis progressivement ils seront remplacés par des organes juvéniles/adultes après dégradation et remodelage tissulaire.
Plasticité phénotypique	**Quels sont les mécanismes qui sont à la base de la croissance et de la mortalité chez les larves ?** (i) Les transitions morphophysiologiques du cycle de la vie sont chacune adaptées à des phases (a) planctonique et (b) benthique. (ii) Le succès métamorphique est variable en fonction de la vitesse de la transition pélago-benthique, le budget énergétique dépendant des réserves énergétiques préférentiellement acquises (acides gras essentiels *vs* lipides totaux) et la tolérance aux agressions bactériennes, la disponibilité trophique et la fuite aux prédateurs. (iii) Un succès métamorphique équivaut à une survie et probablement une chance de transmission des gènes responsables de ce succès.
Adaptation locale	**Quels sont les paramètres qui permettent de comprendre le but et la fonction de la croissance et l'apoptose ?** (i) Le flux de gène est variable en fonction de la qualité et du nombre des descendants. (ii) Un compromis énergétique est nécessaire pour assurer un succès reproducteur sachant qu'il peut être corrélé à la disponibilité trophique. (iii) L'adaptation morphophysiologique contribue à une sélection génomique dépendemment des pressions sélectives environnementales de populations résistantes ou sensibles. (iv) La sélection naturelle des populations repose sur la valeur sélective inclusive des gènes de croissance et de mortalité.
Plasticité génomique	**Pourquoi les précurseurs des eicosanoïdes peuvent-ils améliorer la croissance ?** **Pourquoi les précurseurs des eicosanoïdes peuvent-ils entraîner une susceptibilité aux agressions ou une amélioration des chances de survie ?** **Quels sont les gènes de croissance et de mortalité chez *M. edulis* ?** (i) Activation de stimulants nécessaires pour l'initiation des transitions morphophysiologiques. (ii) Activation d'inhibiteurs de la phase larvaire a- activités essentielles pour la dégradation des organes larvaires et la différentiation des organes juvéniles/adultes et b- régulateurs géniques modulaires associés à des mécanismes interconnectés à la base de la tolérance cellulaire, du métabolisme et de l'apoptose.

Le modèle de développement larvaire est déterminant pour la variabilité et la flexibilité des stratégies adaptatives morphophysiologiques des histoires de vie des invertébrés marins (Figure 1.1 page 10). Gunnar Thorson a décrit les modèles de reproduction et les catégorisations de base des modèles de développement [Thorson, 1950]. De nouvelles catégories n'ont pas encore été proposées, néanmoins de nouveaux types de larves peuvent présenter un différentiel de traits morphologiques observés dans les catégories originales [Mileikovsky, 1971]. La *larve démersale* par exemple, peut exploiter les environnements détritiques riches et les ressources organiques dissoutes dans les profondeurs marines en échappant à quelques problèmes de dispersion [Pearse, 1969] et la *larve téléplanique planctotrophe* peut être transportée sur de longues distances via les courants océaniques [Scheltema, 1971]. Malgré leur ubiquité, plusieurs généralisations *Thorsoniennes* ont été remises en cause [Hines, 1986], parce que la variabilité et la flexibilité des phases de développement des modèles de l'histoire de la vie défient la simple catégorisation [Marshall and Keough, 2007]. Par exemple la variabilité de la corrélation entre l'énergie et la taille de l'œuf. Les grands œufs d'échinodermes ne comportent pas nécessairement plus d'énergie probablement à cause d'un effet d'hydratation. Effectivement, des concentrations élevées en matière organique peuvent être rencontrées dans les petits œufs [Villinski et al., 2002], comme des concentrations élevées en matière inorganique dans les grands œufs [Lawrence et al., 1984; Strathmann, 1987]. Pour ces raisons, les modèles de développement diffèrent dans les transitions morphophysiologiques en fonction des pressions sélectives et les réponses génomiques dépendent de l'interaction entre le phénotype et l'environnement (Chapitre 5 de [Fox et al., 2001])

1.1.1 *L'ontogenèse des bivalves marins*

Le différentiel ontogénique[6] des larves est inhérent aux modes d'organisation structurelle des organes, de la modulation polygénique phénoménologique[7] et du changement d'habitat. Bien que l'ensemble contribue à une adaptation locale, il entraîne une élimination des génomes à valeur sélective réduite. L'ontogenèse est donc représentative du développement progressif d'un organisme depuis sa conception (fécondation) jusqu'à un stade juvénile mature (Figure 1.1 page 10). Chez la moule bleue, le développement larvaire et juvénile constituent deux périodes de différenciation anatomique et structurelle qui se chevauchent et se succèdent pour former un *assemblage morphologique*[8] nécessaire pour la transition larvaire en une forme juvénile/adulte.

L'histoire de la vie larvaire d'une moule est divisée en trois phases : la larve trocophore, véligère et pédivéligère et l'ensemble de leurs changements physiologiques peut être visualisé à la Figure 1.1. La larve prémétamorphique (*planctonique*) utilise ses réserves énergétiques protéiques (barres noires), lipidiques (barres grises foncées) et en carbohydrates (barres grises claires) pour maintenir sa viabilité (triangle brun clair). La viabilité est analysée en fonction de la taille et du nombre de juvéniles produits (la description de l'analyse se trouve dans la Section 4.5.5 page 164). La taille des larves (ligne verticale pointillée) est indiquée pour chaque phase de développement et exprimée en µm, avec le nombre de jours postfécondation (jpf, y est une variable d'âge constante mesurée en fonction du nombre de mois). La longueur totale des flèches indique la *taille théorique* que les larves peuvent acquérir à la fin de chaque stade de développement. Le nombre de juvéniles est mesuré par rapport à la concentration en ADN qui est corrélée positivement au nombre observé d'individus vivants.

---

6 *Différentiel ontgénique* : la succession des différentes formes physiologiques et les traits fonctionnnels et conséquentiels pendant le développement d'un organisme

7 *Modèle phénoménologique* : modèle d'étude de plusieurs types de structures expérimentales à la recherche de leur intentionnalité (Edmund Husserl).

8 *Assemblage morphologique* : en anglais *body plan* est un regroupement de plusieurs caractères physiologique et phénotypique, caractéristique d'une espèce ou d'un groupe taxonomique.

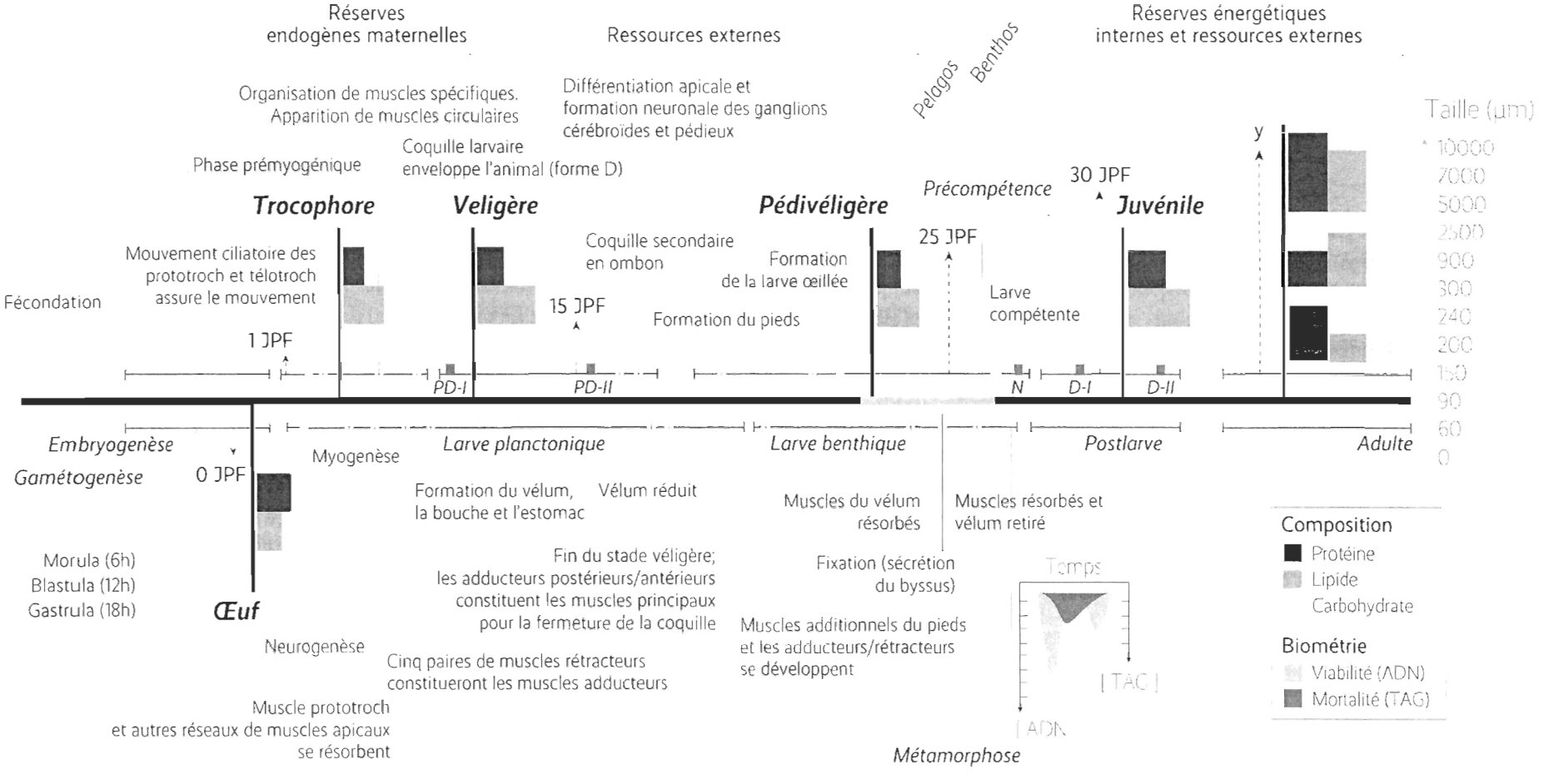

Figure 1.1 : **Schéma synoptique de l'ontogenèse de *Mytilus edulis* : stades de développement différentiels, composition biochimique et réserves énergétiques associées.**

Figure 1.1: (*Continuée*) Le nom des stades larvaires est indiqué à côté du profil énergétique qui comporte une mesure par rapport au poids sec de l'animal et de la concentration en protéines, lipides et carbohydrates. Les mesures énergétiques ont été adaptées de plusieurs travaux publiés, certains détaillés dans le texte. Les lignes diagonales pointillées représentent le changement de milieu de la larve. Les lignes verticales représentent un changement physiologique. La taille maximale des flèches verticales pointillées correspond à la taille maximale que la larve peut acquérir à la fin du stade relatif. La barre grise horizontale (~25 jpf) indique le début et la fin de la période de précompétence métamorphique. La taille y varie en fonction de l'âge de l'adulte. La biométrie comprend la viabilité des larves reportée par rapport à la concentration du milieu en ADN. Un nombre élevé de larves entraîne une concentration élevée en ADN. Il en est de même pour la concentration du milieu en TAG, qui illustre le taux de mortalité reporté. Pour un suivie détaillé du développement neuronal [Voronezhskaya et al., 2008] et musculaire [Dyachuk and Odintsova, 2009] du genre *Mytilus* pendant la métamorphose [Bayne, 1971]. Plus d'informations sur la mobilisation des réserves énergétiques peuvent être trouvées dans [Whyte, 1992]. Prototroch, constitution de cellules ciliées équatoriales dans la région apicale de la larve trocophore. Télotroch, constitution de cellules ciliées dans la région basale. jpf, jours postfécondation ; PDI, prodissoconche I ; PDII, prodissoconche II ; N, népioconche ; DI, dissoconche I ; DII, dissoconche II ; TAG, triacylglycerol.

1.1.1.1 *Plasticité phénotypique*

La machinerie de transformation est responsable de l'assemblage des traits moléculaires, cellulaires, morphologiques et génétiques, nécessaires à l'adaptation et à la résistance face aux changements successifs d'habitat du pelagos au benthos depuis l'embryogenèse jusqu'à la postlarve. L'oogenèse et l'embryogenèse des moules bleues commencent par le développement de la morula (~6 h), de la blastula (~12 h), puis de la gastrula (~18 h) [Cao et al., 2004]. La larve trocophore (~24 h) est pélagique (dispersive). L'apparition d'une coquille embryonnaire (rudimentaire) est associée à la larve véligère prodissoconche I (PDI). PDI et prodissoconche II (PDII) correspondent à deux phases de croissance distinctes du développement planctonique, séparées par une stratégie de croissance et un changement de sculpture [Jablonski, 1986]. Le potentiel de développement est basé sur la capacité de la larve précompétente à se métamorphoser[9] en réponse à une interaction de signaux externes et internes [Bayne, 1965]. L'initiation de la compétence est marquée par la formation de la *larve œillée* (ou *eyespot*) [Evseev et al., 2011]. Le nombre de jours nécessaire pour que l'individu progresse dans son développement dépend directement de la température et de la nature/concentration des nutriments dans la colonne d'eau [Laing et al., 1987; His et al., 1989; Widdows, 1991; Kheder et al., 2010]. Ainsi, la *plasticité du développement* est potentiellement variable en fonction de prédicteurs environnementaux et de leurs interactions avec le modèle dynamique[10] des génomes [Bertram and Strathmann, 1998].

9 *La métamorphose* : l'histoire de vie des invertébrés à *développement indirect* requiert l'initiation d'un processus de morphogenèse métamorphique, la métamorphose responsable de la transition de la larve en juvénile et de l'état végétatif en une forme reproductive (Figure 1.2 page 15).

10 *Modèle dynamique* : le rôle fonctionnel de l'ADN associé au potentiel à coordonner les processus biologiques, moléculaires et cellulaires. Trois aspects sont donc distingués : (i) caractériser les éléments fonctionnels de l'ADN qui peuvent être transcrits en protéines (ii) cribler le génome pour déterminer son potentiel fonctionnel, c'est-à-dire le nombre de protéines et (iii) identifier les éléments régulateurs de la chromatine.

1.1.1.2 *Précompétence métamorphique*

Pendant la prémétamorphose de la forme œillée (~23 jpf), les larves sont susceptibles à un taux croissant de mortalité (la courbe en brun foncé) représenté par la concentration en triacylglycerol (TAG) [Pernet et al., 2004; Gireesh et al., 2009]. Le TAG est un lipide neutre représentatif des réserves énergétiques disponibles. Un taux de mortalité élevée est associé à une diminution de la quantité d'ADN (donc un plus petit triangle) et une réduction de la quantité de TAG (donc une courbe réduite). Les larves qui survivent après le remodelage métamorphique continuent leur différenciation en juvéniles caractérisés par une succession des formes de coquilles : népioconche (N), dissoconche I (DI) et dissoconche II (DII) pour développer finalement la forme adulte. Cette *plasticité phénotypique* est donc associée à (i) la différenciation des traits juvénile/adulte (ii) la dégénérescence des structures larvaires (iii) la fixation (~23 jpf) (iv) l'initiation de la compétence (~25 jpf) (v) la métamorphose (~26 jpf) et (vi) le changement de leur habitat (larves planctoniques *vs* larves ou juvéniles benthiques) [Widdows, 1991].

1.1.1.3 *Interactions écophysiologiques morphofonctionnelles*

Au cours de la prémétamorphose quand les réserves énergétiques commencent à s'épuiser (Figure 1.1 page 10), la métamorphose peut s'initier indépendamment des facteurs qualitatifs physiques et des constituants biologiques d'une surface [Marshall and Keough, 2007; Tremblay et al., 2007b]. Effectivement, la *métamorphose* est un engagement irréversible associé à une transformation tissulaire importante associée à des *signaux environnementaux*[11] ou physiologiques [PAWLIK, 1992; Lau et al., 2005; Zardus et al., 2008; Hadfield, 2011; Thiyagarajan, 2010]. Pour cette raison, la métamorphose est un accélérateur évolutif de la transition larvaire en juvénile [Strathmann, 1993]. La capacité des larves pélagiques à survivre et maintenir une compétence métamorphique en absence de signaux environnementaux est

11 *Signal environnemental* : représente la qualité de l'habitat, par exemple les signaux physiques ou chimiques dérivant de la nature du substrat, indicateurs spécifiques de l'habitat.

néanmoins observée et connue pour être associée à un régime convergent adaptatif, c'est-à-dire la présence entre autres d'une nourriture adéquate [Hadfield et al., 2001]. Ce délai de métamorphose peut durer de plusieurs jours à quelques semaines [Bayne, 1965; Bishop et al., 2006b]. La rétention correspond à la sélection physicochimique et biologique de l'habitat (Section 1.3.3.3 page 36) nécessaire pour produire un phénotype avec une corrélation optimale entre la valeur sélective et l'adaptation locale des larves [Pechenik, 1990b, 2006].

Bien que la larve pélagique peut se fixer et se métamorphoser en fonction du temps qu'elle passe dans l'état de compétence, l'activation de régulateurs géniques est nécessaire pour stimuler la restructuration et inhiber les gènes de maintien de la forme larvaire [Bishop et al., 2006b]. Cette hypothèse a été avancée par Chia et collaborateurs [1978] qui suggèrent que ce mécanisme de régulation active des stimulateurs et inhibiteurs métamorphiques au lieu de conduire à l'éventuelle mort programmée de l'individu (Figure 1.2 page 15). La machinerie de base responsable de l'initiation de la métamorphose est probablement épigénétiquement[12] associé aux processus métaboliques lors d'un épuisement énergétique. Ainsi, plusieurs évolutions indépendantes ont pu favoriser la rétention de différents mécanismes interdépendants de l'activation de la métamorphose, parce que leur activation respective améliore la valeur sélective de l'organisme [Strathmann, 1993]. De ce fait, la forme larvaire est un phénomène dynamique dépendant des facteurs écologiques et regroupe des adaptations morphogénétiques inclusives responsables de son intégration dans un nouvel environnement [Hedgecock, 1986; Sanford and Kelly, 2011].

12 *Epigénétique* : des processus qui dépendent de l'adaptation locale et les histoires de vie individuelle altèrent l'expression génique via des modifications chimiques réversibles, transmissibles d'une génération à l'autre sans l'incorporation de mutations nucléotidiques [Gavery and Roberts, 2014].

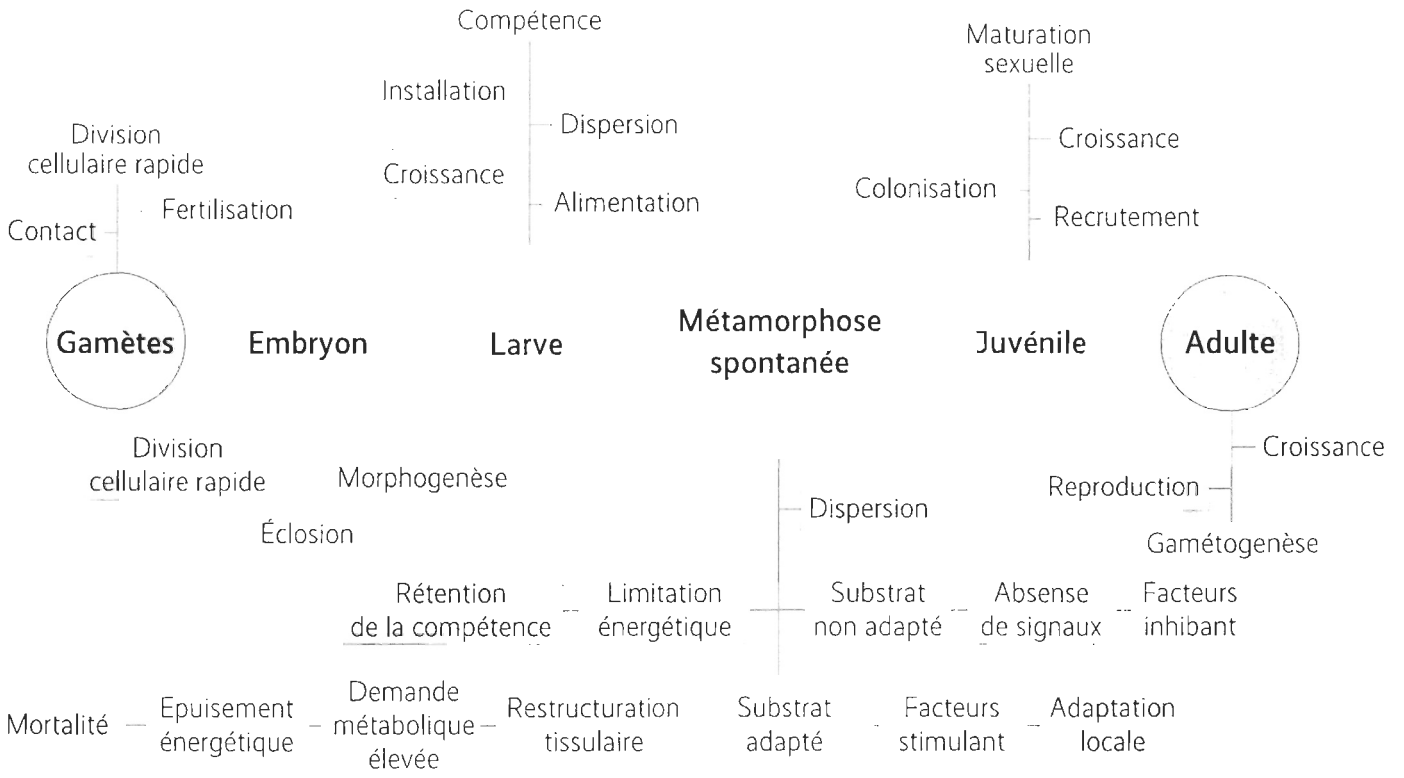

Figure 1.2 : **Comportement larvaire et spécificité des signes de l'initiation de la métamorphose.**

Figure 1.2: (*Continuée*) Les cercles représentent les différents stades du développement de l'histoire de vie larvaire et adulte de la moule bleue *Mytilus edulis*. Les branches verticales comportent les évènements moléculaires clés qui caractérisent chacun des stades. La période métamorphique est placée au milieu du schéma conceptuel. L'avantage des larves dispersives varie en fonction de la présence ou absence de signaux de compétence et d'initiation de la métamorphose (case blanche). Les cases gris foncées regroupent les différentes conséquences à la fin d'un changement physiologique. La mortalité larvaire pendant la phase métamorphique est variable en fonction de la restructuration (complète) des organes, d'un changement d'habitat, de la densité larvaire (élevée) et de la (faible) teneur en réserves lipidiques. Pour plus d'informations sur la corrélation entre la compétence métamorphique et l'environnement plusieurs hypothèses sont discutées dans Bishop et al. [2006b].

1.1.1.4 *La métamorphose*

La métamorphose est spontanée, radicale, mais manifestement coordonnée[13]. Elle implique des changements morphologiques, physiologiques et écologiques, influencés par des facteurs spécifiquement associés à la sélection de l'habitat (par exemple communication/interaction chimique entre adultes, juvéniles et larves dispersives). Quoique, la spontanéité de la métamorphose n'est pas nécessairement liée à des signaux externes, cette transition écomorphologique peut être considérée comme étant une manœuvre pour produire deux formes à valeur sélective différente à partir d'un seul génome [Bishop et al., 2006a]. La métamorphose est généralement attribuée à un sous-ensemble d'organismes, dont les insectes et les amphibiens [Truman and Riddiford, 1999], mais également quelques poissons et plusieurs invertébrés marins [Young, 1990; Bishop et al., 2006a]. L'usage du terme métamorphose en biologie comparative pourrait être également attribué à d'autres organismes incluant les levures, les plantes à fleurs et quelques espèces d'algues marines [Bishop et al., 2006a]. Les insectes holométaboles comme les diptères et les hyménoptères subissent une *métamorphose complète* (transformation radicale), tandis que les hémimétaboles comme les orthoptères acquièrent par métamorphose partielle juste une seule ou quelques structures fonctionnelles. Bien que différentes phases et modes de vie surviennent pendant le cycle de vie d'un organisme, la métamorphose se situe au centre des transitions critiques de l'histoire de vie (Figure 1.2 page précédente).

1.1.1.5 *Coordination génique modulaire*

La transition des phases écomorphologiques est manifestement coordonnée chez les espèces à métamorphose spontanée, mais ceci ne signifie pas que la machinerie de régulation génétique de base est héritable [Heyland and Moroz, 2006]. Effectivement, les régula-

13 *Coordination métamorphique* : la métamorphose n'est pas seulement associée aux changements physiologiques de l'organisme, mais aussi à des changements écologiques et comportementaux où le choix de l'habitat est souvent corrélé aux modifications de la forme larvaire [Hadfield et al., 2001] (Figure 1.1 page 10).

teurs de gènes ont pu probablement moduler chez l'ancêtre commun de multiples transitions avantageuses à la survie de l'espèce, ainsi ils ont été retenus dépendemment de la pression sélective. D'ailleurs, l'importance des voies de signalisation[14] a probablement apparue chez les organismes originellement indépendants grâce à un avantage sélectif des traits homoplasiques[15] [Hodin, 2006]. Les réseaux de régulation transcriptionnelle sont susceptibles d'assister l'émergence de nouveaux processus plus complexes du développement [Schlosser, 2002], voire un raccourcissement de la phase de restructuration larvaire [Hodin, 2006]. Pour ces raisons, les larves affrontent un défi évolutif continu, celui de coordonner le niveau de complexité de deux potentiels régulatoires (i) spécifiques à l'apparition des structures larvaires et (ii) la maturation puis le maintien des structures juvéniles/adultes (Figure 1.1 page 10).

1.1.2 *Régulation polygénique phénoménologique*

L'étude ontogénique permet de caractériser le lien entre les changements physiologiques successifs et le programme génétique qui code pour des traits écophénotypiques sélectifs chez les larves et juvéniles (Section 1.1.1 page 8). La plasticité génomique[16] désigne un champ ou une fenêtre d'expression au sein duquel les gènes peuvent opérer et s'adapter efficacement afin d'améliorer leur valeur sélective [Martin et al., 2011]. Plusieurs mécanismes moléculaires semblent être impliqués en période périmétamorphique[17] ou pendant un déséquilibre du fonctionnement de l'organisme (détaillés dans les Figures 2.1 page 51 et 2.2 page 58) afin de rétablir les processus de maintien de la croissance, la défense contre les

14 *Voie de signalisation* : ensemble d'interactions moléculaires intracellulaires qui s'enchaînent en cascade et sont déclenchées en réponse à la fixation spécifique d'une molécule -ligand- sur son récepteur oligopeptidique membranaire.

15 *Homoplasie* : une similitude interspécifique de traits écocomportementaux ne provenant pas d'un ancêtre commun.

16 *Plasticité génomique* : En présence de facteurs environnementaux particuliers, une sélection adaptative dépend prioritairement sur l'efficacité plastique d'un génome à produire une alternance de formes physiologiques, morphologiques et comportementales [Fraser et al., 2014].

17 *Périmétamorphose* : période temporelle qui regroupe successivement la phase d'initiation de la compétence larvaire prémétamorphique, la métamorphose et la phase postlarvaire postmétamorphique.

agents pathogènes et l'*homéostasie lipidique*[18] [Genard et al., 2012]. Par conséquent, le moment d'initiation de la métamorphose peut différer chez les individus au sein d'une même espèce [Paulet et al., 1988; Pechenik, 1990b]. Cette variation intraspécifique peut être une des raisons expliquant la *diversité* présente parmi les populations [Vogler and Desalle, 1994; Wang et al., 2004].

La cinétique génique différentielle des processus métamorphiques peut être fortement variable en fonction des espèces et des populations. Pourtant, des similitudes peuvent apparaître entre les transcrits (ARN) impliqués dans la réponse de l'organisme aux perturbateurs, à l'immunité et à la mort cellulaire. L'effet additif d'un évènement multistress peut être expliqué par une régulation polygénique qui est à la base d'une coordination de plusieurs mécanismes [Brazhnik et al., 2002]. Chez le pétoncle, la combinaison des facteurs d'acidification et d'hypoxie réduit davantage l'adaptation des larves prémétamorphiques et chez la palourde le niveau de sensibilité de la croissance des larves peut varier en fonction du temps de métamorphose [Gobler et al., 2014]. Certaines hormones et radicaux libres par exemple, jouent un rôle central dans de différentes voies de signalisation et mécanismes métaboliques stress-dépendants [Hodin, 2006]. Ces biomolécules ont été décrites comme régulateurs de la compétence métamorphique chez les larves d'huîtres *Crassostrea gigas, Ostrea edulis* et de moules *Mytilus galloprovincialis* [Beiras and Widdows, 1995; García-Lavandeira et al., 2005], où leur fonction est probablement corrélée à l'efficacité de la croissance des larves avec un soutien différentiel des gènes de la digestion et de l'assimilation des nutriments et liés à l'allocation énergétique [Meyer and Manahan, 2010]. La formation de la coquille chez les larves d'huîtres *C. gigas* nécessite la participation de gènes qui codent pour des protéines structurales comme le collagène [Zhang et al., 2012a]. Ces biomolécules structurales sont impliquées aussi dans l'angiogenèse[19] et sont associées à la matrice extracellulaire (Chapitre 3 page 83 ou les Sections 3.4 page 87 et 3.6 page 117). Un grand nombre de gènes sont impli-

18 *Homéostasie lipidique* : dépend entre autres des réserves énergétiques nécessaires pour la survie postlarvaire.

19 *Angiogenèse* : c'est la néovascularisation à partir de vaisseaux préexistants. Les larves de moules ne possèdent pas de vaisseaux sanguins, donc les gènes impliqués dans ce mécanismes seront probablement codant pour des protéines de transport, ou une autre fonction similaire.

qués dans les mêmes voies d'apoptose et de défense immunitaire, quoique, chez l'huître leur expression s'est avérée interconnectée et plutôt différentielle dépendemment de la nature de la perturbation (Figure 3 de [Zhang et al., 2012a]). Les interactions gène-gène que nous avons étudié dans les Chapitres 2 et 3 sont des processus complexes hautement interconnectés. Ils sont également indispensable pour coordonner une régulation sélective [Alon, 2007] tout le long de l'ontogenèse. Leur implication dans les voies de signalisation et les mécanismes métaboliques sont peu étudiés pendant le développement larvaire.

1.1.3 *Mécanismes moléculaires phénotype-dépendant*

Les processus génomiques et physiologiques sont à l'origine du phénotype et de la distribution d'une espèce dans un habitat [Kingsford et al., 2002]. La croissance peut constituer un proxy déterminant de la valeur sélective [Giménez, 2010; Rico-Villa et al., 2010]. D'autres traits biologiques pertinents, tels que la performance reproductive, la morphologie fonctionnelle larvaire et l'expression génique différentielle en réponse à des perturbateurs devraient aussi être étudié afin de mieux expliquer le lien entre la fonction génique et le maintien de l'homéostasie. Les bases moléculaires et génétiques de la précompétence métamorphique, du recrutement et de métamorphose sont peu étudiées [Heyland and Moroz, 2006], malgré le vaste nombre de travaux qui existe sur la physiologie des premiers stades chez plusieurs espèces marines [Samain, 2011; Terahara and Takahashi, 2008; Danilova, 2006]. Jusqu'à récemment, les outils d'analyses génomiques étaient peu disponibles pour traiter les données des organismes non modèles [Collins et al., 2003]. Toutefois, le développement des techniques de haut débit en transcriptomie permettra une expansion d'une nouvelle classe d'analyse expérimentale (Chapitre 2) capable de traiter les multiples aspects physico-génomiques des espèces [Wang et al., 2009b; Ekblom and Galindo, 2010].

La réponse dynamique du génome est bien visible chez les bivalves marins puisque la transition de la larve pélago-benthique en postlarve exige un différentiel génique de l'expres-

sion à l'échelle entière de l'organisme [Medina, 2009]. Effectivement, les nouvelles études transcriptomiques peuvent être appropriées pour traiter la réponse génomique en raison d'une amélioration de la détection des transcrits rares. Par ailleurs, l'effet de *dilution*[20] *de la cellule* a même été considéré comme prédicteur transcriptionnel et son influence sur l'abondance des transcrits a été récemment reportée [Alon, 2007]. Plusieurs études utilisant des techniques modernes en transcriptomique ont été publiées récemment dont une qui évalue le taux de la transcription intercellulaire au cours du développement [Sasagawa et al., 2013]. Un différentiel d'expression génique a ainsi été détecté entre les cellules de même origine et parmi les différents stades du cycle cellulaire. D'ailleurs, il serait possible de suivre avec précision les différents processus interdépendants de la morphogenèse larvaire comme la neurogenèse (Figure 1.1 page 10) et de corréler leurs effets sur la défense immunitaire et la différenciation cellulaire, deux processus clés de la phase périmétamorphique [Zhang et al., 2012a].

1.2 MÉCANISMES ADAPTATIFS ET ÉVOLUTIFS DIFFÉRENTIELS

1.2.1 *L'adaptation locale*

Plusieurs hypothèses ont déjà été proposées chez certains bivalves, gastéropodes et crustacés pour expliquer l'évolution des modèles de développement larvaire [Strathmann, 1978, 1993; Martel and Chia, 1991]. Les espèces de mollusques à larves dispersives planctotrophes ont une plus grande capacité de dispersion, un brassage génétique plus efficace et une plus longue vie que les espèces lécithotrophes ou à développement direct [Jablonski, 1986; Strathmann, 1987; Shanks and Brink, 2005; Levin, 2006]. La Figure 1.3 montre les différentes zones habitées par l'espèce *M. edulis*. Pourtant des exceptions existent chez certains groupes [Hines, 1986]. La dispersion est probablement une conséquence liée à la forme pélagique et non pas à l'investissement énergétique parental ou l'indépendance trophique des

20 *Effet de dilution* : le contenu en acide nucléaire ribosomal varie en fonction de la croissance cellulaire.

Figure 1.3: **Distribution géographique de *Mytilus edulis*** [Fisheries, 2011].

larves [Palumbi, 1994, 1992]. D'ailleurs, certaines espèces à *développement aplanctonique* ou même direct sont potentiellement capables d'une dispersion géographique et des larves de courte vie peuvent être dispersées loin de la population parentale [Dirnberger, 1993]. En outre, des stratégies de reproduction peuvent varier entre plusieurs épisodes de pontes partielles ou une seule ponte synchronisée par saison. Un phénomène reconnu pour avoir des conséquences sur la survie de la descendance et la dispersion [Fielman and Marsh, 2005; Crisp, 1978]. L'adaptation qui vise à produire de grands œufs ne confère pas seulement une stratégie efficace vis-à-vis du budget énergétique, mais aussi sur la production de plus grandes larves au moment de la fixation, par exemple chez l'huître (*Ostrea chilensis*), la palourde (*Mercenaria mercenaria*) et le pétoncle (*Argopecten irradians*) [Wilson et al., 1996; Kraeuter et al., 1981]. Les plus grandes larves prémétamorphiques de *M. galloprovincialis* présentent parfois des avantages contre la prédation et la sélection de la nourriture [Phillips, 2002].

Tableau 1.2: **Les avantages de dispersion des larves planctoniques**

	Avantages*
1	Réduction de la compétition pour les ressources avec les adultes
2	Réduction temporaire de la mortalité benthique
3	Hausse de la diversité de la structure génétique
4	Réduction de l'endogamie avec effets morphologiques et évolutifs des génomes
5	Aptitude accrue à tolérer l'extinction locale.
6	Réduction du cannibalisme par filtration des larves en suspension par les adultes
7	Ralentissement de la saturation de l'habitat
8	Colonisation de nouveaux habitats et expansion du domaine
9	Hausse du flux génique sur de longue distance
10	Fuite en cas de détérioration des conditions locales
11	Recolonisation après amélioration des conditions
12	Persistence temporelle plus longue

* Les avantages sont cités sans priorisation.

Le compromis entre le nombre et la valeur sélective de la descendance constitue un phénomène central dans l'adaptation locale qui justifie *la théorie de la décision*[21]. À court terme, la vie en groupe dépend du caractère altruiste d'un organisme qui à long terme favorise l'apparition d'un élément génétique égoïste (Section 1.3 page 26) [Hedgecock, 1986]. Les larves pélagiques sont associées à plusieurs avantages sélectifs [Nützel, 2014] résumés dans le Tableau 1.2. Pourtant, l'adaptation des mollusques pélago-benthiques semble être accompagnée par des mortalités larvaires élevées, ce qui signifie que les larves sont d'une manière ou d'une autre désavantagées [Schneider et al., 2003]. Des désavantages associés à la dispersion de la larve planctonique peuvent être donc considérés [Pechenik, 1999] et ils sont résumés dans le Tableau 1.3 page suivante. Pour cette raison, la croissance postmétamorphique et la compétitivité[22] (interindividuelle) seront probablement réduites, un phé-

21 *Théorie de la décision* : la façon rationnelle de survivre et de se reproduire dans un environnement extrême est d'étendre le risque (produire une descendance, coloniser et augmenter le nombre des gènes).

22 *La compétition sélective* : typiquement pour *échapper* à la contre-sélection, la compétition se cerne entre le prédateur et la proie [Darwin, 1872]. Le conflit sera donc une lutte interspécifique pour survivre. Ici, cette lutte sera considérée comme étant plutôt intraspécifique et donc va reposer sur la valeur sélective des gènes

Tableau 1.3: **Les désavantages de dispersion des larves planctoniques**

	Désavantages*
1	Retard de la métamorphose
2	Manque d'habitat convenable pour les postlarves
3	Mauvaise adaptation locale
4	Hausse de la susceptibilité aux perturbateurs chimiques et physiques provenant de différentes zones
5	Réduction de la valeur sélective

* Les désavantages sont cités sans priorisation.

nomène déduit d'une réduction de la valeur sélective corrélée positivement à une rétention de la métamorphose [Pechenik, 1990b; Highsmith and Emlet, 1986; Roberts and Lapworth, 2001; Rahman et al., 2014]. Il est donc important de considérer les avantages et les désavantages associés à l'évolution de la phase dispersive, le maintien de l'homéostasie cellulaire et dans certains cas de sa disparition potentielle des cycles de l'histoire de vie.

Le cycle de vie est naturellement divisé en différentes phases soumises aux pressions sélectives de chaque milieu de développement (Figure 1.1 page 10). En effet, les environnements sont aussi soumis à des modifications liées à la dynamique des populations existantes [Fox et al., 2001]. Une fois perdu, il est peu probable qu'un phénotype puisse réapparaître, en particulier quand cette perte est accompagnée par des simplifications morphologiques substantielles [McEdward and Janies, 1997]. Bien que la rétention de la progéniture près des parents est souvent observée chez les animaux coloniaux, la sélection favorise parfois la fixation de traits dispersifs réduits [Anderson, 1994; Newman and Ross, 1977]. En outre, la nature ancestrale de la phase larvaire a toujours été débattue. A l'heure actuelle, la plupart des indices et les arguments présentés tendent vers une évolution aplanctonique produite à partir d'un ancêtre planctonique [Pechenik, 1999]. Quoique, d'autres positions supportent l'hypothèse que le développement planctonique a évolué à partir d'un ancêtre aplanctonique, le mode de couvaison étant ancestral, la mise en place d'une larve dispersive correspondrait

[Gardner and Welch, 2011] entre individus de la même espèce pour *échapper* à la contre-sélection [Dawkins, 1976].

donc à une dérive éventuelle [Chaffee and Lindberg, 1986]. Effectivement, il est plus facile de perdre une structure complexe que d'en gagner une [Strathmann, 1978, 1985]. La perte de la larve dispersive pourrait parfois être réversible chez les groupes où le développement aplanctonique est juste associé à une réduction de structures superficielles larvaires. Par exemple, la réacquisition de la phase larvaire à partir d'un ancêtre aplanctonique (à couvaison) chez les gastéropodes du genre *Epheria* est possible, car la structure entière du vélum n'a jamais disparu chez l'ancêtre incubateur [Nielsen, 1998]. Au vue de ces controverses, la perte d'une structure n'est pas toujours accompagnée de la perte des gènes qui en sont responsables [Raff, 1987] et par conséquent, le redéploiement serait possible même si la structure larvaire a été considérablement réduite ou entièrement perdue. Les avancées transcriptomiques pourront établir l'ampleur de *l'état silencieux des génomes* et de leurs effets sur la plasticité développementale des larves. Ainsi, les analyses génomiques permettront de résoudre les controverses à propos des changements de direction des modèles reproductifs.

La conservation d'un allèle x_1 dépendemment de son habitat est favorisée seulement s'il provoque l'apparition d'un trait bénéfique à l'organisme. Dans le cas d'une sélection divergente selon les habitats, la transmission de x_1 sera représentée par une fréquence de sélection f_1 élevée. Par ailleurs, un allèle neutre x_2 situé en proximité de x_1 dans la même région du locus sera caractérisé par une fréquence de sélection $f_2 = f_1$ en raison d'un *auto-stop génétique*[23] [Andolfatto, 2001; Rice and Hostert, 1993; Boudry et al., 2002]. Effectivement, la *déviation génétique*[24] sera probablement plus élevée pour les loci neutres non soumis à une pression de conservation [Faure et al., 2008].

23 *Auto-stop génétique* : en anglais *genetic hitchhiking* ou *genetic draft* est un processus pseudo-aléatoire qui augmente la variabilité génétique des génomes. La connectivité allélique peut avoir des conséquences additives en cas de fixation de l'échange génétique, ou délétère sur le phénotype en cas d'une élimination de l'ensemble du produit perdu.

24 *Déviation génétique* : les forces de la sélection divergente sont responsables d'un polymorphisme fonctionnel en raison de la dégénérescence du code génétique, par exemple en conséquence d'un taux élevé de recombinaison dans les régions non codantes de l'ADN.

1.2.2 *Différences génomiques et physiologiques*

Toute population avec une meilleure valeur sélective présente une résistance élevée contre la mortalité alors que les conditions environnementales deviennent extrêmes [Hare and Cowen, 1997; Bertram and Strathmann, 1998; Phillips, 2002; Marshall et al., 2003]. Un *compromis énergétique*[25] favorise le maintien de la croissance, soutenant ainsi un développement continu[26] [Eschweiler and Christensen, 2011]. Bien qu'une induction de l'autophagie chez la larve de Drosophile par exemple entraîne une mort cellulaire programmée et localisée, l'assimilation par les cellules environnantes des nutriments libérés assure un maintien de la croissance des tissus et des structures adultes [Rusten et al., 2004].

Le phénomène de compromis énergétique est donc basé sur un mécanisme de *protection par compensation* contre une perte excessive d'énergie, par exemple pendant la gamétogenèse, en s'engageant avec une force égale contre le déclenchement de ce processus pour réduire le succès reproductif (par exemple [Olofsson et al., 2009]). L'attribution de temps et de ressources pour surproduire des descendants[27] ou pour améliorer les chances de survie des parents afin qu'ils puissent se reproduire ultérieurement, constituent respectivement des perspectifs adaptationistes de *sémelparité*[28] et d'*itéroparité*[29]. L'importance de l'effet du nombre de descendants sur l'étendue de la population des bivalves marins est également corrélée au flux de gènes qui varie en fonction du facteur de dispersion des larves planctoniques.

---

25 *Compromis énergétique* : en anglais *energetic trade-off*, c'est l'introduction d'une inégalité comportementale pour conserver la valeur sélective de l'animal sous des conditions non optimales afin de réduire le coût métabolique et par conséquent rebalancer le flux énergétique associé à la survie et à la capacité reproductive.

26 ***Note*** : l'évolution coopérative ne sera pas discutée ici. Des traits altruistes favorisent la survie des membres d'un groupe qui ne coopèrent pas suivant un compromis énergétique, mais reçoivent l'aide d'autres membres altruistes. Les individus altruistes sont dans ce cas désavantagés [Kreft and Bonhoeffer, 2005].

27 *Surproduire des descendants* : une surproduction à pour but d'augmenter la probabilité de survie de la population, sachant qu'à priori un grand nombre des descendants serait contre-sélectionné à cause d'une faible valeur sélective.

28 *Sémelparité* : un seul évènement reproductif qui peut-être associé à la mort des parents

29 *Itéroparité* : pontes partielles à différentes périodes

Chez les espèces sessiles avec un cycle de vie pélago-benthique, la larve constitue le vecteur de dispersion principal. Par ailleurs, les larves précompétentes peuvent suspendre temporairement leur recrutement et leur métamorphose [Bayne, 1965; Fielman and Marsh, 2005] (Figure 1.2 page 15). Ce retard de la métamorphose augmente la connectivité entre les diverses populations. Pour cette raison, la structure génétique inter et intraspécifique sera variable en fonction du flux des gènes provoqué par cette connexion interpopulationnelle et corrélée positivement à la dispersion des larves [Lenormand, 2002]. Or, une fécondité élevée et une grande taille des descendants[30] augmentent l'homogénéisation de la structure génétique [Hedgecock, 1986; Duran et al., 2004; Levin, 2006]. En absence des processus de décalage des pontes dans un environnement hétérogène, la dispersion pourrait réduire l'adaptation locale [Lenormand, 2002; Levin, 2006]. Les facteurs qui influencent le *polymorphisme adaptatif* d'une espèce sont donc nombreux et variables en fonction des environnements contrastés.

1.3 IMPACTS ENVIRONNEMENTAUX SUR LE DÉVELOPPEMENT (VISION DU NÉO-DARWINISME CONTEMPORAIN)

L'étude de l'histoire de la vie d'une espèce permet d'explorer les causes du mode d'adaptation et de la réponse des génomes. Chez les espèces à développement indirect, la grande majorité de la communauté larvaire est susceptible d'une mortalité différentielle variable en fonction de la pression sélective de l'environnement et de la prédation [Fielman and Marsh, 2005; Genard et al., 2012; Von der Meden et al., 2012]. La survie et la démographie larvaire peuvent être modulées par les facteurs biotiques comme la prédation, la compétition, la disponibilité trophique [Dobretsov and Wahl, 2008], la présence des pathogènes [Fielman and Marsh, 2005] et par les facteurs environnementaux physico-chimiques (abiotiques) comme la température, la salinité [Pfeifer et al., 2005; Wing and Leichter, 2011],

30 *Note* : les individus de grande taille échappent à la contre-sélection naturelle grâce à une valeur sélective élevée (Section 1.1 page 5).

l'acidité [Bechmann et al., 2011] et la pollution anthropique [Fang et al., 2009]. La plupart de ces composants biotiques et abiotiques contribuent à l'apparition d'une compétition entre les différentes structures génétiques et favorisent la sélection du *gène égoïste*[31]. Bien que l'acquisition des ressources trophiques, la recherche d'un refuge, de la lumière et d'un espace pour se reproduire peuvent être considérées comme des processus phénotypiques, la compétition sélective des génomes pour promouvoir une adaptation locale dépend à la base du gène ayant la meilleure valeur sélective inclusive[32].

1.3.0.1 *Les réplicateurs*

Quand un individu place sa vie en danger pour la protection des voisins apparentés ou non (mais porteurs de la même signature génique), il agit dans l'intérêt de ses propres gènes. Richard Dawkins [1976; 2006] dans *The Selfish Gene* a proposé le concept de *réplicateur*[33] la molécule initiale qui a réussi à se caractériser par sa fonctionnalité et son avantage supérieur vis-à-vis des autres molécules de la *soupe primordiale*. Malgré l'utilité de tous les gènes qui favorisent le bon développement de la cellule, certains sont associés à des mécanismes essentiels à la survie et à la reproduction de l'organisme. Ils sont généralement soumis à une pression sélective positive et une transmission prioritaire [Lewontin, 1970]. Pour cette raison, une signature différentielle de l'expression de certains gènes ou marqueurs peut être à la base d'un effet bénéfique ou bien défavorable sur le phénotype [Mousseau et al., 1987; Reed and Frankham, 2003]. Si le produit d'un gène participe à la résistance d'un organisme

31 *Gène égoïste* : en anglais *selfish gene*, une analogie qui essaie d'expliquer le comportement altruiste sans faire intervenir la sélection du groupe de l'évolution coopérative. Le concept se focalise sur le gène et non sur le phénotype comme étant une unité de sélection.

32 *Valeur sélective inclusive* : en anglais *organism's inclusive fitness*. La convention Darwinienne se centre sur l'individu et le phénotype [Darwin, 1872; Hamilton, 1970] pour expliquer l'adaptation, quoique, une vision alternative sur le gène égoïste et la coopération (Hamilton et Dawkins) peut être basée sur le gène comme prédicteur de l'adaptation [Hamilton, 1972; Dawkins, 2006]. Par ailleurs, Nowak et collaborateurs [2010] réfutent le concept de l'*inclusive fitness* dans les populations eusociales avec des traits altruistes.

33 *Réplicateur* : un gène qui fait partie intégrante du génome d'un être qui se reproduis et lutte pour sa survie. Le gène subit une réplication régulée indépendante du *choix* de l'organisme. Ceci est considéré comme étant un conflit d'intérêts entre le vecteur et le réplicateur.

contre une perturbation quelconque, la diffusion de ce gène sera favorisée, car son hôte survivra [Gardner and Welch, 2011]. Un exemple représentatif du gène égoïste est la sélection chez les guppys, des petits poissons à reproduction rapide. Les observations montrent que les femelles sont attirées par les contrastes brillants chez les mâles [Endler, 1991]. Toutefois, dans l'habitat naturel du guppy, ce phénotype est susceptible d'une forte prédation [Godin and Briggs, 1996]. En d'autres termes, les mâles colorés sont plus enclins à se reproduire, mais en même temps attirent davantage les prédateurs. L'importance des gènes de reproduction et des couleurs brillantes est donc corrélée positivement, mais leur conséquence sur le phénotype ne l'est pas. Bien que la transmission des gènes de guppys à la descendance ne dépend pas de la pression exercée par la prédation sur leur phénotype, une *compétition intraspécifique* maximisera la valeur sélective inclusive du gène qui à long terme aura une tendance à s'accroître en nombre [Hamilton, 1970, 1972; Grafen, 2006].

1.3.0.2 *Valeur sélective inclusive des gènes*

Malgré l'influence probablement négative de la dispersion sur la valeur sélective (Section 1.2.1 page 20), la dimension de la descendance est linéairement corrélée au flux de gène engendré [Lenormand, 2002]. La valeur sélective de ces gènes est conservée seulement si la reproduction se réalise entre les organismes qui les partagent. Par ailleurs, un *conflit*[34] pourra se produire entre les réplicateurs et leur hôte, sous l'effet d'une adaptation locale qui limite la biodiversité malgré l'enrichissement de la structure génétique par le nombre élevé de descendances. La transmission de la valeur sélective dépendra d'un réseau de gènes optimal (*Correspondence V* du Tableau 3 de [Gardner and Welch, 2011]). Par exemple, promouvoir une ponte complète, partielle ou aucune ponte chez les espèces *Mytilus* sous des conditions extrêmes, est un compromis adaptatif afin d'empêcher l'épuisement énergétique, l'affaiblissement, puis la mort de l'organisme [Petes et al., 2008]. Cette sélection

34 ***Note*** : le conflit qui survient au sein de l'organisme représente une analogie pour expliquer les répercussions de l'adaptation locale et la contre-sélection sur les génomes.

gène-dépendante concorde avec l'hypothèse que le gène favorable sera retenu puis transmis seulement s'il se révèle. Le concept du gène égoïste considère que les réplicateurs constituent une unité de sélection et que l'adaptation est centrée sur le gène non pas sur l'organisme. Par contre, certains physiologistes considèrent que l'individu est l'unité d'évolution et la sélection étant spontanée, la reproduction et la survie de l'organisme participeront à la modification de la fréquence d'expression du génome et seront à long terme les effecteurs principaux de l'évolution [Gould and Eldredge, 1977; Gould and Lloyd, 1999]. Ainsi, ils mentionnent que les gènes ne sont pas directement visibles par la sélection naturelle, mais que seul le phénotype est l'unité de sélection en vue de son interaction avec l'environnement sous l'opérateur de la sélection naturelle (particulièrement le Tableau 1 de [Gardner and Welch, 2011]). Le concept du gène égoïste est prioritairement considéré dans les prochains chapitres, ce qui permet une évaluation centrée sur le gène et non pas sur la larve comme système évolutif pour l'étude de la fonction et de l'adaptation.

1.3.1 *Sélection naturelle*

La croissance et la survie sont parfois influencées par la taille des larves [Kraeuter et al., 1981; Phillips, 2002], ce qui explique l'importance résiduelle dans l'alternance des formes larvaires et en particulier l'avantage sélectif de chacune des transitions morphologiques vis-à-vis l'adaptation locale des juvéniles. Ainsi, l'individu prépreproductible appartenant à une population constante, adopte une forme et un habitat qui lui permet d'établir un ratio de mortalité et croissance réduit [Strathmann, 1993]. La survie des larves dépend des critères de sélection d'habitats et de dispersion [Gosselin and Qian, 1997], lesquels sont variables en fonction des facteurs biotiques et abiotiques locaux. Si la larve avait seulement évolué par rapport à la sélection de l'habitat, nous pourrions supposer qu'elle serait seulement libérée lors de la phase de compétence afin de pouvoir se fixer au premier endroit convenable rencontré (Section 1.1.3 page 19). Cependant, les larves dispersives passent généralement des

semaines, voire des mois pour acquérir le stade de recrutement (période de précompétence Figure 1.1 page 10) puis initier la compétence métamorphique [Jackson and Strathmann, 1981]. Chez certaines espèces d'échinodermes, d'annélides, de gastéropodes et de bivalves [Bayne, 1965; Bishop et al., 2006b; Williams and Degnan, 2009], la larve est capable de suspendre sa métamorphose (Figure 1.2 page 15) en absence de signaux environnementaux spécifiques. Il est donc possible que la larve dispersive soit retenue dans l'histoire de vie des espèces marines parce que la meilleure façon d'augmenter la valeur sélective des gènes serait d'envoyer les descendants loin du voisinage parental [Strathmann et al., 2010], bien qu'un potentiel élevé de dispersion réduit l'effet adaptatif d'une *métamorphose sélective* (Tableau 1.3 page 23).

La compétence métamorphique est principalement considérée comme étant une condition physiologique, mais l'orchestration à la base se situe au niveau d'une valeur sélective inclusive d'un ou plusieurs réseaux de gènes corrélés ou connectés (Sections 1.1.2 page 17 et 5.3 page 191). Chez les larves de gastéropodes par exemple, le temps d'établissement de la compétence est variable en fonction de la disponibilité trophique [Pechenik et al., 1996] ou de la nature des protéines algales [Morse and Morse, 1984], des neurotransmetteurs chez les larves de moules, d'huîtres et de palourdes [García-Lavandeira et al., 2005; Beiras and Widdows, 1995], ou d'une interaction entre larves et juvéniles [Bishop et al., 2006b]. Chaque étape de métamorphose amène un tissu larvaire à un ordre de différenciation plus élevé que son état précédent (Figure 3.1 page 88). La succession de ces changements écomorphologiques facilite le passage de la vie planctonique fonctionnelle en créant une tendance sélective multifactorielle (biotique et abiotique Section 1.3) envers un milieu particulier, lequel constituera éventuellement l'habitat des juvéniles/adultes [Hadfield et al., 2001]. Comprendre les forces responsables de la répartition actuelle du développement larvaire parmi les invertébrés benthiques et l'impact potentiel des activités anthropiques sur l'évolution des modèles de reproduction nécessite l'étude des sujets suivants :

* La sélectivité des larves dispersives en fonction des forces environnementales.
* La tolérance multistress des larves et l'efficacité de la valeur sélective génique inclusive.
* La prédation et son effet sur les larves pélago-benthiques et juvéniles.
* La relation entre la taille de la larve ou du juvénile sur leur vulnérabilité vis-à-vis des prédateurs pélago-benthiques.
* Les changements morphogénétiques et leurs effets sur la mortalité larvaire.
* Le degré de détournement des pressions sélectives contre les formes larvaires.

1.3.2 *Facteurs abiotiques*

L'adaptation locale des larves peut être limitée par des phénomènes courants tels que l'eutrophisation [Brehmer et al., 2011], les mortalités estivales liées à la reproduction [Xiao et al., 2005] et la température critique [Talmage and Gobler, 2011], la réduction de la biomasse [Manahan, 1990], la mortalité des animaux benthiques [Gosselin and Qian, 1997], les épidémies et la prolifération d'espèces invasives [Kaustuv et al., 2001; Schmidt et al., 2008]. L'ensemble de ces facteurs pose une contrainte sévère à l'ontogenèse et la maturation des larves. La sensibilité des larves aux facteurs abiotiques varie en fonction de leur âge. Les individus matures sont reconnus pour une meilleure résistance et une adaptation avancée aux changements environnementaux [Thorson, 1950]. Ainsi, la tolérance des larves aux perturbateurs physiologiques est probablement variable par rapport aux caractéristiques génétiques de l'espèce [Rumrill, 1990; Luna-Gonzalez et al., 2002].

1.3.2.1 *Impact de la température et la salinité*

Le potentiel adaptatif à la température locale dépend en partie de la diversité de la structure génétique des populations [Rayssac et al., 2010]. Dans une étude récente qui com-

pare le taux de tolérance larvaire de *M. edulis* et *M. trossulus* à différentes températures, les auteurs ont remarqué une corrélation espèce spécifique positive entre la survie, la croissance, la composition en lipides et la tolérance thermique des larves D jusqu'à la phase dissoconche (Figure 1.1 page 10). Les larves de *M. Trossulus* tolèrent préférentiellement des températures comprises entre 10 et 17 °C, ce qui reflète la distribution géographique de l'espèce dans les eaux du Nord, tandis que les larves de *M. edulis* semblent avoir un préférendum thermique pour les températures plus élevées entre 17 et 24 °C (Figure 1.3 page 21). Les premiers stades ontogéniques du développement de *Mytilus* sont soumis à une *sélection thermique adaptative* [Pechenik, 1990a; Manoj Nair and Appukuttan, 2003], ainsi une augmentation de la température entre 10 et 20 °C pendant la période métamorphique favoriserait une transition rapide en postlarve [Yaroslavtseva and Sergeeva, 2006]. Pour cette raison, une maturation complète et rapide dépendra probablement d'un rapport entre l'adaptation sélective des gènes, la synchronisation de la morphogenèse métamorphique et le niveau de résistance aux conditions environnementales.

Le réchauffement journalier et les précipitations atmosphériques entraînent la formation superficielle de différentes microcouches de températures et salinités. Les organismes qui résident au niveau de ces couches seront donc soumis à un impact continu probablement nuisible [Dam, 2013]. En raison des circulations océaniques horizontale et verticale qui créent des déplacements sur de longues distances et profondeurs, les larves doivent survivre aux fluctuations relatives de la température et la salinité. Bien qu'une réduction de la salinité peut provoquer un effet indésirable sur les larves [Qiu et al., 2002; Yaroslavtseva and Sergeeva, 2006], les blastulas in vitro, les trocophores et la phase précoce de la larve D, chez certaines espèces du genre *Mytilus*, se rassemblent au niveau du film supérieur de la surface d'eau [Yaroslavtseva and Sergeeva, 2003] et s'adaptent à long terme à la dessalure. Les trocophores cependant peuvent parfois exhiber une adaptation réduite par rapport aux blastulas et aux véligères. Ainsi, les transitions des formes larvaires peuvent être chacune unique à cause des pressions sélectives spécifiques aux différentes phases de développement

comme les variations thermiques et de salinité [Koehn et al., 1980; Luedeking and Koehler, 2004; Lockwood and Somero, 2011].

1.3.2.2 *Facteurs anthropogénétiques (impacts métalliques et organiques)*

Les polluants environnementaux affectent en particulier les organismes qui n'ont pas encore acquis un certain niveau de maturation pour activer spontanément la défense immunitaire et la détoxication [Tirapé et al., 2007; Fielman and Marsh, 2005; Weng and Wang, 2014; Mohammed, 2013]. Naturellement, ils peuvent s'accumuler dans les cellules ou altérer le fonctionnement du métabolismes, comme le métabolisme lipidique [Perrat et al., 2013], mais aussi leurs effets toxiques peuvent être transmis au phytoplancton qui constitue la nourriture primaire des larves [Basti et al., 2013]. Des polluants comme le plomb, le cadmium, le zinc, le mercure et le 4-nonylphénol ont attiré l'attention de la communauté scientifique à cause de leurs effets directs sur les phases larvaires, voire même leur impact vertical transgénérationnel.

Les observations reflètent souvent des corrélations entre la nature de l'espèce, les sensibilités précoce et spontanée *vs* la valeur de la toxicité chronique et les effets combinés de plusieurs *vs* l'additivité d'un seul composant [Faria et al., 2010]. Les émissions anthropiques du dioxyde de carbone (CO_2) par exemple entraînent une augmentation continue de l'acidité des eaux qui réduit le pH de la surface des océans [Pespeni et al., 2013]. Les populations des hautes latitudes vont être probablement soumises à l'effet de ce changement environnemental en raison de la solubilité élevée du CO_2 à faible température. Bechmann et collaborateurs [2011] ont montré que le pH acide à 7.6 n'avait pas d'effet marquant à court terme sur la fécondation, la durée du développement larvaire, l'apparition d'anomalie de coquille de la larve D ou sur l'alimentation des larves de *M. edulis*. Cependant, après deux mois de traitement, environ 30 % des larves avaient une taille significativement plus petite. Le nonylphénol est probablement un des polluants chimiques les plus dangereux avec des effets irréversibles sur le système endocrinien en réduisant la viabilité des organismes aqua-

tiques [Nice et al., 2000, 2003]. Malgré ses effets observés à l'échelle populationnelle telles l'inhibition de la croissance et les anomalies morphologiques [Liu et al., 2011], un effet sous-létal EC_{50}=200 $\mu g \times L^{-1}$ n'a pas affecté la fécondation ni les premières phases ontogéniques chez *M. edulis* [Granmo et al., 1989]. Pourtant, des EC_{50}=56 $\mu g \times L^{-1}$ peuvent déstabiliser la structure du byssus et l'allocation énergétique, entraînant ainsi des anomalies de croissance chez la moule bleue *M. edulis* ou une dérégulation des processus biologiques chez l'ormeau *Haliotis diversicolor supertexta* tels le métabolisme énergétique, la signalisation cellulaire, le système immunitaire et la réponse aux stress, dont l'ensemble induit un défaut métamorphique [Liu et al., 2011; Granmo et al., 1989].

1.3.3 *Facteurs biotiques*

1.3.3.1 *Induction de la compétence larvaire : le cas d'Haliotis asinina*

Williams et collaborateur [2009] ont exploré l'expression différentielle des gènes larvaires jusqu'à la métamorphose afin de caractériser les mécanismes moléculaires à l'origine du recrutement prémétamorphique. Parmi les 60 % de gènes identifiés, l'activation de certains pourrait être corrélée aux signaux exogènes et à l'initiation de la fixation, sachant que la coralline a été utilisée pour induire la métamorphose chez les larves *H. asinina*. Bien qu'une régulation constante a été observée pendant l'acquisition de compétence et l'induction de la fixation, des gènes associés aux processus biologiques de l'activité motrice des microfilaments et du métabolisme énergétique ont montré un différentiel d'expression spécifique à la stimulation par la coralline. Une connexion entre la transition de l'état larvaire et la nourriture peut donc désormais être considérée. Le cycle de vie biphasique de la larve est loin d'être transcriptionnellement statique, mais comprend au moins une période d'activation de fonctions moléculaires spécifiques appartenant à différents processus biologiques interconnectés. Ces réseaux régulateurs sont à la base de l'assimilation des nutriments es-

sentiels, la mise en place d'un budget énergétique, puis l'orchestration des changements transcriptionnels chargés de maintenir une valeur sélective constante des gènes.

1.3.3.2 *Probiotique et résistance aux stress biotiques en larviculture*

Le terme *probiotique*[35] a été défini pour la première fois dans les années 1965 par Lilly et Stillwell [1965]. Verschuere et collaborateurs [2000] restreignent leur définition à la larviculture dans une perspective d'optimisation des rendements. Ainsi, le probiotique représente l'ensemble des souches microbiennes vivantes (et des additifs nutritionnels) administrées sous forme d'un complément alimentaire, conservées dans leur état actif afin d'améliorer la santé de l'hôte (amélioration générale de la survie et la croissance). L'effet bénéfique d'un probiotique est atteint en fonction de l'ensemble de ces rôles suivant : (i) complément nutritionnel (ii) amélioration de l'alimentation (iii) optimisation du système immunitaire (iv) activité antibactérienne (v) compétence spécifique à des substances et (vi) stimulation des processus biologiques, tels que le recrutement et la métamorphose. Un *biocontrôle* peut donc être développé [Verschuere et al., 2000] pour éliminer ou réduire les agressions de pathogènes [Bachère, 2003] en introduisant un organisme adverse (comme un parasite ou un pathogène antagoniste spécifique) qui pourrait prévenir les dommages et ralentir sa croissance en réduisant la pathogénicité sur l'organisme cible [Prado et al., 2010]. Certaines bactéries peuvent améliorer la digestion chez les larves d'huîtres [Douillet and Langdon, 1994], inhiber les pathogènes retrouvés souvent dans les milieux de culture [Gibson et al., 1998], ou synthétiser des acides gras essentiels [Yazawa, 1996]. Ces facteurs contribuent à une amélioration des chances de survie et de croissance des larves [Watanabe et al., 1992; Gomez-Gil et al., 2000].

35 *Probiotique* : *définition I selon* Lilly and Stillwell [1965] c'est une substance produite par un protozoaire qui stimule la croissance d'autres organismes. *définition II selon* Verschuere et al. [2000] c'est un complément microbien avec des effets bénéfiques sur l'organisme cible, en modifiant la communauté microbienne ambiante associée à ce dernier par le fait d'une amélioration de l'alimentation ou de sa valeur nutritionnelle et d'une réduction de la sensibilisation de l'hôte envers les maladies.

1.3.3.3 *Interaction entre microorganismes et développement larvaire (consortium bactéries/-diatomées)*

Le recrutement et la métamorphose larvaire de plusieurs espèces d'invertébrés marins tels la moule sont facilités par la présence de *biofilm* [Bao et al., 2007; Hadfield, 2011], constitué d'une agrégation de microorganismes, au sein duquel des cellules de différentes sources et origines (bactéries et microalgues) adhèrent les unes aux autres et à la surface. Cette *conglomération extracellulaire* est composée d'acides nucléiques, de protéines et de polysaccharides avec des configurations diverses et variées, dont l'ensemble est couvert d'une matrice gluante formée de *substances polymériques extracellulaires* (SPE) [Qian et al., 2007]. Une couche de molécules organiques qui couvre son interface solide facilite le recrutement réversible de colonisateurs primaires. Des *microcolonies* sont ainsi formées après propagations et synthèse de SPE par les premiers colonisateurs. L'hétérogénéité physiologique des couches micro et macrostructurées du biofilm est variable en fonction de la multiplication des colonisateurs et le nouveau recrutement d'autres microorganismes [Salta et al., 2013].

Des conditions optimales d'un consortium synchronisé entre microorganismes contribueraient à l'initiation de la métamorphose par des stimuli spécifiques provoqués par des molécules organiques (riches en éléments nutritifs), des souches bactériennes non pathogènes, des microalgues nutritionnelles, des polymères extracellulaires et une exposition différentielle à la température [Hadfield, 2011]. La composition biochimique du substrat peut donc avoir un impact sur les larves précompétentes des bivalves marins (Figure 1.1 page 10). Des bactéries marines pigmentées productrices d'un SPE hydrosoluble peuvent favoriser le recrutement des larves d'huîtres prémétamorphiques de *Crassostrea virginica* [Weiner and Colwell, 1982; Weiner et al., 1985]. L'initiation de la métamorphose chez les moules *M. coruscus* et *M. edulis* par des agents neurostimulants dissous et des solvants organiques peut être aussi observée [Yang et al., 2013; Toupoint et al., 2012]. Un biofilm bactérien facilite le recrutement des larves pélagiques de pétoncle *Argopecten purpuratus* et le raccourcissement

de la période de précompétence corrélés à l'activation de marqueurs génétiques spécifiques à la résistance au stress oxydatif [Zapata et al., 2009]. Parson et collaborateurs [1993] ont observé un captage préférentiel des larves de pétoncle *Placopecten magellanicus* sur des substrats hautement colonisés par des microorganismes, suggérant ainsi que le développement de biofilms sur les collecteurs peut améliorer la fixation et la survie larvaire. Malgré l'effet inclusif de leurs prédispositions génétiques, les larves interagissent avec leur environnement par des signaux chimiques [Hay, 2009]. Ces facteurs peuvent modifier les tendances écophysiologiques des larves et la direction sélective de la structure génique des populations [Alfaro et al., 2011]. D'autres facteurs biotiques comme la nature trophique ont aussi des effets sur le phénotype larvaire [Tremblay et al., 2007a]. Cependant, les mécanismes moléculaires et les interactions géniques qui sont à l'origine de la réponse phénotypique résultante sont peu explorés. La réponse de l'organisme est complexe et variée respectivement en fonction de la qualité et de la disponibilité des nutriments et de la nature de l'espèce.

1.4 IMPORTANCES DES PRÉCURSEURS DES EICOSANOÏDES

Les études physiologiques essaient de présenter une explication conceptuelle et réaliste des mécanismes adaptatifs sélectionnés par une modélisation évolutive des traits phénotypiques. Toutefois le but de l'adaptation locale peut varier en fonction de ses propriétés fonctionnelles. Pour les larves de pétoncles *A. purpuratus*, cela a été réalisé en incluant la valeur sélective des parents et les succès de reproduction et de fécondation [Navarro et al., 2000]. Le remodelage tissulaire morphophysiologique, les impacts multistress [Calabrese et al., 1977], la phase planctonique et la précompétence larvaire peuvent aussi avoir des effets sur la croissance et la mortalité [Hadfield et al., 2001; Roberts and Lapworth, 2001]. Malgré les compromis physiologiques entrepris pour l'acquisition de ressources et la résistance aux perturbateurs, l'adaptation locale dépend de la structure génétique d'une population et de la fonction sélective inclusive de ses gènes (Section 1.2.1 page 20). Pour cette raison une relation pourrait exister entre le génotype et le phénotype [Storz and Wheat, 2010].

Les études fonctionnelles qui prennent simultanément en compte les variances génétiques et phénotypiques en fonction de prédicteurs adaptatifs peuvent probablement inférer des interactions plus complètes entre le phénotype et l'environnement.

Malgré la conservation des métabolismes de protéines et des carbohydrates [Holland, 1978], les réserves lipidiques sont prioritairement catabolisés chez la plupart des larves d'invertébrés, afin de satisfaire leur budget énergétique et maintenir leur métabolisme basal[36] [Holland, 1978; Gallager et al., 1986; Whyte, 1987]. Le budget énergétique des larves de bivalves est lié essentiellement aux réserves de lipides neutres et de protéines, quoique, les carbohydrates sont considérés comme des ressources secondaires [Holland and Spencer, 1973; Gallager et al., 1986; Whyte, 1992; Videla et al., 1998; Gireesh et al., 2009; Garcia-Esquivel et al., 2001]. Les lipides des mollusques peuvent être divisés en deux groupes, les lipides polaires, dont les phospholipides et les lipides neutres composés principalement de triglycérides. La régulation fonctionnelle des différents acides gras est encore peu connue chez les bivalves, mais leurs effets sur la physiologie de l'organisme sont bien documentés [Pernet et al., 2005].

La croissance et la survie des larves de bivalves marins dépendent de l'apport en acide gras essentiel[37] (EFA) provenant de la nourriture et leur incorporation dans les membranes cytoplasmiques, nucléaires et mitochondriales des cellules [Hulbert et al., 2007; Waldock and Holland, 1984]. Ces acides gras sont dits essentiels, car ils sont nécessaires pour la survie et la croissance des animaux [Pernet et al., 2004; Marshall et al., 2010]. Bien que les microalgues sont à la base de la chaîne alimentaire constituant la source principale d'EFA (l'acide linoléique, LOA, 18 :2ω6 ; l'acide linolénique, LNA, 18 :3ω3 ; l'acide arachidonique, ARA, 20 :4ω6 ; l'acide eicosapentaénoïque, EPA, 20 :5ω3 ; acide docosahexapentaénoïque, DHA, 22 :6ω3), la plupart des animaux ne peuvent pas les synthétiser [Glencross, 2009].

36 *Métabolisme de base* : le niveau énergétique minimal dépensé pour le maintien des mécanismes cellulaires nécessaires à la survie de l'organisme.

37 *EFA* : en anglais *Essential Fatty Acid*, ces acides gras sont reconnus pour être des composants clés dans plusieurs mécanismes cellulaires nécessaires pour le maintien de l'homéostasie [Brett and Müller-Navarra, 1997; Arts et al., 2001; Furuhashi and Hotamisligil, 2008; Tocher, 2003].

Les fonctions biologiques basées sur les EFA peuvent être classées en deux catégories qui incluent (i) le DHA, un oméga-3 (ω3), impliqué dans le maintien de l'intégrité et de la structure fonctionnelle de la membrane phospholipidique et facilite les changements conformationnels des protéines transmembranaires [Tocher, 2010] et (ii) les acides gras polyinsaturés (PUFA) à longue chaîne comme l'EPA (ω3) et l'ArA (ω6) qui constituent les *précurseurs des eicosanoïdes* [Harbige, 2003]. Les eicosanoïdes sont connus pour leurs actions pro-inflammatoires, quoique, ils sont probablement aussi impliquées dans différentes voies de signalisation (Figure 1.4) [Stanley and Howard, 1998; Sampath and Ntambi, 2005]. La production d'eicosanoïdes comme la prostaglandine et la leukotriene est associée à des situations de perturbations physiologiques et biologiques chez les bivalves marins [Howard and Stanley, 1999; Delaporte et al., 2006; Milan et al., 2013; Canesi et al., 2002]. La concentration élevée en ArA dans la fraction neutre des lipides chez l'huître adulte *Crassostrea corteziensis* suggère un compromis sélectif entre la maturation des gonades et l'activation des effecteurs immunitaires [Hurtado et al., 2009]. La fraction neutre est généralement utilisée comme composante énergétique prioritaire chez les larves d'huîtres *O. edulis* et de pétoncles *Pecten maximus* [Holland and Spencer, 1973; Delaunay et al., 1993]. Par ailleurs, chez le pétoncle *A. purpuratus* la performance larvaire prémétamorphique a été négativement corrélée à un enrichissement nutritionnel en EPA et DHA, mais pas après une supplémentation en ArA [Nevejan et al., 2003]. Toutefois, l'EPA a provoqué une augmentation du nombre et du volume des œufs de palourdes *Macoma balthica* [Hendriks et al., 2003]. Une rétention prioritaire de certains acides gras pourrait être parfois observée, avec une conservation préférentielle de l'ArA durant des périodes de jeûnes [Izquierdo, 1996] et son effet positif sur la croissance et la survie des larves [Pernet et al., 2005] et des postlarves du pétoncle *Placopecten magellanicus* [Milke et al., 2008; Pernet and Tremblay, 2004]. Des effets opposés existent entre l'EPA et le DHA, pour la biosynthèse de phospholipides et entre l'EPA et l'ArA pour la production des eicosanoïdes [Sargent et al., 1993, 1997], sachant que les eicosanoïdes produits par la chaîne de synthétise de l'ArA sont les plus inflammatoires [Harbige, 2003]. Un état oxydatif associé

à la défense cellulaire contre les agressions est induit après activation des cellules immunitaires [Pompeia et al., 2000; Canesi et al., 2002; Hulbert et al., 2007], quoique, les dommages oxydatifs produits dans la cellule peuvent se cumuler irréversiblement et provoquer une apoptose (Figure 2.2 page 58).

Un apport régulier en EFA équilibré spécifiquement pour les différentes phases larvaires constitue une stratégie efficace dans l'amélioration des transitions morphophysiologiques des larves et probablement la réduction des mortalités [Widdows, 1991]. Effectivement, la disponibilité trophique peut être un facteur déterminant dans la vitesse de croissance des larves et la prolongation de leur phase pélagique [Thompson et al., 1993]. Le développement larvaire chez le pétoncle *P. maximus* a montré une métabolisation préférentielle des différents EFA, quoique, la composition en acides gras totaux était corrélée négativement à la croissance [Delaunay et al., 1993]. Bien que la croissance a été améliorée après une supplémentation en EPA chez les postlarves de palourde *Tapes philippinarum* [Caers et al., 1998], une corrélation négative a été observée chez les larves d'huîtres *C. gigas* [Thompson et al., 1996]. Une dissolution de l'ArA dans l'eau a augmenté son incorporation dans la fraction polaire des lipides chez l'huître *C. gigas*, ainsi qu'une activation des mécanismes de phagocytose et d'adhésion héomcytaire [Delaporte et al., 2006]. La valeur sélective d'huîtres adultes *C. gigas* a été étudiée chez deux populations, sensible et résistante aux mortalités [Delaporte et al., 2007]. Après un enrichissement en EFA, la reproduction, le stress oxydatif et la mortalité des individus sensibles étaient tous plus élevés comparé à la population d'huîtres résistantes. Ce qui confirme la présence possible d'une pression sélective sur la valeur inclusive de la transmission des gènes ou du compromis de survie (Section 1.3 page 26). D'ailleurs, la synthèse d'eicosanoïdes dépend majoritairement de l'ArA (en particulier la prostaglandine qui est variable en fonction du niveau de stress). Ainsi, l'effet nutritionnel sur la survie est probablement corrélé à l'activation des mécanismes de défense larvaire et du métabolisme énergétique. L'ensemble de ces résultats attribue une implication possible des EFA en particulier l'ArA dans les mécanismes de défense et le maintien de l'homéostasie nécessaires à la

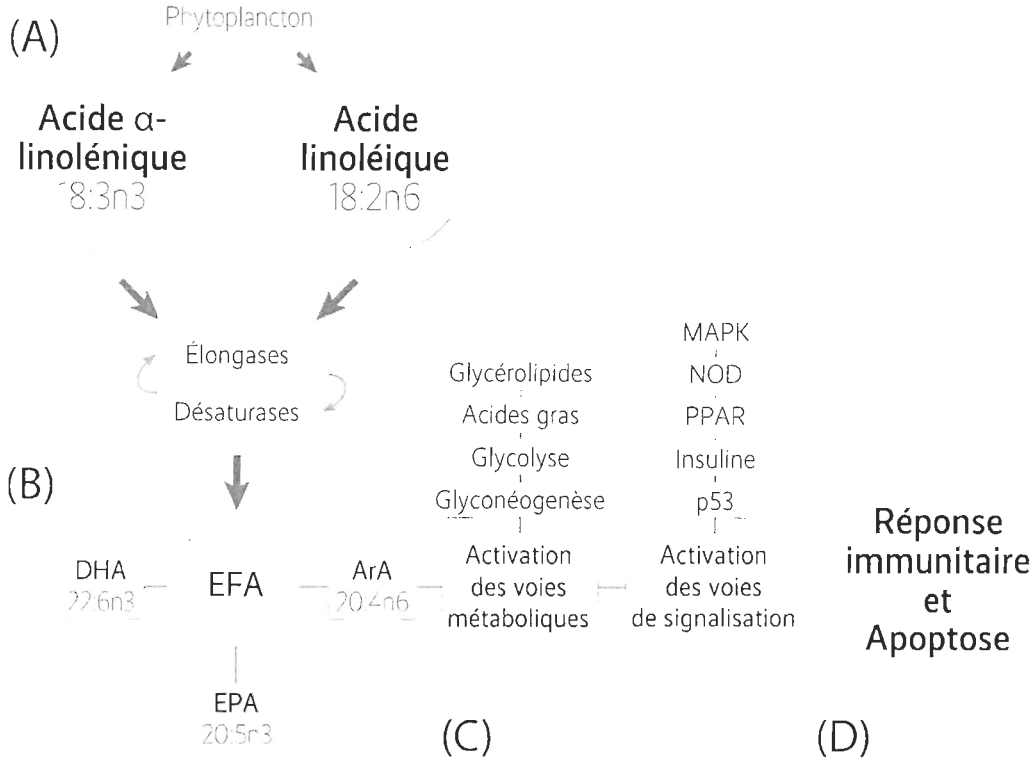

Figure 1.4: **Schéma conceptuel de la voie de synthèse des acides gras essentiels et les effets moléculaires des précurseurs des eicosanoïdes.** (A) La nourriture est la source essentielle de lipides pour les larves. Plusieurs classes d'élongases et désaturases vont assister à la synthèse des acides gras polyinsaturés à longue chaîne. (B) Certains acides gras essentiels vont servir comme précurseurs d'eicosanoïdes ou jouer un rôle dans la structure de la membrane cellulaire. (C) L'ArA est un précurseur d'eicosanoïdes pro-inflammatoires, une fois libérés, des voies métaboliques et de signalisation seront activées. (D) Une inflammation fait partie de la défense immunitaire. Elle aboutit parfois à une mort cellulaire programmée. DHA, acide docosahexaenoïque ; EPA, acide eicosapentaenoïque ; ArA, acide arachidonique ; EFA, essential fatty acid ; MAPK, mitogen-activated protein kinase ; NOD, nucleotide-binding oligomerization domain ; PPAR, peroxisome proliferator-activated receptor (adapté de [Fielman and Marsh, 2005]).

survie des larves de plusieurs espèces marines (Figure 1.4 page précédente), quoiqu'il reste à savoir si l'effet de l'ArA est continu tout au long du développement larvaire. Dans de telles circonstances quelles seraient les bases génétiques responsables du succès physiologique observé ?

Il n'est pas facile de retrouver l'origine d'un trait phénotypique que ce soit un gène, des interactions entre gènes ou une activité protéique. Mais il est possible aujourd'hui d'étudier l'écologie évolutive par des algorithmes génétiques afin de visualiser la dynamique du génome. Une fois caractérisée, cette dynamique pourra éclaircir le fonctionnement des mécanismes responsables de la valeur sélective des organismes. Et dans le cas ou une telle interaction existe entre le génotype et le phénotype pourrons-nous trouver les gènes qui en sont la base ?

1.5 LES OBJECTIFS DE LA THÈSE

Le principal objectif de l'étude est d'enrichir par une approche intégrative et quantitative les connaissances sur les mécanismes moléculaires qui contribuent au succès du développement larvaire de la moule bleue *M. edulis*. Il s'agit plus particulièrement de comprendre comment les gènes interagissent pour améliorer les chances de survie larvaire, la performance et la croissance des postlarves. Les conditions physiologiques des larves, relatives aux facteurs trophiques, représentent un déterminant majeur susceptible de compromettre le succès de la métamorphose, une étape cruciale du développement, ou bien rendre l'individu immunodéprimé et de ce fait susceptible de mortalité. Ainsi, l'hypothèse générale de la thèse est de considérer les effets bénéfiques des précurseurs des eicosanoïdes, l'ArA et l'EPA sur le développement larvaire. L'importance biologique de ces acides gras essentiels sur la croissance et le développement ontogénique semble être confirmée chez *M. edulis* ainsi que d'autres bivalves marins. L'acquisition des réserves énergétiques augmente la valeur sélective des larves et contribue à l'activation des processus métaboliques nécessaires

pour l'organogenèse et le maintien de l'homéostasie cellulaire. Pourtant les voies moléculaires associées au succès métamorphique et à une résistance aux perturbations sont peu documentées et leurs mécanismes d'actions demeurent inconnus.

Les stades de développement périmétamorphique sont caractérisés par une sensibilisation élevée des larves de la moule aux facteurs environnementaux. Pour cette raison, nous allons rechercher pendant les premiers stades de développement, les gènes associés avec une meilleure survie et croissance. Nous essayerons également de caractériser la fonction des processus biologiques et moléculaires qui s'activent pendant l'ontogenèse larvaire. Quels sont les gènes qui sont impliqués dans ces processus et comment ils interagissent pour assurer une meilleure valeur sélective des larves ?

L'objectif de l'évaluation du potentiel fonctionnel du transcriptome de *M. edulis* est donc de caractériser la dynamique du génome pendant les premiers stades de développement. Ces objectifs essayent de cadrer le compromis intraspécifique de la moule bleue *M. edulis* en réponse aux pressions sélectives trophiques et l'adaptation locale relativement à la morphologie fonctionnelle. La planification des activités est structurée de la façon suivante :

* Le profil d'expression génique pendant le développement larvaire et la cinétique différentielle relative.
* Les régulateurs transcriptionnels et leurs rôles dans les processus biologiques du développement larvaire.
* L'importance d'un régime alimentaire riche en précurseurs d'eicosanoïdes sur le développement larvaire, la croissance et la survie.

Deuxième partie

RÉSULTATS DE RECHERCHE

2

L'IMMUNITÉ PENDANT L'ONTOGENÈSE DES BIVALVES : ÉVALUER LE POTENTIEL DES TECHNIQUES DE SÉQUENÇAGE NOUVELLE GÉNÉRATION

Bassim Sleiman[1,2], Genard Bertrand[1], Gauthier-Clerc Sophie[3], Moraga Dario[2], Tremblay Réjean[1]

[1] *Institut des Sciences de la mer de Rimouski, Université du Québec à Rimouski, 310 allée des Ursulines, Rimouski, Québec, G5L 3A1 Canada*

[2] *Laboratoire des Sciences de l'Environnement Marin, Institut Universitaire Européen de la Mer, Université de Bretagne Occidentale, Rue Dumont d'Urville, 29280 Plouzané, France*

[3] *Bureau Recherche Développement Valorisation, Université de Montréal, Montréal, H3C 3J7 Canada*

2.1 RÉSUMÉ

Ce premier article, intitulé *Ontogeny of bivalve immunity : assessing the potential of next-generation sequencing techniques*, fut corédigé par moi-même ainsi que par mes collègues les docteurs Bertrand Génard, Sophie Gauthier-Clerc, Dario Moraga et Réjean Tremblay. Il fut publié début 2014 dans la revue *Reviews in Aquaculture*. En tant que premier auteur, ma contribution à ce travail fut la rédaction de l'article, l'idée originale et la recherche sur l'état de l'art. Le docteur Bertrand Génard a aidé dans la recherche sur la partie des défenses immunitaires et la révision de l'article. Le docteure Sophie Gauthier-Clerc a aidé à la révision de l'article. Les docteurs Dario Moraga et Réjean Tremblay ont aidé dans l'organisation des parties, le choix de l'idée originale et la révision de l'article. Une version abrégée de cet article a été présentée au congrès *Ecotoxicologie des Bivalves Marins* (ECOBIM) à Reims (France) en 2012.

Les organismes vivants sont en constante évolution afin de survivre en s'adaptant au niveau moléculaire et cellulaire. La plupart des bivalves marins commencent leur histoire de vie larvaire par un stade microscopique planctonique puis ils deviennent des individus benthiques après leur fixation. Les larves pélagiques sont généralement plus sensibles aux agressions environnementales et aux agents pathogènes que les adultes de la même espèce. Une meilleure capacité adaptative pourrait améliorer la survie de ces jeunes organismes. Les progrès dans l'exploration des bases de données et les analyses génomiques à grande échelle doivent améliorer notre connaissance des processus biologiques du fonctionnement moléculaire inconnu de ces stades larvaires. Les données actuellement disponibles sur le développement larvaire sont fragmentées, mais l'information sur la biologie de l'individu adulte est plutôt bien documentée. Explorer la diversité par un suivi comportemental des histoires de vie des espèces marines et trouver des solutions pour des problèmes communs, par exemple les mortalités massives des larves, surtout dans les conditions actuelles de changement climatique repose sur notre connaissance des processus moléculaires responsables de la plasticité phénotypique. Bien qu'il soit difficile d'évaluer les mécanismes immunitaires par le suivi de la circulation immunocytaire chez les jeunes larves, les études sur le développement immunitaire sont désormais possible au niveau du transcriptome. Les techniques de nouvelle génération à haut débit offrent d'excellentes solutions pour l'analyse à grande échelle du transcriptome. Nous présentons un bref examen de l'ontogenèse précoce du système immunitaire chez les bivalves marins, avec une attention particulière sur les applications de séquençage de nouvelle génération. Il y a toutefois un compromis entre l'étendue de la couverture d'un sujet et le nombre de détails abordés pour celui-ci. Nous allons donc se concentrer surtout sur l'ontogenèse de la défense immunitaire et les progrès des études moléculaires relatives.

Ontogeny of bivalve immunity: assessing the potential of next-generation sequencing techniques

2.2 ABSTRACT

Living organisms are constantly evolving to secure their survival via adaptations at the molecular and cellular level. Most marine bivalves have microscopic planktonic larval stages until settlement to the benthic environment. These pelagic stages are generally more sensitive than their adult counterparts to environmental and pathogen threats. Adaptive capacities could improve survival of these early stages. Recent advancements in data mining and pipeline analysis should shed light on the currently unknown processes that occur during these first stages. Existing data on early stages are fragmented compared to the abundance of information available for adult. Exploring diversity through aquaculture and lessening the impact of common issues, e.g., massive mortalities of larvae, especially within the current conditions of a changing climate, ultimately rests on our knowledge of the molecular processes responsible for phenotypic plasticity. Although it is somewhat difficult to assess immune mechanisms by tracking circulating immunocytes in larvae, studies on the development of immune processes are now feasible at the transcript level. Next-generation techniques offer outstanding solutions for wide-range transcriptome analysis. We present a short review of the early ontogeny of the immune system in marine bivalves, with particular focus on next-generation sequencing applications. Like all reviews of this nature, there is a trade-off between the depth of the coverage and the number of subjects discussed. We will thus restrict the scope to bivalve immunity and focus on the central concepts across a wide range of topics, i.e., the ontogeny of immunity and advancements in molecular studies. ***Key words:*** Bivalve, immune defence, next-generation sequencing, ontogeny, transcriptomics.

2.3 INTRODUCTION

2.3.1 *A new chapter for marine bivalves*

In this paper, we only consider marine bivalves. These benthic organisms have been studied worldwide, and numerous molecular, cellular, and behavioural approaches have been used to examine the evolutionary bases of defence mechanisms, survival, and organism maintenance (Figure 2.1). More recently, research trends in marine biology have revolved around how these elements interrelate to avoid genetic aberrations and to establish better synchronisation and adaptation of the phenotype [Bieler et al., 2013]. There are many avenues of study, including the investigation of processes involved in lesion repair [Clark et al., 2010], shell formation and repair [Waldbusser et al., 2013], capture [Pales Espinosa et al., 2010] and elimination of toxic particles [Canesi et al., 2012], inter-individual variability in inflammatory responses [Mosca et al., 2011; Dang et al., 2012], immunocell phagocytic activities [García-García et al., 2008], hemocyte number, size, and cell-division rate [Flye-Sainte-Marie et al., 2009], oxyradical production (respiratory burst and nitric oxide (NO) release) [Vouras and Dailianis, 2012], endocrine system perturbations [Hughes Jr et al., 1990; Aarab et al., 2006], and species identification, allele frequencies, and allozyme polymorphisms [Carcamo et al., 2005]. All these topics and more have already captured the interest of many researchers who attempt to understand the evolutionary basis of survival mechanisms on marine coastal ecosystems.

These preliminary studies utilized common biomarkers that have almost always been found to be correlated to the cellular change under investigation. While molecular correlations may not always indicate a direct relationship, gene networks responsible for that cellular change remain unexplored. In addition, studies on early developmental stages and inflammatory responses throughout this sensitive period are scarce. Whether there is an abiotic or biotic threat, larvae need to maintain normal growth and survival through interrelated defence mechanisms. In this regard, we still do not possess all the pieces that will allow us

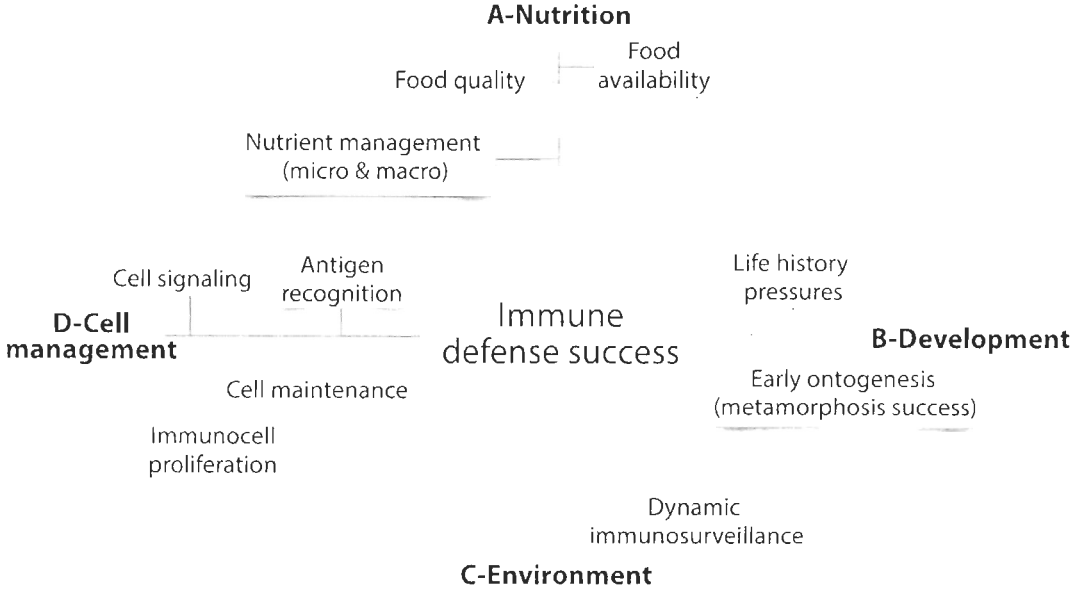

Figure 2.1: **Success rate of the immune system during larval development is dependent on processes occurring at the cellular level.** Each function is dependent on a module of co-regulated genes, and every module interacts with another module inside one or different category (branched network), thus forming a gene network responsible of one specific function. (a) Nutrient acquisition and appropriate use of energy reserves are crucial for sustaining early development and survival. (b) Young stages are transitioned through an intense makeover phase, metamorphosis. Selection is based on the survival of the fittest-concept (risking survival just to pass on genetic code) or survival of the selfish adults (withholding reproduction or partial spawning). (c) Biotic and abiotic factors can render and organism vulnerable to exogenous threats. (d) Immune defence is dependent on a molecular management success, from antigen recognition and elimination, to cellular repair.

to understand how the mechanisms function the way they do. This review unequivocally demonstrates the potential for new scientific discoveries in early ontogenesis that rest on new techniques in data mining and analyses. Emerging comprehensive transcriptome studies will undoubtedly speed up new immune and developmental discoveries and hence lead

to new functional genomic breakthroughs in early ontogenesis. This will shed more light on gene behaviour patterns and expression differentiation between cells.

2.3.2 *Innate defences*

Marine bivalves are dependent on a non-specific defence mechanism known as the innate immune system. The innate immune system is a primitive mechanism that triggers intracellular molecular signalling cascades; these are found in all living multicellular organisms and are conserved across kingdoms [Danilova, 2006; Mydlarz et al., 2006]. The innate immune system promotes generalized protection against pathogenic organisms (e.g., protozoa, bacteria, endoparasites, mycoplasma, and viruses), environmental stressors (e.g., xenobiotic assault and algal toxins, i.e., harmful algal blooms), and other immunomodulating stimuli (e.g., air exposure, mechanical stress, high temperatures, and extreme salinity conditions). Some microorganisms confer a prebiotic or probiotic role to the host without causing disease or mortality, and these are of particular interest in bivalve larviculture [Bachère, 2003; Prado et al., 2010; Ringø et al., 2010; Kesarcodi-Watson et al., 2012]. Other species may compromise or colonise the host, probably to a greater extent when host immunosurveillance is reduced because of unfavourable abiotic factors (e.g., suboptimal temperature or salinity conditions, and organic/inorganic pollutants) [Wang et al., 2012a; Husmann et al., 2013]. An organism's sensitivity to excessive immunodisturbances may be permanent or temporary, thus the existence of resistant or susceptible communities [Bricelj et al., 2010]. Permanent sensitivity is caused by endogenous damage (genetically predisposed or acquired DNA alterations) [Singh and Hartl, 2012; Varotto et al., 2013], while temporary sensitivity is influenced by transient exogenous factors [Gobler and Talmage, 2013] that are related to environmental factors (acquired sensitivity due to the frequency of aggression) or that may have mixed origins (sensitivity related to the intensity of aggression). Protective mechanisms in marine bivalves are based on an immediate cellular and humoral response of short duration.

2.3.3 *Complexity of the innate immune system*

Cellular resistance is based on the chemotactic and chemokinetic activities of phagocytosis [Kuchel et al., 2010; Balseiro et al., 2011], encapsulation [Mladineo et al., 2012], and melanization [Estrada et al., 2010], which can be summarized by the catalytic activity of destruction enzymes and the release of reactive oxygen metabolites. The innate immune system in bivalves displays a wide variety of selective effectors, sensitive receptors, and synergistic GRNs; this multi-layered model involves all cells in the organism [Barmo et al., 2013]. Unlike cellular resistance, the humoral response includes non-specific opsonic processes (enhanced cellular attachment after antigen coating). It is often represented by polymerizing enzymes and by secretion of lytic and cytotoxic peptides, for example, lectins [Yue et al., 2013a], agglutinins and bactericidins [Yue et al., 2013b], and lysosomal proteases [Niu et al., 2013]. This defence system is nevertheless primitive compared to the more sophisticated system of adaptive immunity found in higher species of vertebrates, but it is still effective for protecting the host in a fluctuating environment. While cellular response can be monitored on the protein level, early gene regulation remains largely unknown. Even though some immunocell capabilities have been characterised in studies on *Mytilus edulis* [Dyrynda et al., 1995] and *Crassostrea virginica* [Elston, 1980] larvae for example, current databases lack large-scale transcriptomic data on defence mechanisms during ontogenesis. This information is of critical importance when trying to discover connections between oxidative stress, metabolic cost, and immunity during early life stages of marine bivalves [Estevez-Calvar et al., 2013; Diaz de Cerio et al., 2013]. Two decades ago, the introduction of arrays enabled scientists to overcome previous obstacles regarding gene regulation, which permitted the exploration of the first transcriptomic studies possible [Fodor et al., 1993]. This technology was initially expensive and often requires the complex design of target probes [Pevzner and Lipshutz, 1994]. Regardless of these challenges, arrays became a stable platform and provided a first indication of what kind of data would be expected. Conversely, determining

gene interactions associated with cellular responses in non-genotyped organisms has been difficult and often misleading.

2.3.4 *Functional genomics in the era of next-generation techniques*

Due to the difficulties involved in studying the immune potential of microscopic larvae, NGS techniques provide powerful tools to obtain accurate information on the ontogeny of immune-related genes and their interactions in a manner previously not possible [Balseiro et al., 2013]. Some features of NGS techniques include whole-genome de novo sequencing, alternative splicing and junction mapping, discovery of non-coding RNAs (ncRNAs), gene behaviour profiling, and transcript structure evaluation; every step is done with a high level of accuracy and at a single nucleotide level. For example, dual RNA-seq can effectively sequence transcripts and quantify their abundance in host and pathogen cells simultaneously, without the need to actively separate the two specimens. This provides correlations of gene behaviour during host-pathogen interactions even in a timeseries design, which is very helpful for understanding mortalities occurring throughout early development [Westermann et al., 2012]. Collectively, these applications can be used without limitations on non-model species with sparse genomic resources, specifically on economically important marine bivalves [Ekblom and Galindo, 2010; Shokralla et al., 2012; Tanguy et al., 2013b].

2.3.5 *Why the focus on immune studies*

Dynamic equilibrium is essential for organism maintenance (growth and survival) [Pettersen et al., 2010; Powell et al., 2011], assimilation of food resources, allocation of energy reserves (metabolic mechanisms, dynamic energy budgets), stress tolerance, and disease resistance (defence mechanisms and local adaptation) [Genard et al., 2011; Samain, 2011]. To make advancements in the field of bivalve immunology, we need to improve our un-

derstanding of the factors that influence and stimulate stress responses [Fitzgerald-Dehoog et al., 2012; Guerra et al., 2012]. Mortalities affecting hatchery facilities cause economic hardships [Roch, 1999]; they are usually due to infections associated with microorganisms [Torres et al., 2013; Ren et al., 2013b] and sometimes amplified by genomic aberrations [Martín-Gómez et al., 2013]. To stabilise short-term farm production and promote long-term performance, it is essential to simultaneously improve our knowledge of immune defence mechanisms and successfully develop targeted genetic approaches.

2.3.6 *New strategies for a better aquaculture*

Success of the immune defence in early larvae and postlarvae is highly dependent on energy sources (nutrients) [Marshall et al., 2010], an adequate energy allocation for growth (development) [Bayne, 2004; Wang et al., 2012a], the capability of dynamic adaptation (selective pressures) [Powell et al., 2011; Barros et al., 2013], and the ability to initiate defence and annihilate specific aggressors (cell management) [Yue et al., 2013b] (Figure 2.1). Physico-chemical conditions and environmental threats influence the probability of host survival by reducing the effectiveness of the immune response, which makes the animal immunodepressed [Delaporte et al., 2003; Waldbusser et al., 2013]. In addition, poor quality food (composition) and a low feeding rate (food dilution and starvation) would reduce growth rate and therefore lengthen development processes [Bayne, 2004; Gagné et al., 2010]. This could directly impact mortality and juvenile survival by delaying or hindering larval settlement and lowering spat recruitment success. Collectively, this body of literature suggests the probable existence of an as yet uncharacterised global genetic signature underlying susceptible or resistant organisms [Americo et al., 2013]. This is highly likely since cellular defence and maintenance are dependent on interconnected GRNs variously co-regulated and enhanced in a synchronised linear or in a more complex non-linear way, eg., [Husmann et al., 2013]. These networks have yet to be identified then unambiguously linked to cellular processes.

High periods of larval mortality are often associated with the peri-metamorphic stages [Balseiro et al., 2013]. Larval survival and resistance depend primarily on the maternal contribution of proteins, lipids, and carbohydrates [Yue et al., 2013b], as well as the stage-specific consumption of phytoplankton [Aarab et al., 2013; Genard et al., 2013]. Endogenous energy reserves are required for larvae to survive metamorphosis, as well as the early formation and reorganisation of juvenile/adult organs, i.e., to complete the relatively stressful physiological transformation marked by the change from a pelagic to a benthic lifestyle. In addition, immune-success modulators depend on energy reserves (Figure 2.1A) since feeding is extremely reduced, so these reserves need to be correctly managed to favour survival [Bayne, 2004; Pernet et al., 2004]. While all studies to date have shown without exception the negative effect of suboptimal nutrition on larval and juvenile bivalves, it is unclear still which GRNs have critical impacts on development and which gene interactions are particularly implicated in vulnerability or resistance during later development. In the sections below, we shall investigate the fast-acting components of the main immune processes that lead to the most effective defence response in order to improve our understanding of the phylogenetically conserved defence mechanisms. In addition, we will summarise a few basic challenges that must be overcome to fully understand the determinants of selectivity for profitable energy allocation, growth, and maintenance mechanisms during early ontogenic development [Sokolova, 2013].

2.4 STRUCTURAL COMPLEXITY OF THE BIVALVE IMMUNE DEFENCE

Bivalve immunity depends on a synchronised set of conserved processes that are behavioral, cellular, and molecular. This molecular system is triggered in response to aggressions, examples include, but not limited to physical injury (e.g., cellular debris), pathogens, and biochemical compounds. Foremost, the immune recognition system (Figure 2.2A) triggers a wide array of signalling pathways and the production of specific reactive effectors (Figure 2.2B), leading to neutralisation of dangerous threats through phagocytosis (Figure 2.2C).

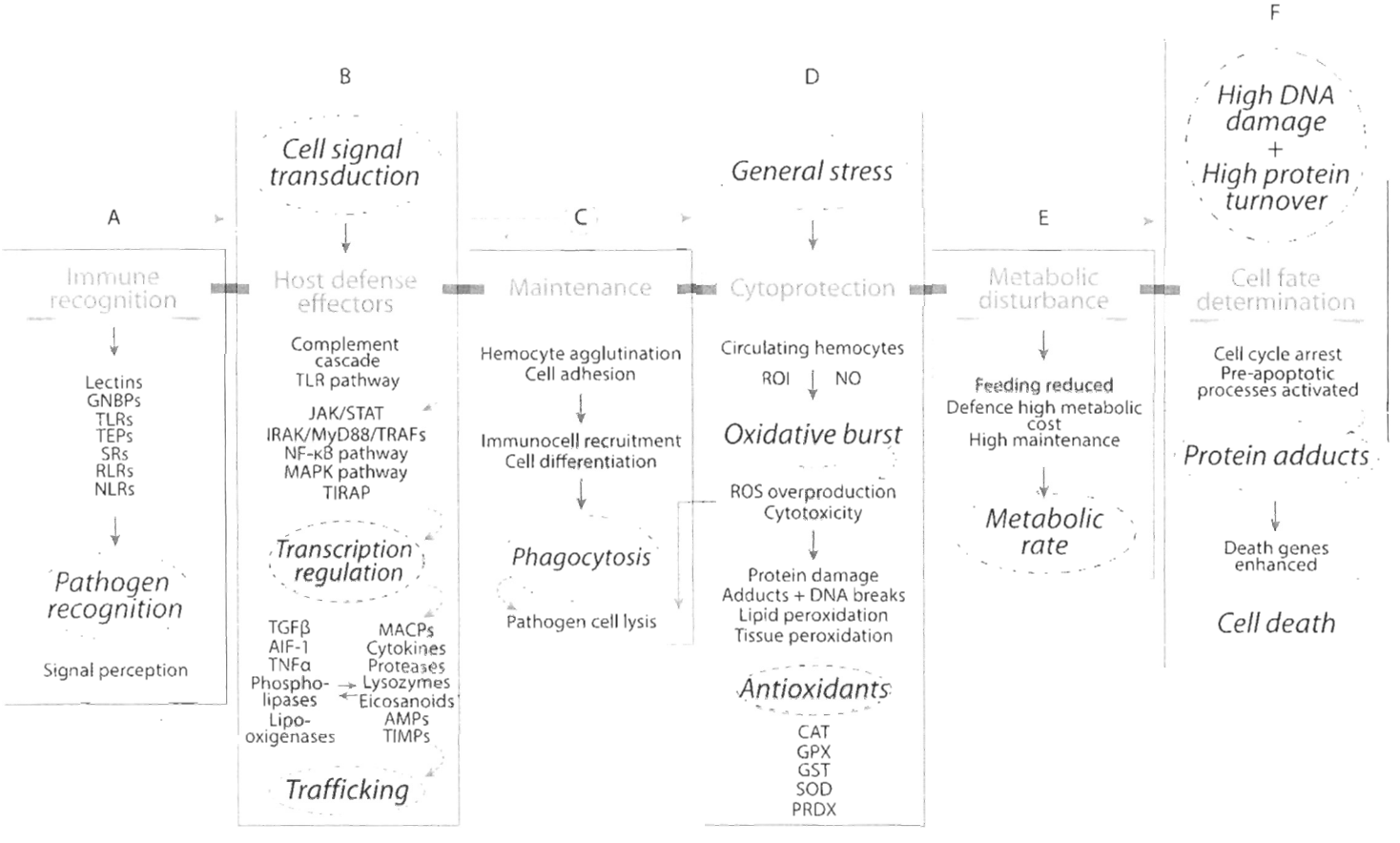

Figure 2.2: **Molecular and physiological determinants of cellular defence mechanisms.**

Figure 2.2: (*Continued*) (a) Threat recognition, (b) defence initiation, (c) immunocell differentiation, (d) immunocell activation, (e) metabolic modulation, and (f) cell damage accumulation then cell death. GNBP, gram-negative binding proteins; TLR, toll-like receptor; TEP, thioester-containing protein; SR, scavenger receptor; RLR, rig-like receptor; NLR, NOD-like receptor; JAK, janus kinase; STAT, signal transducer and activator of transcription; IRAK, IL-1 receptor-associated kinase; MyD88, myeloid differentiation primary response protein 88; TRAF, tumor necrosis factor receptor-associated factor; MAPK, mitogen-activated protein kinase; MACP, membrane attack complex and perforin; AMP, antimicrobial peptide; TIMP, tissue inhibitor of metalloproteinase; TGFβ, transforming growth factor β; AIF-1, allograft inflammatory factor-1; TNFα: tumor necrosis factor α; ROS, reactive oxygen species; CAT, Catalase; GPX, glutathione peroxidase; GST, glutathione-S-transferase; SOD, superoxide dismutase; PRDX, peroxiredoxin.

All of these elements are detailed below for adult and larvae with a particular focus on recent advancements in functional genomics and NGS techniques.

2.4.1 *Immune recognition*

Immune recognition is the first step of the immune response allowing the discrimination of self/not self substances. Antigens are recognised using pattern recognition receptors (PRRs), which is an evolved characteristic of invertebrates. These receptors could be either soluble or expressed on the cell surface [Kawai and Akira, 2010]. PRRs recognise selectively a large family of conserved foreign molecules called pathogen-associated molecular patterns (PAMPs), examples include lipopolysaccharides (LPSs), lipoproteins, peptidoglycans (PGNs), lipoteichoic acids, viral dsRNA, unmethylated bacterial DNA, zymosans, and heat shock proteins (HSPs) [Costa et al., 2009; Green and Montagnani, 2013]. Thus, both the circulating and anchored forms of these receptors play a crucial role in activating the immune system to eliminate pathogens. Several groups of distinct PRRs have been identified in bivalves [Song et al., 2010] and the most researched include lectins, peptidoglycan recognition proteins (PGRPs), gram-negative binding proteinss (GNBPs), and toll-like receptors (TLRs) (summarized in Figure 2.2A). Other groups of PRRs have been found in bivalves adults, these include thioester-containing proteins (TEPs), [Prado-Alvarez et al., 2009; Zhang et al., 2013], scavenger receptors (SRs, [Song et al., 2010; Zhang et al., 2013]), rig-like receptors (RLRs, [Philipp et al., 2012]) and NOD-like receptors (NLRs, [Philipp et al., 2012]) but, no expression profile is available for these receptors at larval stages.

2.4.1.1 *Lectins*

Lectins recognise special carbohydrate patterns on the surface of living cells [Sharon and Lis, 2004]. They form a crucial reversible protein-sugar interaction that allows target selection among captured particles, which is essential in suspension-feeding bivalves [Espinosa

et al., 2009; Jing et al., 2011]. Lectins facilitate mutual exchange between the cells of microorganisms (commensal systems) and trigger host defence mechanisms during opportunistic pathogen aggressions [Adhya et al., 2010; De Lorgeril et al., 2011]. They can be found in soluble form, free in the extracellular matrix, or associated to specific carbohydrate recognition domains (CRDs) in the biological membrane [Espinosa et al., 2010]. They will promote agglutination, immobilisation, or opsonisation mechanisms that are mediated by complementary components [Sharon and Lis, 2004; Danilova, 2006; Tanguy et al., 2008]. A series of transcriptomic studies based on deep mRNA sequencing recently revealed the presence of several lectin sequences in stimulated hemocytes from *Ruditapes philippinarum* [Moreira et al., 2012a], *C. virginica* [Zhang et al., 2013], and *M. edulis* [Tanguy et al., 2013b]. First-generation techniques through cloning and purification enabled the identification of multiple lectin-related transcripts in marine bivalves [Jenny et al., 2002; Tanguy et al., 2008; Fleury et al., 2009]. Studies on development in embryos and larvae detected no change in the behaviour of ficolin3, which stays unfolded during ontogenesis, while a galectin8 expression increases during larval development [Tirapé et al., 2007]. Recent work by Genard et al. [2013] showed an overexpression of lectin-related transcripts in young veliger larvae of *Crassostrea gigas* and *C. virginica* in response to bacterial infection. Similarly, the expression of lectin-3 has been investigated in context of bacterial infection throughout development in *Chlamys farreri* [Yue et al., 2013a], and ficolin-related transcripts have also been detected at the trocophore stage in *Mytilus galloprovincialis* [Balseiro et al., 2013].

2.4.1.2 *Peptidoglycan recognition proteins*

Peptidoglycan recognition proteins bind selectively to PGNs [Ni et al., 2007]; PGNs are structures occurring in the outer layer of the bacterial membranes and are targets for PGRPs that are specifically triggered in order to track bacterial aggression in various tissues. These PGRPs play a diverse but central role in activating certain immune reactions, including melanization cascades, phagocytosis, and several signal transduction pathways. In *C. farreri*,

PGRPs represent a PGN recognition system with hydrolytic [Wei et al., 2012] and antibacterial [Itoh and Takahashi, 2009] activities. PGRPs have been identified in several bivalve species, for example *Argopecten irradians* [Ni et al., 2007], *C. farreri* [Su et al., 2007], *C. gigas* [Itoh and Takahashi, 2009], and *Solen grandis* [Wei et al., 2012]. The implication of these PRRs on the bivalve innate immune system remains unclear [Yang et al., 2010]. However, a series of reported high-throughput transcriptomic analyses combined with homology and domain screens from *R. philippinarum* [Moreira et al., 2012a], *M. edulis* [Philipp et al., 2012; Tanguy et al., 2013b], and *C. virginica* [Zhang et al., 2013] managed to classify the bivalve PGRPs as an important and putatively functional group of the recognition proteins. To date, the only existing records of the behavioural patterns of these recognition proteins during ontogeny are from *C. farreri* and result from a single PGRP-S1 isoform [Yue et al., 2013a].

2.4.1.3 *Gram-negative binding proteins*

Gram-negative binding proteins bind Gram-negative bacteria and LPS- and β-1,3-glucan (LGBP) [Song et al., 2010]. In bivalves, the first reports on GNBP were from *Perna viridis* (βGBP, [Jayaraj et al., 2008]) and *C. farreri* (LPS- and B-1,3-glucan (LGBP), [Wang et al., 2009a; Liu et al., 2010]). Recent studies using NGS techniques enabled the identification of related transcripts to GNBP in *R. philippinarum* [Moreira et al., 2012a], *M. edulis* [Philipp et al., 2012], and *C. virginica* [Zhang et al., 2013]. LGBP is associated with various biological functions, including the activation of the prophenoloxidase system, cytolysis, bacterial aggregation, and opsonisation [Wang et al., 2009a]. At larval stages, a significant increase of LGBP transcript abundance has been measured at the trocophore stage in *C. farreri* [Yue et al., 2013a].

2.4.1.4 *Toll-like receptors*

Toll-like receptors represent a family of PRRs that can recognise and be activated by the presence of PAMPs [Qiu et al., 2007]. A typical TLR structure consists of several extracellular (LRR)

solenoids [Offord et al., 2010]. TLRs are involved in the molecular recognition of pathogens as well as cell adhesion, signal transduction, and cell growth (Figure 2.2A to 2.2C) [Takeda and Akira, 2004; Kim et al., 2011]. The structure also consists of a trans-membrane and an intra-cytoplasmic domain known as the toll/interleukin 1 receptors (TIRs), which is highly conserved in a wide range of organisms, and is responsible for membrane localisation of the TLR and for signal transduction [Zhang et al., 2011; Mikami et al., 2012]. The search for TLRs enabled discoveries in many bivalve models [Tanguy et al., 2004; Song et al., 2006; Qiu et al., 2007; Perrigault et al., 2009; Mateo et al., 2010; Venier et al., 2011; Ren et al., 2013a]. Transcriptome analysis using NGS techniques led to the discovery of a small synexpressed repertoire of putative TLRs encoding sequences for *M. edulis* [Philipp et al., 2012; Tanguy et al., 2013b], *R. philippinarum* [Moreira et al., 2012a], and *C. virginica* [Zhang et al., 2013]. The expression and function of TLRs during early ontogeny are relatively unknown in bivalves. To our knowledge, only the transciptomic study on *Ostrea lurida* larvae provides a mention of the functional existence of TLRs during early ontogenesis [Timmins-Schiffman et al., 2013].

2.4.2 *Immune signalling pathways*

Upon the successful recognition of foreign compounds (Figure 2.2A), signalling cascades trigger the initiation of the cellular defence while promoting gene transcription of inducible immune-related proteins and hæmocyte recruitment (Figure 2.2B and 2.2C). Several signalisation constituents, including TLRs, mitogen-activated protein kinase (MAPK), nuclear factor-κB (NF-κB) pathways, and complement component pathways have all been monitored repeatedly in many marine species for their central function in the innate immune system of bivalves. These highly conserved systems have evolved to protect the cell from infection. We will briefly describe what is known regarding bivalve defence pathways (for a more complete discussion, see the review by [Song et al., 2010]).

2.4.2.1 *The complement system*

The complement pathway relies on several interacting proteins to recognise and then eliminate foreign microorganisms. It plays a pivotal role in the initiation of defence mechanisms, including immunocell homing and trafficking, agglutination, adhesion, opsonisation, and cell lysis. Once activated, the complement system promotes target proteolytic reactions that operate following classical, lectin or alternative pathways [Nonaka and Yoshizaki, 2004]. Mannose-binding protein (MBL), ficolin, and other C1q enzymes that are capable of binding terminal mannose groups of microorganisms play key roles as initiators of the complement pathway [Romero et al., 2011]. When immobilised onto the recognition site, the initiation molecules interact with a protein belonging to the C2-like/Factor B-like amplification factors which trigger the host immune response [Nonaka and Yoshizaki, 2004]. Although the complement pathway has not been widely studied in bivalves [Bado-Nilles et al., 2009; Prado-Alvarez et al., 2009], data mining using first- and next-generation sequencing techniques for functional proteins containing C1q domains, has revealed several conserved homologs in different bivalves [Jenny et al., 2002; Gourdine and Smith-Ravin, 2007; Prado-Alvarez et al., 2009; Gerdol et al., 2011; Moreira et al., 2012a; Zhang et al., 2012a; Philipp et al., 2012; Zhang et al., 2013]. This strongly suggests the presence and conservation of a complement system in bivalves. Additional work on immune-related genes during ontogenesis revealed antigen recognition components of the complement pathway to be triggered as early as the trocophore stage in *M. galloprovincialis* [Balseiro et al., 2013].

2.4.2.2 *Toll signalling pathway*

The activation of TLRs (Figure 2.2A) induces an instant coordination of downstream signalling pathways, including NF-κB, MAPKs and pre-apoptotic processes (Figure 2.2B and 2.2F). The most familiar components that are activated in the TLR pathway of bivalves are the myeloid differentiation primary response protein 88 (MyD88), tumor necrosis factor

receptor-associated factors (TRAFs), IL-1 receptor-associated kinase (IRAK), and adapter-like protein (MAL, also known as TIRAP). These interrelated elements were identified using first-generation pipeline analysis in several bivalve species [Tanguy et al., 2004; Qiu et al., 2009; Wang et al., 2011; Lee et al., 2011; Venier et al., 2011; Green and Montagnani, 2013; He et al., 2013]. Recently, mRNA deep sequencing provided an even larger interconnected picture of the activated functional constituents related to the TLR signalling pathway for even more bivalves [Moreira et al., 2012a; Philipp et al., 2012; Tanguy et al., 2013b; Zhang et al., 2013; Toubiana et al., 2013]. Concerning timeseries monitoring of gene behaviour during ontogenesis, the conventional semiquantitative technique provided exclusively a detection of MyD88 and TRAF3 from embryo to juvenile in *C. gigas* [Tirapé et al., 2007] and after bacterial infection in veliger larvae [Genard et al., 2013].

2.4.2.3 *Mitogen-activated protein kinase pathways*

MAPK signalling cascades are key players in a variety of processes including cell proliferation, growth, differentiation, cell death, innate immunity, and development. Thus, their large range of actions has a great influence on diverse and complex biological and pathological conditions. MAPKs are categorised into three subfamilies that include the extracellular signal-regulated kinases (ERKs), c-jun amino-terminal kinases (JNKs), and p38-MAPKs [Johnson and Lapadat, 2002]. These enzymes function by influencing cell cycle progression. In *M. galloprovincialis*, they can be triggered in response to inflammatory cytokines or during defence [Betti et al., 2006; Ciacci et al., 2010], which is pivotal for stress signalling transduction [Yao and Somero, 2012] and cell cycle arrest [Yao and Somero, 2013]. Several homologs to MAPK pathway constituents have been sequenced in oyster [Tanguy et al., 2004] and Manila clam [Kang et al., 2006]. Recently, NGS analyses on adult *R. philippinarum* [Moreira et al., 2012a] and *C. virginica* [Zhang et al., 2013] have helped to identify a functionally conservative set of regulated transcripts associated with MAPK pathways. During early developmental stages, MAPK-related transcripts have been successfully identified using first-generation

techniques in 12-day-old *C. gigas* larvae in response to a bacterial challenge [Genard et al., 2013], in metamorphic *Crassostrea angulata* larvae [Yang et al., 2012], and in one-year-old spat in response to a herpes virus [Jouaux et al., 2013]. Collectively these studies demonstrate that this stress signalling pathway is operational during early ontogenesis. Furthermore, a series of studies based on NGS larval transcriptomes provided preliminary reports in tracing constituents of these cascades in *Meretrix meretrix* [Huan et al., 2012b] and in *O. lurida* [Timmins-Schiffman et al., 2013].

2.4.2.4 *NF-κB pathway*

Over the past decade, the NF-κB signal transduction pathway appeared to play a central role in the innate immune defence system of bivalves. NF-κB polymerizing enzymes are involved in multiple aspects of cell differentiation and immunity which requires the co-regulation of numerous inflammatory genes during development, maintenance, and defence [Montagnani et al., 2008; Qiu et al., 2009; Vallabhapurapu and Karin, 2009]. NF-κB dimers are usually sequestered in the cytoplasm by an inhibitory protein κB (IκB) that forms an NF-κB:IκB heterodimer. After activation of the TLR pathway, cascade reactions induce the phosphorylation of IκB for ubiquitination, which leads to its degradation. The active NF-κB is then extricated and translocated to the nucleus [Zhou et al., 2013], where it binds to its DNA response element and then initiates the recruitment of general transcription factors and the basal transcription machinery [Ghosh et al., 2012]. In bivalves, components of the NF-κB pathway were firstly been described in *C. gigas* [Montagnani et al., 2004]. The relative transcripts could be detected throughout the early stages of development [Tirapé et al., 2007; Genard et al., 2013], suggesting that this pathway is necessary during development. A series of recent transcriptomic NGS studies enabled the identification of interrelated NF-κB homologs responsible of the activation of TLRs in *M. edulis* [Philipp et al., 2012; Tanguy et al., 2013b], *R. philippinarum* [Moreira et al., 2012a] and *C. virginica* [Zhang et al., 2013]. The selective regulation of NF-κB target genes adds another level of complexity to the study

of its molecular mechanisms in bivalves [Ghosh et al., 2012]. NF-κB may also functions as an inhibitor of serine proteases and metalloproteinases, two key factors in pathogen onset [Johnson, 2013]. In veliger larvae a synergistic regulation of tissue inhibitor of metalloproteinase (TIMP) and NF-κB has been detected just 48 hours after bacterial challenge [Genard et al., 2013]. Taken together, these data illustrate how complex the hidden NF-κB pathways are within maintenance and defence processes, especially those occurring in early development.

2.4.2.5 *Less explored signalling pathways*

The janus kinase/signal transducer and activator of transcription (JAK/STAT) is amongst the many pathways known to be triggered during an immune response in bivalves. The JAK/STAT pathway plays a major role in intracellular signal transduction of cytokine receptors [Shaposhnikov et al., 2013]. The functional presence of this pathway in bivalves has been firstly documented in *M. galloprovincialis* after bacterial challenge, after which a cytokine response was recorder [Canesi et al., 2003]. Recent advancements in transcriptomics and data mining pipelines have enabled the discovery of JAK/STAT homologs in *M. edulis* [Philipp et al., 2012]. It is however to our knowledge that this preliminary study is the only one to acknowledge the activation of this pathway, while to date no other information has yet to be found concerning larval stages.

Moreover, the immune system can be modulated by lipid-related pathways, notably after eicosanoid release that triggers a variety of inflammatory pathways [Smith and Murphy, 2002]. Eicosanoids are highly qualified oxygenated molecules that are recognised as universally important to all life forms. They are implicated in various signalling pathways, tend to promote inflammation, and also play a role in the central nervous system [Smith and Murphy, 2002; Phillis et al., 2006]. Polyunsaturated fatty acids (PUFAs) like eicosapentaenoic acid (EPA) and arachidonic acid (ArA) are precursors of eicosanoids, which include prostaglandins, leukotrienes, and hydroxy derivatives. Although an interrelated activation

of the eicosanoid pathway has been successfully demonstrated through phospholipases and lipoxigenases during immune response in invertebrates [Stanley et al., 2009], molecular studies on eicosanoids and lipid-related pathways are still lagging [Delaporte et al., 2006; Seguineau et al., 2011]. Recent work on *C. gigas* larvae exposed to a bacterial pathogen, brings additional insights about the up-regulation of a phospholipase A2 and a AA15 lipoxigenase correlated with a decline of the eicosanoid precursor AA, all of which highlight the importance of this pathway in the immune response in early development [Genard et al., 2013].

2.4.3 *Immune effectors*

An efficient immune response leads to the production of inflammatory effectors that neutralise and eliminate unfavorable threats. In the following section we give an overview of the most recent accomplishments in the fields of antimicrobial peptides, lysozymes, cytokines and acute phase processes that depend on perforins, immunocell activation, and antioxidant enzymes.

2.4.3.1 *Antimicrobial peptides*

Antimicrobial peptides (AMPs) are amongst the most active effectors of the innate immunity. They display a variety of action mechanisms, including microbe annihilation through membrane disruption or altering metabolic processes related to cell wall synthesis (see review by [Song et al., 2010]). Several antimicrobial peptide (AMP) have been found in bivalves, they include defensins, mytilins, myticins, and mytimycins [Zhao et al., 2010; Tanguy et al., 2013a; Defer et al., 2013; Husmann et al., 2013]. Additional data mining based on NGS analyses have already predicted several nucleotide sequences related to AMP for *M. edulis* [Philipp et al., 2012; Tanguy et al., 2013b] and *C. virginica* [Zhang et al., 2013]. Developmental immunity remains little explored; although mytilin and defensin have both been identified in

peri-metamorphic larvae [Mitta et al., 2000; Tirapé et al., 2007; Genard et al., 2013; Balseiro et al., 2013]. Similarly, a defensin homolog in *M. meretrix* has been identified in a preliminary NGS study where its abundance was interrelated with the progression of larval stages [Huan et al., 2012b].

2.4.3.2 *Lysozymes and proteases*

Lysozymes are ubiquitously efficient bacteriolytic enzymes that act by hydrolysing the β-(1, 4) bonds between N-acetylglucosamine and N-acetylmuramic acid found in the PGN bacterial cell walls [Callewaert and Michiels, 2010]. It has been widely accepted that lysozymes are crucial polymerizing enzymes in innate immunity. Three distinctive groups of lysozymes emerge after sequence analysis; from invertebrates (i-type), chicken type (c-type), and goose type (g-type). Aside from bivalves, only very few invertebrates still carry functional g-type lysozymes, which naturally occurs in scallops for example [Zhao et al., 2007] or even freshwater mussels [Wu et al., 2013]. The activity of several of these lysozymes has already been characterised in bivalves (see review by [Gestal et al., 2008]). High-throughput transcriptome analysis revealed the presence of an operational group of synexpressed transcripts triggered in response to an inflammatory reaction in *M. edulis* [Tanguy et al., 2013b] and in the Baltic clam *Macoma balthica* [Pante et al., 2012]. Lysozyme-like and hydrolytic enzyme activities were previously reported in *C. gigas* larvae and juveniles [Luna-González et al., 2004], while expression patterns of lysozymes-related genes were recently observed in the clam *M. meretrix* [Yue et al., 2013c] and the scallop *C. farreri* [Yue et al., 2013a] during larval development. Finally, lysozyme homologs were found in *M. meretrix* larvae using a deep mRNA sequencing approach [Huan et al., 2012b]. Alongside lysozymes, host defence mechanisms also depend on the synthesis of several lysosomal proteases for pathogen inactivation and cell lysis, which include cathepsins. These serine proteases are triggered in oyster larvae and the adult clams after bacterial infection [Genard et al., 2012; Niu et al.,

2013] and in *M. meretrix* when larvae acquire feeding traits [Wang et al., 2008; Huan et al., 2012b].

2.4.3.3 *Cytokines*

Cytokines are signallisation molecules that are known as mediators and regulators of the immune defence and other cellular processes (e.g., neurogenesis and apoptosis) [Hughes et al., 1992]. A wide array of defence mechanisms in bivalves is modulated by a cytokine-triggered network similarly to that found in vertebrates [Hughes Jr et al., 1990; Ottaviani et al., 2004; Philipp et al., 2012]. Some studies have found actively conserved domains of common cytokines and their receptors in a number of marine bivalves [Renault et al., 2011; Xu et al., 2012; Moreira et al., 2012b]. The tumor necrosis factor α (tumor necrosis factor (TNF)α) is a form of cytokine synthesised by bivalve hemocytes in response to pathogenic microorganisms, notably in the presence of LPS from Gram-negative bacteria [Lelong et al., 2007; Li et al., 2012]. The inflammatory response generated post-LPS induction will trigger multiple signalling pathways, for example MAPK (p38-MAPK, JNK, ERK) and NF-κB [Canesi et al., 2004; Betti et al., 2006; Qiu et al., 2009]. Other than the information from these studies on marine bivalves mentioned above, data on cytokines, their receptors, and the interrelated elements can be limited or hard to find. Deep sequencing on mRNA in adults of *M. edulis*, *R. philippinarum*, and *C. virginica* facilitated recently novel discoveries of multiple cytokines and their putative co-regulation with other inflammatory effectors [Moreira et al., 2012a; Philipp et al., 2012; Zhang et al., 2013; Tanguy et al., 2013b]. The complete functional behaviour of cytokine-related genes at larval stages has not yet been thoroughly described. However, transforming growth factor β (TGFβ) superfamily of ligands, an immuno-inhibitory cytokine, has been found to mediate an important role in embryonic development and metamorphosis of *C. gigas* [Herpin et al., 2005]. The allograft inflammatory factor-1 (AIF-1) is also a cytokine-mediated molecule, it enhances the cell cycle progression and promotes specific immune reactions from gastrula to veliger stages and in *C. farreri*

adults [Wang et al., 2013a]. Lastly, TNF-related transcripts have been predicted in *O. lurida* during a preliminary attempt to enrich the immunome database of environmental model species [Timmins-Schiffman et al., 2013]. They are also thought to play a role in several pre-apoptotic downregulated genes in oysters [Zhang et al., 2012a, 2013] and mussels [Philipp et al., 2012; Tanguy et al., 2013b].

2.4.3.4 *Perforins and attack membrane proteins*

Specific molecules, known as the membrane attack complex and perforin (MACP) family members, are responsible for the formation of holes in the membranes of pathogen cells [Baran et al., 2009]. The cytotoxic activity of perforin is crucial for eliminating infected self-cells. In the presence of calcium, perforins oligomerise at the membrane of the targeted cell and are transformed into a pore-forming aggregate, which causes cell death [Wiens et al., 2005]. A series of available reports describe MACP involvement in anti-viral defences in *C. gigas* after its genetic up-regulation [Renault et al., 2011]. Transcriptome analysis using massive parallel sequencing in *M. edulis* predicted several MACP/perforin domains that might belong to the complement system [Philipp et al., 2012]. Using a proteomic approach, another member of the MACP class called apextrin was found to be up-regulated in adult *M. galloprovincialis* in response to pollutants [Ji et al., 2013b]. It has also been shown that apextrin is expressed at the blastula stage of *M. galloprovincialis*, as well as in adult tissue and hemocytes treated with bacteria and PAMPs [Estévez-Calvar et al., 2011]. In *M. galloprovincialis* a further significant increase in perforin expression has been registered from spatfall to post-larvae and in adult hemocytes [Balseiro et al., 2013]. The LPS-binding/bactericidal-permeability increasings (LBP/BPIs) are membrane attack proteins that permit the opsonisation of small number of Gram-negative bacteria through the increase of their membrane permeability [Gonzalez et al., 2007]. Described firstly in *C. gigas* [Gueguen et al., 2003; Gonzalez et al., 2007; Schmitt et al., 2012], transcripts coding for LBP/BPI have also been found in *M. edulis* using NGS analyses [Philipp et al., 2012]. Their activation during a bacterial

challenge at larval stages in *C.gigas* [Genard et al., 2013] and *C. farreri* [Yue et al., 2013a] indicates that LBP/BPI may promote highly-sensitive pro-inflammatory responses during early development as suggested by Tirape et al. (2007).

2.4.3.5 *Hemocytes and phagocytosis*

Bivalve hemocytes facilitate an integrated host response against pathogens through phagocytosis and cytotoxic reactions, which include hydrolytic enzymes, the coordinated release of different enzymes (e.g., lectins, AMPs, perforins), and free-radical production (Figure 2.2C). Mobilised hemocytes permit the final destruction of foreign organisms [Song et al., 2010]. Immune functions are mainly dependent on the phagocytic activity [Canesi et al., 2012], the number of circulating hemocytes [Malagoli et al., 2007], and the concentration of intracellular Ca^{2+} [Ottaviani et al., 1997; Bettencourt et al., 2009]. While phagocytic cells have been characterised in most adult bivalves, data about differentiated immunocells and their functions during ontogenesis remain constrained to the practical approach used to study them. Elston (1980) have successfully observed immunocells (described as cœlomocytes) containing bacterial remnants in the visceral cavity of *C. virginica* veliger larvae, while Tirape et al. (2007) suggested that hemocytes could be present early during the gastrula/trochophore stages. High-throughput transcriptome sequencing further enabled the identification of hemocyte-specific genes, then traced their expression back to the cleavage and blastula stages, it has been suggested that hemocytes may be the least differentiated cells in the fertilized oyster eggs [Zhang et al., 2012a]

2.4.3.6 *Oxidative burst and cytoprotection*

Phagocytosis is generally accompanied by a production of reactive oxygen species (ROS) or intermediates (ROIs) that are locally produced in the phagolysosome in order to degrade phagocyzed substances, or nitric oxide (NO) released in the extracellular matrix by hemo-

cytes (Figure 2.2D). Upon recognition of foreign threats, the innate immune defence in bivalves is triggered and a respiratory burst is usually registered afterwards [Labreuche et al., 2006; Lambert et al., 2007]. Aside from the degradation of phagocyzed material, ROS manage to induce dangerous intra-cellular damage, notably to complex structures, including lipids, proteins, and DNA. The oxidative burst can nonetheless be reduced before attaining critical levels by means of highly conserved and tightly co-regulated antioxidants, which may include Catalase (CAT), glutathione peroxidase (GPX), glutathione-S-transferase (GST), superoxide dismutase (SOD), and peroxiredoxin (PRDX). Many nucleotide coding sequences associated with these enzymes have been identified in bivalves and their activity is well detailed elsewhere (see review by [Song et al., 2010]). Preliminary studies based on high-throughput sequencing led to new transcript discoveries of antioxidant enzymes in adult *M. edulis* [Tanguy et al., 2013b] and *C. virginica* [Zhang et al., 2013], or at larval stages in *M. meretrix* [Huan et al., 2012a] and *O. Lurida* [Timmins-Schiffman et al., 2013]. In addition to gene behaviour of SODs, CATs, or PRDX has been monitored during early ontogeny in *C. gigas* [Tirapé et al., 2007], *C. virginica* [Genard et al., 2012], and *C. farreri* [Yue et al., 2013a], and their enzymatic activity has been measured in *Crassostrea corteziensis* juveniles [Campa-Córdova et al., 2009], or in challenged oyster larvae [Genard et al., 2011, 2013]. Collectively, the differential and specific occurrence of these molecules highlights their protective importance early in development. Apart from antioxidant defences, bivalves activate various cytoprotective proteins to avoid cell degradation or to repair damaged molecules, among which HSPs, stress-response chaperones, that bind to damaged or misfolded proteins, either promoting their repair or elimination [Daugaard et al., 2007]. Up-regulation of HSPs has been observed after *Vibrio* infection in adult bay scallops [Gao et al., 2008] and in pacific oysters [Fleury et al., 2009; De Lorgeril et al., 2011]. Several homologs of HSPs have been recently identified through next-generation transcriptome analyses in *R. philippinarum* [Moreira et al., 2012a], *M. edulis* [Tanguy et al., 2013b], and in *O. lurida* larvae [Timmins-Schiffman et al., 2013]. Up-regulation of HSP70 after bacterial infection in *C. virginica* and *C. gigas* larvae [Genard

et al., 2012, 2013] implies that HSPs are successfully activated during the early development of the bivalve's defence system.

2.5 OVERVIEW OF MARINE SCIENCE: NGS AS NEW POTENTIAL

High mortality rates occur throughout early development stages. The effects of different factors, which were detailed earlier, render developing larvae susceptible to abnormal growth or reduces the chances of survival. Bivalve larvae alter their morphogenetic structure during metamorphosis, which alters their high mobile form into a typical sessile adult. Knowledge of the mechanisms responsible for maintaining cell homeostasis are of crucial concern if we are to understand how life-history traits interact and evolve.

2.5.1 *Measuring gene expression*

Next-generation methods are being adopted for convenience in assessing large-scale biological processes [Forrest and Carninci, 2009] and for other molecular research, specifically on lower level taxa (Figure 2.3F and Table 2.1) [Shokralla et al., 2012]. Gene behaviour is usually monitored by quantifying transcript abundance, which is a direct measure taken from cells, tissue, or an entire organism. The most reliable genome-wide applications utilized today are microarrays and RNA-seq (and their variants, e.g., chromatin immunoprecipitation, ChIP-seq) [Gilad et al., 2009; Bourlat et al., 2013]. Microarray manufacturers provide catalogue designs, mostly for human and other classic model organisms. Oligoarrays may nevertheless be adequate for immune studies on non-model large-genome organisms [Varotto et al., 2013], although transcript coverage is not as wide compared to that of the NGS [Xu et al., 2013]. Depth of coverage is the first detail to consider before carrying out a transcriptome analysis since gene monitoring and sequencing each have different dependencies on the number of reads required. Although this approach exists still, it might be even more time-consuming and labor-intensive than NGS application [Gilad et al., 2009; Westermann et al.,

2012], not to mention the massive signal-to-noise ratio that can plague the generated redundant datasets, which renders the analysis less informative (Figure 2.3C) [Bourlat et al., 2013].

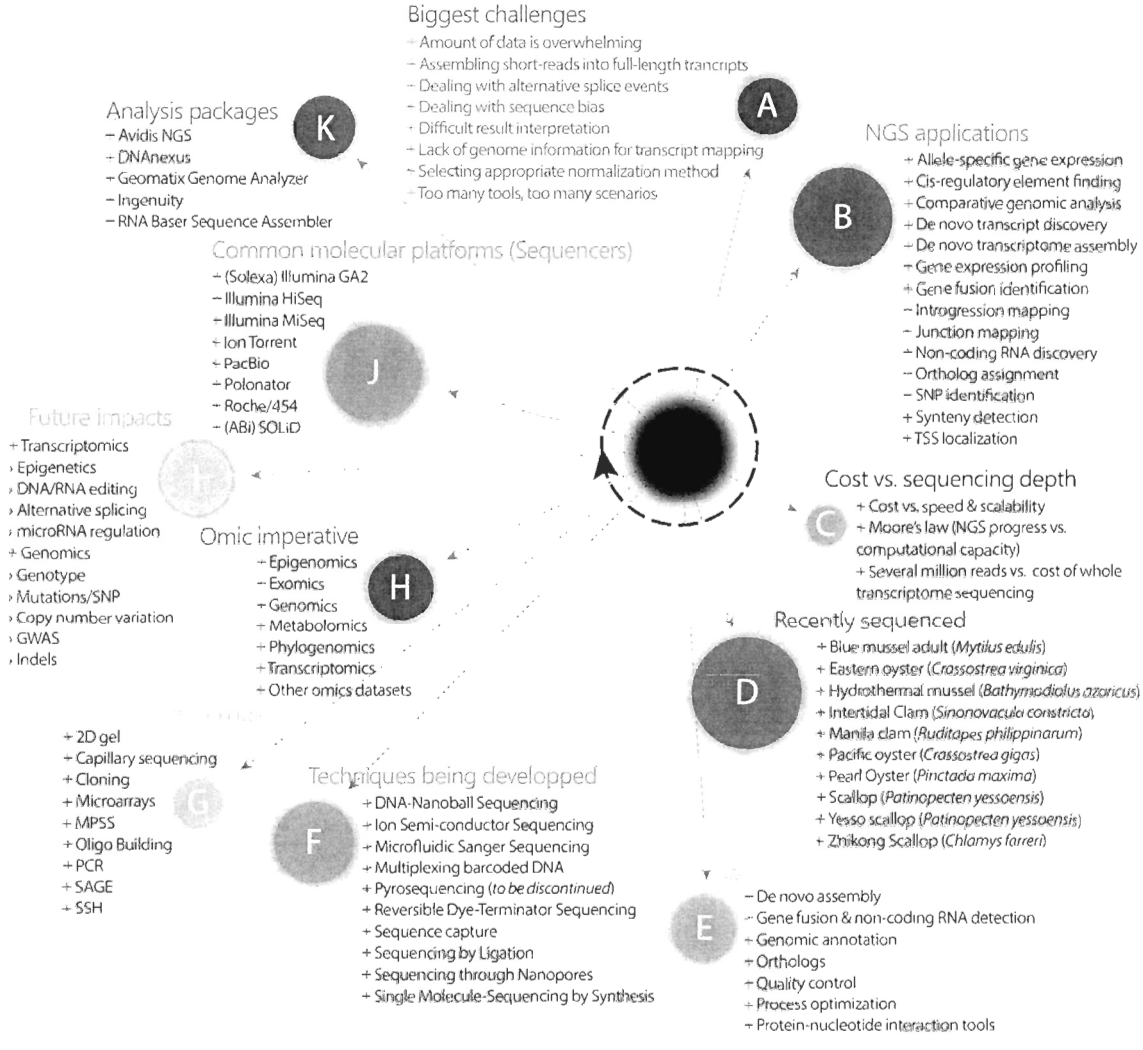

Figure 2.3: **New future avenues of research in marine bivalves.** NGS, next generation sequencing; GWAS, genome-wide association study; MPSS, massive parallel signature sequencing; PCR, polymerase chain reaction; SAGE, serial analysis of gene expression, SNP, single nucleotide polymorphism; SSH, suppression subtractive hybridization; TSS, transcription-start site.

Table 2.1: **Summary of NGS recent immune-related discoveries in marine bivalves**

Species	Defense transcripts identified	NGS	Reference
Crassostrea virginica	Different adult tissues and hemolymph	RNA-seq	[Zhang et al., 2013]
	Adult	RNA-seq	[Gomez-Chiarri, 2013]
Ostrea lurida	Whole larvae	RNA-seq	[Timmins-Schiffman et al., 2013]
Chlamys farreri	Larvæ, and different adult tissues (P)	454	[Wang et al., 2013b]
Argopecten irradians	Adult eye tissue	454	[Pairett and Serb, 2013]
Placopecten magellanicus	Adult eye tissue	454	[Pairett and Serb, 2013]
Mytilus edulis	Adult hemolymph	454	[Tanguy et al., 2013b]
	Different adult tissues and hemolymph	454	[Philipp et al., 2012]
Pinctada spp.	Different adult tissues (P)	RNA-seq	[Huang et al., 2012]
	Adult mantle tissue	454	[Shi et al., 2013]
Pteria penguin	Different adult tissues (P)	RNA-seq	[Huang et al., 2012]
Chamelea gallina	Adult muscle tissue	454	[Coppe et al., 2012]
Meretrix meretrix	Whole multistage larvae (TS)	454	[Huan et al., 2012b]
Crassostrea gigas	Whole multistage larvae, and different adult tissues (TS)	RNA-seq	[Zhang et al., 2012a]
Ruditapes philippinarum	Adult hemolymph	454	[Moreira et al., 2012a]
	Different adult tissues	454	[Milan et al., 2011]
Patinopecten yessoensis	Larvæ, and different adult tissues (P)	454	[Hou et al., 2011]
Laternula elliptica	Adult mantle tissue	454	[Clark et al., 2010]

(TS) timeseries study throughout early development; (P) pooled results from larvae and/or adult tissue; RNA-seq: sequencing mRNAs into short-length reads; 454: pyrosequencing mRNAs into longer, about 400 base nucleotide reads

Meanwhile, many researchers are switching to fast, genome-wide NGS techniques (Table 2.1), including studies about microsatellites and single nucleotide polymorphism identification [Gallardo-Escárate et al., 2013; Jones et al., 2013] (Figure 2.3F). They are discovering the advantages of analyzing whole modules of defence genes [Zhang et al., 2012b] rather than specifically selected orthologs [Niu et al., 2013]. These reports enable advanced genome-wide studies that reconstruct synergistic regulations between operational immune molecules and associate them with tightly co-regulated complex defence mechanisms [Meng et al., 2013]. Indeed, as with the case of all new techniques, scientists can find themselves inundated with a plethora of options to treat their data (Figure 2.3B) [Altschul et al., 2013]. This novel focus on 'omics' enriches the field with the regular appearance of progressive new tools, and different packages (Figure 2.3H). In spite of theses divergent tools, where each case reveals a unique class of information, one can still easily find documented guides to work with NGS datasets. Thus, NGS analysis involves nothing more difficult than to choose from which microarray normalisation step to consider. With all the practical tools available, it is even possible to use upgraded packages of microarray preprocessing for NGS datasets in a pipeline analysis [Soneson and Delorenzi, 2013].

2.5.2 *A new pipeline analysis*

Conventional approaches can only take us so far regarding molecular understanding of cells and interrelated immune mechanisms. Despite their long use, these approached suffer from a design bias (probe-dependent) that is definitely a drawback when trying to cover all aspects of the immunome. These applications, suppression subtractive hybridization (SSH) for example (Figure 2.3G), rely on a singular analytical pipeline (metadata indexing) without the use of heterogeneous validated data (assembly correction). While reverse-transcription quantitative polymerase chain reaction (RT-q(PCR)) assays are strictly dependent on functional coding sequences, they can often poorly reflect the reaction kinetics of interrelated

genes [Du et al., 2013], they are best used to confirm a differential expression in a cost-effective way. This becomes a bottleneck for immune gene discoveries, when publicly indexed genes outweigh undocumented novel ones [Araya et al., 2010]. This is true for two reasons: firstly the depth of coverage and secondly the less-integrative pipeline analysis. Therefore, we will not cover immune mechanisms in detail since inferences would not be sound enough to be considered accurate. In contrast, NGS techniques provide computationally intensive procedures (Figure 2.3F) to correct a transcriptome while assembling it by alignment with already verified libraries, known as reference genomes. Thus they enable the end-user to study inter- and intra-species variations. In addition, the depth of coverage and sequencing length combined with empirical corrections enable a better construction of more accurate and less noisy datasets [Xu et al., 2013]. For example, a recent study demonstrated the function of a novel scaffolder to expand the pearl oyster's draft genome [Takeuchi et al., 2012], which combines gene fragments into larger sequences and thus resolves fragmentation of transcribed RNAs [Xue et al., 2013]. Non-model species without genomic resources will definitely benefit from these data mining capabilities, specifically in larval studies, since full-length transcriptomes can be assembled with an increasing sensitivity [Grabherr et al., 2011]. Furthermore, NGS analytical pipelines are superior to conventional methods for detecting both coding and ncRNA, thus discovering hidden constituent parts of various GRNs [Forrest and Carninci, 2009]. Environmental studies will also benefit, notably with the development of particular DNA-barcoding multiplexing methods [Wong et al., 2013] or sequence-capture templates for identifying DNA targets [DuBose et al., 2013], which can be more cost-effective than other immune molecular methods. Collectively, these new strategies are important in the fields of development and immunity, where data could remain scarce or fragmented otherwise. Consequently, these techniques help to close the gap between co-regulated genes by network inference or discoveries of hidden components.

2.5.3 *Leading the way for new and sensitive discoveries*

NGS have come a long way since their first release [Gilad et al., 2009]. Challenges still exist but can be overcome when it comes to parallel pipeline analyses (Figure 2.3A), which are already being simplified for higher taxonomic levels [Garg and Jain, 2013]. The automation of methods that discover defence-related candidate genes, splice isoforms, and orthologs will introduce us to a new class of immune studies and likely alter our view of the defence in non-model organisms (Figure 2.3I). Understanding the complex regulatory networks based on protein cooperation is necessary to expand our understanding of the cell. These studies are useful for revealing not only the abundance of transcripts in different tissues, but also at a single-cell level [Katayama et al., 2013], or even at the exon level [Bi and Davuluri, 2013].

2.5.4 *Exploring the hidden and various aspects of life-history traits*

Molecular pathways required for early development, including those involved in non-self recognition, signalisation, and organism maintenance (Figure 2.2), are often used again by the innate immune system in adults. Processes related to cell growth may correlate with those of cellular differentiation during tissue repair [Zhang et al., 2012a], apoptosis involved in early embryogenesis (e.g., angiogenesis and neurogenesis) and defence in later stages [Kiss, 2010], or interrelated metabolic and immune signalling pathways (cell stimulation) [Wang et al., 2012b]. Furthermore, interactions between tightly co-regulated synexpressed genes during early ontogenesis are not fully understood. Foremost, gene behaviour during development is often considered as the major obstacle because of the requirement for the time-sensitive capture of differential expression patterns during fast biometric and morphometric changes in larvae [Pernet et al., 2006a]. This demonstrates a putative hidden association of different immune mechanisms that might appear to be co-dependent but are untraceable with a traditional molecular technique.

2.5.5 *The rise of non-model organisms*

As the cost of obtaining next-generation RNA sequence reads continues to fall, it is anticipated that huge amounts of heterogeneous metadata on marine bivalves will become available for integration (Figure 2.3H). This is required if we are to reveal the complete picture of enhanced/silenced processes that regulate cellular cycle, defence, maintenance and death, which will contribute enormously to the immunity field of bivalve larvae (Figure 2.2). Although it is not difficult to obtain hundreds of millions of reads [Hornett and Wheat, 2012; Tanguy et al., 2013b], computational tools to accurately predict and visualise the data are still being developed and refined (Figure 2.3A). It is of the utmost importance to ensure their proper propagation (Figure 2.3K), the best uniformity in data exploration, and elimination of the proliferation in diverging variants, that is gene and protein data fragmentation. Novel computational methods are heavily based on statistical analyses, which can be used to reveal hidden variances and provide accurate hypothesis-based estimates that have largely gone unexplored. For example, tests conducted on the heterogeneity between samples and the mixing effect of cell mRNA, can enrich our repertoire with new ways in inferring the when and why special immune genes respond to cellular threats [Gong et al., 2011]. Consequently, marine biologist will be able to investigate in an unbiased manner tissue differentiation, a process that is tightly-regulated throughout early development and is the basis of larval survival and performance [Dyachuk, 2013]. Some strategies based on conventional approaches (Figure 2.3G) are still used to discover new proteins of the immune defence system during ontogenesis [Huan et al., 2012a], but this is likely because new methods have not become widely known [Timmins-Schiffman et al., 2013]. Functional genomics have been dramatically improved over the last few years by studying transcriptome dynamics and more recently with novel NGS procedures (Figure 2.3H). Regardless of which approach scientists choose, transcriptome analysis will provide a reliable and rich set of data to answer to their biological questions. Even with the various challenges associated with the old and new techniques, it is possible, for example, to develop a custom

workbench based first on de novo next-generation deep sequencing combined with a conventional application for repeated monitoring [Milan et al., 2011]. Finally, gene-oriented low-scale enzymatic studies could then bridge the gap between molecular and physiological data, while ontogeny-related analyses will enable a close monitoring of these genes on a short progressive timescale, which will give more meaningful insights into the formation of cell resistance.

2.6 CONCLUSION

In recent years, interest in bivalve immunity has increased considerably due to the potential of this model organism in environment assessment [Gauthier-Clerc et al., 2013]. NGS technologies such as the Illumina, 454, and Ion Torrent platforms (Figure 2.3J) are providing new opportunities in high-throughput research for functional genomic studies and the interrogation of genomes and transcriptomes at unparalleled resolutions [Mardis, 2008; Romero et al., 2012; Shokralla et al., 2012]. Their analytical pipelines are becoming unbiased multitools that can be used on any dataset and without the need to repeatedly sequence the same transcriptome in order to identify different genetic principles. Early ontogenic studies can definitely profit from these novel methods in order to map cell molecules; the use of first-generation techniques on these unexplored development stages revealed interesting deep molecular connections to phenotypic plasticity related to defence and survival in marine bivalves. NGS techniques also promises future integration of valuable metadata and groundbreaking turnaround tools, all of which can be used to assess the remaining challenges that hinder our understanding about the dark side of the innate immune system of bivalves.

2.7 ACKNOWLEDGEMENTS

This review was written during S. Bassim's Ph.D research and was supported financially by a Strategic Grant from the Natural Sciences and Engineering Research Council of Canada (STPGP 364918). The authors would like to thank Laure Devine for text revision.

3

IDENTIFICATION DES RÉGULATEURS GÉNÉTIQUES AU COURS DU DÉVELOPPEMENT LARVAIRE DE *MYTILUS EDULIS*

Bassim Sleiman[1,2], Tanguy Arnaud[3], Genard Bertrand[1], Moraga Dario[2], Tremblay Réjean[1]
[1] *Institut des Sciences de la mer de Rimouski, Université du Québec à Rimouski, 310 allée des Ursulines, Rimouski, Québec, G5L 3A1 Canada*
[2] *Laboratoire des Sciences de l'Environnement Marin, Institut Universitaire Européen de la Mer, Université de Bretagne Occidentale, Rue Dumont d'Urville, 29280 Plouzané, France*
[3] *UPMC Universite Paris 6, UMR 7144, Genetique et adaptation en milieu extreme, Station biologique de Roscoff*

3.1 RÉSUMÉ

Ce deuxième article intitulé *Identification of Mytilus edulis genetic regulators during early development*, fut corédigé par moi-même ainsi que par mes collègues les docteurs Arnaud Tanguy, Bertrand Génard, Dario Moraga et Réjean Tremblay. Il fut publié pendant le deuxième trimestre 2014 dans la revue *Gene*. En tant que premier auteur, ma contribution à ce travail fut l'idée originale, la recherche sur l'état de l'art, la préparation, le traitement et l'analyse des données, le design expérimental du pipeline, la rédaction de l'article et les informations supplémentaires. Le docteur Arnaud Tanguy a aidé dans l'assemblage du transcriptome, l'idée originale et la révision de l'article. Les docteurs Bertrand Génard, Dario Moraga et Réjean Tremblay ont partagé à la révision de l'article. Une version abrégée de cet article a été présentée au congrès *Ecotoxicologie des Bivalves Marins* (ECOBIM) à Reims (France) en 2013.

Comprendre les mécanismes par lesquels la croissance et la survie conduisent un organisme à son adaptation est fondamental pour répondre aux questions de la biodiversité et de l'évolution, en considérant notamment les changements climatiques globaux. Nous annonçons ici le séquençage et l'assemblage de novo du transcriptome des larves de moules en utilisant un séquençage de nouvelle génération (NGS). Environ 50,383 contigs ont été assemblés. Plus de 15,000 transcrits ont été associés à des protéines fonctionnelles et ensuite ont servi à fabriquer des puces qui ciblent les premiers stades du développement pré et post-métamorphique. Les stades étudiés incluent des transitions majeures dans l'organisation des tissus de la moule bleue, *Mytilus edulis*. Un traitement de séquence de 3,633 transcrits différentiellement exprimés a permis de découvrir 12 clusters de gènes impliqués dans divers changements physiologiques et morphologiques de la larve. Des interactions ont été observées entre les gènes différentiellement exprimés et 16 régulateurs de la transcription. Les réseaux de régulation géniques formés par cette association ont été décrits suivant leur expression différentielle durant l'ontogenèse. C'est la première étude à grande échelle basée sur un assemblage de novo du transcriptome des larves de *M. edulis*. En intégrant des techniques de reconnaissance de modèles d'expression, nous avons associé à l'échelle de l'ensemble du génome le fonctionnement des processus de développement à des réseaux de gènes mutuellement régulés. L'intégration des données NGS et puce nous a fourni l'information clé pour comprendre les mécanismes moléculaires précoces des stades larvaires de la moule bleue.

Identification of *Mytilus edulis* genetic regulators during early development

3.2 HIGHLIGHTS

* The First draft of *Mytilus edulis* transcriptome has been assembled,
* *Mytilus edulis* microarrays are available for timeseries profiling,
* Timeseries datasets for larvae transcriptome are available for predictive analysis,
* New regulation factors in muscle and neural processes has been discovered,
* Inferred networks showed mechanistic regulation of genes and functional morphology,

3.3 ABSTRACT

Understanding the mechanisms that enable growth and survival of an organism while driving it to the full range of its adaptation is fundamental to the issues of biodiversity and evolution, particularly regarding global climatic changes. Here we report the Illumina RNA-sequencing (RNA-seq) and de novo assembly of the blue mussel *Mytilus edulis* transcriptome during early development. This study is based on high-throughput data, which associates genome-wide differentially expressed transcript (differentially expressed transcript (DET)) patterns with early activation of developmental processes. Approximately 50,383 high-quality contigs were assembled. Over 8,000 transcripts were associated with functional proteins from public databases. Coding and non-coding genes served to design customized microarrays targeting every developmental stage, which encompass major transitions in tissue organization. Consequently, multi-processing pattern exploration protocols applied to 3,633 DETs helped discover 12 unique coordinated eigengenes supposedly implicated in various physiological and morphological changes that larvae undergo during early develop-

ment. Moreover, dynamic Bayesian networks (DBNs) provided key insights to understand stage-specific molecular mechanisms activated throughout ontogeny. In addition, delayed and contemporaneous interactions between DETs were coerced with 16 relevant regulators that interrelated in non-random genetic regulatory network (GRNs). Genes associated with mechanisms of neural and muscular development have been characterized and further included in dynamic networks necessary in growth and functional morphology. This is the first large-scale study being dedicated to *M. edulis* throughout early ontogeny. Integration between RNA-seq and microarray data enabled a high-throughput exploration of hidden processes essential in growth and survival of microscopic mussel larvae. Our integrative approach will support a holistic understanding of systems biology and will help establish new links between environmental assessment and functional development of marine bivalves.

Abbreviations: NGS, next-generation sequencing; FDR, false discovery rate; FCM, fuzzy c-means; DET, differentially expressed transcripts; GRN, genetic regulatory networks; ECM, extracellular matrix; GO, gene ontology; DPF, day post-fertilization; NCBI, National Center for Biotechnology Information; PIR, Protein Information Resource; SAM, significance analysis of microarray; ANOVA, one-way analysis of variance; PCA, principal components analysis; AIC, Akaike's information criterion; EBDBN, empirical Bayes dynamic Bayesian network; PC, principal component; RDA, redundancy analysis; BP, biological process; MF, molecular function; CC, cellular component; TF, transcription factor; PIP5K1A, phosphatidylinositol 4-phosphate 5-kinase type-1 alpha; GATAD1, gata zinc finger domain-containing protein 1; ATRX, transcriptional regulator ATRX; CnVAS1, vasa-like protein; GTPase, guanosine triphosphatase; DERL2, derlin-2; ACTL, actin-like protein; Hsp70, heat-shock protein 70; ARF4, ADP-ribosylation factor 4; ROPN1L, ropporin-1-like protein; B9D1, B9 domain-containing protein 1; YWHAZ, 14-3-3 protein zeta; PACRG, Parkin coregulated gene protein; CDC42, cell division control protein 42; CSNK2B, casein kinase II subunit beta isoform 1; SMARCD1, SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily d member 1; Med6, mediator of RNA polymerase II transcription

subunit 6; ANAPC7, anaphase-promoting complex subunit 7; CBX3, chromobox protein 3; TNKS2, a chain human tankyrase 2-catalytic parp domain; Rho1, RAS-like GTP-binding protein Rho1-like; HISTH2A, histone H2A; UBN1, ubinuclein-1; ABCF3, ATP-binding cassette sub-family F member 3; DNMT1, DNA (cytosine-5)-methyltransferase 1; PTPN13, tyrosine-protein phosphatase non-receptor type 13; SWI/SNF, SWItch/Sucrose Non Fermentable; PBX1, pre-B-cell leukaemia transcription factor 1 isoform 2; NRARP, notch-regulated ankyrin repeat-containing protein; ARF6, ADP-ribosylation factor 6.

Keywords: Bivalve; Gene regulatory networks; Development; Next-generation sequencing; Transcriptome

3.4 INTRODUCTION

Many species evolved disparate body plan transitions [Bishop et al., 2006a], driven by coordinated genetic programs and triggered by physiological states of competence. Consequently, researching interlocking mechanisms becomes an essential task to understand this mosaic development. *Mytilus edulis* retained a polymorphic life-cycle during which early developmental stages are planktotrophic [Bayne, 1965]. At the first dispersive pelagic form, mussel larvae are efficient in sustaining fluctuating conditions, begin to feed at the veliger stage thus increasing their energy reserves until the pre-metamorphic competence phase.

Larval development chronologically unfolds following a mechanical remodelling of the form and behaviour, which requires controlled mechanisms engaged in tissue distribution and organ system differentiation [Morse, 1990; Ackerman et al., 1994] (Fig. 3.1A,B). Fertilization occurs externally and embryos are characterized with fast cell division that leads to their segmentation [Cao et al., 2004]. Pre-competent metamorphic larvae are then differentiated into motile forms with successive development of various shell forms [Bayne, 1971].

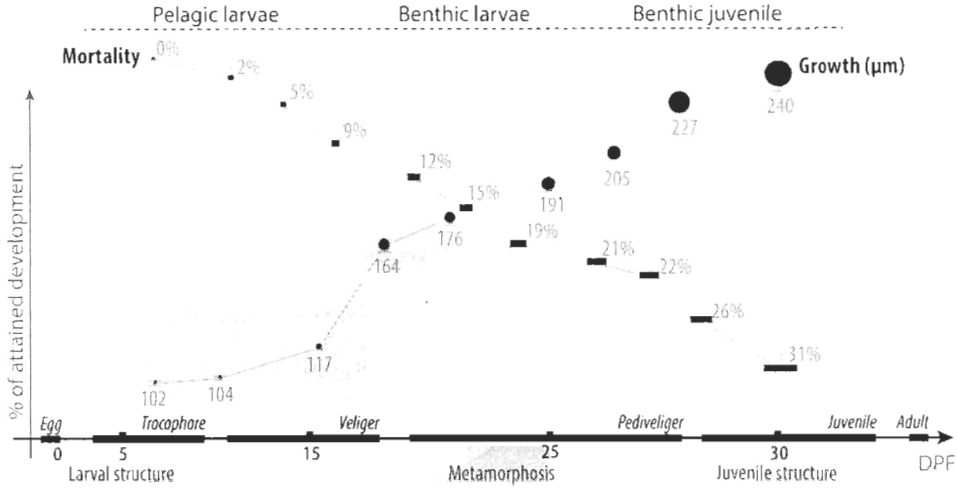

Figure 3.1: **Conceptual and operational frameworks of larval attained development.** Density plots represent major transitions in tissue remodelling and organogenesis expressed through percentage rates of attained (A) larval and (B) postlarval development. The hypothesized increase in attained development (%) sums up the singular peaks (dark vs. light grey) that elicit the end of one phase in physiological change and the beginning of another. Around 15 DPF larval structures were all developed, it was thus marked by a decrease in the attained development. Post-metamorphic pediveligers (>25 DPF) show an end of the development of larval structures and the beginning of the exponential increase in juvenile/adult structure. Mortality and growth were experimental measurements obtained during *Mytilus edulis* rearing. Survival rates were expressed as the total number of individuals less than the cumulative amount of empty shells and based on the first sampling timepoint at 1 DPF (N=3 tanks). Growth was the measure (in µm, N=50) of shell growth. These results were related to the larval body plan during early development. Adult stage was included for reference purposes.

Zhang et al. [2012a] found that shell formation was dependent on genes encoding laminins, fibronectins, and collagens associated also with the extracellular matrix (ECM), which highlights the importance of overlapping mechanisms in bivalves. Furthermore, a relationship between genomes and larval fitness can be visualized through delayed metamorphosis [Pechenik, 1990b, 2006]. For example, loss of the velum and formation of the adult shell can be spontaneously induced by activation of specific genes after a special treatment with coralline microalgae, which is known to induce larval metamorphic competence [Williams et al., 2009]. Behavioural flexibility can henceforth be seen at the beginning of larval formation [Bayne, 2004]. This functional morphology was strongly controlled by

conserved genetic effectors and was specially associated with food availability and space relocation [Pechenik, 1990a; Gosling, 1992; Abrams et al., 1996].

Development plasticity of bivalve larvae involves quick transitions ranging from feeding and nutrient allocation, to metamorphosis and growth, to defence and survival given cumulative tightly coordinated regulations [Roff, 1992]. For example, neuronal interactions mediate sensory mechanisms early in development notably when larvae begin to capture particles and assess their surrounding for proper settling cues [Croll et al., 1997; Voronezhskaya et al., 2008]. Larval neurons are lost after metamorphosis, and then replaced with an entirely different set of neural wiring [Marois and Carew, 1997]. Moreover, muscle cells provide a contractile apparatus responsible for ciliary beating required for the locomotion of free-swimming larvae and a functional foot at the pediveliger stage, all of which are essential to explore the settlement substrate and produce byssus filaments [Ackerman et al., 1994; Dyachuk and Odintsova, 2009]. Accordingly, neural innervation of the velar muscle contributes to the whole larval behaviour [Braubach et al., 2006]. Thus the arrest of the pelagic phase seems to be accompanied by a synchronous loss of the velar muscles and their innervations [Penniman et al., 2013], then development of functional respiratory [Carroll and Catapane, 2007] and papillary [Cannuel et al., 2009] gills. As can be seen, muscle tissue in larvae is surprisingly more complex than that found in adults, whereas it gets simplified after metamorphosis [Odintsova et al., 2010].

Defining boundaries of molecular mechanisms is sometimes problematic [?]. For example, the immune system, activates immunocells as a defence mechanism against threatening pathogens, but they might also release angiogenic factors that attract proliferating endothelial cells, which migrate and form new blood vessels [Revathy et al., 2013]. In addition, angiogenesis is the formation of new blood vessels led by collagen deposition in the ECM, epithelialization, and vascularization of endothelial cells and finally unneeded cells are destroyed [Montagnani et al., 2007; Chappell et al., 2011]. Consequently, a similar mechanism could be expected in embryogenic and early larval development or even activated through

wound healing. For this reason, apoptosis is another tightly regulated mechanism triggered during a defence state and histolysis, or purposely targeting unneeded or less-specialized cells [Kiss, 2010]. Furthermore, the structure of the ECM constitutes a dynamic scaffold-based circuitry with integrative components that promote cellular proliferation, migration, and tissue motion [Green et al., 2013]. As a result, muscle [Dyachuk, 2013] and neural [Odintsova et al., 2010] cell differentiation depend on functional peptides and growth factors found in the ECM, which constitute basic requirements for muscle assembly, neural guidance, and synaptogenesis. In addition, endocrine chemical messengers target special cell receptors and trigger genetic neurohormonal control, therefore mediating an interactive effect on ciliary [Carroll and Catapane, 2007] and cardiac [Kodirov, 2011] functions, signalization [Fabbri and Capuzzo, 2010], and immune defence [Ramos MartiNez et al., 2012].

Processes activated throughout development were mainly sought through microscopic anatomy and immunochemistry. Such works provide precious anatomical and physiological evidence that ought to be related to genomes. Together with the ability to perform genome-wide associations of the larval transcriptome, we can now identify genetic connections required in shaping the early life-history of the blue mussel *M. edulis*, provided that we capture key interlocking mechanisms [?]. In this study we combine functional information of physiological processes with the genomic profiles corresponding to each developmental stage, from fertilization to egg formation until juvenile. For this reason, the RNA-sequenced (RNA-seq) transcriptome of *M. edulis* was de novo assembled, then annotated using the latest heterogeneous metadata available. Furthermore, genetic behaviour of larvae was assessed chronologically with custom-build microarrays. Consequently, expression trends of candidate genes at the fertilized egg, larvae, and postlarvae were analyzed using a multi-processing approach. As a result, interactions between synexpressed group of genes were detected. Finally, a mechanistic assessment of the organism cannot rest on one gene, thus following the dynamic patterns of physiological and phenotypic plasticity and interactions of synergistic genes, we were able to determine critical keypoints of unexplored genetic

regulatory network (GRNs) and less-known functions activated early on in larval development. Here we show a different approach elucidating the mechanistic relationship between genes, cellular responses, and phenotypic changes through functional morphology. Consequently, we have explored the evolutionary characteristics of several developmental traits of *M. edulis* larvae on the basis of gene regulators, further explaining the dynamic response of genomes to the environment.

3.5 MATERIALS AND METHODS

3.5.1 *Biological sampling*

3.5.1.1 *Larval culture*

Spawning adults were obtained from the Havres-aux-Maisons lagoon (Magdalen Islands; 47°25'N, 61°51'W) in Quebec, Canada, which harbours pure population of *M. edulis* [Tremblay et al., 1998]. Transfer permits of Fisheries and Oceans Canada have been obtained for breeders and the studies did not involve endangered or protected species. Species identification was assessed after genomic analysis of the foot DNA with Glu-5' (polyphenolic adhesion protein) as molecular marker (not shown) [Rawson et al., 1996], and only *M. edulis* species was present in sampled specimens. A series of three independent cultures constituting biological replicates were each stocked with fertilized eggs produced from gametes derived from 20 males and 20 females at a ratio of 10 sperms/oocyte and were carried out at the Aquicole Station of Pointe-aux-Peres for the University of Quebec at Rimouski (UQAR), Quebec. Larvae were reared at 18°C with a final concentration of 20 larvae ml^{-1} in 1 µm (pore-size) filtered 150 L seawater in an up-welling flow-through system of 0.1 $L.min^{-1}$ [Andersen et al., 2011]. Pediveliger were transferred using downweller units (125 µm mesh size) and all larvae were fed continuously at 30,000 cells ml^{-1} [Andersen et al., 2011] with a cocktail of three different algae species (*Isochrysis galbana, Pavlova lutherii, Nannochloropsis oculata*). Flow of seawater maintained a constant crawling movement of larvae. Mussel

trocophore, veliger, and pediveliger larvae and juvenile postlarvae were collected from cultures at each development stage beginning at 1 day post-fertilization (DPF). At each sampling time N=50 larvae were measured. Replication consisted of three independent larval tanks (N=3). Shell growth rates were measured with Image-Pro v5.1 an image analysis software by Media Cybernetics and an Olympus BX41 microscope. Morphological criteria (e.g., eyespot and foot formation) helped determine the level of competence for metamorphic larvae [Pechenik, 1980; Beaumont and Budd, 1982]. Survival rates were expressed as the total number of individuals less than the cumulative amount of empty shells and based on the first sampling timepoint at 1 DPF (N=3 tanks). Rates of growth and survivorship were reported as means±SEM (standard error of the mean) and differences were considered significant at $P<0.05$ (Fig. 3.1).

3.5.1.2 *Collection, extraction, and RNA quality control*

Samples containing known number of larvae and postlarvae (150,000 and 20,000) were collected at 0, 1, 18, 28, and 42 DPF, flash-frozen in liquid nitrogen, and stored at −80°C until usage for RNA extraction. Mussel eggs ($\approx 50 \times 10^3$), larvae ($\approx 25 \times 10^3$), and postlarvae ($\approx 10 \times 10^3$) were homogenized and lysed in 500 µL TRIzol reagent (Invitrogen) solution (per $2.5–5 \times 10^6$ cells) using the Precellys homogenizers (Bertin Technologies) designed for fast throughput lysis. Nucleic acid separation and RNA extraction were performed with 200 µL chloroform (Sigma) without alcohol compound stabilizers, further minimizing DNA contamination. Precipitation of RNA was carried out overnight in one volume of 3 M ammonium acetate (Sigma) and 2.5 volume of ice-cold 100% ethanol at −20°C. RNAs were resuspended in an appropriate volume of nuclease free water necessary to achieve the required concentrations for subsequent analysis after a 25 minute centrifugation at 4°C, and washing the pelleted RNAs with 500 µL of ice-cold 80% ethanol. Quality and quantity assessments of the RNA were performed using the NanoDrop ND-8000 UV spectrophotometer designed for high throughput sample evaluation and RNA integrity checks were performed with the

Bioanalyzer and the RNA6000 Nano assay kit (Agilent Technologies) designed for minimal sample consumption.

3.5.2 *De novo transcriptome assembly of M. edulis*

3.5.2.1 *Illumina HiSeq sequencing*

Two libraries of transcriptional read sets of *M. edulis* larvae and postlarvae were sequenced using a paired-end Illumina HiSeq2000 system (Genome Quebec Innovation Center and McGill University, Montreal, Canada). To maximize coverage of transcripts that belong to early ontogenesis of the mussel transcriptome one RNA sample from each developmental interval was pooled into a single sample. Since the transcriptome was spanning multiple correlated timepoints with high temporal resolution and no gene expression was required, it was wasteful to introduce biological replicates [Stewart-Oaten et al., 1986]. Moreover, technical replicates from the same RNA library were required to evaluate biological variability of the multiple development intervals being sequenced. The pooled sample was prepared for RNA-seq sequencing according to the manufacturers protocol. Extensive sequencing was required since the purpose of the experiments was discovering novel transcribed elements and strong quantification of known transcript isoforms. Furthermore, the sensitivity of detection was important, for this reason a minimum depth of 100–200 million pair-end reads of length 2×76 base pair (bp) was produced. No RNA fragmentation was carried out to increase coverage along the entire gene body. No normalization preventing the alteration of the sample preparation was employed. As a result, detection was improved (cf. Table 3.1 for details of the assembled library specs and read sets produced by the different sequencing runs). The initial short read dataset from this study was deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession number SRP043335.

Table 3.1: **RNA-seq paired-end library sequencing and gene annotation detailed counts**

Assembly and annotation summary	Number of sequences
HiSeq total number of bases	55,539,936,200
HiSeq total number of reads	277,699,681
HiSeq total number of sequences	50,383
BLAST low-hits	34,926
BLAST top-hits	15,257
Significant mapping	12,094
GO annotated (cutoff$\leqslant 10^{-6}$)	3,815
Biological process (BP) nodes	6,448
Molecular function (MF) nodes	1,746
Cellular component (CC) nodes	897
InterPro annotations	2,294
Enzyme code mapping	4,703
% annotated	30

3.5.2.2 *De novo assembly of RNA-seq reads*

An in-house Perl script was used on the Forward and Reverse reads to remove low-quality sequences from raw data sequences. The high-quality reads were then assembled using Velvet (v1.1.07; Kmer=51) and Oases (v0.1.11) to construct unique consensus sequences [Zerbino and Birney, 2008; Schulz et al., 2012]. Finally, the high-throughput sequencing generated over 270 million cDNAs (Fig. 3.2B). Additional information related to data assembly can be found in Table 3.1.

3.5.2.3 *Functional annotation*

High-throughput sequence annotation and biological function association were carried out using the functional algorithms implemented in the professional version (v2.7.0) of Blast2GO (B2G) application [Conesa et al., 2005].

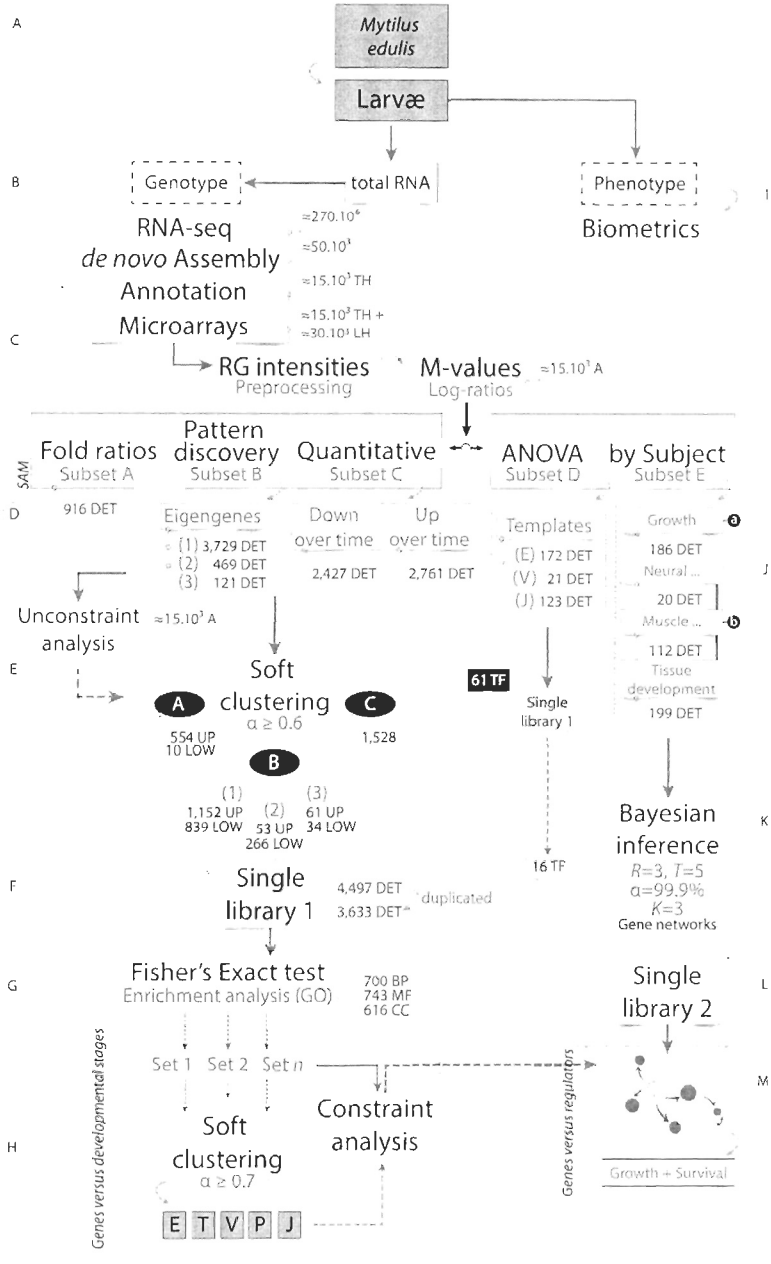

Figure 3.2: **Analytical pipeline for effective coercion of genetic regulatory networks.**

Figure 3.2: (*Continued*) (A) Pooled samples of larvae were (B) RNA-sequenced generating around 270 million reads. Transcriptome annotation identified 15,257 top-hits (TH) and ~30,000 low-hits (LH) transcripts, both (C) printed onto 15 microarrays. (D) Significance analysis of microarrays (SAM) enabled extraction of subsets A, B, and C. (J) Analysis of variance (ANOVA) resulted in subset D. Biologically significant genes (E-value$\leqslant 10^{-6}$) were assigned to subset E. Related expression patterns were clustered into "up" and "low" over time gene sets at (E) ($\alpha \geqslant 0.6$) and (H) ($\alpha \geqslant 0.7$) thresholds. (F) Library 1 contained the highest significant co-regulated genes (FDR 2% and $P \leqslant 0.001$). (F) Library 1 contained the highest significant co-regulated genes (FDR 2% and $P \leqslant 0.001$). (F) Library 1 contained the highest significant co-regulated genes (FDR 2% and $P \leqslant 0.001$). (G) Fisher's exact test ($P < 0.05$) helped visualize the main ontologies of significant genes that cover the main three entries for the biological process (BPs), molecular functions (MFs), and cellular components (CCs). (K-L) Subset D was used as a training set of the 61 regulators (TF), then (M) Bayesian regulatory networks were inferred between selected TFs and subset E.

Biological dependencies were assessed for the number of annotated sequences (Table 3.1) to which, Gene Ontology (GO) terms were obtained by assignment to unknown query sequences based on similarity conducted against the nr (non-redundant) database. Two rounds of annotation were performed under an *E*-value cutoff threshold of 10^{-6}. The minimum needed coverage between a significant alignment and high-scoring segment pairs (HSPs) would have to cover at least 60% of the longitude of that high scoring sequence (Fig. 3.3D) because *M. edulis* was under genotyped.

Moreover, it was clear (Fig. 3.3C) that a significant part of the mapping was based on model species, some not being invertebrates. In addition, low-complexity regions were assigned for removal to assist in the matching of blast results, since they can cause artefacts in the similarity search. In contrast, queries with low number of high-complexity regions had low scoring hits and thus get filtered out. Collectively, this pipeline proved to be a suitable approach for analyzing poorly characterized non-model organisms with up-to-date heterogeneous datasets. Furthermore, B2G was used as an exploratory tool to monitor the range and map gene identifiers to GO terms of the novel transcriptome mussel dataset. Because we were working with a de novo assembly of a non-genotyped species, no microarray specific annotation packages where available from Bioconductor, otherwise BioMart could have been used. Therefore, a custom annotation list (gene-to-GO) was defined then loaded after specific mapping of characterized gene IDs. Identified functional proteins relative to the mapped gene IDs were retrieved from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. InterPro scans and enzyme codes were hence generated and data-mining of indexed sequences was performed with MG-RAST (Metagenomics RAST) using its metagenomic analysis server (metagenomics.anl.gov) for overall quality control (Table A.1) as well as to characterize and remove prokaryotic sequences from the preprocessing analysis [Meyer et al., 2008].

3.5.3 *Gene expression kinetics*

3.5.3.1 *Microarray chip construction*

Mussel microarray chips were designed for each of the three biological replicates at these five timepoints: 0, 1, 12, 28, and 42 DPF.

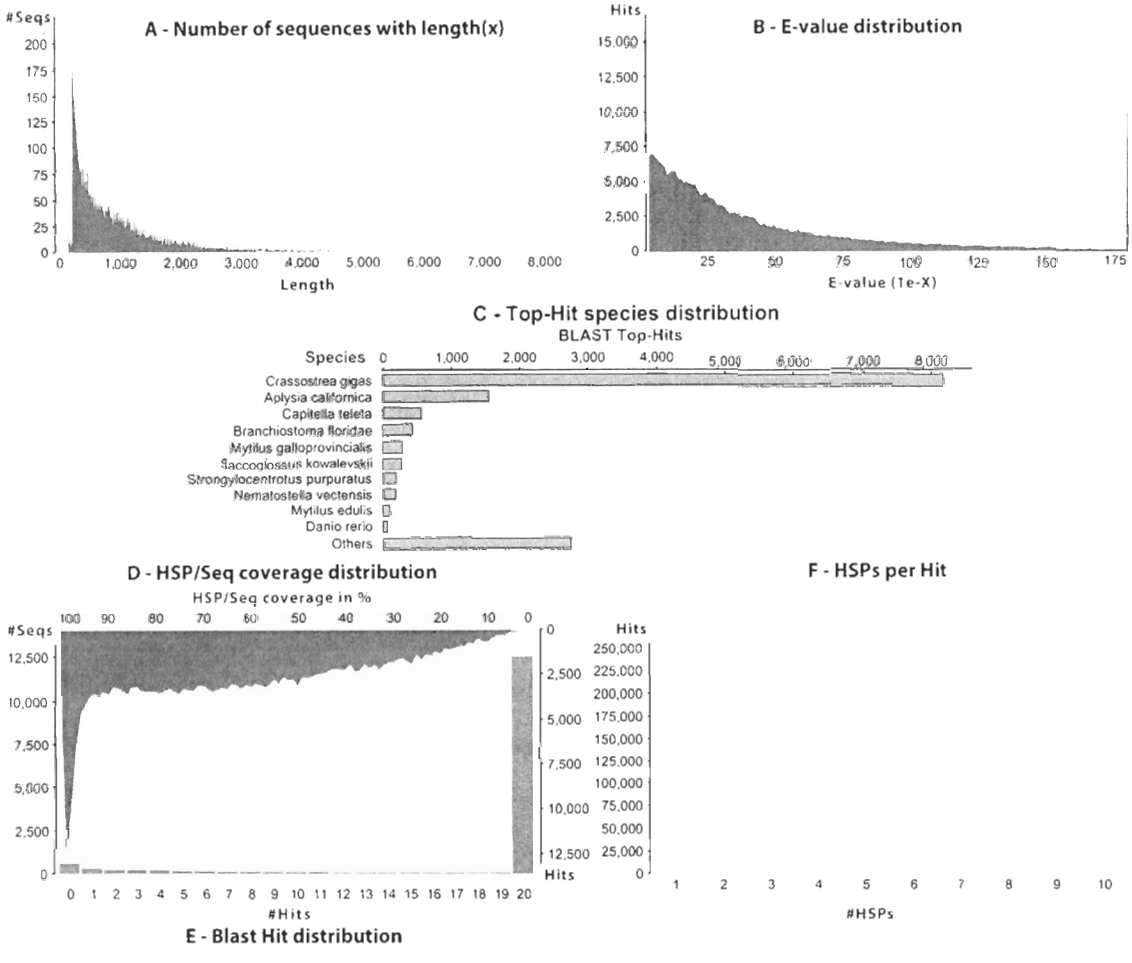

Figure 3.3: **Overview of the quality control data of 15,257 top-hit transcripts annotated and GO-mapped after RNA-sequencing and transcriptome assembly.** (A) The range of the length of sequences, (B) the *E*-value of the best hits during annotation, (C) the top 10 species contributing to the annotation, (D) high-scoring segment pairs (HSPs) covering top-hit sequences, (E) maximum number of hits retrieved per sequence, (F) maximum number of HSPs per hit.

The experimental design generated raw data consisting of 15 measurements. Gene expression was monitored during early development stages from egg to postlarvae by hybridizing total RNA to custom-designed 8x60K Agilent microarrays. There were 15,257 annotated mussel transcripts each covered twice per array with the exception of a few sequences that belong to (i) positive housekeeping genes each covered ten times, (ii) negative *Arabidopsis thaliana* genes each covered ten times, and (iii) 26,038 non-annotated sequences covered once per array. A ladder of commercial RNA spike-ins was included to calibrate quantification, sensitivity, coverage, linearity, and enabled weighting in spots throughout the preprocessing of the raw data. Gene boundaries were determined by analyzing the open reading frames (ORFs) of sequences, thus improving predictive analysis through orientation of coding regions. This process was performed with a web-tool ORF-Predictor (proteomics.ysu.edu/tools/OrfPredictor.html) using default parameters. The whole design resulted in a custom gene expression glass slide, formatted with a total of around 60,000 custom designed probes spotted on eight high-definition 60K arrays (Agilent SurePrint G3 format). High-quality oligonucletides were *in situ* synthesized and printed using Agilent's 60-mer SurePrint inkjet technology. Complementary RNA (cRNA) was synthesized with the Illustra CyScribe GFX purification kit (GE Healthcare) according to the manufacturers instructions. Probes were labelled with an Agilent's Low Input Quick Amp labelling kit (Agilent) for two-channel microarray-based gene expression analysis, following the manufacturer's instructions. Quantity assessment of cRNA was done using NanoDrop ND-8000 UV spectrophotometer. The common reference design was dye-labelled with cyanine 3 and experimental samples were dye-labelled with cyanine 5. Isolated RNAs from all developmental stages were pooled to prepare the reference sample. Quality control was performed and images were quantified in Agilent's Feature Extraction software (Agilent, version 10.7.3.1).

3.5.3.2 *Preprocessing of microarray data*

Preprocessing of microarray data (Fig. 3.2C) was performed with the available or adapted functions from the `limma` package from Bioconductor [Smyth, 2004]. A minimum background correction was applied to the five-group series. Probes known in advance to be non-differentially expressed (control probes and titration probes) were given more weight to assist in the normalization process. Those known to produce a differential expression were down-weighted. Normalization within arrays was implemented to uncover underlying differential gene expression [Smyth and Speed, 2003], which employed the predefined contrast of spot quality weights. The time series dataset was subjected to log conversion so that it was similar in terms of central and dispersion tendencies. The different arrays were not expected to have similar M-value ranges, thus between-array scale normalization was not considered. Linear models were computed on the basis of a design matrix without the use of a contrast-matrix and results were compared to all the subsets of the filtered DETs. All the coefficients in the design matrix provided estimates of the log-ratios between every stage of development [Smyth, 2004]. An empirical Bayes method was used to moderate the standard errors of the log-fold changes. An estimate of array weights was assessed in order to control the array quality during preprocessing [Ritchie et al., 2006b]. More details about the microarray experiment and R scripts are available from the corresponding author upon request. Raw microarray data and the preprocessed expression matrices are available at the NCBI Gene Expression Omnibus (GEO) accession number GSE55580.

3.5.3.3 *Differential expression and functional analysis*

Differential gene expressions between all developmental stages were compared to select a set of candidate transcripts, based on significant statistical imputations. The list of DETs (ranked first according to their adjusted q-values <0.03) was compared in a one class timecourse [Storey, 2002]. Significance analysis of microarray (SAM) data was used for com-

putation of gene correlations with response to all timepoints. The samr package in R was used on the array data [Tibshirani et al., 2011] and no further normalization was performed. First, slopes (least squares slope of gene expression versus time) were carried out to find significant DETs with consistent increase or decrease over time (subset A, Fig. 3.2D). Second, pattern discovery was performed with multiple requested eigengenes. As a result, false discovery rate (FDR) values were considered to be an effective indicator for choosing the best requested eigengene, thus estimation of the powerful significant genes (subset B). Finally, additional timecourse quantitative SAM grouping was done with preselected patterns. We chose to search for specific patterns, taking into account both positive and negative trends in gene behaviour. Therefore, targeted genes were either specific to early, middle or advanced stages of ontogenesis and were grouped in subset C. Moreover, genes whose expression varied significantly between the larval and post-larval stages were characterized by a one-way ANOVA with an overall *P*-value threshold $\leqslant 0.0007$. Consequently, template matching (similar to a *t*-test) was applied to shape the samples and search for group of genes particularly expressed during early or late larval development [Pavlidis, 2003]. In addition, FDRs (5.7%) were used to follow the trends and set the *P*-value threshold [Benjamini and Hochberg, 1995]. As a result the measured gene expression in this model was assumed to be only the result of the effect of time on development (subset D).

3.5.3.4 *Construction of library 1*

Library 1 contained significant DETs co-expressed at every developmental stage from egg to juvenile, chosen from subsets A, B, and C after one round of cluster analysis. The first round of clustering retrieved co-expressed DETs from the initial collection of 15,257 transcripts, which formed library 1 (Fig. 3.2E). The second round grouped only the highly synergistic and enriched genes from library 1 (Fig. 3.2F). Consequently, iterations were parameterized to find DETs with cluster memberships higher than 0.6 or 0.7 for the first or second round respectively. Implementing these filters was essential for the selection of high-

membership genes to the identified clusters as well as for increasing the sensitivity of the inferential algorithms used for constructing the GRNs. Briefly, filtering out low-membership genes increased the identification sensitivity of connected genes, which was due to the removal of noise. However potential loss of information ensued. Although, we can not provide an absolute qualification of the gene networks, we can only confirm that the results presented here are the best we could infer using numerous iterative settings. Moreover, since we are working with high-throughput parametric models, we did not only change the parameters and thus the sensitivity of the filters, but we implemented also different sets of genes to improve the precision of the predictions. Finally, it should be noted that not all transcripts from library 1 were GO-mapped (Fig. 3.2G). Public sequence libraries used in annotation and mapping did not seem to index more than 40% of raw data. Although those libraries contain high number of sequences, mostly redundant, which was one of the causes behind the diversity of the sequence-discovery approach used here [Zhang et al., 2012a]. For that reason, FDR cutoffs throughout the analysis were set as low as 2% to reduce noise through elimination of potential low-membership genes, while maintaining a high number of selected genes (higher than 2,000 genes) for further processing.

SUBSET A: ONE CLASS TIME COURSE DATA. Subset A contained genes with a positive or negative expression over the timecourse of larval development. Rankings were computed over 100 requested permutations. FDR of genes were controlled with a value of Δ=0.58 yielding 10 down-regulated and 554 up-regulated significant genes with an FDR of 2.1% on the average. First, this data was generated as a single group timecourse design where three replicates per timepoint were processed, giving upper and lower score cutpoints of ± 2. The SAM score (d_i) was the T-statistic value of the ith gene that represents the power of the relationship between genes during the early development stages. Second, the highest ranked value of d_i=3.8 was called statistically significant for positively regulated genes (up-regulated) correlated with development, and a value of d_i=-3.5 was found for negatively regulated genes

(down-regulated). Finally, the least significant q-value for both ranked genes was q=0.02 (Fig. 3.2D).

SUBSET B: UNKNOWN PATTERN DISCOVERY. Subset B contained significant DETs extracted using a pattern discovery scheme over 100 permutations. Predominant patterns have not been selected a priori. Consequently, significant eigengene (a representative gene module of the expression profile) delivered relatively highly correlated DETs. First, only 3,729 significant genes with an FDR of 2.1%, giving upper and lower score cutpoints (Δ=1.18) of $\pm$2 were selected respectively (Fig. 3.2D). Higher ranked SAM scores of selected genes had a d_i=6.6 for up-regulated DETs, and a d_i=-8.3 for down-regulated DETs. The q-value was highly significant for first ranked genes at <0.001 and went up to 0.026 for lower ranked genes of both expression sets. The second and third eigengenes delivered at an FDR of 2% a total number of 469 and 121 significant genes, for upper and lower score cutpoints (Δ=1.14 and 1.12) of $\pm$3 respectively. The d_i were 6.3 and 4.99 for up-regulated DETs, and -7.7 and -5.9 for down-regulated DETs at a $q \leqslant 0.022$. Finally, subsequent eigengenes delivered high FDRs therefore we stopped at the third component. The first eigengene was characterized by a module of genes significantly expressed at early and late stages of the development of mussel larvae (can be further visualized in Fig. 3.4). The second eigengene was more specific to the stages where larvae were around metamorphosis. Subsets A and B contained genes with a consistent increase or decrease over time.

SUBSET C: QUANTITATIVE. Analysis of subset C focused only on unique expression trends over the time of development. Since developmental transformation was closely regulated, young mussels undergo special physiological changes [Widdows, 1991]. For this reason, molecular interactions regulated the loss of larval organs and the appearance of juvenile/adult structures (Fig. 3.1A,B). A crucial transition from a primitive to a more sophisticated organism, capable of faster defence and a higher-level of fitness [Pechenik, 1990a].

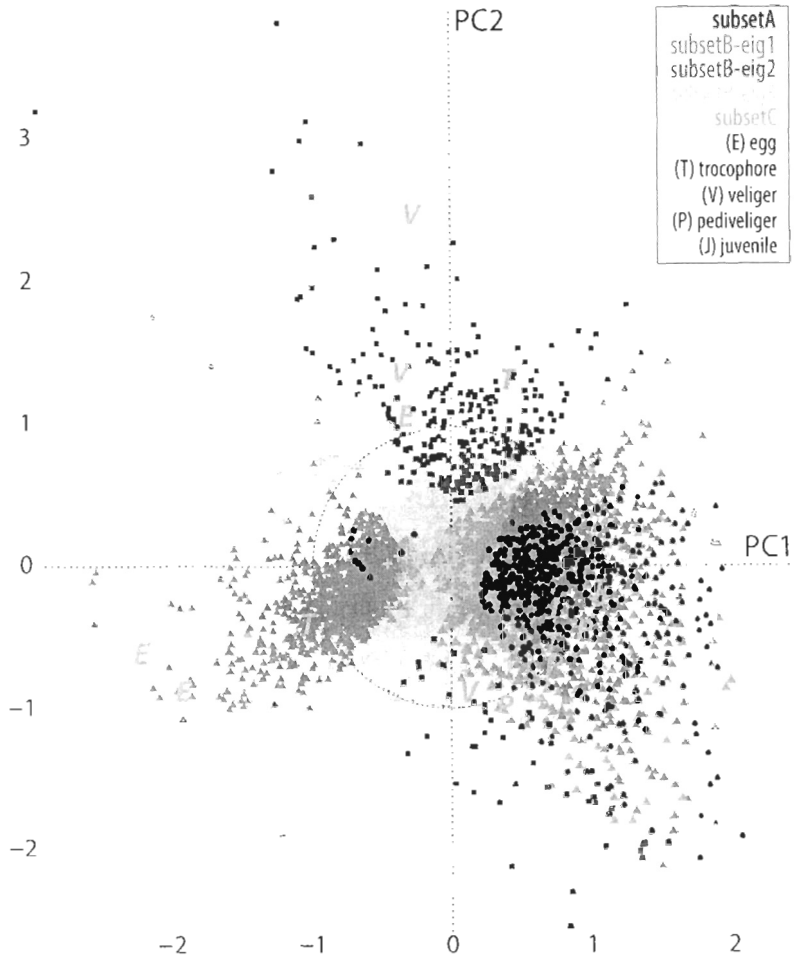

Figure 3.4: **Inter-variable covariance structure of the expression data of *Mytilus edulis*.** Data includes 15,257 transcripts in principal components and developmental stages in standard component.

Accordingly, it was at this point where selected genes were up-regulated either pre- or post-metamorphosis [Balsciro et al., 2013]. Therefore, subset C contained genes that matched five pre-selected patterns grouped into two setups (not shown). We searched for genes that can be described with singular expression peaks, either during the egg and trocophore stages (setup 1), or during the pediveliger and early juvenile stages (setup 2). Multiple quantitative search inquiries based on specific pattern discovery templates resulted

in 2,427 DETs for setup 1 and 2,761 DETs for setup 2 with an average FDR of 7% and 2.5% respectively.

3.5.3.5 *Construction of library 2*

Library 2 grouped a list of the different networks (GRNs) inferred from subset E in addition to the 16 regulators (Fig. 3.2K) chosen from library 1. Among the 61 regulators extracted from library 1, only 16 were retained for further dynamic network inference. Subset D (Fig. 3.5, and Table A.2) was used to select these 16 most interactive hubs (GRNs not shown). These hubs belong to regulators with at least 50 outward connections (Fig. 3.6, and Table A.3). Subset D did not contain any expression structure since ANOVA tests only delivered significant DETs without indicating at which stage they were significant. Thus it was used as a training set to predict potential interactive regulators among TFs of library 1 (Fig. 3.2K).

SUBSET D: TRAINING SET FOR THE REGULATORS. Subset D contained highly significant genes ($P \leqslant 0.0007$) across all five developmental stages, selected based on a one-way ANOVA. Next, we used pattern matching templates that identified DETs in pre-metamorphosis versus post-metamorphosis stages [Pavlidis, 2003]. A Benjamini-Hochberg FDR threshold were defined at 5.7%. At this FDR, we identified 188 highly significant DETs ($P \leqslant 0.0007$). Among these transcripts 172 exhibited a significant change in expression during the egg phase, 21 at the veliger, and 123 at the juvenile stages. The reason we did not include subset D in library 1 was because clustered DETs of library 1 were chosen according to an expression pattern contrasting the progression of the development stages, whereas subset D contained solely selected DETs according to their statistical significance in all the samples.

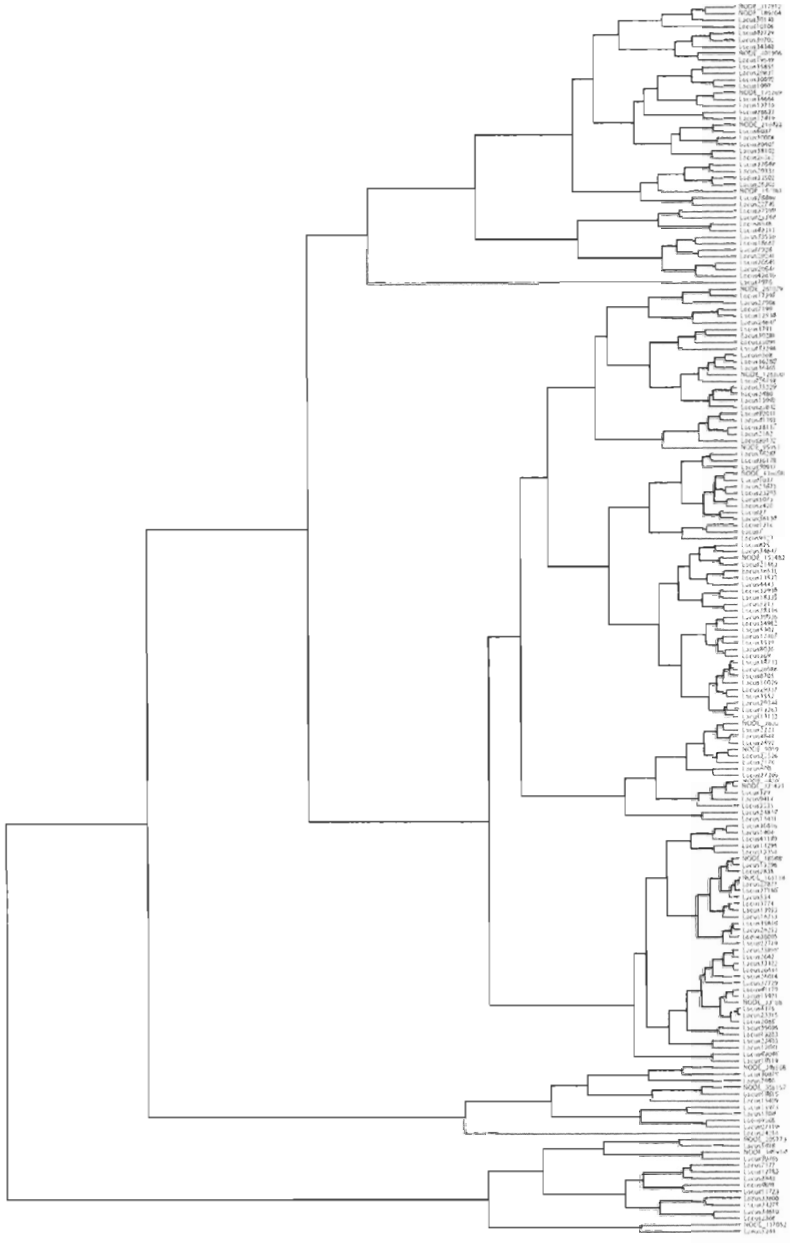

Figure 3.5: **Hierarchical clustering of subset D containing expressed genes during early ontogenesis.** Red bars represent significance at $P \leqslant 0.05$ assigned after 2000 bootstrapping estimations. The original full heatmap view associated with the tree can be found in the manuscript. The following is an augmentation to accurately visualize gene associations. Each Locus or NODE is assigned to an annotated transcript, described in Table A.2.

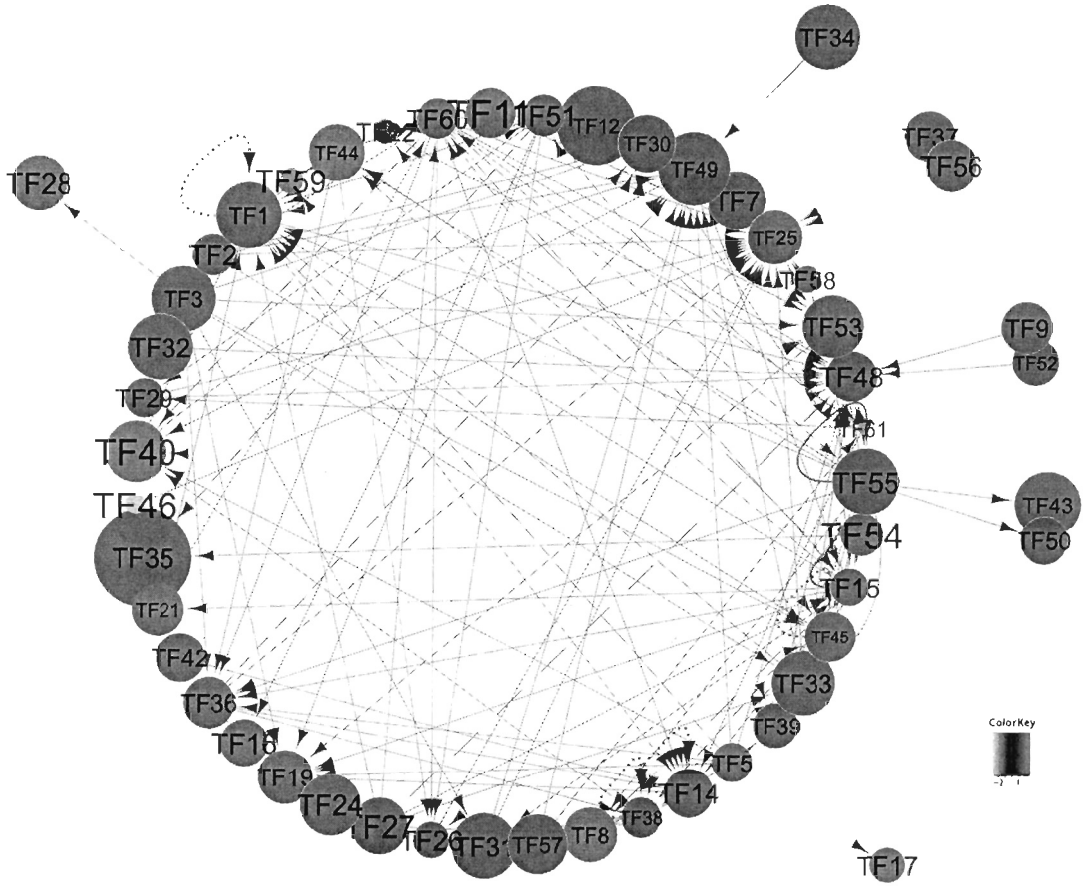

Figure 3.6: **Differential genetic interaction between 61 significant transcription regulators expressed at the egg, pediveliger, and juvenile stages of development.** Nodes are identified by regulator (TF) numbered index. IDs, descriptions, and UniProt entries are found in Table A.3. Regulators are selected from Library 1 based on their function (transcription factor, kinase, phosphatase, or chromatin remodeler) and their expression (a minimum of two fold change). Redundancy may occur when assigning TF to categories. Node color, label size, and node width represent the genes log expression ratios during the egg phase, pediveliger and juvenile stages respectively. Edge line type corresponds to either a negative (dotted blue line) or positive (solid red line) interaction between two genes computed by Bayesian network fitting model. Edge width corresponds to the betweenness coefficient of gene-to-gene interaction. The stronger an established interaction is, the bigger the outward connecting edges are.

SUBSET E: LIBRARY 2. Subset E was designed without an a priori knowledge of the significant biologic connection between selected transcripts. First, we chose transcripts that have a relationship to BP involved in growth (118 DETs), development (699 DETs), locomo-

tion (137 DETs), and rhythmic process (23 DETs) and MFs in neural (20 DETs), muscle (112 DETs), and tissue development (199 DETs). Second, assuming relevance for growth, development, locomotion, and rhythmic processes, these DETs were pooled together then were identified among library 1 candidate genes. Consequently, set a (Fig. 3.2J,a) incorporated 55 significantly up-regulated DETs extracted from library 1 (Table A.5), the same way the 61 regulators were selected (mentioned above and listed in Table A.3). Additionally, set b (Fig. 3.2J,b) was based on genes related to neural, muscular, and tissue development mechanisms. They were pooled together and reduced to incorporate only up-regulated 155 DETs (Table A.4). As a result, subset E (set a and b) was constructed based on enriched BPs related to growth, which consisted on interrelated MFs associated to neural and muscle development essential for the swimming and feeding larvae. Finally, the pre-selection follows pre-defined sets of functional DETs that were selected based on their significantly enriched GO-terms.

3.5.4 *Data mining and big data analysis*

3.5.4.1 *Fuzzy clustering*

Fuzzy clustering defined a membership value to the clustered transcript (Fig. 3.7). The membership was then partitioned among clusters by assigning weights that sum up for each transcript to one unit. Membership values were not an indication of the DET effectiveness or statistical significance. It only constituted a fidelity score of the gene expression from one time to the next, which represented the overall structure of the cluster. Soft partitional clustering enabled the pre-grouped subsets A, B, and C to be split without the definition of cluster hierarchy. Gene expression data was converted into *ExpressionSet* objects with `Biobase` from Bioconductor [Gentleman et al., 2004]. Replicated arrays (three per stage for five timepoints) were averaged to their means prior to clustering by the `mfuzz` function in R [Kumar and Futschik, 2007].

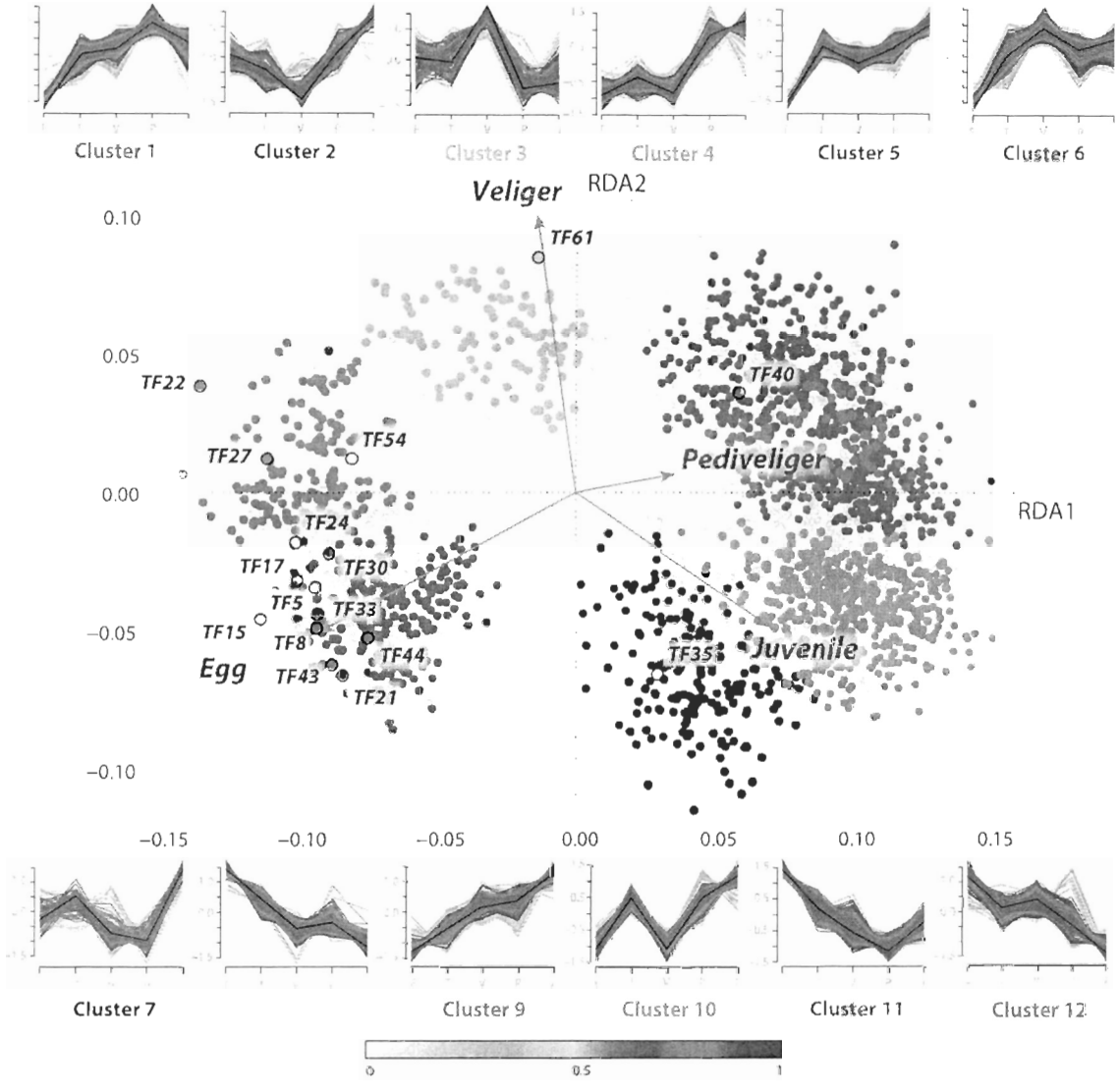

Figure 3.7: **Library 1 projecting the relationship between 3,633 clustered DETs at early development (5 timepoints, N=3).** Stages were fitted as explanatory variables. Grey empty circles are low membership DETs. Numbers can be traced back to Figures 3.8B and 3.8C, each was a regulator pointed out by a black outliner, coloured at the centre according to cluster membership, and identified as regulators in the inferred GRNs of Table A.3 and Fig. 3.6. The 12 temporal expression patterns represent library 1 clustered genes (clusters 1 through 12). Green, blue, and red/purple lines indicate low, medium, and high membership DETs respectively. The solid black line is the core of the cluster, varying at ±1.5. FCM analysis was performed on 3,633 DETs of library 1 at c=12 and m=1.5.

The algorithm implemented a fuzzy version of the *K*-means clustering protocol, providing furthermore functions based on *C*-means methodologies and the fuzzifier *m* that was chosen accordingly to prevent random clustering; an option lacking in *K*-means clustering [Schwämmle and Jensen, 2010]. Data standardization was performed prior to clustering on highly significant genes to emphasize the relative change in expression. Genes based on their strong membership scores contributed to the structure of one cluster. Its stability was ranked by setting the FCM (fuzzy *c-means*) clustering parameter *m* to 1.5 [Kumar and Futschik, 2007]. Finally, all clusters were stable at higher *m* values, which indicated powerful clustering.

Likewise, non-random genes maintained their core which validated the strength of the cluster generated in the Euclidean space. How closely a gene expression followed the pattern of a cluster, its core, was the underlying feature of membership computations and expression kinetics for structuring of genes [Futschik and Carlisle, 2005]. Filtering off non-significant genes (first round of clustering) led to an increase in sensitivity for co-regulated genes at a higher α-threshold (α-cores $\geqslant 0.7$), which means a noise-reduced cluster structure. Subsets from library 1 were clustered (second round) in search for co-expressed DETs with strong membership values. The FCM clustering parameter (m=1.5) revealed 9 to 20 clusters for all subsets. Increasing *m* downweighed the influence of low-membership genes (α-cores were set to sort and filter high-membership genes). Cluster stability was assessed while tweaking the *m* parameter resulting in the selection of compact clusters. The closer genes were to the core, the higher were the membership values. Stable clusters were isolated and compact without losing their structure. Their overall structure was examined by coupling similar clusters so that shared genes get identified, resulting in the inferential of only 12 clusters in Fig. 3.7.

3.5.4.2 *Principal component analysis*

Since each DET was assigned to a cluster with a membership value extending between [0,1], only those with an α-threshold $\geqslant$0.6 (first round of clustering), the closest to the cluster centroid (α-core), were selected. These DETs share the most stable relationship within a cluster. They were projected onto a principal components analysis (PCA) covariance biplot to get a visual of the internal cluster structure. Consequently, the principal coordinates where assigned to gene expression to visualize the covariance structure of the variables. It was done by a singular value decomposition of the 15,257 gene expression in an $n \times m$ matrix after centring and normalization. Let n be the number of replicate per stage and m the number of genes. As a result, the biplot represents how the different developmental stages spread out with respect to one another. The closer a developmental sample was to the unit circle of the biplot, the better it was being displayed since the standardized cases have lengths equal to 1 or less. Finally, differences between developmental stages were the average squared differences between all transcripts that had the weighted sum of squares equal to the squared singular values [Greenacre and Blasius, 2006].

3.5.4.3 *Hierarchical clustering*

Hierarchical clustering was only employed on subset D after data mining had been carried out on significant DETs grouped in library 1 and verified by linear model fitting. Subset D contained 188 significant DETs ($P \leqslant 0.0007$). First, logarithmic ratios were centred and scaled before complete clustering of genes by Pearson correlation and biological samples (3 replicates of 5 timepoints, N=15) by Spearman correlation. Second, significance (cutoff $\leqslant$0.05 threshold) of clusters were retrieved after 2,000 bootstrap repetitions. Complete bootstrap analysis was carried out using `pvclust` package in R. Finally, the hierarchical tree was cut at specific heights, showing colours corresponding to clusters, over the heatmap colour bar (Fig. 3.8A and Fig. 3.5).

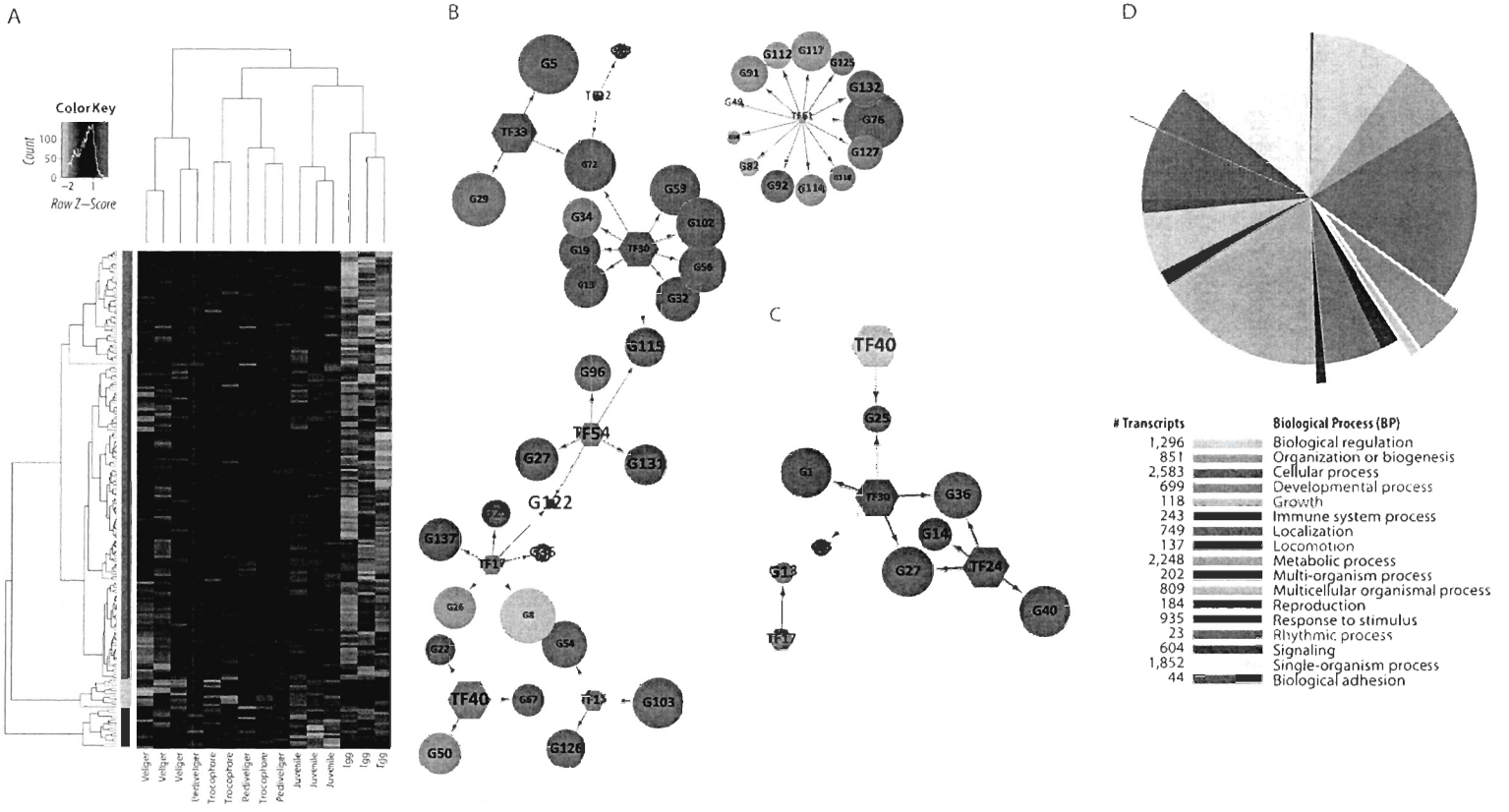

Figure 3.8: **Hierarchical clustering of subset D and network inference of subset E containing expressed genes during early ontogenesis.** Nodes are identified by gene (G) or regulator (TF) numbered indices. Gene IDs, description, and UniProt entries are found in Tables A.3 for regulators, and A.4 and A.5 for networks B and C respectively. Node colour, label size, and node width represent each of the genes log expression ratios during the egg phase, pediveliger, and juvenile stages respectively. Colour key is shared among subfigures A, B, and C. Bigger label size and node width indicate gene over-expression. Highlighted hubs (yellow) are discussed in the text. Blue dashed and red solid edges denote a negative and positive coupling (transcriptional regulation, protein-protein interaction, or protein modification) between two nodes respectively, estimated (99.9% cutoff) by an empirical Bayes procedure of the posterior mean and variance of the differentially expressed genes (N=3 larval tanks over 5 timepoints). Biological processes of the 15,257 annotated larvae transcripts are represented in subfigure D.

3.5.4.4 *Enrichment analysis*

An enrichment analysis of Fisher's exact test presented a thorough digest of GO-term integration of clustered sets with significant genes instantiated in library 1. Only non-random or significant modules (P<0.05) were presented. Tests were performed using the `topGO` package in R [Alexa and Rahnenfuhrer, 2010]. Enrichment tests presented here were based on gene counts, but Kolmogorov-Smirnov (KS) tests based on gene scores and methods accounting for the GO topology were also employed in the process of comparing the results (not shown). GO categories were mapped to protein IDs using the functional information found in public databases provided by NCBI, PIR (Protein Information Resource), and GO that were used to retrieve GO-terms stored in GO repositories. An Evidence Code (not shown) has been used to monitor the quality of this functional assignment and irrelevant mappings were henceforward removed. No BioMart annotated genes specific to Mollusca, Bivalvia or *M. edulis* could have been adopted for term-gene linkages. Likewise, no microarray specific annotation packages were available from Bioconductor for this species. In this regard, our custom arrays made use of specific mapping between genes and GO terms, with most up-to-date annotations, implemented with a GO hierarchical structure (topology – not shown) covering thus around 60% of the annotated sequences (15,257 transcript) with specific BP, MF, and CC ontologies (Table 3.1 and Fig. 3.2G).

3.5.4.5 *Dynamic Bayesian networks*

Network structure of co-regulated genes was inferred using an empirical Bayes dynamic Bayesian network (EBDBN) method and was applied during the early five stages of larval development, on the basis of three biological replicates. Pre-selected genes of subset E were inferred after quantile normalization [Bolstad et al., 2003] (Fig. 3.8B,C). Interactive GRNs were built using all data of the analysis described in this paper combined with overall functional assessment of genes predicted from GO mapping and literature data. First,

the estimation procedure was based on a linear dynamical system and was accomplished with the ebdbNet package in R [Rau et al., 2010]. Consequently, the model selection relied on a block-Hankel matrix of autocovariances of the timeseries genes expression data and on estimated hidden states. Additionally, the time lag between a gene and its regulator was chosen to be m=1 and a cutoff of 99.9% for the Z-scores of the edges (positive and negative regulation) of the gene-gene interactions. As a result, the Z-scores represented an empirical Bayes estimation of the inferred posterior mean and variance of the network parameters [Rau et al., 2010]. Second, at m=1, synexpressed genes were directly connected to their specific regulators without additional chaperones. Moreover, since we were assessing timeseries observations, we considered that a genetic regulator can dynamically interacts with different proteins at overlapping timepoints in a non-linear way. Third, hidden variables can be estimated using a Kalman filter/smoother and were set at a dimension K=3 at a 90% cutoff, then iterated according to the hyperparameters values and posterior distributions relative to gene-gene interactions in each of the GRNs being discovered. These hidden variables can elicit a priori biological information that cannot be encoded into the experiment, for instance, effects of siRNA, levels of protein turn-over, or dilution effect of cell growth [Beal et al., 2005; Alon, 2007]. Thus, when we followed the reaction kinetics from one timepoint to another, we assumed that these hidden factors substituted non-measured observations. Collectively, these parameters yielded graphical models known as dynamic Bayesian networks (DBNs). Finally, the posterior distribution of the network described the direct gene-to-gene interactions representing each edge. Positive and negative edges were interpreted as activations and inhibitions respectively.

Network visualization and manipulation was established in Cytoscape (v3.0.2) [Shannon et al., 2003]. All molecular interaction data were further loaded for advanced analysis. Data were comprised of gene expression, gene-to-gene (for feedback GRNs) or gene-to-TF (for input GRNs) interactions, clustering, and data related to each gene (e.g., GO and UniProt). Confidence value of each interaction was calculated using the network analyzer

implemented in Cytoscape. A wide variety of significant coefficients (e.g., neighbourhood connectivity, edge betweenness, interactions, shortest path, and partner node pairs) were attributed to every network, which aided in the visualization of the GRNs. Moreover, feedback GRNs were assigned as undirected networks and inferred from library 2 alone without fitting TFs to the calculation of interactions. In contrast, input GRNs were those with TF rules. Feedback GRNs were inferred at a 99.9% confidence level for the Z-scores of the posterior distribution and convergence criterion of Δ_1=0.15, Δ_2=0.01, and Δ_3=0.001. Input GRNs were inferred at a 99.9% confidence level for the Z-scores of the posterior distribution and convergence criterion of Δ_1=0.15, Δ_2=0.05, and Δ_3=0.001.

3.5.5 *Constrained ordination analysis*

Redundancy analysis (RDA) was employed to assess the possible relationship between DETs and the physiological changes during each of the developmental stages. Vectors were set by the square root of the correlation coefficient with coordinates equal to their linear regression coefficients. The fitted vectors points to the direction of the developmental stages and vectors strength was measured by their length. First, significance for this constraint ordination with five variables was as follows: for the egg phase (P=0.03, R^2=0.80), trocophore (P=0.828, R^2=0.04), veliger (P=0.007, R^2=0.46), pediveliger (P=0.133, R^2=0.31), and juvenile (P=0.001, R^2=0.89). Second, predictors were then reduced from five to four, in a stepwise approach. Consequently, an Akaike's information criterion (AIC) was also implemented to generate, using 2,000 random permutations, penalized goodness-of-fit values for every automatic selected and grouped constraints. The `vegan` [Dixon, 2003] and `MASS` [Venables and Ripley, 2002] packages in R were used for all ordination functions and for fitting the community data. Finally, data matrices ($n \times m$ matrix, where n=15 – three replicates per timepoint for five timepoints – and m=3,633 DETs) were originally M-values (R/G intensities, normalized and corrected; R (red) was the amount of gene-specific RNA in the experimental

sample correlated to G (green) amount of RNA from the reference sample). Response composition data (columns) were therefore already dimensionless (no units) and the explanatory variables were of the same unit but both were scaled into a [0,1] interval. This type of standardization provided a wider spread of transcripts in the two dimensional space and reduced the effect of outliers.

3.5.6 *Alignment and sequence mapping*

In order to verify the functional domains contained in the sequenced transcripts, we selected amongst others the following reviewed and genotyped species to represent the reference models: *Aplysia spp., Argopecten spp., Bos taurus, Branchiostoma floridae, Crassostrea gigas, Danio rerio, Drosophilia melanogaster, Gallus gallus, Haliotis spp., Homo sapiens, Mus musculus, M. edulis, and Xenopus spp.* Since genomic metadata and genetic material were still fragmented and hard to find for *M. edulis*, we invested in an alternative approach for mapping the identified genes to other model organisms, which helped narrow down on the implicated biological traits when associations were found. This gene-candidate approach was solely based on sequence similarities depending on conserved protein domains. Functional regions found in key genes or group of genes were associated with functional morphology traits (e.g., growth, locomotion, and rhythmic processes). This helped classify genes into conserved families [Bateman et al., 2004] and confirmed interactions between genes and/or their regulators. Functional mapped sequences were further checked against InterPro databases for family classification and provided prediction of response sites in regulated genes. The `blastx` algorithm was also used for domain mapping and for search of signature regions between two interactive genes. Furthermore, several significant DETs were not characterized through mapping (mentioned as NAs) or they were assigned a redundant annotation that did not identify the sequence functional affiliation. Although those sequences could be considered as assembly artefacts and thus eliminated from the analysis, data min-

ing and `blast` heterologous metadata screening of only the highly interactive ones helped identify them as orthologous genes (not shown). *M. edulis* GC (%) content of codon usage was a good indicator to actively narrow down the possibilities of a sequence being part of *Mytilus* spp. Standard genetic code of *M. edulis* was characterized by 45.23% of coding GC at the final codon usage; 51.22% for the first letter, 51.61% for the second, and 32.87% for the third (NCBI-Genbank data source). The mRNA sequences were translated in every reading frame, which efficiently helped identify best-hit homologous queries in the next step [Plot, 2000]. Next, `tblastn` was used to search for related protein against the NCBI non-redundant (nr/nt) nucleotide database [Altschul et al., 1997]. This approach was used in order to reduce loss in sensitivity to distant sequence relationships. The `blastx` algorithm could have also been used solely to acquire functional annotation of NA nucleotide sequences, however `blastx` was predefined for faster computations which might jeopardize the quality of the emerging hits. Indeed, some of the results showed high-sequence homologies to queries belonging to close-related species or to new non-categorized sequences in close-related datasets (i.e., sequences being identified and assembled using other transcriptome techniques).

3.6 RESULTS AND DISCUSSION

We have characterized the connection between transcriptional regulators and target DETs throughout the early developmental stages of the blue mussel larvae *M. edulis*. The major setback we first encountered was the lack of sequence metadata specific to *M. edulis*. High-throughput sequencing paired with microarray technologies allowed us to identify previously unknown effectors for multiple important biological processes (Fig. 3.8D). These players were essential to understand how larval cell responses could be triggered by exogenous biotic and abiotic cues. Our large-scale multi-processing approach grouped co-regulated genes and helped predict their associations with transcription factors, phosphatases, kinases, and chromatin remodelers (Fig. 3.9).

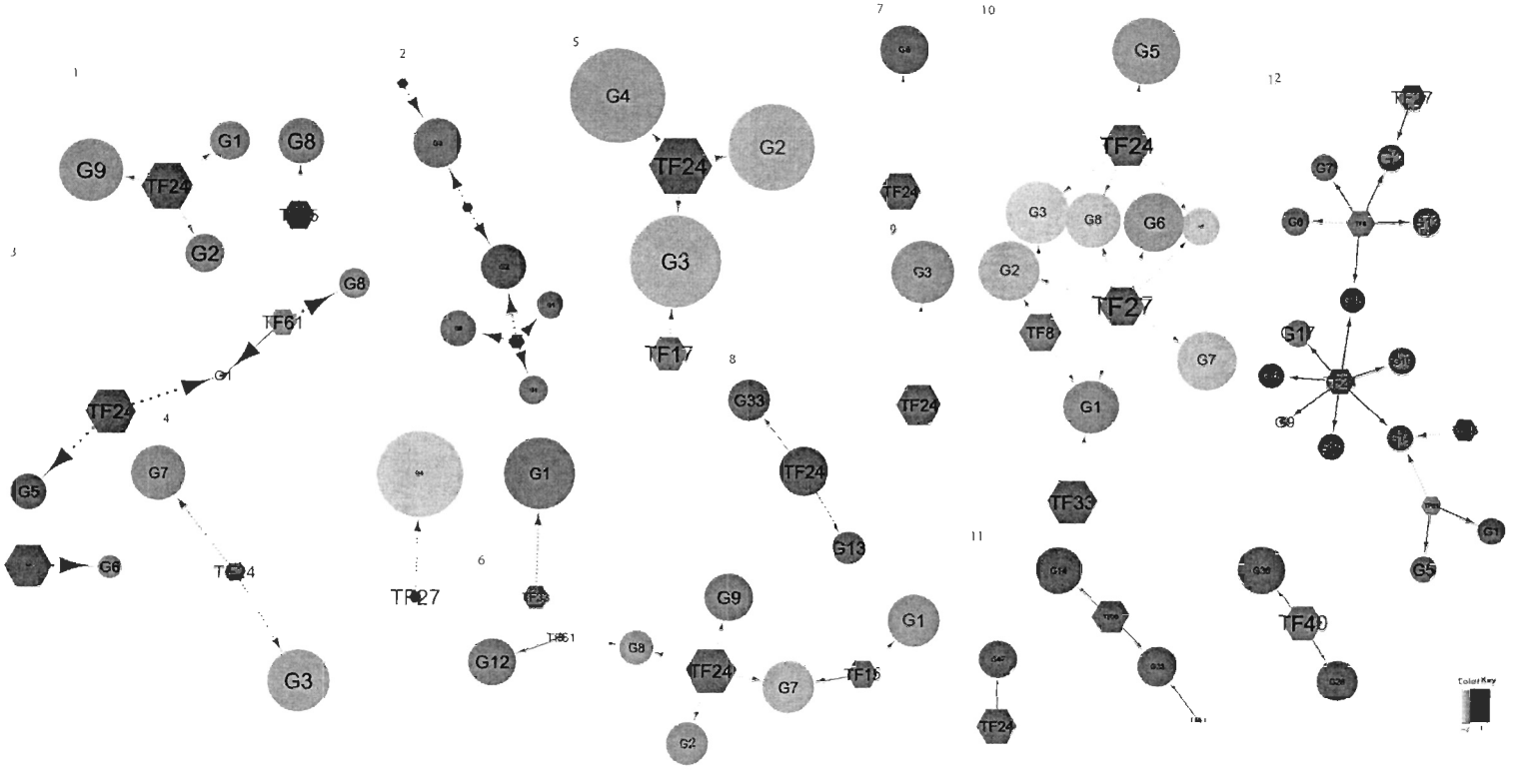

Figure 3.9: **Differential genetic interaction between 16 significant transcription regulators expressed at the egg, pediveliger, and juvenile stages of development and the 12 clusters of Library 1.**

Figure 3.9: (*Continued*) Nodes are identified by gene (G) or regulator (TF) numbered index. IDs, descriptions, and UniProt entries are found in Table A.3 for regulators and in Table A.6 for clustered genes. Regulators (hexagones) are selected from Library 1 based on their function (transcription factor, kinase, phosphatase, or chromatin remodeler) and their expression (a minimum of two fold change). Redundancy may occur when assigning TF to categories. Clusters are extracted based on expression templates, either highly expressed at the egg or juvenile stages (described in Figure 3.2 cf. manuscript). The number figuring in the upper left corner of the dotted box, that group a small set of genes or hubs, represent the cluster number (from 1 to 12). Timeseries expression of each cluster and constrained interaction between clusters can be visualized in Figure 3.7 (cf. manuscript). Node color, label size, and node width represent the genes log expression ratios during the egg phase, pediveliger and juvenile stages respectively. It should be noted that the size of one same node may change from one cluster to another. This would permit better visualization of relative expression between nodes of the same cluster. Nevertheless, each one node has the same expression across all clusters. Edge line type corresponds to either a negative (dotted blue line) or positive (solid red line) interaction between two genes computed by Bayesian network fitting model.

The synergistic use of these methods provided an efficient analytical pipeline for the identification of novel and uncharacterized developmental gene modules.

3.6.1 *Subsetting genes to reduce inferential variation*

Throughout the manuscript, genes that were selected from the output of specialized algorithms were grouped into five subsets (Fig. 3.2D,J). Each of the A, B, and C subsets contained a module of highly regulated genes expressed over the time of the experiments (42 DPF). The difference between these subsets resided in the expression patterns of the selected eigengenes. After the first round of clustering of these subsets, we were able to identify highly coordinated expressions (Fig. 3.2E). As a result, library 1 contained all the closely characterized synexpressed DETs (Fig. 3.2F). Furthermore, all the regulation factors identified through library 1 were coerced by significant DETs from subset D into selecting the highly connected regulators. Finally, GRNs were inferred on the basis of these regulators and the developmental genes from subset E (Fig. 3.2K). Consequently, library 2 contained regulation factors connected through activation or inhibition of DETs implicated in the functional morphology of *M. edulis* larvae (Fig. 3.2L).

3.6.2 *Regulatory genes represent the pre- and post-metamorphic transformation*

In order to understand the molecular functioning of larvae which was central to understand their physiological growth and adaptation, we had to distinguish which genes play central roles, when, and to which extent (Fig. 3.1). For this reason we proceeded by registering several transcriptional profiles of larvae throughout their development to expose the underlying sources of variation in gene expression. Next, timeseries transcription variables were reduced in a PCA to distinguish between decisive expression patterns justifying the prominent directions of the variance-covariance structure of the high-dimensional data (Fig.

3.4). The covariance PCA biplot exhibited a between-sample dispersion and a tight ordering of gene modules, all of which were clustered according to particular temporal patterns of expression. As a result, genes displayed an overall strong homogeneity in expression coordination during each stage. In that case, the variance accounted for 43.5% of the first two principal components (PC1 and PC2) with 15,257 transcripts. The observed variation was lower than expected due to noise in the non-structured data [de Haan et al., 2007]. Furthermore, the number of measured genes was substantial (Table 3.1) and only a fraction was expected to be essential (coloured dots in Fig. 3.4 and Table A.1). For this reason, we reduced the number of genes and the amount of noise to increase the represented variance (Fig. 3.2D). Significance analysis of microarray (SAM) was performed on preprocessed microarray data. As a result, three subsets A, B, and C (Fig. 3.4) were extracted using this technique and combined together into library 1, which included the most significant up- and down-regulated genes from one stage to the next (Fig. 3.2F). Moreover, clustering of library 1 introduced a second dimension to exploring the property and homogeneity of co-regulated eigengenes (Fig. 3.2E). Consequently, after noise-reduction and structuring, the final centroid biplot in Fig. 3.4 represented only by the three significant subsets explained 92.5% of the variance. Accordingly, the five stages were arranged on the PC1 axis in a developmental order that reflected the association of PC1 to the differential expression related to larval maturation. Finally, a proportion of the prior (V) and posterior (P and J) metamorphic DETs (SubsetB-eig2, Fig. 3.4) were inversely correlated to the increase of the PC2 score, which indicated that the metamorphic transition was mainly reflected in PC2.

3.6.3 *Synexpressed genes are susceptible to an interlocking regulation*

Regulatory systems can depend on interlocking components (e.g., transcription factors, enzymes, small molecules) implicated in the functioning of one or several GRNs [Alon, 2007]. For this reason, a second round of clustering was aimed to group tightly co-regulated

eigengenes with similar time-dependent expression patterns, which evidently demonstrated a synergistic regulation of cellular response (following also the guilt-by-association method [Chu et al., 1998; Brazhnik et al., 2002]). As a result, the interaction between clusters further revealed interlocking mechanisms of regulation between unique eigengenes and positive or negative system regulators. Fig. 3.7 illustrated the different cluster structures for library 1. For example, clusters 3, 4, 7, and 10 exhibited a high expression periodicity, which demonstrated how much each cluster tolerated noise and standout being co-regulated under a specific physiological condition (Fig. 3.7). Noise tolerance found in microarray data can be further seen in cluster 10, where a small group, at least 10 transcripts, standout with a non-influential distinct pattern than the overall order of the general cluster. Furthermore, the α-core of the isolated cluster (black single line) was not affected by the low-membership transcripts.

3.6.3.1 *Differential expression reveals larval developmental changes*

Next step consisted on evaluating how the diverse clusters correlated with the developmental changes through dimension reduction and constrained projection. Consequently, analysis was assessed under 999 random permutations for the whole 3,633 transcripts and the resulting constrained triplot explained 59% of the variance of the transcript abundance of library 1 (P=0.002, *adj*-R^2=0.43) (Fig. 3.7). Transcripts with low membership scores were displayed in grey. Transcripts with a potential role in transcription regulation were marked by a borderline black circle (Table A.3). On the whole, two distinct sets of clusters exhibited a strong bond (Fig. 3.7) due to interlocking relationships or to recurring patterns shared between adjacent clusters of the same set. Accordingly, clusters 3, 8, 11 and 12 make the first set and the remaining clusters make the second set (clusters 1, 2, 4, 5, 6, 7, 9, and 10).

3.6.3.2 *Interactive clusters involve eigengenes tied to complementary processes*

Transcript distribution in library 1 covered mainly MFs ($P<0.001$) related to cell adhesion (GO:0050839), hydrolase (GO:0016787) and hormone activity (GO:0005179), and myosin (GO:0017022) and calmodulin binding (GO:0005516). The most enriched BPs ($P<0.001$) were distributed between maintenance-terms related to circulatory system (GO:0003013), actin movement (GO:0030048), heart (GO:0060047) and striated muscle contraction (GO:0006941), regulation of the cell cycle (GO:0010564), neuromuscular junction development (GO:0007528) and catabolic processes associated to organic acid (GO:0016054), fatty acid (GO:0009062), and lipid modification (GO:0030258).

Different cluster schemes were tested then dimensionally reduced in a constrained ordination (Fig. 3.7). We noticed that whatever the arrangement of the variables was, three trends were always depicted: i- transcripts followed a bi-dimensional pattern that was either bias towards pre- (left) or post- (right) metamorphosis, ii- clusters 2 and 7 were usually grouped around the juvenile stage, and iii- the trocophore were consistently placed opposite to the juvenile's vector. Trending MFs for cluster 2 were mostly relevant to the response for inorganic substances (GO:0010035), hormone stimulus (GO:0009725), and phototransduction (GO:0007603), whereas cluster 7 related to organic substance metabolic processes (GO:0071704) and fatty acid oxidation (GO:0019395). Repeated pseudo-patterns were observed in some highly-stable periodic clusters, e.g., clusters 3, 4, and 7, where gene expression changed after the veliger stage and at peri-metamorphosis. The relationship with cluster 7 further incorporated enriched MFs of cluster 3 related to several defence processes (GO:0006952), regulation of cytokine production (GO:0001817), and muscle cell development (GO:0055001) and of cluster 4 related to renal (GO:0001822), liver (GO:0001889), and vasculature development (GO:0001944), and response to oxidative stress (GO:0006979).

Moreover, cluster 4 contained synexpressed DETs only enhanced at post-metamorphosis, and reflected a coordinated relation with cluster 10 which was enriched with aminoglycan

catabolic process (GO:0006026), ribosome biogenesis (GO:0042254), ncRNA (GO:0034660) and rRNA (GO:0016072) metabolic processes, and translation (GO:0006412). Furthermore, cluster 3 consisted of up-regulated transcripts at the veliger stage. Although cluster 6 and 12 were closer to cluster 3 in gene behaviour, they showed a non-exclusive up-regulation at the juvenile stage and egg phase respectively (Fig. 3.7). Cluster 6 was enriched with regulation of hormone secretion (GO:0046883) and transport (GO:0009914), and neurotransmitter transport (GO:0006836), whereas cluster 12 was enriched with Golgi vesicle transport (GO:0048193), regulation of synapse structure and activation (GO:0050803), and regulation of skeletal muscle development (GO:0048742). Clusters 3, 6, and 12 must have adhered to different regulation mechanisms at the veliger stage because they were distant.

3.6.4 *Expression coordination reveals timeseries genetic regulatory networks*

Biochemical interactions are the only means by which cells promote their growth, maintenance, and tissue development, the reason why dynamic regulatory transcriptional networks had to be reverse engineered, thus providing a thorough insights into cellular response [Murphy et al., 1999]. For this reason, interrelated DETs were further assessed using an empirical Bayes estimation method [Rau et al., 2010] by inferring the structure of GRNs during the five stages of early larval development of *M. edulis*. Regulation factors that assumed a function in transcription regulation, phosphatase and kinase activities, and chromatin remodelling were chosen for this work. Subset D (Table A.2 and Fig. 3.8A) was used for training of the 61 predicted regulators (Table A.3) found in library 1, among which 16 highly interactive hubs where selected since they occupied a crucial position in the network (Fig. 3.7-TFs). These hubs were used to assess GRN topologies of the 12 clustered DETs (Fig. 3.7) and subset E (Fig. 3.2J), which included predicted DETs related to tissue development (Table A.4) and growth (Table A.5), visualized in Figures 3.8B and 3.8C respectively.

3.6.4.1 *Coordinated regulation reveals clustering of active developmental processes*

In a network of direct connections between the 16 predicted regulators and subset D, three transcriptional interaction motifs occupied much of the connections (Fig. 3.10), including *PIP5K1A* (TF17), *GATAD1* (TF24), and *ATRX* (TF27) (Table A.3). Sequence alignment confirmed the presence of many conserved domains (not shown), but can be explored following the UniProt links in Table A.3. *PIP5K1A* is a kinase (GO:0016307) member of cluster 8, *GATAD1* is a transcription regulator (GO:0003700), and *ATRX* is a phosphatase (GO:0017111), both members of cluster 12 (Fig. 3.7). First, *PIP5K1A* was negatively connected to the phosphatase *CnVAS1* (TF5, GO:0017111). However *GATAD1* and *ATRX* were both strongly connected to 5 and 12 other TFs respectively (Fig. 3.6). Second, according to Fig. 3.9 only *GATAD1* showed significant positive connections with two DETs in cluster 8, and 7 DETs in cluster 12. Moreover, cluster 8 was enriched (P<0.001) in GO-terms related to GTPase regulator activity (GO:0030695), nucleoside-triphosphatase regulator activity (GO:0060589), and ion binding (GO:0043167). Finally, *ATRX* positively interacted with one DET in cluster 12 and among all the 12 clusters (Fig. 3.9), *PIP5K1A* was bounded by a singular negative connection in cluster 5 (Table A.6), which was enriched (P<0.02) in GO-terms related to steroid dehydrogenase (GO:0016229), intramolecular oxidoreductase (GO:0016860), and endopeptidase (GO:0004175) activities.

3.6.4.2 *Same interlocking regulators engage in different developmental processes*

Likewise, when it came to inferring genes implicated in neural and muscle processes or tissue development, TF17, TF30, and TF61 appeared to be the most connected (Fig. 3.8B). *PIP5K1A* (TF17) was covered above, TF30 (GO:0033613, cluster 11) and TF61 (GO:0008134, cluster 3) represented *DERL2* and *ACTL* respectively. The latter contained a conserved domain that belonged to the Actin subfamily of ACTIN/mreB/sugarkinase/Hsp70 superfamily and both included GO-terms related to regulation of transcription (Table A.3). Furthermore,

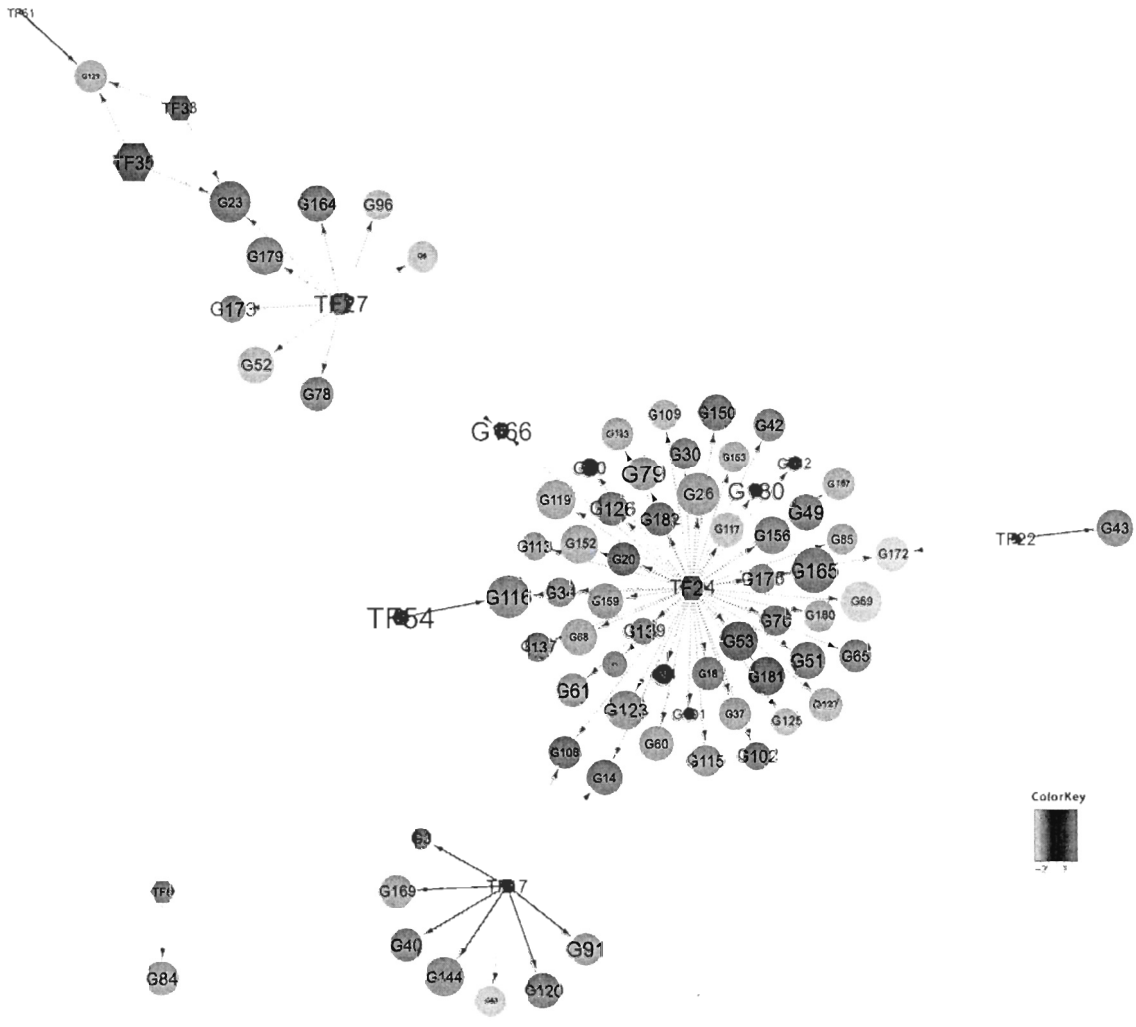

Figure 3.10: **Differential genetic interaction between 16 significant transcription regulators from Library 1 and 188 enriched genes of subset D expressed at the egg, pediveliger, and juvenile stages of development.** Nodes are identified by gene (G) or regulator (TF) numbered index. IDs, descriptions, and UniProt entries are found in Table A.2 for genes and Table A.3 for regulators. Regulators (hexagones) belong to either category of transcription factors, kinases, phosphatases, or chromatin remodelers as described by their GO terms. Redundancy may occur when assigning TF to categories. Node color, label size, and node width represent the genes log expression ratios during the egg phase, pediveliger and juvenile stages respectively. Edge line type corresponds to either a negative (dotted blue line) or positive (solid red line) interaction between two genes computed by Bayesian network fitting model.

PIP5K1A interacted with 6 DETs, among those, one was implicated in neural development (G36), two in muscle process (G8, G21), and three in tissue development (G26, G122, G137) (Table A.4). Similarly, *DERL2* was implicated in 9 interactions, two of them were related to muscle process (G13, G72) and the remaining seven belonged to tissue development. In addition, among the 13 DETs connected to *ACTL*, one DET possessed a function in neural (G49), five DETs in muscular (G38, G76, G117, G118, G125), and the remaining seven were related to tissue development processes (Table A.4).

Furthermore, *GATAD1* (TF24) and *DERL2* (TF30) were strongly connected to DETs whose functions included growth, development, locomotion, and rhythmic processes (Fig. 3.8C). First, *DERL2* encodes an endoplasmic reticulum component targeting unfolded proteins (all references are supplemented through the UniProt links in Table A.4). According to Fig. 3.8B, it interacted positively with *ARF4* (G19) involved in vesicle formation and transport, *ROPN1L* (G32) a kinase regulator and *B9D1* (G102) a member of the tectonic-like complex, both involved in cilium motility and assembly respectively. Additionally, *DERL2* was connected to *YWHAZ* (G34) involved in various signalling pathways and is implicated in suppression of apoptotic processes. With this in mind, apoptosis could also be suppressed by the product of the coding gene *PACRG* (G56) positively associated with *DERL2*. Second, *CALM2* (G13) and *CDC42* (G72), regulators of cytokinesis, were positively connected to *DERL2* (TF30). Consequently, the former plays a role in immunity, synaptogenesis, and cardiac muscle contraction as do *CDC42*, which may be additionally involved in Wnt signalling pathways like *CSNK2B* (G59), cell polarity, regulation of neural apoptotic processes, and angiogenesis (Table A.4). According to Fig. 3.6, *DERL2* might also be involved in coordination of chromatin remodelling, essential in neural post-mitotic mechanisms, which included *SMARCD1* (TF3) and actins. Moreover, *DERL2* could play a role as a scaffold of *Med6* (TF55) a coactivator of transcription bridging gene-specific regulatory proteins to the basal transcriptional machinery. Finally, *ANAPC7* (TF49), a cell-cycle ligase involved in protein degradation, was negatively

connected to *DERL2* which in turn was also negatively associated to *CBX3* (TF60) involved in chromatin remodelling for transcription repression (Table A.3).

DERL2 exhibited an inward negative connection from *GATAD1* inferred in Fig. 3.6. According to Fig. 3.8C *GATAD1* was positively associated with *TNKS2* (G14) involved in activation of the Wnt signalling pathway that is recognized for its function in embryonic development and cell polarity and to *Rho1* (G27) involved in mechanistic transitions of the embryonic body plan through cytoskeletal organization (Table A.5). Consequently, we showed here that *GATAD1*, a component of a chromatin complex specific to target methylated histone sites [Vermeulen et al., 2010], was associated to the core component *HISTH2A* (G36) of the nucleosome. Furthermore, *UBN1* (G40) was also positively associated to *GATAD1* and can encode a senescence regulator by forming a complex that includes histone H3.3. Finally, Fig. 3.6 showed negative connections between *GATAD1* and *ABCF3* (TF25) an ATPase involved in viral cellular response, *ARF6* (TF15) a GTPase involved in apoptosis, cell adhesion, cell polarity, and cytoskeletal organization, *DNMT1* (TF45) which encodes a methyltransferase involved in histone methylation essential for epigenetic inheritance, and *PTPN13* (TF19) a phosphatase that negatively regulates apoptotic mechanisms (Table A.3).

3.6.4.3 *Development is dependent on conserved networks through chromatin remodelers and transcription regulators*

Looking deeper into molecular functions related to tissue development and growth, we have characterized coordinated connections between regulators and several candidate genes. One such example was *GATAD1* (TF24, Figures 3.8C and 3.6), a member of the 16 regulators selected from library 1, encodes a protein that interacts with transcription repressors [Vermeulen et al., 2010]. These repressors were involved in histone modifications through deacetylation and demethylation, which enabled the regulation of transcription, chromatin condensation, and DNA repair [Kouzarides, 2007]. Mechanisms of chromatin remodelling were dependent on the recruitment of catalytic enzymes like ATPases which correlated with

the connection found between *GATAD1* and *ABCF3* (Fig. 3.6). Moreover, *Rho1* encodes a GT-Pase that played a role in actin regulation [Magie et al., 1999] which was fundamental to further modify chromatin [Olave et al., 2002] and also shared a similar connection. *GATAD1* was further associated to *DNMT1* a methyltransferase and *HISTH2A*, both hypothetically involved in effective recruitment of non-histone proteins and methylation maintenance of the chromatin [Leu et al., 2003; Kouzarides, 2007]. Additionally, *GATAD1* seemed to be connected to *DERL2* which in turn was associated with *SMARCD1* and *Med6* (Fig. 3.6). SMARCD1 is an ATP-dependent chromatin remodelling (SWI/SNF) complex [Ho and Crabtree, 2010] and *Med6* encodes a cofactor to the RNA polymerase II [Lee et al., 1997]. We hypothesize that *GATAD1* was largely involved in gene regulation of early development of blue mussels. The above-described connections provide clues as how gene regulation was tightly coordinated, so much as at the chromatin level.

Additional data mining helped characterize several other connections one in particular between *PBX1* (TF33), *WNT16* (G5), and *NRARP* (G29) (Fig. 3.8B). The former encodes a transcription factor and the latter acts downstream of the Notch signalling pathway, involved in blood vessel endothelial cell proliferation (Table A.4). *WNT16* is considered to be an early developmental protein, and its coordinated expression with *NRARP* is central in vessel density during sprouting angiogenesis [Phng et al., 2009]. Furthermore, *ARF6* (TF15), discussed above, seemed to be connected to a calmodulin (G54), a histone chaperone (G103), and a collagen (G126) (Fig. 3.8B). Although calmodulin-related proteins are known for their ion-sensor activity and Ca^{2+} transport, they can also play a role in the calcification processes, biomineralization, epithelial cell differentiation [Li et al., 2005], muscle elasticity, and differentiation [Gregorio et al., 2005]. Collagen and other extracellular components participate in muscle differentiation [Dyachuk, 2013], blood-vessel development [Malfait et al., 2007], and forms the basis of the byssus threads released at the pediveliger stage [Lucas et al., 2002]. Collectively, these non-random associations between regulators and relevant genes

demonstrated the central role of differential connectivities of regulatory networks during early development.

3.7 CONCLUSION

Interactions between transcriptional regulators and their target genes have been thoroughly investigated. This helped us define coordinated expressions between genes and the physiological changes underlying larval body-plan transitions. Therefore, data mining of the newly built transcriptome of *M. edulis* and particularly the large-scale examination of gene expression provided new insights on the potential existence of an environmental impact on the dynamic of larval transcriptome. Our approach was based on multi-processing modules and inferential protocols that helped predict how co-regulated genes cluster together at different timepoints and how co-regulation may overlap due to evolutionary pressures. As a result, statistical modelling enabled us to map the whole transcriptome of *M. edulis* and to identify new genetic regulators. Biological processes implicated in several neural and muscular mechanisms seemed to be regulated in a coordinated way throughout early development. Our strategy combined clustering of co-regulated genes and inference of genetic regulatory networks and was thereby particularly suited as an integrative analytical pipeline for the investigation of direct regulation of gene modules.

3.8 ACKNOWLEDGEMENTS

The work presented in this paper was carried out during S. Bassim's Ph.D research at the University of Quebec at Rimouski and was supported financially by a Strategic Grant from the Natural Sciences and Engineering Research Council of Canada (STPGP 364918) and from the Réseau Aquaculture Québec (RAQ). The authors would like to thank Dr. Andrea Rau for much helpful advice concerning implementation of the ebdbNet package.

3.9 SUPPLEMENTARY INFORMATION

Comprehensive analysis of the different methods associated with our analytical pipeline i.e., larval production, sequencing, assembly, microarray data processing, data integration, and dynamic Bayesian network inferences. This appendix also includes all the GRNs and tables containing the functional mapping of the genes being coerced and annotated.

3.10 AUTHORS' CONTRIBUTIONS

SB carried out the sampling, data preparation and processing, pipeline design, data analysis and interpretation, biological knowledge, drafted and maintained the manuscript and the supporting information (SI). AT helped with the coordination of the de novo assembly of the transcriptome, participated in the microarray design and execution, and helped in drafting the manuscript. BG, DM, and RT critically improved the manuscript. All authors read and approved the final manuscript.

4

IMPACT GÉNÉTIQUE DES ACIDES GRAS SUR LA CROISSANCE ET LA MORTALITÉ PENDANT LES PREMIERS STADES DU DÉVELOPPEMENT LARVAIRE CHEZ *MYTILUS EDULIS*

Bassim Sleiman[1,2], Chapman W. Robert[3], Tanguy Arnaud[4], Moraga Dario[2], Tremblay Réjean[1]

[1] *Institut des Sciences de la mer de Rimouski, Université du Québec à Rimouski, 310 allée des Ursulines, Rimouski, Québec, G5L 3A1 Canada*

[2] *Laboratoire des Sciences de l'Environnement Marin, Institut Universitaire Européen de la Mer, Université de Bretagne Occidentale, Rue Dumont d'Urville, 29280 Plouzané, France*

[3] *Marine Resources Research Institute, South Carolina Department of Natural Resources and Hollings Marine Laboratory, 331 Ft. Johnson Road, Charleston, SC 29412, USA*

[4] *UPMC Université Paris 6, UMR 7144, Génétique et adaptation en milieu extrême, Station biologique de Roscoff*

4.1 RÉSUMÉ

Ce troisième article, intitulé *Genetic impact of dietary fatty acids on growth and mortality during early Mytilus edulis larval development*, fut corédigé par moi-même ainsi que par mes collègues les docteurs Robert W. Chapman, Arnaud Tanguy, Dario Moraga et Réjean Tremblay. Il fut soumis pour publication pendant le deuxième trimestre 2014 à la revue (Gene). En tant que premier auteur, ma contribution à ce travail fut l'idée originale, la recherche sur l'état de l'art, la préparation, le traitement et l'analyse des données, le design expérimental du pipeline, et la rédaction de l'article et les informations supplémentaires. Le docteur Robert W. Chapman a aidé dans le choix des analyses et modèle de prédiction et la révision de l'article. Les docteurs Arnaud Tanguy, Dario Moraga et Réjean Tremblay ont participé à la révision de l'article.

L'identification de la signature génétique de la croissance et de la mortalité au cours du développement larvaire est cruciale pour la production de juvéniles et dans la conception de meilleures techniques pour la sélection de génomes. Nous avons utilisé un protocole de prédiction automatique basé sur des méthodes de *Machine Learning* afin de caractériser la régulation génique en réponse à deux traitements diététiques à concentrations variables en acides gras (AG), pendant une période de temps de 42 jours après la fécondation chez les larves de *Mytilus edulis*. Les larves ont été nourries soit avec un mélange (Co) de microalgues ou une diète monoalgale (Ti). Celles élevées sous le traitement Co ont montré une accumulation progressive de lipides et AG, un faible taux de mortalité et une augmentation de la fixation des naissains et du recrutement de postlarves. En revanche, les larves élevées sous le traitement Ti ont montré un taux élevé de mortalité et une croissance réduite de la coquille. La performance des postlarves a été positivement corrélée avec une baisse en acide arachidonique et en acide eicosapentaénoïque. Les modèles d'apprentissage (Ensemble Methods) ont été combinés. Ils regroupent en premier un modèle linéaire généralisé (GLM) renforcé par une fonction de rééchantillonnage (boosted) et en deuxième un réseau de neurones artificiel (ANN). L'ensemble a permis d'obtenir les meilleures prédictions parmi plusieurs modèles testés. La combinaison du boosted-GLM et le ANN ont été choisis en fonction du taux d'erreur (RMSE) mesuré. Les prédictions de gènes ont permis d'identifier 29 marqueurs génétiques différentiellement corégulés avec une précision de classification minimale de 87 %. La réponse physiologique qui constitue la base d'apprentissage du modèle correspond à une variable binomiale qui suit le développement larvaire sous les deux traitements diététiques Co et Ti. Les marqueurs candidats identifiés sont impliqués dans les mécanismes stress-dépendant trouvés dans d'autres études.

Genetic impact of dietary fatty acids on growth and mortality during early *Mytilus edulis* larval development

4.2 ABSTRACT

Identifying the genetic signature of growth and mortality during early development is crucial to spat production and in the design of better genome selection techniques. We employed a Machine Learning framework to detect gene response to variable FA concentrations, over a time course of 42 days post-fertilization in *Mytilus edulis* larvae. Larvae were fed either a diet supplied in a rich mixture (Co) of microalgae or a deficit unialgal diet (Tiso). The Tiso diet held low values in two essential fatty acids, arachidonic acid and eicosapentaenoic acid. Co-reared larvae showed a progressive accumulation of lipids and FAs, low mortality rates, and increased spatfall and postlarvae recruitment. In contrast, Tiso-reared lavae showed higher mortality rates and reduced shell growth and postlarvae performance, all of which were correlated with the decline in arachidonic acid and eicosapentaenoic acid. Ensemble methods based on a boosted generalized linear model and a neural network performed best amongst several tested base learners and were chosen in relation to their root-square-mean error rates. Inferential predictions identified 29 gene markers differentially co-regulated with a minimum prediction accuracy of 87 % of the physiological biclass outcome of Co- and Tiso-reared larvae, some of which were implicated in stress-induced mechanisms found in other studies.

Keywords: Microarray, Machine Learning, Ensemble Methods, data mining, fatty acids, larvae, growth, mortality

4.3 INTRODUCTION

Energy reserve-depletion relating mainly to lipids is responsible for mussel larvae mortalities throughout early ontogeny [Rayssac et al., 2010]. In contrast, the performance of mussel larvae, given their fatty acid (FA) content and composition of their lipid classes, promoted a behavioral success [Pernet et al., 2004; Aranda-Burgos et al., 2014], which suggests that FAs can be efficient modulators for larval competence and juvenile performance [Pernet et al., 2005]. While these FAs could enhance growth, they are sensitive to irreversible cellular oxidative damage, especially in premature ectotherms that possess high metabolic rates [Sánchez-Lazo and Martínez-Pita, 2014].

As a result, early larval and metamorphic stages are sensitive keypoints [Bassim et al., 2014a] during which hatchery-produced larvae suffer heavy mortalities [Pernet and Tremblay, 2004; Pettersen et al., 2010]. Cellular membranes and signal transduction pathways all have indispensable FAs as essential components, which constitutes key molecules during early ontogeny [Soudant et al., 1998]. Polyunsaturated fatty acids (polyunsaturated fatty acids (PUFAs)) are primary structural biomo-lecules of microalgae lipids particularly 20:4ω6 (ArA), 20:5ω3 (EPA), and 22:6ω3 (docosahexaenoic acid (DHA)). They constitute therefore nutritional FAs with the potential to modulate larval performance and mortality [Delaunay et al., 1993]. As larvae are incapable of producing sustainable quantities of these FAs, they require a balanced feeding regimen [Hemaiswarya et al., 2011]. ArA and EPA can serve as an energy source for bivalves, as structural component of membranes, or as precursors to active eicosanoids [Pernet and Tremblay, 2004]. Eicosanoids are conserved biomolecules, integrated in a wide network of various biochemical pathways [De Petrocellis and Di Marzo, 1994; Stanley, 2006]. They can cover many housekeeping roles while keeping a homeostatic balance or they can become critical effectors during stressful timepoints throughout development [Knight et al., 2000]. It has been argued that a temporal imbalance in eicosanoids might influence the stability of the organism without causing an ever lasting damage [Howard and

Stanley, 1999]. However, in the case of a shift of the organism from its normal state, a serious deleterious effect can be registered, which is the case of metamorphic marine larvae [Pernet et al., 2005]. This study was set to examine the effect of dietary eicosanoid precursors on genomic regulation.

The blue mussel *Mytilus edulis* (Linnaeus, 1758) is a marine organism with fast-evolved adaptive behaviors. At early ontogenesis, competent larvae are transformed, on the basis of a successful genetic coordination, into juveniles fit to tolerate a wide range of ecological conditions, survive benthic life, and maintain cellular balance [Williams and Somero, 1996]. *M. edulis* have a polymorphic life-cycle, during which the early developmental stages are planktotrophic [Bayne, 1965]. The dispersive pelagic first larval form (trochophore stage) is able to sustain fluctuating conditions, feeds (veliger stage), and increases its energy reserves until the pre-metamorphic stage [Sánchez-Lazo and Martínez-Pita, 2012]. Metamorphosis radically alters the morphogenetic structure of larvae, resulting in a regulated anatomical and structural differentiation, after which the organism is transitioned into a juvenile/adult form [Williams and Degnan, 2009]. The metabolic requirements of larvae are entirely dependent on allocated energy reserves, constituted mainly by neutral lipids [Baker and Mann, 1994]. During this period, feeding is reduced until complete development of gills [Cannuel et al., 2009]. Moreover, immune functions are preferentially activated early in development [Xing et al., 2014], which aid in the elimination of primitive tissue and debris during metamorphosis or the resistance to benthic pathogens [Dyrynda et al., 1995]. Faster development and growth during larval ontogenesis might depend in part on the feeding-time and composition of the diet [Ben Kheder et al., 2010], but do these traits come at the expense of genetic flexibility?

Plasticity of larvae to shift energy resources from feeding and nutrient allocation, to maintenance and growth, to defense and survival, is usually subject to cumulative and synergistic genotypic variations [Rolf, 1992]. While studies to date have shown without exception the negative effects of suboptimal nutrition on larval and juvenile bivalves [Sánchez-Lazo and

Martínez-Pita, 2014], it is unclear which gene networks have critical impacts on development and which gene interactions are particularly implicated in vulnerability or resistance on later development. Despite their ecological and economical importance, functional genomics of bivalves started to get more attention only in recent years (eg., RNAi-mediated gene-targeting [You et al., 2012] and de novo genome annotation [Zhang et al., 2012a]). Despite the growing interest in environmental monitoring and molecular studies of non-model species, marine sequence databases are still relatively fragmented [Forrest and Carninci, 2009], making it hard to carry out thorough studies [Nikinmaa and Rytkönen, 2012] or go beyond just the ecological observations [Samain, 2011]. Gene regulation underlying the early stages of ontogenesis are yet to be discovered, facilitated only by the technological advances in molecular genetics, data-sharing repositories, and complementary pipeline analyses [Yip et al., 2013].

In this work we extended a previous work (HiSeq sequencing and de novo transcriptome annotation) with an additional set of custom-built microarrays that correspond to each developmental stage of the blue mussel *M. edulis*. We report information regarding two dietary conditions, the first is identified as a rich nutrient-value mixed microalgae and the second as an essential fatty acid (EFA)-deficient single-algal diet, used on larvae throughout the first 42 day post-fertilization (DPF). Experiments were conducted on hatchery-reared animals and the goal was to induce a deficiency in ArA and EPA, which can lower larval performance and help us explore the dynamic patterns of the physiological plasticity and interactions of synergistic genes. We would therefore be able to determine keypoints in interrelated gene regulation in early larval development. The biological background assessed here is correlated with metabolism and growth. Accordingly, we set up an analytical pipeline to accurately identify the genes behind growth and mortality, during early development of *M. edulis*, then try to predict physiological behavioral variations due to aging and dietary supplementation using their expression profiles.

4.4 RESULTS

4.4.1 *Fatty acid composition of blue mussel larvae*

Both Co and Tiso diets used in this study expressed various effects on the FA composition of larvae (Figure 4.1).

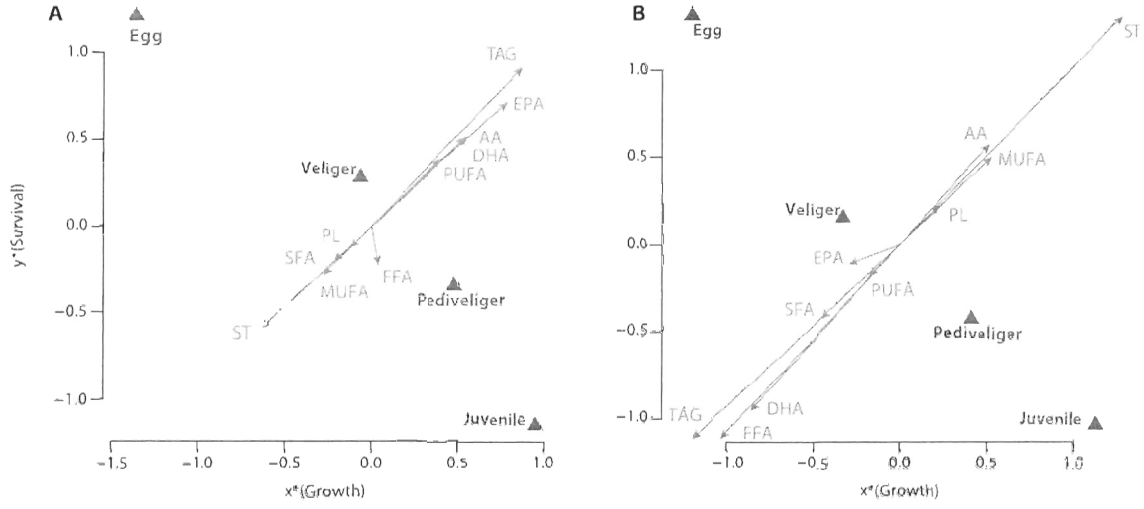

Figure 4.1: **Variation of fatty acid content as a function of growth and survival rates.** Analyses were done during early larval development (egg, veliger, pediveliger, and juvenile), given a linear regression function that combines standardized growth and survival rates for the (A) cocktail- and (B) Tiso-reared larvae. TAG, triacylglycerol; EPA, eicosapentaenoic acid; ArA, arachidonic acid; DHA, docosahexaenoic acid; PUFA, polyunsaturated fatty acid; FFA, free fatty acid; MUFA, monounsaturated fatty acid; ST, sterol; SFA, saturated fatty acid; PL, phospholipid.

Dietary requirements of Tiso-reared larvae were partially met and their needs were below the physiological threshold set by the Co diet. The eicosanoid precursors arachidonic acid (ArA) and eicosapentaenoic acid (EPA) were significantly low in the Tiso diet (Table 4.5 and Figure 4.1B). According to Table 4.2, ArA and EPA were significantly affected by the diet ($P<0.001$) but EPA was also time-dependent, specifically post-metamorphosis ($P=0.003$), whereas ArA shows a selective retention throughout development ($P=0.470$). The ratio EPA/ArA was strongly balanced between diets but was perturbed after metamorphosis ($P=0.022$). Al-

though the ratio $\Sigma\omega3/\Sigma\omega6$ was the same between diets, Tiso-reared larvae were prone to higher mortality rates (Figure 4.2) with time as the significant main effect, especially before and after metamorphosis (P=0.002).

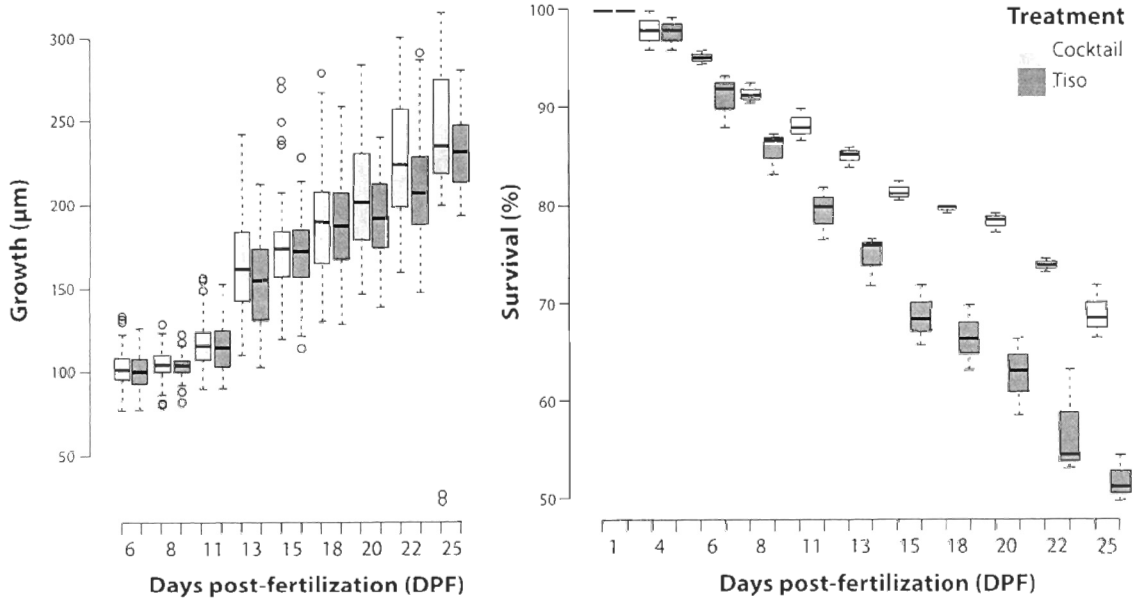

Figure 4.2: **Growth and survival of blue mussel larvae, reared under controlled conditions using two dietary treatments (cocktail and Tiso).** Left: growth as a function of days post-fertilization . On average, growth increases with the aging larvae for both treatments, with a similar median for every stage. Right: survival as a function of days post-fertilization. On average, larval mortality increases with the deficiency of the dietary treatment on essential FAs.

The Tiso diet was characterized by a high docosahexaenoic acid (DHA) levels, which was seen from the increased DHA content in post-veliger larvae (Table 4.1 and Figure 4.1B).

DHA content was significantly affected by the diet (P=0.028) and time (P<0.001) specifically in pre-settled larvae (Table 4.2). An effect can be visualized between growth, mortality, and FA contents around 25 DPF (Figure 4.3). Growth exhibited a nonlinear response (adj-R^2=0.720) to time (P<0.001), diet (P<0.001), and the combination of both (P=0.028). Mortality was linearly affected (adj-R^2=0.971) by time (P<0.001), diet (P<0.001), and the combination of both (P<0.001). Peri-metamorphic Tiso-reared larvae showed higher mortality rates and

Table 4.1: **Mass % of selected fatty acid groups of blue mussel larvae as a function of early developmental stages (and days post-fertilization) and diet (means±SD).**

Days post-fertilization	0	15		22		30	
Stages of development	Egg	Veliger		Pediveliger		Juvenile	
Diet	–	Co	Tiso	Co	Tiso	Co	Tiso
FA composition (%)							
12:0	4.4±2.2	1.4±2.4	1.1±1.9	0.2±0.4	1.3±1.2	0.6±1.0	1.1±1.8
14:0	0.0±0.0	1.6±2.8	1.1±1.8	0.2±0.3	0.0±0.0	0.6±1.1	0.0±0.0
15:0	2.4±0.7	2.4±0.4	1.5±1.0	1.3±0.6	2.2±1.6	3.2±1.8	2.9±1.1
16:0	15.0±1.9	17.7±1.1	15.1±2.0	19.5±8.9	20.5±5.5	17.2±1.4	21.1±8.5
18:0	3.9±4.0	1.4±2.5	3.1±2.9	0.2±0.4	3.9±5.5	0.0±0.0	4.5±5.1
20:0	1.2±0.5	1.2±0.1	0.9±0.2	0.5±0.3	1.1±1.0	1.6±1.0	1.4±0.7
ΣSFA	27.0±5.4	25.7±1.9	22.7±2.7	21.9±8.9	29.0±5.8	23.1±1.8	31.0±10.8
15:1	0.0±0.0	0.0±0.0	0.4±0.7	2.2±1.9	1.5±1.4	0.0±0.0	1.2±2.0
16:1	13.3±1.4	14.3±1.8	15.9±2.2	15.0±8.9	15.1±9.9	10.5±6.2	12.2±6.3
17.1	2.3±0.5	1.9±0.1	2.0±1.3	10.3±13.4	7.5±8.6	3.5±0.6	6.2±6.5
18:1ω9	22.9±5.8	22.4±1.7	22.0±6.4	11.6±3.1	13.3±0.6	13.6±2.9	12.5±3.7
20:1	2.4±2.0	0.8±0.7	3.1±1.5	1.9±2.0	0.6±0.7	1.4±0.9	0.8±1.0
22:1ω9	0.0±0.3	0.0±0.4	0.3±0.9	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
ΣMUFA	40.7±11.8	39.2±0.6	43.7±6.6	41.0±7.5	38.0±11.7	29.1±2.1	32.9±8.5
18:2ω6	6.1±2.0	5.9±0.1	7.7±0.1	2.1±0.3	3.2±1.3	4.1±1.3	2.1±1.9
18:3ω3	2.7±0.7	2.6±0.1	6.0±5.5	1.0±0.6	0.7±0.7	1.5±2.0	1.3±1.2
18:3ω6	3.2±1.4	3.1±0.3	1.9±1.0	1.1±0.4	3.3±3.3	4.7±2.8	4.5±3.9
18:4ω3	3.2±1.4	3.1±0.4	1.6±1.4	1.0±0.5	3.1±3.2	4.6±2.9	4.5±4.0
20:4ω6 (ArA)	2.3±0.8	2.8±0.6	1.3±0.3	4.8±1.6	1.2±0.6	4.7±2.4	1.0±0.4
20:5ω3 (EPA)	4.7±1.4	6.7±1.4	1.2±0.2	10.2±1.3	1.5±1.0	12.5±2.3	3.4±1.6
22:6ω3 (DHA)	10.1±1.5	10.9±0.7	13.9±2.2	16.9±2.0	20.1±3.1	15.8±2.8	19.4±4.7
ΣPUFA	32.3±6.9	35.1±1.2	33.6±3.9	37.1±5.0	33.0±8.3	47.8±3.2	36.1±8.1
Σω3	20.7±4.6	23.3±0.9	22.8±4.9	29.1±3.4	25.3±4.6	34.4±3.6	28.5±5.8
Σω6	11.6±2.7	11.8±0.3	10.9±1.3	8.0±1.7	7.6±4.0	13.4±1.9	7.6±2.4
Σω3/Σω6	1.8±0.3	2.0±0.1	2.2±0.7	3.7±0.5	3.7±1.1	2.6±0.5	3.9±0.5
DHA/EPA	2.3±0.6	1.7±0.4	12.0±1.0	1.7±0.3	22.6±21.1	1.3±0.3	7.5±5.3
EPA/ArA	2.1±0.2	2.5±0.5	1.0±0.3	2.3±0.6	1.3±0.7	2.9±0.8	3.7±2.4
DHA+EPA	14.8±2.5	17.6±1.3	15.1±2.3	27.1±2.5	21.6±2.2	28.3±4.0	22.8±3.7
EPA+ArA	7.0±2.1	9.5±1.7	2.4±0.2	15.0±2.8	2.6±1.4	17.1±4.7	4.4±1.7

All values represent the means±SD., N=3 larval tanks by treatment. Reported FAs contribute to 1 % or more of the ensemble.

equivalently Co-reared larvae had a faster growth. We therefore distinguished two larval populations, before and after 25 DPF, different by the means of their physiological performance.

Co-reared larvae exhibited higher concentration of PUFAs (P=0.05), especially in later larval stages (around 15 DPF) and after metamorphosis (>22 DPF, Table 4.1), which was translated in higher growth rates (Figure 4.3). PUFAs are structural biomolecules, presumably essential for the maintenance and neo-synthesis of cellular membranes [Hulbert et al., 2007]. The proportion of these FAs increases as a consequence of sustaining the ever increasing growth rate. The demand for structural components is dependent on PUFAs, thus the concentration

Table 4.2: **Summary of factorial MANOVAs and multiple comparisons test predicting the effect of dietary treatment (two levels) and time (three levels) on FA composition in blue mussel larvae.**

Model	Fatty acid	MANOVA			Multiple comparisons*		
		Diet	Time	Diet×Time	Diet**	Time**	Diet×Time
MANOVA I	Overall***	0.230	0.150	0.670			
	SFA	0.220	0.790	0.300			
	MUFA	0.634	0.062	0.624			
	PUFA	0.050	0.077	0.316			
MANOVA II	Overall***	**<0.001**	**<0.001**	0.480			
	EPA	**<0.001**	**0.003**	0.392	Tiso<Co	V^{b} P^{ab} J^{a}	
	DHA	**0.028**	**0.004**	0.989	Co<Tiso	V^{b} P^{a} J^{a}	
	Σω3	0.110	**0.016**	0.602		V^{b} P^{ab} J^{a}	
	ArA	**<0.001**	0.470	0.210	Tiso<Co		
	Σω6	**0.048**	**0.048**	0.144	Tiso<Co	V^{b} P^{a} J^{ab}	
MANOVA III	Overall***	**<0.001**	**<0.001**	0.394			
	EPA/ArA	0.082	**0.022**	0.128		V^{b} P^{b} J^{a}	
	DHA/EPA	**<0.001**	0.275	0.377	Co<Tiso		
	Σω3/Σω6	0.167	**0.002**	0.220		V^{b} P^{a} J^{a}	

Significant probabilities are in **bold** ($P<0.05$).

* Tukey's HSD multiple comparisons test was used to predict which means significantly differ. Different letters equate for significant differences $P<0.05$.

** Abbreviations: Co, cocktail; Tiso, Tiso; V, veliger; P, pediveliger; J, juvenile.

*** Wilks' Lambda exact statistic

of accumulated neutral lipids as in monounsaturated fatty acids (MUFAs) and saturated fatty acids (SFAs) decreases. As a result, Co-reared larvae had a steep decline in ΣMUFA and ΣSFA at 30 DPF (Table 4.1), which can be visualized in a fast increase in growth rate in Figure 4.3 at the end of the metamorphic transition. SFAs, especially 16:0 was assessed as a promoter of larval growth [Costa et al., 2012]. Their levels were sufficiently balanced throughout development under both diet treatments (Table 4.1). Furthermore, 18:1ω9 was found to be implicated in mechanisms of starving larvae [Costa et al., 2012]. In our study, we confirm the reduced level of this FA during peri-metamorphosis (22-30 DPF) in both treatment conditions (Table 4.1). Rapid growth of mussel larvae will be considered in this work as an indicator of successful use of energy reserves. Shell growth and FA composition form both a reliable relationship, which describes the low mortality threshold and high performance (i.e.,

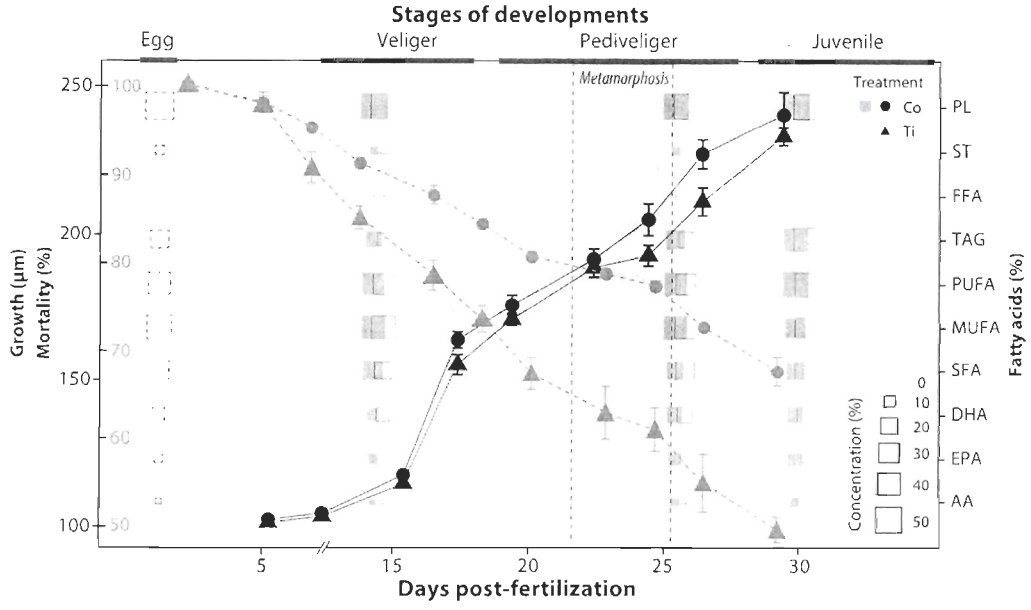

Figure 4.3: **Growth, mortality, and fatty acid content during early *M. edulis* larval stages of development.** The stages are shown in the upper side of the main grid, in addition to the black bars representing the duration of each developmental stage. Metamorphosis is represented through a gray period. The days post-fertilization (DPF) are in the lower side of the main grid. Shell growth measurements were taken from 5 DPF (in µm, $N \leqslant 100$) and are represented by solid black lines. Survival rate (mortality) was measured from 1 DPF and is the number of individuals in the sample relative to the stock culture (multiplied by 100) during each timepoint, represented by dotted gray lines. Both treatments were depicted, Co for cocktail- and Tiso for Tiso-reared larvae respectively. Selected FA classes were included in the plot (means, N=3 larval tanks, means±SD can be found in Table 4.1), in which the surface area of each related rectangle represents the concentration of that FA for both treatments. During the egg stage, no treatment was assigned to the samples. Growth (black) and mortality (gray) are projected onto the left y axis and FA contents are projected onto the right y axis.

settlement, metamorphic rates, and spat resistance) between both Co and Tiso treatments [Marshall et al., 2010].

4.4.2 *Data visualization*

Gene-ranking properties were generated from mean-log gene expression (e.g., *P*-values adjusted for multiple testing, average fold change, and moderated *T*-statistics [Smyth et al., 2005]), and a binomial categorical variable corresponding to the larval phenotypic state: a faster growth or a high mortality, each representing an experimental condition either at a rich or deficit diet respectively. Rows of the data matrix represent the samples where $0<i\leqslant n_s$ and columns represent genes where $0<j\leqslant n_g$ in (Equation 4.1).

$$X_{ij} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n_g} \\ x_{21} & x_{22} & \cdots & x_{2n_g} \\ . & . & . & . \\ . & . & . & . \\ x_{n_s1} & . & . & x_{n_sn_g} \end{pmatrix} \tag{4.1}$$

Ranking of top significant 4,000 genes was parametrized with a set of contrast matrices (26.67 % of the dataset, $P\leqslant 0.06$), which ensured pairwise feature-comparison (Figure 4.13) to distinguish relationship of significant genes with the five larval stages, the treatment main effect, and both their interactions, which is a direct interest to our study. In Figure 4.13 setups I and II significantly represent differentially expressed genes ($P<0.05$) of Co- and Tiso-reared larvae respectively. Setup III is the differential ranking when the effect of both treatments is combined with all five developmental stages. Setups IV and V are relative to the main effects of treatment and time respectively. A significant amount of variability in growth and mortality rates was associated with each stage of development, and so developmental stages alone did not provide an accurate prediction of larval physiological success. FA concentrations were typically greater for Co-reared larvae. As a result, at peri-metamorphic

Table 4.3: **Number of annotated transcripts used to study gene expression of *Mytilus edulis* larvae**

	Datasets	Percentage
Transcriptome	50,383	
Functional Transcripts	15,257	30.28 %
Catalytic Function	4,703	30.83 %
RNA Features	1,254	8.22 %
Protein Features	9,300	60.96 %

stages, when larvae started to feed, lipids especially EFAs tended to have higher concentrations than Tiso-reared larvae (Figure 4.3 and Table 4.1). In contrast, mortality rates of Co-reared larvae were significantly lower than Tiso-reared larvae. Hence, the most accurate prediction of larval physiological success will clearly be obtained by studying the genomic response during both Co×Tiso treatments. Although gene-ranking was not used directly in machine learning, it provided a first global picture of the collinearity and redundancy of gene expression. Setup 3 confirms the mutual effect of diet enrichment and the transitioning aging larvae. It shows also the peri-metamorphic separate differential expression. Furthermore, 619 genes were preferentially distinguished in the third setup, the highest amongst all. This confirms, the special and combined effect of diet and developmental stage on larval performance.

4.4.3 *Setting the learning rules*

Substantial amount of sequences (91.78 %) and gene expression profiles were generated from the de novo assembly of *M. edulis* transcriptome, which were unequivocally central to hypothesis testing (Table 4.3). Gene expression data was divided into validation and generalization matrices, where the former was used for training the classifiers and the latter was used as a dedicated testing data for an unbiased evaluation of the prediction estimates of the ensemble methods (Figure 4.4).

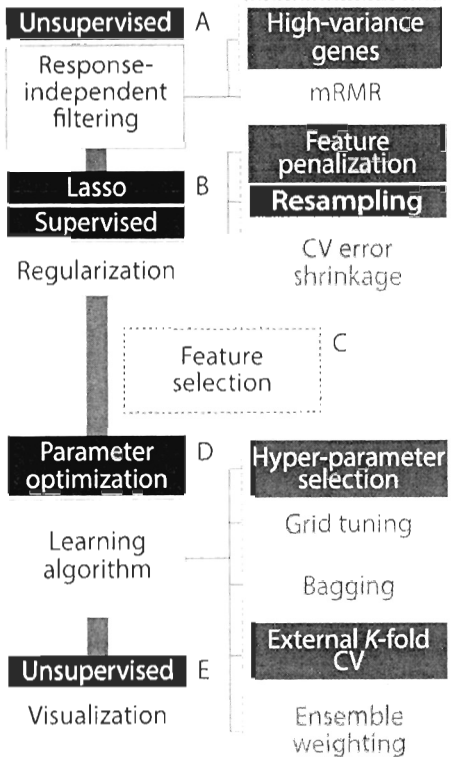

Figure 4.4: **Workflow machine learning analysis for prediction of most related genes to a particular treatment.** (A) Unsupervised feature elimination was based on high-variance and minimum redundant maximum relevant (MRMR) rules for model reduction. (B) Supervised model selection was based on penalized likelihood (least absolution shrinkage and selection operator, Lasso) for variance reduction of model fitting. Nested cross-validation (CV) was used for penalty estimation and (C) features were selected after re-fitting all the model. (D) Before bagging (bootstrap aggregating), hyper-parameters tuning of multiple learning algorithms ensued. Root-mean-square error (RMSE) was used for model selection and generalization was done using ensemble learning weighed predictions to increase estimation accuracy. (E) One-way clustering grouped co-regulated genes, which reduced the dimensionality of the dataset.

Given that machine learning algorithms depend on the classifier(s)' efficiency and accuracy, a gene selection step was implemented to reduce the number and dispose of irrelevant or less-significant genes (Figure 4.4A-B). Inference can not be formed using all the features p under the prediction rules, because the noise associated with some genes will substantially affect the variance of the estimated predictions [Ambroise and McLachlan, 2002]. High-

ranked minimum redundant maximum relevant (MRMR) genes were used as targets to select 500 subgenes from an initial set of 2,000 high-variant genes that account for 5 % of the whole microarray dataset. The target genes were tested for prediction accuracy and their averaged root-mean-square error (RMSE) were registered using multiple base learners after least absolution shrinkage and selection operator (lasso) supervised feature selection (Figure 4.5). Gene I provided the best prediction estimates and was used next to select the best eigengenes that show correlation with the target gene but had less correlation amongst each other. Subset I was chosen after validation was done on different base learners (Figure 4.6).

The supervised and unsupervised combination of feature-subset selection helped select a compact subset of genes (Figure 4.4C) that optimizes the given classification without significant loss of the predictive accuracy [Li et al., 2014]. Faster computations and interpretability was additionally achieved. Amongst the tested classifiers, generalized linear models (GLM) with a component-wise boosting algorithm [Buehlmann, 2006] increased its power when combined with the second classifier, a nonlinear feed-forward ANN with 17 units in the hidden layer [Venables and Ripley, 2002]. When averaging both of the classifiers while upweighing the ANN predictions (method III), the ensemble methods delivered the best RMSE estimates for our high-dimensional data (Figure 4.7) and a minimum of 87 % classification accuracy.

4.4.4 *Gene markers for FA deficiency*

In Figure 4.8 we showed the heatmap of the 29 marker genes selected for inferential classification of transcriptional expression datasets from Co- and Tiso-reared larvae monitored throughout ontogenesis. Red, green, and black spots represent high, low, and baseline gene expressions respectively. The relationship between these markers and the response outcome was the main goal in this study.

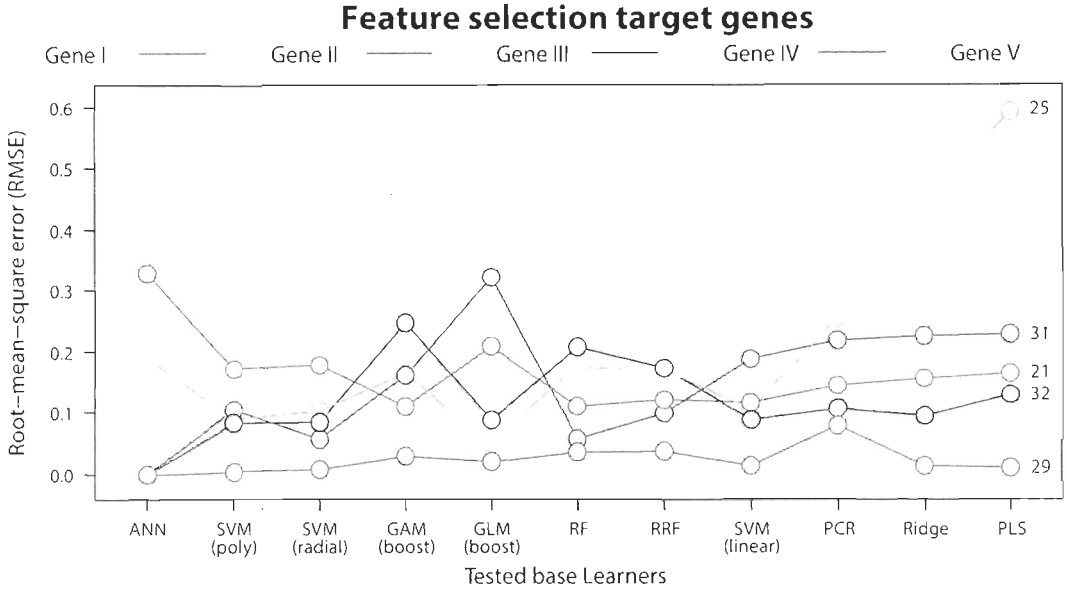

Figure 4.5: **Minimum redundancy feature selection using a filter approach.** The search for a low redundancy feature set uses high ranked genes as targets for feature selection. We assessed the root-mean-square error (RMSE) of the classification of the first five high-ranked genes. These genes will help filter highly-correlated subgenes to the target genes (shown here) but are a priori less-correlated amongst each other [Ding and Peng, 2005]. Gene I showed the lowest RMSE estimates across all used base learners. As a result, Gene I was selected as target gene for further minimum redundancy maximum relevance MRMR unsupervised gene filtering. Spearman was used for estimation of correlation. The unsupervised gene filtering method were carried out on 2,000 high-variance genes. Microarrays helped register the profile of 40,000 genes across development and across two dietary treatments (cocktail, rich diet; Tiso, deficient diet), 15,000 of which were annotated. Seventy five percent of these annotated high-variance genes were screened for MRMR. Double numbers in the right side represent the final lasso (least absolution shrinkage and selection operator) feature selected, trained by a base learner, and validated then generalized on a holdout test set. ANN, multilayer perceptron artificial neural network; SVM, support vector machines; GAM, generalized additive models ; GLM, generalized linear models; RF, random forest; RRF, regularized random forest; PCR, principal components regression; Ridge, ridge regression; PLS, partial least squares. All functions were iterated using a workstation cluster for parallel computations in R.

Coordinated regulation during the separate phases of larval development influenced by an exogenous FA deficiency were further documented in Table 4.4, along with the marker de-

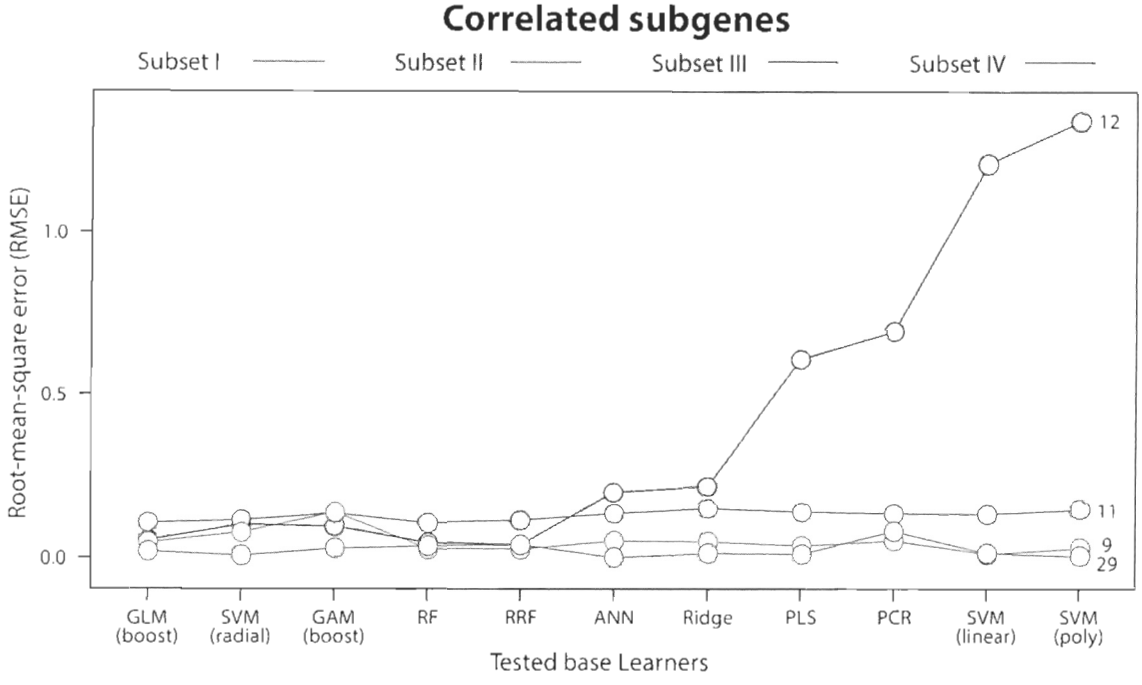

Figure 4.6: **Filtered highly-correlated subgenes to the target genes (Figure 4.5) but less correlated amongst each other.** These subgenes represent the last step of the feature selection filter approach by comparing the prediction generated from the filtered subgenes. Amongst all the five subsets related to Gene I shown here, Subset I retained the lowest root-mean-square error (RMSE) prediction error rate. Double numbers in the right side represent the final lasso (least absolution shrinkage and selection operator) feature selected, trained by a base learner, and validated then generalized on a holdout test set. ANN, multilayer perceptron artificial neural network; SVM, support vector machines; GAM, generalized additive models ; GLM, generalized linear models; RF, random forest; RRF, regularized random forest; PCR, principal components regression; Ridge, ridge regression; PLS, partial least squares. All functions were iterated using a workstation cluster for parallel computations in R.

scription, UniProt entry, and the modulated modularity clustering (MMC) modules. Gene expression was described for the egg (E), trochophore (T), Co-reared veliger (VC), pediveliger (PC), and juvenile (JC), and Tiso-reared veliger (VT), pediveliger (PT), an juvenile (JT).

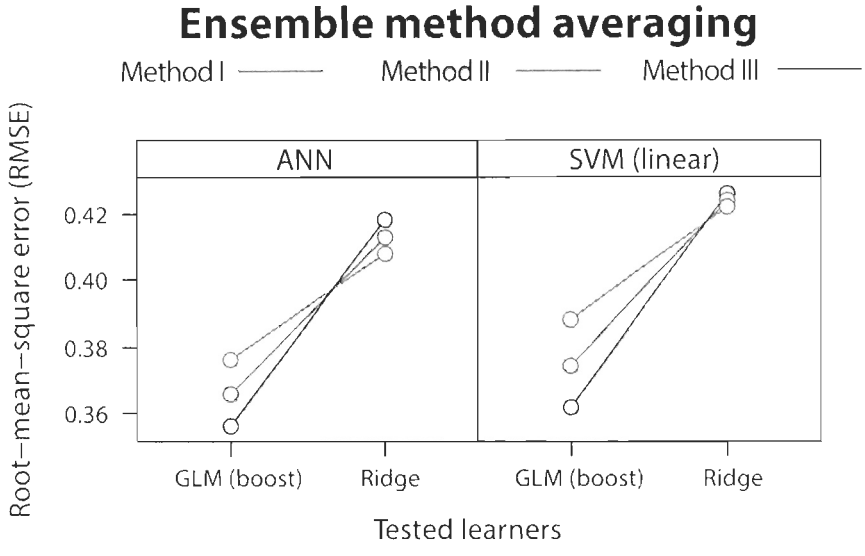

Figure 4.7: **Ensemble method averaging across the base learners with the best prediction estimates and the lowest root-mean-square error (RMSE).** Three methods were used for ensembling. Method I, baseline without weighing; Method II, ANN and SVM were down-weighed; Method III, ANN and SVM were up-weighed. Method III for ANN vs. GLM was selected for reducing the RMSE threshold to the maximum. The ensemble methods were based on a simple GLM boosted function combined with a more complex ANN learner for most accurate predictions. ANN, multilayer perceptron artificial neural network; SVM, support vector machines; GLM, boosted generalized linear model; Ridge, ridge regression.

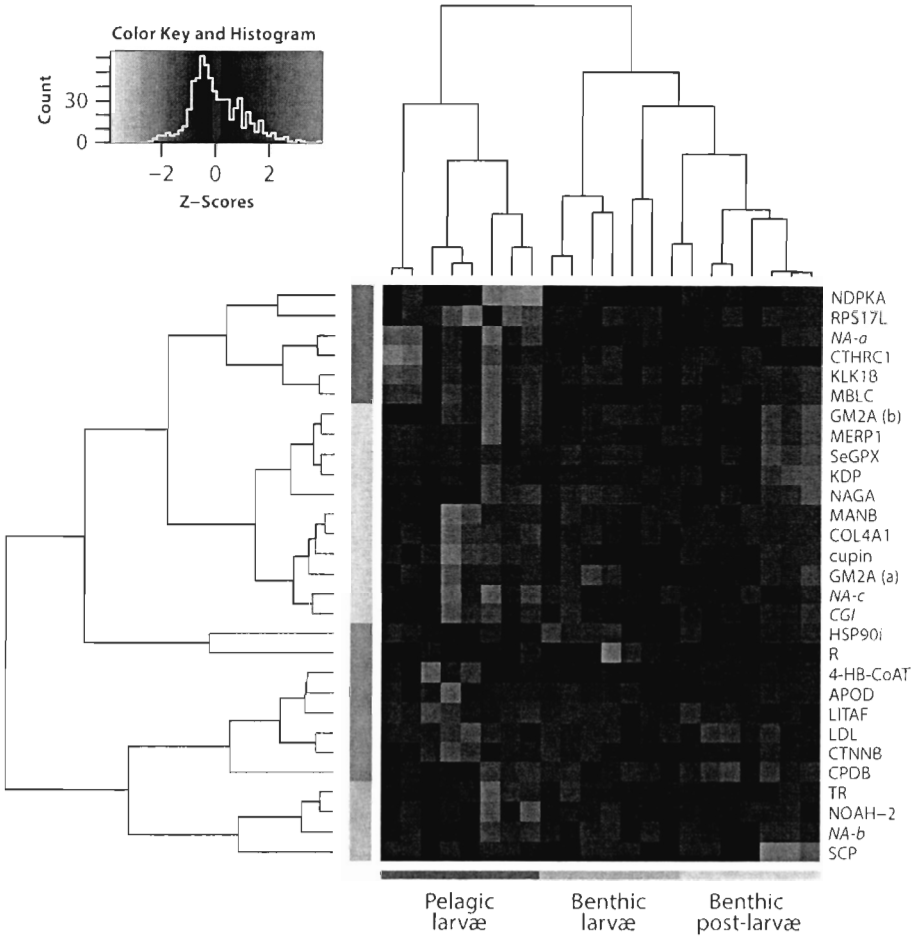

Figure 4.8: **Differential expression of the 29 marker genes selected for each diet treatment and the stage of development of *Mytilus edulis* larvae.** Red and green spots represent the up- and down genetic expression. Black is the baseline of gene expression. Gene symbols are shown on the right side of the heatmap. Dendrograms of gene clustering on the left were tested for significance (cutoff $\leqslant 0.05$ threshold) using 2000 bootstraps. Dendrograms at the top represent clustering of larval and postlarval stages.

Table 4.4: **Predicted genes and their implication in larval growth and mortality during the early stages of development.**

			Differential expression**									
					Cocktail			T-iso				
Dataset index	Symbol	MMC*	E	T	VC	PC	JC	VT	PT	JT	Protein description	UniProt***
Locus14902	*MERP1*	1	○	○	○	○	●	○	●	○	Mammalian ependymin-related protein 1	Q96J80
Locus18467	*KDP*	1	○	○	○	○	●	○	●	○	Kazal domain-containing peptide	B6DDZ6
Locus12217	*CTHRC1*	2	○	○	○	●	●	○	●	○	collagen triple helix repeat containing protein	Q96CG8
Locus39409	*NA-a*	2	○	○	●	●	●	○	●	○	*Non-annotated*	
NODE_102703	*MBLC*	2	○	○	○	●	●	○	●	○	Mannose-binding protein C-like isoform 6	
Locus15939	*NA-b*	3	○	○	●	○	○	○	○	○	*Non-annotated*	
Locus2223	*RPS17L*	3	○	○	○	○	●	●	●	●	40s ribosomal protein S17-like	P0CW22
Locus20448	*LITAF*	4	●	●	○	○	○	○	○	○	Lipopolysaccharide-induced tumor necrosis factor-alpha factor-like protein	Q99732
Locus2768	*CTNNB*	4	●	●	○	○	○	○	○	○	Beta-catenin	P35222
Locus4075	*LDL*	4	●	●	○	○	○	○	○	○	Low density lipoprotein	
NODE_72817	*APOD*	4	●	●	○	○	○	○	○	○	Apolipoprotein D	P05090
NODE_98355	*4-HB-CoAT*	4	●	●	○	○	○	○	○	○	4-hydroxybutyrate coenzyme A transferase	Q898H0
Locus1692	*CGI*	5	○	○	○	○	●	○	●	●	hypothetical protein CGI_10028077	
Locus1990	*SeGPX*	5	○	○	○	○	●	○	●	○	Selenium-dependent glutathione peroxidase	
Locus26575	*GM2A* (a)	5	○	○	○	●	●	○	●	●	Ganglioside gm2 activator	P17900
NODE_153482	*cupin*	5	○	○	○	○	●	●	●	●	Cupin family protein	F4IQK5
NODE_67785	*KLK1B*	5	○	○	○	○	●	○	●	●	Plasma kallikrein	P03952
Locus12375	*NOAH-2*	6	○	○	●	○	○	○	○	○	nompA-homolog 2	Q9XUE5
Locus34728	*NDPKA*	6	○	○	○	●	●	○	●	●	Nucleoside diphosphate kinase	P15531
NODE_274308	*TR*	6	○	○	●	○	○	○	○	○	Family transcriptional regulator	
Locus12314	*R*	7	○	○	○	●	○	○	○	○	R domain-containing protein	
Locus23055	*GM2A* (b)	7	○	○	○	○	●	○	●	○	Ganglioside gm2 activator	P17900
Locus27574	*SCP*	7	○	●	●	○	○	●	○	○	Sarcoplasmic calcium-binding protein	P04570
Locus33335	*CPDB*	7	●	●	○	○	○	○	○	○	Bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase precursor	P08331
Locus36544	*NAGA*	7	○	○	○	○	●	○	●	○	Alpha-N-acetylgalactosaminidase-like isoform 4	
Locus22673	*HSP90i*	8	○	○	○	●	○	●	○	○	A chain tricyclic series of HSP90 inhibitors	
Locus2838	*COL4A1*	8	○	○	○	●	●	○	●	●	Collagen alpha-(IV) chain	P02462
Locus37848	*NA-c*	8	○	○	○	●	●	●	●	●	*Non-annotated*	
Locus7	*MANB*	8	○	○	○	○	●	○	●	●	Lysosomal alpha-mannosidase	Q8VHC8

* MMC, represents the modulated modularity clustering of the 29 markers. For additional analysis of MMC, the reader should report to Figure 4.9 for a multicomparison visualization of the 29 markers and Figure 4.11 for a global clustering of the correlated genes.

** Differential expression is detailed for E, egg; T, trochophore; Co-reared VC, veliger; PC, pediveliger; JC, juvenile respectively; Tiso-reared VT, veliger, PT, pediveliger; JT, juvenile respectively. White and black circles represent down- (and baseline) or up-regulation of genes respectively.

*** UniProt entries correspond to the cataloged proteins in the Universal Protein Resource repository.

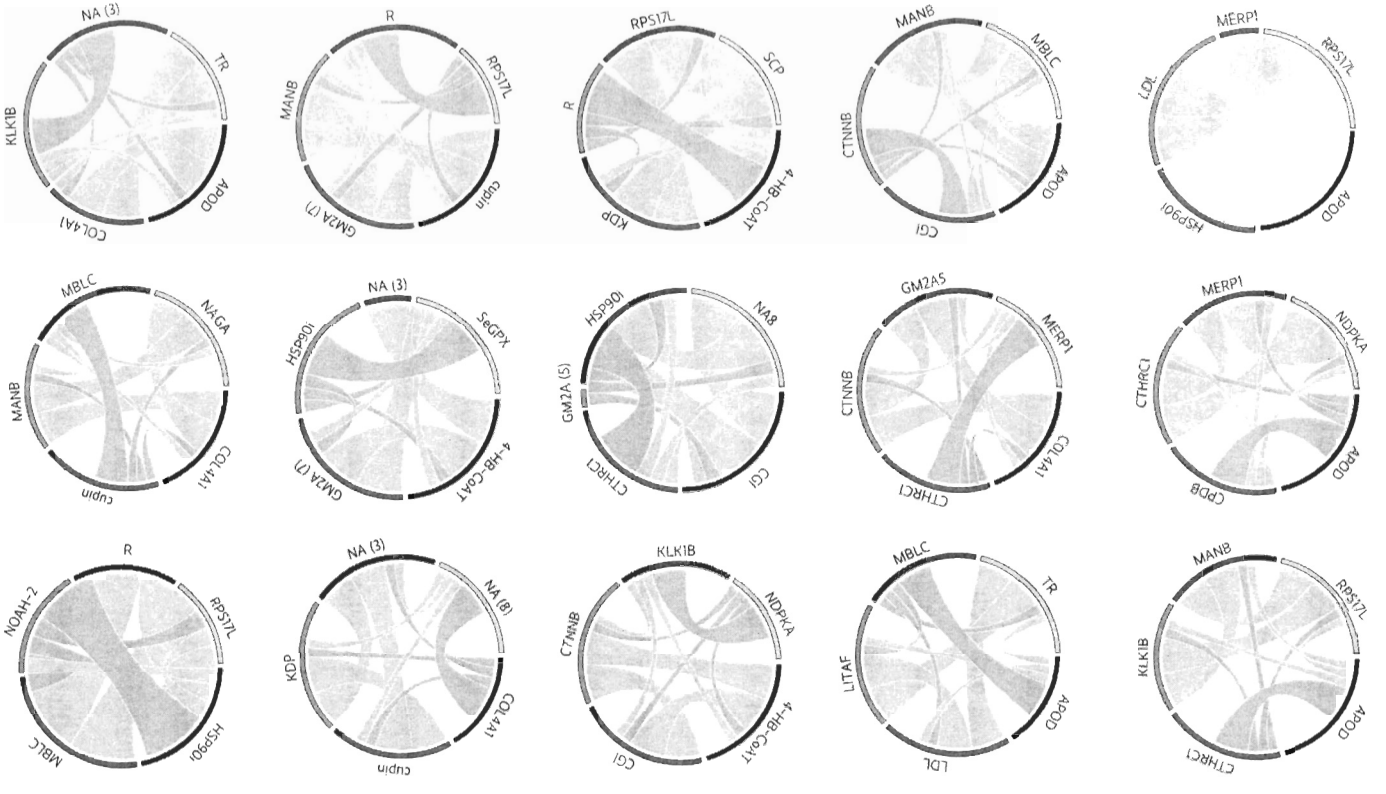

Figure 4.9: **Relationship of several gene markers according to their correlated coefficients.** Genes are represented by the separate external sections. Genes in each "circos" or circle are chosen at random from the set of 29 marker genes of Table 4.4. Green and orange links correspond to negative and positive expression correlations between genes. Coefficients are ANOVA residuals calculated after fitting for each gene a linear model function of dietary treatment (Cocktail and Tiso) and the stage of development (egg, trochophore, veliger, pediveliger, and juvenile). Description of each gene can be found in Table 4.4.

The reported MMCs show increasing similarity with the hierarchical clustering of Figure 4.8, for the same set of marker genes. The underlying modulars (i.e., network of genes with small number of connections) are not necessarily aimed to reconstruct a detailed heavily connected network, but rather, to extract maximum number of information from the 29 marker set. For example, *4-HB-CoAT*, *APOD*, *LITAF*, *LDL*, *CTNNB* clustered together in both methods with a preferential up-regulation in E and T. *CGI*, *SeGPX*, *GM2A*, and *cupin* also clustered and were up-regulated in JC, PT, and JT.

When we considered the MMC modules of Table 4.4 several signature genes had higher expression in later Co-reared larvae but their expression was up-regulated in earlier stages of Tiso-reared larvae. Module 1 comprising of *MERP1* and *KDP* followed this same pattern of expression. Modules 2, 5, and 7 also showed same trends. A preferential VC gene up-regulation was also noted for genes NA-a, NA-b (nonannotated), *NOAH-2* and *TR. SCP* was preferentially expressed only in T, VC, and VT. During E and T stages, *LITAF*, *CTNNB*, *LDL*, *APOD*, and *4-HB-CoAT* were also preferentially expressed. Nonannotated transcripts lack the gene identification entries in sequence databases. *ganglioside gm2 activator (GM2A)* gene had two queries in Table 4.4 which could result from either an alternative splicing mechanism of the same gene or transcription from two separate genes.

The biological relationship among the selected markers was assessed using an enrichment analysis for the different GO classifications. Biological process (BP), molecular function (MF), and cellular component (CC) Fisher's enrichments are reported in Tables B.1, B.2, and B.3 respectively. Each table included a list of GO-terms shared by the marker genes. For each GO-term, the number of entries found in the whole GO-mapped transcriptome of *M. edulis* and the number of markers sharing that term were also reported. It is evident from the tables that the inferred gene markers were involved in similar BPs, MFs, and CCs.

For the purpose of illustration, the transcriptional data of the 29 markers was further used to project the classification of the different larval and postlarval conditions. The first three principal components explained 80.9 % of the variation (Figure 4.10).

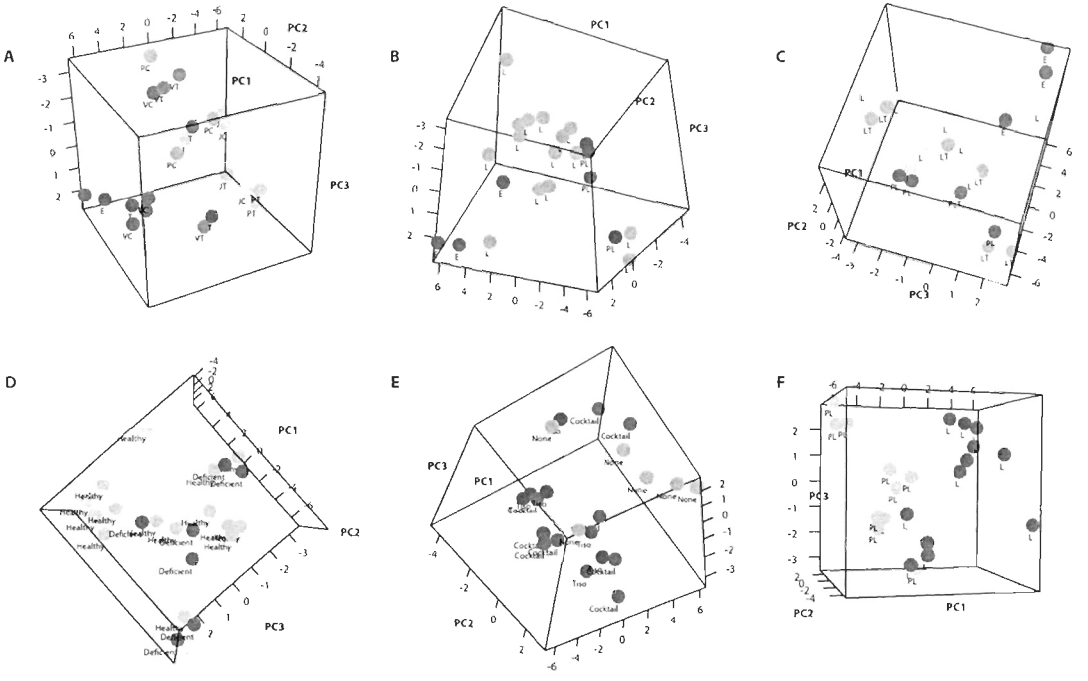

Figure 4.10: **Principal components analysis of sampled blue mussel larvae projected according to the transcriptional profiles of the 29 selected marker genes of Table 4.1.** The first three principal components explained 80.9% of the variance from the differential gene expression due to treatment (cocktail, Co; Tiso, Tiso) and time effects. The six differentially oriented plots provide a cluster visualization of predicted samples using those 29 genes. (A) Structural projection of samples provided the following labels: egg (E), trochophore (T), Co-reared veliger (VC), pediveliger (PC), and juvenile (JC), and Tiso-reared veliger (VT), pediveliger (PT), and juvenile (JT). (B) Structural projection of samples provided the following labels: egg (E), larvae (L), and postlarvae (PL). (C) Structural projection of samples provided the following labels: egg (E), larvae (L), postlarvae (PL), deficient larvae (LT), and deficient postlarvae (PLT). (D) Structural projection of samples provided the following labels: Co-reared larvae (Healthy) and Tiso-reared larvae (Deficient). (E) Structural projection of samples provided the following labels: egg and trochophore (None), veliger, pediveliger, and juvenile (Cocktail and Tiso). (F) Structural projection of samples provided the following labels: larvae (L) and postlarvae (PL).

The different classifications were compared using several labeling schemes for larval or postlarval stages and rich (Co) or deficient (Tiso) diets. Same samples were grouped together following preferential patterns when genetic data of the marker genes were evaluated in Figure 4.10. It is evident from the principal components regression that the marker genes are able to predict the association of the samples *n* to their relative classes level.

4.5 MATERIAL AND METHODS

4.5.1 *Larval collection and sampling*

Adult blue mussels were obtained from the Havres-aux-Maisons lagoon, Magdalen Islands (47°25'N, 61°51'W) in Quebec, Canada harboring a pure population of *M. edulis* [Tremblay et al., 1998]. Transfer permits of Fisheries and Oceans Canada have been obtained for breeders and the studies did not involve endangered or protected species. Gametes from different individual parents were used in a pool-cross design to produce one random larval family and was conducted at the Aquicole Station of Pointe-aux-Peres for the University of Quebec at Rimouski (UQAR), Quebec. Cultured larvae and postlarvae were reared in a flow-through system following the method described in [Bassim et al., 2014b]. Larvae were fed continuously at 30,000 cells ml^{-1} [Andersen et al., 2011] either with a cocktail (Co) of three different algae species (*Isochrysis galbana, Pavlova lutherii, Nannochloropsis oculata*) or with a Tiso diet (*Nannochloropsis oculata*). Microalgal stocks were obtained from the Center for Culture of Marine Phytoplankton (CCMP), Bigelow Laboratory for Ocean Sciences and were batch-cultured in F/2 medium [Guillard, 1975] and silicates with 0.1 µm filtered and UV treated sea water at 20°C under continuous illumination. Mussel larvae and postlarvae were obtained at discrete timepoints of development: trochophores, veligers, pediveliger larvae, and juvenile postlarvae, beginning at 1 day post-fertilization (DPF). Morphological criteria enabled the level of competence to be determined for metamorphosis of larvae dependent on eyespot

and foot formation [Pechenik, 1980; Beaumont and Budd, 1982]. Replication consisted on three independent tanks (N=3) for every diet treatment.

4.5.2 *Fatty acids and RNA quality control*

Samples containing known number of eggs, larvae, and postlarvae were collected at 0, 1, 4, 6, 11, 13, 15, 18, 20, 22, 25, 30, and 42 DPF, flash-frozen in liquid nitrogen and stored at −80°C until usage for FA and RNA extraction. Mussel eggs (≈30 x10^3), larvae (≈15 x10^3), and postlarvae (≈5 x10^3) were sampled during ontogenesis between April and June (2011-2012). Three biological replicates (N=3 larval tanks per diet treatment) during each stage were sampled on pre-combusted Whatman GF/F 25 mm filters for determination of lipid class and FA composition. Individuals at 0, 15, 22, and 30 DPF were sampled and stored in lipid-free amber glass vials with Teflon™-lined caps under nitrogen in 1 mL dichloromethane at −80°C. Lipids were extracted then neutral and polar fractions separated, and FA profile characterized by GC-MS, as described in [Toupoint et al., 2012]. FA composition was expressed as the percentage of total lipids as µg of FA per mg of dry mass (~ $ng.\mu m^{-3}$). The ratio $\omega 3/\omega 6$ was calculated using the sum of molar proportions (%) of the polar lipid fraction. Total RNA was extracted following the protocol described in [Bassim et al., 2014b]. Quality and quantity assessment of RNA was performed using the NanoDrop ND-8000 UV spectrophotometer designed for high throughput sample evaluation and RNA integrity checks were performed with the Bioanalyzer and the RNA6000 Nano assay kit (Agilent Technologies) designed for minimal sample consumption.

4.5.3 *Transcriptome analysis*

Our dataset includes more than 40,000 features (p) and 22 microarrays (n). Amongst these predictors (transcripts), some exhibit large correlation with the outcome response (Figure

4.11). We had the advantage of choosing either genes with high-variance or high-similarity rates. The former includes genes that vary in expression levels throughout the sampled stages. The latter includes genes with matching patterns of expression and correlated functions. Hence, the workflow was built on the notion of the former rule, since high-similarity gene pattern recognition were discussed in [Bassim et al., 2014b]. Sequencing, assembly, and annotation procedures for the transcriptome of *M. edulis* and microarray chip construction were all covered in detail in [Bassim et al., 2014b].

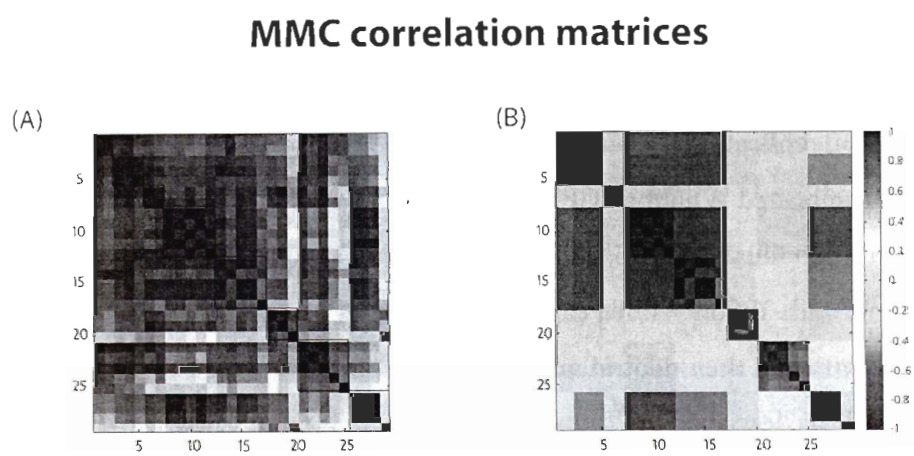

Figure 4.11: **Modulated Modularity Clustering (MMC) analysis of *Mytilus edulis* genetic data.** Blue mussel larvae were sampled throughout ontogenesis and under two dietary treatments that differ in fatty acid (FA) composition ratios (cocktail, Co FA rich diet; Tiso, Tiso FA deficient diet). (A) represent the reordered pairwise Pearson correlated transcriptional profiles and (B) when smoothing was applied on the selected 29 markers. analysis of variance (ANOVA) residuals were used to infer the modulated relationship between genes when both the dietary treatments were taken into account with all five developmental stages (egg, trochophore, veliger, pediveliger, and juvenile). The clusters identified by the MMC are arranged in Table 4.4. These 29 tabulated genes are those selected from the machine learning process described in Figure 4.5 Gene I, Figure 4.6 Subgene I, and Figure 4.14.

In this study we extended our previous work [Bassim et al., 2014b] with a set of new Agilent 8x60K microarrays that target early larval stages, given a second dietary treatment condition. Preprocessing of microarrays was equivalent to a first step of an adaptive background subtraction for differential expression [Ritchie et al., 2006a], then a loess normaliza-

tion method, which gave us the best reduction in variability along low and high intensities within each array (density plots in Figure 4.12).

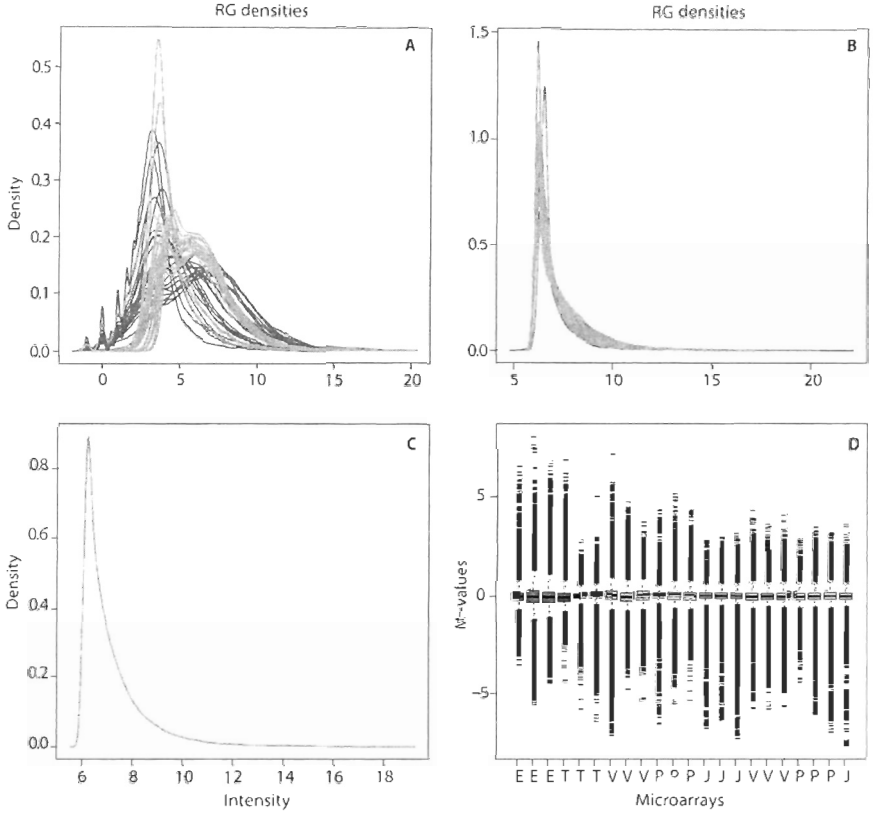

Figure 4.12: **Quality assessment of the preprocessing of two-color microarray (MA) data.** (A) Before background correction, (B) after background correction before loess within array normalization, and (C) after within and between array normalization. (D) Boxplots of the M-values of all MAs used in this work of Co-reared (cocktail diet) and Tiso-reared (Tiso dit) larvae. E, egg; T, trochophore; V, veliger; P, pediveliger; J, juvenile. The last seven MAs (D) were hybridized with Tiso-reared total RNA larvae. All MAs were printed on 8x60k Agilent MAs with de novo HiSeq assembled transcriptome of the blue mussel *Mytilus edulis* [Bassim et al., 2014b]. MA can be acquired following the Gene Expression Omnibus (GEO) accession number GSE55580. Background correction was considered since missing values were encountered. Noise substraction was efficient in correcting negative or zero-corrected intensities. Correction was done on the M-values for each array without taking into account the nature (Co or Tiso) of the array. Normalization was done on the A-values which makes the red and green distributions the same across arrays (C). Overall boxplot assessment shows a somewhat similar spreads of M-values across MAs.

All arrays were normalized together, irrespective of the treatment and the larval stage using the `limma` package in R from Bioconductor [Smyth, 2004]. Normalization is dependent on the working datasets, which in our case proved often best when reducing the log2-ratios for very low intensity spots. Background correction method was included as a trade-off to avoid errors during quantile normalization, even though results elsewhere where optimized without it [Zahurak et al., 2007] or used with a minimal offset [Bassim et al., 2014b]. Microarray data are available on Gene Expression Omnibus (GEO) using the following accession number GSE55580.

Gene selection is a crucial step before any pattern classification and discovery as it reduces the dimensionality for the problem by eliminating gene profiles that are extraneous to the phenomenon under investigation. Hence, we used specialized rules for appropriate preprocessing categorization of our experiments [Smyth et al., 2005], which extended our analysis to include hypothesis testing for gene significance, all of which are based on an empirical Bayes moderated T-statistics that model linearly the gene expression and permit to cross-compare genes [Smyth, 2004].

Genes selected in Figure 4.13 were coerced using a false discovery rate and P-value cutoffs (P<0.05) adjusted for multiple testing using a Benjamini and Hochberg procedure [Benjamini and Hochberg, 1995]. The FDR were controlled at the gene level only in contrast to the gene and developmental stage levels. This gave us a greater coverage over genes that are experimentally significant across multiple stages. Hierarchical clustering was only employed on the final selected markers (29 transcripts) after data mining been carried out on the whole dataset. Logarithmic transcriptional data were centered and scaled before complete clustering of genes by Pearson complete correlation. Significance (cutoff $\leqslant$0.05 threshold) of clusters were retrieved after 2000 bootstrap repetitions. The hierarchical tree was cut at specific heights, showing colors corresponding to clusters, over the heatmap color bar. An enrichment analysis of Fisher's exact test was realized in order to integrate GO-terms with significant genes. Enrichment tests presented here (Tables B.1, B.2, and B.3) are

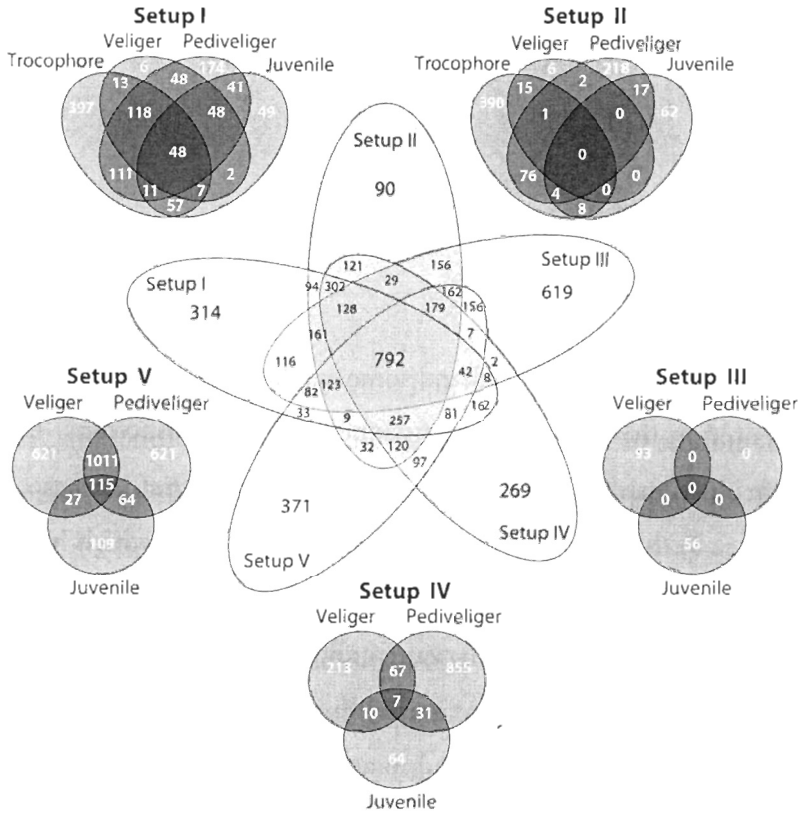

Figure 4.13: **Differentially expressed genes in response to five tested hypotheses.** Setups I and II described significant genes differentially expressed in Co- and Tiso reared larvae respectively. Genes that respond differently over time and between the two treatments simultaneously were represented in setup III. Setups IV and V correspond to selected genes, when the treatment or stage of development were considered as main effects respectively. The false discovery rate were controlled by Benjamini and Hochberg's adjustment for multiple testing (P<0.05) across contrasted stages for each hypothesis [Benjamini and Hochberg, 1995].

based on gene counts and methods accounting for the GO topology. Additional procedures were carried out following Bassim et al., [2014b] related work on the whole transcriptome. A principal components regression (PCR) was used to reduce the hypothesis space and to assess the differential gene expression due to treatment (cocktail, Co; Tiso, Tiso) and time effects from deriving the 29 genes of Table 4.4 from the entire dataset. The six differentially

oriented plots provided a cluster visualization of predicted samples using those 29 genes (Figure 4.10).

4.5.4 *Ensemble learning (EL)*

4.5.4.1 *Feature reduction*

Many of the genes are highly correlated and some can be associated to known mechanisms that are documented in the literature and assigned to known functions. Others are more obscure and difficult to explain. Consequently, we examined first the association of 2,000 high-variance genes to their expression levels and to each other, which helped select in an unsupervised approach (independently of the response outcome) a subset of 500 of these genes that can be used to explain the larval condition (Figure 4.4A). The selected feature set were sampled according to their maximized Euclidean distances or to their minimized Spearman pair-wise correlations [Ding and Peng, 2005]. An implementation of the MRMR algorithm from the mRMRe R library served for the unsupervised filtering process. A MMC was generated on the selected features before and after MRMR filtering (Figure 4.11). Inferential clustering was accomplished by MMC [Stone and Ayroles, 2009] and validated the hierarchical advantage of the selected features (http://mmc.gnets.ncsu.edu/). The MMC modules included in Table 4.4 were based on fitted coefficients of the 29 eigengenes with a differential expression modulated in relation to the combined response of the Co×Tiso diets and the five stages of development.

4.5.4.2 *Model selection*

The high-dimensional dataset ($p \gg n$) was shrunk using the lasso technique by considering a penalized likelihood that leads to a significant decrease in variance of the estimate at the expense of a slight bias [Tibshirani, 2011]. As model selection is our primary concern

here, minimizing the variance of the estimates leads to the best pruning of the hypothesis space. Model interpretability was increased by forcing some of the coefficient estimates of all the fitted features to be equal to zero, while optimizing the training criterion in an inner nested loop [Ambroise and McLachlan, 2002]. In general, we expect that among the features fitted in the lasso model there would be a non-random set of predictors with substantial coefficients that exhibits a relatively direct impact on the response variable.

Ten-fold nested cross-validation (CV) was applied to select an ℓ_1-regularized parameter (λ), which involved the smallest prediction error and can cope with over-fitting [Bishop, 1995; Cawley and Talbot, 2010] by controlling the bias-variance trade-off [Hastie et al., 2009]. The original dataset was split into training (2/3) and testing (1/3) datasets to calculate the classification error rate on an independent evaluation set. In general the test error is positively correlated to the increase in dimensionality, unless the added predictors are truly meaningful for the response. In this regard, RMSE measures were recorded for different parameter regularization before choosing the model with an optimized shrinkage. The selected model with all the available observations was re-fit and adjusted with the tuned parameters [Cawley and Talbot, 2010]. Prediction accuracy was estimated on the dedicated evaluation set (Figure 4.4B).

4.5.4.3 *Ensemble architecture*

A combination of different base learning algorithms were evaluated to construct an ensemble learning pipeline for hypothesis testing. The selected base learners (ANN, linear, polynomial, or radial kernels for SVM, boosted generalized additive models (GAM), boosted GLM, random forest (RF), regularized random forest (RRF), principal components regression (PCR), ridge regression (Ridge), partial least squares (PLS)) were tested through iterated runs. We first found the features that deliver predictions with the lowest RMSEs, combined with the highest accuracy estimates (see above). These steps were intended to help pick the model that deals best with overfitting and the selection bias, while disposing of noisy predictors

[Ambroise and McLachlan, 2002]. High-variable genes from the filtering process (Figure 4.4A) were used to select an independent evaluation set using several linear and nonlinear learners (Figure 4.4D and 4.5). Hyper-parameters were then tuned on the selected set of features using a nested 10-fold CV without repetition following a grid-search. Tuned hyper-parameters were then tested for optimization and their estimator performance was evaluated. The performance of the selected hyper-parameters was measured on a dedicated evaluation set (29 % of the total dataset) that was not used during the model selection step. Following a grid-tuning approach and a bagging (boostrap aggregating, [Breiman, 1996]) implementation of the ensemble (with repetition), learners were evaluated for performance using an external 10-fold CV, then their subresults were matched into a final evaluation, which was accomplished through averaging of the model estimate weights [Braga-Neto and Dougherty, 2004]. Overall, our ensemble model combines supervised classifiers (Figure 4.4D) covering different subdomains of our problem and complement each other for higher machine learning accuracy assumptions. All R libraries and the underlying engines of machine learning are part of several R packages (see Section B.4 for the R session information).

4.5.5 *Biometrical analytics*

Shell growth rates were measured with an image analysis software (Image-Pro v5.1 by Media Cybernetics) and an Olympus BX41 microscope. Fifty individuals were measured at each sampling time and for every dietary treatment. Survivorship was expressed as the total number of individuals less than the cumulative amount of empty shells and based on the first sampling timepoint at 1 DPF (N=3). Growth rates and survivorship are reported as means±SEM (or stated otherwise) and differences were considered significant at a P<0.05 and multivariate normality was tested using Shapiro-Wilk statistics. Homogeneity of variance-covariance matrices were graphically assessed using Brown-Forsyth method and further tested using a non-parametric Figner-Killeen test for the equality of variances. When necessary data were

arcsine square-root transformed to achieve homogeneity of variances [Fowler et al., 2013]. The effects between time (developmental stage) and diet (Co and Tiso) on larval growth and mortality were tested using one-way ANOVAs. When necessary additional weights were used to assess nonlinearity (quadratic terms) in the fitted models.

GLMs were used on continuous growth and survival data, taking into account the nonlinear effect between lipid data. The molar proportions (%) of the response variables were transformed making the data more symmetrically distributed. Next, data were fit to the standardized and centred predictors y^*("survival") and x^*("growth"). The length of the vectors corresponds to the increase of the mean of the variable originating from the center of the map. The coefficients of the model were estimated using maximum likelihood estimators. Biplot with a nonlinear calibration was included as higher order (interaction) effects were to be selected for further screening of gene co-regulation.

Multiple analyses of variances (MANOVAs) were applied to determine differences among FA between each of the five developmental stages of the blue mussel and as a function of the dietary treatment. Three separate two-way MANOVAs were performed on FA profiles which helped characterize more response variables than cases [Tabachnick et al., 2001]. The first MANOVA was dependent on the sum of saturated (ΣSFA), monounsaturated (ΣMUFA), and polyunsaturated (ΣPUFA) FAs as response variables. The second grouped essential FAs particularly EPA, DHA, $\Sigma\omega3$, ArA, and $\Sigma\omega6$. The third focused on ratios for EPA/ArA, DHA/EPA, and $\Sigma\omega3/\Sigma\omega6$. Subgroup differences of the main effect of time, treatment, and their interaction on lipid fractions were evaluated using a *post hoc* Tukey's multiple comparisons test. Different letters indicate statistically significant differences with *P*-values been adjusted with the Bonferroni method (Tables 4.5 and 4.2).

The analytical pipeline was built in R (v3.0.2) [R Core Team, 2013] with up-to-date packages. Reliability and accuracy were challenged with multiple published microarrays and custom-built benchmark datasets (not shown). Algorithm integrity can be reviewed with published

Table 4.5: **Fatty acid composition of microalgal species used in diets for treatment Co and Tiso.**

	Co	Tiso
FA composition (%)		
12:0	$0.4^{b}\pm0.3$	$0.9^{a}\pm0.1$
14:0	$6.3^{b}\pm2.5$	$11.7^{a}\pm2.2$
15:0	0.7 ± 0.2	0.9 ± 0.1
16:0	14.2 ± 4.7	17.1 ± 1.4
18:0	$1.8^{b}\pm0.2$	$2.8^{a}\pm0.1$
20:0	$1.2^{b}\pm0.2$	$2.0^{a}\pm0.2$
ΣSFA	24.5 ± 7.6	35.5 ± 0.9
16:1	$13.8^{b}\pm1.0$	$2.9^{a}\pm0.1$
17:1	$0.0^{b}\pm0.0$	$1.4^{a}\pm0.2$
18:1ω7	2.2 ± 0.3	3.1 ± 0.5
18:1ω9	$6.9^{b}\pm0.5$	$15.1^{a}\pm0.7$
20:1	$0.9^{b}\pm0.1$	$1.4^{a}\pm0.3$
22:1ω9	$0.9^{b}\pm0.1$	$1.5^{a}\pm0.1$
ΣMUFA	24.7 ± 2.0	25.5 ± 0.9
18:2ω6	$3.1^{b}\pm0.3$	$5.9^{a}\pm0.2$
18:3ω3	$1.9^{b}\pm0.2$	$4.5^{a}\pm1.1$
18:4ω3	$4.8^{b}\pm0.4$	$7.4^{a}\pm0.3$
20:4ω6 (ArA)	$4.8^{b}\pm0.4$	$0.9^{a}\pm0.1$
20:5ω3 (EPA)	$20.8^{b}\pm3.7$	$2.1^{a}\pm0.2$
22:6ω3 (DHA)	15.4 ± 2.2	18.3 ± 0.0
ΣPUFA	$50.8^{b}\pm5.9$	$39.0^{a}\pm0.9$
Σω3	$42.9^{b}\pm5.2$	$32.3^{a}\pm1.0$
Σω6	$7.9^{b}\pm0.7$	$6.7^{a}\pm0.2$
Σω3/Σω6	5.4 ± 0.3	4.8 ± 0.3
DHA/EPA	$0.8^{b}\pm0.1$	$8.9^{a}\pm0.9$
EPA/ArA	$4.4^{b}\pm0.5$	$2.5^{a}\pm0.2$
DHA+EPA	$36.2^{b}\pm5.5$	$20.4^{a}\pm0.2$
EPA+ArA	$25.6^{b}\pm4.0$	$2.9^{a}\pm0.2$
Total FA ($ng.\mu m^{-3}$)	73.1 ± 11.2	78.4 ± 13.9

All values represent the means±SD., N=3 larval tanks by treatment. Reported FAs contribute to 1 % or more of the ensemble. Different letters equate for significant differences P<0.05.

work cited for each of the used function (see Section B.4 for R session information). The R commands are available from corresponding author upon request.

4.6 DISCUSSION

This work helped us understand the association between genes and physiological performance, provided that we learn which genes play crucial roles in larval growth and mortality and how interrelated genes have coordinated expression given the stage of development. In this context, we constructed an ensemble method that can predict with a minimum of 87 % accuracy the stage of development and the physiological condition of larvae. The different classes that can be used to describe larval physiological states can be visualized in Figure 4.10. The ensemble is feature-dependent (29 gene markers reviewed in Table 4.4) and is classified into a number ($d \geqslant 2$) of separate larval conditions that fall either before or after 25 DPF. The features (p) represented the transcriptional expression of a very large number of genes, obtained from 22 microarrays of 22 larvae samples. Each microarray experiment gave the expression levels of the p genes of the dth stage.

Using a supervised feature elimination approach based on the squared ℓ_1 norm penalty term [Bishop, 1995], we yielded sparse regularized models that forces irrelevant genes out and as a result were easier to interpret (Figure 4.6). Internal and external CVs were used to compare the performance of the trained classifier [Ambroise and McLachlan, 2002]. However, in the case of high-dimensional data ($p \gg n$), CV can overfit. By reducing the number of features we did not need to include an exponentially increasing number of data points to increase our classifier's performance (Figure 4.5). Hence our sequential implementation of a filtering method gave us better control over this high-complexity data [Xing et al., 2001]. We incorporated furthermore an ensemble learning method by serially combining two classifiers (Figure 4.7), selected amongst several that were tested for performance and prediction accuracy (Figure 4.14).

The first model was an ANN with adjustable weight and bias that would adapt to our data. Multiple training runs were performed to evaluate the system. As ANNs are generally less sensitive to data deficiency, collinearity, and data fluctuation [Kotsiantis, 2007], pattern dis-

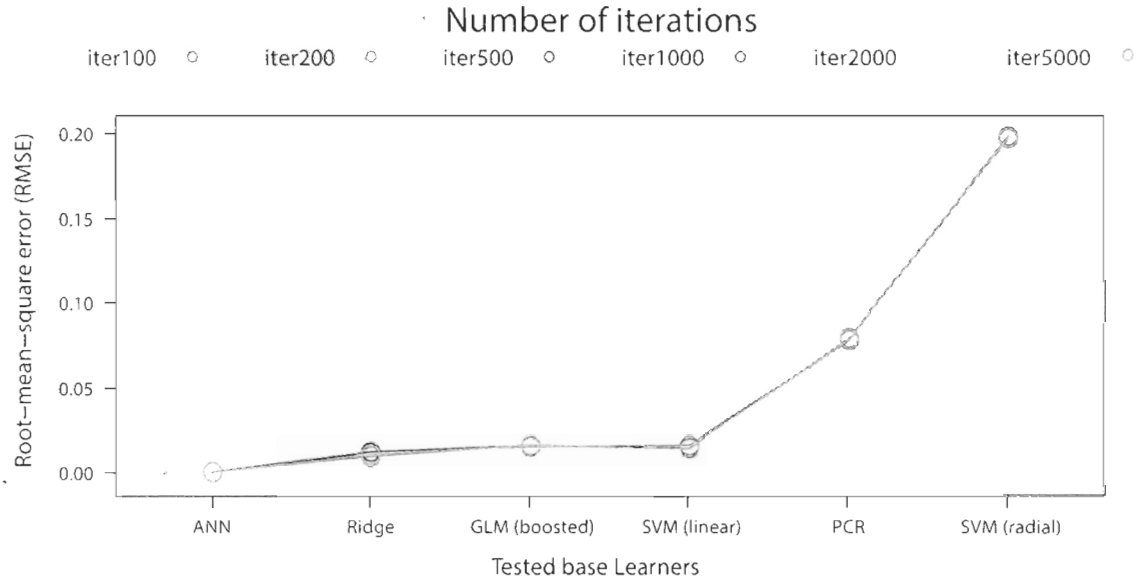

Figure 4.14: **Error rate estimates across selected base learners using bagging (bootstrap aggregating).** After the feature selection procedures, the model was trained with six different base learners for classification estimation then generalized on a dedicated holdout set for root-mean-square error (RMSE) estimation. Bagging was used under different iterations (100, 200, 500, 1K, 2K, 5K). ANN, multilayer perceptron artificial neural network; SVM, support vector machines; GLM, generalized linear models; PCR, principal components regression; Ridge, ridge regression; PLS, partial least squares. All functions were iterated using a workstation cluster for parallel computations in R.

covery given our small number of training samples yielded good estimations. We note that this is a general feature of machine learning tools including SVMs. A linear function with a boosting algorithm, GLM, inferred a significant portion of the response variance and as a result helped reduce the classification error rate of the ensemble (Figure 4.7).

GLM was used for its simplicity not for its prediction complexity [Meyer et al., 2003]. Inferential learning was our main concern [Hastie et al., 2005], coordinated regulation of the 29 markers helped therefore explore in small details the transcriptional effect of the diet treatment and the stage of development. These two models are specialized in different subdomains of our classification problem. They complemented each other in different instances. As far as GLM goes, reducing the high dimensional space of our data helped reduce compu-

tation time and with many converging outputs, ANNs induced a greater accuracy in the assumptions. The simplicity of these classifiers contributed in combination with each other to a better prediction estimates than the individual learners. However, it should be noted that the hyper-parameters of the ANN's were extensively tuned on every run by a grid search and the model was generalized by choosing the hyper-parameters that minimized the testing RMSE [Meyer et al., 2003].

With so many frameworks for pattern recognition and identification of differential gene expression, one might questions the reliability of the selected genes. In this regard, it is worth noting that genomic datasets are based on standardized profiling of gene expression [Barrett et al., 2009]. Our microarray datasets have been deposited at the GEO library (accession number GSE55580) and hence can be assessed freely for further analysis. Secondly, microarrays deliver high number of gene profiles that account for almost all of the transcriptome [Gracey and Cossins, 2003]. As a result, we are able to browse this snapshot of the transcriptome looking for multiple molecular mechanisms, which are mostly dependent on various clusters of co-regulated genes. These groups or eigengenes are not few, they can be interrelated or serially activated [Bassim et al., 2014b]. As a result, each classifier would find different sets of genes, which will be matched to its systematic complexities and the learning task characteristics [Kotsiantis, 2007]. Consequently, we opted for an ensemble learning framework by averaging the predictions of two complementary base learners to reduce the variance and take better control of the decisions [Krogh and Sollich, 1997]. Moreover, the number of samples n did not provide sufficient parametrization for choosing a single best classifier. The ensemble however, produced good approximations from combining the two differently specialized models.

Among the gene markers of Table 4.4, many were already been validated and were associated with stress response and development. Ganglioside GM2 activator (*GM2A*) for example, was induced in response to benzo[a]pyrene-treated adult mussels [Brown et al., 2006], involved in lipid metabolism in adult hypoxic oysters [Sussarellu et al., 2010], or nickel and

heat stressed mussels [Mohamed et al., 2014]. Two *GM2A* homologues were also been associated to hyperlipidosis syndrome of chemically exposed adult mussels [Dondero et al., 2011]. Moreover, *GM2As* were also oppositely differentially expressed in 17β-Estradiol adult treated mussels [Canesi et al., 2004]. Sarcoplasmic calcium-binding protein (*SCP*) was described only in invertebrates, predominantly in muscle and neurones [Hermann and Cox, 1995] or implicated in biomineralization processes [Shi et al., 2013]. It was also identified as an intracellular trafficking and signaling component and activated in response to heat stress in adult mussels [Truebano et al., 2010]. Permanent inactivation of protein nompA-homolog 2 (*NOAH-2*) in worm was lethal for larval transgenic animals [Tsang et al., 2007]. The mammalian ependymin-related protein 1 (*MERP1*) is associated with cell communication, membrane receptor activity, and immunity, and was up-regulated in response to hydrocarbon contamination in adult oysters [Boutet et al., 2004] or to bacterial infections in adult mussels [Ji et al., 2013a]. mammalian ependymin-related protein 1 (MERP1) is a conserved protein with biomineralization properties in bivalves [Miyamoto et al., 2013]. Apolipoprotein D (*APOD*) codes for a peptide derived from high density lipoproteins and is associated with inflammation, growth arrest, and several stress-induced conditions [Do Carmo et al., 2007]. The precursor of eicosanoid ArA can binds with high-affinity to the product of *apolipoprotein D (APOD)*, which can slow its release and prevent eicosanoids synthesis, after the activation of the immune system and membrane remodeling pathways [Rassart et al., 2000; Do Carmo et al., 2007]. A relative increase in membrane PUFAs was associated to a greater membrane fluidity, however with the presence of PUFAs it is expected that lipid peroxidation increases and causes irreversible damage [Hulbert et al., 2007]. In summary, our approach have identified many candidate gene markers that are also shown to be associated with developmental, stress-induced, lipid homeostasis conditions. Therefore, it will be interesting to extend these results with additional molecular experimentation to investigate the roles of these marker genes in growth and mortality during larval stages.

The molecular abundance (i.e., protein presence, enzyme activity, and turnover rate) of gene products was not assessed as we focused particularly on the fast cellular response to growth and mortality. Although our data can predict a change in the level of cell protein, expression kinetics can only estimate 87 % of the cell protein levels [Newman et al., 2006; Chapman et al., 2011]. However, it is only with transcriptome mining of fast-response differential expressions could we determine the topology of complex regulatory genetic networks. Then combined with physiological data that we could determine their biological impact on living cells. To our knowledge, our workflow combines both computational and biological evidence and presents a clear hypothesis about 29 marker genes that can be further tested experimentally.

4.7 ACKNOWLEDGEMENTS

The work presented in this paper was carried out during S. Bassim's Ph.D research at University of Quebec at Rimouski and was supported financially by a Strategic Grant from the Natural Sciences and Engineering Research Council of Canada (STPGP 364918) and funds from Ressources Aquatiques Québec (RAQ). The manuscript is contribution number (*will be updated*) to the Marine Resources Division of the South Carolina Department of Natural Resources.

Troisième partie

DISCUSSION ET CONCLUSION GÉNÉRALE

5

DISCUSSION ET PERSPECTIVES

La forme planctonique peut disposer de traits non avantageux qui rendent les larves susceptibles d'une contre-sélection au cours de l'histoire de vie des bivalves marins (Section 1.2.1 page 20). Toutefois, la dispersion larvaire augmente le niveau de la biodiversité intraspécifique et favorise à long terme l'apparition de nouvelles structures génétiques distinctes (Section 1.2.2 page 25). Pour ces raisons, j'ai considéré l'importance de la dynamique du génome et son impact fonctionnel sur la physiologie des premiers stades du développement d'un bivalve marin soit la moule bleue *M. edulis*. C'est dans cet aspect que j'introduis un lien entre la valeur sélective des réseaux génétiques et leurs répercussions directes sur la croissance de l'organisme et sa survie.

La période larvaire regroupe divers changements morphophysiologiques afin d'accommoder deux phases de vie (i) pélagique et (ii) benthique dans l'environnement marin (Figure 1.1 page 10). Une transition entre ces deux formes est corrélée à une restructuration tissulaire radicale de la larve primitive et sa transformation en une forme sexuée capable de se reproduire et transmettre ses gènes. Or, de fortes mortalités sont souvent observées pendant les premiers stades du développement larvaire en particulier durant la période péri-métamorphique (Figures 3.1 page 88 et 4.3 page 143). En effet, un compromis peut apparaître et favoriser une prolongation de la période pélagique qui augmente les chances des larves à trouver le meilleur substrat pour entrer en compétence métamorphique, quoique, la reproduction sera retardée (Figure 1.2 page 20). La portée principale de la présente thèse est donc de mettre en évidence l'effet de la disponibilité et de la qualité nutritionnelle sur la dynamique du transcriptome larvaire (Figure 1.4 page 41). Bien que les réserves énergétiques

acquises influencent l'adaptation locale des larves de *M. edulis*, les transitions morphophysiologiques sont généralement coordonnées par une plasticité génotypique de l'organisme. Pour cette raison, les hypothèses développées ont permis de caractériser l'impact de la régulation génique sur l'adaptation qui est observée en terme de croissance et de survie larvaire.

L'état physiologique des larves est variable en fonction des conditions biotiques locales, capable de compromettre le succès métamorphique et rendre l'organisme immunodéprimé susceptible de mortalité. Pour cette raison, la première partie (Section 2.4 page 56) du Chapitre 2 examine en détail les différents effecteurs immunitaires souvent utilisés comme marqueurs génétiques pendant le développement de la défense larvaire. Dans la seconde partie (Section 2.5 page 73) je considère le potentiel qui est encore peu exploré des études fonctionnelles du développement en biologie marine. Ainsi, les différents travaux qui suivent dans la présente thèse exploitent les différentes techniques utilisées en génomique fonctionnelle, mais aussi les effets de *nouveaux réseaux* de gènes sur le développement larvaire chez *M. edulis* (Section 2.3.6 page 55 et Figure 2.1 page 51). Une inférence interactive des réseaux géniques permet une meilleure visualisation de l'action coordonnée du génome sur la réponse cellulaire en réponse aux conditions environnantes.

Un deuxième axe de recherche est plutôt orienté sur les mécanismes moléculaires impliqués dans les processus biologiques du développement. Caractériser l'activité spatio-temporelle des fonctions moléculaires pendant l'ontogenèse larvaire permet d'explorer l'étendue de la relation entre la transition des formes larvaires et la valeur sélective inclusive des gènes. Pour cette raison le Chapitre 3 se centre sur la recherche et la reconnaissance des réseaux de régulation génique ou *genetic regulatory networks* (GRN). Ces GRN ont été caractérisés pendant le développement larvaire et ils sont spécifiques à la moule bleue, inférés à la base par l'analyse de la cinétique d'expression génique précoce de *M. edulis* (Section 3.5.2.1 page 93). Des effecteurs impliqués dans plusieurs voies de signalisation ont été ainsi caractérisés. Ils interagissent avec des régulateurs transcriptionnels, des phosphatases, des kinases et des remodeleurs de la chromatine, dépendemment du stade de développement (Figure

Tableau 5.1: **Conclusion de tous les résultats discutés et décrits dans les différents chapitres de cette thèse.**

	Description
Chapitre 2	**L'immunité pendant l'ontogenèse des bivalves : évaluer le potentiel des techniques de séquençage nouvelle génération** Une technique NGS couplée à une méthode classique en transcriptomie produise des données multidimensionnelles et de haute résolution.
Chapitre 3	**Identification des régulateurs génétiques au cours du développement larvaire de *Mytilus edulis*** (i) Le premier transcriptome des larves de *M. edulis* a été assemblé. (ii) Des oligoarrays sont désormais disponibles pour le suivi de la cinétique de l'expression génique pendant les différents stades de la transition larvaire. (iii) Le transcriptome de *M. edulis* comprend au moins 8,000 transcrits fonctionnels (mRNA polyA). (iv) Des régulateurs transcriptionnels avec une expression génique différentielle relative aux stades de développement ont été caractérisés. (v) Plusieurs réseaux de gènes impliqués dans le développement neuronal, musculaire et autres tissus interagissent ensemble et leur co-régulation varie en fonction du stade de développement.
Chapitre 4	**Impact génétique des acides gras sur la croissance et la mortalité pendant les premiers stades du développement larvaire chez *Mytilus edulis*** (i) La croissance et la mortalité chez les larves sont influencées par la qualité trophique. (ii) Les gènes à la base de la croissance et la mortalité chez les larves sont associés à la voie des précurseurs des eicosanoïdes. (iii) Les gènes impliqués dans la tolérance aux stress chez les larves constituent aussi des marqueurs géniques de la croissance et la mortalité. (iv) Une connexion potentielle existe entre les gènes de résistance aux stress et ceux de l'homéostasie lipidique.

3.8 page 108). Leur expression est corrélée à l'activation des processus biologiques de la croissance cellulaire et du développement neuronal, musculaire et des autres tissus.

Le milieu marin est riche en facteurs variables qui influencent la biologie de l'animal (Section 1.3 page 26), mais pendant le développement larvaire de *M. edulis* les acides gras essentiels (EFA) constituent une source indispensable pour maintenir l'homéostasie et améliorer la valeur sélective de l'organisme. Pour cette raison, le Chapitre 4 entreprend une approche multidisciplinaire pour associer l'effet des acides arachidonique (ArA) et eicosa-

pentaenoïque (EPA) sur la dynamique des génomes des larves de la moule bleue. L'ArA et l'EPA ont des effets sur la croissance, l'immunité et la résistance aux perturbations environnementales (Section 1.4 page 37). Compte tenu du stress engendré pendant la transition en juvénile, les larves ont une capacité réduite pour se nourrir après la dégradation du velum et la maturation des branchies [Cannuel et al., 2009], les réserves énergétiques endogènes leurs sont donc essentielles pour survivre. Nous avons utilisé des modèles de prédiction qui considèrent tout d'abord les signatures adaptatives retrouvées dans l'expression différentielle des gènes par stade de développement (Section 4.5.4 page 162) et par traitement nutritionnel afin d'extraire ensuite des décisions relatives aux effets de l'ArA et l'EPA sur les transitions morphogénétiques (Tableau 4.2 page 142). Plusieurs gènes candidats ont été caractérisés comme étant des cibles directes ou indirectes de certains EFA (Section 4.6 page 167 et Tableau 4.4 page 152). Ces gènes sont impliqués dans plusieurs mécanismes du métabolisme lipidique et dans les voies de résistances aux stress et peuvent être considérés comme marqueurs potentiels de la croissance et de la mortalité des larves.

5.1 L'ÉTUDE POLYGÉNIQUE : COMMENT PROCÉDER ?

Au cours des dernières années, l'intérêt pour la transcriptomie chez les bivalves a considérablement augmenté en raison de leur potentiel en tant que modèle d'évaluation environnementale (Figure 5.1 page ci-contre). Les études des premiers stades ontogéniques peuvent certainement être exploitées, parce que cette phase peu explorée de l'histoire de vie a montré un potentiel intéressant pour la découverte de marqueurs génétiques (chez les larves [Hens et al., 2006; Zapata et al., 2009; Andersen et al., 2009; Je Lee and Boulding, 2009; Genard et al., 2012, 2013] et chez l'adulte [Gueguen et al., 2003; Tanguy et al., 2008; Wang et al., 2009a; Perrigault et al., 2009; Tanguy et al., 2013a]).

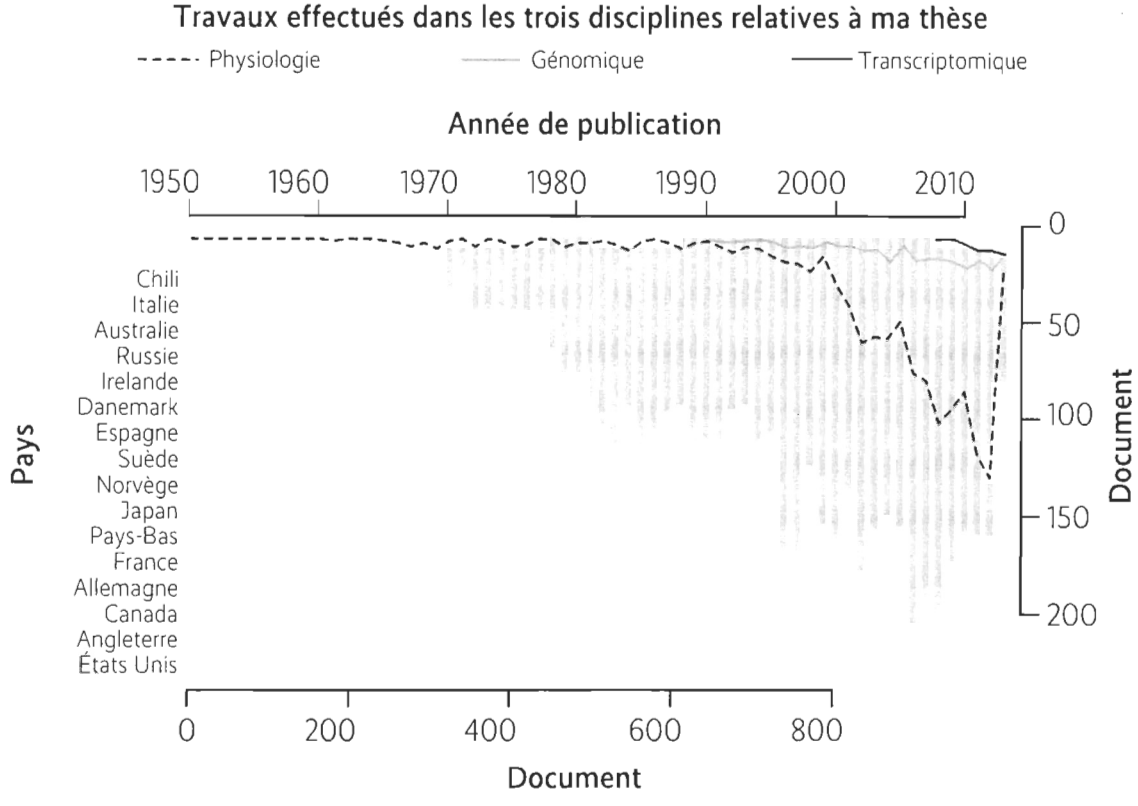

Figure 5.1: **Analyse exploratoire qui résume le nombre, la nature et l'origine des travaux publiés en relation avec la physiologie, la génomique et la transcriptomie des bivalves pendant les 63 dernières années.**

5.1.1 *Une approche transcriptomique : assemblage fonctionnel*

Une façon de procéder à l'assemblage des transcriptomes, que ce soit pour les techniques classiques ou de nouvelles générations, est d'utiliser les *données hétérogènes* des différents organismes dont les métadata sont disponibles dans les bases de séquences publiques et privées. Effectivement, des séquences basées sur des génomes de références ont permis d'améliorer la taille fonctionnelle de notre transcriptome de *M. edulis* après séquençage. Cependant, la structure fonctionnelle des métadata est parfois mal explorée (Section 3.5.4.4 page 113), malgré la disponibilité de données comparables entre les taxons clés. Les contrariétés créées par le manque d'une méthodologie sensible et standardisée pour les techniques

classiques entraînent une expansion non contrôlée de librairies parfois non utilisables (Section 2.5.5 page 80). Par exemple, après examen des séquences immunitaires publiques sans l'utilisation d'un *pipeline intégratif* (par exemple le pipeline analytique de la larve *M. edulis* Figure 3.2 page 96B et Section 3.5.2 page 93) pour l'assemblage du transcriptome chez deux espèces de palourde, les analyses transcriptomiques étaient redondantes en ce qui concerne l'implication immunitaire de NF-κB et sa voie de signalisation [Moreira et al., 2012b]. Il se peut que pendant le processus d'annotation d'une matrice nucléotidique inconnue, en utilisant les outils analytiques classiques, de faux positifs apparaissent (Section 5.1.2 page suivante). En effet, ce nombre élevé de faux positifs en raison des librairies mal assemblées et mal annotées entraînera un classement biaisé de cette matrice inconnue (Section 3.5.2.3 page 94) [Ekblom and Galindo, 2010; Chang et al., 2014]. Les méthodologies classiques ne permettent pas d'apprécier la *structure analytique intégrative* afin d'identifier les faux positifs introduits dans une librairie (Section 2.5.2 page 77), pour cette raison le transcriptome de *M. edulis* dans ce travail a été assemblé de novo par séquençage direct de l'ARN sans se baser sur des séquences hétérogènes. Une fois assemblé, le transcriptome a été amélioré par traitement comparatif avec des séquences hétérogènes. Par ailleurs, l'interrogation des banques pour une annotation de plusieurs milliers de séquences a besoin d'algorithmes rapides, des machines puissantes ainsi que des banques volumineuses pour stocker le flux intrinsèque de données. La rapidité des algorithmes est souvent corrélée avec la puissance des machines utilisées (Figure 2.3 page 75C), mais leur efficacité et leur sensibilité ne l'est pas souvent [Perkins et al., 2014]. Ainsi, la méthodologie NGS qui a été utilisée ici pendant l'assemblage fonctionnel nous a permis une exploitation paramétrisée efficace avec un contrôle progressif de la sensibilité d'indexation et de la puissance du calcul (Section 3.5.2.2 page 94 et Tableau A.1 page 205). Pour cette raison, s'assurer de la qualité des librairies de séquences doit être une priorité pour les nouveaux utilisateurs, en particulier les chercheurs qui travaillent sur les espèces non modèles.

Tableau 5.2: **Récapitulatif du nombre de larves et quantité de matrice en acide nucléique pour différents protocoles transcriptomique et génomique.**

Analyse moléculaire	Protocole	Nombre de larves*	ARN_T^{**}	ADN_g^{***}
		10,000–50,000†	500–5,000Δ	200–2,000
Contrôle qualité	Nanodrop		1	
État de l'ARN	Bioanalyzer		100–1,000	
Transcriptomie	RNA-seq		100•	
Oligoarrays	2 Channels		300○	
Expression génique	RT-qPCR		10-50‡	
Test d'espèce	PCR			50-100‡

† Nombre d'œufs ou trocophores (50,000) ; véligères ou pédivéligères (25,000) ; juvéniles (10,000).
Δ Concentration d'ARN_T initiale collectée après extraction.
‡ Concentration d'ADNc ou d'ADN par échantillon par gène.
• Concentration d'ARN par échantillon par stade de développement (N=10).
○ Concentration d'ARN par échantillon par stade de développement et par réplicat (N=23).

Note : Le nombre de larves est soumis aux conditions expérimentales conçues pour ce travail de thèse (Sections 3.5.1.1 page 91 et 4.5.1 page 156). Le nombre et les concentrations sont indiqués par échantillons (par stade de développement). RT-q(PCR), reverse-transcription quantitative polymerase chain reaction.
* Élevage d'œufs et des larves de trocophore, véligère, pédivéligère et juvénile. Plus la larve est grande en taille, moins de produits biologiques sera nécessaires.
** L'ARN total mesuré en ng/µl pour les œufs et les larves. Une concentration de 2 ng/larve est attendue suivant nos protocoles d'extraction.
D'après Cox [2004] la concentration d'ARN bactérien est de 20 fg/cellule ou 10 pg/cellule souche embryonnaire d'après Sasagawa [2013]. Donc nos analyses expérimentales sont basées sur les transcriptomes de 10×10^6–10^8 cellules.
*** L'ADN génomique mesuré en ng/µl pour du tissu adulte. D'après Cooper [1968] la concentration d'ADN (bactérien) est de 4 fg/cellule.

5.1.2 *Une approche génomique : annotation fonctionnelle*

L'ADN définit la structure génétique d'une population et permet d'établir une cartographie des bornes de l'étendue de la présence d'une espèce, ainsi nous pourrons approximer ses dérivés et mesurer avec exactitude sa biodiversité [St-Onge et al., 2013]. L'étude des séquences d'ADN transcrit (ARN) permet l'identification des changements précoces au niveau cellulaire, signatures d'une adaptation locale de l'espèce ou même de prédire son niveau de résistance contre une perturbation environnementale. L'expression différentielle des gènes de *M. edulis* a été mesurés par des puces oligonucléotidiques spécialement conçues pour ci-

bler chacun des stades larvaires. Ainsi, nous signalons que les oligopuces peuvent être couplées à une technique NGS pour délivrer un rendement suffisamment fiable pour l'analyse de la réponse cellulaire précoce. Le terme *ressource* ne fait pas seulement allusion au prix d'un séquençage ou de la production d'une base de données, mais cela implique également l'accès au matériel biologique (ARN et ADN) et la capacité de le conserver.La conservation est déterminante dans les travaux qui impliquent des cultures larvaires, lorsque l'échantillonnage et la conservation du matériel biologique sont sensibles aux variations techniques (Tableau 5.2 page précédente). Ainsi, nous pouvons observer que la RNA-seq (*vs* RT-qPCR) considère la gravité du problème et utilise une concentration d'ARN minimale pour effectuer le séquençage des larves de *M. edulis*. Le séquençage a été réalisé sur au moins 10×10^6 cellules provenant de larves à différentes structures morphophysiologiques (Tableau 5.2 page précédente). Ultimement, le défi pour le biologiste marin est d'adapter les outils conçus pour étudier des organismes modèles, pour lesquels un nombre significatif de séquences publiées existe déjà (Tableau C.1 page 224). Ensuite, il sera en mesure de développer un programme compatible avec la biologie des espèces non modèles.

Les études basées sur la production d'*expressed sequence tags* (EST) sont relativement indispensables pour le processus d'annotation des librairies nouvellement assemblées, afin d'identifier des marqueurs nucléaires (des microsatellites et single nucleotide polymorphisms, SNP) ou pour produire des puces polygéniques [Boutet et al., 2005; Fleury et al., 2009; Tanguy et al., 2008]. Les bases de séquences des espèces non modèles manquent encore d'informations fonctionnelles (Section 3.5.6 page 116). Chez *M. edulis*, la majorité des séquences sont désignées comme inconnues ou protéines putatives et cela réduit l'utilité de sa banque génomique pour l'ensemble des études fonctionnelles possibles (NCBI, *Taxonomy ID* :6550). Effectivement, les séquences non annotées (*NA*) observées après notre assemblage et annotation du transcriptome constituent environ 70 % de la librairie des 50,383 transcrits séquencés (Tableau 3.1 page 94). Le manque de métadata dans les banques de séquences interrogées pendant l'étape d'annotation du transcriptome de *M. edulis* (à date de l'année 2013) a réduit le

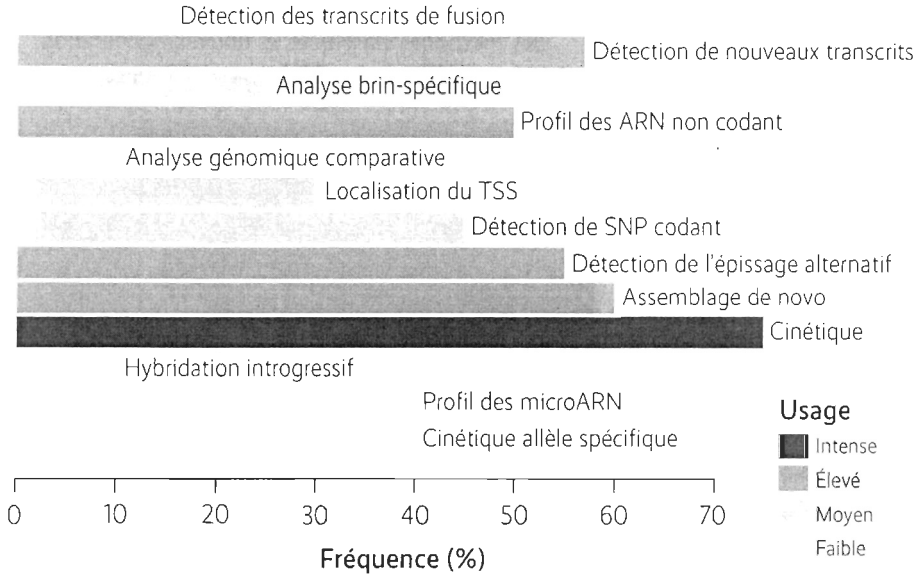

Figure 5.2: **La fréquence d'utilisation des différentes applications moléculaires par le moyen de la technologie de séquençage nouvelle génération (NGS).** Recensement conceptuel approximatif de plusieurs travaux génomiques et transcriptomiques publiés pour l'année 2012 . TSS, transcription-start site ; SNP, single nucleotide polymorphism.

rendement des résultats fonctionnels. Cependant, parmi les transcrits annotés, 13,978 (91.62 %) sont prédits comme étant des produits géniques à activité protéique (56.22 %) ou enzymatique (49.58 %). Par ailleurs, les analyses d'alignement de séquence ont permis d'effectuer des prédictions sur les transcris à la recherche d'une fonction moléculaire par localisation de domaines enzymatiques conservés (Section 3.5.2.3 page 94). Ainsi parmi les séquences *NA*, 25,923 (74.22 %) ont montré une fonctionnalité cellulaire putative (Tableau A.1 page 205). Or, la qualité de la librairie séquencée n'est pas entièrement influencée par ces limitantes puisque seulement 2.40 et 1.66 % des séquences *NA* présentent respectivement une activité protéique et enzymatique potentielle (Tableau A.1 page 205).

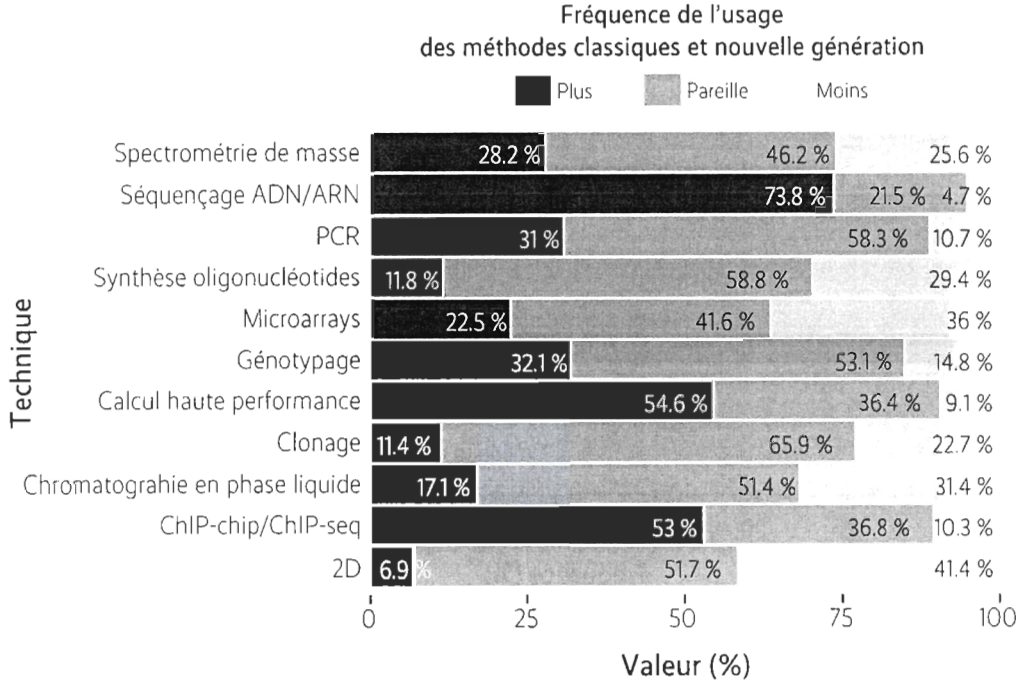

Figure 5.3: **La tendance d'utilisation des techniques moléculaires dans environ 150 laboratoires internationaux pendant l'année 2013.** Données adaptées du recensement effectué sur RNA-seq blog. L'ISMER et le LEMAR sont inclue.

5.1.3 *Une approche analytique : prédiction transcriptionnelle*

Avec la baisse des prix des méthodes de génomique fonctionnelle et l'expansion du marché international en biostatisticiens expérimentés dans la transcriptomie et autres disciplines relatives aux études omiques (Figure 2.3H page 75), les travaux sur les espèces modèles ont permis un enrichissement des bases de séquences. Par conséquent, l'intérêt de la recherche a récemment migré vers les espèces non modèles comme les bivalves (Tableau 2.1 page 76, Figure 2.3D page 75 et 5.4 page suivante).

Les analyses génomiques à grande échelle produisent des résultats qui peuvent couvrir le polymorphisme d'une base unique d'un gène jusqu'à la variation du nombre de copies de ses transcrits. Pour cette raison la technique NGS utilisée au cours de cette thèse a délivré un niveau élevé de couverture et d'interprétabilité. La résolution des données générées avec

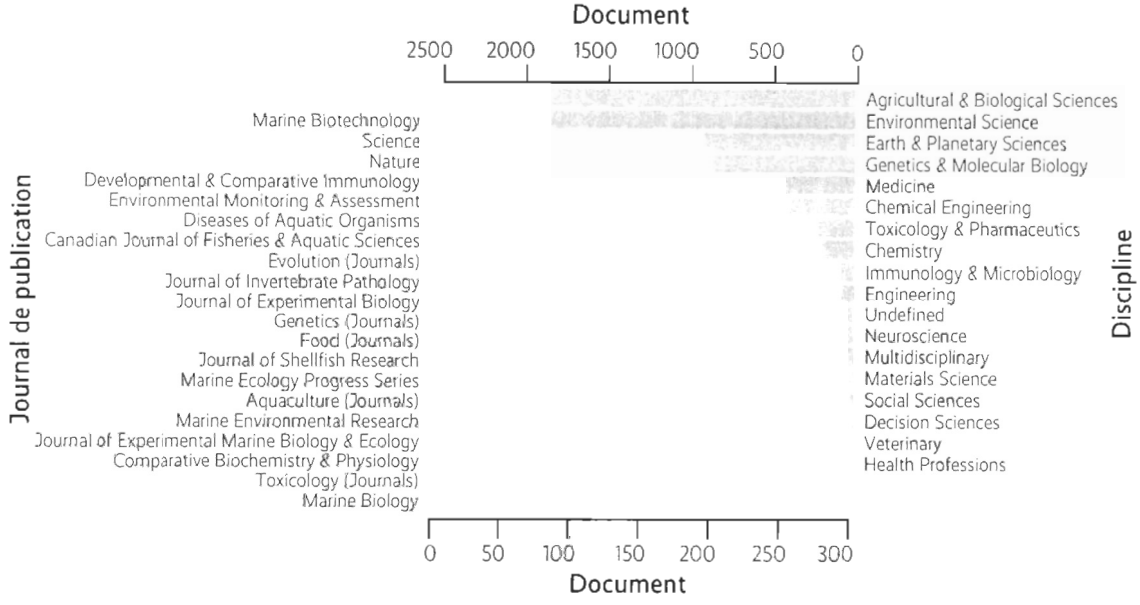

Figure 5.4: **Analyse qui résume le nombre des travaux publiés sur les bivalves en relation avec la discipline et le journal de publication pendant les 63 dernières années.**

nos *pipelines analytiques* va au-delà de l'approche réductionniste insuffisante qui est encore utilisée parfois avec les outils de première génération (Figure 5.3 page ci-contre). Ceci a amélioré nos tentatives d'intégration, de prédiction et de découverte des interactions entre nouvelles composantes fonctionnelles et qualitatives (Chapitre 4). Enfin, les assemblages des transcriptomes basés sur des techniques NGS sont quasi-standard, mais le potentiel est vaste si nous considérons la dynamique des génomes.

En considérant la coordination différentielle de la régulation génique, c'est-à-dire la recherche de groupes de gènes co-régulés (clusters corrélés Tableau A.6 page 213 et 4.4 page 152 ; Figures 4.9 et 4.11 page 158), est-ce que la perte d'information suite à une réduction de l'efficacité du processus d'annotation peut influencer l'authenticité et la certitude des résultats ? L'authenticité et la certitude des clusters inférés ne doivent pas être affectées par le nombre de transcrits NA. Pourtant, il faut absolument accorder du temps pour inspecter les paramètres d'assemblage et s'assurer que le processus suit la *règle de l'usage des codons* de l'organisme en question. Par exemple le pourcentage en GC (les nucléobases guanine,

G et cytosine, C) de *M. edulis* (Section 3.5.6 page 116) a été bien représenté pendant nos traitements de séquence (Tableau A.1 page 205). Ensuite, il faut se rappeler que les analyses transcriptomiques par définition sont basées sur l'ensemble des transcrits [Alon, 2007]. En d'autres termes, nous ne trouverons pas *un* seul cluster de gène représentatif d'une fonction moléculaire ou d'un processus biologique [Brazhnik et al., 2002; Baginsky et al., 2010; Zhang et al., 2012b]. Effectivement, à la recherche de GRN activés pendant le développement larvaire par 16 régulateurs transcriptionnels, nous avons trouvé 12 clusters de gènes qui présentent un différentiel d'expression corrélé à l'activation de processus moléculaires comme la croissance, la neurogenèse, la myogenèse et le développement d'autres tissus (Tableau A.6 page 213). En variant grosso modo les paramètres de l'interprétation des gènes candidats, nous sommes capables d'alterner la sélection des différents groupements de gènes (eigengenes) représentatifs d'une réponse physiologique [Brazhnik et al., 2002]. Le but est donc de trouver les meilleurs paramètres. En effet, nous avons réussi à corréler la corégulation de 29 gènes candidats impliqués dans des mécanismes variés en transduction du signal, maintien de l'intégrité de l'ADN, épigénétique, réplication de l'ADN et transport chez *M. edulis* (Tableaux B.1 page 219, B.2 et B.3). Ces gènes se regroupent sous 8 clusters qui permettent de prédire une interaction entre mécanismes de maintien de l'ADN, de défense et de l'homéostasie lipidique (Figure 4.9 page 153).

5.1.4 *Nouvelles directions en transcriptomie*

Tout d'abord, il faut noter que l'ARN est une structure moins stable que l'ADN et donc extraire de l'ARN de bonne qualité n'est pas une tâche simple. Récemment, des chercheurs ont standardisé une technique d'extraction de l'ARN adaptée pour au moins 300,000 espèces de plantes [Yockteng et al., 2014]. La standardisation ne permet pas seulement de simplifier une tâche, mais également favorise les analyses comparatives non biaisées. Le protocole utilisé au cours de la thèse pour l'extraction d'ARN a été optimisé pour les meilleurs ren-

dements après plusieurs tentatives d'adaptation des différents paramètres de manipulations des acides nucléiques (Section 3.5.1.2 page 92). De ce fait, une standardisation des techniques d'extraction de l'ARN larvaire fera l'objet d'une étude comparative intéressante pour les prochaines cinq années. Tandis qu'à long terme les techniques basées sur l'ARN unicellulaire prendront un point culminant en transcriptomie [Chi, 2014; Shapiro et al., 2013].

Concernant l'assemblage de novo des transcriptomes séquencés en particulier chez les espèces non modèles, notre base de données constitue un point de départ vers plusieurs avenues pour exploiter les interactions phénotype-génotype (Figure 5.2 page 183). La librairie regroupe un grand nombre de gènes codant et non codant qui couvrent la majorité du génome de la moule bleue *M. edulis*. Il est donc possible de l'utiliser pour améliorer l'annotation d'autres transcriptomes [Hornett and Wheat, 2012] ou pour effectuer de l'assemblage de novo d'espèces non modèles apparentées. Notre stratégie de travail a incorporé deux technologies qui génèrent chacune une quantité substantielle de données. Il nous a pris environ une année complète pour effectuer le séquençage, l'assemblage, l'annotation et la production des puces ADN. Nous sommes désormais en mesure de réduire cette période de moitié si d'autres applications sont considérées (Figure 5.2 page 183) puisque l'assemblage a déjà été réalisé. Toutefois, il n'est pas nécessaire de suivre la même méthodologie. Nous sommes capables d'effectuer une analyse différentielle de l'expression génique avec un séquençage par synthèse NGS, ce qui remplacera l'étape de production des puces. La durée n'excédera plus les trois mois. Ainsi, nous serons en mesure d'effectuer des analyses différentielles, en utilisant la même librairie, pour caractériser par exemple des éléments de régulation à épissage alternatif et leurs interactions [Wen et al., 2013], un phénomène probablement impliqué dans les processus d'adaptation et de spéciation.

L'identification de microRNA et de leurs gènes cibles sera aussi réalisable [Liu et al., 2014]. Introduire des ARN interférence (RNAi) dans *Caenorhabditis elegans* via l'alimentation a été une méthode efficace pour promouvoir une variation différentielle de l'expression génique chez la larve et l'adulte [Timmons et al., 2001]. Une fois reconnu et leur cible caracté-

risée chez *M. edulis*, ces ARN non codant peuvent servir de modèle élégant pour moduler les différents stades du développement larvaire en variant l'effet des gènes et l'interprétation des signaux de métamorphose (Figure 1.2 page 15) ou pour réduire la mortalité et favoriser une croissance rapide via l'influence de la nature du biofilm bactérien (Section 1.3.3.3 page 36), tout en ciblant par exemple avec de l'ARN double brin (dsRNA) un ou plusieurs gènes candidats du Tableau 4.4 page 152. Une façon pour procéder sera d'utiliser des biosphères lipidiques avec des précurseurs RNAi artificiels ou endogènes [De Fougerolles et al., 2013] qui seront délivrées directement à la larve en utilisant du triacylglycérol [Pernet et al., 2004]. D'autres techniques de délivrance des RNAi ont été déjà testé efficacement sur des bivalves marins, soit par injection directe dans l'organe cible de l'adulte [Fang et al., 2011], soit dans le milieu de culture des cellules primaires des larves ou l'adulte [You et al., 2012].

5.2 CROISSANCE ET MORTALITÉ

Les modèles de classement utilisés dans le Chapitre 4 ont servi à représenter les tendances physiologiques (croissance et mortalité larvaire) en fonction du temps et du traitement nutritionnel. Une non-linéarité est observée entre la croissance et l'interaction partagée du stade de développement et la composition du régime alimentaire en ArA et EPA. Par contre, le taux de mortalité a été affecté linéairement par la combinaison du stade de développement et la composition nutritionnelle en ArA et EPA (Section 4.4.1 page 139). Ces deux tendances de la croissance et de la mortalité ont été modélisées efficacement afin d'inférer la connexion entre la réponse génique et le phénotype observé (Figure 4.7 page 150). Les observations résumées dans le Tableau 4.2 page 142 montrent que les larves et les postlarves cultivées sous deux régimes nutritionnels différentiellement riches en ArA et EPA sont capables de résister légèrement aux fluctuations nutritionnelles durant la progression du développement. Cependant, l'ArA et l'EPA confèrent une résistance supplémentaire aux larves et la composition du régime alimentaire riche en ces précurseurs d'eicosanoïdes contribue à une réduction du taux de mortalité larvaire (Figure 4.2 page 140). Tout au long de la période

d'observation, l'ArA et l'EPA ont eu un effet direct sur la vitesse moyenne de la croissance de la larve périmétamorphique (Figure 4.3 page 143). L'effet de la qualité et la disponibilité de la nourriture sur le développement est connu [Pernet et al., 2005; Pernet and Tremblay, 2004; Pernet et al., 2006b; Gagné et al., 2010; Genard et al., 2011]. Effectivement, l'effet d'une carence en ArA et EPA sur les larves de la moule bleue a été aussi caractérisé sur l'expression différentielle des gènes.

5.2.1 *Nouvelles directions en biométrie*

Dans le Chapitre 3 nous avons considéré le développement larvaire afin de caractériser l'interaction génique et la coordination différentielle de la régulation. Les gènes sélectionnés sont impliqués dans les processus de croissance et de développement neuronal, musculaire et autres tissus. Toutefois, au Chapitre 4, la croissance et la mortalité larvaire ont été cette fois considérées comme variables physiologiques. La démarche qui aurait été peut-être plus satisfaisante pour complémenter les résultats de ces deux chapitres aurait été de choisir dès le début des gènes impliqués dans les processus de défense immunitaire et d'apoptose. En effet, ces deux mécanismes ont été considérés dans l'introduction (Section 3.4 page 87) du Chapitre 3. D'ailleurs, plusieurs gènes impliqués dans ces mécanismes ont été caractérisés et présentés dans les Tableaux A.2 page 208 et A.3 page 209. Ensuite, avec une approche d'apprentissage automatisée et paramétrisée, suivant la réponse physiologique larvaire sous les deux conditions de traitement nutritionnel, nous aurons pu exploiter les composants moléculaires responsables de la défense et la mort cellulaire pendant la métamorphose.

Pourquoi l'apoptose et la défense immunitaire n'ont pas été considérées dès le début de notre étude, sachant que les taux de mortalité et de croissance larvaire ont montré une significativité en fonction des concentrations en ArA et EPA (Section 4.4.1 page 139) pendant la phase périmétamorphique ? Pour répondre à cette question, il faut considérer la stratégie d'échantillonnage employée (Section 3.5.1.1 page 91 et 4.5.1). La récolte de larves

(Tableau 5.2 page 181) a été réalisée sur une période entre 0 à <40 jours. Cinq stades de développement ont produit cinq points d'échantillonnage. Toutefois, pour mieux cibler les mécanismes d'apoptose et de défense immunitaire au moment de la métamorphose larvaire, plusieurs points d'échantillonnage encadrant cette transition auront dû être effectués. Par conséquent, l'expression différentielle précoce aurait été enregistrée en profondeur et les interactions entre les gènes d'apoptose et d'immunité auraient été entièrement inférées sans biais. Pour cette raison, le fait d'avoir sélectionné dans le Chapitre 3 les voies moléculaires de croissance et du développement, nous étions en mesure d'inférer avec le minimum de biais des résultats non redondants qui décrivent toutes les transitions à partir de l'œuf, passant par la métamorphose jusqu'à la postlarve.

Ainsi, une nouvelle étude qui cible la phase spontanée de métamorphose pourrait être envisagée, puisque le pipeline analytique a été déjà mis en place (Figure 4.4 page 146) et confirmé par des résultats cohérents et corrélés avec des données déjà publiées. En considérant les différentes approches détaillées dans la Section 5.1.4 page 186 et dans la Figure 5.2 page 183, une nouvelle stratégie d'échantillonnage nous permettrait de cibler une phase complexe et cruciale du développement ontogénique. Au lieu d'étudier l'effet de l'ArA et l'EPA sur tous les stades du développement (comme ce qui a été fait durant ce travail de thèse), se pencher seulement sur la réponse métamorphique des larves et la dynamique de leur génome pendant cette période constitueraient une démarche ciblée et non biaisée pour discerner en profondeur l'effet des précurseurs des eicosanoïdes sur la mortalité. Par ailleurs, si la complexité de la technique NGS et du traitement des données post-séquençage devient difficile pour effectuer cette tâche plutôt intensive, des oligopuces basées sur les sondes des 15,457 transcrits déjà annotés (Tableau 3.1 page 94) peuvent délivrer en une période de deux mois des profils d'expression génique différentiel de larves périmétamorphiques. Suivant la stratégie précédente, l'activité des régulateurs transcriptionnels identifiés au Chapitre 3 peuvent être étudié en profondeur et leurs gènes cibles caractérisés par immunoprécipitation de la chromatine (ChIP) de la technique NGS (ChIP-seq) [Metzker, 2010].

5.3 MARQUEURS GÉNÉTIQUES

Il y a une décennie, nous croyions encore qu'un seul marqueur moléculaire pourrait être à l'origine d'une réponse cellulaire. Après avoir démêlé ici la complexité du système de développement et de la régulation génique, croyons-nous encore que la connectivité entre les gènes n'est pas un effecteur essentiel pour comprendre la dynamique des génomes ? L'organisme subit des modulations moléculaires en réponse à toute variation de l'homéostasie (Section 1.1.2 page 17). Restaurer l'équilibre cellulaire dépend entre autres de l'activation de plusieurs voies de signalisation et de l'expression de régulateurs transcriptionnels spécifiques essentiels dans la régulation génique ciblée (Tableau A.3 page 209). La réponse regroupe des effecteurs à activités catalytiques ou structurales (Figure 5.5 page suivante). La figure 5.5 illustre l'expression différentielle caractérisée pendant la recherche de gènes de régulation impliqués dans le développement chez les larves de *M. edulis*. La description de certains de ces gènes se retrouve dans le Tableau A.2 page 208. Il est possible d'observer une corrélation associée à une corégulation (Figure 3.7 page 109) qui illustre des interactions intergénique ou génophénotypique[1] potentielles. Par ailleurs, l'effet des précurseurs des eicosanoïdes sur la croissance et la mortalité des larves et postlarves a été associé à 29 marqueurs géniques avec des expressions différentielles corrélées entre gènes et stade de développement (Tableau 4.4 page 152). L'ensemble de ces résultats permet de conclure que la recherche *du gène* de croissance ou *du gène* de survie est loin d'être une approche fiable ou productive [Brazhnik et al., 2002; McKinney et al., 2006].

Compte tenu de la logique précédente, je conclus que seulement un nombre spécifique de gènes est activé ou inhibé en réponse à une variation homéostatique. Effectivement, les réseaux géniques à corégulation sélective forment un *modèle phénoménologique* qui pourrait expliquer l'interaction entre l'adaptation locale et la valeur sélective inclusive des gènes [Brazhnik et al., 2002]. La transcription génique est souvent sous un contrôle

1 *Génophénotype* : la connexion qui existe entre l'expression d'un gène et le trait phénotypique qui en résulte. Cette expression génique dépend de la séquence ADN qui peut varier au sein d'une même espèce.

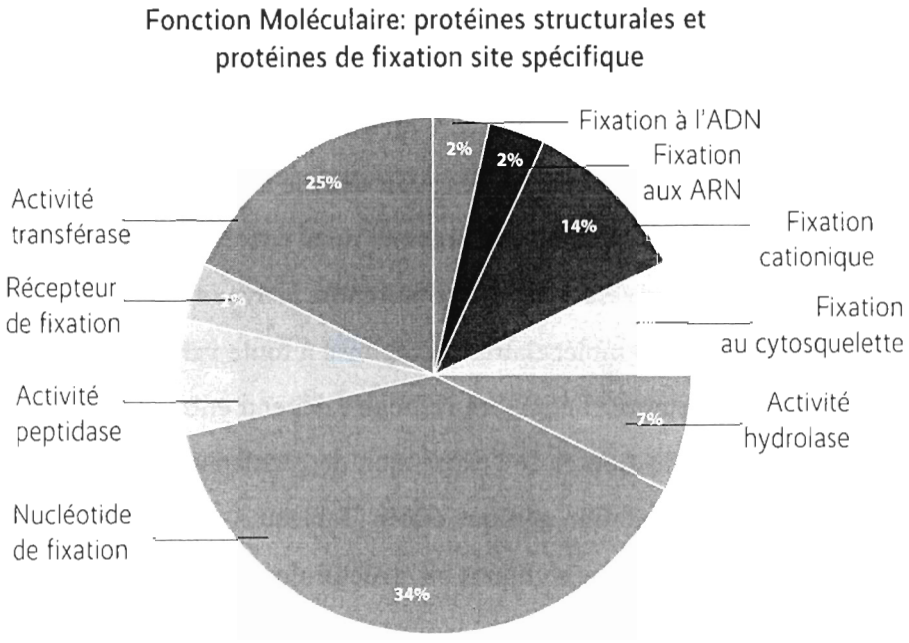

Figure 5.5: **Répartition de quelques fonctions moléculaires des gènes séquencés et annotés de *Mytilus edulis*.** Les Tableaux 3.1 page 94 et A.1 page 205 et la Figure 3.3 page 98 comportent les informations essentielles sur les bases de séquences, leur contenu fonctionnel et leur qualité.

rigoureux puisque c'est un processus complexe, dépendant de l'intervention de plusieurs composants cellulaires [Alon, 2007]. En effet, les gènes candidats caractérisés au Chapitre 4 possèdent un rôle clé dans différents mécanismes complémentaires des précurseurs des eicosanoïdes. Leur implication dans des processus de détoxication, métabolisme lipidique, biominéralisation, neurogenèse, mortalité, immunité et du cycle cellulaire justifie la présence d'une connexion entre les réseaux géniques à valeur sélective inclusive, déterminante de l'adaptation locale chez les larves. Si l'activation génique se produit en parallèle, l'interaction est corrélée, sinon elle sera complémentaire pour les gènes activés séquentiellement. Ces deux classes d'interactions sont suffisantes pour expliquer l'étendue de la régulation génique malgré le fait que le réseautage et la corégulation génique sont parfois denses en connexions complexes [Brazhnik et al., 2002]. Les gènes corrélés peuvent appartenir à une même voie de signalisation ou en gros, dans un même mécanisme de régulation [Chu et al.,

1998] (Section 3.6.3 page 121). Les gènes complémentaires sont par contre différentiellement impliqués dans de vastes voies de réponses (Figures 4.9 page 153 et 4.11 page 158B). Ceci pourrait expliquer pourquoi une variation d'un paramètre environnemental sera à la base d'une activation de plusieurs mécanismes moléculaires à la fois complémentaires et corrélés [Zhang et al., 2012a]. Effectivement, la recherche de GRN impliqués dans les processus biologiques du développement a montré un enrichissement de gènes responsables dans le maintien de l'homéostasie, la myogenèse, la neurogenèse et le métabolisme lipidique chez les larves (Section 3.6.3.2 page 123).

Les marqueurs génétiques regroupent des gènes avec une variable fonctionnelle et résultante. La variable fonctionnelle représente l'efficacité du produit génique à activer un mécanisme moléculaire. C'est-à-dire le poids de la contribution de l'interaction gène-protéine sur le mécanisme activé [Nitsch et al., 2009]. La variable résultante représente le facteur d'impact de ce mécanisme sur la réponse cellulaire en fonction de la cinétique du gène. Les gènes de la Figure 5.5 page précédente ont été retirés préférentiellement de la *Subset D* dans la Figure 3.2 page 96. Nous observons deux régulations complémentaires dans l'expression génique pendant les mêmes stades larvaires qui séparent les gènes en deux groupes et des régulations corrélées négative et positive de gènes impliqués dans différents mécanismes pour chacun des stades larvaires. Pour cette raison, je propose que l'identification de marqueurs génétiques soit basée principalement sur la variable fonctionnelle du marqueur et l'ensemble de ses interactions complémentaires et corrélées.

5.3.1 *Nouvelles directions en prédiction génique*

L'identification de variantes génétiques spécifiques à d'autres traits complexes que la croissance et la mortalité, est désormais faisable à partir de nos banques de transcrits séquencées, annotées et dont l'expression est enregistrée (Chapitre 3 et 4). Nos banques ne sont pas seulement de hautes dimensions (x>15,000 à 45,000 transcrits) mais également peuvent être

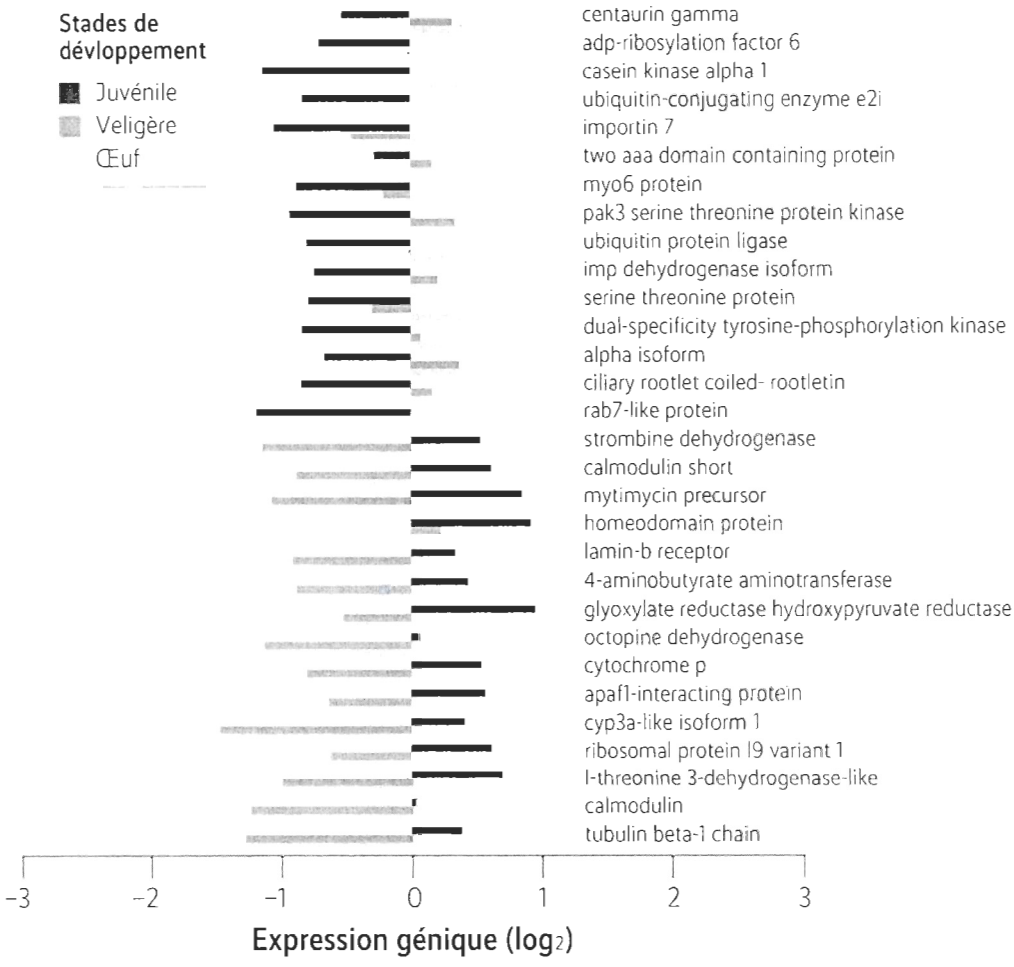

Figure 5.6: **Expression génique différentielle de quelques gènes dont le produit protéique représente une fonction moléculaire de fixation à l'ADN pendant l'ontogenèse de *Mytilus edulis*.**

utilisées pour inférer des problèmes multivariés. Pour cette raison, l'approche utilisée pour identifier les 29 gènes détaillés dans le Chapitre 4 est relative à l'interaction génique avec une variable binomiale qui regroupe le changement physiologique en réponse à un effet nutritionnel. D'ailleurs, si le nombre de larves échantillonnées n avait été plus élevé, par exemple si n=10 par stade (donc N=100 pour deux traitements nutritionnels), d'autres variables y auraient pu être ajoutées afin de retirer des décisions encore plus complexes. Notre implémentation des modèles *Machine Learning* dans le Chapitre 4 a pris en compte l'effet

de l'ArA et l'EPA pendant les périodes de croissance rapide et de mortalité élevée. Hors, avec un nombre n plus élevé il serait possible de caractériser l'effet de l'ArA et l'EPA sur chaque stade de développement, voire la métamorphose et la restructuration tissulaire en suivant leurs effets sur un intervalle t court, par exemple t=2 pour une période totale T=20 jours postfertilisation (JPF).

Par ailleurs, le nombre de gènes caractérisés dans notre étude est inféré à partir d'une prédiction physiologique basée sur seulement 29 gènes (Chapitre 4). Toutefois, certains chercheurs pensent que le nombre suffisant pour prédire un trait complexe doit être de l'ordre de 1,000 gènes, voire plus [Ding and Peng, 2005]. Certainement, plus le nombre de gènes est élevé plus les interactions gène-gène sont importantes, mais plus l'interprétabilité est réduite. Pour cette raison, le Chapitre 3 est basé sur des modèles Bayésiens empiriques qui considèrent à priori des interactions fonctionnelles suivant une expression différentielle corrélée d'un nombre élevé de gènes. Ainsi, d'après le pipeline analytique résumé dans la Figure 4.4 page 146, l'identification de variantes génétiques dépendant de traits complexes serait faisable si nous pouvions utiliser un nombre élevé de gènes (500 à 2,000 gènes).

En outre, une approche intéressante que nous n'avons pas encore utilisée sera une étude centrée sur les SNP. Typiquement, 50,383 transcrits comprennent plusieurs milliers de SNP. Certains d'entre eux ne sont pas représentatifs des traits analysés. Notre pipeline intégratif utilisé au Chapitre 4 dépend d'un modèle de filtration des gènes redondants et un deuxième qui élimine par pénalisation les gènes non significativement liés à la réponse y désirée. La même stratégie pourrait être utilisée pour identifier les SNP des larves *M. edulis*, une fois que les 50,383 transcrits auront été traités pour contre sélectionner par filtration les SNP non représentatifs de l'y choisi.

5.4 CONCLUSION

Le travail de thèse sur *Mytilus edulis* a été effectué pendant une durée de trois ans en plus de deux ans pour les analyses de données et la rédaction des papiers. L'approche que nous avons utilisée traite d'une façon différente l'effet de l'environnement sur la réponse génomique. Une stratégie multidisciplinaire nous a aidés à développer de nouvelles observations en testant plusieurs idées écologique, génomique et physiologique. Le stade larvaire de la moule bleue est susceptible à de fortes mortalités, quoique, un grand nombre de larves réussissent la transition métamorphique et développent des organes juvéniles. Nous avons démontré que les gènes impliqués dans la croissance et le développement pendant cette transition de la larve en juvénile sont régulés par des patrons génétiques conservés qui peuvent activer aussi des voies de défense et du métabolisme énergétique. Effectivement, le budget énergétique des larves est une déterminante de leur croissance et de leur survie et est variable en fonction du stade ontogénique et de la disponibilité nutritionnelle en acides gras essentiel. Les eicosanoïdes sont des hormones pro-inflammatoires synthétisées par les voies de certains acides gras essentiels. D'après les observations obtenues, la croissance et la mortalité larvaire sont influencées par la quantité des acides gras essentiels et leur disponibilité relative au stade ontogénique. La base de cette réponse phénotypique adaptative a une origine génétique. Les gènes impliqués dans le maintien de l'homéostasie et la tolérance aux stress ont été aussi corégulés en fonction de la quantité de ressources énergétiques en acides gras essentiels. Toutefois, ces résultats expliquent seulement une partie des mécanismes activés pendant la progression des transitions morphogénétiques. Nos analyses descriptives confirment la complexité des régulations géniques et la dimension de leur connectivité.

Je suis convaincu que la cause principale qui ralentit le progrès de la découverte scientifique est l'obligation d'une conformité de pensée. Afin qu'une théorie soit acceptée, elle doit être au début subjectivement persuasive. Mais je suis aussi certain que cela ne tiendra pas pour longtemps. Le progrès technologique ouvre de nouvelles avenues d'expérimentation.

Que ça soit pour tester des théories statistiques longtemps connues pour leur complexité, comme les modèles bayésiens en génétique et la sélection des génomes ou pour explorer l'étendue des interactions moléculaires des transcriptomes par le développement d'outils interactifs qui ciblent un ou plusieurs composants du génome. La complémentarité des tests basés sur les observations biologiques et les prédictions d'hypothèses offre un nouvel espoir. Effectivement, plusieurs études en plus de la nôtre ont entrepris une analyse à grande échelle pour la recherche des marqueurs génétiques de l'adaptation larvaire, non basée sur des hypothèses subjectives. Un jour, il nous sera possible d'explorer tous les mystères cachés derrière la complexité de l'évolution des génomes.

Quatrième partie

ANNEXE

A

SUPPLEMENTAL INFORMATION I

Identification of *Mytilus edulis* genetic regulators during early development

GENETIC REGULATORY NETWORKS (GRNS)

General overview

The arrow direction (edge) indicates the interactive effect (outward connection) of a regulator (TF) on a child-transcript (G) and the edge itself constitutes a connection between two genes (inward connection). Each transcript (node) can be translated into one protein/enzyme (hence the assigned functional description). Further, the transcript may constitutes a part of a complex macromolecule construct, which means that one transcript might represent a structural (e.g., stabilizer) or functional (e.g., catalytic) part of a whole biomolecule. In this latter case, the described transcript should not be considered capable on its own, of an enzymatic activity. Blue edges represent a negative effect of the regulator (TF), or it can be viewed as a suppression of the gene(s) represented as the child-transcript. The opposite is true for red edges. Red edges represent an activation between two nodes. All of the 61 regulators (Table A.3) have been chosen after being biologically and statistically validated, then GO mapped in Library 1.

$$\text{ARN} \xrightarrow[\text{Filtrate}]{\text{Non random edges}} \text{Library 1} \xrightarrow[\text{Integrate}]{\left(\frac{\text{Clustering}}{\text{Enrichment}}\right)} \text{Library 2} \xrightarrow[\text{Superimpose}]{\text{Regulators}} \text{Networks}$$

Choosing regulators

Regulators have been chosen in a four step approach:

1. All transcripts GO-mapped with a semantic similarity to at least one of the following GO-terms were selected amongst the 15,257 annotated top-hits; 803 regulators in total.
   * Transcription regulator (193 transcripts)
   * Phosphatase (419 transcripts)
   * Kinase (263 transcripts)
   * Chromatin remodeler (56 transcripts)
2. Only 186 regulators found in Library 1 were kept for further tests.
3. Amongst these regulators, 61 with at least one two-fold up-regulation at any time-point of the time series data were picked.
4. Subset D was used as a training set for the 61 regulators (Figure 3.2J cf. manuscript) in order to select the ones with the highest interaction coefficient (network not shown). Sixteen predicted regulators have been used for network inference at a 99.9% cutoff.

Regulatory networks $\xrightarrow[\text{Interactive(pathways+complexes)}]{\text{Network(structure+function)}}$ ***Interactive genetic modules***
Visualize

Networks summary

The following networks represent:

FIRST The inference between the 16 regulators (Table A.3) and the 188 DETs (Figure 3.5 and Table A.2) of subset D (Figure 3.10). This network shows the relationship between the regulators and statistically significant DETs ($P \leqslant 0.0007$ and 5.7% FDR). According to Figure 3.8 gene expression during all experimental conditions of subset D applies

to 3 expression patterns. These patterns imply the existence of gene-specific expression trends. This network (Figure 3.10) helps predict significant gene-gene interactive changes of expression profiles.

SECOND The dynamic of recurring interaction patterns between statistically significant ($P \leqslant 0.0007$ and 5.7% FDR) 188 genes of subset D (Figure 3.5 and Table A.2). This network (Figure A.1) emphasize the key hubs (top connected nodes) of subset D and shows overall integral feedback loops. Integral feedback loops are based on the connection of two genes. The increase of gene A leads to the production of gene B which causes gene A to drop to its basal levels. Not all of the 188 DETs are visualized since network inference was set to 99.9% significance threshold.

THIRD Recurring interaction patterns between the set of 61 regulators (Table A.3). This network (Figure 3.6) shows how the different regulators connect to each other following an integral feedback loop. Not all of the 61 regulators are visualized since network inference was set to 99.9% significance threshold.

FORTH Circuits composed of network motifs between the 16 regulators (Table A.3) and each of the 12 clusters of Library 1 (Figure 3.7 cf. manuscript). Each cluster interact differently to one or several regulator. This network (Figure 3.9) emphasize the synergistic expression trends from one time to the next of the candidate regulator and its target gene.

A.1 FIGURES

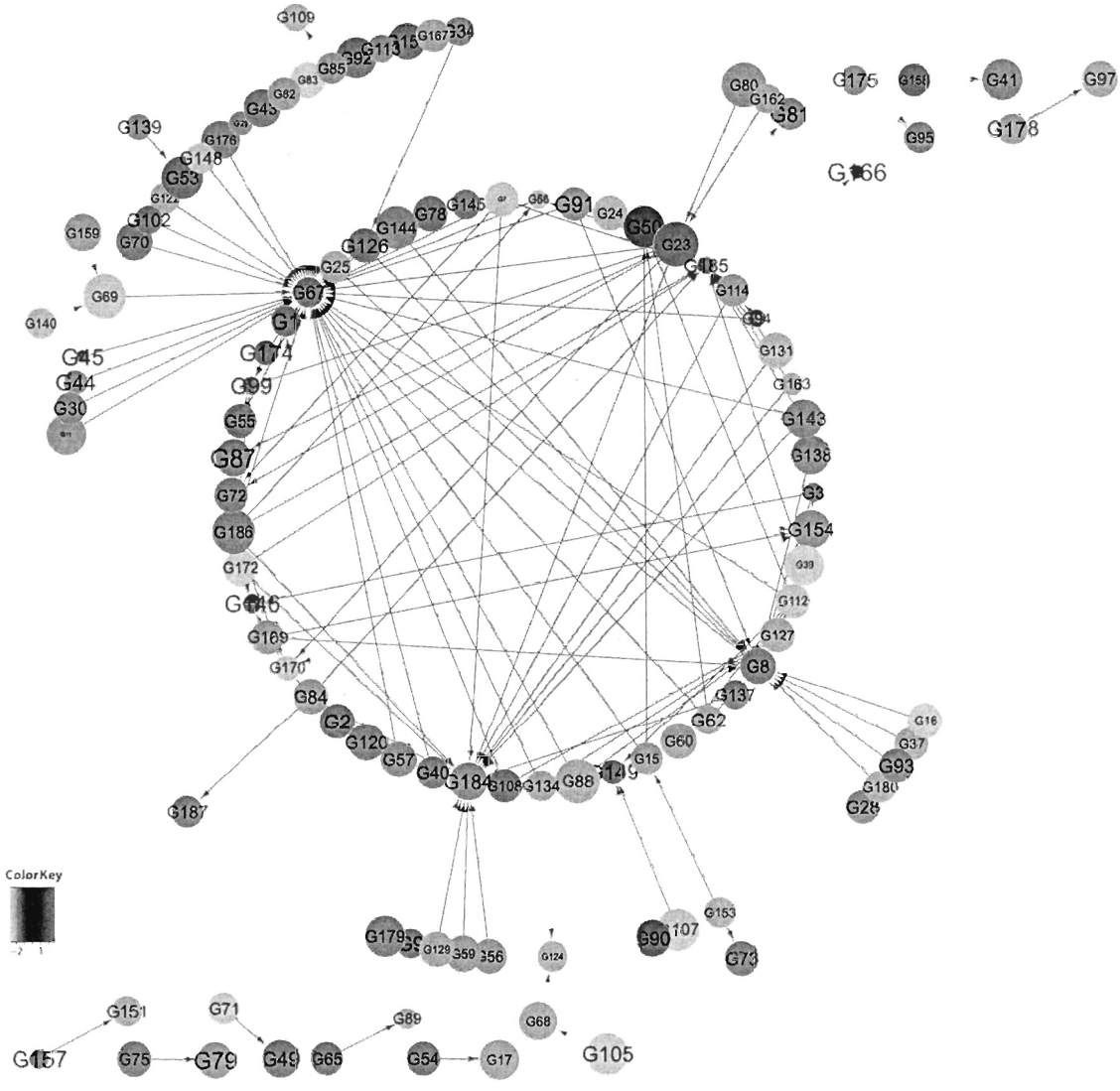

Figure A.1: **Differential genetic interaction between 188 significant genes of subset D expressed at the egg, pediveliger, and juvenile stages of development.** Nodes are identified by gene (G) numbered index. IDs, descriptions, and UniProt entries are found in Table A.2. Node color, label size, and node width represent the genes log expression ratios during the egg phase, pediveliger and juvenile stages respectively. Edge line type corresponds to either a negative (dotted line) or positive (solid line) interaction between two genes computed by Bayesian network fitting model. Edge width and color correspond to the betweenness coefficient of gene-to-gene interaction. The stronger an established interaction is, the bigger the outward connecting edges are.

A.2 TABLES

Table A.1: **MG-RAST analysis of assembled RNA-seq library**

	Metagenome overview			
	Annotated		Non-annotated	
Processed Material	**Raw***	**Post-QC****	**Raw***	**Post-QC****
Size (bp)	10,003,680	7,972,434	19,40,309	15,952,438
Sequence count	15,257	14,516	34,926	33,338
Mean sequence length (bp)	647±600	549±379	546±468	478±282
Mean GC %	38±4	38±4	30±4	30±4
Predicted protein features		13,978		25,923
Predicted rRNA features		1,254		7,505
Identified protein features		8,577		837
Identified rRNA features		35		0
Identified functional categories		7,564		581

* Prior to sequence quality control analysis.

** Quality control filtering

Table A.2: **Significant genes of Subset D at 5.7% FDR and $P \leqslant 0.0007$ as represented in Figure 3.10 and A.1**

Index	Locus	ID	Description	UniProt
G1	Locus10106	SLC25A15	solute carrier family 25 (mitochondrial carrier ornithine transporter) member 15	Q9Y619
G2	Locus1037	RPS13	40s ribosomal protein s13	P62277
G3	Locus11722	ELL	rna polymerase ii elongation factor ell	P55199
G4	Locus12061	MRC2	c-type mannose receptor 2	Q9UBG0
G5	Locus1216	RPS2	ribosomal protein s2	P15880
G6	Locus12283	alpi2	alkaline phosphatase	Q4QRJ9
G7	Locus12354	SCP	sarcoplasmic calcium-binding protein	P04570
G8	Locus12534		*Non-annotated*	
G9	Locus12710	PCBD2	pterin-4-alpha-carbinolamine dehydratase-like	Q9H0N5
G10	Locus12782	DNAJA1	dnaj homolog subfamily a member 1	P31689
G11	Locus1304	CAV1	caveolin-1	Q03135
G12	Locus13133	CTSL	cathepsin L	P07711
G13	Locus13263	CYS-PIN	cysteine protease inhibitor	Q03196
G14	Locus13431	ARL6IP4	adp-ribosylation factor-like protein 6-interacting protein 4	Q96PJ3
G15	Locus13523	A-ferritin	ferritin heavy chain oocyte isoform	Q7SXA6
G16	Locus13953	beg1	Endo-beta-1,3-1,4-glucanase	Q45691
G17	Locus1404	XBP1	x-box binding	P17861
G18	Locus14294	LAMP1	lysosome-associated membrane glycoprotein 1	P11279
G19	Locus15296	MGL-2	lysozyme 2	A5LHX1
G20	Locus15409	CSAD	cysteine sulfinic acid decarboxylase	Q9Y600
G21	Locus15921	cba1	zinc carboxypeptidase a 1-like	Q29NC4
G22	Locus15942	nhaX	stress response protein nhax	O07552
G23	Locus15973	SLC27A3	long-chain fatty acid transport protein 3	Q5K4L6
G24	Locus16026	CTSL	cathepsin L	P07711
G25	Locus16753	C1q3	c1q domain containing protein 1q3	F0V440
G26	Locus17297	FCER2	low affinity immunoglobulin epsilon fc receptor	P06734
G27	Locus17407	RBKS	ribokinase	Q9H477
G28	Locus17419	CALM	calmodulin	P62158
G29	Locus18119	CTSC	cathepsin c	P53634
G30	Locus18335	ADAMTSL3	adamts-like protein 3	P82987
G31	Locus18615	odh1	octopine dehydrogenase	Q9BHM6
G32	Locus18662	WDR3	wd repeat-containing protein 3	Q9UNX4
G33	Locus19549	RWDD4	rwd domain-containing protein 4	Q6NW29
G34	Locus1997	PRKAR2B	camp-dependent protein kinase regulatory subunit	P31323
G35	Locus20132	cre-gst-11	cre-gst-11 protein	E3LEU5
G36	Locus20407	C1QL4	complement c1q-like protein 4-like	Q86Z23
G37	Locus20647	HSPA12A	heat shock 70 kda protein 12a	O43301
G38	Locus21463	MCCA	methylcrotonoyl-CoA carboxylase subunit beta mitochondrial-like	Q96RQ3
G39	Locus2162	COL4A4	collagen alpha-4 chain	P53420
G40	Locus2170	RPL10	ribosomal protein l10	P27635
G41	Locus2223	RPS17L	40s ribosomal protein s17-like	P0CW22
G42	Locus22403	L3HYPDH	Trans-L-3-hydroxyproline dehydratase	Q96EM0
G43	Locus22719	CD207	c-type lectin domain family 4 member k-like	Q9UJ71
G44	Locus22745	CACNG5	voltage-dependent calcium channel gamma-5 subunit	Q9UF02
G45	Locus23275	CHDC2	calponin homology domain-containing protein 2	Q8N9S7
G46	Locus23293	RPL9	60s ribosomal protein l9	P32969
G47	Locus23315	vdg3	Developmentally regulated vdg3	Q5D218
G48	Locus2366	CDC42	Cell division control protein 42	P60953
G49	Locus2420	RPS15	ribosomal protein s15	P62841
G50	Locus24214	SDH	saccharopine dehydrogenase	K1PXH5
G51	Locus24647	ANKRD28	serine threonine-protein phosphatase 6 regulatory ankyrin repeat subunit a	O15084
G52	Locus2480	COL4A4	collagen alpha-4 chain	P53420
G53	Locus24874	DIEXF	digestive organ expansion factor homolog	Q68CQ4
G54	Locus2492	RPS3A	40s ribosomal protein s3a	P61247
G55	Locus25349	ZNF830	zinc finger protein 830-like	Q96NB3
G56	Locus25526	RPLP0	60s acidic ribosomal protein p0	P05388
G57	Locus25673	RPL7	ribosomal protein l7	P18124
G58	Locus25842	GAPDH	glyceraldehyde-3-phosphate dehydrogenase	P04406
G59	Locus26014		perlucin 6	G5CT94
G60	Locus26253	CHIA	acidic mammalian chitinase-like	Q9BZP6
G61	Locus26303	TEP	thioester-containing protein	Q9GYW4
G62	Locus26362	GALM	aldose 1-epimerase	Q96C23
G63	Locus2642	algL	alginate lyase	Q9L7P2
G64	Locus26434	beg1	Endo-beta-1,3-1,4-glucanase	Q45691
G65	Locus26643	MPP	mitochondrial-processing peptidase subunit alpha	P29677
G66	Locus2668	CTSL	cathepsin L1 partial	P07711
G67	Locus26759	PSPH	phosphoserine phosphatase	P78330
G68	Locus26988	ACAN	aggrecan core protein	E7ENV9
G69	Locus27	FABP7	fatty acid-binding brain	O15540
G70	Locus27119	SLC27A2	very long-chain acyl-CoA synthetase-like	O14975
G71	Locus27180	lysg2	goose-type lysozyme 2	Q7YXC2
G72	Locus27306	DDX1	ATP-dependent RNA helicase ddx1	Q92499
G73	Locus27390	PSMF1	proteasome inhibitor pi31 subunit	Q92530
G74	Locus27877	lysg2	goose-type lysozyme 2	Q7YXC2

G75	Locus27908	dsbA	disulfide bond formation protein a	P0A6M2
G76	Locus28316	BYSL	bystin	Q13895
G77	Locus2838	COL4A4	collagen alpha-4 chain	P53420
G78	Locus28623	NEF-sp	rna exonuclease nef-sp	Q96IC2
G79	Locus28886	CHRNB4	neuronal acetylcholine receptor subunit beta-4	P30926
G80	Locus29241	GAT2	sodium and chloride-dependent gaba transporter 2-like	A6MKG6
G81	Locus29333	CYB561D2	cytochrome b561 domain-containing protein 2-like	O14569
G82	Locus29337	CTSB	cathepsin b	P07858
G83	Locus29344	Cvsi2	serine protease inhibitor cvsi-2	Q30HU9
G84	Locus29837	Plcd1	1-phosphatidylinositol-4 5-bisphosphate phosphodiesterase delta-1	P10688
G85	Locus30008	PHYH	phytanoyl-CoA dioxygenase	O14832
G86	Locus30092	GMDS	gdp-mannose 4,6 dehydratase	O60547
G87	Locus30140	nAChR	nicotinic acetylcholine receptor subunit type e	Q2XWK8
G88	Locus30281	CALM1	calmodulin 1	P62158
G89	Locus30702		*Non-annotated*	
G90	Locus30875	CASP2	caspase-2 isoform x1	P42575
G91	Locus30917	ARRDC3	arrestin domain-containing protein 3	Q96B67
G92	Locus32011	coll4	short-chain collagen c4	P18503
G93	Locus32298	METTL24	methyltransferase-like protein 24-like	Q5JXM2
G94	Locus32765	PIWIL1	piwi-like protein 1	Q96J94
G95	Locus329	RPL27	60s ribosomal protein l27	P61353
G96	Locus32918	2OG	2og-fe oxygenase	D5VA68
G97	Locus33422		perlucin 6	P82596
G98	Locus33502	tinc	protein tincar-like isoform x1	Q86B91
G99	Locus33600	SEC23IP	sec23-interacting protein	Q9Y6Y8
G100	Locus34340		*Non-annotated*	
G101	Locus34610	gusB	lysosomal beta-glucuronidase partial	P08236
G102	Locus34647	MNN9	mannan polymerase complex subunit mnn9	P53697
G103	Locus34664	nrf-6	nose resistant to fluoxetine protein 6-like	Q09225
G104	Locus34982	eIF-3	eukaryotic translation initiation factor 3 subunit h-like	Q9Y262
G105	Locus35094	LRP	low-density lipoprotein receptor-related	Q07954
G106	Locus35096	CCOAOMT	caffeoyl-CoA o-methyltransferase	Q9XGD6
G107	Locus35329	COL4A4	collagen alpha-4 chain	P53420
G108	Locus3535	RPL14	60s ribosomal protein l14	P50914
G109	Locus3539	FCGBP	IgGFc-binding protein	Q9Y6R7
G110	Locus3552		Fibrinolytic chymotrypsin-like serine proteinase	
G111	Locus35558		*Non-annotated*	
G112	Locus35816	GH	glycoside hydrolase	
G113	Locus35855	AGA	N(4)-(Beta-N-acetylglucosaminyl)-L-asparaginase	P20933
G114	Locus35893	FIBCD1	fibrinogen c domain-containing protein 1	Q8N539
G115	Locus36134	tdh	l-threonine 3-dehydrogenase	A0Q5K3
G116	Locus36178	CYGB	cytoglobin	Q8WWM9
G117	Locus36267		perlucin 1	P82596
G118	Locus36280	RPL36A	ribosomal protein l36a	P83881
G119	Locus36465	FCGBP	IgGFc-binding protein	Q9Y6R7
G120	Locus36511	PHGDH	3-phosphoglycerate dehydrogenase	O43175
G121	Locus36616	Glu I	beta-1,3-glucanase	Q7XJ83
G122	Locus369	SELENBP1	selenium-binding protein 1-like	Q13228
G123	Locus37649	GRHPR	glyoxylate reductase/hydroxypyruvate reductase	Q9UBQ7
G124	Locus37729	C1QL3	complement c1q-like protein 3	Q5VWW1
G125	Locus3774	CTRB1	chymotrypsinogen b-like	P17538
G126	Locus3791		*Non-annotated*	
G127	Locus38005	Cpb1	zinc carboxypeptidase B-like	P19223
G128	Locus38102	LRP2	low-density lipoprotein receptor-related protein 2	P98164
G129	Locus38117		*Non-annotated*	
G130	Locus3843	CDCA4	cell division cycle-associated protein 4	Q9BXL8
G131	Locus38733	SIGLEC	sialic acid binding lectin	Q9Y286
G132	Locus39936	WRAP53	telomerase cajal body protein 1	Q9BUR4
G133	Locus41179	COL1A2	collagen alpha-2 chain	P08123
G134	Locus41189	c1q3	c1q domain containing protein 1q3	F0V440
G135	Locus41393		*Non-annotated*	
G136	Locus42044	unc-9	innexin unc-9	O01393
G137	Locus42313	pol	endonuclease-reverse transcriptase	Q9N9Z1
G138	Locus42636	AANAT	serotonin n-acetyltransferase	Q16613
G139	Locus42729	nhaX	stress response protein nhax	O07552
G140	Locus4376	vdg3	developmentally-regulated vdg3	Q5D218
G141	Locus4443	nrf-6	nose resistant to fluoxetine protein 6	Q09225
G142	Locus4691	PSAP	proactivator polypeptide	P07602
G143	Locus4844	RPL3	60s ribosomal protein l3	P39023
G144	Locus5075	RPS15A	40s ribosomal protein s15a-like	P62244
G145	Locus5213	RPL18	60s ribosomal protein l18	Q07020
G146	Locus5244	DNAH7	dynein heavy chain 7 axonemal	Q8WXX0
G147	Locus5302	NACA	nascent polypeptide-associated complex subunit alpha	Q13765
G148	Locus554	ODZ3	Odd oz ten-m homolog 3	G3CAS9
G149	Locus5616	COL4A3BP	collagen type iv alpha-3-binding protein	Q9Y5P4
G150	Locus600	RPL18A	60s ribosomal protein l18a	Q02543
G151	Locus6087	PLCD3	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-3	Q8N3E9
G152	Locus6568		temptin	B6RB31
G153	Locus7	MAN2B1	lysosomal alpha-mannosidase	O00754
G154	Locus7199	XBP1	x-box binding	P17861

G155	Locus7577	TRHDE	thyrotropin-releasing hormone degrading enzyme-like	Q9UKU6
G156	Locus7928	CETN1	centrin-1	Q12798
G157	Locus7976	SAMD8	sphingomyelin synthase-related 1-like	Q96LT4
G158	Locus7990	LGALS8	lectin galactose binding soluble 8-like	O00214
G159	Locus8036	DAO	d-amino acid oxidase	P14920
G160	Locus825	INTS9	integrator complex subunit 9	Q9NV88
G161	Locus8548	RERG	ras-related and estrogen-regulated growth inhibitor	Q96A58
G162	Locus8705	CTSB	cathepsin b	P07858
G163	Locus9323	SBSN	suprabasin isoform 2	Q6UWP8
G164	Locus9414	RPL22L1	60s ribosomal protein l22-like 1	Q6P5R6
G165	Locus9568	CLEC4F	c-type lectin domain family 4 member f	Q8N1N0
G166	NODE_117052	pol	endonuclease-reverse transcriptase	Q9N9Z1
G167	NODE_126300	norA	norsolorinic acid reductase-like isoform x2	Q00049
G168	NODE_150381		*Non-annotated*	
G169	NODE_153482		cupin family protein	
G170	NODE_163118	lysg2	goose-type lysozyme 2	Q7YXC2
G171	NODE_175269	CYP4F22	cytochrome p450 4f22	Q6NT55
G172	NODE_18688	GH	glycoside hydrolase	
G173	NODE_189264	NOS	nitric oxide synthase	P29474
G174	NODE_205772	NPHP4	nephrocystin-4	O75161
G175	NODE_219422		*Non-annotated*	
G176	NODE_258157	apelB	apextrin-like protein	G0YL86
G177	NODE_261079		fad binding domain-containing protein	Q22YW5
G178	NODE_317912	CHRDL2	chordin-like protein 2-like	Q6WN34
G179	NODE_321421	RpL34a	ribosomal protein l34a	C1LA13
G180	NODE_33188		Serine proteinase	
G181	NODE_348568		perlucin 6	B6RAZ3
G182	NODE_3800	RPS20	40s ribosomal protein s20	P60866
G183	NODE_401906	GPX1	glutathione peroxidase 1	P07203
G184	NODE_438658	RPL9	60s ribosomal protein L9	P32969
G185	NODE_445914	ATAD2B	atpase family aaa domain-containing protein 2b-like	Q9UL10
G186	NODE_4820	Rpl34	60s ribosomal protein l34-like isoform 1	B2RZD4
G187	NODE_9059	RPL44	60s ribosomal protein l44	Q96499
G188	NODE_95951	BDH2	3-hydroxybutyrate dehydrogenase	Q9BUT1

Note: NAs are introduced in a third round of annotation (protein sequence similarity search) whenever top hits depart from the required significant *E*-value threshold, or present low-shared homology and no functional annotation is found (e.g., hypothetical or uncharacterized proteins). Their score might be high but this is due to high number of regions with low-complexity nucleotide composition.

Table A.3: **Regulators of transcription or translation extracted from Library 1 and interactively visualized in Figure 3.6**

Index	Locus	ID	Description	UniProt	GO
TF1	Locus10228	RIOK2	serine threonine-protein kinase rio2	Q9BVS4	K
TF2	Locus11050	src1-PTK	src-family protein tyrosine kinase		K
TF3	Locus1154	SMARCD1	swi/snf-related matrix-associated actin-dependent regulator of chromatin sub-family d member 1	Q96GM5	T
TF4	Locus11591	ALPL	alkaline phosphatase tissue-nonspecific isozyme	P05186	P
TF5	Locus1293	CnVAS1	vasa-like protein	Q6TEC0	P
TF6	Locus12973	CAMKK2	calcium calmodulin-dependent protein kinase kinase 2	Q96RR4	K
TF7	Locus14026	TUBG1	tubulin gamma-1 chain	P23258	P
TF8	Locus14184	KIAA0101	pcna-associated factor	Q15004	C
TF9	Locus1420	CMPK1	ump/cmp kinase-like	Q69SQ0	K
TF10	Locus1483	INSIG2	insulin-induced gene 2 protein	Q9Y5U4	T
TF11	Locus14986	H-RAS	v-ha-ras harvey rat sarcoma viral oncogene-like protein	P5ANT1	P
TF12	Locus15776	MCM3	dna replication licensing factor mcm3	P25205	C P
TF13	Locus16380	trc	serine threonine-protein kinase tricorner-like	Q9NBK5	K
TF14	Locus17608	CDC2	Cyclin dependent kinase 1	P06493	K
TF15	Locus18282	ARF6	adp-ribosylation factor 6	P62330	P
TF16	Locus18945	CYTH1	cytohesin-1-like isoform 2	Q15438	P
TF17	Locus20926	PIP5K1A	phosphatidylinositol 4-phosphate 5-kinase type-1 alpha	O99755	K
TF18	Locus21041	CAD	cad protein	P27708	K
TF19	Locus21779	PTPN13	tyrosine-protein phosphatase non-receptor type 13	Q12923	P
TF20	Locus218	ABCF2	atp-binding cassette sub-family f member 2	Q9UG63	P
TF21	Locus23557	GINS3	dna replication complex gins protein psf3	Q9BRX5	C
TF22	Locus2366	CDC42	Cell division control protein 42	P60953	P
TF23	Locus24964	ARHGEF4	rho guanine nucleotide exchange factor 4	Q9NR80	P
TF24	Locus25746	Gatad1	gata zinc finger domain-containing protein 1	Q8WUU5	T
TF25	Locus26087	ABCF3	atp-binding cassette sub-family f member 3	Q9NUQ8	P
TF26	Locus26696	MARK1	kinase associated domain 1 family protein	Q9P0L2	K
TF27	Locus2777	ATRX	transcriptional regulator atrx	P46100	P
TF28	Locus28025	RBBP5	retinoblastoma-binding protein 5 isoform 3	Q15291	T
TF29	Locus28427	ARL3	adp-ribosylation factor-like protein 3	P36405	P
TF30	Locus29915	DERL2	derlin-2	Q9GZP9	T
TF31	Locus30074	CSNK1G2	casein kinase i isoform gamma-2	P78368	K
TF32	Locus31482	Rho1	ras-like gtp-binding protein rho1-like	P48148	K P
TF33	Locus31749	PBX1	pre-b-cell leukemia transcription factor 1 isoform 2	P40424	T
TF34	Locus31911	TLK1	serine threonine-protein kinase tousled-like 1	Q9UKI8	K
TF35	Locus3194	eIF4A	atp-dependent rna helicase eif4a	P10081	P
TF36	Locus32533	GNAI1	guanine nucleotide-binding protein g subunit alpha-1	P63096	P
TF37	Locus33249	TOP2A	dna topoisomerase 2-alpha	P11388	K P C
TF38	Locus3390	DNM1L	dynamin-1-like protein	O00429	P
TF39	Locus35083	SRSF2	serine arginine-rich splicing factor 2	Q01130	T
TF40	Locus35615	TUBB4B	Tubulin beta-4B chain-like	P68371	P
TF41	Locus35747	GTF2A2	Transcription initiation factor iia subunit 2	P52657	T
TF42	Locus36874	TK1	thymidine kinase cytosolic	P04183	K
TF43	Locus36966	PDCD11	protein rrp5 homolog	Q14690	T
TF44	Locus39725	hells	hells protein	A8WFW2	C
TF45	Locus40231	DNMT1	DNA (cytosine-5)-methyltransferase 1	P26358	T
TF46	Locus40308	ACTC1	Actin, alpha cardiac muscle 1	P68032	P
TF47	Locus41949	IPO7	importin-7	O95373	P
TF48	Locus42107	MAP4K3	mitogen-activated protein kinase kinase kinase kinase 3-like	Q8IVH8	K P
TF49	Locus42233	ANAPC7	anaphase-promoting complex subunit 7	Q9UJX3	P
TF50	Locus4798	RAP1B	ras-related protein rap-1b	P61224	P
TF51	Locus6106	KCBP	kinesin-like calmodulin-binding	Q7Y1U0	P
TF52	Locus6911	TEF	thyrotroph embryonic factor	Q10587	T
TF53	Locus7360	gro	groucho protein	P16371	K
TF54	Locus8485	act1	actin 1like		T K
TF55	Locus9449	Med6	mediator of rna polymerase ii transcription subunit 6	B0BNE1	T
TF56	NODE_108674	CABIN1	calcineurin-binding protein cabin-1	Q9Y6J0	P
TF57	NODE_142152	POLD1	dna polymerase delta catalytic subunit	P28340	C
TF58	NODE_197218	Prkcd	protein kinase c delta type	P28867	K
TF59	NODE_260401	MINK1	misshapen-like kinase 1	Q8N4C8	K P
TF60	NODE_287829	CBX3	chromobox protein 3	Q13185	T C
TF61	NODE_34619	ACTL	actin-like protein		T

Note: GOs are based on GO-term mapping. Redundant annotation may occur. T: transcription regulator; P: phosphatase; K: kinase; C: chromatin remodeler.

Table A.4: **Unique hubs of inferred genes of Subset E associated with the following GO semantic similarities, neural, muscle, and tissue development, with a minimum two-fold expression change in Figure 3.8B (cf. manuscript)**

Index	Locus	ID	Description	UniProt
G102	Locus37989	B9D1	b9 domain-containing protein 1	Q9UPM9
G103	Locus3814	ASF1	histone chaperone asf1-like	P32447
G112	Locus40374	WNT2B	protein wnt-2b	Q93097
G114	Locus41355	GFI1	zinc finger protein gfi-1	Q99684
G115	Locus41777	PLAC8L1	plac8-like protein 1	A1L4L8
G117	Locus41953	CALM2	calmodulin 2	P62158
G118	Locus41961	CALM1	calmodulin 1	P62158
G122	Locus42556	TMSB10	thymosin beta-10	P63313
G125	Locus489	CALM3	calmodulin 3	P62158
G126	Locus5616	COL4A3BP	collagen type iv alpha-3-binding protein	Q9Y5P4
G127	Locus6335	PTPRM	receptor-type tyrosine-protein phosphatase mu	P28827
G13	Locus14354	CALM2	calmodulin 2	P62158
G131	Locus8254	TARDBP	tar dna-binding protein 43-like	Q13148
G132	Locus8422	WNT4	protein wnt-4	P56705
G137	Locus9306	PAF1	rna polymerase ii-associated factor 1 homolog	Q8N7H5
G19	Locus15032	ARF4	adp-ribosylation factor 4	P18085
G21	Locus16021	SRSF1	serine arginine-rich splicing factor 1-like	Q07955
G22	Locus16319	RPS6KB1	ribosomal protein s6 kinase beta-1 isoform x2	P23443
G23	Locus1661	MYL12B	myosin regulatory light chain 12b	O14950
G26	Locus1694	Sec61a2	transport protein sec61 subunit alpha 2	Q9JLR1
G27	Locus16972	MED28	mediator of rna polymerase ii transcription subunit 28	Q9H204
G29	Locus17155	NRARP	notch-regulated ankyrin repeat-containing protein	Q7Z6K4
G32	Locus1791	ROPN1L	ropporin-1-like protein	Q96C74
G34	Locus19052	YWHAZ	14-3-3 protein zeta	P63104
G36	Locus19781	MIB1	e3 ubiquitin-protein ligase mib1	Q86YT6
G38	Locus2071	ACTA1	alpha skeletal muscle	P68133
G49	Locus21605	FOXB1	fork-head box a b transcription factor	Q99853
G5	Locus11263	WNT16	protein wnt-16	Q9UBV4
G50	Locus21687	CHRM3	muscarinic acetylcholine receptor m3-like	P20309
G54	Locus2303	CALM1	calmodulin 1	P62158
G56	Locus2435	PACRG	parkin coregulated gene protein	Q96M98
G59	Locus2494	CSNK2B	casein kinase ii subunit beta isoform x1	P67870
G67	Locus26630	SES1	serine–trna ligase cytoplasmic-like	P07284
G72	Locus28093	CDC42	cdc42 protein	P60953
G76	Locus2915	HDAC1	histone deacetylase 1	Q13547
G8	Locus11703	CALM1	calmodulin 1	P62158
G82	Locus30501	FOXC2	forkhead box protein c2	Q99958
G91	Locus32058	FOXD3	forkhead box d3	Q9UJU5
G92	Locus34072	NR2E1	nuclear receptor subfamily 2 group e member 1	Q9Y466
G96	Locus35680	mei4	meiosis-specific transcription factor	Q13606

Note: GOs are based on GO-term mapping. Redundant annotation may occur.

Table A.5: **Unique hubs of inferred genes of Subset E associated with the following GO semantic similarities, growth, development, locomotion, and rhythmic behaviors, with minimum two-fold expression change in Figure 3.8C (cf. manuscript)**

Index	Locus	ID	Description	UniProt
G14	Locus19939	TNKS2	a chain human tankyrase 2 - catalytic parp domain in complex with an inhibitor	Q9H2K2
G15	Locus21041	CAD	cad protein	P27708
G36	Locus40425	HISTH2A	histone H2A	Q8I0T3
G40	Locus7645	UBN1	ubinuclein-1	Q9NPG3
G27	Locus31482	Rho1	ras-like gtp-binding protein rho1-like	P48148
G25	Locus28427	ARL3	adp-ribosylation factor-like protein 3	P36405
G1	Locus1154	SMARCD1	swi snf-related matrix-associated actin-dependent regulator of chromatin subfamily d member 1	Q96GM5
G13	Locus19781	MIB1	e3 ubiquitin-protein ligase mib1	Q86YT6

Note: GOs are based on GO-term mapping. Redundant annotation may occur.

Table A.6: **Unique hubs of inferred clustered genes from Library 1 constrained in Figure 3.7 (cf. manuscript) with a minimum two-fold expression change and shown in Figure 3.9**

Index	Locus	ID	Description	UniProt
Cluster 1				
G1	Locus1502	CLSTN1	syntenin-1-like protein	O94985
G2	Locus1534	RBM38	rna-binding protein 38	Q9H0Z9
G8	NODE_407056	ENAH	protein enabled homolog isoform x5	Q8N8S7
G9	NODE_47662	BANF1	barrier-to-autointegration factor	O75531
Cluster 2				
G1	Locus14955	CALML3	calmodulin 3	P27482
G2	Locus1830	PEX7	peroxisomal targeting signal 2 receptor	O00628
G3	Locus26079	EMX1	homeobox protein emx1-like	Q04741
G5	Locus34194	Cyp3A2	cytochrome p450	B9X1A2
G6	Locus7280	LOX	protein-lysine 6-oxidase	P28300
Cluster 3				
G1	Locus1105	furin	furin-like convertase	Q75WU0
G5	Locus26008	APOD	apolipoprotein d	P05090
G6	Locus26216	AQP4	aquaporin-4 isoform x1	P55087
G8	Locus30907	CASP8	caspase apoptosis-related cysteine peptidase	Q66I19
Cluster 4				
G1	Locus15232	CYP17A1	steroid 17-alpha-hydroxylase/17,20 lyase	P05093
G3	Locus25755	DMRTA2	doublesex- and mab-3-related transcription factor a2	Q96SC8
G4	Locus29734	SOD1	superoxide dismutase	P00441
G7	Locus4136	EDIL3	egf-like repeat and discoidin i-like domain-containing protein 3	O43854
Cluster 5				
G2	Locus21345	AGRN	Agrin	O00468
G3	Locus37649	GRHPR	glyoxylate reductase hydroxypyruvate reductase	Q9UBQ7
G4	Locus5075	RPS15A	40s ribosomal protein s15a-like	P62244
Cluster 6				
G1	Locus14384	CSRP3	cysteine and glycine-rich protein 3	P50461
G2	Locus1636	Mlc-c	nonmuscle myosin essential light chain	P54357
G7	Locus27512	APOD	apolipoprotein d	P05090
G8	Locus27739	DUSP10	dual specificity protein phosphatase 10	Q9Y6W6
G9	Locus39860	IIID1	protein hid-1 homolog	Q8IV36
G12	Locus8878	ZC4H2	zinc finger c4h2 domain-containing protein	Q9NQZ6
Cluster 7				
G6	Locus3688	ALDH6A1	methylmalonate-semialdehyde dehydrogenase	Q02252
Cluster 8				
G13	Locus18000	CUL3	cullin-3	Q13618
G33	Locus27541	U2SURP	u2 snrnp-associated surp motif-containing protein isoform x3	O15042
Cluster 9				
G3	Locus7928	CETN1	centrin-1	Q12798
Cluster 10				
G1	Locus12409	APOD	apolipoprotein d-like	P05090
G2	Locus15311	AMY1A	alpha-amylase	P04745
G3	Locus16176	CYP2U1	cytochrome p450 2u1-like	Q7Z449
G4	Locus18485	Lectin-galC1	galactose-specific c-type	Q94881
G5	Locus23406	pnbA	para-nitrobenzyl esterase-like	P37967
G6	Locus36105	01 Mar	mosc domain-containing protein mitochondrial	Q5VT66
G7	Locus36280	RPL36A	ribosomal protein l36a	P83881
G8	Locus673	MOXD1	dbh-like monooxygenase protein 1	Q6UVY6
Cluster 11				
G14	Locus17187	RBMX2	rna-binding motif x-linked 2	Q9Y388
G23	Locus24054	APEX1	dna-(apurinic or apyrimidinic site) lyase-like	P27695
G28	Locus2658	HM13	histocompatibility 13	Q8TCT9
G36	Locus32818	OTUB1	ubiquitin thioesterase otub1	Q96FW1
G47	Locus40401	APITD1	centromere protein s	Q8N2Z9
Cluster 12				
G1	Locus12149	BTBD2	BTB/POZ domain-containing protein 2	Q9BX70
G2	Locus12973	CAMKK2	calcium calmodulin-dependent protein kinase kinase 2	Q96RR4
G5	Locus14828	CDK9	cyclin-dependent kinase 9	P50750
G6	Locus1483	INSIG2	insulin-induced gene 2 protein	Q9Y5U4
G7	Locus1529	TBL1XR1	f-box-like wd repeat-containing protein tbl1xr1	Q9BZK7
G9	Locus18396	Spr	cell wall-associated hydrolase	Q8RBE1
G10	Locus18840	STX5	syntaxin 5	Q13190
G11	Locus20545	RPL38	60s ribosomal protein l38	P63173
G13	Locus28025	RBBP5	retinoblastoma-binding protein 5 isoform 3	Q15291
G14	Locus30708	SDC2	Syndecan 2	P34741
G15	Locus35371	KANK1	kn motif and ankyrin repeat domain-containing protein 1	Q14678

G17	Locus4424	SNF7	snf7-like protein	P39929
G18	Locus8995	CUX1	homeobox protein cut-like 1 isoform x9	P39880

Note: After clustering, the 12 sets of genes have been further compacted with an enrichment analysis of the top GO-terms (P<0.05), and were subjected next to network inference, each against the 16 most significant regulators of Table A.3.

A.3 AUTHORS' CONTRIBUTIONS

SB carried out the sampling, data preparation and processing, pipeline design and data mining, developed the bioinformatics methods and data analysis, prepared the figures and tables, drafted and maintained the manuscript and the supporting information (SI). AT helped with the coordination of the de novo assembly of the transcriptome, participated in the microarray design and execution, and helped in drafting the manuscript. BG, DM, and RT critically improved the manuscript. All authors read and approved the final manuscript.

A.4 R SESSION INFORMATION

All analyses were performed in R v3.0.2 [R Core Team, 2013]. The R commands are available from corresponding author upon request.

* R version 3.0.2 (2013-05-16), `x86_64-w64-mingw32`
* Base packages: base, datasets, graphics, grDevices, grid, methods, parallel, stats, tcltk, utils
* Other packages: AnnotationDbi 1.22.6, Biobase 2.20.1, BiocGenerics 0.6.0, caTools 1.14, class 7.3-9, corpcor 1.6.6, DBI 0.2-7, DynDoc 1.38.0, e1071 1.6-1, ebdbNet 1.2.3, fdrtool 1.2.11, gdata 2.13.2, GeneNet 1.2.8, GO.db 2.9.0, gplots 2.11.3, graph 1.38.3, gtools 3.1.0, igraph 0.6.5-2, impute 1.34.0, KernSmooth 2.23-10, lattice 0.20-24, limma 3.16.8, longitudinal 1.1.8, MASS 7.3-29, matrixStats 0.8.12, Mfuzz 2.18.0, permute 0.7-0, plyr 1.8, pvclust 1.2-2, RColorBrewer 1.0-5, RSQLite 0.11.4, samr 2.0, SparseM 1.03, statmod 1.4.18, topGO 2.12.0, vegan 2.0-9, widgetTools 1.38.0, xtable 1.7-1
* Loaded via a namespace (and not attached): bitops 1.0-6, IRanges 1.18.4, R.methodsS3 1.5.2, stats4 3.0.2, tkWidgets 1.38.0, tools 3.0.2

B

SUPPLEMENTAL INFORMATION II

Genetic impact of dietary fatty acids on growth and mortality during early *Mytilus edulis* larval development

B.1 FIGURES

B.1.1 *Machine Learning*

The main goal of our modeling scheme was to characterize the relationship between the n_g predictors and the outcome y. The response outcome, as described above, was a two-class physiological classification problem. Larvae will either fall in the class of fast growth and normal mortality rates, or in the second class of normal growth and high-mortality rates. Modeling for inference was evaluated on the basis of the lowest averaged RMSE, next to the highest accuracy and best performance. Figure 4.14 uses the model selected after feature elimination for training six base learners chosen due to their capability of best fitting the dataset averaging a low RMSE threshold.

B.1.1.1 *Ensemble methods tuning*

Extracting information from biological data requires sophisticated computational frameworks [Yip et al., 2013]. Pattern discovery was used to search for non-random gene signatures in order to extract a relative decision about the physiological outcome y. The hyper-

parameters of the selected base learners (ANN, linear, polynomial, or radial kernels for SVM, boosted GAM, boosted GLM, RF, RRF, principal components regression (PCR), ridge regression (Ridge), PLS) were adjusted sequentially through iterated tests. Ten-fold CV was used as an external resampling (without repetition) method accuracy estimation of each predicted model. The classifiers with the best performance, after hyper-parameter optimization, were retained for a final learning round, given that bagging (random subsampling with repetition of rows) was used in this second accuracy evaluation process. Different learners with best performance (ANN, Ridge regression, boosted GLM, SVM (linear and radial kernels), and a PCR) were used for final classification estimation. Figure B.1 shows the computational running time of the system of these learners, under six iteration intervals. It should be noted that time requirement for all these algorithms is usually more because of the tuning of the hyper-parameters performed during each resampling execution. However, we reduced execution time by tuning the models before bagging. The model settings were tuned first using CV then further trained using the best hyper-parameter estimates. Ensemble method was build by combining the highest performance classifiers by averaging their predictive estimates. Variance and stability of the ensemble were tested using several iterated votes and Figure 4.7 shows the four finally selected ensembles with the best predictive accuracies. Amongst these classifiers, GLM with a component-wise boosting algorithm [Buchlmann, 2006] increased its power when combined with the second classifier, a nonlinear feed-forward ANN with 17 units in the hidden layer [Venables and Ripley, 2002]. When averaging both of the classifiers while up-weighing the ANN predictions (method III), the ensemble methods delivered the best RMSE estimates for our high-dimensional data (Figure 4.7) and a minimum of 87% classification accuracy.

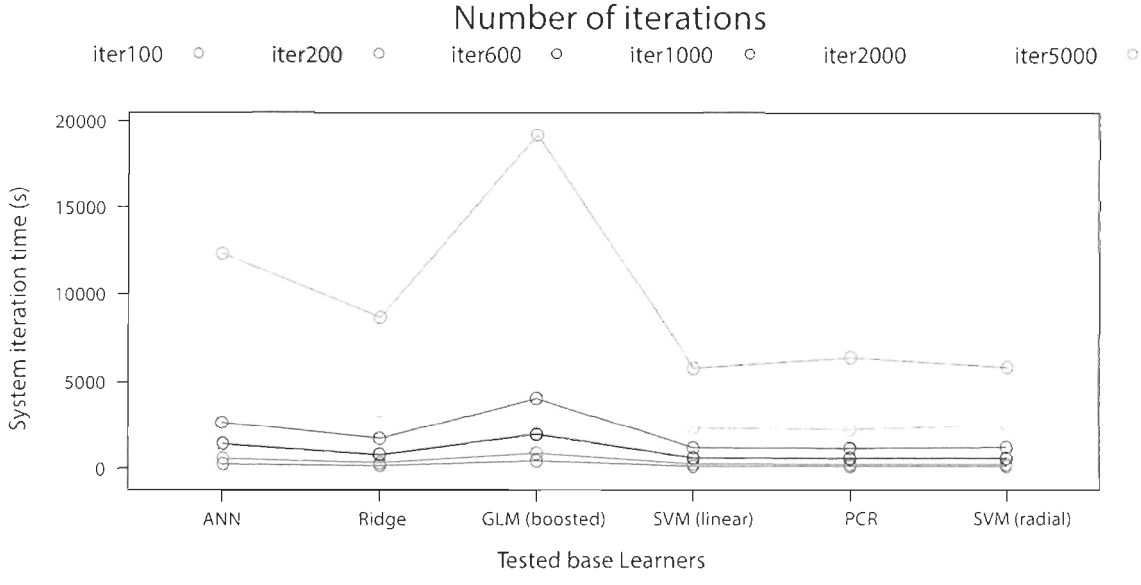

Figure B.1: **System running time for the different bagging iterations from Figure** 4.14. Bagging was used under different iterations (100, 200, 500, 1K, 2K, 5K). ANN, multilayer perceptron artificial neural network; SVM, support vector machines; GLM, generalized linear models; PCR, principal components regression; Ridge, ridge regression; PLS, partial least squares. All functions were iterated using a workstation cluster for parallel computations in R and executed on an Intel Core i5 1.7 GHz multi-core machine with 4 GB memory having Windows 7 operating system.

B.2 TABLES

An enrichment analysis of Fisher's exact test was realized in order to integrate GO-terms with significant genes. Enrichment tests are based on gene counts and methods accounting for the GO topology. GO categories were mapped to protein IDs using the functional information found in public databases provided by NCBI (National Center for Biotechnology Information), PIR (Protein Information Resource), and GO (Gene Ontology) that were used to retrieve GO-terms stored in GO repositories. An Evidence Code (not shown) has been used to monitor the quality of this functional assignment and irrelevant mappings were henceforward removed. No BioMart annotated genes specific to Mollusca, Bivalvia or *Mytilus edulis* could have been adopted for term-gene linkages. Likewise, no microarray specific annotation packages were available from Bioconductor for this species. In this regard, our custom arrays made use of specific mapping between genes and GO terms, with most up-to-date annotations, implemented with a GO hierarchical structure covering thus around 60% of the dataset with specific Biological Process, Molecular Function, and Cellular Component ontologies.

Table B.1: **Enrichment analysis of Biological Processes GO-mapped transcripts against the whole annotated *Mytilus edulis* transcriptome**

GO ID	Term	Annotated	Significant	Expected	Classic Fisher
GO:0043966	histone H3 acetylation	11	1	0.02	0.022
GO:0032392	DNA geometric change	13	1	0.03	0.026
GO:0032508	DNA duplex unwinding	13	1	0.03	0.026
GO:0034645	cellular macromolecule biosynthetic proc...	750	4	1.49	0.037
GO:0044743	intracellular protein transmembrane impo...	19	1	0.04	0.037
GO:0009059	macromolecule biosynthetic process	759	4	1.51	0.039
GO:0006626	protein targeting to mitochondrion	20	1	0.04	0.039
GO:0065002	intracellular protein transmembrane tran...	21	1	0.04	0.041
GO:0071806	protein transmembrane transport	21	1	0.04	0.041
GO:0072655	establishment of protein localization to...	23	1	0.05	0.045
GO:0070585	protein localization to mitochondrion	24	1	0.05	0.047
GO:0009615	response to virus	25	1	0.05	0.049
GO:0090305	nucleic acid phosphodiester bond hydroly...	26	1	0.05	0.051
GO:0006839	mitochondrial transport	28	1	0.06	0.055
GO:0006302	double-strand break repair	31	1	0.06	0.060
GO:0010467	gene expression	874	4	1.74	0.063
GO:0071103	DNA conformation change	34	1	0.07	0.066
GO:0016573	histone acetylation	37	1	0.07	0.072
GO:0018393	internal peptidyl-lysine acetylation	37	1	0.07	0.072
GO:0018394	peptidyl-lysine acetylation	37	1	0.07	0.072

Note: The terms show the top 20 Gene Ontology (GO) nodes for Biological processes (BPs). BP GO-terms were mapped to 1,707 nodes and 3,468 edges of GO-entries from *Mytilus edulis* transcriptome (Bassim et al., *in prep*). Significance is measured by Fishers classic test.

Table B.2: **Enrichment analysis of Molecular Functions GO-mapped transcripts against the whole annotated *Mytilus edulis* transcriptome**

GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
GO:0002135	CTP binding	1	1	0.00	0.0020
GO:0017098	sulfonylurea receptor binding	1	1	0.00	0.0020
GO:0001884	pyrimidine nucleoside binding	2	1	0.00	0.0040
GO:0002134	UTP binding	2	1	0.00	0.0040
GO:0032551	pyrimidine ribonucleoside binding	2	1	0.00	0.0040
GO:0019103	pyrimidine nucleotide binding	3	1	0.01	0.0059
GO:0032552	deoxyribonucleotide binding	3	1	0.01	0.0059
GO:0032554	purine deoxyribonucleotide binding	3	1	0.01	0.0059
GO:0032557	pyrimidine ribonucleotide binding	3	1	0.01	0.0059
GO:0032558	adenyl deoxyribonucleotide binding	3	1	0.01	0.0059
GO:0032564	dATP binding	3	1	0.01	0.0059
GO:0030911	TPR domain binding	4	1	0.01	0.0079
GO:0016787	hydrolase activity	687	4	1.36	0.0264
GO:0030235	nitric-oxide synthase regulator activity	16	1	0.03	0.0313
GO:0004553	hydrolase activity, hydrolyzing O-glycos...	28	1	0.06	0.0542
GO:0051082	unfolded protein binding	30	1	0.06	0.0579
GO:0044325	ion channel binding	32	1	0.06	0.0617
GO:0016798	hydrolase activity, acting on glycosyl b...	42	1	0.08	0.0803
GO:0019901	protein kinase binding	58	1	0.11	0.1095
GO:0019900	kinase binding	62	1	0.12	0.1166

Note: The terms show the top 20 Gene Ontology (GO) nodes for Molecular Function (MFs). MF GO-terms were mapped to 1,742 nodes and 2,179 edges of GO-entries from *Mytilus edulis* transcriptome (Bassim et al., *in prep*). Significance is measured by Fishers classic test.

Table B.3: **Enrichment analysis of Cellular Components GO-mapped transcripts against the whole annotated *Mytilus edulis* transcriptome**

GO.ID	Term	Annotated	Significant	Expected	Classic Fisher
GO:0005956	protein kinase CK2 complex	2	1	0.00	0.0044
GO:0031519	PcG protein complex	9	1	0.02	0.0199
GO:0005924	cell-substrate adherens junction	15	1	0.03	0.0329
GO:0005925	focal adhesion	15	1	0.03	0.0329
GO:0030055	cell-substrate junction	16	1	0.04	0.0351
GO:0005912	adherens junction	21	1	0.05	0.0458
GO:0070161	anchoring junction	23	1	0.05	0.0501
GO:0030054	cell junction	58	1	0.13	0.1224
GO:0005886	plasma membrane	333	2	0.74	0.1629
GO:0071944	cell periphery	362	2	0.81	0.1869
GO:0005856	cytoskeleton	367	2	0.82	0.1911
GO:0005874	microtubule	126	1	0.28	0.2497
GO:0005829	cytosol	442	2	0.98	0.2563
GO:0005576	extracellular region	139	1	0.31	0.2722
GO:0005737	cytoplasm	1740	5	3.87	0.3094
GO:0043228	non-membrane-bounded organelle	896	3	1.99	0.3171
GO:0043232	intracellular non-membrane-bounded organ...	896	3	1.99	0.3171
GO:0005840	ribosome	223	1	0.50	0.4044
GO:0015630	microtubule cytoskeleton	251	1	0.56	0.4437
GO:0032991	macromolecular complex	1068	3	2.38	0.4471

Note: The terms show the top 20 Gene Ontology (GO) nodes for Cellular Components (CCs). CC GO-terms were mapped to 892 nodes and 1,724 edges of GO-entries from *Mytilus edulis* transcriptome (Bassim et al., *in prep*). Significance is measured by Fishers classic test.

B.3 AUTHORS' CONTRIBUTIONS

SB carried out the sampling, data preparation and processing, pipeline design and data mining, developed the bioinformatics methods and data analysis, prepared the figures and tables, drafted and maintained the manuscript and the supporting information (SI). AT participated in the microarray design and execution. RC assisted with the Machine Learning analysis and critically reviewed the manuscript. DM, and RT critically improved the manuscript. All authors read and approved the final manuscript.

B.4 R SESSION INFORMATION

All analyses were performed in R v3.0.2 [R Core Team, 2013]. The R commands are available from corresponding author upon request.

* R version 3.0.2 (2013-09-25), `i386-w64-mingw32`
* Locale: `LC_COLLATE=English_Canada.1252`, `LC_CTYPE=English_Canada.1252`, `LC_MONETARY=English_Canada.1252`, `LC_NUMERIC=C`, `LC_TIME=English_Canada.1252`
* Base packages: base, datasets, graphics, grDevices, methods, parallel, splines, stats, utils
* Other packages: caret 5.17-7, cluster 1.14.4, doSNOW 1.0.9, elasticnet 1.1, foreach 1.4.1, gbm 2.1, glmnet 1.9-5, igraph 0.6.6, iterators 1.0.6, kernlab 0.9-19, lars 1.2, lattice 0.20-24, Matrix 1.1-0, mboost 2.2-3, mRMRe 2.0.4, nnet 7.3-7, pls 2.4-3, plyr 1.8, randomForest 4.6-7, reshape2 1.2.2, rgl 0.93.991, RRF 1.6, scatterplot3d 0.3-34, snow 0.3-13, survival 2.37-4
* Loaded via a namespace (and not attached): codetools 0.2-8, compiler 3.0.2, grid 3.0.2, stringr 0.6.2, tools 3.0.2

C

INFORMATION SUPPLÉMENTAIRE III

C.1 GITHUB

GitHub est une banque de code informatique qui héberge les projets de développement et permet leur partage une fois ils sont diffusés publiquement. Avoir accès au projet via GitHub permet de télécharger les scripts qui sont publiquement partagés, les testés ou éventuellement les corrigés. Pour plus d'information concernant GitHub, veuillez suivre le *lien* suivant. Les codes sources pour les trois papiers fournis dans cette thèse (Chapitres 2, 3 et 4), ainsi que les scripts individuels pour certaines figures de la thèse sont rendues publique par le premier auteur de ces manuscrits et de la thèse, S. Bassim. Pour consulter les scripts R dans votre navigateur web veuillez suivre le *lien* suivant si la thèse est visualisée sous format PDF. Dans le cas échéant rendez-vous sur le site GitHub :
github.com/neocruiser/thesis2014

GitHub is a project hosting public repository that allows sharing, downloading, testing or correcting the available scripts. To find more about GitHub please follow this direct *link*. The source codes for the three first manuscripts appearing in the dissertation of the present thesis (Chapters 2, 3, and 4), also scripts for some figures are kindly made public by S. Bassim, the first author of the manuscripts and this thesis. If the manuscript is being assessed in its PDF format, the R scripts can be visualised or downloaded following this *link*. If the paper version is being used please visit the deployed GitHub repository available for this thesis :
github.com/neocruiser/thesis2014

Figure C.1: **QR-code du répertoire des scripts R rendu publique sur GitHub.** (*Accès via smartphone*)

Tableau C.1: **Les différents logiciels utilisés pour le traitement des données et la présentation des résultats.**

Titre	Version	Courte description
Timecult	1.4	Gestionnaire de tâches et suivi de temps
Dropbox	2.4.11	Backup
Git	1.9.2	Version Control Service (VCS)
Xyplorer	13.10	Explorateur de répertoires
Freemind	1.0.0	Visualisation et organisation
Notepad++	6.5.1	Traitement de séquence et gestionnaire
Emacs	24.3	Emulateur R, LaTeX, Markdown, HTML
Past	2.17c	Tabulateur classique
R	3.0.2	Explorateur statistique des bases de données
TexStudio	2.6.2	Processeur LaTeX
Antidote	8	Correction du français
Jabref	2.9.2	Gestionnaire BibTex
LaTeX2RTF	2.3.3	Conversion LaTeX en *.doc et *.odt
Foxit Reader	6.0.7	Outil PDF
Adobe Illustrator	CS6	Traitement vectoriel des figures
Adobe Indesign	CS6	Préparation des présentations
Beamer	3.33	Préparation des présentations sous LaTeX
Prezi		Préparation des présentations interactives
ImageMagick	6.8.8	Traitement de figures et animations via bash
Velvet	1.1.07	Assemblage de novo des contigs
Oases	0.1.11	Assemblage de novo des contigs
Blast2GO	Pro	Annotation et GO-mapping du transcriptome
MEGA5	5.2	Alignement de séquence
ClustalX	2	Alignement de séquence
Cytoskape	3.0.2	Traitement visuel des networks (réseaux géniques)
Cell Designer	4.3	Schématisation des composants de la cellule
Oligo 7	7	Dessin des sondes RT-(q)PCR

BIBLIOGRAPHIE

Aarab, L., Pérez-Camacho, A., del Pino Viera-Toledo, M., de Viçose, G. C., Fernández-Palacios, H., and Molina, L. (2013). Embryonic development and influence of egg density on early veliger larvae and effects of dietary microalgae on growth of brown mussel *Perna perna* (l. 1758) larvae under laboratory conditions. *Aquaculture International*, 21(5) :1065–1076.

Aarab, N., Lemaire-Gony, S., Unruh, E., Hansen, P., Larsen, B., Andersen, O., and Narbonne, J. (2006). Preliminary study of responses in mussel (*Mytilus edulis*) exposed to bisphenol A, diallyl phthalate and tetrabromodiphenyl ether. *Aquatic Toxicology*, 78(Suppl. 1) :S86–S92.

Abrams, P. A., Leimar, O., Nylin, S., and Wiklund, C. (1996). The effect of flexible growth rates on optimal sizes and development times in a seasonal environment. *American Naturalist*, pages 381–395.

Ackerman, J. D., Sim, B., Nichols, S. J., and Claudi, R. (1994). A review of the early life history of zebra mussels (*Dreissena polymorpha*) : comparisons with marine bivalves. *Canadian Journal of Zoology*, 72(7) :1169–1179.

Adhya, M., Choi, K.-S., Yu, Y., and Cho, M. (2010). Expression and localization of MCsialec, a sialic acid-specific lectin in the marine bivalve Manila clam, *Ruditapes philppinarum*. *Journal of Fish Diseases*, 33(11) :889–899.

Alexa, A. and Rahnenfuhrer, J. (2010). topgo : enrichment analysis for gene ontology. *R package version*, 2(1).

Alfaro, A. C., Young, T., and Ganesan, A. M. (2011). Regulatory effects of mussel (*Aulacomya maoriana* Iredale 1915) larval settlement by neuroactive compounds, amino acids and bacterial biofilms. *Aquaculture*, 322(Complete) :158–168.

Alon, U. (2007). Network motifs : theory and experimental approaches. *Nature Reviews Genetics*, 8(6) :450–461.

Altschul, S., Demchak, B., Durbin, R., Gentleman, R., Krzywinski, M., Li, H., Nekrutenko, A., Robinson, J., Rasband, W., Taylor, J., and Trapnell, C. (2013). The anatomy of successful computational biology software. *Nature Biotechnology*, 31(10) :894–897.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped blast and psi-blast : a new generation of protein database search programs. *Nucleic acids research*, 25(17) :3389–3402.

Ambroise, C. and McLachlan, G. J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Sciences*, 99(10) :6562–6566.

Americo, J. A., Dondero, F., Moraes, M. O., Allodi, S., and Freitas Rebelo, M. d. (2013). Gene discovery in the tropical scallop *Nodipecten nodosus* : construction and sequencing of a normalized cDNA library. *Marine Environmental Research*, 91(1) :34–40.

Andersen, A., Torgersen, J. S., Pagander, H. H., Magnesen, T., and Johnston, I. A. (2009). Gene expression analyses of essential catch factors in the smooth and striated adductor muscles of larval, juvenile and adult great scallop (*Pecten maximus*). *Journal of Muscle Research and Cell Motility*, 30(5-6) :233–242.

Andersen, S., Christophersen, G., and Magnesen, T. (2011). Spat prdouction of the great scallop (*Pecten maximus*) a roller coster. *Canadian Journal of Zoology*, 89(7) :579–598.

Anderson, D. T. (1994). *Barnacles : structure, function, development and evolution.* Springer.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics & Development*, 11(6) :635–641.

Aranda-Burgos, J. A., da Costa, F., Nóvoa, S., Ojea, J., and Martínez-Patiño, D. (2014). Effects of microalgal diet on growth, survival, biochemical and fatty acid composition of *Ruditapes decussatus* larvae. *Aquaculture*, 420(Complete) :38–48.

Araya, M. T., Markham, F., Mateo, D. R., McKenna, P., Johnson, G. R., Berthe, F. C., and Siah, A. (2010). Identification and expression of immune-related genes in hemocytes of soft-shell clams, *Mya arenaria*, challenged with *Vibrio splendidus*. *Fish & Shellfish Immunology*, 29(4) :557–564.

Arts, M. T., Ackman, R. G., and Holub, B. J. (2001). "Essential fatty acids" in aquatic ecosystems : A crucial link between diet and human health and evolution. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(1) :122–137.

Bachère, E. (2003). Anti-infectious immune effectors in marine invertebrates : potential tools for disease control in larviculture. *Aquaculture*, 227(1) :427–438.

Bado-Nilles, A., Quentel, C., Auffret, M., Le Floch, S., Gagnaire, B., Renault, T., and Thomas-Guyon, H. (2009). Immune effects of HFO on european sea bass, *Dicentrarchus labrax*, and Pacific oyster, *Crassostrea gigas*. *Ecotoxicology and Environmental Safety*, 72(5) :1446–1454.

Baginsky, S., Hennig, L., Zimmermann, P., and Gruissem, W. (2010). Gene expression analysis, proteomics, and network discovery. *Plant physiology*, 152(2) :402–410.

Baker, S. M. and Mann, R. (1994). Feeding ability during settlement and metamorphosis in the oyster *Crassostrea virginica* (Gmelin, 1791) and the effects of hypoxia on post-settlement ingestion rates. *Journal of Experimental Marine Biology and Ecology*, 181(2) :239–253.

Balseiro, P., Falcó, A., Romero, A., Dios, S., Martínez-López, A., Figueras, A., Estepa, A., and Novoa, B. (2011). *Mytilus galloprovincialis* myticin c : a chemotactic molecule with antiviral activity and immunoregulatory properties. *PloS one*, 6(8) :e23140.

Balseiro, P., Moreira, R., Chamorro, R., Figueras, A., and Novoa, B. (2013). Immune responses during the larval stages of *Mytilus galloprovincialis* : metamorphosis alters immunocompetence, body shape and behavior. *Fish & shellfish immunology*, 35(2) :438–447.

Bao, W.-Y., Yang, J.-L., Satuito, C. G., and Kitamura, H. (2007). Larval metamorphosis of the mussel *Mytilus galloprovincialis* in response to *Alteromonas* sp. 1 : evidence for two chemical cues ? *Marine Biology*, 152(3) :657–666.

Baran, K., Dunstone, M., Chia, J., Ciccone, A., Browne, K. A., Clarke, C. J., Lukoyanova, N., Saibil, H., Whisstock, J. C., Voskoboinik, I., et al. (2009). The molecular basis for perforin oligomerization and transmembrane pore assembly. *Immunity*, 30(5) :684–695.

Barmo, C., Ciacci, C., Canonico, B., Fabbri, R., Cortese, K., Balbi, T., Marcomini, A., Pojana, G., Gallo, G., and Canesi, L. (2013). *In vivo* effects of n-TiO 2 on digestive gland and immune function of the marine bivalve *Mytilus galloprovincialis*. *Aquatic Toxicology*, 132(Complete) :9–18.

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M., Marshall, K. A., et al. (2009). NCBI GEO : archive for high-throughput functional genomic data. *Nucleic Acids Research*, 37(suppl 1) :D885–D890.

Barros, P., Sobral, P., Range, P., Chícharo, L., and Matias, D. (2013). Effects of sea-water acidification on fertilization and larval development of the oyster *Crassostrea gigas*. *Journal of Experimental Marine Biology and Ecology*, 440 :200–206.

Bassim, S., Genard, B., Gauthier-Clerc, S., Moraga, D., and Tremblay, R. (2014a). Ontogeny of bivalve immunity : Assessing the potential of next-generation sequencing techniques. *Reviews in Aquaculture*, 6 :1–21.

Bassim, S., Tanguy, A., Genard, B., Moraga, D., and Tremblay, R. (2014b). Identification of *Mytilus edulis* genetic regulators during early development. *Gene*.

Basti, L., Nagai, K., Tanaka, Y., and Segawa, S. (2013). Sensitivity of gametes, fertilization, and embryo development of the Papanese pearl oyster, *Pinctada fucata martensii*, to the harmful dinoflagellate, *Heterocapsa circularisquama*. *Marine Biology*, 160(1) :211–219.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., et al. (2004). The pfam protein families database. *Nucleic acids research*, 32(suppl 1) :D138–D141.

Bayne, B. (1965). Growth and the delay of metamorphosis of the larvae of *Mytilus edulis* (L.). *Ophelia*, 2(1) :1–47.

Bayne, B. (1971). Some morphological changes that occur at the metamorphosis of the larvae of mytilus edulis. In *Fourth European Marine Biology Symposium*, pages 259–280. Cambridge University Press, Cambridge.

Bayne, B. L. (2004). Phenotypic flexibility and physiological tradeoffs in the feeding and growth of marine bivalve molluscs. *Integrative and Comparative Biology*, 44(6) :425–432.

Beal, M. J., Falciani, F., Ghahramani, Z., Rangel, C., and Wild, D. L. (2005). A Bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, 21(3) :349–356.

Beaumont, A. and Budd, M. (1982). Delayed growth of mussel (*Mytilus edulis*) and scallop (*Pecten maximus*) veligers at low temperatures. *Marine Biology*, 71(1) :97–100.

Bechmann, R. K., Taban, I. C., Westerlund, S., Godal, B. F., Arnberg, M., Vingen, S., Ingvarsdottir, A., and Baussant, T. (2011). Effects of ocean acidification on early life stages of shrimp (*Pandalus borealis*) and mussel (*Mytilus edulis*). *Journal of Toxicology and Environmental Health, Part A*, 74(7-9) :424–438.

Beiras, R. and Widdows, J. (1995). Induction of metamorphosis in larvae of the oyster *Crassostrea gigas* using neuroactive compounds. *Marine Biology*, 123(2) :327–334.

Ben Kheder, R., Quere, C., Moal, J., and Robert, R. (2010). Effect of nutrition on *Crassostrea gigas* larval development and the evolution of physiological indices : Part b : Effects of temporary food deprivation. *Aquaculture*, 308(3) :174–182.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1) :289–300.

Bertram, D. F. and Strathmann, R. R. (1998). Effects of maternal and larval nutrition on growth and form of planktotrophic larvae. *Ecology*, 79(1) :315–327.

Bettencourt, R., Dando, P., Collins, P., Costa, V., Allam, B., and Serrão Santos, R. (2009). Innate immunity in the deep sea hydrothermal vent mussel *Bathymodiolus azoricus*. *Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology*, 152(2) :278–289.

Betti, M., Ciacci, C., Lorusso, L. C., Canonico, B., Falcioni, T., Gallo, G., and Canesi, L. (2006). Effects of tumour necrosis factor α (TNFα) on *Mytilus* haemocytes : role of stress-activated mitogen-activated protein kinases (MAPKs). *Biology of the Cell*, 98(4) :233–244.

Bi, Y. and Davuluri, R. V. (2013). NPEBseq : nonparametric empirical bayesian-based procedure for differential expression analysis of RNA-seq data. *BMC bioinformatics*, 14(1) :1–12.

Bieler, R., Mikkelsen, P. M., and Giribet, G. (2013). Bivalvia-a discussion of known unknowns*. *American Malacological Bulletin*, 31(1) :123–133.

Bishop, C., Erezyilmaz, D., Flatt, T., Georgiou, C., Hadfield, M., Heyland, A., Hodin, J., Jacobs, M., Maslakova, S., Pires, A., et al. (2006a). What is metamorphosis ? *Integrative and Comparative Biology*, 46(6) :655–661.

Bishop, C. D., Huggett, M. J., Heyland, A., Hodin, J., and Brandhorst, B. P. (2006b). Interspecific variation in metamorphic competence in marine invertebrates : the significance for comparative investigations into the timing of metamorphosis. *Integrative and Comparative Biology*, 46(6) :662–682.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.

Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2) :185–193.

Boudry, P., Collet, B., Cornette, F., Hervouet, V., and Bonhomme, F. (2002). High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture*, 204(3) :283–296.

Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., Griffith, J. F., Lettieri, T., Field, D., Benzie, J., et al. (2013). Genomics in marine monitoring : New opportunities for assessing marine health status. *Marine Pollution Bulletin*, 74(1) :19–31.

Boutet, I., Meistertzheim, A.-L., Tanguy, A., Thébault, M.-T., and Moraga, D. (2005). Molecular characterization and expression of the gene encoding aspartate aminotransferase from the Pacific oyster *Crassostrea gigas* exposed to environmental stressors. *Comparative Biochemistry and Physiology Part C : Toxicology & Pharmacology*, 140(1) :69–78.

Boutet, I., Tanguy, A., and Moraga, D. (2004). Response of the Pacific oyster *Crassostrea gigas* to hydrocarbon contamination under experimental conditions. *Gene*, 329(Complete) :147–157.

Braga-Neto, U. M. and Dougherty, E. R. (2004). Is cross-validation valid for small-sample microarray classification ? *Bioinformatics*, 20(3) :374–380.

Braubach, O. R., Dickinson, A. J., Evans, C. C., and Croll, R. P. (2006). Neural control of the velum in larvae of the gastropod, *Ilyanassa obsoleta*. *Journal of Experimental Biology*, 209(23) :4676–4689.

Brazhnik, P., de la Fuente, A., and Mendes, P. (2002). Gene networks : how to put the function in genomics. *TRENDS in Biotechnology*, 20(11) :467–472.

Brehmer, P., Do Chi, T., Laugier, T., Galgani, F., Laloë, F., Darnaude, A. M., Fiandrino, A., and Mouillot, D. (2011). Field investigations and multi-indicators for shallow water lagoon management : perspective for societal benefit. *Aquatic Conservation : Marine and Freshwater Ecosystems*, 21(7) :728–742.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2) :123–140.

Brett, M. and Müller-Navarra, D. (1997). The role of highly unsaturated fatty acids in aquatic foodweb processes. *Freshwater Biology*, 38(3) :483–499.

Bricelj, V., MacQuarrie, S., Doane, J., and Connell, L. (2010). Evidence of selection for resistance to paralytic shellfish toxins during the early life history of soft-shell clam (*Mya arenaria*) populations. *Limnology and Oceanography*, 55(6) :2463–2475.

Brown, M., Davies, I., Moffat, C., and Craft, J. (2006). Application of SSH and a macroarray to investigate altered gene expression in *Mytilus edulis* in response to exposure to benzo[*a*]pyrene. *Marine Environmental Research*, 62(Supplement 1) :S128–S135.

Buehlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, (2) :559–583.

Caers, M., Coutteau, P., Lombeida, P., and Sorgeloos, P. (1998). The effect of lipid supplementation on growth and fatty acid composition of *Tapes philippinarum* spat. *Aquaculture*, 162(3) :287–299.

Calabrese, A., MacInnes, J., Nelson, D., and Miller, J. (1977). Survival and growth of bivalve larvae under heavy-metal stress. *Marine Biology*, 41(2) :179–184.

Callewaert, L. and Michiels, C. W. (2010). Lysozymes in the animal kingdom. *Journal of Biosciences*, 35(1) :127–160.

Campa-Córdova, A. I., González-Ocampo, H., Luna-González, A., Mazón-Suástegui, J. M., and Ascencio, F. (2009). Growth, survival, and superoxide dismutase activity in juvenile *Crassostrea corteziensis* (Hertlein, 1951) treated with probiotics Crecimiento. *Hidrobiológica*, 19(2) :151–157.

Canesi, L., Betti, M., Ciacci, C., Citterio, B., Pruzzo, C., and Gallo, G. (2003). Tyrosine kinase-mediated cell signalling in the activation of *Mytilus* hemocytes : possible role of stat-like proteins. *Biology of the Cell*, 95(9) :603–613.

Canesi, L., Ciacci, C., Betti, M., Lorusso, L. C., Marchi, B., Burattini, S., Falcieri, E., and Gallo, G. (2004). Rapid effects of 17β-estradiol on cell signaling and function of *Mytilus* hemocytes. *General and Comparative Endocrinology*, 136(1) :58–71.

Canesi, L., Ciacci, C., Fabbri, R., Marcomini, A., Pojana, G., and Gallo, G. (2012). Bivalve molluscs as a unique target group for nanoparticle toxicity. *Marine Environmental Research*, 76(Complete) :16–21.

Canesi, L., Scarpato, A., Betti, M., Ciacci, C., Pruzzo, C., and Gallo, G. (2002). Bacterial killing by *Mytilus* hemocyte monolayers as a model for investigating the signaling pathways involved in mussel immune defence. *Marine Environmental Research*, 54(3) :547–551.

Cannuel, R., Beninger, P. G., McCombie, H., and Boudry, P. (2009). Gill development and its functional and evolutionary implications in the blue mussel *Mytilus edulis* (Bivalvia : Mytilidae). *The Biological Bulletin*, 217(2) :173–188.

Cao, L., Kenchington, E., and Zouros, E. (2004). Differential segregation patterns of sperm mitochondria in embryos of the blue mussel (*Mytilus edulis*). *Genetics*, 166(2) :883–894.

Carcamo, C., Comesana, A. S., Winkler, F. M., and Sanjuan, A. (2005). Allozyme identification of mussels (bivalvia : *Mytilus*) on the Pacific coast of south america. *Journal of Shellfish Research*, 24(4) :1101–1115.

Carroll, M. A. and Catapane, E. J. (2007). The nervous system control of lateral ciliary activity of the gill of the bivalve mollusc, *Crassostrea virginica*. *Comparative Biochemistry and Physiology-Part A : Molecular & Integrative Physiology*, 148(2) :445–450.

Cawley, G. C. and Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research*, 99(Complete) :2079–2107.

Chaffee, C. and Lindberg, D. R. (1986). Larval biology of early cambrian molluscs : the implications of small body size. *Bulletin of Marine Science*, 39(2) :536–549.

Chang, Z., Wang, Z., and Li, G. (2014). The impacts of read length and transcriptome complexity for *De Novo* assembly : A simulation study. *PLoS ONE*, 9(4) :e94825.

Chapman, R. W., Mancia, A., Beal, M., Veloso, A., Rathburn, C., Blair, A., Holland, A., Warr, G., Didinato, G., Sokolova, I. M., et al. (2011). The transcriptomic responses of the Eastern oyster, *Crassostrea virginica*, to environmental conditions. *Molecular Ecology*, 20(7) :1431–1449.

Chappell, J. C., Wiley, D. M., and Bautch, V. L. (2011). Regulation of blood vessel sprouting. In *Seminars in cell & developmental biology*, volume 22, pages 1005–1011. Elsevier.

Chi, K. R. (2014). Singled out for sequencing. *Nature Methods*, 11(1) :13–17.

Chia, F.-S. and Rice, M. E. (1978). Settlement and metamorphosis of marine invertebrate larvae. In *Symposium on Settlement and Metamorphosis of Marine Invertebrate Larvae (1977 : Toronto, Ont.)*. Elsevier.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science*, 282(5389) :699–705.

Ciacci, C., Betti, M., Canonico, B., Citterio, B., Roch, P., and Canesi, L. (2010). Specificity of anti-vibrio immune response through p38 MAPK and PKC activation in the hemocytes of the mussel *Mytilus galloprovincialis*. *Journal of Invertebrate Pathology*, 105(1) :49–55.

Clark, M. S., Thorne, M. A., Vieira, F. A., Cardoso, J. C., Power, D. M., and Peck, L. S. (2010). Insights into shell deposition in the Antarctic bivalve *Laternula elliptica* : gene discovery in the mantle transcriptome using 454 pyrosequencing. *Bmc Genomics*, 11(1) :1–12.

Collins, F. S., Green, E. D., Guttmacher, A. E., and Guyer, M. S. (2003). A vision for the future of genomics research. *Nature*, 422(6934) :835–847.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2go : a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18) :3674–3676.

Cooper, S. and Helmstetter, C. E. (1968). Chromosome replication and the division cycle of *Escherichia coli Br*. *Journal of molecular biology*, 31(3) :519–540.

Coppe, A., Bortoluzzi, S., Murari, G., Marino, I. A. M., Zane, L., and Papetti, C. (2012). Sequencing and characterization of striped venus transcriptome expand resources for clam fishery genetics. *PloS one*, 7(9) :e44185.

Costa, F., NóVoa, S., Ojea, J., and Martínez-Patiño, D. (2012). Effects of algal diets and starvation on growth, survival and fatty acid composition of *Solen marginatus* (Bivalvia : Solenidae) larvae. *Scientia Marina*, 76(3) :527–537.

Costa, M., Prado-Alvarez, M., Gestal, C., Li, H., Roch, P., Novoa, B., and Figueras, A. (2009). Functional and molecular immune response of mediterranean mussel (*Mytilus galloprovincialis*) haemocytes against pathogen-associated molecular patterns and bacteria. *Fish & Shellfish Immunology*, 26(3) :515–523.

Cox, R. A. (2004). Quantitative relationships for specific growth rates and macromolecular compositions of *Mycobacterium tuberculosis*, *Streptomyces coelicolor* a3 (2) and *Escherichia coli* B/r : an integrative theoretical approach. *Microbiology*, 150(5) :1413–1426.

Crisp, D. (1978). Genetic consequences of different reproductive strategies in marine invertebrates. *Marine Organisms : Genetics, Ecology and Evolution. Plenum, New York*, pages 257–273.

Croll, R. P., Jackson, D. L., and Voronezhskaya, E. E. (1997). Catecholamine-containing cells in larval and postlarval bivalve molluscs. *The Biological Bulletin*, 193(2) :116–124.

Dam, H. G. (2013). Evolutionary adaptation of marine zooplankton to global change. *Marine Science*, 5.

Dang, C., Tan, T., Moffit, D., Deboutteville, J. D., and Barnes, A. C. (2012). Gender differences in hemocyte immune parameters of bivalves : The sydney rock oyster *Saccostrea glomerata* and the pearl oyster *Pinctada fucata*. *Fish & Shellfish Immunology*, 33(1) :138–142.

Danilova, N. (2006). The evolution of immune mechanisms. *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, 306(6) :496–520.

Darwin, C. (1872). *The origin of species by means of natural selection : Or, the preservation of favoured races in the struggle for life and the descent of man and selection in relation to sex*. Modern Library.

Daugaard, M., Rohde, M., and Jäättelä, M. (2007). The heat shock protein 70 family : Highly homologous proteins with overlapping and distinct functions. *FEBS letters*, 581(19) :3702–3710.

Dawkins, R. (1976). The selfish gene.

Dawkins, R. (2000). *Unweaving the rainbow : Science, delusion and the appetite for wonder*. Houghton Mifflin Harcourt.

Dawkins, R. (2006). *The selfish gene*. Oxford university press.

De Fougerolles, A., Wood, K. M., and Valencia, P. (2013). Formulation and delivery of plga microspheres. US Patent App. 13/897,372.

de Haan, J. R., Wehrens, R., Bauerschmidt, S., Piek, E., van Schaik, R. C., and Buydens, L. M. (2007). Interpretation of ANOVA models for microarray data using PCA. *Bioinformatics*, 23(2) :184–190.

De Lorgeril, J., Zenagui, R., Rosa, R. D., Piquemal, D., and Bachère, E. (2011). Whole transcriptome profiling of successful immune response to *Vibrio* infections in the oyster *Crassostrea gigas* by digital gene expression analysis. *PLoS One*, 6(8) :e23142.

De Petrocellis, L. and Di Marzo, V. (1994). Aquatic invertebrates open up new perspectives in eicosanoid research : biosynthesis and bioactivity. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 51(4) :215–229.

Defer, D., Desriac, F., Henry, J., Bourgougnon, N., Baudy-Floch, M., Brillet, B., Le Chevalier, P., and Fleury, Y. (2013). Antimicrobial peptides in oyster hemolymph : the bacterial connection. *Fish & Shellfish Immunology*, 34(6) :1439–1447.

Delaporte, M., Soudant, P., Lambert, C., Jegaden, M., Moal, J., Pouvreau, S., Dégremont, L., Boudry, P., and Samain, J.-F. (2007). Characterisation of physiological and immunological differences between Pacific oysters (*Crassostrea gigas*) genetically selected for high or low survival to summer mortalities and fed different rations under controlled conditions. *Journal of Experimental Marine Biology and Ecology*, 353(1) :45–57.

Delaporte, M., Soudant, P., Moal, J., Giudicelli, E., Lambert, C., Séguineau, C., and Samain, J.-F. (2006). Impact of 20 :4n- 6 supplementation on the fatty acid composition and hemocyte parameters of the Pacific oyster *Crassostrea gigas*. *Lipids*, 41(6) :567–576.

Delaporte, M., Soudant, P., Moal, J., Lambert, C., Quéré, C., Miner, P., Choquet, G., Paillard, C., and Samain, J.-F. (2003). Effect of a mono-specific algal diet on immune functions in two bivalve species-*Crassostrea gigas* and *Ruditapes philippinarum*. *Journal of Experimental Biology*, 206(17) :3053–3064.

Delaunay, F., Marty, Y., Moal, J., and Samain, J.-F. (1993). The effect of monospecific algal diets on growth and fatty acid composition of *Pecten maximus* (L.) larvae. *Journal of Experimental Marine Biology and Ecology*, 173(2) :163–179.

Diaz de Cerio, O., Hands, E., Humble, J., Cajaraville, M., Craft, J., and Cancio, I. (2013). Construction and characterization of a forward subtracted library of blue mussels *Mytilus edulis* for the identification of gene transcription signatures and biomarkers of styrene exposure. *Marine Pollution Bulletin*, 71(1) :230–239.

Ding, C. and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2) :185–205.

Dirnberger, J. M. (1993). Dispersal of larvae with a short planktonic phase in the polychaete *Spirorbis spirillum* (Linnaeus). *Bulletin of Marine Science*, 52(3) :898–910.

Dixon, P. (2003). Vegan, a package of r functions for community ecology. *Journal of Vegetation Science*, 14(6) :927–930.

Do Carmo, S., Levros Jr, L.-C., and Rassart, E. (2007). Modulation of apolipoprotein d expression and translocation under specific stress conditions. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1773(6) :954–969.

Dobretsov, S. and Wahl, M. (2008). Larval recruitment of the blue mussel *Mytilus edulis* : The effect of flow and algae. *Journal of Experimental Marine Biology and Ecology*, 355(2) :137–144.

Dondero, F., Banni, M., Negri, A., Boatti, L., Dagnino, A., and Viarengo, A. (2011). Interactions of a pesticide/heavy metal mixture in marine bivalves : a transcriptomic assessment. *BMC genomics*, 12(1) :1–18.

Douillet, P. A. and Langdon, C. J. (1994). Use of a probiotic for the culture of larvae of the Pacific oyster (*Crassostrea gigas* thunberg). *Aquaculture*, 119(1) :25–40.

Du, Y., Zhang, L., Xu, F., Huang, B., Zhang, G., and Li, L. (2013). Validation of housekeeping genes as internal controls for studying gene expression during Pacific oyster (*Crassostrea gigas*) development by quantitative real-time PCR. *Fish & Shellfish Immunology*, 34(3) :939–945.

DuBose, A. J., Lichtenstein, S. T., Narisu, N., Bonnycastle, L. L., Swift, A. J., Chines, P. S., and Collins, F. S. (2013). Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene. *Nucleic Acids Research*, 41(6) :e70–e70.

Duran, S., Palacin, C., Becerro, M. A., Turon, X., and Giribet, G. (2004). Genetic diversity and population structure of the commercially harvested sea urchin *Paracentrotus lividus* (Echinodermata, Echinoidea). *Molecular ecology*, 13(11) :3317–3328.

Dyachuk, V. (2013). Extracellular matrix is required for muscle differentiation in primary cell cultures of larval *Mytilus trossulus* (Mollusca : Bivalvia). *Cytotechnology*, 65(5) :1–11.

Dyachuk, V. and Odintsova, N. (2009). Development of the larval muscle system in the mussel *Mytilus trossulus* (Mollusca, Bivalvia). *Development, Growth & Differentiation*, 51(2) :69–79.

Dyrynda, E. A., Pipe, R. K., and Ratcliffe, N. A. (1995). Host defence mechanisms in marine invertebrate larvae. *Fish & Shellfish Immunology*, 5(8) :569–580.

Ekblom, R. and Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1) :1–15.

Elston, R. (1980). Functional morphology of the coelomocytes of the larval oysters (*Crassostrea virginica* and *Crassostrea gigas*). *Marine Biological Association of the United Kingdom*, 60(4) :947–957.

Endler, J. A. (1991). Variation in the appearance of guppy color patterns to guppies and their predators under different visual conditions. *Vision Research*, 31(3) :587–608.

Eschweiler, N. and Christensen, H. T. (2011). Trade-off between increased survival and reduced growth for blue mussels living on Pacific oyster reefs. *Journal of Experimental Marine Biology and Ecology*, 403(1-2) :90–95.

Espinosa, E. P., Perrigault, M., Ward, J. E., Shumway, S. E., and Allam, B. (2009). Lectins associated with the feeding organs of the oyster *Crassostrea virginica* can mediate particle selection. *The Biological Bulletin*, 217(2) :130–141.

Espinosa, E. P., Perrigault, M., Ward, J. E., Shumway, S. E., and Allam, B. (2010). Microalgal cell surface carbohydrates as recognition sites for particle sorting in suspension-feeding bivalves. *The Biological Bulletin*, 218(1) :75–86.

Estévez-Calvar, N., Romero, A., Figueras, A., and Novoa, B. (2011). Involvement of pore-forming molecules in immune defense and development of the Mediterranean mussel (*Mytilus galloprovincialis*). *Developmental & Comparative Immunology*, 35(10) :1017–1031.

Estevez-Calvar, N., Romero, A., Figueras, A., and Novoa, B. (2013). Genes of the mitochondrial apoptotic pathway in *Mytilus galloprovincialis*. *PloS one*, 8(4) :e61502.

Estrada, N., Rodríguez-Jaramillo, C., Contreras, G., and Ascencio, F. (2010). Effects of induced paralysis on hemocytes and tissues of the giant lions-paw scallop by paralyzing shellfish poison. *Marine Biology*, 157(6) :1401–1415.

Evseev, G., Kolotukhina, N., and Kulikova, V. (2011). Morphological structures of the shell of *Mytilus trossulus* and *Crenomytilus grayanus* in early ontogenesis and their importance in the taxonomy of mytilinae (bivalvia : Mytilidae). *Journal of the Marine Biological Association of the United Kingdom*, 91(08) :1655–1664.

Fabbri, E. and Capuzzo, A. (2010). Cyclic amp signaling in bivalve molluscs : an overview. *Journal of Experimental Zoology Part A : Ecological Genetics and Physiology*, 313(4) :179–200.

Fang, D., Xu, G., Hu, Y., Pan, C., Xie, L., and Zhang, R. (2011). Identification of genes directly involved in shell formation and their functions in pearl oyster, *Pinctada fucata*. *PLoS One*, 6(7) :e21860.

Fang, J., Wu, R., Yip, C., and Shin, P. (2009). Power analysis for biomarkers in mussels for use in coastal pollution monitoring. *Marine Pollution Bulletin*, 58(8) :1152–1158.

Faria, M., López, M. A., Fernández-Sanjuan, M., Lacorte, S., and Barata, C. (2010). Comparative toxicity of single and combined mixtures of selected pollutants among larval stages of the native freshwater mussels (*Unio elongatulus*) and the invasive zebra mussel (*Dreissena polymorpha*). *Science of the Total Environment*, 408(12) :2452–2458.

Faure, M. F., David, P., Bonhomme, F., and Bierne, N. (2008). Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evolutionary Biology*, 8(1) :1–14.

Fielman, K. T. and Marsh, A. G. (2005). Genome complexity and repetitive DNA in metazoans from extreme marine environments. *Gene*, 362(Complete) :98–108.

Fisheries, F. (2011). Aquaculture department. 2013. *Global Aquaculture Production Statistics for the year*.

Fitzgerald-Dehoog, L., Browning, J., and Allen, B. J. (2012). Food and heat stress in the california mussel : Evidence for an energetic trade-off between survival and growth. *The Biological Bulletin*, 223(2) :205–216.

Fleury, E., Huvet, A., Lelong, C., De Lorgeril, J., Boulo, V., Gueguen, Y., Bachère, E., Tanguy, A., Moraga, D., Fabioux, C., et al. (2009). Generation and analysis of a 29,745 unique expressed sequence tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database : the gigasdatabase. *BMC genomics*, 10(1) :1–15.

Flye-Sainte-Marie, J., Soudant, P., Lambert, C., Le Goïc, N., Goncalvez, M., Travers, M.-A., Paillard, C., and Jean, F. (2009). Variability of the hemocyte parameters of *Ruditapes philippinarum* in the field during an annual cycle. *Journal of Experimental Marine Biology and Ecology*, 377(1) :1–11.

Fodor, S., Rava, R. P., Huang, X. C., Pease, A. C., Holmes, C. P., and Adams, C. L. (1993). Multiplexed biochemical assays with biological chips. *Nature*, 364(6437) :555–556.

Forrest, A. R. and Carninci, P. (2009). Whole genome transcriptome analysis. *RNA biology*, 6(2) :107–112.

Fowler, J., Cohen, L., and Jarvis, P. (2013). *Practical statistics for field biology*. John Wiley & Sons.

Fox, C. W., Roff, D. A., and Ecology, D. F. E. (2001). *Concepts and case studies*. Oxford University Press.

Fraser, B. A., Janowitz, I., Thairu, M., Travis, J., and Hughes, K. A. (2014). Phenotypic and genomic plasticity of alternative male reproductive tactics in sailfin mollies. *Proceedings of the Royal Society B : Biological Sciences*, 281(1781) :20132310.

Furuhashi, M. and Hotamisligil, G. S. (2008). Fatty acid-binding proteins : role in metabolic diseases and potential as drug targets. *Nature Reviews Drug Discovery*, 7(6) :489–503.

Futschik, M. E. and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. *Journal of Bioinformatics and Computational Biology*, 3(04) :965–988.

Gagné, R., Tremblay, R., Pernet, F., Miner, P., Samain, J.-F., and Olivier, F. (2010). Lipid requirements of the scallop *Pecten maimus* (L.) during larval and post-larval development in relation to addition of *Rhodomonas salina* in diet. *Aquaculture*, 309(1-4) :212–221.

Gallager, S. M., Mann, R., and Sasaki, G. C. (1986). Lipid as an index of growth and viability in three species of bivalve larvae. *Aquaculture*, 56(2) :81–103.

Gallardo-Escárate, C., Valenzuela-Muñoz, V., Núñez-Acuña, G., and Haye, P. (2013). SNP discovery and gene annotation in the surf clam *Mesodesma donacium*. *Aquaculture Research*.

Gao, Q., Zhao, J., Song, L., Qiu, L., Yu, Y., Zhang, H., and Ni, D. (2008). Molecular cloning, characterization and expression of heat shock protein 90 gene in the haemocytes of bay scallop *Argopecten irradians*. *Fish & Shellfish Immunology*, 24(4) :379–385.

García-Lavandeira, M., Silva, A., Abad, M., Pazos, A. J., Sánchez, J. L., and Luz Pérez-Parallé, M. (2005). Effects of GABA and epinephrine on the settlement and metamorphosis of the larvae of four species of bivalve molluscs. *Journal of Experimental Marine Biology and Ecology*, 316(2) :149–156.

García-Esquivel, Z., Bricelj, V. M., and González-Gómez, M. A. (2001). Physiological basis for energy demands and early postlarval mortality in the Pacific oyster, *Crassostrea gigas*. *Journal of Experimental Marine Biology and Ecology*, 263(1) :77–103.

García-García, E., Prado-Álvarez, M., Novoa, B., Figueras, A., and Rosales, C. (2008). Immune responses of mussel hemocyte subpopulations are differentially regulated by enzymes of the PI 3-K, PKC, and ERK kinase families. *Developmental & Comparative Immunology*, 32(6) :637–653.

García-Lavandeira, M., Silva, A., Abad, M., Pazos, A. J., Sánchez, J. L., and Luz Pérez-Parallé, M. (2005). Effects of gaba and epinephrine on the settlement and metamorphosis of the larvae of four species of bivalve molluscs. *Journal of Experimental Marine Biology and Ecology*, 316(2) :149–156.

Gardner, A. and Welch, J. J. (2011). A formal theory of the selfish gene. *Journal of Evolutionary Biology*, 24(8) :1801–1813.

Garg, R. and Jain, M. (2013). RNA-Seq for transcriptome analysis in non-model plants. In *Legume Genomics*, pages 43–58. Springer.

Gauthier-Clerc, S., Boily, I., Fournier, M., and Lemarchand, K. (2013). *In vivo* exposure of *Mytilus edulis* to living enteric bacteria : a threat for immune competency ? *Environmental Science and Pollution Research*, 20(2) :612–620.

Gavery, M. R. and Roberts, S. B. (2014). A context dependent role for DNA methylation in bivalves. *Briefings in functional genomics*, page elt054.

Genard, B., Miner, P., Nicolas, J.-L., Moraga, D., Boudry, P., Pernet, F., and Tremblay, R. (2013). Integrative study of physiological changes associated with bacterial infection in Pacific oyster larvae. *PloS one*, 8(5) :e64534.

Genard, B., Moraga, D., Pernet, F., David, É., Boudry, P., and Tremblay, R. (2012). Expression of candidate genes related to metabolism, immunity and cellular stress during massive mortality in the American oyster *Crassostrea virginica* larvae in relation to biochemical and physiological parameters. *Gene*, 499(1) :70–75.

Genard, B., Pernet, F., Lemarchand, K., Boudry, P., Moraga, D., Tremblay, R., et al. (2011). Physiological and biochemical changes associated with massive mortality events occurring in larvae of American oyster (*Crassostrea virginica*). *Aquatic Living Resources*, 24(3) :247–260.

Gentleman, R. C., Carey, V. J., Bates, D. M., and others (2004). Bioconductor : Open software development for computational biology and bioinformatics. *Genome Biology*, 5 :R80.

Gerdol, M., Manfrin, C., De Moro, G., Figueras, A., Novoa, B., Venier, P., and Pallavicini, A. (2011). The C1q domain containing proteins of the Mediterranean mussel *Mytilus galloprovincialis* : A widespread and diverse family of immune-related molecules. *Developmental & Comparative Immunology*, 35(6) :635–643.

Gestal, C., Roch, P., Renault, T., Pallavicini, A., Paillard, C., Novoa, B., Oubella, R., Venier, P., and Figueras, A. (2008). Study of diseases and the immune system of bivalves using molecular biology and genomics. *Reviews in Fisheries Science*, 16(sup1) :133–156.

Ghosh, G., Wang, V. Y.-F., Huang, D.-B., and Fusco, A. (2012). NF-κB regulation : lessons from structures. *Immunological reviews*, 246(1) :36–58.

Gibson, L., Woodworth, J., and George, A. (1998). Probiotic activity of *Aeromonas media* on the Pacific oyster, *Crassostrea gigas*, when challenged with *Vibrio tubiashii*. *Aquaculture*, 169(1) :111–120.

Gilad, Y., Pritchard, J. K., and Thornton, K. (2009). Characterizing natural variation using next-generation sequencing technologies. *Trends in Genetics*, 25(10) :463–471.

Giménez, L. (2010). Relationships between habitat conditions, larval traits, and juvenile performance in a marine invertebrate. *Ecology*, 91(5) :1401–1413.

Gireesh, R., Biju, A., and Muthiah, P. (2009). Biochemical changes during larval development in the short neck clam, *Paphia malabarica Chemnitz*. *Aquaculture Research*, 40(13) :1510–1515.

Glencross, B. D. (2009). Exploring the nutritional demand for essential fatty acids by aquaculture species. *Reviews in Aquaculture*, 1(2) :71–124.

Gobler, C. and Talmage, S. (2013). Short-and long-term consequences of larval stage exposure to constantly and ephemerally elevated carbon dioxide for marine bivalve populations. *Biogeosciences*, 10(4) :2241–2253.

Gobler, C. J., DePasquale, E. L., Griffith, A. W., and Baumann, H. (2014). Hypoxia and acidification have additive and synergistic negative effects on the growth, survival, and metamorphosis of early life stage bivalves. *PloS one*, 9(1) :e83648.

Godin, J.-G. J. and Briggs, S. E. (1996). Female mate choice under predation risk in the guppy. *Animal Behaviour*, 51(1) :117–130.

Gomez-Chiarri, M. (2013). The transcriptome of the eastern oyster *Crassostrea virginica* in response to bacterial challenge. In *Plant and Animal Genome XXI Conference*. Plant and Animal Genome.

Gomez-Gil, B., Roque, A., and Turnbull, J. F. (2000). The use and selection of probiotic bacteria for use in the culture of larval aquatic organisms. *Aquaculture*, 191(1) :259–270.

Gong, T., Hartmann, N., Kohane, I. S., Brinkmann, V., Staedtler, F., Letzkus, M., Bongiovanni, S., and Szustakowski, J. D. (2011). Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples. *PloS one*, 6(11) :e27156.

Gonzalez, M., Gueguen, Y., Destoumieux-Garzón, D., Romestand, B., Fievet, J., Pugnière, M., Roquet, F., Escoubas, J.-M., Vandenbulcke, F., Levy, O., et al. (2007). Evidence of a bactericidal permeability increasing protein in an invertebrate, the *Crassostrea gigas* Cg-BPI. *Proceedings of the National Academy of Sciences*, 104(45) :17759–17764.

Gosling, E. (1992). *The mussel Mytilus : ecology, physiology, genetics and culture*. Elsevier.

Gosselin, L. A. and Qian, P.-Y. (1997). Juvenile mortality in benthic marine invertebrates. *Marine Ecology Progress Series*, 146(1) :265–282.

Gould, S. J. and Eldredge, N. (1977). Punctuated equilibria : the tempo and mode of evolution reconsidered. *Paleobiology*, 3(2) :115–151.

Gould, S. J. and Lloyd, E. A. (1999). Individuality and adaptation across levels of selection : How shall we name and generalize the unit of darwinism ? *Proceedings of the National Academy of Sciences*, 96(21) :11904–11909.

Gourdine, J.-P. and Smith-Ravin, E. J. (2007). Analysis of a cDNA-derived sequence of a novel mannose-binding lectin, codakine, from the tropical clam *Codakia orbicularis*. *Fish & Shellfish Immunology*, 22(5) :498–509.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7) :644–652.

Gracey, A. Y. and Cossins, A. R. (2003). Application of microarray technology in environmental and comparative physiology. *Annual Review of Physiology*, 65(1) :231–259.

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology*, 238(3) :541–563.

Granmo, Å., Ekelund, R., Magnusson, K., and Berggren, M. (1989). Lethal and sublethal toxicity of 4-nonylphenol to the common mussel (*Mytilus edulis* L.). *Environmental Pollution*, 59(2) :115–127.

Green, D. W., Padula, M. P., Santos, J., Chou, J., Milthorpe, B., and Ben-Nissan, B. (2013). A therapeutic potential for marine skeletal proteins in bone regeneration. *Marine Drugs*, 11(4) :1203–1220.

Green, T. J. and Montagnani, C. (2013). Poly I :C induces a protective antiviral immune response in the Pacific oyster (*Crassostrea gigas*) against subsequent challenge with *Ostreid herpesvirus* (OsHV-1 µvar). *Fish & Shellfish Immunology*, 35 :382–388.

Greenacre, M. and Blasius, J. (2006). *Multiple correspondence analysis and related methods*. CRC Press.

Gregorio, C. C., Perry, C. N., and McElhinny, A. S. (2005). Functional properties of the titin/connectin-associated proteins, the muscle-specific ring finger proteins (murfs), in striated muscle. *Journal of Muscle Research & Cell Motility*, 26(6-8) :389–400.

Gueguen, Y., Cadoret, J.-P., Flament, D., Barreau-Roumiguière, C., Girardot, A.-L., Garnier, J., Hoareau, A., Bachère, E., and Escoubas, J.-M. (2003). Immune gene discovery by expressed sequence tags generated from hemocytes of the bacteria-challenged oyster, *Crassostrea gigas*. *Gene*, 303(1/2) :139–145.

Guerra, C., Zenteno-Savín, T., Maeda-Martínez, A., Philipp, E., and Abele, D. (2012). Changes in oxidative stress parameters in relation to age, growth and reproduction in the short-lived catarina scallop *Argopecten ventricosus* reared in its natural environment. *Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology*, 162(4) :421–430.

Guillard, R. R. (1975). Culture of phytoplankton for feeding marine invertebrates. In *Culture of marine invertebrate animals*, pages 29–60. Springer.

Hadfield, M. G. (2011). Biofilms and marine invertebrate larvae : what bacteria produce that larvae use to choose settlement sites. *Annual Review of Marine Science*, 3(1) :453–470.

Hadfield, M. G., Carpizo-Ituarte, E. J., del Carmen, K., and Nedved, B. T. (2001). Metamorphic competence, a major adaptive convergence in marine invertebrate larvae. *American Zoologist*, 41(5) :1123–1131.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, 228(5277) :1218–1220.

Hamilton, W. D. (1972). Altruism and related phenomena, mainly in social insects. *Annual Review of Ecology and Systematics*, 3(1) :193–232.

Harbige, L. S. (2003). Fatty acids, the immune response, and autoimmunity : a question of n- 6 essentiality and the balance between n- 6 and n- 3. *Lipids*, 38(4) :323–341.

Hare, J. A. and Cowen, R. K. (1997). Size, growth, development, and survival of the planktonic larvae of *Pomatomus saltatrix* (Pisces : Pomatomidae). *Ecology*, 78(8) :2415–2431.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning : data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2) :83–85.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 2. Springer.

Hay, M. E. (2009). Marine chemical ecology : chemical signals and cues structure marine populations, communities, and ecosystems. *Annual Review of Marine Science*, 1(1) :193–212.

He, C.-b., Wang, Y., Liu, W.-d., Gao, X.-g., Chen, P.-h., Li, Y.-f., and Bao, X.-b. (2013). Cloning, promoter analysis and expression of the tumor necrosis factor receptor-associated factor 6 (TRAF6) in Japanese scallop (*Mizuhopecten yessoensis*). *Molecular Biology Reports*, 40(8) :4769–4779.

Hedgecock, D. (1986). Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates ? *Bulletin of Marine Science*, 39(2) :550–564.

Hemaiswarya, S., Raja, R., Kumar, R. R., Ganesan, V., and Anbazhagan, C. (2011). Microalgae : a sustainable feed source for aquaculture. *World Journal of Microbiology and Biotechnology*, 27(8) :1737–1746.

Hendriks, I. E., van Duren, L. A., and Herman, P. M. (2003). Effect of dietary polyunsaturated fatty acids on reproductive output and larval growth of bivalves. *Journal of Experimental Marine Biology and Ecology*, 296(2) :199–213.

Hens, M. D., Fowler, K. A., and Leise, E. M. (2006). Induction of metamorphosis decreases nitric oxide synthase gene expression in larvae of the marine mollusc *Ilyanassa obsoleta* (Say). *The Biological Bulletin*, 211(3) :208–211.

Hermann, A. and Cox, J. A. (1995). Sarcoplasmic calcium-binding protein. *Comparative Biochemistry and Physiology Part B : Biochemistry and Molecular Biology*, 111(3) :337–345.

Herpin, A., Lelong, C., Becker, T., Rosa, F., Favrel, P., and Cunningham, C. (2005). Structural and functional evidences for a type 1 TGF-β sensu stricto receptor in the lophotrochozoan *Crassostrea gigas* suggest conserved molecular mechanisms controlling mesodermal patterning across bilateria. *Mechanisms of Development*, 122(5) :695–705.

Heyland, A. and Moroz, L. L. (2006). Signaling mechanisms underlying metamorphic transitions in animals. *Integrative and Comparative Biology*, 46(6) :743–759.

Highsmith, R. C. and Emlet, R. B. (1986). Delayed metamorphosis : effect on growth and survival of juvenile sand dollars (echinoidea : Clypeasteroida). *Bulletin of Marine Science*, 39(2) :347–361.

Hines, A. H. (1986). Larval problems and perspectives in life histories of marine invertebrates. *Bulletin of Marine Science*, 39(2) :506–525.

His, E., Robert, R., and Dinet, A. (1989). Combined effects of temperature and salinity on fed and starved larvae of the Mediterranean mussel *Mytilus galloprovincialis* and the Japanese oyster *Crassostrea gigas*. *Marine Biology*, 100(4) :455–463.

Ho, L. and Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, 463(7280) :474–484.

Hodin, J. (2006). Expanding networks : Signaling components in and a hypothesis for the evolution of metamorphosis. *Integrative and Comparative Biology*, 46(6) :719–742.

Holland, D. (1978). Lipid reserves and energy metabolism in the larvae of benthic marine invertebrates. *Biochemical and Biophysical Perspectives in Marine Biology*, 4(1) :85–123.

Holland, D. and Spencer, B. (1973). Biochemical changes in fed and starved oysters, *Ostrea edulis* l. during larval development, metamorphosis and early spat growth. *Journal of the Marine Biological Association of the United Kingdom*, 53(2) :287–298.

Hornett, E. A. and Wheat, C. W. (2012). Quantitative RNA-Seq analysis in non-model species : assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. *BMC genomics*, 13(1) :1–16.

Hou, R., Bao, Z., Wang, S., Su, H., Li, Y., Du, H., Hu, J., Wang, S., and Hu, X. (2011). Transcriptome sequencing and de novo analysis for Yesso scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PloS one*, 6(6) :e21560.

Howard, R. W. and Stanley, D. W. (1999). The tie that binds : Eicosanoids in invertebrate biology. *Annals of the Entomological Society of America*, 92(6) :880–890.

Huan, P., Wang, H., Dong, B., and Liu, B. (2012a). Identification of differentially expressed proteins involved in the early larval development of the Pacific oyster *Crassostrea gigas*. *Journal of Proteomics*, 75(13) :3855–3865.

Huan, P., Wang, H., and Liu, B. (2012b). Transcriptomic analysis of the clam *Meretrix meretrix* on different larval stages. *Marine Biotechnology*, 14(1) :69–78.

Huang, X.-D., Zhao, M., Liu, W.-G., Guan, Y.-Y., Shi, Y., Wang, Q., Wu, S.-Z., and He, M.-X. (2012). Gigabase-scale transcriptome analysis on four species of pearl oysters. *Marine Biotechnology*, pages 1–12.

Hughes, T., Smith, E., Leung, M., Stefano, G., et al. (1992). Immunoreactive cytokines in *Mytilus edulis* nervous and immune interactions. *Acta Biologica Hungarica*, 43(1-4) :269–273.

Hughes Jr, T. K., Smith, E. M., Chin, R., Cadet, P., Sinisterra, J., Leung, M. K., Shipp, M. A., Scharrer, B., and Stefano, G. B. (1990). Interaction of immunoactive monokines (interleukin 1 and tumor necrosis factor) in the bivalve mollusc *Mytilus edulis*. *Proceedings of the National Academy of Sciences*, 87(12) :4426–4429.

Hulbert, A., Pamplona, R., Buffenstein, R., and Buttemer, W. (2007). Life and death : metabolic rate, membrane composition, and life span of animals. *Physiological Reviews*, 87(4) :1175–1213.

Hurtado, M. A., Reza, M., Ibarra, A. M., Wille, M., Sorgeloos, P., Soudant, P., and Palacios, E. (2009). Arachidonic acid (20 : 4 *n*- 6) effect on reproduction, immunology, and prostaglandin e2 levels in *Crassostrea corteziensis* (Hertlein, 1951). *Aquaculture*, 294(3) :300–305.

Husmann, G., Abele, D., Rosenstiel, P., Clark, M., Kraemer, L., and Philipp, E. (2013). Age-dependent expression of stress and antimicrobial genes in the hemocytes and siphon tissue of the Antarctic bivalve, *Laternula elliptica*, exposed to injury and starvation. *Cell Stress and Chaperones*, 19(1) :15–32.

Itoh, N. and Takahashi, K. G. (2009). A novel peptidoglycan recognition protein containing a goose-type lysozyme domain from the Pacific oyster, *Crassostrea gigas*. *Molecular immunology*, 46(8) :1768–1774.

Izquierdo, M. S. (1996). Essential fatty acid requirements of cultured marine fish larvae. *Aquaculture Nutrition*, 2(4) :183–191.

Jablonski, D. (1986). Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Science*, 39(2) :565–587.

Jackson, G. A. and Strathmann, R. R. (1981). Larval mortality from offshore mixing as a link between precompetent and competent periods of development. *American Naturalist*, pages 16–26.

Jayaraj, S., Thiagarajan, R., Arumugam, M., and Mullainadhan, P. (2008). Isolation, purification and characterization of β-1, 3-glucan binding protein from the plasma of marine mussel *Perna viridis*. *Fish & Shellfish Immunology*, 24(6) :715–725.

Je Lee, H. and Boulding, E. G. (2009). Spatial and temporal population genetic structure of four northeastern Pacific littorinid gastropods : the effect of mode of larval development on variation at one mitochondrial and two nuclear DNA markers. *Molecular Ecology*, 18(10) :2165–2184.

Jenny, M. J., Ringwood, A. H., Lacy, E. R., Lewitus, A. J., Kempton, J. W., Gross, P. S., Warr, G. W., and Chapman, R. W. (2002). Potential indicators of stress response identified by expressed sequence tag analysis of hemocytes and embryos from the American oyster, *Crassostrea virginica*. *Marine biotechnology*, 4(1) :81–93.

Ji, C., Wu, H., Wei, L., Zhao, J., Wang, Q., and Lu, H. (2013a). Responses of *Mytilus galloprovincialis* to bacterial challenges by metabolomics and proteomics. *Fish & Shellfish Immunology*, 35(2) :489–498.

Ji, C., Wu, H., Wei, L., Zhao, J., and Yu, J. (2013b). Proteomic and metabolomic analysis reveal gender-specific responses of mussel *Mytilus galloprovincialis* to 2, 2', 4, 4'-tetrabromodiphenyl ether (BDE 47). *Aquatic Toxicology*, 140-141(1) :449–457.

Jing, X., Espinosa, E. P., Perrigault, M., and Allam, B. (2011). Identification, molecular characterization and expression analysis of a mucosal C-type lectin in the Eastern oyster, *Crassostrea virginica*. *Fish & Shellfish Immunology*, 30(3) :851–858.

Johnson, C. N. (2013). Fitness factors in vibrios : a Mini-review. *Microbial Ecology*, 65(4) :826–851.

Johnson, G. L. and Lapadat, R. (2002). Mitogen-activated protein kinase pathways mediated by erk, jnk, and p38 protein kinases. *Science*, 298(5600) :1911–1912.

Jones, D. B., Jerry, D. R., Forêt, S., Konovalov, D. A., and Zenger, K. R. (2013). Genome-Wide SNP validation and mantle tissue transcriptome analysis in the silver-lipped Pearl oyster, *Pinctada maxima*. *Marine Biotechnology*, 15(6) :1–12.

Jouaux, A., Lafont, M., Blin, J.-L., Houssin, M., Mathieu, M., and Lelong, C. (2013). Physiological change under OsHV-1 contamination in Pacific oyster *Crassostrea gigas* through massive mortality events on fields. *BMC genomics*, 14(1) :590.

Kang, Y.-S., Kim, Y.-M., Park, K.-I., Kim Cho, S., Choi, K.-S., and Cho, M. (2006). Analysis of EST and lectin expressions in hemocytes of Manila clams (*Ruditapes philippinarum*)(bivalvia : Mollusca) infected with *Perkinsus olseni*. *Developmental & Comparative Immunology*, 30(12) :1119–1131.

Katayama, S., Töhönen, V., Linnarsson, S., and Kere, J. (2013). SAMstrt : Statistical test for differential expression in single-cell transcriptome with spike-in normalization. *Bioinformatics*, 29(22) :2943–2945.

Kaustuv, R., Jablonski, D., and Valentine, J. W. (2001). Climate change, species range limits and body size in marine bivalves. *Ecology Letters*, 4(4) :366–370.

Kawai, T. and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity : update on toll-like receptors. *Nature immunology*, 11(5) :373–384.

Kesarcodi-Watson, A., Miner, P., Nicolas, J.-L., Robert, R., et al. (2012). Protective effect of four potential probiotics against pathogen-challenge of the larvae of three bivalves : Pacific oyster (*Crassostrea gigas*), flat oyster (*Ostrea edulis*) and scallop (*Pecten maximus*). *Aquaculture*, 344(Complete) :29–34.

Kheder, R. B., Moal, J., and Robert, R. (2010). Impact of temperature on larval development and evolution of physiological indices in *Crassostrea gigas*. *Aquaculture*, 309(1-4) :286–289.

Kim, M. H., Yoo, D. S., Lee, S. Y., Byeon, S. E., Lee, Y. G., Min, T., Rho, H. S., Rhee, M. H., Lee, J., and Cho, J. Y. (2011). The TRIF/TBK1/IRF-3 activation pathway is the primary inhibitory target of resveratrol, contributing to its broad-spectrum anti-inflammatory effects. *Die Pharmazie-An International Journal of Pharmaceutical Sciences*, 66(4) :293–300.

Kingsford, M. J., Leis, J. M., Shanks, A., Lindeman, K. C., Morgan, S. G., and Pineda, J. (2002). Sensory environments, larval abilities and local self-recruitment. *Bulletin of Marine Science*, 70(Supplement 1) :309–340.

Kiss, T. (2010). Apoptosis and its functional significance in molluscs. *Apoptosis*, 15(3) :313–321.

Knight, J., Rowley, A. F., Yamazaki, M., and Clare, A. S. (2000). Eicosanoids are modulators of larval settlement in the barnacle, *Balanus amphitrite*. *Journal of the Marine Biological Association of the UK*, 80(01) :113–117.

Kodirov, S. A. (2011). The neuronal control of cardiac functions in molluscs. *Comparative Biochemistry and Physiology-Part A : Molecular & Integrative Physiology*, 160(2) :102–116.

Koehn, R., Bayne, B., Moore, M., and Siebenaller, J. (1980). Salinity related physiological and genetic differences between populations of *Mytilus edulis*. *Biological Journal of the Linnean Society*, 14(3-4) :319–334.

Kotsiantis, S. B. (2007). Supervised machine learning : a review of classification techniques. *Informatica (03505596)*, 31(3) :249–268.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4) :693–705.

Kraeuter, J. N., Castagna, M., and van Dessel, R. (1981). Egg size and larval survival of *Mercenaria mercenaria* (L.) and *Argopecten irradians* (Lamarck). *Journal of Experimental Marine Biology and Ecology*, 56(1) :3–8.

Kreft, J.-U. and Bonhoeffer, S. (2005). The evolution of groups of cooperating bacteria and the growth rate versus yield trade-off. *Microbiology*, 151(3) :637–641.

Krogh, A. and Sollich, P. (1997). Statistical mechanics of ensemble learning. *Physical Review E*, 55(1) :811.

Kuchel, R. P., Aladaileh, S., Birch, D., Vella, N., and Raftos, D. A. (2010). Phagocytosis of the protozoan parasite, *Marteilia sydneyi*, by sydney rock oyster (*Saccostrea glomerata*) hemocytes. *Journal of Invertebrate Pathology*, 104(2) :97–104.

Kumar, L. and Futschik, M. E. (2007). Mfuzz : a software package for soft clustering of microarray data. *Bioinformation*, 2(1) :5–7.

Labreuche, Y., Lambert, C., Soudant, P., Boulo, V., Huvet, A., and Nicolas, J.-L. (2006). Cellular and molecular hemocyte responses of the Pacific oyster, *Crassostrea gigas*, following bacterial infection with *Vibrio aestuarianus* strain 01/32. *Microbes and Infection*, 8(12-13) :2715–2724.

Laing, I., Utting, S., and Kilada, R. (1987). Interactive effect of diet and temperature on the growth of juvenile clams. *Journal of Experimental Marine Biology and Ecology*, 113(1) :23–38.

Lambert, C., Soudant, P., Jegaden, M., Delaporte, M., Labreuche, Y., Moal, J., and Samain, J.-F. (2007). *In vitro* modulation of reactive oxygen and nitrogen intermediate (ROI/RNI) production in *Crassostrea gigas* hemocytes. *Aquaculture*, 270(1) :413–421.

Lau, S. C., Thiyagarajan, V., Cheung, S. C., and Qian, P.-Y. (2005). Roles of bacterial community composition in biofilms as a mediator for larval settlement of three marine invertebrates. *Aquatic Microbial Ecology*, 38(1) :41–51.

Lawrence, J. M., McClintock, J. B., and Guille, A. (1984). Organic level and caloric content of eggs of brooding asteroids and an echinoid (Echinodermata) from Kerguelen (South Indian Ocean). *International Journal of Invertebrate Reproduction and Development*, 7(4) :249–257.

Lee, Y., Whang, I., Umasuthan, N., De Zoysa, M., Oh, C., Kang, D.-H., Choi, C. Y., Park, C.-J., and Lee, J. (2011). Characterization of a novel molluscan MyD88 family protein from manila clam, *Ruditapes philippinarum*. *Fish & Shellfish Immunology*, 31(6) :887–893.

Lee, Y. C., Min, S., Gim, B. S., and Kim, Y.-J. (1997). A transcriptional mediator protein that is required for activation of many RNA polymerase II promoters and is conserved from yeast to humans. *Molecular and Cellular Biology*, 17(8) :4622–4632.

Lelong, C., Badariotti, F., Le Quere, H., Rodet, F., Dubos, M., and Favrel, P. (2007). Cg-TGF-β, a TGF-β activin homologue in the Pacific oyster *Crassostrea gigas* , is involved in immunity against Gram-negative microbial infection. *Developmental & Comparative Immunology*, 31(1) :30–38.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4) :183–189.

Leu, Y.-W., Rahmatpanah, F., Shi, H., Wei, S. H., Liu, J. C., Yan, P. S., and Huang, T. H.-M. (2003). Double RNA interference of DNMT3b and DNMT1 enhances DNA demethylation and gene reactivation. *Cancer Research*, 63(19) :6110–6115.

Levin, L. A. (2006). Recent progress in understanding larval dispersal : new directions and digressions. *Integrative and Comparative Biology*, 46(3) :282–297.

Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, 1(1) :1–18.

Li, H.-j., Yang, Q., Gao, X.-g., Su, H., Wang, J., and He, C.-b. (2012). Identification and expression of a putative LPS-induced TNF-α factor from Asiatic hard clam *Meretrix meretrix*. *Molecular Biology Reports*, 39(2) :865–871.

Li, S., Xie, L., Ma, Z., and Zhang, R. (2005). cDNA cloning and characterization of a novel calmodulin-like protein from pearl oyster *Pinctada fucata*. *Febs Journal*, 272(19) :4899–4910.

Li, Z., Yang, A., Chen, X., Zeng, L., and Cao, T. (2014). A composite method for feature selection of microarray data. *Journal of Computational and Theoretical Nanoscience*, 11(2) :472–476.

Lilly, D. M. and Stillwell, R. H. (1965). Probiotics : growth-promoting factors produced by microorganisms. *Science*, 147(3659) :747–748.

Liu, H., Qin, C., Chen, Z., Zuo, T., Yang, X., Zhou, H., Xu, M., Cao, S., Shen, Y., Lin, H., et al. (2014). Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. *BMC genomics*, 15(1) :15–25.

Liu, W., He, C., Li, W., Zhou, Z., Gao, X., and Fu, L. (2010). Discovery of host defence genes in the japanese scallop *Mizuhopecten yessoensis* jay by expressed sequence tag analysis of kidney tissue. *Aquaculture Research*, 41(11) :1602–1613.

Liu, Y., Tam, N. F., Guan, Y., Yasojima, M., Zhou, J., and Gao, B. (2011). Acute toxicity of nonylphenols and bisphenol A to the embryonic development of the abalone *Haliotis diversicolor supertexta*. *Ecotoxicology*, 20(6) :1233–1245.

Lockwood, B. L. and Somero, G. N. (2011). Transcriptomic responses to salinity stress in invasive and native blue mussels (genus Mytilus). *Molecular Ecology*, 20(3) :517–529.

Lucas, J. M., Vaccaro, E., and Waite, J. H. (2002). A molecular, morphometric and mechanical comparison of the structural elements of byssus from *Mytilus edulis* and *Mytilus galloprovincialis*. *Journal of Experimental Biology*, 205(12) :1807–1817.

Luedeking, A. and Koehler, A. (2004). Regulation of expression of multixenobiotic resistance (MXR) genes by environmental factors in the blue mussel *Mytilus edulis*. *Aquatic toxicology*, 69(1) :1–10.

Luna-Gonzalez, A., Maeda-Martínez, A., Sainz, J., and Ascencio-Valle, F. (2002). Comparative susceptibility of veliger larvae of four bivalve mollusks to a *Vibrio alginolyticus* strain. *Diseases of aquatic organisms*, 49(3) :221–226.

Luna-González, A., Maeda-Martınez, A. N., Ascencio-Valle, F., and Robles-Mungaray, M. (2004). Ontogenetic variations of hydrolytic enzymes in the Pacific oyster *Crassostrea gigas*. *Fish & Shellfish Immunology*, 16(3) :287–294.

Magie, C. R., Meyer, M. R., Gorsuch, M. S., and Parkhurst, S. M. (1999). Mutations in the rho1 small gtpase disrupt morphogenesis and segmentation during early drosophila development. *Development*, 126(23) :5353–5364.

Malagoli, D., Casarini, L., Sacchi, S., and Ottaviani, E. (2007). Stress and immune response in the mussel *Mytilus galloprovincialis*. *Fish & Shellfish Immunology*, 23(1) :171–177.

Malfait, F., Symoens, S., De Backer, J., Hermanns-Lê, T., Sakalihasan, N., Lapiere, C. M., Coucke, P., and De Paepe, A. (2007). Three arginine to cysteine substitutions in the pro-alpha (i)-collagen chain cause ehlers-danlos syndrome with a propensity to arterial rupture in early adulthood. *Human Mutation*, 28(4) :387–395.

Manahan, D. T. (1990). Adaptations by invertebrate larvae for nutrient acquisition from seawater. *American Zoologist*, 30(1) :147–160.

Manoj Nair, R. and Appukuttan, K. (2003). Effect of temperature on the development, growth, survival and settlement of green mussel *Perna viridis* (Linnaeus, 1758). *Aquaculture Research*, 34(12) :1037–1045.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24(3) :133–141.

Marois, R. and Carew, T. J. (1997). Ontogeny of serotonergic neurons in *Aplysia californica*. *Journal of Comparative Neurology*, 386(3) :477–490.

Marshall, D. J. and Bolton, T. F. (2007). Effects of egg size on the development time of non-feeding larvae. *The Biological Bulletin*, 212(1) :6–11.

Marshall, D. J., Bolton, T. F., and Keough, M. J. (2003). Offspring size affects the post-metamorphic performance of a colonial marine invertebrate. *Ecology*, 84(12) :3131–3137.

Marshall, D. J. and Keough, M. J. (2007). The evolutionary ecology of offspring size in marine invertebrates. *Advances in Marine Biology*, 53(Complete) :1–60.

Marshall, R., McKinley, S., and Pearce, C. M. (2010). Effects of nutrition on larval growth and survival in bivalves. *Reviews in Aquaculture*, 2(1) :33–55.

Martel, A. and Chia, F.-S. (1991). Drifting and dispersal of small bivalves and gastropods with direct development. *Journal of Experimental Marine Biology and Ecology*, 150(1) :131–147.

Martin, S., Richier, S., Pedrotti, M.-L., Dupont, S., Castejon, C., Gerakis, Y., Kerros, M.-E., Oberhänsli, F., Teyssié, J.-L., Jeffree, R., et al. (2011). Early development and molecular plasticity in the mediterranean sea urchin *Paracentrotus lividus* exposed to CO2-driven acidification. *The Journal of Experimental Biology*, 214(8) :1357–1368.

Martín-Gómez, L., Villalba, A., Carballal, M. J., and Abollo, E. (2013). Identification of relevant cancer related-genes in the flat oyster *Ostrea edulis* affected by disseminated neoplasia. *Marine Biotechnology*, 15(2) :159–174.

Mateo, D. R., Greenwood, S. J., Araya, M. T., Berthe, F. C., Johnson, G. R., and Siah, A. (2010). Differential gene expression of gamma-actin, toll-like receptor 2 (TLR-2) and interleukin-1 receptor-associated kinase 4 (IRAK-4) in *Mya arenaria* haemocytes induced by *in vivo* infections with two *Vibrio splendidus* strains. *Developmental & Comparative Immunology*, 34(7) :710–714.

McEdward, L. R. and Janies, D. A. (1997). Relationships among development, ecology, and morphology in the evolution of echinoderm larvae and life cycles. *Biological Journal of the Linnean Society*, 60(3) :381–400.

McKinney, B. A., Reif, D. M., Ritchie, M. D., and Moore, J. H. (2006). Machine learning for detecting gene-gene interactions. *Applied bioinformatics*, 5(2) :77–88.

Medina, M. (2009). Functional genomics opens doors to understanding metamorphosis in nonmodel invertebrate organisms. *Molecular Ecology*, 18(5) :763–764.

Meng, J., Zhu, Q., Zhang, L., Li, C., Li, L., She, Z., Huang, B., and Zhang, G. (2013). Genome and transcriptome analyses provide insight into the euryhaline adaptation mechanism of *Crassostrea gigas*. *PloS one*, 8(3) :e58563.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11(1) :31–46.

Meyer, D., Leisch, F., and Hornik, K. (2003). The support vector machine under test. *Neurocomputing*, 55(1) :169–186.

Meyer, E. and Manahan, D. T. (2010). Gene expression profiling of genetically determined growth variation in bivalve larvae (*Crassostrea gigas*). *Journal of Experimental Biology*, 213(5) :749–758.

Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., et al. (2008). The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, 9(1) :1–8.

Mikami, T., Miyashita, H., Takatsuka, S., Kuroki, Y., and Matsushima, N. (2012). Molecular evolution of vertebrate toll-like receptors : evolutionary rate difference between their leucine-rich repeats and their tir domains. *Gene*, 503(2) :235–243.

Milan, M., Coppe, A., Reinhardt, R., Cancela, L. M., Leite, R. B., Saavedra, C., Ciofi, C., Chelazzi, G., Patarnello, T., and Bortoluzzi, S. (2011). Transcriptome sequencing and microarray development for the Manila clam, *Ruditapes philippinarum* : genomic tools for environmental monitoring. *BMC Genomics*, 12(1) :234.

Milan, M., Pauletto, M., Patarnello, T., Bargelloni, L., Marin, M. G., and Matozzo, V. (2013). Gene transcription and biomarker responses in the clam *Ruditapes philippinarum* after exposure to ibuprofen. *Aquatic Toxicology*, 126(Complete) :17–29.

Mileikovsky, S. (1971). Types of larval development in marine bottom invertebrates, their distribution and ecological significance : a re-evaluation. *Marine Biology*, 10(3) :193–213.

Milke, L. M., Bricelj, V. M., and Parrish, C. C. (2008). Biochemical characterization and nutritional value of three *Pavlova* spp. in unialgal and mixed diets with *Chaetoceros muelleri* for postlarval sea scallops, *Placopecten magellanicus*. *Aquaculture*, 276(1) :130–142.

Mitta, G., Hubert, F., Dyrynda, E. A., Boudry, P., and Roch, P. (2000). Mytilin B and MGD2, two antimicrobial peptides of marine mussels : gene structure and expression analysis. *Developmental & Comparative Immunology*, 24(4) :381–393.

Miyamoto, H., Endo, H., Hashimoto, N., limura, K., Isowa, Y., Kinoshita, S., Kotaki, T., Masaoka, T., Miki, T., Nakayama, S., et al. (2013). The diversity of shell matrix proteins : Genome-wide investigation of the pearl oyster, *Pinctada fucata*. *Zoological science*, 30(10) :801–816.

Mladineo, I., Petrić, M., Hrabar, J., Bočina, I., and Peharda, M. (2012). Reaction of the mussel *Mytilus galloprovincialis* (Bivalvia) to *Eugymnanthea inquilina* (Cnidaria) and *Urastoma cyprinae* (Turbellaria) concurrent infestation. *Journal of Invertebrate Pathology*, 110(1) :118–125.

Mohamed, B., Hajer, A., Susanna, S., Caterina, O., Flavio, M., Hamadi, B., and Aldo, V. (2014). Transcriptomic responses to heat stress and nickel in the mussel *Mytilus galloprovincialis*. *Aquatic Toxicology*, 148(Complete) :104–112.

Mohammed, A. (2013). Why are early life stages of aquatic organisms more sensitive to toxicants than adults ? *New Insights into Toxicity and Drug Testing, InTech, Rijeka, Croatia*.

Montagnani, C., Avarre, J.-C., De Lorgeril, J., Quiquand, M., Boulo, V., and Escoubas, J.-M. (2007). First evidence of the activation of *Cg-timp*, an immune response component of Pacific oysters, through a damage-associated molecular pattern pathway. *Developmental & Comparative Immunology*, 31(1) :1–11.

Montagnani, C., Kappler, C., Reichhart, J., and Escoubas, J. (2004). *Cg*-Rel, the first Rel/NF-κb homolog characterized in a mollusk, the Pacific oyster *Crassostrea gigas*. *FEBS letters*, 561(1) :75–82.

Montagnani, C., Labreuche, Y., and Escoubas, J. (2008). Cg-IκB, a new member of the IκB protein family characterized in the Pacific oyster *Crassostrea gigas*. *Developmental & Comparative Immunology*, 32(3) :182–190.

Moreira, R., Balseiro, P., Planas, J. V., Fuste, B., Beltran, S., Novoa, B., and Figueras, A. (2012a). Transcriptomics of *in vitro* immune-stimulated hemocytes from the manila clam *Ruditapes philippinarum* using high-throughput sequencing. *PLoS ONE*, 7(4) :e35009.

Moreira, R., Balseiro, P., Romero, A., Dios, S., Posada, D., Novoa, B., and Figueras, A. (2012b). Gene expression analysis of clams *Ruditapes philippinarum* and *Ruditapes decussatus* following bacterial infection yields molecular insights into pathogen resistance and immunity. *Developmental & Comparative Immunology*, 36(1) :140–149.

Morse, A. N. and Morse, D. E. (1984). Recruitment and metamorphosis of *Haliotis* larvae induced by molecules uniquely available at the surfaces of crustose red algae. *Journal of Experimental Marine Biology and Ecology*, 75(3) :191–215.

Morse, D. E. (1990). Recent progress in larval settlement and metamorphosis : closing the gaps between molecular biology and ecology. *Bulletin of Marine Science*, 46(2) :465–483.

Mosca, F., Narcisi, V., Cargini, D., Calzetta, A., and Tiscar, P. G. (2011). Age related properties of the Adriatic clam *Chamelea gallina* (L. 1758) hemocytes. *Fish & Shellfish Immunology*, 31(6) :1106–1112.

Mousseau, T. A., Roff, D. A., et al. (1987). Natural selection and the heritability of fitness components. *Heredity*, 59(Pt 2) :181–197.

Murphy, K., Mian, S., et al. (1999). Modelling gene expression data using dynamic Bayesian networks. Technical report, Technical report, Computer Science Division, University of California, Berkeley, CA.

Mydlarz, L. D., Jones, L. E., and Harvell, C. D. (2006). Innate immunity, environmental drivers, and disease ecology of marine and freshwater invertebrates. *Annual Review of Ecology, Evolution, and Systematics*, 37(Complete) :251–288.

Navarro, J., Leiva, G., Martinez, G., and Aguilera, C. (2000). Interactive effects of diet and temperature on the scope for growth of the scallop *Argopecten purpuratus* during reproductive conditioning. *Journal of Experimental Marine Biology and Ecology*, 247(1) :67–83.

Nevejan, N., Saez, I., Gajardo, G., and Sorgeloos, P. (2003). Energy vs. essential fatty acids : what do scallop larvae (*Argopecten purpuratus*) need most ? *Comparative Biochemistry and Physiology Part B : Biochemistry and Molecular Biology*, 134(4) :599–613.

Newkirk, G. F. (1980). Review of the genetics and the potential for selective breeding of commercially important bivalves. *Aquaculture*, 19(3) :209–228.

Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., and Weissman, J. S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 441(7095) :840–846.

Newman, W. A. and Ross, A. (1977). A living tesseropora (Cirripedia : Balanomorpha) from Bermuda and the Azores : First records from the Atlantic since the Oligocene. San Diego Society of Natural History.

Ni, D., Song, L., Wu, L., Chang, Y., Yu, Y., Qiu, L., and Wang, L. (2007). Molecular cloning and mRNA expression of peptidoglycan recognition protein (PGRP) gene in bay scallop (*Argopecten irradians*, Lamarck 1819). *Developmental & Comparative Immunology*, 31(6) :548–558.

Nice, H., Thorndyke, M., Morritt, D., Steele, S., and Crane, M. (2000). Development of *Crassostrea gigas* larvae is affected by 4-nonylphenol. *Marine Pollution Bulletin*, 40(6) :491–496.

Nice, H. E., Morritt, D., Crane, M., and Thorndyke, M. (2003). Long-term and transgenerational effects of nonylphenol exposure at a key stage in the development of *Crassostrea gigas*. possible endocrine disruption ? *Marine Ecology Progress Series*, 256(Complete) :293–300.

Nielsen, C. (1998). Origin and evolution of animal life cycles. *Biological reviews*, 73(2) :125–155.

Nikinmaa, M. and Rytkönen, K. T. (2012). From genomes to functions in aquatic biology. *Marine Genomics*, 5(Complete) :1–6.

Nitsch, D., Tranchevent, L.-C., Thienpont, B., Thorrez, L., Van Esch, H., Devriendt, K., and Moreau, Y. (2009). Network analysis of differential expression for the identification of disease-causing genes. *PloS one*, 4(5) :e5526.

Niu, D., Jin, K., Wang, L., Sun, F., and Li, J. (2013). Identification of cathepsin B in the razor clam *Sinonovacula constricta* and its role in innate immune responses. *Developmental & Comparative Immunology*, 41(2) :94–99.

Nonaka, M. and Yoshizaki, F. (2004). Primitive complement system of invertebrates. *Immunological reviews*, 198(1) :203–215.

Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466(7310) :1057–1062.

Nützel, A. (2014). Larval ecology and morphology in fossil gastropods. *Palaeontology*.

Odintsova, N. A., Dyachuk, V. A., and Nezlin, L. P. (2010). Muscle and neuronal differentiation in primary cell culture of larval *Mytilus trossulus* (Mollusca : Bivalvia). *Cell and Tissue Research*,

339(3) :625–637.

Offord, V., Coffey, T., and Werling, D. (2010). Lrrfinder : a web application for the identification of leucine-rich repeats and an integrative toll-like receptor database. *Developmental & Comparative Immunology*, 34(10) :1035–1041.

Olave, I. A., Reck-Peterson, S. L., and Crabtree, G. R. (2002). Nuclear actin and actin-related proteins in chromatin remodeling. *Annual Review of Biochemistry*, 71(1) :755–781.

Olofsson, H., Ripa, J., and Jonzen, N. (2009). Bet-hedging as an evolutionary game : the trade-off between egg size and number. *Proceedings of the Royal Society B : Biological Sciences*, 276(1669) :2963–2969.

Ottaviani, E., Malagoli, D., and Franchini, A. (2004). Invertebrate humoral factors : cytokines as mediators of cell survival. In *Invertebrate Cytokines and the Phylogeny of Immunity*, pages 1–25. Springer.

Ottaviani, E., Sassi, D., and Kletsas, D. (1997). PDGF- and TGF-beta-induced changes in cell shape of invertebrate immunocytes : effect of calcium entry blockers. *European Journal of Cell Biology*, 74(4) :336–341.

Pairett, A. N. and Serb, J. M. (2013). De novo assembly and characterization of two transcriptomes reveal multiple light-mediated functions in the scallop eye (bivalvia : *Pectinidae*). *PloS one*, 8(7) :e69852.

Pales Espinosa, E., Perrigault, M., and Allam, B. (2010). Identification and molecular characterization of a mucosal lectin (MeML) from the blue mussel *Mytilus edulis* and its potential role in particle capture. *Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology*, 156(4) :495–501.

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution*, 7(4) :114–118.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, 25(1) :547–572.

Pante, E., Rohfritsch, A., Becquet, V., Belkhir, K., Bierne, N., and Garcia, P. (2012). SNP detection from de novo transcriptome sequencing in the bivalve *Macoma balthica* : Marker development for evolutionary studies. *PloS one*, 7(12) :e52302.

Parsons, G., Dadswell, M., and Roff, J. (1993). Influence of biofilm on settlement of sea scallop, *Placopecten magellanicus* (Gmelin, 1791), in Passamaquoddy Bay, New Brunswick, Canada. *Journal of Shellfish Research*, 12(2) :279–283.

Paulet, Y., Lucas, A., and Gerard, A. (1988). Reproduction and larval development in two *Pecten maximus*(L.) populations from brittany. *Journal of Experimental Marine Biology and Ecology*, 119(2) :145–156.

Pavlidis, P. (2003). Using ANOVA for gene selection from microarray studies of the nervous system. *Methods*, 31(4) :282–289.

PAWLIK, J. (1992). Chemical ecology of the settlement of benthic marine invertebrates. *Oceanography and Marine Biology*, 30(Complete) :273–335.

Pearse, J. (1969). Slow developing demersal embryos and larvae of the Antarctic sea star *Odontaster validus*. *Marine Biology*, 3(2) :110–116.

Pechenik, J. (1990a). The influence of food concentration and temperature on growth and morphological differentiation of blue mussel *Mytilus edulis* l. larvae. *Journal of Experimental Marine Biology and Ecology*, 136(1) :47–64.

Pechenik, J. A. (1980). Growth and energy balance during the larval lives of three prosobranch gastropods. *Journal of Experimental Marine Biology and Ecology*, 44(1) :1–28.

Pechenik, J. A. (1990b). Delayed metamorphosis by larvae of benthic marine invertebrates : does it occur ? Is there a price to pay ? *Ophelia*, 32(1-2) :63–94.

Pechenik, J. A. (1999). On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. *Marine Ecology Progress Series*, 177(1) :269–297.

Pechenik, J. A. (2006). Larval experience and latent effects – metamorphosis is not a new beginning. *Integrative and Comparative Biology*, 46(3) :323–333.

Pechenik, J. A., Hammer, K., and Weise, C. (1996). The effect of starvation on acquisition of competence and post-metamorphic performance in the marine prosobranch gastropod *Crepidula fornicata* (L.). *Journal of Experimental Marine Biology and Ecology*, 199(1) :137–152.

Penniman, J. R., Doll, M. K., and Pires, A. (2013). Neural correlates of settlement in veliger larvae of the gastropod, *Crepidula fornicata*. *Invertebrate Biology*, 132(1) :14–26.

Perkins, J. R., Antunes-Martins, A., Calvo, M., Grist, J., Rust, W., Schmid, R., Hildebrandt, T., Kohl, M., Orengo, C., McMahon, S. B., et al. (2014). A comparison of RNA-seq and exon arrays for whole genome transcription profiling of the L5 spinal nerve transection model of neuropathic pain in the rat. *Molecular Pain*, 10(1) :7–43.

Pernet, F., Bricelj, V., and Parrish, C. (2005). Effect of varying dietary levels of omega6 polyunsaturated fatty acids during the early ontogeny of the sea scallop,. *Journal of Experimental Marine Biology and Ecology*, 327(2) :115–133.

Pernet, F., Bricelj, V. M., and Cartier, S. (2006a). Lipid class dynamics during larval ontogeny of sea scallops, *Placopecten magellanicus*, in relation to metamorphic success and response to antibiotics. *Journal of Experimental Marine Biology and Ecology*, 329(2) :265–280.

Pernet, F., Pelletier, C. J., and Milley, J. (2006b). Comparison of three solid-phase extraction methods for fatty acid analysis of lipid fractions in tissues of marine bivalves. *Journal of Chromatography A*, 1137(2) :127–137.

Pernet, F. and Tremblay, R. (2004). Effect of varying levels of dietary essential fatty acid during early ontogeny of the sea scallop *Placopecten magellanicus*. *Journal of Experimental Marine Biology and Ecology*, 310(1) :73–86.

Pernet, F., Tremblay, R., Langdon, C., and Bourget, E. (2004). Effect of additions of dietary triacylglycerol microspheres on growth, survival, and settlement of mussel (*Mytilus* sp.) larvae. *Marine Biology*, 144(4) :693–703.

Perrat, E., Couzinet-Mossion, A., Fossi Tankoua, O., Amiard-Triquet, C., and Wielgosz-Collin, G. (2013). Variation of content of lipid classes, sterols and fatty acids in gonads and digestive glands of *Scrobicularia plana* in relation to environment pollution levels. *Ecotoxicology and Environmental Safety*, 90(Complete) :112–120.

Perrigault, M., Tanguy, A., and Allam, B. (2009). Identification and expression of differentially expressed genes in the hard clam, *Mercenaria mercenaria*, in response to quahog parasite unknown (QPX). *BMC Genomics*, 10(1) :377–394.

Pespeni, M. H., Sanford, E., Gaylord, B., Hill, T. M., Hosfelt, J. D., Jaris, H. K., LaVigne, M., Lenz, E. A., Russell, A. D., Young, M. K., et al. (2013). Evolutionary change during experimental ocean acidification. *Proceedings of the National Academy of Sciences*, 110(17) :6937–6942.

Petes, L. E., Menge, B. A., and Harris, A. L. (2008). Intertidal mussels exhibit energetic trade-offs between reproduction and stress resistance. *Ecological Monographs*, 78(3) :387–402.

Pettersen, A. K., Turchini, G. M., Jahangard, S., Ingram, B. A., and Sherman, C. D. (2010). Effects of different dietary microalgae on survival, growth, settlement and fatty acid composition of blue mussel (*Mytilus galloprovincialis*) larvae. *Aquaculture*, 309(1) :115–124.

Pevzner, P. A. and Lipshutz, R. J. (1994). Towards DNA sequencing chips. In *Mathematical Foundations of Computer Science 1994*, pages 143–158. Springer.

Pfeifer, S., Schiedek, D., and Dippner, J. W. (2005). Effect of temperature and salinity on acetylcholinesterase activity, a common pollution biomarker, in *Mytilus* sp. from the south-western baltic sea. *Journal of Experimental Marine Biology and Ecology*, 320(1) :93–103.

Philipp, E. E., Kraemer, L., Melzner, F., Poustka, A. J., Thieme, S., Findeisen, U., Schreiber, S., and Rosenstiel, P. (2012). Massively parallel RNA sequencing identifies a complex immune gene repertoire in the lophotrochozoan *Mytilus edulis*. *PLoS One*, 7(3) :e33091.

Phillips, N. E. (2002). Effects of nutrition-mediated larval condition on juvenile performance in a marine mussel. *Ecology*, 83(9) :2562–2574.

Phillis, J. W., Horrocks, L. A., and Farooqui, A. A. (2006). Cyclooxygenases, lipoxygenases, and epoxygenases in CNS : Their role and involvement in neurological disorders. *Brain Research Reviews*, 52(2) :201–243.

Phng, L.-K., Potente, M., Leslie, J. D., Babbage, J., Nyqvist, D., Lobov, I., Ondr, J. K., Rao, S., Lang, R. A., Thurston, G., et al. (2009). Nrarp coordinates endothelial notch and wnt signaling to control vessel density in angiogenesis. *Developmental Cell*, 16(1) :70–82.

Plot, C. (2000). The sequence manipulation suite : Javascript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28(6).

Pompeia, C., Lopes, L., Miyasaka, C., Procopio, J., Sannomiya, P., and Curi, R. (2000). Effect of fatty acids on leukocyte function. *Brazilian Journal of Medical and Biological Research*, 33(11) :1255–1268.

Powell, E. N., Morson, J., and Klinck, J. M. (2011). Application of a gene-based population dynamics model to the optimal egg size problem : Why do bivalve planktotrophic eggs vary in size ? *Journal of Shellfish Research*, 30(2) :403–423.

Prado, S., Romalde, J. L., and Barja, J. L. (2010). Review of probiotics for use in bivalve hatcheries. *Veterinary microbiology*, 145(3) :187–197.

Prado-Alvarez, M., Rotllant, J., Gestal, C., Novoa, B., and Figueras, A. (2009). Characterization of a C3 and a factor B-like in the carpet-shell clam, *Ruditapes decussatus*. *Fish & Shellfish Immunology*, 26(2) :305–315.

Qian, P.-Y., Lau, S., Dahms, H.-U., Dobretsov, S., and Harder, T. (2007). Marine biofilms as mediators of colonization by marine macroorganisms : implications for antifouling and aquaculture. *Marine Biotechnology*, 9(4) :399–410.

Qiu, J.-W., Tremblay, R., and Bourget, E. (2002). Ontogenetic changes in hyposaline tolerance in the mussels *Mytilus edulis* and *M. trossulus* : implications for distribution. *Marine Ecology Progress Series*, 228(Complete) :143–152.

Qiu, L., Song, L., Xu, W., Ni, D., and Yu, Y. (2007). Molecular cloning and expression of a toll receptor gene homologue from Zhikong scallop, *Chlamys farreri*. *Fish & Shellfish Immunology*, 22(5) :451–466.

Qiu, L., Song, L., Yu, Y., Zhao, J., Wang, L., and Zhang, Q. (2009). Identification and expression of TRAF6 (TNF receptor-associated factor 6) gene in Zhikong scallop *Chlamys farreri*. *Fish & Shellfish Immunology*, 26(3) :359–367.

R Core Team (2013). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raff, R. A. (1987). Constraint, flexibility, and phylogenetic history in the evolution of direct development in sea urchins. *Developmental biology*, 119(1) :6–19.

Rahman, M. A., Yusoff, F. M., Arshad, A., and Uehara, T. (2014). Effects of delayed metamorphosis on larval survival, metamorphosis, and juvenile performance of four closely related species of tropical sea urchins (genus echinometra). *The Scientific World Journal*, 2014(Complete) :918028.

Ramirez Llodra, E. (2002). Fecundity and life-history strategies in marine invertebrates. *Advances in Marine Biology*, 43 :87–170.

Ramos MartiNez, J., González-Riopedre, M., and Barcia, R. (2012). Role of protein kinases c (pkc) in the relationship between the neuroendocrine and immune systems in marine mussels : The model of *Mytilus galloprovincialis* Lamark (1819). *Italian Journal of Zoology*, 79(2) :162–168.

Rassart, E., Bedirian, A., Do Carmo, S., Guinard, O., Sirois, J., Terrisse, L., and Milne, R. (2000). Apolipoprotein d. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, 1482(1) :185–198.

Rau, A., Jaffrézic, F., Foulley, J.-L., and Doerge, R. W. (2010). An empirical Bayesian method for estimating biological networks from temporal microarray data. *Statistical Applications in Genetics and Molecular Biology*, 9(1) :Article–9.

Rawson, P. D., Joyner, K. L., Meetze, K., and Hilbish, T. J. (1996). Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. *Heredity*, 77(6) :599–607.

Rayssac, N., Pernet, F., Lacasse, O., and Tremblay, R. (2010). Temperature effect on survival, growth, and triacylglycerol content during the early ontogeny of *Mytilus edulis* and *M. trossulus*. *Marine Ecology Progress Series*, 417 :183–191.

Reed, D. H. and Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation biology*, 17(1) :230–237.

Ren, Q., Zhong, X., Yin, S.-W., Hao, F.-Y., Hui, K.-M., Zhang, Z., Zhang, C.-Y., Yu, X.-Q., and Wang, W. (2013a). The first Toll receptor from the triangle-shell pearl mussel *Hyriopsis cumingii*. *Fish & Shellfish Immunology*, 34(5) :1287–1293.

Ren, W., Chen, H., Renault, T., Cai, Y., Bai, C., Wang, C., and Huang, J. (2013b). Complete genome sequence of acute viral necrosis virus associated with massive mortality outbreaks in the c4hinese scallop, *Chlamys farreri*. *Virology journal*, 10(1) :110–117.

Renault, T., Faury, N., Barbosa-Solomieu, V., and Moreau, K. (2011). Suppression substractive hybridisation (SSH) and real time PCR reveal differential gene expression in the Pacific cupped oyster, *Crassostrea gigas*, challenged with *Ostreid* herpesvirus 1. *Developmental & Comparative Immunology*, 35(7) :725–735.

Revathy, K. S., Premachandra, H., Umasuthan, N., Whang, I., and Lee, J. (2013). Structural characterization and immunomodulation of a vertebrate beta-thymosin homologue Tβ12 in disk abalone, *Haliotis discus discus*. *Gene*, 527(1) :376–383.

Rice, W. R. and Hostert, E. E. (1993). Laboratory experiments on speciation : what have we learned in 40 years. *Evolution*, 47(6) :1637–1653.

Rico-Villa, B., Bernard, I., Robert, R., and Pouvreau, S. (2010). A dynamic energy budget (deb) growth model for Pacific oyster larvae, *Crassostrea gigas*. *Aquaculture*, 305(1) :84–94.

Ringø, E., Olsen, R., Gifstad, T., Dalmo, R., Amlund, H., Hemre, G.-I., and Bakke, A. (2010). Prebiotics in aquaculture : a review. *Aquaculture Nutrition*, 16(2) :117–136.

Ritchie, M., Diyagama, D., Neilson, J., van Laar, R., Dobrovic, A., Holloway, A., and Smyth, G. (2006a). Empirical array quality weights in the analysis of microarray data. *BMC bioinformatics*, 7(1) :261.

Ritchie, M. E., Diyagama, D., Neilson, J., van Laar, R., Dobrovic, A., Holloway, A., and Smyth, G. K. (2006b). Empirical array quality weights in the analysis of microarray data. *BMC Bioinformatics*, 7(1) :261–278.

Roberts, R. D. and Lapworth, C. (2001). Effect of delayed metamorphosis on larval competence, and post-larval survival and growth, in the abalone *Haliotis iris* gmelin. *Journal of Experimental Marine Biology and Ecology*, 258(1) :1–13.

Roch, P. (1999). Defense mechanisms and disease prevention in farmed marine invertebrates. *Aquaculture*, 172(1) :125–145.

Roff, D. A. (1992). *Evolution of life histories : theory and analysis*. Springer.

Romero, A., Dios, S., Poisa-Beiro, L., Costa, M. M., Posada, D., Figueras, A., and Novoa, B. (2011). Individual sequence variability and functional activities of fibrinogen-related proteins (FREPs) in the mediterranean mussel (*Mytilus galloprovincialis*) suggest ancient and complex immune recognition models in invertebrates. *Developmental & Comparative Immunology*, 35(3) :334–344.

Romero, A., Novoa, B., and Figueras, A. (2012). Genomics, immune studies and diseases in bivalve aquaculture. *Idaho State Journal*, 9(1) :110–121.

Rumrill, S. S. (1990). Natural mortality of marine invertebrate larvae. *Ophelia*, 32(1-2) :163–198.

Rusten, T. E., Lindmo, K., Juhász, G., Sass, M., Seglen, P. O., Brech, A., and Stenmark, H. (2004). Programmed autophagy in the *Drosophila* fat body is induced by ecdysone through regulation of the pi3k pathway. *Developmental cell*, 7(2) :179–192.

Salta, M., Wharton, J. A., Blache, Y., Stokes, K. R., and Briand, J.-F. (2013). Marine biofilms on artificial surfaces : structure and dynamics. *Environmental Microbiology*, 15(11) :2879–2893.

Samain, J.-F. (2011). Review and perspectives of physiological mechanisms underlying genetically-based resistance of the Pacific oyster *Crassostrea gigas* to summer mortality. *Aquatic Living Resources*, 24(03) :227–236.

Sampath, H. and Ntambi, J. M. (2005). Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annual Review of Nutrition*, 25(Complete) :317–340.

Sánchez-Lazo, C. and Martínez-Pita, I. (2012). Biochemical and energy dynamics during larval development of the mussel *Mytilus galloprovincialis* (lamarck, 1819). *Aquaculture*, 358(Complete) :71–78.

Sánchez-Lazo, C. and Martínez-Pita, I. (2014). Effects of different mono, bi and trispecific microalgal diets on survival, growth, development, settlement and fatty acid composition of mussel *Mytilus galloprovincialis* (Lamarck, 1819) larvae. *Aquaculture*, 426 :138–147.

Sanford, E. and Kelly, M. W. (2011). Local adaptation in marine invertebrates. *Annual Review of Marine Science*, 3(Complete) :509–535.

Sargent, J., Bell, J., Bell, M., Henderson, R., and Tocher, D. (1993). The metabolism of phospholipids and polyunsaturated fatty acids in fish. *Coastal and Estuarine Studies*, 43(Complete) :103–124.

Sargent, J., McEvoy, L., and Bell, J. (1997). Requirements, presentation and sources of polyunsaturated fatty acids in marine fish larval feeds. *Aquaculture*, 155(1) :117–127.

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K. D., Imai, T., and Ueda, H. R. (2013). Quartz-Seq : a highly reproducible and sensitive single-cell RNA-seq reveals non-genetic gene expression heterogeneity. *Genome Biology*, 14(4) :R31–48.

Scheltema, R. S. (1971). Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. *The Biological Bulletin*, 140(2) :284–322.

Schlosser, G. (2002). Modularity and the units of evolution. *Theory in Biosciences*, 121(1) :1–80.

Schmidt, A., Wehrmann, A., and Dittmann, S. (2008). Population dynamics of the invasive Pacific oyster *Crassostrea gigas* during the early stages of an outbreak in the Wadden sea (Germany). *Helgoland Marine Research*, 62(4) :367–376.

Schmitt, P., Lorgeril, J. d., Gueguen, Y., Destoumieux-Garzón, D., and Bachère, E. (2012). Expression, tissue localization and synergy of antimicrobial peptides and proteins in the immune response of the oyster *Crassostrea gigas*. *Developmental & Comparative Immunology*, 37(3-4) :363–370.

Schneider, D. W., Stoeckel, J. A., Rehmann, C. R., Douglas Blodgett, K., Sparks, R. E., and Padilla, D. K. (2003). A developmental bottleneck in dispersing larvae : implications for spatial population dynamics. *Ecology Letters*, 6(4) :352–360.

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases : robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, 28(8) :1086–1092.

Schwämmle, V. and Jensen, O. N. (2010). A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics*, 26(22) :2841–2848.

Seguineau, C., Racotta, I. S., Palacios, E., Delaporte, M., Moal, J., and Soudant, P. (2011). The influence of dietary supplementation of arachidonic acid on prostaglandin production and oxidative stress in the Pacific oyster *Crassostrea gigas*. *Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology*, 160(1) :87–93.

Shanks, A. L. and Brink, L. (2005). Upwelling, downwelling, and cross-shelf transport of bivalve larvae : test of a hypothesis. *Marine Ecology. Progress Series*, 302(Complete) :1–12.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape : a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11) :2498–2504.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, 14(9) :618–630.

Shaposhnikov, A., Komarkov, I., Lebedeva, L., and Shidlovskii, Y. V. (2013). Molecular components of JAK STAT signaling pathway and its interaction with transcription machinery. *Molecular Biology*, 47(3) :343–351.

Sharon, N. and Lis, H. (2004). History of lectins : from hemagglutinins to biological recognition molecules. *Glycobiology*, 14(11) :53R–62R.

Shi, Y., Yu, C., Gu, Z., Zhan, X., Wang, Y., and Wang, A. (2013). Characterization of the pearl oyster (*Pinctada martensii*) mantle transcriptome unravels biomineralization genes. *Marine Biotechnology*, 15(2) :175–187.

Shokralla, S., Spall, J. L., Gibson, J. F., and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21(8) :1794–1805.

Singh, R. and Hartl, M. G. (2012). Fluctuating estuarine conditions are not confounding factors for the comet assay assessment of DNA damage in the mussel *Mytilus edulis*. *Ecotoxicology*, 21(7) :1998–2003.

Smith, W. L. and Murphy, R. C. (2002). The eicosanoids : cyclooxygenase, lipoxygenase, and epoxygenase pathways. *New Comprehensive Biochemistry*, 36(Complete) :341–371.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1) :Article3.

Smyth, G. K., Michaud, J., and Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, 21(9) :2067–2075.

Smyth, G. K. and Speed, T. (2003). Normalization of cDNA microarray data. *Methods*, 31(4) :265–273.

Sokolova, I. M. (2013). Energy-limited tolerance to stress as a conceptual framework to integrate the effects of multiple stressors. *Integrative and Comparative Biology*, 53(4) :597–608.

Soneson, C. and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, 14(1) :91–109.

Song, L., Wang, L., Qiu, L., and Zhang, H. (2010). Bivalve immunity. In *Invertebrate Immunity*, pages 44–65. Springer.

Song, L., Xu, W., Li, C., Li, H., Wu, L., Xiang, J., and Guo, X. (2006). Development of expressed sequence tags from the bay scallop, *Argopecten irradians irradians*. *Marine Biotechnology*, 8(2) :161–169.

Soudant, P., Marty, Y., Moal, J., Masski, H., and Samain, J. F. (1998). Fatty acid composition of polar lipid classes during larval development of scallop *Pecten maximus* (L.). *Comparative Biochemistry and Physiology-Part A : Molecular & Integrative Physiology*, 121(3) :279–288.

St-Onge, P., Sévigny, J.-M., Strasser, C., and Tremblay, R. (2013). Strong population differentiation of softshell clams (*Mya arenaria*) sampled across seven biogeographic marine ecoregions : possible selection and isolation by distance. *Marine Biology*, 160(5) :1065–1081.

Stanley, D. (2006). Prostaglandins and other eicosanoids in insects : biological significance. *Annual Review of Entomology*, 51(1) :25–44.

Stanley, D., Miller, J., and Tunaz, H. (2009). Eicosanoid actions in insect immunity. *Journal of Innate Immunity*, 1(4) :282–290.

Stanley, D. W. and Howard, R. W. (1998). The biology of prostaglandins and related eicosanoids in invertebrates : cellular, organismal and ecological actions. *American Zoologist*, 38(2) :369–381.

Stewart-Oaten, A., Murdoch, W. W., and Parker, K. R. (1986). Environmental impact assessment :" pseudoreplication" in time ? *Ecology*, 67(4) :929–940.

Stone, E. A. and Ayroles, J. F. (2009). Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS genetics*, 5(5) :e1000479.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 64(3) :479–498.

Storz, J. F. and Wheat, C. W. (2010). Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, 64(9) :2489–2509.

Strathmann, R. (1987). Echinoderm larval ecology viewed from the egg. *Echinoderm Studies 2 (1987)*, 2 :55.

Strathmann, R. R. (1978). The evolution and loss of feeding larval stages of marine invertebrates. *Evolution*, 32(4) :894–906.

Strathmann, R. R. (1985). Feeding and nonfeeding larval development and life-history evolution in marine invertebrates. *Annual Review of Ecology and Systematics*, 16(1) :339–361.

Strathmann, R. R. (1993). Hypotheses on the origins of marine larvae. *Annual Review of Ecology and Systematics*, 24(1) :89–117.

Strathmann, R. R., Strathmann, M. F., Ruiz-Jones, G., and Hadfield, M. G. (2010). Effect of plasticity in hatching on duration as a precompetent swimming larva in the nudibranch *Phestilla sibogae*. *Invertebrate Biology*, 129(4) :309–318.

Su, J., Ni, D., Song, L., Zhao, J., and Qiu, L. (2007). Molecular cloning and characterization of a short type peptidoglycan recognition protein (CfPGRP-S1) cDNA from Zhikong scallop *Chlamys farreri*. *Fish & Shellfish Immunology*, 23(3) :646–656.

Sussarellu, R., Fabioux, C., Le Moullac, G., Fleury, E., and Moraga, D. (2010). Transcriptomic response of the Pacific oyster *Crassostrea gigas* to hypoxia. *Marine genomics*, 3(3) :133–143.

Tabachnick, B. G., Fidell, L. S., and Osterlind, S. J. (2001). Using multivariate statistics.

Takeda, K. and Akira, S. (2004). Microbial recognition by Toll-like receptors. *Journal of Dermatological Science*, 34(2) :73–82.

Takeuchi, T., Kawashima, T., Koyanagi, R., Gyoja, F., Tanaka, M., Ikuta, T., Shoguchi, E., Fujiwara, M., Shinzato, C., Hisata, K., et al. (2012). Draft genome of the pearl oyster *Pinctada fucata* : a platform for understanding bivalve biology. *DNA research*, 19(2) :117–130.

Talmage, S. C. and Gobler, C. J. (2011). Effects of elevated temperature and carbon dioxide on the growth and survival of larvae and juveniles of three species of northwest Atlantic bivalves. *PloS one*, 6(10) :e26941.

Tanguy, A., Bierne, N., Saavedra, C., Pina, B., Bachère, E., Kube, M., Bazin, E., Bonhomme, F., Boudry, P., Boulo, V., et al. (2008). Increasing genomic information in bivalves through new EST collections in four species : development of new genetic markers for environmental studies and genome evolution. *Gene*, 408(1) :27–36.

Tanguy, A., Boutet, I., Laroche, J., and Moraga, D. (2004). Molecular identification and expression study of differentially regulated genes in the Pacific oyster *Crassostrea gigas* in response to pesticide exposure. *FEBS Journal*, 272(2) :390–403.

Tanguy, M., McKenna, P., Gauthier-Clerc, S., Pellerin, J., Danger, J.-M., and Siah, A. (2013a). Functional and molecular responses in *Mytilus edulis* hemocytes exposed to bacteria, *Vibrio splendidus*. *Developmental & Comparative Immunology*, 39(4) :419–429.

Tanguy, M., McKenna, P., Gauthier-Clerc, S., Pellerin, J., Danger, J.-M., and Siah, A. (2013b). Sequence analysis of a normalized cDNA library of *Mytilus edulis* hemocytes exposed to *Vibrio splendidus* LGP32 strain. *Results in Immunology*, 3(1) :40–50.

Terahara, K. and Takahashi, K. G. (2008). Mechanisms and immunological roles of apoptosis in molluscs. *Current Pharmaceutical Design*, 14(2) :131–137.

Thiyagarajan, V. (2010). A review on the role of chemical cues in habitat selection by barnacles : new insights from larval proteomics. *Journal of Experimental Marine Biology and Ecology*, 392(1) :22–36.

Thompson, P., Guo, M., and Harrison, P. (1993). The influence of irradiance on the biochemical composition of three phytoplankton species and their nutritional value for larvae of the Pacific oyster (*Crassostrea gigas*). *Marine Biology*, 117(2) :259–268.

Thompson, P. A., Guo, M.-x., and Harrison, P. J. (1996). Nutritional value of diets that vary in fatty acid composition for larval Pacific oysters (*Crassostrea gigas*). *Aquaculture*, 143(3) :379–391.

Thorson, G. (1950). Reproductive and larval ecology of marine bottom invertebrates. *Biological Reviews*, 25(1) :1–45.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso : a retrospective. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 73(3) :273–282.

Tibshirani, R., Chu, G., Narasimhan, B., and Li, J. (2011). *samr : SAM : Significance Analysis of Microarrays*. R package version 2.0.

Timmins-Schiffman, E. B., Friedman, C. S., Metzger, D. C., White, S. J., and Roberts, S. B. (2013). Genomic resource development for shellfish of conservation concern. *Molecular Ecology Resources*, 13(2) :295–305.

Timmons, L., Court, D. L., and Fire, A. (2001). Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene*, 263(1) :103–112.

Tirapé, A., Bacque, C., Brizard, R., Vandenbulcke, F., and Boulo, V. (2007). Expression of immune-related genes in the oyster *Crassostrea gigas* during ontogenesis. *Developmental & Comparative Immunology*, 31(9) :859–873.

Tocher, D. (2003). Metabolism and functions of lipids and fatty acids in teleost fish. *Reviews in Fisheries Science*, 11(2) :107–184.

Tocher, D. R. (2010). Fatty acid requirements in ontogeny of marine and freshwater fish. *Aquaculture Research*, 41(5) :717–732.

Torres, M., Romero, M., Prado, S., Dubert, J., Tahrioui, A., Otero, A., and Llamas, I. (2013). *N*-acylhomoserine lactone-degrading bacteria isolated from hatchery bivalve larval cultures. *Microbiological Research*, 168(9) :547–554.

Toubiana, M., Gerdol, M., Rosani, U., Pallavicini, A., Venier, P., and Roch, P. (2013). Toll-like receptors and MyD88 adaptors in *Mytilus* : complete cds and gene expression levels. *Developmental & Comparative Immunology*, 40(2) :158–166.

Toupoint, N., Mohit, V., Linossier, I., Bourgougnon, N., Myrand, B., Olivier, F., Lovejoy, C., and Tremblay, R. (2012). Effect of biofilm age on settlement of *Mytilus edulis*. *Biofouling*, 28(9) :985–1001.

Tremblay, R., Cartier, S., Miner, P., Pernet, F., Quéré, C., Moal, J., Muzellec, M.-L., Mazuret, M., and Samain, J. F. (2007a). Effect of *Rhodomonas salina* addition to a standard hatchery diet during the early ontogeny of the scallop *Pecten maimus*. *Aquaculture*, 262(2-4) :410–418.

Tremblay, R., Myrand, B., and Guderley, H. (1998). Temporal variation of lysosomal capacities in relation to susceptibility of mussels, *Mytilus edulis*, to summer mortality. *Marine Biology*, 132(4) :641–649.

Tremblay, R., Olivier, F., Bourget, E., and Rittschof, D. (2007b). Physiological condition of *Balanus amphitrite* cyprid larvae determines habitat selection success. *Marine Ecology Progress Series*, 340(Complete) :1–8.

Truebano, M., Burns, G., Thorne, M. A., Hillyard, G., Peck, L. S., Skibinski, D. O., and Clark, M. S. (2010). Transcriptional response to heat stress in the Antarctic bivalve *Laternula elliptica*. *Journal of Experimental Marine Biology and Ecology*, 391(1) :65–72.

Truman, J. W. and Riddiford, L. M. (1999). The origins of insect metamorphosis. *Nature*, 401(6752) :447–452.

Tsang, H. S., Hui, A. K., and Chow, K. L. (2007). noah-2 : new component in ray morphogenesis of *C. elegans*. In *International Worm Meeting*.

Vallabhapurapu, S. and Karin, M. (2009). Regulation and function of NF-κb transcription factors in the immune system. *Annual Review of Immunology*, 27(Complete) :693–733.

Vance, R. R. (1973). On reproductive strategies in marine benthic invertebrates. *American Naturalist*, pages 339–352.

Varotto, L., Domeneghetti, S., Rosani, U., Manfrin, C., Cajaraville, M. P., Raccanelli, S., Pallavicini, A., and Venier, P. (2013). DNA damage and transcriptional changes in the gills of *Mytilus galloprovincialis* exposed to nanomolar doses of combined metal salts (Cd, Cu, Hg). *PloS one*, 8(1) :e54602.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition.

Venier, P., Varotto, L., Rosani, U., Millino, C., Celegato, B., Bernante, F., Lanfranchi, G., Novoa, B., Roch, P., Figueras, A., et al. (2011). Insights into the innate immunity of the Mediterranean mussel *Mytilus galloprovincialis*. *BMC Genomics*, 12(1) :1–19.

Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K. K., Olsen, J. V., Hyman, A. A., Stunnenberg, H. G., et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, 142(6) :967–980.

Verschuere, L., Rombaut, G., Sorgeloos, P., and Verstraete, W. (2000). Probiotic bacteria as biological control agents in aquaculture. *Microbiology and Molecular Biology Reviews*, 64(4) :655–671.

Videla, J., Chaparro, O., Thompson, R., and Concha, I. (1998). Role of biochemical energy reserves in the metamorphosis and early juvenile development of the oyster *Ostrea chilensis*. *Marine Biology*, 132(4) :635–640.

Villinski, J. T., Villinski, J. C., Byrne, M., and Raff, R. A. (2002). Convergent maternal provisioning and life-history evolution in echinoderms. *Evolution*, 56(9) :1764–1775.

Vogler, A. P. and Desalle, R. (1994). Diagnosing units of conservation management. *Conservation Biology*, 8(2) :354–363.

Von der Meden, C. E., Porri, F., and McQuaid, C. D. (2012). New estimates of early post-settlement mortality for intertidal mussels show no relationship with meso-scale coastline topographic features. *Marine Ecology Progress Series*, 463(Complete) :193–204.

Voronezhskaya, E. E., Nezlin, L. P., Odintsova, N. A., Plummer, J. T., and Croll, R. P. (2008). Neuronal development in larval mussel *Mytilus trossulus* (Mollusca : Bivalvia). *Zoomorphology*, 127(2) :97–110.

Vouras, C. and Dailianis, S. (2012). Evidence for phosphatidylinositol-3-OH-kinase (PI3-kinase) involvement in Cd-mediated oxidative effects on hemocytes of mussels. *Comparative Biochemistry and Physiology Part C : Toxicology & Pharmacology*, 155(4) :587–593.

Waldbusser, G. G., Brunner, E. L., Haley, B. A., Hales, B., Langdon, C. J., and Prahl, F. G. (2013). A developmental and energetic basis linking larval oyster shell formation to acidification sensitivity. *Geophysical Research Letters*, 40(10) :1–6.

Waldock, M. and Holland, D. (1984). Fatty acid metabolism in young oysters, *Crassostrea gigas* : polyunsaturated fatty acids. *Lipids*, 19(5) :332–336.

Wang, J., Zhang, H., Wang, L., Qiu, L., Yue, F., Yang, C., and Song, L. (2013a). Molecular cloning and transcriptional regulation of an allograft inflammatory factor-1 (AIF-1) in Zhikong scallop *Chlamys farreri*. *Gene*, 532(2) :178–184.

Wang, L., Song, L., Zhao, J., Qiu, L., Zhang, H., Xu, W., Li, H., Li, C., Wu, L., and Guo, X. (2009a). Expressed sequence tags from the Zhikong scallop (*Chlamys farreri*) : Discovery and annotation of host-defense genes. *Fish & Shellfish Immunology*, 26(5) :744–750.

Wang, M., Yang, J., Zhou, Z., Qiu, L., Wang, L., Zhang, H., Gao, Y., Wang, X., Zhang, L., Zhao, J., et al. (2011). A primitive Toll-like receptor signaling pathway in mollusk Zhikong scallop *Chlamys farreri*. *Developmental & Comparative Immunology*, 35(4) :511–520.

Wang, S., Hou, R., Bao, Z., Du, H., He, Y., Su, H., Zhang, Y., Fu, X., Jiao, W., Li, Y., et al. (2013b). Transcriptome sequencing of Zhikong scallop (*Chlamys farreri*) and comparative transcriptomic analysis with Yesso scallop (*Patinopecten yessoensis*). *PloS one*, 8(5) :e63927.

Wang, X., Liu, B., Wang, G., Tang, B., and Xiang, J. (2008). Molecular cloning and functional analysis of cathepsin B in nutrient metabolism during larval development in *Meretrix meretrix*. *Aquaculture*, 282(1) :41–46.

Wang, X., Wang, L., Yao, C., Qiu, L., Zhang, H., Zhi, Z., and Song, L. (2012a). Alternation of immune parameters and cellular energy allocation of *Chlamys farreri* under ammonia-N exposure and *Vibrio anguillarum* challenge. *Fish & Shellfish Immunology*, 32(5) :741–749.

Wang, X., Wang, L., Zhang, H., Ji, Q., Song, L., Qiu, L., Zhou, Z., Wang, M., and Wang, L. (2012b). Immune response and energy metabolism of *Chlamys farreri* under *Vibrio anguillarum* challenge and high temperature exposure. *Fish & Shellfish Immunology*, 33(4) :1016–1026.

Wang, Y., Xu, Z., and Guo, X. (2004). Differences in the rDNA-bearing chromosome divide the Asian-Pacific and Atlantic species of *Crassostrea* (Bivalvia, Mollusca). *The Biological Bulletin*, 206(1) :46–54.

Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-Seq : a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1) :57–63.

Watanabe, K., Sezaki, K., Yazawa, K., and Hino, A. (1992). Nutritive fortification of the rotifer *Brachionus plicatilis* with eicosapentaenoic acid-producing bacteria. *Nippon Suisan Gakkaishi*, 58(2) :271–276.

Wei, X., Yang, J., Yang, D., Xu, J., Liu, X., Yang, J., Fang, J., and Qiao, H. (2012). Molecular cloning and mRNA expression of two peptidoglycan recognition protein (PGRP) genes from mollusk *Solen grandis*. *Fish & Shellfish Immunology*, 32(1) :178–185.

Weiner, R. M. and Colwell, R. R. (1982). *Induction of settlement and metamorphosis in Crassostrea virginica by a melanin-synthesizing bacterium.* Sea Grant Program, University of Maryland.

Weiner, R. M., Segall, A. M., and Colwell, R. R. (1985). Characterization of a marine bacterium associated with *Crassostrea virginica* (the Eastern oyster). *Applied and environmental microbiology*, 49(1) :83–90.

Wen, J., Chen, Z., and Cai, X. (2013). A biophysical model for identifying splicing regulatory elements and their interactions. *PloS one*, 8(1) :e54885.

Weng, N. and Wang, W.-X. (2014). Improved tolerance of metals in contaminated oyster larvae. *Aquatic Toxicology*, 146(Complete) :61–69.

Westermann, A. J., Gorski, S. A., and Vogel, J. (2012). Dual *RNA-seq* of pathogen and host. *Nature Reviews Microbiology*, 10(9) :618–630.

Whyte, J. (1992). Compositional changes in the larva to juvenile development of the scallop *Crassadoma gigantea* ray. *Journal of Experimental Marine Biology and Ecology*, 163(1) :13–29.

Whyte, J. N. (1987). Biochemical composition and energy content of six species of phytoplankton used in mariculture of bivalves. *Aquaculture*, 60(3) :231–241.

Widdows, J. (1991). Physiological ecology of mussel larvae. *Aquaculture*, 94(2-3) :147–163.

Wiens, M., Korzhev, M., Krasko, A., Thakur, N. L., Perović-Ottstadt, S., Breter, H. J., Ushijima, H., Diehl-Seifert, B., Müller, I. M., and Müller, W. E. (2005). Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a *MyD88*-dependent signaling pathway induction of a perforin-like molecule. *Journal of Biological Chemistry*, 280(30) :27949–27959.

Williams, E. and Somero, G. (1996). Seasonal-, tidal-cycle-and microhabitat-related variation in membrane order of phospholipid vesicles from gills of the intertidal mussel *Mytilus californianus*. *Journal of Experimental Biology*, 199(7) :1587–1596.

Williams, E. A., Degnan, B. M., Gunter, H., Jackson, D. J., Woodcroft, B. J., and Degnan, S. M. (2009). Widespread transcriptional changes pre-empt the critical pelagic-benthic transition in the vetigastropod *Haliotis asinina*. *Molecular Ecology*, 18(5) :1006–1025.

Williams, E. A. and Degnan, S. M. (2009). Carry-over effect of larval settlement cue on postlarval gene expression in the marine gastropod haliotis asinina. *Molecular Ecology*, 18(21) :4434–4449.

Wilson, J., Chaparro, O., and Thompson, R. (1996). The importance of broodstock nutrition on the viability of larvae and spat in the Chilean oyster *Ostrea chilensis*. *Aquaculture*, 139(1) :63–75.

Wing, S. R. and Leichter, J. J. (2011). Variation in environmental conditions in a subtidal prey refuge : effects of salinity stress, food availability and predation on mussels in a fjord system. *Marine Ecology Progress Series*, 422(Complete) :201–210.

Wong, K. H., Jin, Y., and Moqtaderi, Z. (2013). Multiplex illumina sequencing using DNA barcoding. *Current Protocols in Molecular Biology*, 40(10) :7–11.

Wu, D., Hu, B., Wen, C., Lin, G., Tao, Z., Hu, X., and Xie, Y. (2013). Gene identification and recombinant protein of a lysozyme from freshwater mussel *Cristaria plicata*. *Fish & Shellfish Immunology*, 34 :1033–1041.

Xiao, J., Ford, S. E., Yang, H., Zhang, G., Zhang, F., and Guo, X. (2005). Studies on mass summer mortality of cultured Zhikong scallops (*Chlamys farreri* Jones et Preston) in China. *Aquaculture*, 250(3) :602–615.

Xing, E. P., Jordan, M. I., Karp, R. M., et al. (2001). Feature selection for high-dimensional genomic microarray data. In *ICML*, volume 1, pages 601–608. Citeseer.

Xing, J., Tang, X., Ni, Y., and Zhan, W. (2014). Application of monoclonal antibody against granulocytes of scallop *Chlamys farreri* on granulocytes occurrence at different developmental stages and antigenic cross-reactivity of granulocytes in five other bivalve species. *Fish & Shellfish Immuno-*

logy, 36(1) :315–319.

Xu, T., Yang, S., Xie, J., Ye, S., Luo, M., Zhu, Z., and Wu, X. (2012). HMGB in mollusk *Crassostrea ariakensis* gould : Structure, pro-inflammatory cytokine function characterization and anti-infection role of its antibody. *PloS one*, 7(11) :e50789.

Xu, X., Zhang, Y., Williams, J., Antoniou, E., McCombie, W., Wu, S., Zhu, W., Davidson, N. O., Denoya, P., and Li, E. (2013). Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics*, 14(Suppl 9) :S1.

Xue, W., Li, J.-T., Zhu, Y.-P., Hou, G.-Y., Kong, X.-F., Kuang, Y.-Y., and Sun, X.-W. (2013). L_RNA_scaffolder : scaffolding genomes with transcripts. *BMC Genomics*, 14(1) :1–14.

Yang, B., Qin, J., Shi, B., Han, G., Chen, J., Huang, H., and Ke, C. (2012). Molecular characterization and functional analysis of adrenergic like receptor during larval metamorphosis in *Crassostrea angulata*. *Aquaculture*, 366(Complete) :54–61.

Yang, J., Wang, W., Wei, X., Qiu, L., Wang, L., Zhang, H., and Song, L. (2010). Peptidoglycan recognition protein of *Chlamys farreri* (CfPGRP-S1) mediates immune defenses against bacterial infection. *Developmental & Comparative Immunology*, 34(12) :1300–1307.

Yang, J.-L., Li, S.-H., Li, Y.-F., Liu, Z.-W., Liang, X., Bao, W.-Y., and Li, J.-L. (2013). Effects of neuroactive compounds, ions and organic solvents on larval metamorphosis of the mussel *Mytilus coruscus*. *Aquaculture*, 396(Complete) :106–112.

Yao, C.-L. and Somero, G. N. (2012). The impact of acute temperature stress on hemocytes of invasive and native mussels (*Mytilus galloprovincialis* and *Mytilus californianus*) : DNA damage, membrane integrity, apoptosis and signaling pathways. *The Journal of Experimental Biology*, 215(24) :4267–4277.

Yao, C.-L. and Somero, G. N. (2013). Thermal stress and cellular signaling processes in hemocytes of native (*Mytilus californianus*) and invasive (*M. galloprovincialis*) mussels : Cell cycle regulation and DNA repair. *Comparative Biochemistry and Physiology Part A : Molecular & Integrative Physiology*.

Yaroslavtseva, L. and Sergeeva, E. (2003). Response of the larvae of *Mytilus trossulus* (Bivalvia, Mytilidae) to desalination and temperature rise on the surface of the water column. *Biologiya Morya*, 29(3) :184–188.

Yaroslavtseva, L. M. and Sergeeva, E. P. (2006). Adaptivity of the bivalve *Mytilus trossulus* larvae to short-and long-term changes in water temperature and salinity. *Russian Journal of Marine Biology*, 32(2) :82–87.

Yazawa, K. (1996). Production of eicosapentaenoic acid from marine bacteria. *Lipids*, 31(1) :S297–S300.

Yip, K. Y., Cheng, C., and Gerstein, M. (2013). Machine learning and genome annotation : a match meant to be ? *Genome Biology*, 14(5) :1–10.

Yockteng, R., Almeida, A. M., Yee, S., Andre, T., Hill, C., and Specht, C. D. (2014). A method for extracting high-quality RNA from diverse plants for Next-Generation Sequencing and gene expression analyses. *Applications in Plant Sciences*, 1(12) :1–7.

You, Y., Huan, P., and Liu, B. (2012). RNAi assay in primary cells : a new method for gene function analysis in marine bivalve. *Molecular Biology Reports*, 39(8) :8209–8216.

Young, C. M. (1990). Larval ecology of marine invertebrates : a sesquicentennial history. *Ophelia*, 32(1-2) :1–48.

Yue, F., Shi, X., Zhou, Z., Wang, L., Wang, M., Yang, J., Qiu, L., and Song, L. (2013a). The expression of immune-related genes during the ontogenesis of scallop *Chlamys farreri* and their response to bacterial challenge. *Fish & Shellfish Immunology*, 34(3) :855–864.

Yue, F., Zhou, Z., Wang, L., Ma, Z., Wang, J., Wang, M., Zhang, H., and Song, L. (2013b). Maternal transfer of immunity in scallop *Chlamys farreri* and its trans-generational immune protection to offspring against bacterial challenge. *Developmental & Comparative Immunology*, 41(4) :569–577.

Yue, X., Huan, P., Xiao, G., and Liu, B. (2013c). Expression patterns of an i-type lysozyme in the clam *Meretrix meretrix* along with larval development. *Developmental & Comparative Immunology*, 41(1) :27–32.

Zahurak, M., Parmigiani, G., Yu, W., Scharpf, R. B., Berman, D., Schaeffer, E., Shabbeer, S., and Cope, L. (2007). Pre-processing agilent microarray data. *BMC Bioinformatics*, 8(1) :1–13.

Zapata, M., Tanguy, A., David, E., Moraga, D., and Riquelme, C. (2009). Transcriptomic response of *Argopecten purpuratus* post-larvae to copper exposure under experimental conditions. *Gene*, 442(1-2) :37–46.

Zardus, J. D., Nedved, B. T., Huang, Y., Tran, C., and Hadfield, M. G. (2008). Microbial biofilms facilitate adhesion in biofouling invertebrates. *The Biological Bulletin*, 214(1) :91–98.

Zerbino, D. R. and Birney, E. (2008). Velvet : algorithms for de novo short read assembly using de bruijn graphs. *Genome research*, 18(5) :821–829.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., et al. (2012a). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, 490(7418) :49–54.

Zhang, L., Hou, R., Su, H., Hu, X., Wang, S., and Bao, Z. (2012b). Network analysis of oyster transcriptome revealed a cascade of cellular responses during recovery after heat shock. *PloS one*, 7(4) :e35484.

Zhang, Q., Zmasek, C. M., Cai, X., and Godzik, A. (2011). Tir domain-containing adaptor SARM is a late addition to the ongoing microbe-host dialog. *Developmental & Comparative Immunology*, 35(4) :461–468.

Zhang, R., Zhang, L.-l., Ye, X., Tian, Y.-y., Sun, C.-f., Lu, M.-x., and Bai, J.-j. (2013). Transcriptome profiling and digital gene expression analysis of nile tilapia (*Oreochromis niloticus*) infected by *Streptococcus agalactiae*. *Molecular Biology Reports*, 40(10) :5657–5668.

Zhao, J., Li, C., Chen, A., Li, L., Su, X., and Li, T. (2010). Molecular characterization of a novel big defensin from clam *Venerupis philippinarum*. *PloS one*, 5(10) :e13480.

Zhao, J., Song, L., Li, C., Zou, H., Ni, D., Wang, W., and Xu, W. (2007). Molecular cloning of an invertebrate goose-type lysozyme gene from *Chlamys farreri*, and lytic activity of the recombinant protein. *Molecular Immunology*, 44(6) :1198–1208.

Zhou, Z., Wang, M., Zhao, J., Wang, L., Gao, Y., Zhang, H., Liu, R., and Song, L. (2013). The increased transcriptional response and translocation of a Rel/NF-κB homologue in scallop *Chlamys farreri* during the immune stimulation. *Fish & shellfish immunology*, 34(5) :1209–1215.

INDEX