CHRISTIAN ROBERGE

ÉVOLUTION CONTEMPORAINE ET PLASTICITÉ PHÉNOTYPIQUE CHEZ LE SAUMON ATLANTIQUE SOUS LA LUNETTE DU TRANSCRIPTOME

Thèse présentée
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de doctorat en Biologie
pour l'obtention du grade de Philosophiae Doctor (Ph.D.)

DÉPARTEMENT DE BIOLOGIE
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2009

Résumé court

Les principaux objectifs de ces travaux étaient d'étudier la transcriptomique de l'évolution contemporaine et de la plasticité phénotypique chez le saumon atlantique. L'utilisation de bio-puces a permis d'identifier plusieurs gènes dont le niveau de transcription présente des différences héritables entre saumons d'élevage et sauvages. Ces résultats nous renseignent sur les mécanismes génétiques de l'évolution parallèle et permettent d'évaluer l'ampleur des différences génétiques entre saumons d'élevage et sauvages. En outre, l'estimation du Q_{st} des niveaux de transcription nous a permis d'identifier des gènes pour lesquels le contrôle génétique du niveau de transcription aurait évolué en 6 générations sous l'effet de la sélection chez deux sous-populations de saumon en nature. Nous avons également identifié des gènes vraisemblablement impliqués dans l'immunité ainsi que dans la perte plastique de dominance sociale chez de jeunes saumons mis en présence de truites, contribuant à développer les connaissances sur les liens entre transcription génique et plasticité phénotypique.

Résumé long

L'un des principaux objectifs des travaux présentés dans cette thèse était d'étudier les mécanismes moléculaires de l'évolution contemporaine chez le saumon atlantique (*Salmo salar*) à l'aide de la technique des bio-puces. La création de lignées de saumon d'élevage est un exemple d'évolution contemporaine dirigée par une activité humaine. La comparaison des niveaux de transcription de milliers de gènes entre saumons d'élevage, saumons sauvages et leurs hybrides nous a permis d'en identifier certains dont le niveau de transcription présentait des différences héritables entre ces groupes. Ces résultats nous renseignent sur les mécanismes génétiques de l'évolution parallèle et permettent d'évaluer l'ampleur des différences génétiques accumulées en 25 à 35 ans entre saumons d'élevage et sauvages ainsi que l'importance relative de l'additivité dans le contrôle génétique des niveaux de transcription géniques. Ils appuient l'idée que des mesures limitant les échappées de saumons d'élevage doivent être prises rapidement. Dans d'autres travaux, l'estimation du Q_{st} ainsi que de l'héritabilité du niveau de transcription de milliers de gènes nous a permis d'appliquer une méthode nouvelle, le « balayage transcriptomique », afin d'identifier des gènes pour lesquels le contrôle génétique du niveau de transcription a potentiellement évolué en 6 générations sous l'effet de la sélection naturelle chez deux sous-populations de saumon de la rivière Ste-Marguerite. Un autre objectif de ces travaux était l'étude des mécanismes moléculaires de la plasticité phénotypique. Ainsi, la comparaison des niveaux de transcription géniques chez des juvéniles de saumons sains ou atteints de saprolegniose nous a permis d'identifier plusieurs des acteurs potentiels de la réponse immunitaire à cette infection, incluant notamment plusieurs gènes codant pour des protéines de la réponse immunitaire non-spécifique. Enfin, nous avons identifié, en comparant le niveau de transcription de 16006 gènes dans les cerveaux de saumons juvéniles dominants ou subordonnés en présence ou en l'absence de truite arc-en-ciel, certains acteurs moléculaires potentiellement impliqués dans la perte plastique de dominance sociale chez de jeunes saumons mis en présence de truites, contribuant ainsi à développer les connaissances sur les liens entre la transcription génique et la plasticité comportementale dans le contexte d'interactions compétitives entre espèces invasives et espèces indigènes.

Abstract

One of the main objectives of the work presented here was to study the molecular mechanisms of contemporary evolution in Atlantic salmon (*Salmo salar*) using microarrays. The creation of salmon breeding lines through artificial selection is an example of contemporary evolution driven by human activities. Comparing the transcription levels of thousands of genes between farmed salmon, wild salmon and their hybrids allowed us to identify genes of which the transcription level showed heritable differences between these groups. The results inform us on the genetic mechanisms of parallel evolution and allow us to evaluate the extent of the genetic differences accumulated in only 25 to 35 years of artificial selection between wild and farmed salmon as well as the prevalence of additivity in the genetic control of gene transcription. These results also support the idea that measures to markedly reduce escapes of farmed salmon and their reproduction in the wild are urgently needed. In another study, estimating gene transcription level Q_{st} and heritability for thousands of genes allowed us to apply a new method, the «transcriptome scan», to identify genes for which the genetic control of transcription is likely to have evolved in only 6 generations under the effect of directional selection in two Atlantic salmon sub-populations from rivière Ste-Marguerite. Another goal of this thesis was to study the molecular mechanisms of phenotypic plasticity. Hence, the comparison of transcript levels from saprolegniosis-affected or healthy salmon juveniles allowed us to identify several potential actors of the immune response to this infection, including several genes coding for proteins of the acute-phase response. Finally, we compared the transcription levels of thousands of genes in the brains of socially dominant or subordinate salmon juveniles in absence or presence of rainbow trout (*O. mykiss*), which allowed us to identify several potential molecular actors in the plastic loss of social dominance hierarchies in salmon in presence of trout. This contributes to a better understanding of the relation between gene transcription and behavioural plasticity in the context of competitive interactions between invasive and native species.

Avant-Propos

Remerciements

Je remercie mon directeur Louis Bernatchez pour son enthousiasme scientifique, ses commentaires judicieux et pour avoir gardé confiance en moi dans mes moments de déroute ou d'isolement. Je remercie aussi ma co-directrice Helga Guderley et les autres membres de mon comité d'encadrement, Alan Anderson et Julian Dodson pour les nombreux et intéressants échanges scientifiques que j'ai eus avec eux.

Je remercie également tous mes collègues et amis des laboratoires Bernatchez et Dodson, Sigurd Einum, le personnel du LARSA et de la station du NINA à Ims, ainsi que le FQRNT, le CRSNG, la Bourse Richard Bernard, le RAQ, Québec-Océan, nos partenaires du cGRASP, l'AELIÉS et le CIRSA pour leur support moral, financier ou scientifique.

Enfin, je remercie ma famille et mes proches.

Organisation de la thèse

Le corps de cette thèse (chapitres 2 à 6) est constitué de cinq chapitres en anglais publiés (ou en voie de l'être) sous forme d'articles dans des revues scientifiques. Mon directeur Louis Bernatchez ainsi que ma co-directrice Helga Guderley sont co-auteurs sur chacun de ceux-ci puisqu'ils ont rendu ces travaux possibles en obtenant le financement nécessaire à leur réalisation, mais aussi parce qu'ils y ont contribué par d'innombrables idées, commentaires et suggestions.

Le chapitre 2 est publié sous la référence: Roberge, C., Einum, S., Guderley, H. et Bernatchez, L. 2006. Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. *Molecular Ecology* (Fast Track) **15,** 9-20.

Le chapitre 3 est publié sous la référence: Roberge, C., Normandeau, É., Einum, S., Guderley, H. et Bernatchez, L. 2008. Genetic consequences of interbreeding between farmed and wild Atlantic salmon : insights from the transcriptome. *Molecular Ecology* **18,** 314-324.

Le co-auteur Sigurd Einum y a contribué en faisant faire des croisements entre saumons d'élevages et saumons sauvages norvégiens de même qu'en faisant plusieurs suggestions pertinentes lors de la rédaction des articles.

Le co-auteur Éric Normandeau a fait les expériences de PCR quantitative présentées dans le troisième chapitre et a participé à la rédaction de la portion de l'article concernant ces résultats.

Le chapitre 4 est publié sous la référence: Roberge, C., Guderley, H. et Bernatchez, L. 2007. Genomewide identification of genes under directional selection : Gene transcription Q_{st} scan in diverging Atlantic salmon subpopulations. *Genetics* **177,** 1011-1022.

Une image présentant ce travail a d'ailleurs fait la page couverture de ce numéro de *Genetics*.

Le chapitre 5 est publié sous la référence: Roberge, C., Páez, D.J., Rossignol, O., Guderley, H., Dodson, J. et Bernatchez, L. 2007. Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology* **44,** 1374-1383.

Les co-auteurs David Páez et Orlane Rossignol y ont contribué en prenant soin des poissons au LARSA. David Páez a en outre échantillonné les poissons sains et malades dans les bassins. Le co-auteur Julian Dodson est le directeur de David Páez et a donné ses commentaires sur le texte de l'article.

Le chapitre 6 est publié sous la référence : Roberge, C., Blanchet, S., Guderley, H., Dodson, J. et Bernatchez, L. 2008. Disturbance of social hierarchy by an invasive species : a gene transcription study. *PLoS ONE* **3**, e2408, doi:10.1371/journal.pone.0002408.

Le co-premier auteur Simon Blanchet a participé à l'élaboration du plan de l'expérience et en a réalisé la partie comportementale (expérience, analyse, écriture). Le co-auteur Julian Dodson est le directeur de Simon Blanchet et a donné ses commentaires sur le texte de l'article.

À ma chère maman,

Table des matières

Liste des tableaux

Liste des figures

Chapitre 1 : Introduction générale

1.1 Évolution contemporaine, activités humaines et plasticité phénotypique

1.1.1 Évolution contemporaine et néo-darwinisme

La mise en évidence de cas d'évolution adaptative à petite échelle temporelle (entre une génération et quelques siècles; on parle alors d'évolution contemporaine, voir Stockwell et collaborateurs (2003)) en nature constitue un appui de taille à la théorie néo-darwiniste (Ridley 2004), mais reste une tâche ardue qui souvent se solde par l'échec, en partie à cause de la nature fluctuante des conditions environnementales et des pressions de sélection (Merila et coll. 2001). L'exemple le mieux documenté d'évolution adaptative récente en nature demeure ainsi celui du changement de la taille moyenne du bec de pinsons des îles Galápagos (Grant et Grant 1995) en réponse à la distribution de taille des graines disponibles. En outre, bien que la plupart des architectes de la théorie néo-darwinistes s'entendent à dire que la microévolution, soit l'évolution à l'intérieur ou entre les populations d'une même espèce (Hendry et Kinnison 2001), et la macroévolution, soit l'évolution conduisant à la dichotomie de groupes taxonomiques supérieurs, seraient le résultat des mêmes mécanismes appliqués soit sur une échelle de temps humaine (microévolution), soit sur une échelle de temps géologique (macroévolution) (Dobzhansky 1977), d'autres chercheurs suggèrent que ces deux types d'évolution pourraient ne pas différer seulement dans leur degré mais aussi dans la nature même des mécanismes évolutifs impliqués dans l'un et l'autre (Arthur 1997, Ohno 1970, Carroll 2000).

1.1.2 Plasticité phénotypique

Si les populations naturelles peuvent réagir à des changements environnementaux par des changements évolutifs, leurs individus peuvent également réagir en modifiant leur phénotype seulement. Ainsi, la plasticité phénotypique se définit comme la capacité d'un organisme d'un génotype donné à produire des phénotypes différents dans différentes conditions environnementales (Pigliucci 2001). Ceci peut inclure des différences comportementales, morphologiques ou physiologiques, lesquelles peuvent prendre place

durant le développement ou plus tard dans la vie de l'individu. Bien qu'évolution et plasticité phénotypique soient souvent présentées comme opposées, il existe des liens clairs entre la plasticité phénotypique et le potentiel évolutif d'une espèce ou d'une population. Ainsi, la plasticité phénotypique augmente la valeur adaptative des individus et un certain niveau de plasticité serait essentiel pour qu'une population soumise à des changements environnementaux ou introduite dans un nouvel environnement puisse persister (Price et coll. 2003). Les traits phénotypiques présentant de la plasticité pourraient alors évoluer par le mécanisme d'assimilation génétique (Waddington 1961). Par contre, un niveau de plasticité très grand pourrait, tout en assurant la persitance de la population dans de nouvelles conditions, réduire sa capacité à évoluer, à se spécialiser (Price et coll. 2003). Enfin, dans la mesure où il existe une variance génétique de la plasticité phénotypique pour un caractère phénotypique donné, c'est-à-dire un effet d'interaction entre génotype et environnement significatif influençant ce caractère, la plasticité phénotypique elle-même pourra évoluer pour ce caractère (Pigliucci 2001).

1.1.3 Évolution en milieu anthropique et élevage

On estime que les deux tiers de la surface non-immergée du globe sont actuellement occupés ou modifiés par l'homme et que la surface des habitats naturels des populations animales et végétales rétrécit encore d'un pourcent par an en moyenne (Smith et Bernatchez 2008). Les activités humaines engendrent en conséquence des pressions de sélection diverses et souvent très fortes sur des populations d'organismes vivants des quatre coins du monde; l'évolution contemporaine procède d'ailleurs à un taux plus important dans les contextes anthropiques que strictement naturels (Hendry et coll. 2008). Les exemples d'études dans lesquelles on a pu mettre en évidence des changements évolutifs en réponse à des activités humaines sont nombreux (Smith et Bernatchez 2008) et incluent certains des cas les plus célèbres d'évolution contemporaine, comme celui du mélanisme industriel chez *Biston betularia* (Kettlewell 1973) ainsi que ceux de l'émergence de la résistance aux insecticides chez les insectes, aux antibiotiques chez les bactéries, aux antiviraux chez les virus et aux herbicides chez les mauvaises herbes (Palumbi 2001).

Certains travaux suggèrent toutefois que beaucoup de populations naturelles s'éteindront faute d'avoir pu s'adapter aux changements rapides causés plus ou moins directement dans leurs habitats naturels par les activités humaines (incluant, notamment, les effets du réchauffement climatique) (Smith et Bernatchez 2008).

L'élevage est l'une des activités humaines qui entraînent des changements évolutifs importants, à la fois directement chez les populations soumises à la sélection artificielle et indirectement chez d'autres populations ou espèces, par l'intermédiaire des changements qu'il entraîne dans l'environnement. On parle ici de sélection artificielle plutôt que de sélection naturelle puisque, contrairement aux cas où l'homme provoque une réponse évolutive en exerçant, comme un prédateur classique, une sélection contre le phénotype des individus qu'il exploite préférentiellement (Coltman et coll. 2003, Conover et Munch 2002), il assure plutôt, dans le cas de l'élevage, la reproduction préférentielle de ces individus. Dans la mesure où il existe une base génétique aux caractères phénotypiques désirables, une population d'élevage répondra à la sélection artificielle selon l'une des équations fondamentales de la génétique quantitative :

$$R = h^2 S$$

où R est la réponse à la sélection, h^2 l'héritabilité au sens strict du caractère phénotypique sous sélection et S le coefficient de sélection (Falconer et Mackay 1996). L'élevage est pratiqué par l'homme depuis des millénaires et, dans bien des cas, les populations, lignées ou espèces domestiques sont maintenant fort différentes des populations naturelles qui leur sont le plus apparentées. S'il a permis une augmentation considérable de la production agricole (Ridley 2004), l'élevage charrie aussi son lot de d'embarras. Ainsi, la libération en nature d'animaux domestiqués peut représenter une menace pour les populations naturelles (Randi 2008). Aussi, la diversité génétique des populations d'élevage est souvent réduite; par exemple, plusieurs souches bovines comptant plusieurs millions d'individus ont une taille effective d'à peine quelques dizaines d'individus (Taberlet et coll. 2008). Ce fort niveau de consanguinité rend les individus de souches d'élevage plus sensibles à certaines maladies, moins fertiles et plus souvent affectés par des problèmes congénitaux que ceux

des populations sauvages et fait surgir un doute quand à la viabilité à moyen ou long terme de bon nombre de souches d'élevage (Rauw et coll. 2000, van der Waaij 2004, Taberlet et coll. 2008). Des souches sélectionnées pour une productivité accrue peuvent en outre perdre des caractéristiques qu'on les aurait voulu voir conserver. Par exemple, plusieurs cultivars de céréales et de légumes sélectionnés pour une croissance accélérée auraient vu, comme effet secondaire inopiné de cette sélection, leur valeur nutritive (contenu en protéines, vitamines et minéraux) diminuer au cours des 50 dernières années (Davis et coll. 2004).

1.2 Évolution contemporaine chez le saumon atlantique (*Salmo salar*)

1.2.1 L'élevage du saumon atlantique

1.2.1.1 Historique de l'élevage en Norvège et au Canada

C'est en Norvège que, suite au premier succès commercial de l'aquaculture de saumon atlantique en cages marines, débuta en 1971 une expérience de sélection artificielle dont les souches fournissent aujourd'hui plus de 80% des œufs de saumons destinés à l'élevage en Norvège (Gjedrem et coll. 1991). Quatre stocks de reproducteurs furent constitués en croisant 12 femelles et 4 mâles de chacun de 41 sites ou rivières échantillonnés à travers la Norvège (afin de s'assurer de pouvoir alimenter la sélection d'une large variation génétique). Toutefois, les descendants des géniteurs d'une à trois rivières dominèrent chaque souche dès la première génération de sélection. Par exemple, la première souche d'élevage norvégienne était, dès sa troisième génération, dominée à plus de 70% par la descendance des géniteurs de la rivière Namsen. Sachant que plus de 90% de la variance génétique additive des saumons norvégiens se retrouve à l'intérieur des populations plutôt qu'entre celles-ci (Gjedrem et coll. 1991, Gjoen et Bentsen 1997) et que des registres des croisements sont tenus pour limiter les croisements consanguins, les responsables du programme espèrent pouvoir éviter la dépression de consanguinité qui guette en principe les individus d'élevage (Gjoen et Bentsen 1997). L'existence de quatre stocks permet

d'approvisionner en œufs l'industrie de l'aquaculture chaque année (le temps de génération du saumon d'élevage est de trois à quatre ans).

Le régime de sélection artificielle initial du programme norvégien incluait à la fois une sélection familiale et individuelle à deux stades : 190 jours après la première alimentation (réduction du nombre annuel de familles de 150-200 à 120 en éliminant les familles où la survie et la croissance étaient faibles) et après deux ans en mer (indice de sélection tenant compte de la taille et de la famille). Ce régime de sélection artificielle s'est ensuite raffiné pour inclure d'autres caractères pouvant présenter un intérêt commercial (âge à la maturité sexuelle (sélection contre les grilses), résistance à la furonculose puis à l'anémie infectieuse du saumon, couleur de la chair, teneur en graisse de la chair). Pour chacun de ces caractères, la réponse obtenue à la sélection est de l'ordre de 10% par génération (Gjoen et Bentsen 1997).

Ainsi, en aussi peu que 6 générations de sélection, le taux de croissance des saumons d'élevage norvégiens a augmenté de 84% (en moyenne 14% par génération) par rapport à celui des saumons de la population sauvage à l'origine de l'élevage (voir http://www.akvaforsk.no/). La proportion de mâles à maturité sexuelle précoce a aussi considérablement diminué, tandis que la résistance aux maladies a été accrue et la coloration ainsi que la teneur en gras de la chair modifiées (Gjoen et Bentsen 1997). Outre ces caractères, pour lesquels il y avait sélection intentionnelle et intense, plusieurs autres changements phénotypiques à base génétique ont pu être mis en évidence entre les saumons d'élevage et les saumons sauvages : Fleming et Einum (1997) ont ainsi observé des différences morphologiques (taille des nageoires, forme du corps) et comportementales (agressivité, dominance, méfiance face aux prédateurs potentiels) alors que Thorstad et collaborateurs (1998) ont observé des différences dans le comportement migratoire. Fleming (2002) rapporte en outre une augmentation des niveaux plasmatiques et pituitaires de l'hormone de croissance chez les saumons d'élevage (et ce à trois stades de vie), ce qui s'accorde bien avec le taux de croissance supérieur de ceux-ci. Également en accord avec ce fait, Thodesen et collaborateurs (Thodesen et coll. 1999) ont mis en évidence que les

saumons d'élevage s'alimentaient davantage et faisaient une utilisation plus efficace de la nourriture ingérée.

Au début des années 80, l'ASF (*Atlantic Salmon Federation*; St-Andrews, Nouveau-Brunswick), en collaboration avec des chercheurs et des entreprises de l'est canadien, a lancé le SGRP (*Salmon Genetic Research Program*), un programme de sélection artificielle inspiré du programme norvégien et destiné à alimenter l'industrie de l'aquaculture de la baie de Fundy en saumons d'élevage sélectionnés à partir d'une population locale (afin, entre autres choses, de limiter les conséquences néfastes des échappées de saumon). Les saumons de la rivière St-Jean, qui s'étaient montrés particulièrement performants en élevage et où la proportion de grilses était plus faible que dans d'autres rivières du système (la rivière Big Salmon, par exemple), furent choisis pour géniteurs des stocks d'élevage. La sélection des géniteurs se faisait initialement à deux moments (à l'atteinte de la taille de mise en marché et à la maturité sexuelle) à l'aide d'indices combinant et pondérant les valeurs individuelles et familiales de masse corporelle aux divers stades et la proportion familiale de juvéniles ayant smoltifié (Friars et coll. 1995). Deux autres critères de sélection familiale, soit le pourcentage de non-grilses par famille et la résistance de «membres-cobayes» de chaque famille à la maladie bactérienne du rein (causée par la bactérie *Renibacterium salmoninarum*), ont été ajoutés par la suite (O'Flynn et coll. 1999). Le SGRP est devenu en 1998 l'ASBDP (*Atlantic Salmon Broodstock Development Program*) (Quinton et coll. 2004).

1.2.2.2 Conséquences pour les populations naturelles

On aurait pu s'attendre à ce que l'élevage et l'aquaculture du saumon atlantique améliorent la condition des populations sauvages en les soulageant d'une partie des pressions dues à la pêche commerciale. En effet, dans l'Atlantique nord, la production de saumon d'aquaculture a atteint 800 000 tonnes en 2004 tandis que la masse totale de saumons pêchés est passée de 10 000 tonnes par an dans les années 70 à 2100 tonnes en 2005 (Hindar et coll. 2006). Malgré cette diminution de la quantité de saumons pêchés, plusieurs chercheurs ont mis en évidence que la façon dont se pratique présentement l'élevage du

saumon représente une sérieuse menace pour les populations sauvages (McGinnity et coll. 2003, Gross 1998, Hindar et coll. 2006, Naylor et coll. 2005). La menace vient de ce qu'environ 2 millions de saumons d'élevage s'échappent chaque année dans la nature, alors que le nombre d'adultes sauvages est évalué à 4 millions pour l'espèce entière (McGinnity et coll. 2003); en effet, les saumons d'élevage sont engraissés dans des cages marines qui sont souvent endommagées lors de tempêtes ou par d'autres animaux marins. Dans certaines rivières situées près d'installations d'aquaculture, on trouve ainsi jusqu'à dix fois plus de saumons d'élevage fugitifs que de saumons sauvages. Ces saumons d'élevage menacent ceux des populations sauvages de plusieurs façons: en entrant en compétition avec ces derniers pour les ressources en rivière (espace, proies), en agissant comme vecteurs à la transmission de maladies (par exemple, la furonculose du saumon) et en se reproduisant avec eux (Naylor et coll. 2005).

La reproduction entre saumons d'élevage et saumons sauvages est problématique dans la mesure où les saumons d'élevage ont une valeur adaptative (*relative estimated lifetime success*: produit des taux de survie mesurés à différents stades du développement ontogénique) de l'ordre de 2 % de celle des poissons sauvages (McGinnity et coll. 2003), et que leur hybridation à ces derniers engendre une progéniture chez laquelle la mortalité embryonnaire est forte et dont la valeur adaptative est inférieure à celle des souches sauvages (McGinnity et coll. 2003), diluant l'adaptation des saumons sauvages à leur environnement. Les travaux de McGinnity et collaborateurs les amènent ainsi à conclure que les échappées de saumons d'élevage et les interactions génétiques de ceux-ci avec les individus de populations naturelles contribuent à accélérer la réduction des effectifs de ces populations naturelles, et ce sans doute jusqu'à leur extinction éventuelle (Hindar et coll. 2006, McGinnity et coll. 2003). En outre, il semble que les mâles à maturité sexuelle précoce descendant de souches d'élevage aient un succès reproducteur plus élevé que les mâles à maturité sexuelle précoces sauvages (Garant et coll. 2003b), ce qui pourrait favoriser l'introgression des caractères domestiques dans les populations naturelles en permettant à des individus descendant de souches d'élevage d'échapper à la mortalité importante à laquelle ils feraient face en mer. En effet, le taux de retour de la progéniture

des saumons d'élevage après un passage en mer est bien inférieur à celui des saumons sauvages (McGinnity et coll. 2003). Enfin, l'aquaculture du saumon atlantique se pratique également dans des régions d'où il n'est pas natif, comme sur la côte ouest du Canada et au Chili. Les saumons d'élevage fugitifs sont alors une espèce invasive qui perturbe les écosystèmes locaux (Naylor et coll. 2005).

1.2.2 Microévolution en nature : introduction du saumon atlantique sur une nouvelle portion de la branche nord-est de la rivière Sainte-Marguerite

L'introduction, intentionnelle ou non, d'individus d'une espèce donnée dans un habitat qu'elle n'occupait pas au préalable est souvent l'occasion d'études micro-évolutives, comme dans le cas pour les saumons sockeye (*Oncorhynchus nerka*) introduits dans le lac Washington dans les années 1910 (Hendry et coll. 2000), des ombres (*Thymallus thymallus*) introduites dans divers lacs norvégien entre 1880 et 1920 (Koskinen et coll. 2003) ou des saumons chinook (*Oncorhynchus tshawytscha*) introduits dans des rivières néo-zélandaises (Quinn et coll. 2001). Ces études profitent de ce que l'histoire évolutive et le temps de divergence entre les populations occupant diverses portions du nouvel habitat sont connus pour évaluer les taux évolutifs dans ces populations et les effets de la sélection naturelle et de la dérive génétique sur l'évolution des caractères phénotypiques mesurés.

La région du Saguenay nous offre un site naturel idéal à ce type d'étude. En effet, la branche nord-est de la rivière Sainte-Marguerite (48º20'N, 70º00'W) comporte plusieurs segments isolés par des chutes infranchissables pour les saumons. L'un de ces segments leur a récemment été rendu accessible par l'installation d'une passe migratoire (il y a 30 ans, ou environ 6 générations). Une nouvelle population s'y est formée qui semble s'être en partie isolée au niveau reproducteur de la population de l'aval de la chute par sa fidélité à ce nouveau site de frai (Garant et coll. 2000). Certaines données obtenues par des chercheurs du CIRSA suggèrent en outre une différence dans la proportion de mâles à maturité sexuelle précoce entre les saumons de la population de l'amont de la chute et ceux de la population de l'aval (Aubin-Horth et coll. 2006).

1.3 Scruter les niveaux d'ARN messagers: la technique des bio-puces et son application à des questions évolutives

1.3.1 Survol de la technique

La technique centrale aux travaux présentés dans cette thèse est celle des bio-puces, aussi appelées puces à l'ADN, *microarrays* ou *DNA chips* (Knudsen 2002, Bowtell et Joseph 2003, Mills et coll. 2001, Schulze et Downward 2001). Plusieurs variantes en existent, mais l'idée de base reste toujours de permettre la comparaison des niveaux de transcription de plusieurs milliers de gènes dans plusieurs échantillons. Pour ce faire, de l'ADN représentant les milliers de gènes dont on veut tester en parallèle les niveaux relatifs de transcription est déposé puis fixé à des positions connues sur une lamelle de verre par un dispositif robotisé; il s'agit le plus souvent de portions d'ADN complémentaires amplifiées par PCR (0.5 à 1 kb). Il reste ensuite à préparer les « sondes complexes ». Pour ce faire, on extrait l'ARN (total ou poly-adénylé) des échantillons entre lesquels on veut comparer les niveaux de transcription des gènes représentés sur la puce, puis on procède à la transcription inverse de cet ARN et à son marquage (le plus souvent par des composés fluorescents). Chaque sonde complexe obtenue est donc un ensemble de molécules marquées dérivées des ARNm d'un échantillon et qui en conserve, en principe, les proportions. Après incubation de la bio-puce et des sondes, l'intensité du signal fluorescent aux points où les sondes se sont fixées est détectée par un scanner, lequel transmet une numérisation de ces résultats à un ordinateur. L'intensité de la fluorescence en un point est liée au niveau de transcription du gène représenté à ce point dans l'échantillon correspondant.

Filles d'une course à la miniaturisation, les bio-puces sont très sujettes aux variations de conditions expérimentales, à la fois durant leur «impression» (dépôt des ADNc amplifiés par PCR sur la lame) et durant les hybridations, d'où la nécessité d'en normaliser les résultats avant de les analyser (Draghici 2003, Knudsen 2002, Leung et Cavalieri 2003, Quackenbush 2002, Speed 2003). Les stratégies de design expérimental et d'analyse

statistique des résultats de bio-puces, d'abord simplistes (par exemple: ne considérer que les gènes montrant une différence moyenne de deux fois ou plus du niveau de transcription entre les types d'échantillons étudiés), se sont développées rapidement et une vaste de gamme de logiciels pour l'analyse statistique de données de bio-puces est maintenant disponible (Tusher et coll. 2001, Kerr et coll. 2002, Kerr et coll. 2000, Storey et Tibshirani 2003, Storey 2002, Meiklejohn et Townsend 2005, Shieh et coll. 2008, Hsiao et coll. 2005, Liu et coll. 2003, Dudoit et coll. 2003).

Les puces utilisées dans ce projet ont été développées par des chercheurs du cGRASP (*Consortium Genomic Research on All Salmon Project*) dans le cadre d'un projet financé par Génome Canada et sont de deux types : 3557 clones d'ADNc de salmonidés sont déposés sur les unes et 16 006 sur les autres. Elles sont le fruit du développement d'une banque détaillée d'EST (*Expressed Sequence Tag*) séquencés dans 35 banques d'ADNc, dont 23 de divers tissus de saumon atlantique (Rise et coll. 2004b) et représentent l'un des rares exemples de développement de ce type de ressource chez une espèce dite «non-modèle». D'autres exemples incluent notamment la puce du téléoste *Fundulus fundulus* (Oleksiak et coll. 2001, Oleksiak et coll. 2002) et les puces d'abeille mellifère (Grozinger et Robinson 2002, Whitfield et coll. 2002).

1.3.2 Applications des bio-puces à des questions évolutives

1.3.2.1 Pertinence

Des gènes à l'activité des protéines qu'ils encodent, plusieurs points de contrôle existent dont chacun pourra avoir une influence sur le phénotype. Si des mutations survenant dans des gènes de structure peuvent alimenter la sélection naturelle en diversité génétique, des chercheurs ont, il y a longtemps déjà, lancé l'idée que les mutations survenant dans des gènes régulateurs ou des régions régulatrices non transcrites pourraient jouer un rôle évolutif encore plus important que les mutations survenant dans des gènes de structure (voir, par exemple, (King et Wilson 1975). Dans ce contexte, l'utilisation de bio-puces, lesquelles avaient d'abord été développées pour et appliquées à des études biomédicales

(toxicologie, immunologie, cancérologie), a récemment permis de mettre en évidence (1) que les profils de transcription présentaient une forte variation naturelle interindividuelle à l'intérieur de populations naturelles (Townsend et coll. 2003, Oleksiak et coll. 2002, Oleksiak et coll. 2005), (2) que ces profils étaient en partie héritables (Brem et coll. 2002) et (3) qu'ils pouvaient répondre rapidement à la sélection (Ferea et coll. 1999). Rassemblés, ces trois éléments suggèrent que la variation naturelle des profils de transcriptions géniques pourrait participer de façon importante à l'évolution en nature (Townsend et coll. 2003). En guise d'exemple, l'humain et les grands singes présentent un très fort degré de similitude dans leurs séquences codantes d'ADN génomique mais d'importantes différences phénotypiques (tout spécialement au niveau des facultés cognitives). L'utilisation de bio-puces a récemment permis de résoudre en partie cette apparente contradiction en mettant en évidence des changements importants dans les niveaux de transcription géniques entre humains et grands singes, et ce particulièrement au cerveau (Enard et coll. 2002, Uddin et coll. 2004).

Stern (2000) propose en outre que des changements accumulés dans les séquences régulatrices en *cis* de gènes du développement, en particulier de gènes *Hox*, pourraient expliquer l'évolution graduelle, par accumulation de «micro-mutations», de plans corporels différents (sans avoir, donc, à invoquer l'occurrence de « macro-mutations»).

1.3.2.2 Transcription et évolution parallèle

L'évolution parallèle se définit comme l'évolution indépendante d'un même caractère phénotypique dans des populations apparentées (Futuyma 1986) et, lorsqu'associée à un changement similaire d'environnement, est généralement considérée comme une manifestation de l'effet de la sélection naturelle (Schluter et coll. 2004). Haldane (1932) a proposé que l'évolution parallèle puisse être le résultat non seulement de pressions de sélection similaires, mais également d'une production de variation héritable plus semblable entre groupes apparentés qu'entre groupes distants. Dans la même ligne de pensée, West Eberhard (2003) propose que l'évolution parallèle puisse être le résultat de contraintes développementales partagées entre groupes apparentés, impliquant également l'effet d'un déterminisme génétique dans l'évolution.

Si l'évolution parallèle est en partie causée par la production de variation héritable plus semblable entre groupes apparentés qu'entre groupes distants, on s'attendra à ce que les mêmes gènes soient modifiés de façon plus fréquente qu'attendu au hasard dans des cas indépendants d'évolution d'un même phénotype (Schluter et coll. 2004). En fournissant un vaste tableau des différences d'expression génique entre diverses souches apparentées, les bio-puces peuvent permettre d'évaluer l'importance d'un déterminisme génétique du niveau d'expression des gènes dans divers cas d'évolution contemporaine en parallèle. Ferea et collaborateurs (1999) ont ainsi pu mettre en évidence des différences dans les niveaux de transcription de plusieurs centaines de gènes suite à l'adaptation de trois souches de levures à un milieu pauvre en glucose. Chose intéressante, les niveaux de transcription de beaucoup de gènes du métabolisme énergétique étaient modifiés de la même façon dans ces trois souches (par rapport à la souche de base non-adaptée). Des résultats similaires ont depuis été obtenus chez *Escherichia coli* (Cooper et coll. 2003). Ces exemples suggèrent que des variations dans l'expression des gènes peuvent, lorsque soumises à une sélection, être à la base d'une évolution adaptative rapide. Ils suggèrent en outre un certain déterminisme génétique des les mécanismes de l'adaptation à un milieu pauvre en glucose dans ces deux systèmes.

1.3.2.3 Évolution des profils de transcription en nature : neutralité, sélection directionnelle ou sélection stabilisatrice?

Les résultats de plusieurs études suggèrent que les niveaux de transcription de la plupart des gènes seraient sous forte sélection stabilisatrice en nature (Denver et coll. 2005, Rifkin et coll. 2005, Gilad et coll. 2006, Lemos et coll. 2005). Se basant sur la corrélation entre la divergence observée des profils de transcription et celle observée pour des marqueurs génétiques neutres, d'autres auteurs proposent quant à eux que les niveaux de transcription de la majorité des gènes évolueraient plutôt par dérive génétique (Khaitovich et coll. 2005). Une divergence adaptative des profils de transcription en réponse à une sélection directionnelle a également pu être mise en évidence, mais seulement pour une minorité de gènes en général (Rifkin et coll. 2003, Lemos et coll. 2005). Par exemple, Bochdanovits et collaborateurs (2003) ont pu mettre en évidence une forte corrélation entre le patron de

transcription de 19 gènes chez les larves et la masse des adultes dans deux populations naturelles de *Drosophila melanogaster*. Chez les drosophiles, l'augmentation de la masse aux latitudes plus septentrionales représenterait une adaptation au froid. Ces résultats appuient ainsi l'hypothèse de l'évolution adaptative de ce caractère phénotypique quantitatif dans les populations naturelles par modification des profils de transcription génique. Cette étude a également permis d'estimer les contributions relatives de la plasticité phénotypique et des changements génétiques dans les modifications de profils d'expression observées.

1.3.2.4 Génétique quantitative de la transcription

Les bio-puces offrent en outre un moyen d'explorer à grande échelle l'héritabilité, l'additivité et l'architecture génétique de la transcription génique prise comme phénotype (Gibson et Weir 2005). Quelques études ont pu obtenir des estimés d'héritabilité des profils de transcription chez divers organismes modèles (souris : (Schadt et coll. 2003), (Chesler et coll. 2005), (Cui et coll. 2006), humain :(Monks et coll. 2004), levure : (Brem et coll. 2002), mais avec des résultats fort divers. Ainsi, lorsqu'indiquée, la proportion des gènes testés dont le niveau de transcription avait une héritabilité significative était de l'ordre de 30% et l'héritabilité médiane du niveau de transcription pour ces gènes variait de 0.11 (Chesler et coll. 2005) à 0.84 (Brem et coll. 2002).

Afin de déterminer si la variabilité des niveaux d'expression est généralement additive ou non, certaines études récentes ont utilisées des bio-puces pour comparer, chez diverses espèces animales ou végétales, les niveaux de transcription entre des souches parentales et divers croisements obtenus à partir de celles-ci. Dans la plupart de ces études, on a observé, pour une majorité de gènes, des patrons de transmission du niveau de transcription qui suggèrent une base génétique non-additive (Gibson et coll. 2004, Vuylsteke et coll. 2005, Auger et coll. 2005, Hedgecock et coll. 2007). Par exemple, Gibson et collaborateurs (2004) ont observé de nombreux gènes dont le niveau de transcription en F1 était soit plus fort ou plus faible que celui observé dans les deux souches parentales. D'autres auteurs ont

quant à eux conclu qu'une majorité des profils de transcriptions observés étaient sous contrôle génétique additif (Swanson-Wagner et coll. 2006, Cui et coll. 2006).

Enfin, d'autres études ont été entreprises pour évaluer à l'échelle du génome complet la liaison génétique entre des marqueurs génétiques neutres et les loci responsables de la variation du niveau de transcription de milliers de gènes (Brem et coll. 2002, Chesler et coll. 2005, Yvert et coll. 2003, Kirst et coll. 2004, Morley et coll. 2004, Hubner et coll. 2005, Potokina et coll. 2008, West et coll. 2007); les régions génomiques présentant de telles associations sont nommées eQTL (*expression quantitative trait loci*). La plupart de ces études ont conclu que, bien que le niveau de transcription soit en général affecté par les effets de plusieurs loci, les eQTL expliquant une large portion (25 à 50%) de la variation des profils de transcription géniques sont fréquents. On pourrait ainsi s'attendre à ce que le niveau de transcription des gènes ne suive pas, dans une population, une distribution normale, unimodale par définition, mais arbore plutôt un nombre de modes discrets correspondant aux allèles du ou des eQTL dont les effets sont les plus importants (Gibson et Weir 2005).

1.4 Objectifs de recherche

L'un des objectifs centraux de cette thèse sera d'étudier les mécanismes moléculaires de l'évolution contemporaine chez le saumon atlantique à l'aide de bio-puces. L'utilisation de bio-puces pour l'étude d'espèces dites «non-modèles», chez lesquelles il est possible d'étudier en nature les processus évolutifs, est souvent compliquée par le fait que peu de données de séquence existent chez ces espèces (Feder et Mitchell-Olds 2003). Pour les travaux qui seront présentés dans cette thèse, j'ai eu la chance de pouvoir disposer de bio-puces développées par des chercheurs du cGRASP (*Consortium Genomic Research on All Salmon Project*) et sur lesquelles 3557 et 16 006 gènes de salmonidés sont représentés.

Dans la première partie (Chapitre 2), je présente une étude visant à évaluer l'importance des changements héritables du niveau de transcription entraînés par l'élevage au Canada et

en Norvège et à identifier les principaux gènes pour lesquels l'élevage a entraîné un changement détectable du niveau de transcription. Pour ce faire, j'ai comparé les niveaux moyens de transcription de 3557 gènes dans la progéniture de géniteurs d'élevage ou de géniteurs sauvages provenant de la principale rivière d'origine de chacune des souches d'élevage, et ce pour une souche d'élevage canadienne ainsi que pour une souche norvégienne. Les comparaisons ont été effectuées chez des alevins au stade de l'émergence (résorption du sac vitellin) et ayant grandi dans des conditions identiques. L'importance de la croissance aux jeunes stades de vie dans le développement ontologique chez les salmonidés (Metcalfe et coll. 1989) rend ces stades particulièrement pertinents à notre étude. L'élevage du saumon est un cas d'évolution contemporaine très récent: pour cette étude, nous avons utilisé une souche d'élevage canadienne ayant été soumise à 5 générations de sélection et une souche norvégienne soumise à 7 générations de sélection. L'importance de comprendre les bases génétiques des changements entraînés chez le saumon par la domestication est accrue du fait que les saumons d'élevage constituent vraisemblablement une menace pour les populations naturelles de saumon. Une meilleure connaissance des bases génétiques des différences entre saumons d'élevage et sauvages pourrait permettre de mieux évaluer les risques des échappées de saumons d'élevage (Bentsen 1991). De plus, étant donné la similitude des régimes de sélection artificielle norvégien et canadien et le parallélisme constaté dans l'évolution du taux de croissance et de la proportion de grilses chez les souches d'élevage norvégiennes et canadiennes (Gjoen et Bentsen 1997, O'Flynn et coll. 1999), notre étude représente un modèle d'évolution parallèle permettant d'évaluer les bases génétiques de ce phénomène et, notamment, le déterminisme génétique de la variation du taux de croissance. Nos hypothèses de travail pour cette première partie du projet étaient donc que (1) 5 à 7 générations de sélection artificielle ont entraîné des changements héréditaires et détectables des profils d'expression de saumons d'élevage canadiens et norvégiens par rapport à leurs populations sauvages d'origine et que (2) les changements d'expression détectés chez les saumons d'élevage canadiens et norvégiens reflètent en partie le parallélisme des pressions de sélection artificielle dans les deux systèmes.

Dans le troisième chapitre de la présente thèse, je présente une étude complémentaire à celle décrite au deuxième chapitre. Les niveaux de transcription de 16006 gènes y sont comparés entre la progéniture de rétro-croisements d'hybrides «sauvage X domestique» norvégiens et celle d'individus sauvages de la rivière Namsen. Les comparaisons ont encore ici été effectuées chez des alevins au stade de l'émergence (résorption du sac vitellin) ayant grandi dans des conditions identiques. Le rétrocroisement d'individus hybrides à des saumons sauvages est le parcours obligé pour l'introgression de caractères domestiques dans les populations naturelles. Or, si l'on peut s'attendre à ce que les différences génétiques pour la plupart des caractères phénotypiques quantitatifs s'estompent suite au croisement répété d'individus hybrides à des individus sauvages (voir, par exemple, McGinnity et coll. 2003, Tymchuk et coll. 2006), les bases génétiques des profils de transcription pourraient être largement non-additives (Gibson et coll. 2004, Hedgecock et coll. 2007). L'exploration de l'additivité des bases génétiques des niveaux de transcription dans ce système pourrait, en plus de présenter un intérêt théorique, permettre de mieux évaluer les risques encourus par les populations naturelles interagissant avec des saumons d'élevage fugitifs. Nos hypothèses de travail pour cette deuxième partie du projet étaient ainsi que (1) la progéniture de rétro-croisements d'hybrides «sauvage X domestique» à des individus sauvages conservera des différences détectables de profils de transcription par rapport à celle de poissons sauvages et que (2) nous observerons des patrons de transmission du niveau de transcription incompatibles avec une base génétique additive. Ce chapitre présente en outre la vérification, par la technique de la PCR quantitative, des résultats obtenus pour deux gènes dans le deuxième chapitre.

Dans le quatrième chapitre, je présente une étude visant à identifier des gènes montrant, entre deux sous-populations de saumon de la rivière Sainte-Marguerite, des différences héréditaires dans leurs niveaux de transcription et, plus précisément, des gènes pour lesquels ces différences seraient le résultat de la sélection naturelle. Pour ce faire, nous proposons une nouvelle méthode, le balayage transcriptomique («*Q_{st} scan*»), dont le principe est de rechercher les gènes pour lesquels le Q_{st} du niveau de transcription est particulièrement élevé. Le Q_{st} (Spitze 1993) est un indice permettant d'évaluer le degré de

différentiation entre populations pour un ou plusieurs caractères phénotypiques. Il se définit ainsi :

$$Q_{st} = \sigma^2_{GB} / (\sigma^2_{GB} + \sigma^2_{GW})$$

où σ^2_{GB} et σ^2_{GW} représentent respectivement les composantes inter- et intra-population de la variance génétique additive. Ce balayage du transcriptome se base sur le même principe que l'approche du «balayage du génome» (*genome scan*), dans laquelle on recherche, pour de nombreux marqueurs génétiques neutres, les loci présentant des patrons atypiques de variation (Storz 2005). En effet, puisque l'on s'attend à ce que le flux génique affecte tout le génome et la sélection seulement certains loci spécifiques (Luikart et coll. 2003), les régions génomiques ou gènes présentant une différentiation particulièrement forte ou faible entre populations auront potentiellement été sous l'effet de la sélection directionnelle ou stabilisatrice, respectivement. Dans ce contexte, l'estimation de la variance génétique additive des niveaux de transcription à l'aide d'un modèle animal (Kruuk 2004) nous permet d'estimer pour chaque gène le Q_{st} du niveau de transcription. Elle nous permet également d'évaluer l'héritabilité des niveaux de transcription, puisque l'héritabilité au sens strict se définit comme :

$$h^2 = V_A/V_P$$

où V_A et V_P représentent la variance génétique additive et la variance phénotypique, respectivement (Falconer et Mackay 1996).

Par ailleurs, Gibson et Weir (2005) suggèrent que, puisque les QTL d'expression à effets majeurs sont courants, la distribution de données de transcription pour un gène donné dans une population devrait souvent s'éloigner de la distribution normale pour prendre plutôt la forme d'une distribution bi- ou multi-modale. Notre étude vise donc aussi à évaluer la normalité ou la bi- ou multi-modalité des donnés de transcription. Pour ce faire, les niveaux de transcription de 16006 gènes ont été comparés chez 80 individus issus de croisements entre géniteurs de deux sous-populations de la rivière Sainte-Marguerite ayant commencé à

divergé il y a seulement 6 générations. Nos hypothèses de recherche pour cette étude étaient ainsi que (1) certains gènes présenteront des différences significatives du niveau de transcription entre les deux sous-populations à l'étude, que (2) les niveaux de transcription de certains gènes auront une héritabilité significative, que (3) les gènes dont le niveau de transcription a une valeur de Q_{st} particulièrement forte ont potentiellement vu leur niveau de transcription modifié par la sélection directionnelle en nature et que (4) la distribution des niveaux de transcription de certains gènes à travers tous les individus s'éloignera d'une distribution normale pour plutôt arborer une distribution bi- ou multi-modale. Cette étude vise ainsi à comprendre les mécanismes de l'évolution contemporaine au niveau du transcriptome dans un cas très récent de divergence en nature.

Les deux derniers chapitres de cette thèse présentent l'étude au niveau transcriptomique de deux cas de plasticité phénotypique chez le saumon atlantique dans le contexte d'environnements modifiés par l'homme. Le cinquième chapitre est habité par une étude constituant un exemple plus classique d'utilisation des bio-puces, puisqu'elle vise à identifier des gènes impliqués dans la réponse immunitaire du saumon à une mycose cutanée courante en milieu d'aquaculture, la saprolegniose. Ainsi, la forte densité de poissons en milieu d'aquaculture favorise l'émergence et la prolifération de diverses maladies (Naylor et coll. 2005). La saprolegniose cause d'importantes pertes économiques pour l'aquaculture partout à travers le monde, et les salmonidés y sont particulièrement vulnérables (Tampieri et coll. 2003, Hughes G.C. 1994). Dans cette étude, j'ai comparé les niveaux de transcription de 16006 gènes chez des juvéniles de saumons sains ou atteints de saprolegniose. Nos hypothèses de recherche pour cette section étaient (1) qu'il existe des différences détectables dans les niveaux de transcription de certains gènes entre des saumons sains et des saumons atteints de saprolegniose, que (2) ces différences sont en partie le résultat d'une réponse immunitaire des saumons et que (3) certains gènes précédemment décrits comme présentant des différences d'expression chez des poissons affectés par d'autres maladies en présenteront également ici.

Enfin, le sixième chapitre présente une étude visant à identifier les bases moléculaires de la plasticité phénotypique d'un comportement social chez saumon. Les activités humaines

facilitent l'établissement d'espèces dans des habitats qu'elles n'occupaient pas au préalable et ces espèces, que l'on qualifie alors d'invasives ou d'envahissantes, peuvent avoir d'importants effets évolutifs et écologiques sur les écosystèmes qu'elles «envahissent» (Mack et coll. 2000, Strauss et coll. 2006, Suarez et Tsutsui 2008) et en menacer la biodiversité (Clavero et Garcia-Berthou 2005). Ainsi, la truite arc-en-ciel (*Oncorhynchus mykiss*) est native des rivières tributaires de l'océan pacifique mais a été introduite par l'homme dans plusieurs autres environnements (Leprieur et coll. 2008), dont plusieurs rivières de la côte est canadienne. Blanchet et collaborateurs (2007) ont récemment mis en évidence que la présence de truites arc-en-ciel exotiques perturbe fortement la hiérarchie sociale chez des juvéniles de saumon atlantique. Les juvéniles de saumon atlantique sont territoriaux et forment spontanément des hiérarchies sociales en nature et en captivité (Overli et coll. 1999); le statut de dominance d'un saumon juvénile aurait une forte influence sur sa stratégie reproductive future (Metcalfe et coll. 1989). Ceci constitue un cas de plasticité phénotypique comportementale puisque les saumons modifient leur comportement en réponse à un environnement social différent; la plasticité phénotypique peut résulter en évolution moléculaire via le phénomène d'assimilation génétique (West-Eberhard 2005, Price et coll. 2003). Afin d'identifier des acteurs moléculaires impliqués dans la perte plastique de dominance sociale chez les jeunes saumons en présence de truite, j'ai comparé le niveau de transcription de 16006 gènes dans les cerveaux de saumons dominants ou subordonnés en présence ou en l'absence de truite arc-en-ciel. Nos hypothèse de recherche pour cette section étaient ainsi (1) qu'il existe des différences détectables dans le niveau d'expression de certains gènes entre saumons dominants et subordonnés et que (2) les profils d'expression chez les saumons dominants et subordonnés seront modifiés par la présence de truite arc-en-ciel.

Chapitre 2 : Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon

2.1 Résumé

On a sélectionné le saumon d'élevage pour en augmenter le taux de croissance ainsi que pour en modifier d'autres caractères phénotypiques d'intérêt commercial, mais les quelque 2 millions de saumons d'élevage qui s'échappent chaque année augmenteraient les risques d'extinction de plusieurs populations sauvages de saumon atlantique via leurs interactions génétiques et écologiques avec ceux-ci. Dans cette étude, nous comparons les niveaux de transcription de 3557 gènes entre la progéniture de saumons d'élevage et sauvages canadiens et norvégiens grandie en conditions contrôlées et mettons en évidence que 5 à 7 générations de sélection artificielle ont mené à des changements héritables du niveau de transcription pour 1.4% et 1.7% des gènes exprimés chez les saumons juvéniles de Norvège et du Canada, respectivement. La magnitude moyenne des différences observées est de 25% en Norvège et de 18% au Canada. De plus, tous les gènes dont le niveau de transcription est significativement modifié dans les deux souches d'élevage (16%) y montrent un changement parallèle. Ces résultats, de même que l'identification de plusieurs gènes dont le niveau de transcription a été modifié par la sélection artificielle, permettent de mieux comprendre les bases moléculaires de l'évolution parallèle dans ce système et appuient l'idée que le flux génique entre populations d'élevage et sauvages pourrait affecter l'intégrité génétique des populations sauvages.

2.2 Abstract

Farmed salmon strains have been selected to improve growth rates as well as other traits of commercial interest but the 2 million farmed salmon escaping annually may enhance the risk of extinction of wild populations through genetic and ecological interactions. Here, we compare the transcription profiles of 3557 genes in the progeny of farmed and wild Atlantic salmon from Norway and Canada grown in controlled conditions, and demonstrate that five to seven generations of artificial selection led to heritable changes in gene transcription profiles, the average magnitude of the differences being 25% and 18% for at least 1.4% and 1.7% of the expressed genes in juvenile salmon from Norway and Canada, respectively. Moreover, genes showing significant transcription profile differences in both farmed strains (16%) all exhibited parallel changes. These findings, along with the identification of several genes whose expression profiles were modified through artificial selection, provide new insights into the molecular basis of parallel evolution, and suggest how gene flow from farmed escapees may affect the genetic integrity of wild populations.

2.3 Introduction

Consumers and environmentalists are increasingly concerned about the potential health and environmental consequences of producing genetically modified organisms. Yet, some of the criticisms of transgenic plants and animals may also apply to artificially selected breeds. In animals, strains selected for high production efficiency may be more susceptible to behavioural, physiological and immunological problems than wild populations (Rauw *et al*. 2000, Rauw *et al*. 1998, van der Waaij 2004). Also, many of the genetic changes accumulated in breeding strains are likely to be unknown.

Selective breeding of Atlantic salmon (*Salmo salar*) was initiated in Norway in 1971 and is now intensively practised in Chile, United Kingdom, United States and Canada. In Norway, artificial selection was first limited to improving growth rate, but currently also targets traits such as age at sexual maturity, bacteria resistance, fat content and flesh colour. The selection response on each of these traits has been approximately 10% per generation (Gjoen and Bentsen 1997). However, phenotypic changes not specifically selected for have also been observed in Norwegian farmed salmon, including increased fat content in flesh (Rye and Gjerde 1996), poorer performance in the wild (McGinnity *et al*. 2003, Fleming *et al*. 2000) morphological and behavioural changes (Fleming and Einum 1997), increased growth hormone (GH) and insulin-like growth factor (IGF) levels (Fleming *et al*. 2002), as well as a higher feeding rate and food conversion efficiency (Thodesen *et al*. 1999). In Canada, the first breeding programme was initiated during the late-eighties, and targeted growth rate, age at sexual maturity, and more recently bacteria resistance (Friars *et al*. 1995, O'Flynn *et al*. 1999).

The last decade has seen the world-wide production of farmed Atlantic salmon outstrip that of fisheries (FAO 2004) and the problem of farmed salmon escapees has taken alarming proportions, with about two million farmed salmon escaping annually from their sea cages while the natural populations in the North Atlantic are estimated to comprise approximately four million individuals (McGinnity *et al*. 2003). Escaped farmed salmon are thought to greatly enhance the wild populations' risk of extinction through genetic and ecological interactions (McGinnity *et al*. 2003, Fleming *et al*. 2000), as well as by spreading diseases

(Naylor *et al*. 2005). Knowledge of evolutionary changes induced by salmon breeding is crucial for assessing the consequences of genetic interactions between wild and escaped farmed salmon (Bentsen 1991). Namely, it is increasingly acknowledged that evolutionary changes may strongly depend on alterations in gene regulation (Wilson 1976), and the microarray technology offers the possibility of a genome-wide scan for gene transcription differences between farmed and wild salmon. Yet, the oldest salmon breeding programme has only operated for eight generations and no microarray study in any species has assessed the accumulation of heritable expression differences following so few generations of divergence. Recent studies in salmonids suggest that evolution may be rapid for phenotypic traits under strong selection (Hendry *et al*. 2000, Koskinen *et al*. 2003), but the genomic basis for such rapid changes is unknown.

Due to the similarities in controlled selective regimes, breeding strains also represent a powerful system for studies on parallel evolution. Parallel phenotypic evolution is considered one of the most eloquent manifestations of the role of selection in driving adaptive change (Harvey and Pagel 1991). Yet the genetic bases of parallel evolution are poorly understood (Schluter *et al*. 2004). Genetic changes implicated in natural phenotypic variation and evolution of a given trait often involve a disproportionate use of the same genes in independent parallel instances of phenotypic evolution instead of affecting genes randomly picked from the pool of all genes influencing this trait (Schluter *et al*. 2004, Remington and Purugganan 2003). This could either be due to the similarly biased production of genetic variation in close relatives (Schluter *et al*. 2004, Haldane 1932) or to a limited number of genes being actually important in phenotypic variation and evolution for a given trait (Remington and Purugganan 2003, Stern 2000). To our knowledge, very few published studies have specifically investigated parallelism in gene transcription (Cooper *et al*. 2003, Ferea *et al*. 1999).

Here, we used a 3557 gene cDNA array (Rise *et al*. 2004b) to compare levels of gene transcription in the progeny of Atlantic salmon from the Norwegian National breeding program first brood line (DOM1 or “population 1” sensu (Gjedrem *et al*. 1991)) and its main population of origin (River Namsen, N 64°27'0" E 11°28'0), and from the Canadian

breeding programme (strain 84JC, Saint John River, from the Atlantic Salmon Broodstock Develoment Program in Saint Andrews, New Brunswick) and its population of origin (St-John River, N 50°17'0" W 64°20'0). Salmon of wild and farmed origin were reared under identical conditions within each country and transcription profiles of fish from the farmed strains and their wild counterparts were contrasted on 23 microarrays (Canada: 13 involving 26 fry, Norway: 10 involving 20 fry).

2.4 Materials and methods

2.4.1 Fish crosses

Twenty wild and twenty farmed genitors were used to produce ten full-sibs pure wild families and ten pure farmed families in both Norway and Canada, for a total of 40 families. Norwegian genitors were from the seventh generation of the Norwegian breeding program first brood line and Canadian ones from the fifth generation of brood strain 84JC. Fertilised eggs of both types were kept under identical controlled conditions within systems (24 h darkness; yet temperatures were slightly different between countries: Norway: mean temperature (range) 3.9 °C (0.8-11.1), Canada: mean water temperatures 0.5 to 1°C from January to March, gradually rising in April to 4°C, and from 4°C to 9°C by the first of June). Because juvenile characteristics play a determinant role in the expression of life history traits at later stages (Metcalfe *et al*. 1989), this study focuses on young of the year (fry stage). Sexually undifferentiated fry were sampled at the yolk-sac resorption stage, before exogenous feeding. The number of families used is relatively high compared to current standards for microarray experiments (see for instance MacKay *et al*. (2005)), minimising biases associated with individual parental effects, either by randomised sampling (Canada) or by using the same number of fry from each family (Norway). For the Canadian fish, however, some of the families may have been over-represented (and others under-represented) in the random sample. Genitors of wild origin were either grown from the egg in controlled conditions (Norway) or caught in the wild and kept in controlled conditions for several months until spawning (Canada), allowing us to control for environmental effects. The rearing of the parents in a controlled environment is also

expected to considerably reduce environmental maternal effects (where offspring from different females have different characteristics due to their mothers having lived in different environments). Hence, differences in gene transcription related to maternal effects are expected to be mainly due to genetically based maternal effects, which are heritable and may respond to selection. These may thus be an important source of evolutionary changes among offspring of wild and farmed stocks, at least during early stages, when genetic maternal effects have most influence in salmonids (Perry *et al*. 2005).

2.4.2 RNA extraction, labelling and cDNA hybridisation

Following sampling, all experiments were performed by the same person (CR), in the same laboratory (Pavillon Marchand, Université Laval, Québec). Juveniles that had been frozen in liquid nitrogen were homogenised individually in 4 ml TRIZOL@Reagent using a Polytron homogeniser, and total RNA was extracted in four separate assays (to limit the error associated with the variability in the RNA extraction efficiency). Briefly, 200 µL chloroform (Sigma) was added to each ml of fish homogenate in Trizol. After mixing and centrifuging (12 000 rpm, 0°C, 15 min), the aqueous layers were transferred in new tubes and 1ml isopropanol (Sigma) was added. The samples were then stored overnight at -80°C. The following day, the samples were centrifuged one hour (12 000 rpm, 0°C) and the isopropanol discarded. The pellets were washed with 1ml 70% ethanol, dried for 15 min at room temperature, resuspended in 40µl non-DEPC treated nuclease-free water (Ambion) and spiked with 1 µl RNAse inhibitor (Ambion). RNA integrity was verified with a 2100 BioAnalyzer (Agilent). For each sample, 15 µg of the pooled RNA of the four separate extractions was then retro-transcribed and labelled using Genisphere 3DNA Array 50 kit, Invitrogen's Superscript II retro-transcriptase and Cy3 and Alexa 647 dyes (Genisphere). The detailed protocol of the retro-transcription, labelling and hybridisation procedures can be found at http://web.uvic.ca/cbr/grasp/ (Genisphere Array 50 Protocol). Briefly, 15 µg total RNA were reverse-transcribed by using special oligo d(T) primers with 5' unique sequence overhangs for the labelling reactions. Microarrays were prepared for hybridisation by washing twice for 5 minutes in 0.1% SDS, washing five times for 1 minute in MilliQ H_2O, immersing 3 minutes in 95°C MilliQ H_2O, and drying by centrifugation (5 minutes at

800 RCF in 50 ml conical tubes). The cDNA was hybridized to the microarrays in a formamide-based buffer (25% formamide, 4X SSC, 0.5% SDS, 2X Denhardt's solution) with competitor DNA (LNA dT blocker, (Genisphere), human COT-1 DNA (Sigma)) for 16 h at 51°C in a humidified hybridisation oven. The arrays were washed once for 5 min at 45°C (2X SSC, 0.1% SDS), twice for 3 min in 2X SSC, 0.1% SDS at room temperature (RT), twice for 3 min in 1X SSC at RT, two times for 3 min in 0.1X SSC at RT, and dried by centrifugation. The Cy3 and Alexa 647 fluorescent dye attached to DNA dendrimer probes (3DNA capture reagent, Genisphere) were then hybridized to the bound cDNA on the microarray using the same hybridisation solution as earlier; the 3DNA capture reagents bound to their complementary cDNA capture sequences on the oligo d(T) primers. This second hybridisation was done for 2 h at 51°C in a humidified hybridisation oven. The arrays were then washed and dried as before.

Transcription profiles of fish from the farmed strains and their wild counterparts were contrasted on 23 microarrays (Canada: 13 involving 26 fry, Norway: 10 involving 20 fry). Since our main interest was in detecting changes driven by artificial selection, one farmed and one wild juvenile of the paired populations were compared directly on each microarray. This simple design maximises the statistical power for detection of such differences for a given number of microarrays. Dye-sample coupling was balanced between fry of wild and domestic origins in both systems (dye flip on the biological replicates). The cDNA microarrays used here were obtained through the Genomic Research on Atlantic Salmon Project (GRASP, available from Ben F. Koop, bkoop@uvic.ca) and comprises 3557 clones from 18 high complexity salmonid cDNA libraries, each printed as double, side-by-side spots on Telechem Superamine slides (6440 Atlantic salmon and 916 rainbow trout (*Oncorhynchus mykiss*) elements or spots; Rise *et al.* 2004). Detailed information on the clones printed on the microarray can be found at http://web.uvic.ca/cbr/grasp/. Rise *et al.* (2004) provided information on the relative representation of diverse molecular functions in several cDNA libraries. However, they did not provide such information for the sequences printed on the cDNA microarray. Using the web-based tool FatiGO (Al Shahrour *et al.* 2004) as well as information available on the GRASP website, we calculated the proportions of genes on the array involved in several biological processes (Supplementary

figure 2.1). Though calculated with all the information available, the results might not be quite accurate since gene ontology (GO) information regarding biological process at level 4 was only available for 18% (631 genes) of the sequences on the array.

2.4.3 Signal detection, data preparation and statistical analysis

Signal was detected using a ScanArray scanner from Packard BioScience. Spots were located and quantified with the QuantArray software, using the adaptive circle spot quantification method and keeping the mean value of intensity for each spot. Local background and the data from bad spots were removed. Missing data were then imputed using the K-nearest neighbours imputer in SAM (Tusher *et al*. 2001) (15 neighbours). Signals from pairs of neighbouring replicate spots (PNRS) were averaged and data were normalised by dividing by the channel mean. To assess differences between wild and farmed salmon, data from Canada and Norway were analysed separately using a mixed model of ANOVA (Wolfinger *et al*. 2001) and the MAANOVA R package (Kerr *et al*. 2002, Kerr *et al*. 2000). Genes with low intensity data (mean intensity smaller than the mean intensity of the empty spot controls plus twice its standard deviation in both channels) were removed from the analysis, leaving 3058 and 2552 detected clones for the Norwegian and Canadian experiments, respectively. Under the assumption that all null hypotheses are true, 31 and 26 false positives at the $P < 0.01$ significance threshold are expected by chance alone. Prepared data were corrected for intensity-linked distortion using a regional LOWESS algorithm and fitted to a mixed model of variance using the R/MAANOVA package (http://www.jax.org/staff/churchill/labsite/software/Rmaanova/index.html). The model included the "array" term as a random term and the "sample type" (farmed or wild) and "dye" terms as fixed terms. A permutation based F test (F2, with 1000 sample ID permutations) was then performed, and restricted maximum likelihood was used to solve the mixed model equations. Variance homoscedasticity was tested prior to each ANOVA using the F-test function in the Excel spreadsheet software.

While our main interest was in detecting changes in farmed versus wild salmon, we were also interested in using our data to assess for differences in transcription levels between

salmon from Norway (Namsen river) and Canada (St-John river). To do this, normalised intensity data obtained previously on separate arrays for wild individuals of both origins were analysed together using an AVOVA model with “dye” and “country” as fixed terms. The “array” term had to be removed from this model because experimental variance associated with the array term would have been confounded with that linked to the “country” term. This simpler model fits the data without notable bias, as verified by plotting the residuals versus the fitted value for each gene in R/MAANOVA (not shown). Since the array-linked experimental variance is considered as random (Draghici 2003), removing the “array” term from our model is not expected to bias the conclusions of the analysis. Yet, it is not ideal since the array-linked experimental variance will further inflate random error, making the test less powerful. Hence, while our experimental design was well suited for our main question of interest, this led to use a less powerful test to assess for differences between fish from Norway (river Namsen) and Canada (St-John river).

2.5 Results

At a significance threshold for permutation corrected P-values of $P < 0.01$, 68 and 74 pairs of neighbouring replicate spots (PNRS) on the array represented genes with significantly different transcription levels in the progeny of wild and farmed populations from Canada and Norway, respectively (Supplementary tables 2.1 and 2.2). After correcting for the expected number of false positives, this represented 1.4% and 1.7% of all detected PNRS for Norwegian and Canadian juveniles, respectively. These values represent conservative estimates of the proportion of genes for which the transcription level was modified by five to seven generations of domestication, which may include the effects of artificial selection, founder effects, genetic drift and inadvertent selection due to artificial rearing or to the selection of traits correlated to the traits of interest.

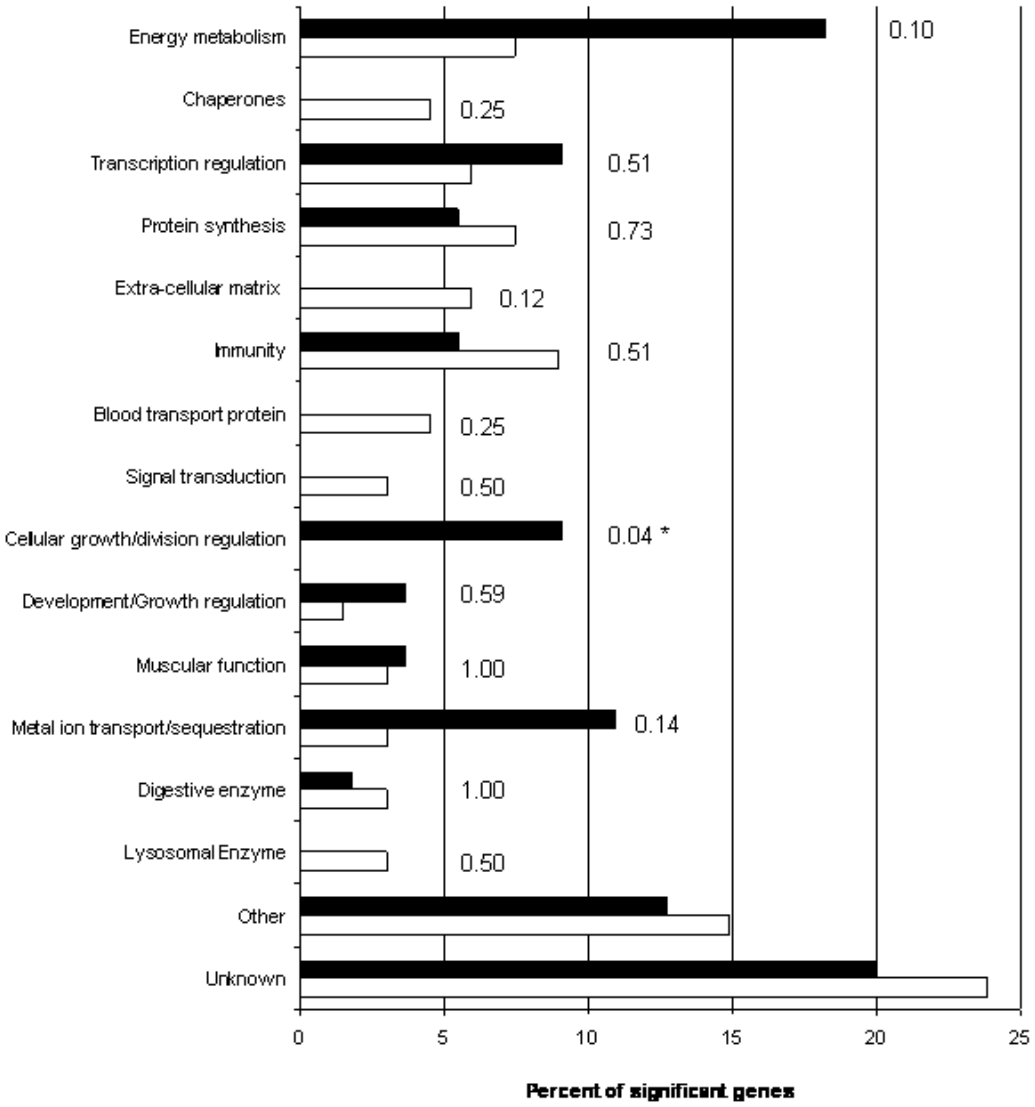


Figure 2.1 Distribution of the differentially expressed genes in various functional classes.

Percent of all genes showing significant expression differences ($P < 0.01$, permutation corrected P-values) between farmed and wild salmon originating from the St-John River (black) and from the River Namsen (white) in each of 16 functional classes defined from the literature. P-values from Fisher's exact bilateral test for equality of proportions (Fisher and van Belle 1993; Draghici 2003) were calculated in the R environment (R Development Core Team, 2002) to assess for the significance of the differences in proportion between the Canadian and Norwegian systems observed among the diverse categories and added to the figure.

Potential caveats of this study include the difference in genetic diversity between farmed and wild salmon as well as the use of imputed data. First, a reduced genetic diversity is expected in farmed versus wild salmon and was documented for Norwegian fish (e.g. Norris *et al*. 1999). This difference could be a problem when using ANOVA. Thus, F-tests for variance homoscedasticity in gene transcription profiles between farmed and wild salmon were performed for each detected gene on the prepared data before to perform the ANOVA. In both the Canadian and Norwegian systems, less significant genes than expected by chance alone were observed (at $P < 0.05$, 13 genes showed significant differences over 2552 tests in Canada and 20 genes over 3058 tests in Norway). These were considered as false positives. Only one of them (clone SK1-0197, similar to Beta actin 1) was present in the lists of genes showing significant differences between farmed and wild salmon (Norwegian system). Hence, reduced genetic diversity in farmed populations does not warn against the use of ANOVA in this study. Second, most microarray data analysis methods require that all the data be present. Missing data must therefore be imputed (here by using the K-nearest neighbours imputer in SAM (Tusher *et al*. 2001)), which could influence the results. In this study, only a very small subset of the data were missing and had to be imputed (0.08% for the Canadian system and 0.12% for the Norwegian one). Only two of the genes for which a significant difference in transcription profile was observed between farmed and wild salmon had received an imputed value; both of these were in the Canadian system (nwh^18^87^nucleoside diphosphate kinase and pitl^504^374^ ZipA). In these two cases, one value over 26 (13 arrays, 2 side-by-side replicates averaged) was imputed. As for nucleotide diphosphate kinase, two other non-imputed clones showed equivalent results. Thus, data imputation did not have much influence on our results.

2.5.1 Functional classification of the differentially expressed genes

Changes in transcription profiles were observed in different functional classes (Fig. 2.1). The proportion of significant changes for genes implicated in protein synthesis and muscle function was similar in Norway and Canada. In contrast, the proportion of changes for genes implicated in energy metabolism and metal ion sequestration was greater in the

Canadian than in the Norwegian salmon, whereas the opposite pattern was observed for genes coding for molecular chaperones, blood transport proteins and cellular matrix proteins. A Fisher exact test for equality of proportions (Fisher and van Belle 1993, Draghici 2003) was applied to these data. Only the over-representation of genes implicated in regulation of cellular division and growth in the Canadian versus Norwegian salmon significant genes was significant at P>0.05 (Fig. 2.1). Tools allowing more reliable functional analysis of microarray data, such as David/Ease (Dennis *et al*. 2003, Hosack *et al*. 2003) or Onto-Express (Khatri *et al*. 2002) are presently poorly adapted to the analysis of data from non-model species custom cDNA microarrays for several reasons, the first one being the difficulty to get meaningful functional annotations for all the clones on such arrays. Hence, we were only able to get GO information for a subset (<18%) of the sequences on the array, using BLAST-gathered information available on the GRASP website. Protein accession numbers were converted to Unigen clusters numbers using David web-based conversion tool and these were imputed in Ease to assess for differences in GO representation between the array and each of the two sets of significant genes (using Fisher's exact test). The results obtained partly reflect those presented in figure 2.1. When compared to the array, the "extracellular", "extracellular space", "extracellular matrix" and "protein binding" categories were all significantly over-represented in the Norwegian significant genes list (P-value: 0.007, 0.018, 0.023, 0.045, respectively), while the "regulation of cell proliferation", "regulation of cellular process" and "regulation of biological process" categories were significantly over-represented in the Canadian significant genes list (P-value: 0.033, 0.039, 0.039, respectively). For the Canadian significant genes list, it is interesting to note that the "transition metal ion binding" and the "iron binding" categories also had low P-values (0.051, 0.091, respectively). However, none of these changes would still be significant following a Bonferroni correction and the partial coverage of the array content requires conservative interpretation of the results.

2.5.2 Genes with multiple significant PNRS

The specific design of the microarrays we used generated additional information (Table 2.1), which further supported the results obtained for several of the significant genes

Table 2.1 Genes for which multiple non-neighbour PNRS showed significant transcription level differences between the progeny of wild and domestic salmon.

Origin	Genes with multiple significant PNRS	Average fold Change	P-value
Canada	Cytochrome C	0.8	3.3×10^{-5}
		0.8	1.9×10^{-4}
		0.9	6.7×10^{-3}
	Metallothionein A	0.7	1.2×10^{-6}
		0.8	1.1×10^{-4}
		0.8	1.1×10^{-4}
		0.8	1.4×10^{-4}
		0.7	2.8×10^{-6}
		0.9	1.8×10^{-3}
	GAPDH	0.9	4.4×10^{-4}
		0.8	1.2×10^{-5}
		0.9	2.8×10^{-3}
		0.9	8.4×10^{-4}
	Nucleoside diphosphate kinase	0.9	8.3×10^{-4}
		0.9	1.2×10^{-3}
		0.9	6.7×10^{-3}
	HES1	1.2	9.2×10^{-3}
		1.2	5.5×10^{-3}
	Brain protein (44-like)	0.9	2.1×10^{-3}
		0.9	4.0×10^{-3}
Norway	NADH dehydrogenase subunit 5	1.3	3.6×10^{-4}
		1.3	4.1×10^{-3}
	Calmodulin	1.3	5.0×10^{-4}
		1.3	6.7×10^{-3}
		1.2	6.0×10^{-3}
		1.2	9.0×10^{-3}
	Clone SS1-0555	0.8	2.3×10^{-3}
		0.9	9.9×10^{-3}
	Protein synthesis initiation factor 4A	1.2	2.4×10^{-3}
		1.2	6.7×10^{-3}
	Brain acidic ribosomal phosphoprotein P0	1.2	3.2×10^{-3}
		1.2	3.5×10^{-3}
	α-globin	1.3	4.1×10^{-3}
		1.2	5.4×10^{-3}

For each PNRS (pairs of neighbouring replicate spots) of genes with multiple significant PNRS at the $P < 0.01$ threshold, the average fold change in farmed relative to wild individuals from 13 (Norwegian) or 10 (Canadian) arrays and the permutation corrected (1000 permutations) P-value from the F-test on the mixed model of ANOVA is given.

(Supplementary tables 2.1 and 2.2). Namely, two types of replicates were present according to their relative positions on the array: neighbouring replicates for each spot on the chip (PNRS; which were averaged) and additional non-neighbour replicates, often representing different cDNA clones of the same mRNA, for a limited proportion of all the cDNAs represented on the array (which were analysed separately without averaging). Hence, within the lists of significant genes, 12 genes were represented more than once. For instance, the metallothionein A-coding gene was represented by six separate pairs of spots, which all showed significant differences in transcription level in the same direction at $P < 0.01$. Such results substantially reduce the probability that these genes represent false positives due to technical artefacts.

2.5.3 Parallel changes

Many changes in gene transcription levels observed here were strain-specific. Yet, our results also illustrate at least four distinct ways by which directional selection could act on gene expression to result in parallel phenotypic evolution. First, equivalent changes could occur in the same genes in populations submitted to similar selective regimes. Hence, seven genes showed significant expression differences ($P < 0.01$) between farmed and wild salmon in both Norway and Canada, and all changed in a parallel direction (Fig. 2.2). The probability that 7 genes show parallel changes by chance alone is $P = 7.8 \times 10^{-3}$, as evaluated using the binomial law with $\pi = 0.5$ (which is equivalent to the probability of at least seven successes over seven iterations in a sign test with $\pi = 0.5$). Given that some of the parallel genes were down-regulated and others up-regulated in the farmed strains, that they belong to very different functional groups and that no relation was found between them either through clustering analysis or with the PathwayAssist software (Ariadne Genomics), we consider that the assumption of independence between the genes required for this test is reasonable. These seven genes represent 16 % of the expected number of true positives in each population. It is noteworthy that three of these genes had unknown functions. A second level of parallelism was exemplified by the parallel changes observed in the expression of ATP synthase coding genes (Fig. 2.2). Here, parallel change did not affect the same gene but genes coding for different subunits of the same functional entity,

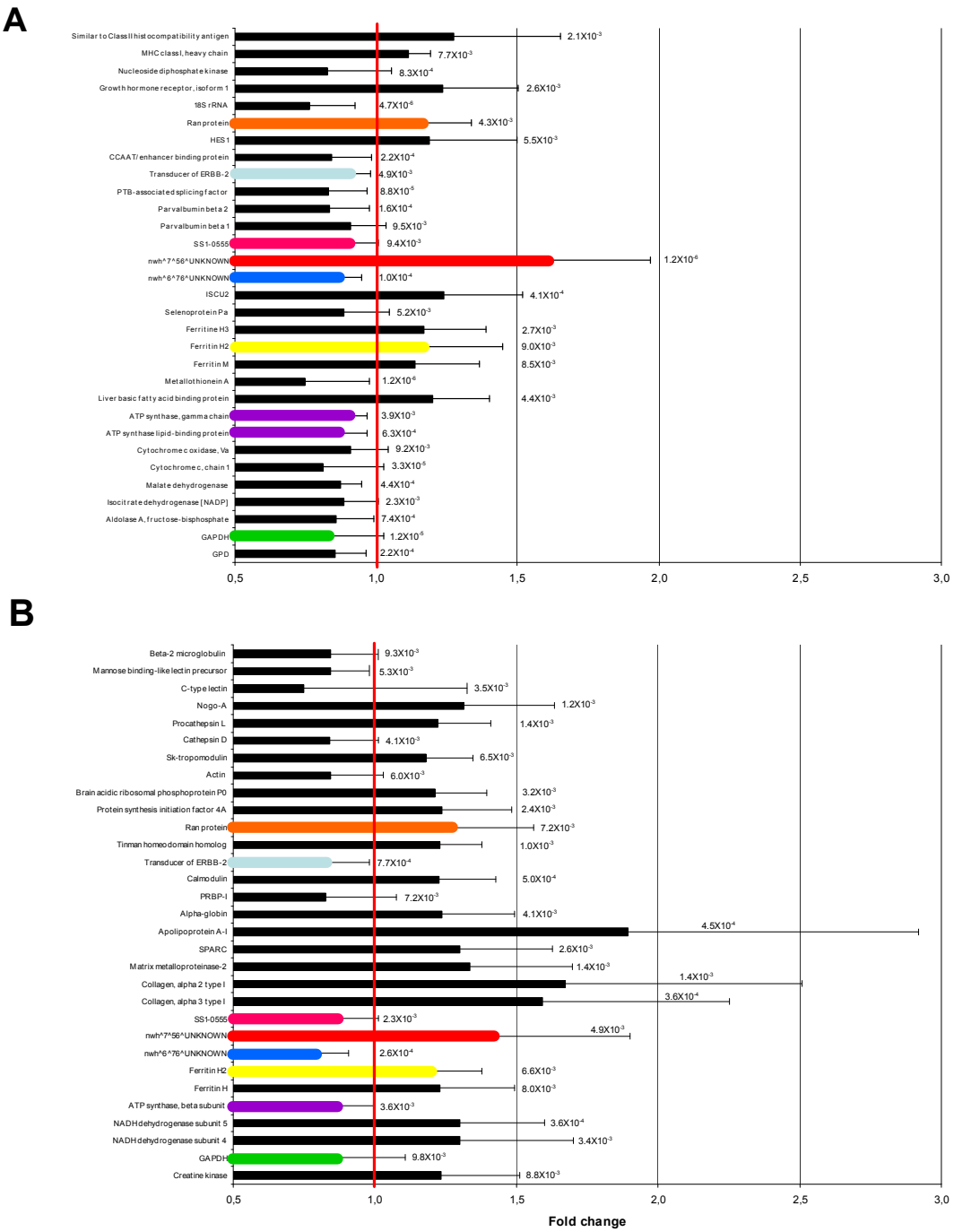


Figure 2.2 Average fold change and standard deviation for a selection of genes differentially expressed between the progeny of farmed and wild salmon originating from (A) the St-John River, Canada and (B) the River Namsen, Norway.

For genes with multiple PNRS, only the smallest P-value observed is given. A value above one (bold line) for a given gene represents over-expression in farmed relative to wild salmon. Colours indicate parallel changes between Canadian and Norwegian salmon, either in genes coding for different subunits of the same functional entity (ATP synthase) or in the same gene (the seven other instances). The genes in this figure were chosen either for their high level of significance, their particular contextual interest or because they showed parallel changes in expression in both systems.

ATP synthase. Third, equivalent changes in different protein-coding genes within a same pathway were observed. Thus, Canadian farmed salmon showed 23% over-expression of the growth hormone (GH) receptor gene, which corroborates the increased levels of GH reported in Norwegian farmed salmon by Fleming *et al.* (2002). However, this study did not confirm their observation since the transcription level of the GH coding genes was not significantly different between Norwegian farmed and wild salmon. Many explanations could account for this: lack of statistical power, different life-stages used (Fleming *et al.* : 1 and 2+ parr, here: alevins before first feeding) and more importantly, the fact that plasmatic levels of GH are regulated post-transcriptionally by several mechanisms (see, for instance, Muller *et al*. 1999). Fourth, equivalent changes in gene transcription profiles could also occur in different pathways affecting a same biological function. Hence, many genes showing significant changes were in similar functional classes but belonged to different molecular pathways, including energy metabolism, transcription regulation, protein synthesis, immunity, muscle function and digestion, in both farmed strains, suggesting the importance of these functions in the response to the similar selection regimes (Fig. 2.1)**.**

2.5.4 Differences between Norwegian and Canadian wild salmon populations

To assess for a potential source of the strain-specific differences in transcription profiles caused by domestication, we investigated transcription profile differences between the wild fry from river Namsen (Norway) and those from St-John river (Canada). Of 3147 detected PNRS (mean intensity over that of the empty spot controls plus twice its standard deviation) in the grouped normalised data from 12 wild Canadian and 10 wild Norwegian fry, 944 (30%) exhibited significantly different variances in the detected signals between countries at $P < 0.05$. For the remaining 2203 genes, data were fitted to an ANOVA model and a permutation based F-test was performed. While 22 false positives are expected by chance at the $P < 0.01$ threshold, 171 PNRS exhibited significant differences between North American and European salmon (Supplementary table 2.3). After subtracting the expected number of false positives, this represented 6.8% of the detected genes. The average fold change in transcription profiles within the set of significant genes for $P < 0.01$

Table 2.2 Relation between country-specific changes caused by domestication and between-country differences in transcription profiles.

Gene	Between countries	Norway	Canada	
HMG-1	2.29	1.27		*
Malate dehydrogenase 2, NAD (mitochondrial)	1.85		0.87	*
Parvalbumin beta 1	2.57		0.91	*
Ferritin, middle subunit	3.96		1.14	
Salmo salar cDNA clone SK1-0455	1.69		1.16	
Mannose binding-like lectin	0.57	0.84		*
Salmo salar cDNA clone SK1-0473 similar to Beta actin 1	2.15	1.64		*
Salmo salar cDNA clone SS1-0134	1.73	0.83		
Salmo salar cDNA clone SS1-0466 similar to Ependymin	0.68		1.23	*
Salmo salar cDNA clone lna^9^95^UNKNOWN	0.66	0.85		*
Nucleoside diphosphate kinase	1.67		0.86-0.90	*
Parvalbumin beta 1	1.89		0.91	*
NADH dehydrogenase subunit 4	1.41	1.30		*
Actin	2.23	0.84		
Fatty acid binding protein	1.60		1.16	
Glyceraldehyde 3-phosphate dehydrogenase	1.41	0.83	0.86	
Collagen alpha 3(I)	1.69	1.59		*
LYSOZYME G	0.48	0.85		*
Salmo salar cDNA clone SNN-1095	0.55	0.81		*
APOLIPOPROTEIN A-I-1 PRECURSOR	1.45	1.89		*
Creatine kinase (EC 2.7.3.2)	1.93	1.23		*
Parvalbumin beta 2	1.52		0.83	*
Ferritin middle subunit	1.32		1.14	
P0 acidic ribosomal phophoprotein	1.71	1.21-1.22		*
Collagen alpha 2(I) chain precursor	1.93	1.67		*
Secreted protein, acidic, rich in cysteine (SPARC)	1.53	1.30		*
Beta-2 microglobulin	1.23	0.84		
Salmo salar cDNA clone lna^9^26^UNKNOWN	0.69		0.89	
Ferritin H-2	1.42	1.20	1.16	
High choriolytic enzyme 1 precursor	1.49		1.20	
Selenoprotein precursor	1.28		0.89	*
Lysozyme	0.56	0.85		*

Genes for which both a country-specific domestication changes and between-country differences in gene transcription profiles were observed at the $P < 0.01$ significance threshold are presented with their average fold change between and within countries. A value superior to 1 in the "between countries" column represents an over-transcription in the Canadian compared to Norwegian salmon. A value above 1 in the country-specific column represents an over-transcription in the domestic population compared to its wild population of origin. Asterisks indicate cases where the between-country differences seem partly counter-balanced by domestication-linked changes.

was 63 %. Several of the continent-specific changes induced by domestication involved the same genes for which transcription differed between wild salmon from Norway and Canada. Hence, of the 39 and 41 genes (expected number of true positives) for which transcription profiles were modified differentially by domestication in Norway and Canada respectively (Supplementary tables 2.1 and 2.2), 17 (43%) and 13 (32%) also differed significantly between continents (Table 2.2). Of these, genes transcribed at higher levels in wild Norwegian salmon were generally either down-regulated in farmed Norwegian salmon (e.g. mannose binding lectin, lysozyme G) or up-regulated in farmed Canadian salmon (e.g. ependymin). Reciprocally, genes transcribed at higher levels in wild Canadian salmon were generally either up-regulated in farmed Norwegian salmon (e.g. Collagen, apolipoprotein A-I, NADH dehydrogenase) or down-regulated in Canadian farmed salmon (e.g. nucleoside diphosphate kinase, parvalbumin beta 1 and 2, malate dehydrogenase; Table 2.2).

2.6 Discussion

Elucidating the nature of changes in gene expression levels in farmed salmon may contribute to better assess the possible consequences of genetic interactions between wild and escaped farmed salmon. It is also of direct relevance to the understanding of contemporary evolution driven by human alterations. Here, we demonstrated that only five to seven generations of domestication led to significant changes in gene expression, the average magnitude of the observed differences being approximately 20% for at least 1.4 and 1.7 % of the expressed genes at the juvenile stage. As evidenced by the parallel changes observed, some of these changes were generated by directional artificial selection for traits of commercial interest. In addition, other factors including founder effects, genetic drift and inadvertent selection due to artificial rearing or to the selection of traits correlated to the traits of interest may have contributed to the observed differences.

Changes in the transcription levels of genes involved in energy metabolism are especially interesting since they may influence the most important traits targeted by artificial selection: growth rate and age at sexual maturity. With the exception of the gene coding for

NADH dehydrogenase and creatine kinase in Norwegian salmon, genes coding for enzymes involved in energy metabolism, including glycolysis, citric acid cycle and oxidative phosphorylation were all under-transcribed in farmed salmon of both strains (Fig. 2.2). Decreased expression of these genes could lower the functional capacity in these pathways or may reflect reduced turnover of their components. The increased expression of creatine kinase is not in contradiction with a diminution in metabolic capacities since this enzyme acts downstream of ATP production and catalyses both ATP and creatine formation from phosphocreatine and ADP and the reverse reaction (Voet and Voet 2004). Under the assumption that over-transcription of the creatine kinase gene translates into increased enzymatic activity, this could represent a more efficient ATP buffering by phosphocreatine in the Norwegian farmed salmon relative to wild ones. Alternatively, heritable increase in the relative mass of skeletal muscle, with its moderate aerobic capacity and high levels of creatine kinase, could underlie the patterns observed in farmed salmon. In accordance with the increased fat levels in the flesh of farmed salmon (Rye and Gjerde 1996), the gene for Apolipoprotein A-I, a constituent of both chylomicrons and HDL, was strongly over-transcribed in the farmed progeny from Norway, and the transcription level of the gene for fatty acid binding protein was increased in Canadian farmed salmon. Altogether, these patterns are consistent with a reduction in basal metabolic rate and an increased metabolic efficiency of farmed juvenile salmon favouring allocation of resources towards growth and fat deposition. Knowing that resistance to infection has been selected both in Norway and Canada farmed salmon, it is also noteworthy that two genes coding for MHC antigens were over-transcribed in farmed salmon from Canada (Fig. 2.2).

We found evidence for parallelism in transcriptome evolution at several distinct levels in this study. Thus, up to 16 % of all transcription profile changes between farmed populations and their wild populations of origin occurred for the same genes and in parallel for both farmed populations. In these cases, there is therefore strong evidence for the role of selection in driving phenotypic changes. Assuming that a large number of genes contribute to the expression of the natural variation in polygenic traits under selection (e.g. growth rate and age at sexual maturity), parallel phenotypic evolution should rarely involve the same genes (Schluter 2004). Three major non-mutually exclusive explanations were

proposed to account for changes in the same genes in two independent lineages having evolved in parallel: (1) the similarly biased production of genetic variation in close relatives (Schluter *et al*. 2004, Haldane 1932), (2) a limited number of genes being actually important in phenotypic variation and evolution for a given trait (Remington and Purugganan 2003, Stern 2000) or (3) developmental constraints shared between closely related groups (e.g. West-Eberhard 2003). It is noteworthy that three of the parallel genes we observed have unknown functions: their parallel expression changes in both breeding strains suggest they might affect growth rate or age at sexual maturity, giving a first glimpse of their potential functions. Yet, we also observed parallel changes in the expression of genes coding for different subunits of the ATP synthase (Fig. 2.2), equivalent changes for different genes within pathways and the representation of common functional classes in the significant changes. These levels of parallelism not necessarily involving the same genes, as well as the fact that most observed changes are strain-specific, suggest that selection also acted on genes for which genetic variation was not biased in a similar way in Norwegian and Canadian salmon. Accordingly, significant differences in both gene expression variance and mean between the wild Norwegian and Canadian salmon were observed, which could modify their responses to similar selection regimes. Also, 43% (Canada) and 32% (Norway) of the genes for which transcription profiles differed between farmed and wild fish also differed significantly between the wild populations of river Namsen (Norway) and St-John river (Canada). Of these, genes transcribed at higher levels in wild Norwegian salmon were generally either down-regulated in farmed Norwegian salmon (e.g. mannose binding lectin, lysozyme G) or up-regulated in farmed Canadian salmon (e.g. ependymin). Reciprocally, genes transcribed at higher levels in wild Canadian salmon were generally either up-regulated in farmed Norwegian salmon (e.g. Collagen, apolipoprotein A-I, NADH dehydrogenase) or down-regulated in Canadian farmed salmon (e.g. nucleoside diphosphate kinase, parvalbumin beta 1 and 2, malate dehydrogenase; Table 2.2). This illustrates how pre-existing differences in gene expression profiles can modify the way by which two related populations respond to selection. It also suggests that genetic or developmental constraints can limit the contribution of a given gene to the evolution of a phenotype under directional selection.

Substantial genetic (based on neutral markers) and heritable phenotypic differences were observed between Atlantic salmon from Europe and North America (Ståhl 1987, Phillips and Hartley 1988, Verspoor and McCarthy 1997). Accordingly, we observed differences both in gene expression variance (30% of the detected genes had non-homogenous variances) and means (average change of 63%, for 6.8% of the detected genes with homoscedastic variances) between the progeny of wild individuals from Norway and Canada. These differences could be causally linked with some of the strain-specific differences in the domestic strains (Table 2.2). Strain-specific changes between wild and domestic salmon may also be associated with differences in the selection regimes applied in Norway and Canada.

In addition, this study provided a first identification of many structural genes for which average transcription levels differ significantly between European and North American salmon (Supplementary table 2.3). Among the most interesting of these are fish allergens (type I collagen and parvalbumines), key genes of energy metabolism (malate dehydrogenase, NADH dehydrogenase, creatine kinase, GAPDH) and several ribosomal protein coding genes (Supplementary table 2.3). However, these results must be interpreted cautiously since they were obtained from only two wild populations. Clearly, studies involving more populations from both continents will be needed to confirm whether the observed differences truly reflect general inter-continental differences of expression at the early fry stage.

Finally, our results also exemplify how founder effects, genetic drift or inadvertent selection may have caused non-desirable evolutionary changes in farmed salmon (McGinnity *et al*. 2003). Hence, previous studies revealed a positive correlation between growth rate and fat content in Atlantic salmon (Quinton *et al*. 2004) and showed that farmed salmon have a higher fat content and eat more frequently (Thodesen *et al*. 1999). Our results provided strong evidence for a 21% under-expression (average of six PNRS) of the metallothionein (MT) A gene in Canadian farmed salmon. Interestingly, it has been shown that MT-null mouse are obese, as a result of a higher food intake and of an abnormal energy balance (Beattie *et al*. 1998), which points towards a role for MT in the regulation

of energy balance. MT is also a key factor in adaptation to heavy-metal environments (Posthuma and Vanstraalen 1993). Therefore, introgression of this reduced MT expression from farmed into wild populations could reduce their resistance to environmental pollutants**.** Admittedly, the phenotypic importance of the transcription level changes observed in this study must be interpreted cautiously since the link between transcription level and phenotypic expression is subjected to other levels of regulation. Studying the actual impact of the transcription profile changes observed on phenotype variation and fitness will therefore represent a necessary and exciting challenge in future studies.

2.7 Acknowlegments

This work was supported by a research grant from Science and Engineering Research Canada (NSERC strategic programme) to L.B., H.G. and J.J. Dodson, a postgraduate scholarship from the *Fonds Québécois de la Recherche sur la Nature et les Technologies* (FQRNT) to C.R., and funding from the Research Council of Norway to S.E.. This study is a contribution to the research programmes of CIRSA (*Centre Inter-universitaire de Recherche sur le Saumon Atlantique*) and *Québec-Océan*. We thank Trevor Goff and the staff from the Mactaquac Fish Hatchery, Department of Fisheries and Oceans (DFO) Canada and the staff at the NINA Research Station, Ims, Norway for their assistance in rearing fish. We are also grateful to Brian Glebe (DFO) for providing fish from the Canadian farmed strain, to Sean Rogers, Jon Slate and three anonymous referees for their valuable input on an earlier version of the manuscript, as well as to Hélène Crépeau and Gary Churchill for their comments on statistical issues.

Chapitre 3 : Genetic consequences of interbreeding between farmed and wild Atlantic salmon : insights from the transcriptome

3.1 Résumé

Le nombre important de saumons atlantiques d'élevage s'échappant chaque année augmente les risques d'extinction des saumons de populations naturelles par leurs interactions génétiques et écologiques avec ceux-ci. Récemment, nous avons présenté des résultats identifiant des changements évolutifs de la transcription génique entre saumons d'élevage et sauvages après seulement 5 et 7 générations de sélection artificielle. Si l'on s'attend à ce que les différences engendrées par l'élevage pour la plupart des caractères phénotypiques quantitatifs s'étiolent suite à des rétrocroisements répétés aux saumons de populations naturelles, il semble que les bases génétiques de la transcription génique soient largement non-additives et les croisements hybrides montrent souvent des patrons de transmission phénotypique inattendus. Il est ainsi difficile d'évaluer à quel point la reproduction entre saumons d'élevage et sauvages est susceptible de modifier la composition génétique des populations sauvages. Dans la présente étude, nous avons comparé le transcriptome de la progéniture de saumons sauvages norvégiens à celui de la progéniture d'un rétrocroisement de saumons hybrides (domestiques X sauvages) à des saumons sauvages. Plus de 6% (298, q-value < 0.01) des gènes exprimés montrent des niveaux de transcription significativement différents et l'étendu et la magnitude moyenne de ces différences sont plus grandes que pour les différences décrites précédemment entre des souches d'élevage et sauvages pures. La plupart des différences observées semblent résulter d'interactions génétiques non-additives. Ces résultats suggèrent que la reproduction de saumons d'élevage fugitifs avec des saumons sauvages en nature pourrait modifier de façon substantielle le contrôle génétique de la transcription génique dans des populations naturelles exposées à une forte migration provenant d'installations d'élevage, menaçant potentiellement la survie de ces populations. Ceci supporte encore davantage la position selon laquelle des mesures doivent être prises rapidement pour réduire considérablement le nombre de saumons d'élevage qui s'échappent ainsi que leur reproduction en nature.

3.2 Abstract

Large annual escapes of farmed Atlantic salmon enhance the risk of extinction of wild populations through genetic and ecological interactions. Recently, we documented evolutionary change in gene transcription between farmed and wild Atlantic salmon after only five generations of artificial selection. While differences for most quantitative traits are expected to gradually dilute through repeated backcrossing to wild populations, the genetic basis of gene transcription has been shown to be largely non-additive and hybrid crosses may display unexpected inheritance patterns. This makes it difficult to predict to what extent interbreeding between farmed and wild individuals will change the genetic makeup of wild salmon populations. Here, we compared the genome-wide gene transcription profiles of Norwegian wild salmon to that of a second generation hybrid cross (backcross: (Farmed X Wild) X Wild). Over 6% (298, q-value < 0.01) of the detected genes exhibited highly significantly different transcription levels and the range and average magnitude of those differences was strikingly higher than previously described between pure farmed and wild strains. Most differences appear to result from non-additive gene interactions. These results suggest that interbreeding of fugitive farmed salmon and wild individuals could substantially modify the genetic control of gene transcription in natural populations exposed to high migration from fish farms, resulting in potentially detrimental effects on the survival of these populations. This further supports the idea that measures to considerably reduce the number of escaped farmed salmon and their reproduction in the wild are urgently needed.

3.3 Introduction

Since the 1970's, Atlantic salmon successfully responded to intense artificial selection aimed at improving growth rates as well as other traits of commercial interest (Gjoen and Bentsen 1997). Genetically-based phenotypic changes not specifically selected for have also resulted from such selection programs, including increased fat content in flesh (Rye and Gjerde 1996), poorer performance in the wild (McGinnity *et al*. 2003), physiological, morphological and behavioural changes (Fleming *et al*. 2002, Fleming and Einum 1997).

Farmed salmon are considered an important problem for natural populations when they escape from sea-cages and enter streams and rivers in significant proportions. About two million escape annually, whereas the total numbers of wild adult salmon are estimated at about four million (McGinnity *et al*. 2003). These fugitives are thought to greatly enhance the risk of extinction of wild populations through genetic and ecological interactions (McGinnity *et al*. 2003), as well as by spreading diseases (Naylor *et al*. 2005). Monitoring in a range of Norwegian wild populations indicated that escaped farmed fish constitute on average close to 20% of the adult population during the breeding season (Fiske *et al.* 2001). Experimental evidence and observations in nature also suggest that most successful breeding of farmed salmon occurs through hybridization with wild fish (Fleming *et al.* 2000). A pressing question is therefore to what extent interbreeding between farmed and wild individuals will change the genetic make-up of wild Atlantic salmon populations.

It is increasingly acknowledged that evolutionary changes may strongly depend on alterations of gene transcription regulation (Wilson 1976). Using the microarray technology, which offers the possibility of a genome-wide scan for gene transcription differences, we recently compared patterns of gene expression in the progeny of farmed and wild Atlantic salmon from Norway and Canada grown in controlled conditions. Our results demonstrated that five to seven generations of artificial selection led to heritable changes in gene transcription profiles, the average magnitude of the differences being 25% and 18% for at least 1.4 % and 1.7% of the expressed genes in juvenile salmon from Norway and Canada, respectively (Roberge *et al*. 2006). Moreover, genes showing significant

transcription profile differences in farmed strains from both Canada and Norway (16% of the differentially expressed genes) all exhibited parallel changes, providing evidence for the role of artificial selection in driving evolutionary changes of gene transcription levels between wild and farmed salmon.

Genetic changes for most quantitative phenotypic traits in farmed salmon can be expected to gradually dilute through repeated backcrossing to wild populations (if non-native immigration ceases) due to their additive genetic basis (Anderson 1949, Tymchuk *et al*. 2006, McGinnity *et al*. 2003). However, the genetic basis of gene transcription appears to be largely non-additive (Gibson *et al*. 2004, Hedgecock *et al*. 2007; but see Swanson-Wagner *et al*. 2006, Cui *et al*. 2006) and hybrid crosses sometimes display non-additive inheritance patterns for quantitative traits, such as outbreeding depression (Gharrett *et al*. 1999) and heterosis (Lippman and Zamir 2007). Moreover, in the absence of strict policies to considerably reduce the number of escaped farmed salmon, this number is expected to increase due to the rapid development of the aquaculture industry, further challenging the genetic integrity of natural populations. Sea farming is also rapidly developing for other marine fish species (*e.g.* Atlantic cod and halibut; see Naylor *et al.* 2005), such that similar risks could threaten their natural populations as well.

Here we compared, using a 16 000-gene microarray, the genome-wide gene transcription profile of the progeny (fry stage) of Norwegian wild salmon to that of a second generation hybrid cross (backcross: (Farmed X Wild) X Wild) to test whether genetically-based differences in gene transcription persisted in these hybrids. We also compared our results to those of a previous study on genome-wide transcription divergence between pure farmed and wild salmon (Roberge *et al*. 2006) in order to gain insights into the relative importance of additive *versus* non-additive genetic interactions in controlling gene transcription in salmon. In this comparison, we addressed four main questions: **(1)** How do the number of significant gene transcription differences observed and **(2)** the average magnitude and range of these differences compare between the two studies? In other words, are there fewer and smaller differences between backcross and wild salmon than between farmed and wild salmon? **(3)** How does the gene transcription difference between backcross and wild

salmon compare to that observed between wild and farmed salmon for genes that had shown significant transcription level differences between wild and farmed salmon and **(4)** for genes that showed significant transcription level differences between hybrid backcross and wild salmon? Finally, we present quantitative PCR confirmation of some of our earlier results on gene transcription differences observed in Canadian farmed *versus* wild salmon. By documenting genome-wide effects of interactions between farmed and wild salmon genomes on patterns of gene expression, our main goal was to contribute to a better understanding of the evolutionary impact of farm escapees on wild salmon populations.

3.4 Methods

3.4.1 Fish crosses and samples

Twenty wild genitors from river Namsen (64°27'N, 11°28'E) kept in captivity for one generation (in the absence of deliberate selection) were used to generate ten full-sib wild families at the NINA (Norwegian Institute for Nature Research) research station in Ims (Norway) in the fall of 2004. Ten hybrid (wild X farmed) genitors were crossed with ten wild genitors to generate ten full-sib backcross families (5 families had wild mothers, and 5 had hybrid mothers). Norwegian farmed genitors were taken from the seventh generation of the Norwegian breeding programme's first brood line ("population 1", sensu Gjedrem *et al.* 1991), which mainly originated from river Namsen (Gjedrem *et al.* 1991, Fleming and Einum 1997). Fertilized eggs were all kept under identical controlled conditions (24 h darkness; mean temperature (range) 3.9°C (0.8–11.1)). The progeny of the ten full-sib families of each type was mixed for rearing and a hundred fry of each type were randomly sampled. Because juvenile characteristics play a determinant role in the expression of life history traits at later stages (Metcalfe *et al*. 1989), and mortality and natural selection is commonly intense during early life (Einum and Fleming 2000a, b; Einum *et al*. 2001), this study focuses on the fry stage (young of the year). Sexually undifferentiated fry were sampled at the yolk-sac resorption stage, before exogenous feeding and immediately frozen in liquid nitrogen. Sampling stage was thus the same as in Roberge *et al.* (2006).

3.4.2 RNA extraction, labelling and cDNA hybridization

Frozen fry were homogenised individually in TRIZOL@Reagent (Invitrogen), and total RNA was extracted as previously described (Roberge *et al*. 2007, Roberge *et al*. 2006). For each sample, 15µg total RNA was retro-transcribed and labelled. Transcription profiles of 23 of pure wild and 23 backcross fry were contrasted on 23 microarrays, always using both types on each array. Dye and sample cross type coupling was flipped between biological replicates. The cDNA microarrays used here were obtained through the consortium Genomic Research on All Salmon Project (cGRASP, available from Ben F. Koop, bkoop@uvic.ca), which comprises 16,006 salmonid cDNA clones (von Schalburg *et al*. 2005).

3.4.3 Signal detection, data preparation, statistical analysis and functional classification

Signal detection and data preparation was as previously reported (Roberge *et al*. 2006). Spots with mean signal intensities smaller than the mean intensity of control empty spots plus twice its standard deviation in both channels were removed from the analysis, leaving a total of 4618 detected spots. To assess differences in gene transcription between wild and backcross fish, data were analysed using a mixed model of ANOVA (Wolfinger *et al*. 2001) and the MAANOVA R package (Kerr *et al*. 2002, Kerr *et al*. 2000). The model included the "array" term as a random term and the "sample type" (pure wild or backcross) and "dye" terms as fixed terms. A permutation-based F-test (Fs, with 1000 sample ID permutations) was then performed, and restricted maximum likelihood was used to solve the mixed model equations. To correct for multiple testing, q-values were calculated from the permutation based p-values with the Q-value R package (Storey 2002). Since samples from a same family cannot be considered as independent, our model should ideally have included a family term (Miller and Anderson, 2004). However, space limitations prevented us from rearing family groups separately and familial information on the samples was lost. We therefore had to deal with a certain level of pseudo-replication in these data, which may

inflate statistical power (Hurlbert *et al*., 1984). Consequently, we used a conservative significance threshold of q-value < 0.01. Functional classification and assessment of significant differential representation of functional classes were performed in the DAVID/EASE environment (http://david.niaid.nih.gov/david/). DAVID 2.1 (beta version) gene accession conversion tool was first used to convert gene ontology-linked identifications of various types gathered in the GRASP 16 000-gene microarray gene identification file to UNIGEN clusters. Assessment of significant differential representation of functional classes was performed with EASE 2.0.

3.4.4 Comparison with previous results

Comparison of the results of the present study and that of Roberge *et al*. (2006) must take into account two main potential caveats. First, they were performed using different cDNA microarrays, though both were prepared by the same group (cGRASP), using similar protocols (see http://web.uvic.ca/cbr/grasp). The microarray used in this study represents an "upgraded" version which comprises substantially more features (17 328 different features) than the one used in our earlier study (8 736 features with neighbor replicates, 4 368 different features). Only about 25% of the cDNA clones printed on the microarray used in the present study were also printed on the earlier microarray platform and, inversely, some of the cDNA clones printed on the earlier one are no longer printed on the new microarray (including 18 "unknown" cDNAs for which significant expression differences were detected in Roberge *et al.* 2006). Consequently, we limited our comparisons to genes represented on both array platforms, keeping for each gene the data from the cDNA clone showing the most significant hybridization signal difference on each array type (smallest associated p- and q-values), in cases where a gene was represented by several different cDNA clones. Second, the number and proportion of genes showing significant transcription level differences at a given significance threshold is not readily comparable between studies because the statistical power to detect significant changes varies with sample size and experimental variance, notably (Draghici 2003). Also, if more cDNAs are printed on the array, more statistical tests are performed and more false positives are then expected. Experimental variance is expected to be similar in both studies since all

hybridizations and extractions were performed by the same person, CR, and using the same material, reagents and protocols. Yet, over twice as many hybridizations were performed in the present study compared to Roberge *et al.* (2006) (n = 46 instead of n = 20 for the Norwegian system alone). To compare the average magnitude and range of the significant gene transcription differences between the two studies, we thus focused on the 1.4% cDNA clones representing the most significantly differentially expressed genes (*i.e.* the proportion of significant genes observed in Roberge *et al.* 2006). A common proportion was chosen instead of a given significance threshold, since with different statistical powers, a same threshold would have represented different stringencies, which could in turn have influenced the average.

3.4.5 Quantitative PCR validation of microarray results

Quantitative RT-PCR was used here to validate earlier microarray results (Roberge *et al*. 2006) for two candidate genes which had shown significant gene transcription differences between wild and farmed Atlantic salmon strains from Canada. The first candidate is an "unknown" gene (geneBank accession: CA039081), meaning the corresponding cDNA clone sequence did not generate any BLAST hits with e-value < 1 X 10^{-15} and an informative name. The second gene codes for metallothionein A. Seventy-nine frozen fry sampled at the yolk sac resorption stage from families of farmed (39) and wild salmon (40) were used (see Roberge *et al.* 2006). These individuals had not been used in our previous microarray experiments. Total RNA was extracted from the whole individuals using a standard Trizol extraction protocol, mRNA was then retro-transcribed using the *High-capacity cDNA archive kit* from Applied Biosystems. TaqMan assay primers and probes were designed with the *Primer Express 1.5* software and then tested in a validating experiment to ensure comparable efficiencies. The control gene (*peptidyl-prolyl-isomerase A*) was chosen because it showed minimal variance in its transcription level, both within and between strains. Quantitative RT-PCR experiments were performed on an ABI 7000 instrument connected to a computer running version 1.2.3 of the ABI Prism 7000 SDS Software (Applied Biosystems). During the quantitative RT-PCR, the recommended default setups from Applied Biosystems were used. Thus, a two min. step at 50ºC was programmed

to ensure activation of the No AmpErase® UNG enzyme, which was followed by a 10 min step at 95ºC to activate the AmpliTaq® Gold enzyme. After those initial steps, 40 cycles were run, which consisted of 15 sec. at 95ºC followed by one min. at 60ºC for hybridization, elongation and detection of fluorescence signal. All PCR reactions were scaled down from 50µl to 25µl and the proper detection volumes entered in the software prior each run. For the unknown gene (CA039081), the sequences and hybridization temperature of the primers used for amplification were: 5'-GATAGCTTCCAGAATACTACAGTGACAATT-3' (59°C) and 5'-CGTGGCCTTTTTTCGACTGA-3' (60°C) , while the sequence and hybridization temperature of the TaqMan probe was 5'-TGCAAAACCAGACTATTAT-3' (69°C). For metallothionein, the sequences of the primers were 5'-GCCTCACTGACAACAGCTGGTA-3' (59°C) and 5'-TGTGCTTCAGGCTGTGTGTGTGT-3' (59°C) and that of the probe 5'-CACAGGTCTTGCCC-3' (68°C). Finally, for Peptidyl-prolyl isomérase A, the sequences of the primers were 5'-AAGAACTGGGACCCGTTGGT-3' (59°C) and 5'-GCATGGGCTGTCTGTCCAT-3' (58°C) and that of the probe 5'-TTAGGGCCAGCGTTGG-3' (70°C). Following the experiments, the transcription levels of the studied genes for each individual were normalized by first subtracting the transcription level of the control gene in that individual and, second, by subtracting the difference between the transcription level of the control gene in the individual as compared to a reference individual that was maintained throughout the experiment. All statistical tests were run in R (version 2.3.1) using these normalized transcription levels. Normality of the dataset was rejected by the Shapiro-Wilk test while homogeneity of variances was accepted as tested by the Ansari-Bradley test. Wilcoxon's rank sum test was therefore used to determine the *P* values associated with the differences in normalized transcription levels between the two strains for the two studied genes. Since this experiment was aimed at validating differences already observed independently, a one-sided test was performed.

3.5 Results

Under assumptions of an additive genetic basis, more phenotypic variance is expected in second generation hybrids than in the parental populations or species (Anderson 1949).

However, F-test for variance homoscedasticity did not reveal more significant differences in the variance of gene transcription levels between the progeny of hybrid and wild salmon than expected by chance. 6.4% (298) of all detected cDNA clones represented genes which showed significantly different transcription levels (q-value < 0.01) between the progeny of wild and hybrid (backcross) salmon. DAVID/EASE analysis for over-representation of gene ontology categories among those genes when compared to all genes represented on the microarray revealed several significantly over-represented categories (Fig. 3.1, EASE score < 0.025). Each of these over-represented categories was directly related to one of four main functional axes: cytoskeleton/cellular organisation, muscle development, oxidative phosphorylation/ATP and extracellular component. Strikingly, the vast majority of the genes falling in the first category (which largely overlapped with the second) appeared over-transcribed in hybrid *versus* wild salmon, including for instance cytoskeletal keratin (10 significant cDNA clones, average: 44% over-expression), cofilin (51% over-expression) and gelsolin (47% over-expression). Inversely, all gene coding for ATP synthase subunits (main genes belonging to the third functional axis) were under-transcribed in hybrid *versus* wild salmon (3 subunits, 12 cDNA clones, average: 31% under-expression). Data for genes from the fourth functional axis (extra-cellular component) showed conflicting trends which reflected the wider array of functions this axis comprised. For instance, while all collagen coding genes were under-transcribed in hybrid relative to wild salmon (12 cDNA clones, average: 28% under-expression), two different genes whose products are involved in antigen presentation showed over-transcription in hybrids (MHC class II antigen IE-α, 4 cDNA clones, average: 52% over-transcription; β-2 microglobuline, 5 cDNA clones, 61% over-transcription). However, since functional annotation was possible for only a portion of the genes (40% of the significant cDNA clones and 26% of all clones on the microarray), results from this functional over-representation analysis might not be representative of the results in their entirety.

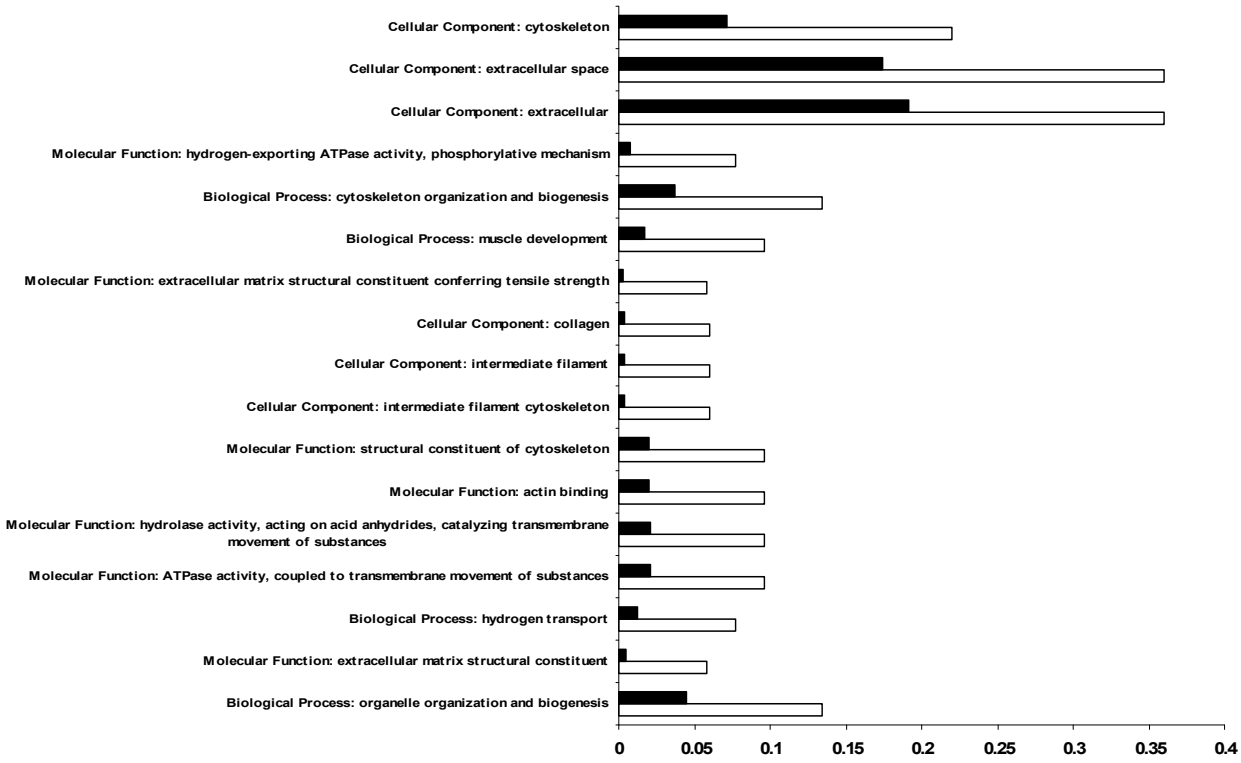


Figure 3.1 Significant over-representation of gene ontology categories among genes which showed significantly different transcription levels between the progeny of wild and hybrid backcross salmon when compared to all genes represented on the microarray.

For each gene ontology category listed, the black bar indicates the proportion of genes in this category among all known genes on the microarray while the white bar indicates the proportion of genes in this category among genes which showed significantly different transcription levels (q-value < 0.01) between the progeny of wild and hybrid backcross salmon.

Known genes for which the transcription level differences between hybrid and wild salmon fry were significant even when using a Bonferroni-corrected threshold are presented in Table 3.1. Data for 62 cDNA clones satisfied this very conservative criterion; these represented 24 genes with functional annotation and 18 "unknown" genes. Among these, the magnitude of the gene transcription differences averaged almost two fold (1.98, either over- or under-transcription in hybrid relative to wild salmon) and ranged from a 7.15 fold under-transcription to a 7.18 fold over-transcription in hybrid *versus* wild salmon.

At the $P < 0.01$ threshold on permutation-corrected p-values, 74 significant cDNA clones were observed in Roberge *et al.* (2006) compared to 656 in the present study. This represents respectively 1.4% and 13.2% of all detected clones after subtracting the expected number of false positives in each case. To evaluate how much of this difference is explainable by increased statistical power due to the greater number of replicates in the present study, we assessed the number of significant clones at $P < 0.01$ in a random subsample of 20 comparisons from the original sample of 46. The proportion and number of cDNAs representing differentially expressed genes (7.5%, 392, $P < 0.01$) was still substantially higher than those observed between farmed and wild Norwegian salmon in Roberge *et al.* (2006).

The average magnitude (for the 1.4% cDNA clones representing the most significantly differentially expressed genes in both studies) of the differences between backcross and wild salmon was 76% (in either direction), which is over three times that observed between wild and farmed strains (25%). Significant gene expression differences in backcross *versus* wild individuals ranged from a 7.15 fold under-transcription to a 7.18 fold over-transcription, while the highest fold difference observed between pure wild and pure farmed Norwegian salmon strains was of 1.67 fold.

Various candidate genes whose transcription levels appear to have evolved as a result of artificial selection and domestication were identified in Roberge *et al.* (2006). What happened to these differences after two generations of interbreeding between farmed and wild salmon? Table 3.2 contrasts transcription level fold changes for those genes when

Table 3.1 Gene products corresponding to known genes that showed significant gene transcription differences between the progeny of hybrid backcross and wild salmon when using a Bonferroni-corrected significance threshold (P < 1.1X10-5, all in bold in the "BC/wild" column).

Gene product	# clones	BC/wild	Farmed/wild	Transmission
ADP,ATP carrier protein 2	2	**0.75**	0.91	TMUT
C1q-like adipose specific protein	2	**1.10**	NA	NA
Fc receptor beta subunit family	2	**2.24**	NA	NA
Low affinity immunoglobulin epsilon Fc receptor	2	**0.14**	NA	NA
Anterior gradient-like protein	5	**2.81**	0.93	TMOT
Parvalbumin	4	**1.14**	0.82	TMOT
Beta-actin	2	**1.22**	**1.36**	A
Lysozyme G	2	**0.85**	**0.85**	D
Thioredoxin-like protein p19 precursor	2	**3.49**	0.93	TMOT
Collagen alpha 1(X) chain	3	**0.50**	1.01	TMUT
Troponin I, fast skeletal muscle	2	**1.37**	1.13	TMOT
Invariant chain S25-7	1	**1.55**	NA	NA
Vitellogenin	1	**1.26**	1.13	TMOT
Ependymin	2	**0.50**	0.88	TMUT
Deoxyribonuclease I-like 3	2	**1.49**	NA	NA
Phosphoglycerate kinase 1	1	**0.31**	NA	NA
Gastrotropin	1	**0.57**	NA	NA
Claudin	1	**2.09**	1.03	TMOT
alpha-globin and beta-globin	1	**0.68**	**1.29**	TMUT
Ribosomal protein S3a	1	**1.80**	1.07	TMOT
ATP synthase lipid-binding protein	2	**0.66**	1.11	TMUT
Matrix metalloproteinase-2	1	**1.07**	**1.24**	A
Myosin regulatory light chain 2	1	**0.62**	1.08	TMUT
Hemopexin	1	**0.66**	0.85	TMUT

The first column presents the number of different cDNA clones among the 62 showing significant differences at this threshold corresponding to each gene. Transcription level fold changes for those genes when compared between the progeny of wild and farmed salmon (Roberge *et al.* 2006) or wild and hybrid backcross (BC) salmon (this study) are presented for comparison, as well as our interpretation of the transmission pattern observed. A: additive transmission, meaning that the BC value is intermediate between farmed and wild values. D: dominant transmission, meaning that the BC value is near to that observed in farmed salmon. TMUT : transgressive mean under-transcription, meaning that the value observed in BC is lower than in both parental strains. TMOT : transgressive mean over-transcription, meaning that the value observed in backcross is substantially higher than in both parental strains. Fold changes in bold in the "farmed/wild" column were significant in Roberge *et al.* (2006, P < 0.01). NA indicates cases were clones representing a gene could not be identified on the microarray used in Roberge *et al.* (2006).

compared between the progeny of wild and farmed salmon on the one hand, and that of wild and backcross salmon on the other hand. The difference in gene transcription observed between pure wild and pure farmed strain (Roberge *et al.* 2006) was either significant but smaller or not significant between wild and backcross hybrids for 67% of the 33 different known genes that could be compared (Table 3.2). This suggested that the genetic control of the expression of those genes was additive (Gibson *et al.* 2004). The remaining genes showed genetic control of gene transcription more akin to low- or high-parent dominance (9%), or to other patterns (24%) which were previously described as gene expression over- or under-dominance (*e.g.* Gibson *et al.* 2004). Here we did not adopt this terminology since over- or under-dominance refers to one locus inter-allelic interactions, while gene transcription is expected to be generally controlled by several loci (Brem *et al*. 2002). Nor do we consider the heterosis/outbreeding depression terminology (*e.g.* (Vuylsteke *et al*. 2005) as appropriate here since we do not know how the transgressive mean fold-changes observed affect growth or other performances. Instead we used the terminology "transgressive mean under-" and "over-transcription" (TMUT and TMOT).

If most differences in gene transcription previously observed between wild and farmed salmon were reduced when comparing hybrid and pure wild salmon, where do the differences observed in the present study come from? Table 3.1 presents data for the 24 different known genes which showed significant gene transcription difference between hybrid and wild salmon under a Bonferroni-corrected threshold. For comparison, we included fold change data from (Roberge *et al*. 2006) in the 17 cases for which the genes were also represented on the earlier array. Strikingly, control of gene transcription by additive gene interactions was compatible with only 12% of the cases while most (82%) of the gene transcription differences observed between hybrids and wild salmon was transgressive (either TMUT or TMOT).

Gene transcription level differences obtained with qRT-PCR assays were highly significant for the two candidate genes tested and concordant the microarray results (Fig. 3.2), although with different average fold changes (metallothionein: 21% under-expression from the microarray assays and 36% under-expression with the qRT-PCR; CA039081: 58%

Table 3.2 Gene products corresponding to known genes that showed significant gene transcription differences between the progeny of Norwegian farmed and wild salmon in Roberge et al, 2006 ($P < 0.01$; all in bold in the "farmed/wild" column).

Gene product	Farmed/wild	BC/wild	q-value	Transmission
Creatine kinase	**1.23**	1.05	$1.20X10^{-1}$	A
Glyceraldehyde phosphate dehydrogenase	**0.83**	**0.87**	$1.34X10^{-2}$	A or D
NADH dehydrogenase subunit 5	**1.31**	0.95	$2.07X10^{-1}$	A*
NADH dehydrogenase subunit 4	**1.30**	0.98	$4.99X10^{-1}$	A*
ATP synthase beta-subunit	**0.84**	**0.78**	$2.93X10^{-2}$	D or TMUT
Ferritin H	**1.23**	**1.30**	$7.15X10^{-2}$	A*
Chaperonin containing T-complex polypeptide 1, epsilon subunit	**1.29**	**1.35**	$1.14X10^{-2}$	D or TMOT
Ribosomal protein L10	**1.22**	**1.08**	$4.08X10^{-2}$	A
Ran protein	**1.24**	**1.14**	$4.20X10^{-2}$	A
Elongation factor 1 alpha	**1.24**	0.93	$3.80X10^{-1}$	A*
Protein synthesis initiation factor 4	**1.24**	**1.04**	$3.66X10^{-3}$	A
Alpha 3 type I collagen	**1.59**	**0.82**	$2.38X10^{-2}$	TMUT
Alpha 2 type I collagen	**1.67**	**0.79**	$3.04X10^{-3}$	TMUT
Matrix metalloproteinase-2	**1.34**	**1.07**	$7.17X10^{-4}$	A
Secreted protein, acidic, rich in cysteine (SPARC)	**1.30**	**0.81**	$1.81X10^{-3}$	TMUT
Mannose binding-like lectin	**0.84**	1.00	$1.80X10^{-1}$	A*
Pentraxin	**0.86**	1.16	$6.10X10^{-1}$	A*
Lysozyme G	**0.85**	0.85	0	D
BA1 beta-2 microglobulin	**0.84**	**1.72**	$1.02X10^{-3}$	TMOT
alpha-globin	**1.28**	**0.68**	$3.52X10^{-4}$	TMUT
Apolipoprotein A-I	**1.89**	**0.71**	$1.06X10^{-2}$	TMUT
Plasma retinol-binding protein 1	**0.83**	1.32	$1.58X10^{-1}$	A*
Calmodulin	**1.29**	1.13	$2.15X10^{-1}$	A
Transducer of ERBB-2	**0.81**	0.97	$2.58X10^{-1}$	A
Beta actin 1	**1.64**	1.22	0	A
Chitinase	**1.30**	**1.26**	$2.02X10^{-2}$	A ou D
Cathepsin D	**0.84**	0.92	$1.56X10^{-1}$	A
Cathepsin L	**1.22**	**0.73**	$4.57X10^{-3}$	TMUT
Similar to Ornithine decarboxylase antizyme	**1.24**	1.04	$6.10X10^{-2}$	A
High mobility group-T protein	**1.27**	1.08	$3.93X10^{-1}$	A
Nogo-A	**1.31**	1.08	$3.79X10^{-1}$	A
Dihydropyrimidine Dehydrogenase, Chain A	**1.19**	0.92	$5.98X10^{-1}$	A*
Cofilin	**1.21**	**1.51**	$6.15X10^{-3}$	TMOT

Transcription level fold changes for those genes when compared between the progeny of wild and farmed salmon (Roberge et al, 2006) or wild and hybrid backcross (BC) salmon (this study) are presented for comparison. The q-value indicating the significance of the fold-change observed here between wild and hybrid backcross salmon is also included, as well as our interpretation of the transmission pattern observed. A: additive transmission, meaning that the BC value is intermediate between farmed and wild values or that a significant difference in Roberge et al (2006) is not significant anymore in the present study (A*). D: dominant transmission, meaning that the BC value is similar to that observed in farmed salmon. TMUT : transgressive mean under-transcription, meaning that the value observed in backcross is substantially lower than in both parental strains. TMOT : transgressive mean over-transcription, meaning that the value observed in backcross is substantially higher than in both parental strains. Fold changes in bold in the "BC/wild" column were significant at the q-value < 0.05 threshold. Q-value represent FDR-corrected p-values.

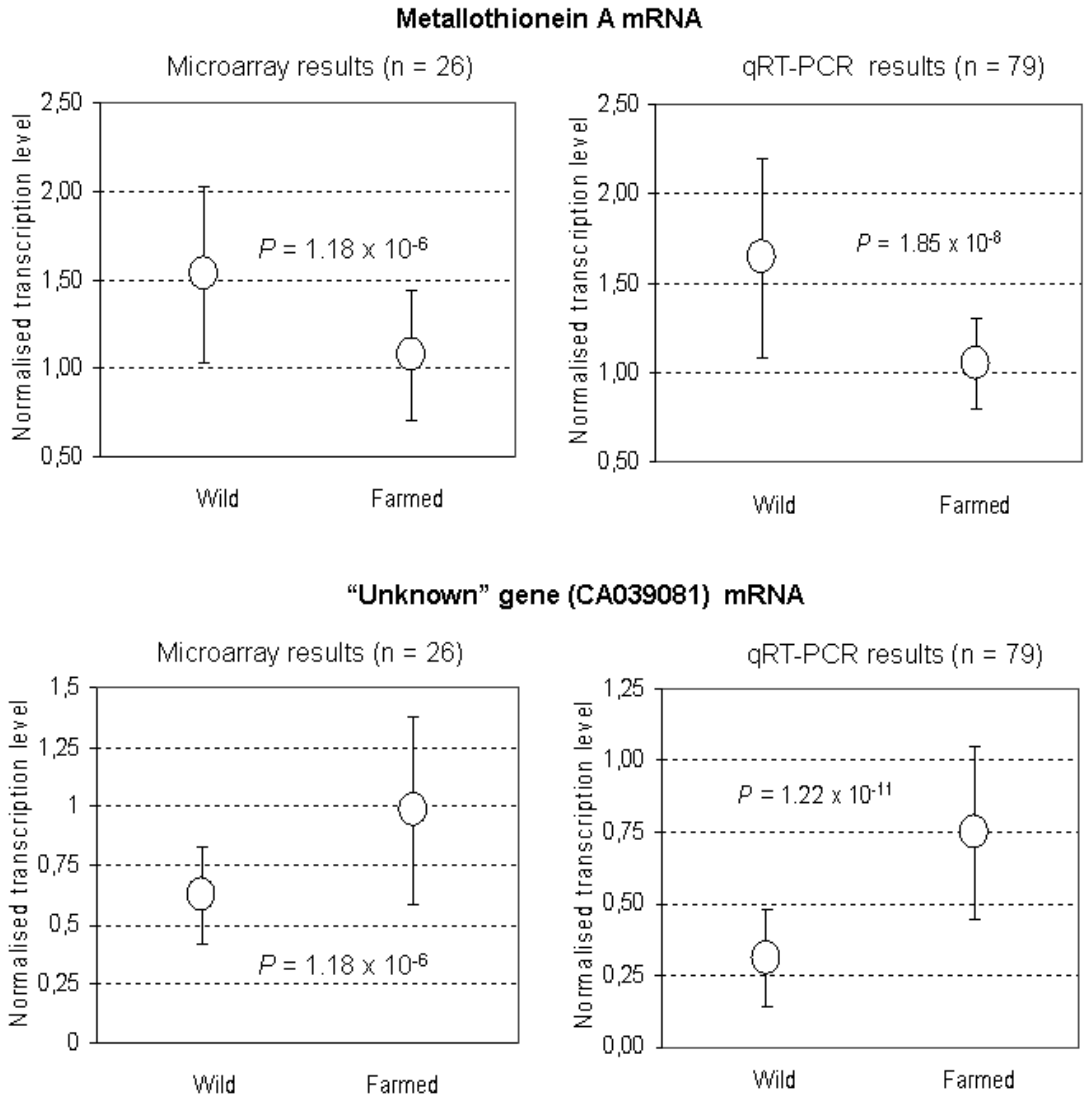


Figure 3.2 Quantitative real-time PCR (qRT-PCR) verification of previous results for two differentially expressed genes (metallothionein A-coding gene and the "unknown" gene cDNA clone CA039081) between the progeny of wild and farmed Canadian salmon.

Microarray results (n = 26) for those two candidates are presented on the left and qRT-PCR results (n = 79) on the right. Dots indicate the normalised average transcription level. Bars indicate the standard deviation of the normalized gene transcription levels. For metallothionein microarray results, results from the most significantly differentially expressed cDNA clone are presented (six metallothionein A clones showed significant gene transcription difference in the same direction between farmed and wild salmon).

over-expression from the microarray assays and 141% over-expression expression with the qRT-PCR). Within group standard deviation was comparable between the microarray and qRT-PCR results (Fig. 3.2), even though sample size was smaller for the microarray experiment (n = 26 *versus* n = 79). This suggests that the microarray platform and protocols we used produced very reproducible results. Yet, qRT-PCR on a larger number of biological samples provides more statistical power to detect differences. It is therefore useful as a validation technique, in particular in cases where the significance of the microarray observed differences is not conservatively high.

3.6 Discussion

Many genes (n = 298) with highly significant transcription level differences between the progeny of Norwegian wild salmon and that of a second generation hybrid backcross of farmed to wild salmon were identified here. Over-representation of genes implicated in cytoskeleton/cellular organisation, muscle development, oxidative phosphorylation and extracellular components (collagen matrice, antigen presentation) among those genes suggests that important physiological and morphological differences may well be persisting between wild and hybrid farmed salmon, even though genetic differences for most phenotypic traits tested were shown to be milder between backcross hybrids and wild than between farmed and wild salmonids (Anderson 1949, Tymchuk *et al*. 2006, McGinnity *et al*. 2003).

The proportion and number of differentially expressed genes was substantially higher here than observed between farmed and wild Norwegian salmon in Roberge *et al.* (2006), even when correcting for the increased statistical power due to larger sampling. This observation is somewhat analogous to that of Gibson *et al.* (2004), where 33% of the genes showed significant transcription differences between a hybrid *Drosophila* strain and its parental isogenic lines, while only 25% of the genes showed significant gene transcription differences between the two parental strains. However, we cannot rule out the possibility that other factors could have increased statistical power in the present study (*e.g.*

pseudoreplication (as discussed above) or an increased constancy in the cDNA printing process, reducing experimental variance). We thus conservatively conclude that the number and proportion of genes showing significant gene expression differences between hybrid backcross and wild salmon are at least as important as observed in the comparison of pure wild and farmed salmon of the same strain.

The range and average magnitude of the significant differences observed here between the hybrid and wild salmon were strikingly higher than previously observed between pure farmed and wild individuals. Differences of this magnitude are comparable to those observed between wild salmon from Norway and Canada (Roberge *et al*. 2006), except that in this case, a large proportion of the gene expression profiles showed non-homogenous variances, which was surprisingly not the case in the present study. We interpret these results as evidence for the prevalence of non-additive genetic control of gene transcription (see below). However, the fact that more genes were represented on the microarray used in this study could have increased the odds of observing extreme differences and hence be responsible for part of the observed results. Yet, the observation of high fold differences for genes which were also represented on the earlier microarray platform (see table 3.1) does not support this possibility. For instance, thioredoxin-like protein 19-coding gene was represented on both microarrays and showed highly significant transcriptional change of high magnitude (3.49 fold) in the present study but no significant difference in Roberge *et al.* (2006).

A majority (67%) of the genes which had shown significant transcription differences between farmed and wild salmon (Roberge *et al.* 2006) showed smaller or non-significant differences here, which concords with additivity of their gene transcription levels. However, the vast majority of the genes which showed significant differences in the present study (82%) had average normalized gene transcription levels either higher or smaller than those of both pure farmed and wild salmon, which suggests a non-additive genetic control of gene transcription. This observation indicates that the genetic basis of transcriptional control is likely to have been modified through artificial selection and domestication for substantially more genes than we previously detected when comparing pure strains only

(Roberge *et al.* 2006). A hypothetical explanation for the high prevalence of new gene transcription changes with a non-additive genetic basis in hybrids is that pleiotropic changes may have had balancing effects on the gene transcription levels of several genes in farmed salmon. This balance would have been lost in segregating hybrids, revealing additional gene transcription differences. Overall, these results support the view that the genetic architecture of gene transcription in farmed salmon has rapidly evolved under artificial selection, such that genetic interactions between farmed and wild salmon may generate unpredicted phenotypes by inducing novel patterns of gene expression.

The result that only a minority of the gene transcription differences observed between hybrid backcross and wild salmon showed transmission patterns compatible with additive gene interaction has also been reported in studies performed on other organisms (*Drosophila*: Gibson *et al*. 2004, <2% additivity of gene transcription; *Arabidopsis*: Vuylsteke *et al*. 2005, 27 to 37.5% additivity, maize: Auger *et al*. 2005, 37% additivity; oyster: Hedgecock *et al*. 2007, 2% additivity). In contrast, other studies have observed a majority of transcription profiles under additive genetic control (maize: Swanson-Wagner *et al*. 2006, 78% additivity; mice: Cui *et al*. 2006. At the level of morphological or physiological traits, additivity is generally admitted as the norm, yet non-additive interactions resulting in either heterosis or outbreeding depression are nonetheless widespread (Rieseberg *et al*. 1999, Lippman and Zamir 2007). In natural conditions, Atlantic salmon farmed X wild hybrids (F1 and backcross) were shown to have phenotypes intermediate to their wild and farmed parental lineages for several life history traits (growth, survival and parr maturity rates), suggesting that the genetic basis for these traits is mainly additive (McGinnity *et al*. 2003). These results were also confirmed for other salmonids in laboratory studies where farmed and wild lineages of rainbow trout (*Oncorhynchus mykiss*) (Tymchuk and Devlin 2005) and Coho salmon (*Oncorhynchus kisutch*) (Tymchuk *et al*. 2006) were crossed and the various hybrids compared to their parental lineages for several traits. On the other hand, non-additive genetic interactions resulting in heterosis and outbreeding depression has also been reported for salmonids (Leary *et al*. 1983, Gharrett *et al*. 1999, Gilk *et al*. 2004). As pointed out by Gibson *et al.* (2004), most gene transcription profiles appear as non-additive traits while most traditional

phenotypic traits (morphological, behavioural, physiological) appear as additive traits. This suggests that the apparent additivity at the phenotypic level might result from complex interactions between transcripts whose expression is controlled in a non-additive manner. To clarify this important issue, further studies should aim at establishing a causal link between transgressive phenotypes (including heterosis and heterozygote depression) and gene transcription, as exemplified by Hedgecock *et al.* (2007) and Meyer *et al.* (2007). Such studies would also allow a better understanding of the potential fitness consequences of non-additive control of gene transcription in post F1 farmed X wild salmon hybrids.

3.6.1 Conservation implications

The interbreeding of escaped farmed salmon with wild individuals is considered a serious threat to natural populations (McGinnity *et al.* 2003). A pressing question is therefore to what extent this interbreeding can be expected to modify the genetic make-up of wild Atlantic salmon populations. Here, we compared the genome-wide gene transcription profiles of wild salmon and second generation hybrids (backcross: (farmed X wild) X wild). Strikingly, we observed more and larger differences than we had previously observed between the progeny of pure farmed and wild genitors (Roberge *et al*. 2006). This suggests that interbreeding of fugitive farmed salmon and wild individuals could substantially modify the genetic control of gene transcription in natural populations exposed to high migration from fish farms, resulting in unpredictable and potentially detrimental effects on the survival of these populations. Moreover, in the absence of strict policies to considerably reduce the number of escaped farmed salmon, this number can be expected to increase due to the rapid development of the aquaculture industry, further modifying the genetic makeup of natural populations. Also, sea farming is developing rapidly for other marine fish species (*e.g.* Atlantic cod and halibut, see Naylor *et al.* 2005). Therefore, similar risks could eventually threaten natural populations of those species as well. Hence, results from the present study further support the idea that measures to considerably reduce the number of escaped farmed salmon and their reproduction in the wild are urgently needed (e.g. Naylor *et al.* 2005; McGinnity *et al.* 2003).

3.7 Acknowledgments

This work was supported by a research grant from Science and Engineering Research Canada (NSERC strategic programme) to L.B. and H.G., a postgraduate scholarship from the *Fonds Québécois de la Recherche sur la Nature et les Technologies* (FQRNT) to C.R., and funding from the Research Council of Norway to S.E.. We also thank Tom Smith, Sébastien Renaut and two anonymous referees for their constructive comments. This study is a contribution to the research programmes of CIRSA (*Centre Inter-universitaire de Recherche sur le Saumon Atlantique*). We thank the staff at the NINA Research Station, Ims, Norway for their assistance in rearing fish.

Chapitre 4 : Genome-wide identification of genes under directional selection: gene transcription Q_{st} scan in diverging Atlantic salmon subpopulations

4.1 Résumé

La génomique évolutive a bénéficié de méthodes permettant l'identification, à l'échelle du génome entier, de régions génomiques importantes pour l'évolution, comme le «balayage génomique» (*genome scan*) et la cartographie de QTL. Récemment, l'utilisation de bio-puces a permis d'identifier, toujours à l'échelle du génome entier, des différences de transcription génique entre espèces ou populations. L'identification de gènes présentant des différences d'expression entre taxons ne suffit toutefois pas en elle-même à conclure que ces différences sont le résultat de la sélection naturelle. Nous proposons ici une approche, le «balayage transcriptomique», permettant d'évaluer le rôle de la sélection dans la transformation des profils de transcriptions entre populations. Nous avons comparé les transcriptomes de deux sous-populations de saumon atlantique qui ont commencé à diverger il y a au plus six générations. Après avoir testé, gène par gène, la normalité et l'unimodalité des données de transcription, la base génétique additive de la transcription génique a été estimée en utilisant un modèle animal. Les estimés d'héritabilité du niveau de transcription étaient significatifs pour 16% (1044) de tous les gènes exprimés. Nous avons identifié, à partir de la distribution des Q_{st} d'expression (estimés à partir des composantes intra- et inter-sous-population de la variance génétique), 16 gènes présentant des valeurs extrêmes de Q_{st} d'expression (Q_{st} d'expression moyen : 0.11). Le niveau de transcription de ces gènes pourrait avoir évolué sous l'influence de la sélection directionnelle en 6 générations seulement. Cette étude contribue ainsi à l'exploration des données de transcription génique par la génétique quantitative, et ce tant au niveau empirique que méthodologique.

4.2 Abstract

Evolutionary genomics has benefited from methods that allow identifying evolutionarily important genomic regions on a genome-wide scale, including genome scans and QTL mapping. Recently, genome-wide scanning by means of microarrays has permitted assessing gene transcription differences between species or populations. However, the identification of differentially transcribed genes does not in itself suffice to measure the role of selection in driving evolutionary changes in gene transcription. Here, we propose and apply a "transcriptome scan" approach to investigate the role of selection in shaping differential profiles of gene transcription between populations. We compared the genome-wide transcription levels between two Atlantic salmon sub-populations that have been diverging for only six generations. Following assessment of normality and unimodality on a gene *per* gene basis, the additive genetic basis of gene transcription was estimated using the animal model. Gene transcription h^2 estimates were significant for 1044 (16%) of all detected cDNA clones. In an approach analogous to that of genome scans, we used the distribution of the Q_{st} values estimated from intra- and inter-subpopulation additive genetic components of the transcription profiles to identify 16 outlier genes (average Q_{st} estimate = 0.11) whose transcription levels are likely to have evolved under the influence of directional selection within six generations only. Overall, this study contributes both empirically and methodologically to the quantitative genetic exploration of gene transcription data.

4.3 Introduction

According to the theory of adaptive radiation, colonisation of new environments may promote rapid population divergence as a by-product of local adaptation to differential selective regimes (Schluter 2000). Elucidating the genetic changes underlying phenotypic evolution resulting from this process is a central objective of evolutionary biology. Reaching this goal has recently greatly benefited from the development of various experimental strategies, including genome scans, QTL analysis and gene expression QTL (eQTL) mapping (Vasemagi and Primmer 2005). Genome-wide scanning for gene transcription differences between species or populations using microarrays is increasingly being performed on a wide range of organisms (*e.g.* Jin *et al*. 2001, Brem *et al*. 2002, Townsend *et al*. 2003, Bochdanovits *et al*. 2003, Oleksiak *et al*. 2002, Roberge *et al*. 2006, Derome *et al*. 2006, Giger *et al*. 2006). However, the identification of genes showing significant differences in transcription levels between groups does not in itself suffice to conclude that such changes were driven by divergent selection, even when environmental conditions were controlled. Thus, the role of selection in shaping patterns of gene transcription is still contentious. For instance, Khaitovich *et al.* (Khaitovich *et al*. 2005) suggested that most inter-specific differences in transcript levels evolve neutrally, whereas others argued that strong stabilizing selection dominates the evolution of transcriptional change (Denver *et al*. 2005, Rifkin *et al*. 2005). Theory predicts that both the magnitude and speed of phenotypic change (including change in gene transcription level) in response to selection depend on the heritability of a trait (Falconer and Mackay 1996). Yet, only a handful of studies have formally estimated h^2 of gene transcription (reviewed in Stamatoyannopoulos 2004, Gibson and Weir 2005).

Because transcription profiles can be considered as reflecting variation at both phenotypic and genomic levels (Whitehead and Crawford 2006), the synergy of methods that can detect the role of selection in shaping differences between populations at the phenotypic and genetic level may represent an efficient means to investigate evolutionary changes in gene transcription profiles. On the one hand, comparing the extent of quantitative genetic differences of phenotypic traits between populations (Q_{st}) (Spitze 1993) with that observed at neutral marker loci (F_{st}) has been widely used to assess the relative contributions of

selection and drift to phenotypic divergence between populations (Koskinen *et al*. 2003, Merila and Crnokrak 2001). Q_{st} is defined as:

$$\sigma^2_{GB}/ (\sigma^2_{GB}+ 2\sigma^2_{GW}) \qquad (1)$$

where σ^2_{GB} and σ^2_{GW} represent the among and average within population components of the genetic variance for quantitative traits, respectively. There is no theoretical constraint for applying the Q_{st} framework to gene transcription profiles (Gibson and Weir 2005). Yet, this has never been done to our knowledge. On the other hand, genome scans, which consist in analysing many (several hundreds to several thousands) molecular markers to reveal patterns of genetic differentiation at the genome scale, have become a popular means for identifying genes evolving under the effect of selection between populations. The underlying principle of this approach is that gene flow will affect all loci across the genome, whereas selection will act locally on specific genes (Luikart *et al*. 2003). As a result, only a few loci will display an atypical pattern of variation caused by the influence of locus-specific forces (Storz 2005). It is then possible to reveal these "outlier" loci by comparison with the rest of the genome. Namely, outlier genes showing the highest levels of differentiation between populations will represent those that are most likely to evolve under directional selection whereas those with the lowest level of differentiation are more likely to be under the effect of stabilising selection (Beaumont 2005).

In this study, we propose and apply a "transcriptome scan" approach combining both the Q_{st} and genome scan frameworks to investigate the role of selection in shaping differential profiles of gene transcription between populations. Recently diverging (six generations) subpopulations of Atlantic salmon (*Salmo salar*), which respectively reproduce in an upstream and downstream stretch of the Sainte-Marguerite River, Canada, were used as a model system. More precisely, a 16 006-gene cDNA microarray was used for comparing the genome-wide gene transcription data of the progeny from several half-sib families from these two subpopulations reared in a controlled environment. Using a quantitative genetic approach based on the animal model (Kruuk 2004), we estimated both gene transcription heritability and Q_{st} for each gene represented and detected on the microarray. We then performed a "transcriptome scan" by which the distribution of individual Q_{st} estimates for all genes with significant transcription profile h^2 was obtained and genes in the upper tail of

the distribution identified as potentially under directional selection. Genes with high transcription profile Q_{st} estimates might not be genetically linked to the causal genetic sequence differences. Nevertheless, high transcription Q_{st} estimates are expected to pinpoint the most likely genes for which transcription profiles evolved as a result of selection, and thus provide valuable information on the actual functions that have been targeted by selection.

4.4 Materials and methods

4.4.1 Habitat characteristics, sampling, and crossing design

The Sainte-Marguerite River (48°20'N, 70°00'W) is located 250 km northeast of Quebec City, Canada, and is subdivided into two main branches. The northeast branch is 85 km long, but access to salmon is limited to the lower 35 km by waterfalls and was limited to the lower 6.5 km until a fish ladder was installed at the level of a waterfall in 1981. In 1997, highly significant genetic differentiation at microsatellite loci (F_{st} = 0.028 and 0.036 (two upstream sites), $P < 0.0006$) was revealed between salmon subpopulations from downstream and upstream of the fish ladder (Garant *et al*. 2000), indicating partial reproductive isolation between them. Moreover, Aubin-Horth *et al.* (Aubin-Horth *et al*. 2006) also reported a significantly higher proportion of precocious male parr (early maturing males) and a smaller size threshold for male parr to mature upstream than downstream from the fish ladder. Environmental conditions also differ between river sections. The downstream stretch is surrounded by bare and unstable clay cliffs. Heavy rainfall leaches the clay from these cliffs into the river, resulting in different water coloration and chemistry relative to the upper reaches (CIRSA, unpubl. data). Overall, evidence for restricted gene flow as well as differences in habitat characteristics and life histories suggest that upstream and downstream salmon subpopulations may be in the course of becoming genetically adapted to their respective environments. Eight male and two female adult salmon from each of the upstream and downstream populations were captured in the summer of 2003 and 2004, and factorial crosses within subpopulations were made at the provincial fish hatchery in Tadoussac (15 km from the Ste-Marguerite River)

during the fall of those years. Crosses should have yielded a total of 64 half sib families, but two families were lost in the first year and seven in the second, for a total of 55 families. Fertilised eggs were incubated in controlled conditions (identical for all families) and alevins were sampled at the yolk-sac resorption stage. Four individuals from each family were collected and immediately frozen in liquid nitrogen.

4.4.2 RNA extraction, labelling and cDNA hybridisation

Whole frozen alevins (immature fish still living on food reserves of their yolk-sacs; here resources from the yolk-sacs were completely or almost completely depleted) were homogenised individually in 1 ml TRIZOL@Reagent (Invitrogen) using a Diax 100 homogeniser (Heidolph instruments), and total RNA was extracted as previously described (Roberge *et al*. 2006). Briefly, 200 µl chloroform (Sigma) was added to each ml of fish homogenate in Trizol. After mixing and centrifuging (12 000 X *g*, 0°C, 15 min), the aqueous layers were transferred into new tubes where 1ml isopropanol (Sigma) was added. Samples were then stored overnight at -80°C. The following day, they were centrifuged for one hour (12 000 X *g*, 0°C) and the isopropanol discarded. The pellets were washed with 1 ml 70% ethanol, dried for 15 min at room temperature, re-suspended in 40µl non-DEPC, treated nuclease-free water (Ambion) and spiked with 1 µl RNAse inhibitor (Ambion). For each sample, 15 µg of the pooled RNA from the four separate extractions was retro-transcribed and labelled using Genisphere 3DNA Array 50 kit, Invitrogen's Superscript II retro-transcriptase and Cy3 and Alexa 647 dyes (Genisphere). The detailed protocol of the retro-transcription, labelling and hybridisation procedures can be found at **http://web.uvic.ca/cbr/grasp/** (Genisphere Array 50 Protocol). Briefly, 15 µg total RNA were reverse-transcribed using special oligo d(T) primers with 5' unique sequence overhangs for the labelling reactions. Microarrays were prepared for hybridisation by washing twice for 5 min in 0.1% SDS, five times for 1 min in MilliQ H_2O, immersing 3 min in 95°C MilliQ H_2O, and drying by centrifugation (5 min at 800 X *g* in 50 ml conical tubes). The cDNA was hybridised to the microarrays in a formamide-based buffer (25% formamide, 4X SSC, 0.5% SDS, 2X Denhardt's solution) with competitor DNA (LNA dT

bloker, (Genisphere), human COT-1 DNA (Sigma)) for 16 h at 51°C in a humidified hybridisation oven. The arrays were washed once for 5 min at 45°C (2X SSC, 0.1% SDS), twice for 3 min in 2X SSC, 0.1% SDS at room temperature (RT), twice for 3 min in 1X SSC at RT, again twice for 3 min in 0.1X SSC at RT, and finally dried by centrifugation. The Cy3 and Alexa 647 fluorescent dye attached to DNA dendrimer probes (3DNA capture reagent, Genisphere) were then hybridized to the bound cDNA on the microarray, using the same hybridisation solution as earlier: the 3DNA capture reagents bound to their complementary cDNA capture sequences on the oligo d(T) primers. This second hybridisation was carried out over 2 h at 51°C in a humidified hybridisation oven. The arrays were then washed and dried as mentioned before.

The design for the experimental microarray hybridisation differed for the two years. In 2003, we were mainly interested in assessing the functional genomic basis of sexual precocity (Roberge *et al.*, unpublished data), such that the progeny of one anadromous male and one precocious parr from the same subpopulation and sharing the same mother were always coupled on a same microarray, for a total of 24 microarrays and 48 offspring analysed. For 2004, one offspring from each the upstream and downstream subpopulation were always coupled on a same microarray, for a total of 21 microarrays and 42 offspring analysed. Thus, transcription profiles could be obtained for a total of 90 offspring representing 50 halfsib families.

4.4.3 Signal detection, data preparation and ANOVA

Hybridisation signals were detected using a ScanArray scanner (Packard BioScience). Spots were located and quantified with the QuantArray 3.0 software, using the histogram quantification method and keeping the mean value of intensity for each spot. Local background and the data from bad spots were removed. Missing data were then imputed using the K-nearest neighbours imputer in SAM (Tusher *et al*. 2001) (15 neighbours). Data were normalised by dividing by the channel mean. Genes with mean intensity smaller than the mean intensity of control empty spots plus twice its standard deviation in both channels were discarded, leaving data from a total of 6484 detected clones to be analysed. Prepared data were corrected for intensity-linked distortion using a regional LOWESS algorithm using the R/MAANOVA package (**http://www.jax.org/staff/churchill/labsite/software**

/Rmaanova). The significance of the observed differences in transcription level between subpopulations was assessed using the R/MAANOVA package (Kerr *et al*. 2002, Kerr *et al*. 2000). The ANOVA model included the "site" (above or below the ladder), "sire type" (anadromous or precocious), "year" and "dye" terms as fixed terms. A permutation based F test (Fs, with 1000 sample ID permutations) was then performed, and restricted maximum likelihood was used to solve the mixed model equations. Best linear unbiased estimates (BLUE) of the "dye" effect were subtracted from the normalised data for each gene and dye; the obtained data were used to assess normality before to be fitted into an animal model for heritability and Q_{st} estimation (see below).

4.4.4 Normality and uni/multimodality assessment

Normality is generally assumed in microarray data but rarely verified because of the small sample sizes typical in such experiments (Draghici 2003). Here, we used an R script to obtain, for each of the 6484 detected cDNA clones, Pearson's correlation test P-values and R^2 between the observed Q-Q plot of the 90 gene transcription data points and that expected for a normal distribution, using the R program (version 2.3.0). A similar strategy was used by Giles and Kipling (Giles and Kipling 2003) as an alternative to the Shapiro-Wilk normality test, which they considered over-stringent (in the sense that, even with a small sample size, it has sufficient power to detect as significant very slight departures from normality). We also performed the Shapiro-Wilk normality test on the gene transcription data of each detected cDNA clone in R. Unimodality of gene transcription data was tested for each of the 6484 detected cDNA clones using an R script and the Diptest package (version 0.25-1). The Diptest package (Maecher and Ringach 2004) computes Hartigan's dip statistic for an empirical distribution, which is the maximum difference between this empirical distribution function and the unimodal distribution function that minimises that maximum difference; it is consistent for testing any unimodal against any multimodal distribution (Hartigan and Hartigan 1985). Interpolating from the table of quantiles from a large simulation for Hartigan's Dip test (qDiptab) in the Diptest package for R, we found that a dip statistics of 0.054 corresponds to a tail probalility of 5% (for n = 90). Hence, we

interpreted transcription profiles of genes with dip > 0.054 ($P < 0.05$) as showing significant departure from unimodality.

4.4.5 Heritability and Q_{st} estimation

Components of the genetic variance in the gene transcription level for each detected gene were estimated from an animal model (Kruuk 2004) using the PEST (Parameter Estimation Software 3.0, (Groeneveld *et al*. 1990) and VCE (Variance Component Estimator 5.1, (Kovac *et al*. 2002) programs. PEST was used for recoding phenotypic and pedigree information into a form usable by VCE (Perry *et al*. 2005). VCE was then used to estimate genetic and phenotypic variance within and between subpopulations using restricted maximum likelihood (REML) to solve a linear model of the form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 a} + \boldsymbol{Z_2 b} + \boldsymbol{Z_3 m} + \boldsymbol{Z_4 t} + \boldsymbol{e} \qquad (2)$$

where $\boldsymbol{y}$ is a vector of the transcription levels for a given gene (n = 90 individuals), $\boldsymbol{X}$ the design matrix relating the appropriate fixed effect ($\boldsymbol{\beta}$) to each individual and $\boldsymbol{Z}\mathbf{i}$ refers to design matrices relating the appropriate random genetic effects ($\boldsymbol{a}$: vector of the additive genetic effect, $\boldsymbol{b}$: vector of subpopulation effect, $\boldsymbol{m}$: vector of maternal effect, $\boldsymbol{t}$: vector of sampling year effect) to each individual, and $\boldsymbol{e}$ is the error term. A Perl script was used to run the programs sequentially and collect the information for all genes into a single output file (available upon request). Heritability estimates and their standard errors were collected directly from VCE output files, whereas Q_{st} values were estimated from equation (1), using the animal genetic variance in gene transcription to represent σ^2_{GW} and the variance from the parental subpopulation (upstream or downstream) as σ^2_{GB}. Significance of heritability and σ^2_{GB} estimates for each gene was assessed by performing the analysis on 500 random permutations of the data, generating neutral distributions for these two quantities. Significance of σ^2_{GB} estimates was used to evaluate the significance of Q_{st} for genes with significant heritability estimates. Indeed, permutation test on Q_{st} were much less accurate, generating hundreds of false positives with non-significant σ^2_{GB} and/or σ^2_{GW} but very high and significant Q_{st} estimates.

4.4.6 Functional classification

Functional classification and assessment of significant differential representation of functional classes between subgroups of cDNA clones and all clones analysed were performed in the DAVID/EASE environment (http://david.niaid.nih.gov/david/). DAVID 2.1 (beta version) gene accession conversion tool was first used to convert gene ontology-linked identifications of various types gathered in the GRASP 16 000-gene microarray gene identification file to UNIGEN clusters. Assessment of significant differential representation of functional classes between subgroups of cDNA clones representing genes with significant transcription profile heritability or genes showing significant differences between subpopulations, and the group of all analysed cDNA clones was performed with EASE 2.0.

4.4.7 Transcriptome scan and neutrality test

The "transcriptome scan" was performed by obtaining the distribution of individual transcription level Q_{st} values for all detected cDNA clones corresponding to genes with significant transcriptional heritability and by identifying those in the upper (1.5%) tail of the distribution as those for which transcription was potentially under directional selection. Additionally, we performed Lande's neutrality test (Lande 1976, 1977; Koskinen *et al*. 2003) on gene transcription data for all detected genes with significant transcriptional heritability. The null-hypothesis of evolution by random drift was tested as follows: $F = (Ne\ \sigma^2_{GB})/(h^2\ \sigma^2_{GW}\ t)$. The numerator and denominator degrees of freedom are the number of populations compared minus one and infinity, respectively. *Ne* was estimated from previously published microsatellite data (Garant *et al.* 2000) using the NeEstimator software and the linkage desequilibrium method (Peel *et al.* 2004).

For both the Q_{st} scan and neutrality test, we obtained more realistic results when considering only genes with significant heritability estimates. This criterion makes sense from a theoretical point-of-view since genes with non-significant transcriptional h^2 or intra-populational genetic variance of transcription levels are not theoretically expected to respond rapidly to selection and evolve different expression levels (in a few generations only). We considered the complete exhaustion of genetic variance for gene expression by

selection unlikely in six generations. Some genes with non-significant transcriptional h^2 may have high transcriptional Q_{st} estimates as a result of leftover environmental variance (though this would mainly come from environmental maternal effects; we included a term for maternal effects in the animal model we used), however we were mainly interested here in identifying genes for which transcription evolved as a result of directional selection. Perhaps even more importantly, using only genes with significant transcriptional h^2 allowed us to correct for artefacts resulting from very small σ^2_{GW} estimates. Hence, both the Q_{st} and Lande's neutrality test equations contain a h^2 and/or σ^2_{GW} term in their denominator. As shown in figure 4.2, h^2 estimates (and hence σ^2_{GW}) are virtually null for a majority of the 6484 detected genes. For Q_{st}, considering the genes with nearly-null transcriptional heritability artefactually generates very high estimates, even in cases where σ^2_{GW} is not significant. For Lande's test, this generates F statistics several orders of magnitude above the threshold F value for significance for a majority of the detected genes where there should in fact be an indeterminate form ($F = (Ne \times \sigma^2_{GW}) / 0$). Finally, there is an increasing number of reports about the predominance of non-additive interactions at the transcriptome levels (Gibson *et al*. 2004, Auger *et al*. 2005, Roberge *et al.* 2008). Estimation of Q_{st} for genes under non-additive control would not be valid since the Q_{st} framework is based on the premise of additivity. Of course, such genes are also more likely to have low or non-significant heritability values. Hence, pre-selecting genes with significant transcriptional h^2 minimized the impacts of several problems and possible artefacts on our results and interpretations.

4.5 Results

4.5.1 Hybridisation design

The design for the experimental microarray hybridisation differed for the two sampling years. In 2003, we were mainly interested in assessing the functional genomic basis of sexual precocity (Roberge *et al.*, unpublished data), such that the progeny of one anadromous male and one precocious parr from the same subpopulation and sharing the same mother were always coupled on a same microarray, for a total of 24 microarrays and

48 offspring analysed. For 2004, one offspring from each the upstream and downstream subpopulation were always coupled on a same microarray, for a total of 21 microarrays and 42 offspring analysed. Thus, transcription profiles were obtained for a total of 90 offspring representing 50 halfsib families.

4.5.2 Normality and uni/multi-modality assessment

Pearson's correlation tests between normalised gene transcription data and a normal distribution revealed a highly significant correlation to a normal distribution for all detected genes (the highest observed P-value being 1.57×10^{-8}), with Pearson's determination coefficient generally very close to one (average 0.97, figure 4.1). Results from the Shapiro-Wilk test were less categorical: 10% of the genes showed a significant departure from normality, even when using a Bonferroni-corrected significance threshold of 7.7×10^{-6}. The distributions of the transcription profiles for the three genes that showed the most significant departures from normality according to the Shapiro-Wilk test all included outlier data points (Suppl. Fig. 4.1), suggesting that the apparent departure from normality was the result of the test's high sensitivity to the presence of outliers. Fig. 4.1 A shows the relationship between the log of the P-value from the Shapiro-Wilk test and Pearson's determination coefficients for all detected genes. While normality does not have to be assumed for performing the permutation-based ANOVA, it is assumed in the procedure we used for assessing genetic parameters. However, restricted maximum likelihood estimators are robust to deviations from the assumption of normality (Kruuk 2004), and therefore we believe that the observed departure from normality for a minority of genes (in the Shapiro-Wilk test only) should not significantly impact our main conclusions. The distribution of the dip statistic from Hartigan's Dip test for the 6484 detected genes is presented in Fig. 4.1 B. While the transcripts from the right tail of the distribution in Fig. 4.1 B show expression patterns across all samples which suggest multi-modality (four examples are presented in Fig. 4.1 C), only 175 transcripts (including those presented in Fig. 4.1 C) showed statistical evidence of multimodality ($P < 0.05$, $dip > 0.054$), while 324 (0.05×6484) could have been expected by chance alone. Thus, there was no strong evidence for the occurrence of bi- or multi-modality of transcription profiles in our data set.

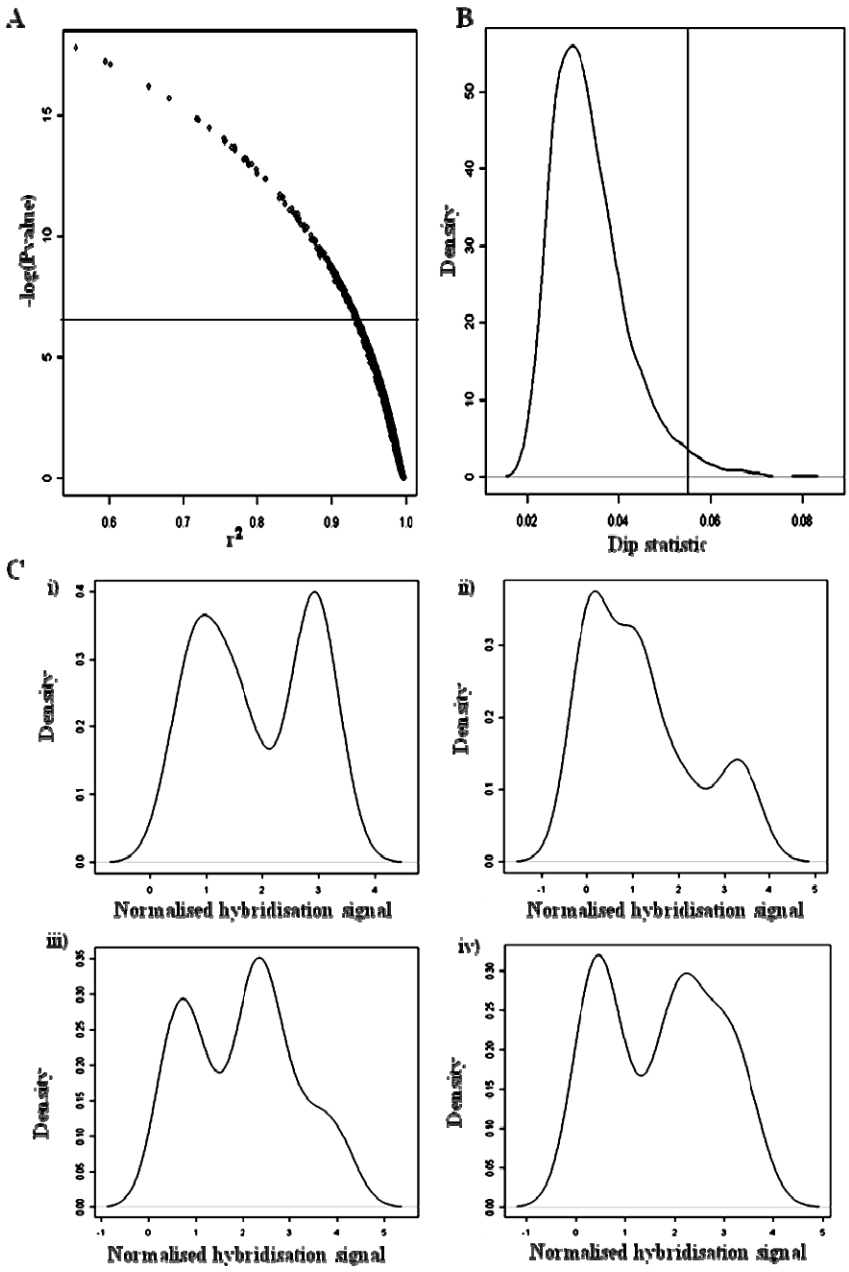


Figure 4.1 Assessment of normality and unimodality.

A) Relationship between Pearson's determination coefficient between a normal distribution and the observed distribution of normalized gene transcription data on the one hand, and the inverse of the log of the P-value from the Shapiro-Wilk normality test of gene transcription data on the other hand, for each of the 6484 cDNA clones corresponding to genes for which expression was detected. The horizontal line indicates the position of the Bonferroni-corrected significance threshold for the Shapiro-Wilk test. B) Density curve of Hartigan's Dip statistic testing for unimodality of the gene transcription data for all detected genes. Gene transcription data from genes at the right of the vertical line (Dip > 0.054, $P < 0.05$) show significant departure from unimodality. **C)** Density curves of the normalised gene transcription data of the 90 sampled individuals for the 4 genes with the highest Dip statistic. The corresponding gene products are: i) an unknown gene product (GeneBank accession number of the cDNA sequence: CA059553, ii) ribosomal protein L35, iii) ubiquitin, iv) α-actin 2.

4.5.3 Heritability estimates, Q_{st} estimation and transcriptome scan

Figure 4.2 presents the distribution of heritability estimates of the normalised transcription profiles for the genes corresponding to the 6484 cDNA clones analysed and their associated standard error. Significant ($P \leq 0.05$) transcriptional h^2 estimates were obtained for 1044 genes (16% of all detected genes) . Transcriptional h^2 estimates for these varied between 0.121 and 0.996 , and averaged 0.409. A test conducted in EASE 2.0 revealed significant over-representation of genes in several gene ontology categories (9 with EASE-score < 1 X 10^{-3}) within the genes with significant transcriptional h^2 when compared to all genes on the microarray. Most of these categories were related to oxidative phosphorylation, including for instance primary active transporter activity (GO molecular function, EASE score = 1.6 X 10^{-4}) and hydrogen ion transporter activity (GO molecular function, EASE score = 7.3 X 10^{-4}).

Transcriptional Q_{st} were only estimated for the 1044 cDNA clones corresponding to genes with significant transcriptional h^2 for reasons explained previously. Most of those genes had low transcriptional Q_{st} estimates that are unlikely to differ from 0 (Fig. 4.3). The 1.5% upper outliers (16 cDNA clones) of this distribution (Table 4.1) are genes for which transcription levels are the most likely to have diverged between the two subpopulations as a result of directional selection. Transcriptional Q_{st} estimates for these outliers ranged from 0.07 (tartrate-resistant acid phosphatase type 5) to 0.19 ("unknown" gene), and averaged 0.11. The identity of ten of these outliers was unknown, meaning that they did not generate any BLAST hits with e-values less than 1 X 10^{-15} and an informative name (for a detailed description of the array annotation process, see http://web.uvic.ca/cbr/grasp). The remaining six outlier cDNA clones represented genes with various functions (Table 4.1). No functional group appeared significantly over-represented within these genes when compared to all genes represented on the microarray. It is noteworthy that all of these outliers were over-transcribed in the upstream sub-population (average = 29% over-expression relative to the downstream subpopulation). The over-transcription ranged from 16% (unknown gene CB513854) to 46% (nucleolar RNA helicase II). Gene transcription h^2 among these genes ranged from 0.14 (transmembrane protein 14C) to 0.44 (pterin-4-alpha-

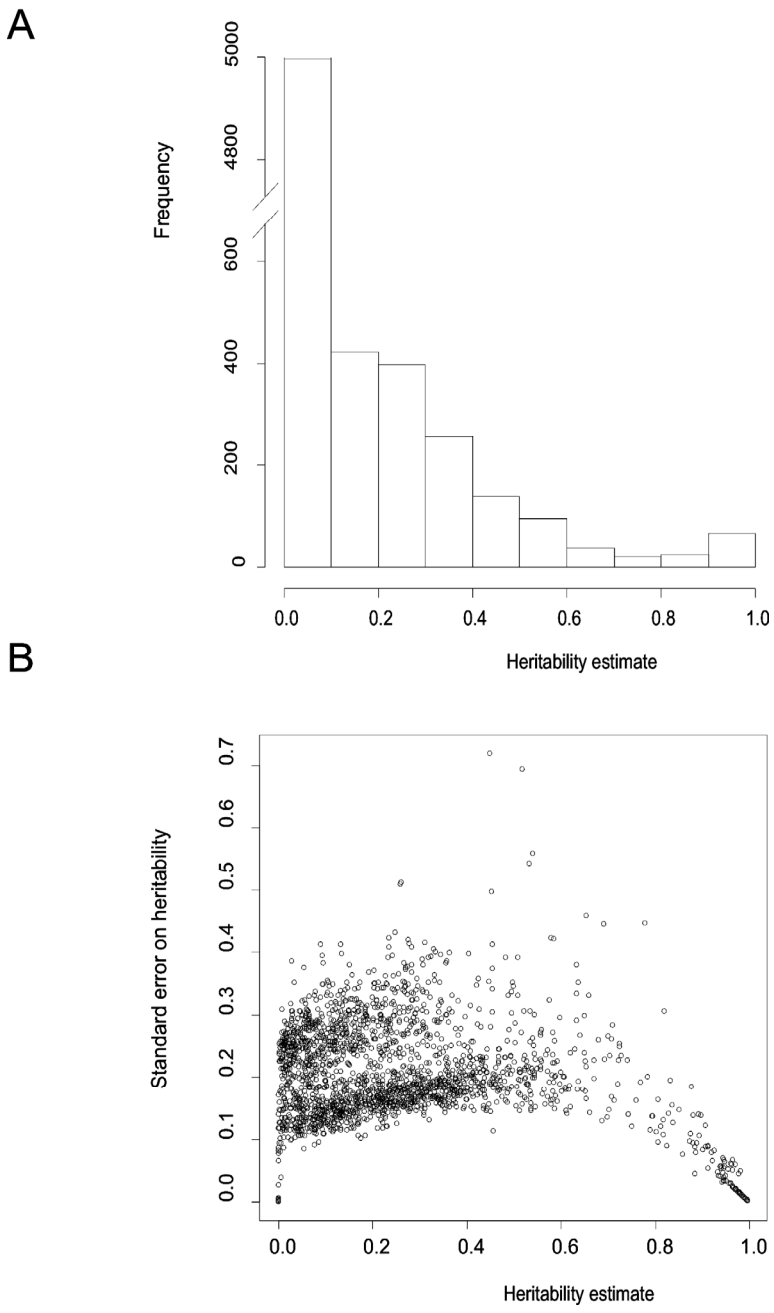


Figure 4.2 Gene transcription heritability.

A) Distribution of heritability estimates of the normalised transcription profiles for the 6484 cDNA clones corresponding to genes which expression was detected B) relationship between h^2 and its standard error for each of the 6484 detected cDNA clones.

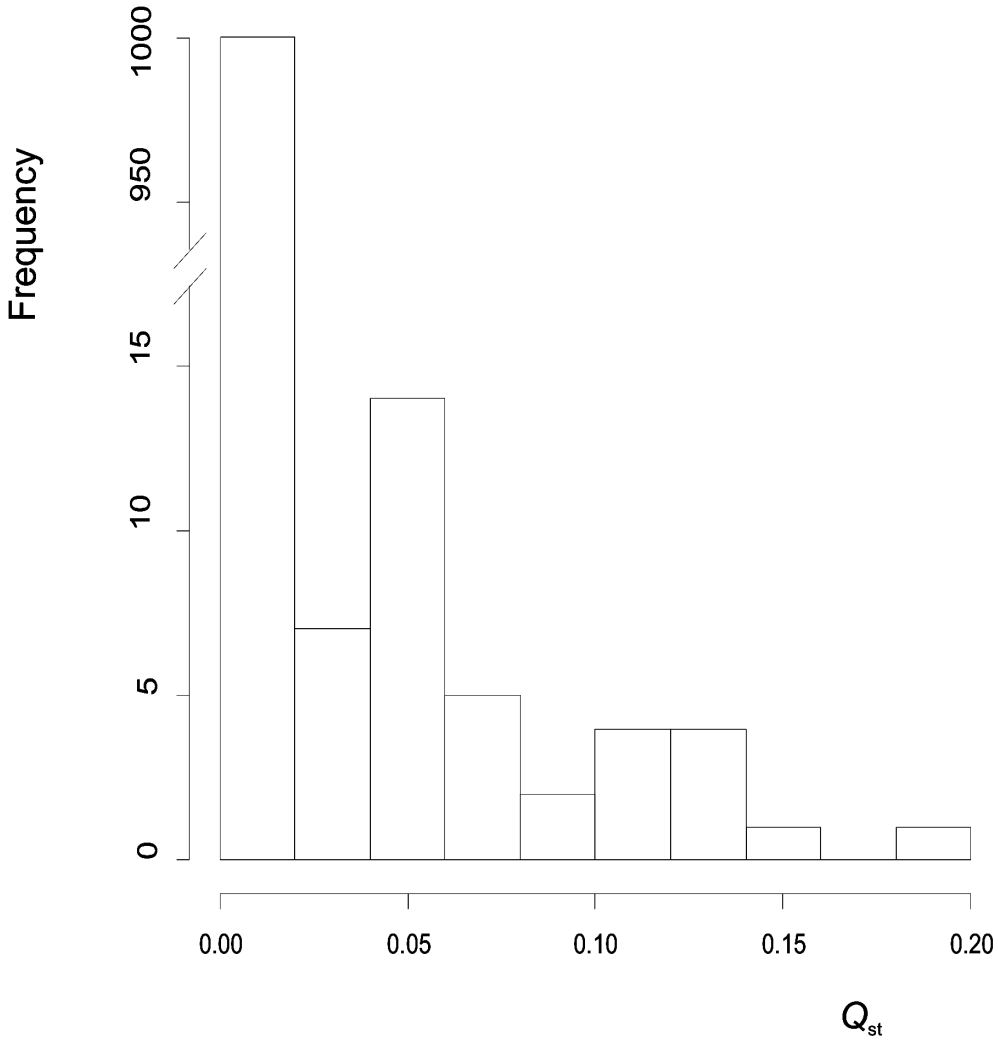


Figure 4.3 Gene transcription Q_{st}.

Distribution of Q_{st} estimates of the normalised transcription profiles of the 1044 cDNA clones representing genes with significant gene transcription heritability ($P \leq 0.05$)

carbinolamine dehydratase), and averaged 0.28. The inter-population component of additive genetic variance (σ^2_{GB}) was significant for all 16 outliers ($P < 0.05$). We used significance of the σ^2_{GB} estimates to test for significance of Q_{st} estimates since the permutation test on Q_{st} estimates was prone to false positives (see methods).

Lande's test, which we also performed on all cDNA clones corresponding to the 1044 genes with significant transcriptional h^2, provided evidence of non-neutral evolution for the gene transcription levels of 24 of these genes only, which nevertheless included 15 of the 16 genes identified by the Q_{st} scan (Table 4.1). The nine remaining genes were also all over-transcribed in the upstream sub-population. Five of these had no known function, while the remaining four were a salmonid toxin-like gene (toxin-1, CB493361), P0-like glycoprotein, myc-associated factor X and glutamyl-prolyl-tRNA synthetase.

4.5.4 ANOVA results on gene expression divergence between subpopulations

In parallel to our quantitative genetic approach, we also tested the significance of gene transcription differences between subpopulations in a simpler linear model where the pedigree information was not considered, as usually done in other studies. The simulation-based ANOVA revealed highly significant ($P < 0.0001$) differences in gene transcription between the progeny of fish from the two sub-populations for 12 cDNA clones (Table 4.2). The magnitude of the observed significant changes varied between a 30% under-transcription (thioredoxin-like protein p19 coding gene) to a 66% over-transcription (creatine kinase coding gene) in the upstream sub-population (average fold difference = 33%). Genes that differed significantly in expression between the two subpopulations were involved in several molecular and biological functions, including energy metabolism and transcription (Table 4.2). Moreover, a test conducted in EASE 2.0 revealed a highly significant (EASE-score = 8.14×10^{-3}) over-representation of genes of the immune response gene ontology category (biological process) among the genes significantly differentially transcribed at $P < 0.005$ when compared to all genes represented on the microarray. Notably, four different clones representing major histocompatibility class II-

Table 4.1 Results from the transcriptome scan. cDNA clones corresponding to the 16 genes for which transcription level Q_{st} estimates were in the upper 1.5 % of the Q_{st} distribution for 1044 cDNA clones representing genes with significant transcription level heritability.

Gene product	GeneBank ID	h^2	Q_{st}	P-value h^2	P-value σ^2_{GB}	Fold difference	F neutrality test	Functional information
Transmembrane protein 14C	CA042090	0.14	0.11	3.8×10^{-2}	1.0×10^{-2}	1.21	18.94*	Unknown function
Selenoprotein P, 1b	CB509685	0.21	0.11	4.8×10^{-2}	8.0×10^{-3}	1.31	12.28*	Oxidant defense, thyroid hormone metabolism, defense against viral infections
Nucleolar RNA helicase II	CB505664	0.32	0.08	2.0×10^{-2}	0	1.46	6.03*	Ribosomal RNA processing, transcription regulation
Pterin-4-alpha-carbinolamine dehydratase	CB497855	0.44	0.07	8.0×10^{-3}	2.0×10^{-3}	1.19	3.80 (NS)	Tetrahydrobiopter in recycling and biosynthesis, HNF-1-alpha regulation
Lysozyme type II	CA037907	0.28	0.07	2.6×10^{-2}	1.8×10^{-2}	1.29	5.76*	Bacteriolytic function
Tartrate-resistant acid phosphatase type 5	CA045149	0.37	0.07	1.2×10^{-2}	2.0×10^{-3}	1.39	4.22*	Cytochemical marker of macrophages, osteoclasts and dendritic cells; unknow biological function
UNKNOWN	CB496434	0.22	0.19	3.8×10^{-2}	0	1.40	22.40*	
UNKNOWN	CA060749	0.23	0.16	5.0×10^{-2}	0	1.31	17.34*	
UNKNOWN	CB514770	0.38	0.14	1.0×10^{-2}	0	1.18	9.30*	
UNKNOWN	CB514561	0.27	0.13	2.4×10^{-2}	0	1.25	12.39*	
UNKNOWN	CB513854	0.20	0.13	2.6×10^{-2}	6.0×10^{-3}	1.16	16.42*	
UNKNOWN	CA060904	0.21	0.13	4.6×10^{-2}	6.0×10^{-3}	1.26	15.59*	
UNKNOWN	CB517440	0.44	0.11	1.8×10^{-2}	0	1.35	6.13*	
UNKNOWN	CA056630	0.23	0.11	3.2×10^{-2}	4.0×10^{-3}	1.30	11.03*	
UNKNOWN	CA041889	0.31	0.07	3.2×10^{-2}	1.6×10^{-2}	1.25	5.18*	
UNKNOWN	CA056760	0.23	0.10	3.8×10^{-2}	6.0×10^{-3}	1.37	9.88*	

Corresponding Q_{st} and heritability estimates (with their P-values from the permutation test), F statistic from Lande's neutrality test (stars indicate significance at the $P < 0.05$ threshold), fold difference between subpopulations (upstream/downstream), GeneBank accession number and functional information from the literature are presented.

associated invariant chains showed consistent under-transcription (average 17%) in the progeny of the upstream sub-population. A nuclease sensitive element binding protein 1 (Y-box-binding protein 1) coding gene was also significantly under-transcribed in the upstream sub-population. It is noteworthy that this gene was originally identified as coding for a DNA-binding protein that regulates MHC class II genes (Didier *et al*. 1988).

4.6 Discussion

Here, we applied a "transcriptome scan" approach combining both the Q_{st} and genome scan frameworks to investigate the role of selection in shaping differential profiles of gene transcription between recently diverging (six generations) subpopulations of Atlantic salmon. The additive genetic basis of gene transcription was estimated for genes corresponding to all detected cDNA clones (6484) and gene transcription heritability estimates were significant for 1044 (16%) of these. The scan of Q_{st} values estimated from intra- and inter-population additive genetic components of the transcription profiles identified 16 outlier cDNA clones (average Q_{st} estimate = 0.11) representing the most likely genes for which transcription level have evolved under the influence of directional selection within only approximately six generations. Additional evidence of non-neutral evolution of the transcription levels for 15 of those 16 genes was obtained by performing Lande's neutrality test on gene transcription data. Overall, this study shows that the transcriptome scan approach can be used to identify genes which transcription profiles are likely to have evolved as a result of directional selection, even at small temporal and spatial scales.

4.6.1 Gene transcription normality and multimodality

When assessing the genetic parameters of gene transcription profiles taken as phenotypic traits, the question of the normality of the data becomes more than a mere to-use-or-not-to-use-parametric-statistics concern: it may reflect either the qualitative or quantitative nature

Table 4.2 Results from the simulation-based ANOVA.

Gene product	P-value	Fold difference	GeneBank accession	Functional information
Creatine kinase (EC 2.7.3.2)	9.87E-06	1.66	CK990405	Energy metabolism
Thioredoxin-like protein p19	9.87E-06	0.70	CB498161	Redox regulation, protein refolding, transcription factors regulation
Alpha-globin and beta-globin	8.00E-05	1.48	CA051720	Oxygen transport
Phosphoglycerate mutase 2 (EC 5.4.2.1)	2.16E-04	1.47	CB497792	Regulation of glycolysis, cell proliferation
Invariant chain INVX	2.65E-04	0.85	CB503772	MHC class II-associated invariant chain
Invariant chain INVX	2.91E-04	0.87	CK990275	MHC class II-associated invariant chain
NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3)	3.27E-04	1.34	CN442556	Energy metabolism
Nuclease sensitive element binding protein 1	3.28E-04	0.82	CB511419	Transcription factor
Invariant chain S25-7	4.16E-04	0.86	CB513162	MHC class II-associated invariant chain
Invariant chain S25-7	5.04E-04	0.75	CB502487	MHC class II-associated invariant chain
Nucleolar RNA helicase II	5.64E-04	1.46	CB505664	Transcription
Acidic mammalian chitinase (EC 3.2.1.14)	7.32E-04	1.44	CB505509	Chitin catabolism

cDNA clones representing genes with significant ($P < 0.0001$) differences in transcription level between the upstream and downstream subpopulations. The corresponding gene identities, P-values, average fold difference (upstream/downstream) and Genebank accessions are provided for each significant cDNA clone, along with brief functional annotations from the literature.

of the traits. Hence, Gibson and Weir (2005) noted that in expression QTL mapping studies, eQTLs accounting for 25%–50% of transcriptional variation prevail. They suggested that since major-effect eQTLs are common, gene transcription data should often depart from a normal distribution and exhibit bi- or multi-modal distributions, which would make the genetic structure of gene transcription profiles more akin to qualitative than to quantitative phenotypic traits. Yet, very few studies have tested the normality of gene transcription data on a gene per gene basis (but see Giles and Kipling 2003) and to our knowledge, none have examined the modality of such data. Here, we observed strong and highly significant correlations between gene transcription data and the normal distribution, and that for all detected genes (cDNA clones with low intensity signal were filtered prior to this analysis). However, results from the Shapiro-Wilk test suggest that the transcription data of 10% of the detected genes departed significantly from normality. The Shapiro-Wilk test appears very sensitive to outliers (Fig S4.1), which might increase the number of false positives. Moreover, Giles and Kipling (2003) argued that even with small sample sizes, the Shapiro-Wilk normality test has the power to detect very slight departures from normality and score them as significant. We also tested the evidence of uni versus bi- or multi- modality in the gene transcription data distributions of each of the 6484 detected cDNA clone. We found little evidence for bi- or multi-modality since only 175 clones showed significant departure from unimodality while 324 were expected by chance alone. Yet, we cannot rule out the possibility that some of these 175 clones represent true positives (genes with bi- or multi-modal transcription level distribution, Fig 4.1C). In summary, there was little evidence of significant departure from normality in transcription profiles, with the vast majority of the genes showing unimodal distribution. This suggests that most transcription profiles behaved as quantitative rather than qualitative traits in this system. Hence, the high prevalence of major effect eQTL in eQTL mapping studies (Gibson and Weir 2005) may correspond more to a detection bias towards major effect eQTL than to a biological fact. However, of the minority of transcripts showing significant departure from either normality or unimodality (Fig 4.1C), some might yet correspond to genes which transcription depends on major effect eQTLs.

4.6.2 Gene transcription heritability

Only a handful of studies have formally estimated the genome-wide heritability of gene transcription profiles, and these concerned only model species (mice: Schadt *et al*. 2003, Cui *et al*. 2006, Chesler *et al*. 2005, human: Monks *et al*. 2004, yeast: Brem *et al*. 2002). Results were quite diverse as the median h^2 among genes with significantly heritable transcription profiles ranged from 0.11 (Chesler *et al*. 2005) to 0.84 (Brem *et al*. 2002), whereas the average h^2 among genes with significantly heritable transcription profiles in the present study was of 0.41. When indicated, the proportion of genes with significant gene transcription h^2 in previous studies was around 30% (Schadt *et al*. 2003, Monks *et al*. 2004), while it was 16% here. This proportion varies with the statistical power to detect truly significant h^2 and is therefore not formally comparable between studies. Neither are the average or median gene transcription h^2 among genes with significant transcriptional h^2, since they are likely to be inflated when the power to detect small h^2 estimates as significant is low. Heritability estimation is also highly model-dependant. To illustrate this point, we re-analysed the same data with a different animal model that did not take into account the sampling year and maternal effects (Supplementary Figure 4.2). The proportion of h^2 estimates detected as significant rose to 39% (2543), which shows the importance of choosing an appropriate model (accounting, notably, for maternal effects) to reduce biases in estimating h^2 (Kruuk 2004).

Here, the statistical power to detect significant transcriptional h^2 was relatively low given the low genetic variance for gene expression in our system. Hence, the crosses we used were performed among genitors of a same subpopulation (one of two small subpopulations). Clearly, then, the additive genetic and phenotypic variances cannot be expected to be as high here as in systems where crosses were performed between genitors of different strains or populations. This low power likely caused the average h^2 estimates among genes with significant transcription heritability to be higher than in some previous studies assessing gene transcription h^2 (Monks *et al*. 2004, Chesler *et al*. 2005). Higher average h^2 estimates may also reflect genuine differences between species (salmon, mouse and human) or result from the different statistical models used. Though our sample size was small compared to those commonly used when quantifying heritability for "classical"

phenotypic traits, it still ranks among the highest used to date in any study that investigated transcriptional heritability.

While the 1044 genes with significant gene transcription h^2 do not necessarily represent genes for which directional selection was evidenced in this study, they likely represent genes for which levels of transcription have a high potential to respond to selection. Functional analysis showed significant over-representation of genes categories related to oxidative phosphorylation among this group of genes when compared to all genes represented on the chip. These categories were not significantly over-represented in the list of genes that showed significant differences in gene transcription between subpopulations (based on ANOVA), nor among the trancriptional Q_{st} distribution upper outliers. This might suggest that genes from this basic pathway, though their transcription profiles often appear heritable, were not particularly involved in the recent phenotypic divergence between the salmon sub-populations we studied.

4.6.3 Q_{st} estimates and transcriptome scan

As illustrated from the distribution of transcription profiles Q_{st} (Fig. 4.3), transcriptional Q_{st} were virtually null for the great majority of the genes tested. While Q_{st} were expected to be low given that divergence between the two subpopulations under study has only been possible since 1981 (about six generations), transcription level Q_{st} for 97% of the genes with significant transcriptional h^2 were under (in some cases by orders of magnitude) the F_{st} values (average = 3.2%) estimated in a previous study based on neutral markers (Garant *et al*. 2000). This observation is consistent with the results of recent studies (Denver *et al*. 2005, Rifkin *et al*. 2005) in which the authors concluded that transcription profiles of a large proportion of all genes appear to be under the effect of stabilizing selection in natural systems. However, Beaumont and Balding (2004) recently used a simulation-based approach to show that current statistical methods could not accurately identify loci under balancing or stabilising selection in a genome scan framework, especially when populations are weakly differentiated. Yet, results from Lande's neutrality test suggest that for 1020 (97.7%) of the genes tested, neutral evolution of gene transcription profiles cannot be

rejected. An alternative explanation for the high number of very small Q_{st} estimates in this study could be the lack of genetic variance for gene transcription profiles (Merila et Crnokrak 2001), as discussed above.

Conversely, by targeting genes in the upper outliers of the transcriptional Q_{st} distribution, the "transcriptome scan" approach identified 16 cDNA clones (Table 4.1) representing the most likely genes for which transcription levels may have diverged between the two subpopulations as a result of directional selection in two contrasting habitats. Additional evidence for the role of selection in driving transcriptional divergence for 15 of those 16 genes (plus 9 others) comes from the results of Lande's neutrality test (Table 4.1). Transcriptional Q_{st} estimates for these genes were high (0.07-0.19, mean = 0.11), given the short time frame in which divergence is likely to have occurred between the two sub-populations. Transcriptional Q_{st} estimates for these genes were also all above the F_{st} estimated between the sub-populations using neutral markers. It is noteworthy that 62% of the genes identified as outliers have no known function, which suggests that selection acted on unpredictable targets. The other candidate genes identified with the transcriptome scan represented various functional groups (Table 4.1). Moreover, the differential transcription of some of these genes could potentially affect gene expression of other genes at different levels. Hence, the nucleolar RNA helicase II is a multifunctional protein that is notably implicated in rRNA biosynthesis and in the regulation of c-Jun-mediated gene expression (Yang *et al*. 2005). Pterin-4-alpha-carbinolamine dehydratase, apart from its role in tetrahydrobiopterin biosynthesis, is also known as a dimerisation factor of the hepatocyte nuclear factor (HNF) 1 α, one of the master regulators of hepatocyte and pancreatic islets transcription (Resibois *et al*. 1999).

4.6.4 Between subpopulations ANOVA and comparison with the results of the transcriptome scan and gene transcription heritability

The between sub-populations ANOVA identified 12 cDNA clones corresponding to genes for which the transcription level diverged between sub-populations, albeit apparently not necessarily through the effect of selection (Table 4.2). Except for a nucleolar RNA helicase

II-coding gene, none of these genes overlapped with outliers identified by the transcriptome scan approach and the null hypothesis of neutral evolution could not be rejected for any of them by Lande's test. Namely, these differences could be the result of six generations of phenotypic divergence in different parts of the Sainte-Marguerite River caused by genetic drift. Yet they still represent genes that potentially affect differentially several physiological functions between the two sub-populations. Of particular interest is the gene coding for nucleolar RNA helicase II , which was both significantly differently transcribed under a conservative significance threshold and exhibited substantial genetic variance in gene transcription between relative to within subpopulation (Table 4.1 and 4.2)**.**

The fact that except for one gene, there was no overlap between the results of the transcriptome scan and those of the ANOVA may seem paradoxal and therefore deserves explanations. First, the absence of most of the transcriptome scan candidate genes from the list of genes showing significant transcription differences between subpopulations (Table 4.2) is largely explained by the stringency of the significance threshold chosen for the ANOVA. Had we raised this threshold to 0.05, differences in transcription for most genes in table 4.1 would have been significant in the ANOVA analysis. Moreover, heritability estimates for eight of the genes harbouring the most significant differences in gene transcription between subpopulations were not significant. Hence, these were not considered in the Q_{st} analysis and could therefore not be detected as outliers in the transcriptome scan. With the exception of nucleolar RNA helicase II, the remaining genes in table 4.2 (Y-box transcription factor, invariant chain INVX and invariant chain S25-7) had significant gene transcription h^2 and showed highly significantly different expression profiles between subpopulations. Yet, they had small transcriptional Q_{st} and hence were not identified as outliers in the transcriptome scan. We propose two interpretations for this result. Either the transcription profiles of these genes evolved neutrally between subpopulations (e.g. by genetic drift, which might be supported by the results of Lande's test) or they evolved as a consequence of selection but their transcription levels have a non-additive genetic basis. Indeed, recent studies have provided clear evidence that a non-additive genetic basis for gene transcription is common

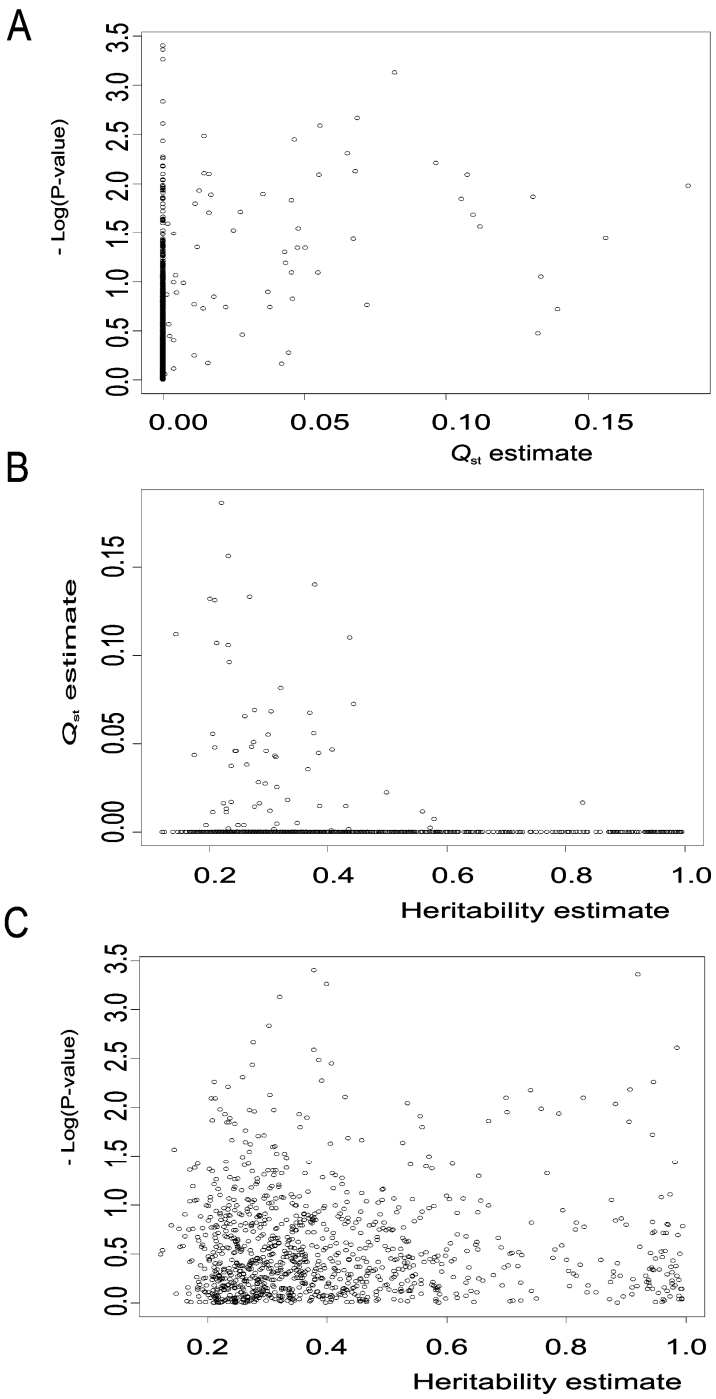


Figure 4.4 Relationships between gene transcription Qst, heritability and the significance from the ANOVA for the 1044 genes with significant heritability estimates.

Scatterplots illustrating the relationships between (A) gene transcription Qst estimates and the inverse of the log of the P-values from an ANOVA testing for significant gene transcription level differences between the progeny of fish from the two subpopulations, (B) gene transcription Qst and heritability estimates and (C) gene transcription heritability estimates and the inverse of the log of the P-values from the above-mentioned ANOVA.

(Gibson *et al*. 2004, Auger *et al*. 2005). Here, a possible example of the latter case may be given by genes related to the major histocompatibility complex (MHC), three of which had significant transcription levels h^2 but very low estimated Q_{st}. Interestingly, (Landry et Bernatchez 2001) showed that directional selection has been acting at the antigen presentation sites of MHC class II B locus on the subpopulations of this same river. The coding sequence of MHC class II B coding genes could have evolved under the effect of selection while the transcription levels of other MHC-related genes evolved neutrally. Alternatively, both could have evolved through selection on traits harbouring non-additive genetic variance. Another general observation on the comparison of the results of the ANOVA and transcriptome scan is that all of the upper outlier genes of the transcriptional Q_{st} distribution were over-transcribed in the upstream versus downstream population (Table 4.1, average fold-change: 29%). This was however not observed for the genes that showed significant transcription differences between the two subpopulations in the ANOVA (Table 4.2), in which a similar number of genes were over- and under-transcribed for a given subpopulation. This raises the hypothesis that selection may have favoured enhanced general metabolism and gene transcription in the upstream subpopulation, which is also characterized by higher growth rate and smaller size threshold at precocious sexual maturation, resulting in a higher proportion of sexually precocious parr relative to the downstream subpopulation (Aubin-Horth *et al*. 2006). Gene over-expression associated with higher metabolic demands have been observed in other salmonids. For instance, in a study on the transcriptomics of population divergence involving dwarf and normal whitefish species pairs (*Coregonus* sp.), Derome *et al*. (2006) observed that higher swimming activity in the dwarf whitefish was consistently associated with over-expression of genes involved in muscle contraction as well as energy metabolism.

No correlation was observed between the P-values from the ANOVA comparing the subpopulations, the estimated gene transcription h^2 and the transcriptional Q_{st} estimate for the 1044 genes with significant gene transcription h^2 (Fig. 4.4), nor for all detected genes (not shown). However, the relationship between transcriptional Q_{st} and h^2 estimates was marked by a mutual exclusion: few genes had both high transcriptional high Q_{st} and h^2. This may be at least partly explained by the fact that the animal genetic variance, which we

used to estimate σ^2_{GW} , is the numerator of h^2 formula and also contributes to the denominator of the Q_{st} formula. High within population genetic variance naturally imposes a superior limit to values that Q_{st} can reach (see Hedrick 2005) for analogous arguments regarding F_{st} estimation). Inversely, very small between subpopulation genetic variance can result in very high Q_{st} estimates for genes with very low within population genetic variance, which could represent a technical artefact inflating Q_{st} estimates for a number of genes. Here, this potential problem was controlled for by considering only genes with significant h^2 estimates for the transcriptome scan. Yet, one could argue that genes with high gene transcription Q_{st} estimates and very low genetic variance estimates illustrate the erosion of the within-subpopulation genetic variance as a result of the action of divergent selection. Although this possibility cannot be ruled out, we do not deem this is very likely considering the short timeframe (six generations) involved in the divergence of the salmon subpopulations studied here.

To conclude, this study represents to our knowledge the first attempt to translate differences in gene transcription into Q_{st} estimates in order to identify genes for which transcription levels are potentially under directional selection and one of the few studies to evaluate transcription level h^2 and its error on a genome-wide scale. It thus contributes both empirically and methodologically to the evolutionary quantitative genetic exploration of transcription data by showing that a combination of Q_{st} and genome scan framework can efficiently identify genes for which transcription may have evolved under the effect of directional selection, even at the scale of very recently diverging subpopulations.

4.7 Acknowledgments

This work was supported by a research grant from Natural Science and Engineering Research Council (NSERC strategic program) to L.B., H.G. and J.J. Dodson, and a scholarship from the *Fonds Québécois de la Recherche sur la Nature et les Technologies* (FQRNT) to C.R. This study is a contribution to the research program of the *Centre Inter-universitaire de Recherche sur le Saumon Atlantique* (CIRSA). We thank François Martin,

David Páez, Orlane Rossignol, Serge Higgins and the staff at the *LAboratoire de Recherche en Sciences Aquatiques* (LARSA), Université Laval, Québec for their assistance in fish rearing and sampling. We are also grateful to Stéphane Larose for Perl scripting, Aurélie Labbe for help with R scripting and Guy Perry for help with PEST/VCE. We also acknowledge the constructive comments of G. Perry, M. Veuille and three anonymous reviewers on a previous version of the manuscript. Finally, we thank the staff at the provincial fish hatchery in Tadoussac for fish crossing and rearing.

Chapitre 5 : Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon

5.1 Résumé

Les saprolegniacées pathogènes sont parmi les plus importantes causes de maladie chez les salmonidés d'élevage et chez les poissons d'eau douce en général. Des études récentes ont utilisé des bio-puces afin d'identifier les acteurs potentiels de la défense immunitaire des poissons face à divers virus et bactéries, mais la réponse des poissons à des pathogènes de type fongique est peu documentée. Dans cette étude, nous avons utilisé une bio-puce de salmonidés pour identifier des gènes dont le niveau de transcription est modifié chez des saumons atlantiques juvéniles atteints de saprolegniose. Nos résultats confirment l'importance de l'immunité non-spécifique dans la réponse des poissons aux infections par des saprolegniacées et suggèrent qu'il existe à la fois des différences et des similitudes entre les réponses immunitaires des poissons à des oomycètes et à des infections bactériennes ou virales. De plus, plusieurs gènes aux fonctions inconnues montraient une sur-transcription chez les poissons infectés; ces gènes représentent potentiellement des acteurs non encore caractérisés de l'immunité chez les poissons.

5.2 Abstract

Pathogenic saprolegniaceae species are amongst the major disease-causing agents in farmed salmonids and in freshwater fishes in general. Recent studies have used high-throughput cDNA-based methods to identify new potential actors of fish defence systems against various bacteria and viruses. However, the response of fish to fungal or fungus-like pathogens is still poorly documented. Here, we used a 16 006-gene salmonid cDNA microarray to identify genes for which transcription levels are modified in juvenile Atlantic salmon (*Salmo salar*) affected with saprolegniasis compared to healthy fish from the same families. Our results confirmed the importance of non-specific immunity in the response of fish to saprolegniaceae infections and identified both similarities and differences in their genome-wide transcriptional responses to oomycetes compared with their response to bacterial or viral infections. Moreover, several clones with no known homologues were shown to be over-transcribed in infected fish, which may represent as yet unidentified immune-relevant genes in fish.

5.3 Introduction

The fish farming industry as well as the interest for comparative and evolutionary immunology has motivated the edification of a substantial body of knowledge in fish immunology (Iwama and Nakanishi 1996). Hence, the activity of non-specific immunity, and in particular that of the alternative complement pathway (Yano 1996), was found to be extremely high in fish compared to that of mammals (Bayne and Gerwick 2001, Ellis 2001). Various high throughput methods, including the creation of EST libraries enriched in immune relevant transcripts (Nam *et al*. 2000, Kono *et al*. 2004, Kono et Sakai 2001, Savan and Sakai 2002, Kocabas *et al*. 2002), subtractive EST libraries (Tsoi *et al*. 2004, Bayne *et al*. 2001, Fujiki *et al*. 2003, Fujiki *et al*. 2001) and microarrays (Ewart *et al*. 2005, Meijer *et al*. 2005, Rise *et al*. 2004a, Kurobe *et al*. 2005), have recently been employed in attempts to identify new molecular actors of the immune response in various teleost fish and clarify host-pathogen interactions. Microarrays are particularly well suited for this task (Meijer *et al*. 2005) and have been employed to characterize fish responses to various bacteria (*Piscirickettsia salmonalis, Aeromonas salmonicida, Mycobacterium marinum*) and viruses (viral hemorrhagic septicemia virus, hirame rhabdovirus). However, no such study has assessed the overall gene expression response when fishes are exposed to fungi or fungus-like protists, though various lines of evidence suggest that fish, like humans, show distinct acute phase responses to viral, bacterial and fungal agents (Ellis 2001, Bayne *et al*. 2001).

The saprolegniaceae family comprises closely related water moulds ubiquitous in freshwater and usually associated with dead tissues and fish eggs (Hart and Reynolds 2002). This family belongs to the oomycetes, a group of fungus-like protists sharing similarities with fungi and brown algae. Other members of this group are well known plant and mammal pathogens (Kamoun 2003). Phylogenetically very different from eumycotan fungi, oomycetes likely evolved different mechanisms for interaction with plants and animals (Kamoun 2003, Torto-Alalibo *et al*. 2005). Furthermore, some plant pathogenic oomycetes suppress host defence (Kamoun 2003), a property that has been suggested to be shared by *Saprolegnia* (Alvarez *et al*. 1995, Alvarez *et al*. 1988).

Pathogenic saprolegniaceae species are known to cause saprolegniasis, a fish tegumentary mycosis characterized by visible patches of filamentous mycelium covering the epidermal tissues (Beakes *et al*. 1994). If untreated, saprolegniasis leads to death, presumably by heamodilution (Hatai and Hoshia 1994). Saprolegniasis is one of the most important freshwater fish diseases, both in nature and aquaculture, where it causes economically important losses worldwide (Tampieri *et al*. 2003, Hughes 1994). This problem has achieved new heights recently (Torto-Alalibo *et al*. 2005) since malachite green, a triarylmethane dye efficiently used to control saprolegniasis since the mid 1930' (Alderman 1994), was prohibited in most producer countries because of its suspected teratogenicity and toxicity (Meyer and Jorgenson 1983, Alderman and Polglase 1984). Pathogenic saprolegniacea can act either as secondary or as primary pathogens (Neish and Green 1977, Whisler 1996). Fish stressed by diverse factors including handling, other infections, mechanical damage, sexual maturity, temperature changes, poor hygiene or social interaction are more susceptible to develop saprolegniasis (Pickering 1994), and salmonids are known to be particularly vulnerable (Hughes 1994). Here, we used a 16 006-gene salmonid cDNA microarray to identify genes for which transcription levels are modified in juvenile Atlantic salmon (*Salmo salar*) affected with saprolegniasis compared with healthy fish.

5.4 Methods

5.4.1 Fish samples

Six months post-hatching, Atlantic salmon juveniles grown from eggs in controlled conditions were collected from freshwater laboratory holding tanks during a non-provoked episode of saprolegniasis. Five alevins from four distinct full-sib families showing clear symptoms of the disease were collected and frozen along with five healthy alevins from the same families. Saprolegniasis was confirmed by a fish veterinarian and hypha were observed both under optical and electron microscopy (fig. 5.1).

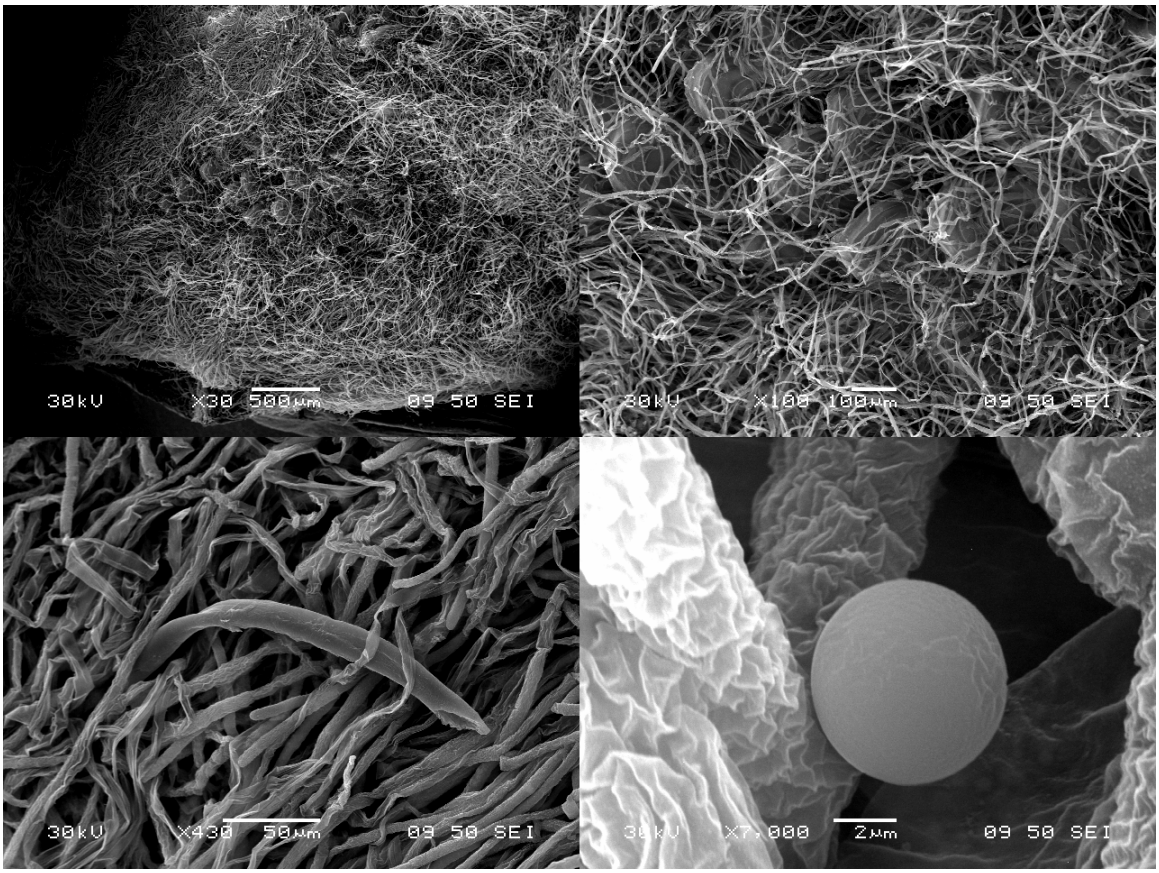


Figure 5.1 Scanning electron micrographs of the skin surface of a fish severely affected with saprolegniasis.

The top pictures are general views of the mycelium-infected epidermis. Bottom left is an empty zoosporangium among hypha and bottom right is a cyst. Scale, acceleration voltage and magnification are indicated on each image. Samples were fixed, dehydrated and gold stained prior to observation.

5.4.2 RNA extraction, labelling and cDNA hybridisation

Juvenile salmon that had been frozen at -80°C were homogenised individually in 10 ml TRIZOL@Reagent (Invitrogen), and total RNA was extracted in ten separate assays as previously described (Roberge *et al*. 2006). For each sample, 15 µg of the pooled RNA from the ten separate extractions was then retro-transcribed and labelled (Roberge *et al*., 2006). Transcription profiles of five healthy and five saprolegniasis-affected fish were contrasted on five microarrays, always comparing an infected fish to a healthy fish of the same family on a given array. Dye-sample coupling was flipped between biological replicates. The cDNA microarrays used here were obtained through the Genomic Research on Atlantic Salmon Project (GRASP, available from Ben F. Koop, bkoop@uvic.ca), which comprises 16 006 salmonid cDNA clones (von Schalburg *et al*. 2005).

5.4.3 Signal detection, data preparation and statistical analysis

Signal detection and data preparation was as previously reported (Roberge *et al*., 2006). Genes with mean intensities smaller than the mean intensity of control empty spots plus twice its standard deviation in both channels were removed from the analysis, leaving 7987 detected clones. To assess differences between healthy and saprolegniasis-affected fish, data were analysed using a mixed model of ANOVA (Wolfinger *et al*. 2001) and the MAANOVA R package (Kerr *et al*. 2002, Kerr *et al*. 2000). The model included the "array" term as a random term and the "sample type" (infected or healthy) and "dye" terms as fixed terms. A permutation based F test (Fs, with 1000 sample ID permutations) was then performed, and restricted maximum likelihood was used to solve the mixed model equations. Hierarchical clustering analysis of Pearson correlation distance between genes and between treatments was run using the GeneSight 3.5 software (BioDiscovery).

Table 5.1 Complement proteins as well as other acute phase protein coding genes that were significantly over-transcribed in saprolegniasis-affected fish.

Gene	P-value†	cDNA clone	Fold change	Reference*
Complement C1r	1.60E-05	3	31.98	
Complement C2	1.01E-03	1	1.79	
Complement C3	2.00E-03	7	2.38	1,2,4
Complement C6	2.68E-03	1	2.25	4
Haptoglobin	5.32E-05	1	37.96	1,2,3,6
Precerebellin-like protein	2.24E-04	2	11.42	1,2,3,6
Differentially regulated trout protein 1	1.60E-04	4	14.24	1,2,3,6
Hemopexin	4.63E-03	1	2.80	2,6
Ceruloplasmin	4.33E-03	1	2.36	2,6
Serotransferrin	3.19E-03	3	2.90	2,6
Mannose-binding protein C	1.72E-03	2	2.19	1,2,3,6
C type lectin receptor A	7.45E-05	2	14.40	
CD209 antigen-like protein E	6.29E-04	1	8.31	

P-values from the ANOVA permutation-based F-test, the number of significant cDNA clones, the average fold change (expression level in infected fish over that in healthy fish) for all clones and references in which the gene differential expression was previously observed in fish infected with other agents are presented. A supplementary online table (Supp. Table 5.1) presents this information in a more detailed form, with individual fold changes, P-values and Q-values for each cDNA clone as well as GeneBank accessions.

* 1: Tsoi *et al.* (2004), 2: Baynes (2001), 3: Ewart *et al.* (2005), 4: Meijer *et al.* (2005), 5: Kurobe *et al.* (2005), 6: Rise *et al.* (2004), 7: Byon *et al.* 2005.

†In the case of genes with multiple significant cDNA clones, only the smallest observed P-value is presented.

5.5 Results

At a significance threshold of P<0.005, the transcription level of genes corresponding to 430 cDNA clones printed on the array differed significantly between saprolegniasis-affected and healthy fish. Among the 430 significant clones, 144 (33%) were marked as "unknown" on the microarray used in this study, meaning that they did not generate any BLAST hits with e-values less than $1X10^{-15}$ and an informative name (for a detailed description of the annotation process, see http://web.uvic.ca/cbr/grasp). Each of these represents a potential unidentified molecular actor of immunity in fish (Supplementary Tab. 5.6). The sequences of most of these cDNA clones are available in GeneBank. To correct for multiple testing, q-values were calculated from the p-values with the Q-value R package (Storey 2002). The number of significant genes increased to 1158 when applying the Q<0.05 significance threshold on the obtained q-values.

Hierarchical clustering analysis was run using the normalized gene expression data from the 25 genes for which the transcription level differences between sick and healthy fish were the most significant, excluding the unknown genes (Fig. 5.2). Clustering analysis on experimental conditions (Fig. 5.2, columns) showed a clear dichotomy between healthy (left vertical cluster) and infected (right vertical cluster) fish. Moreover, 24 of these 25 genes were over-transcribed in diseased fish and most significant clones at P < 0.005 were also over- (67%) rather than under- (33%) transcribed in infected fish**,** which echoed previous observations with different infectious agents (Ewart *et al.* 2005).

Genes coding for several acute phase proteins and, in particular, complement proteins were over-transcribed in infected fish (Table 5.1). Significant over-transcription was detected from a number of different clones corresponding to the same genes for C1r, C3, mannose-binding protein C, transferrin (serotransferrin) as well as for the fish APPs precerebellin-like protein and differentially regulated trout protein 1 (a snake neurotoxin homologue). The average fold induction (9.7-fold) was very high for the genes in table 1, some such as haptoglobin (38-fold) and the complement component C1r (32-fold) reaching particularly high values. Numerous additional clones corresponding to genes shown to be over-

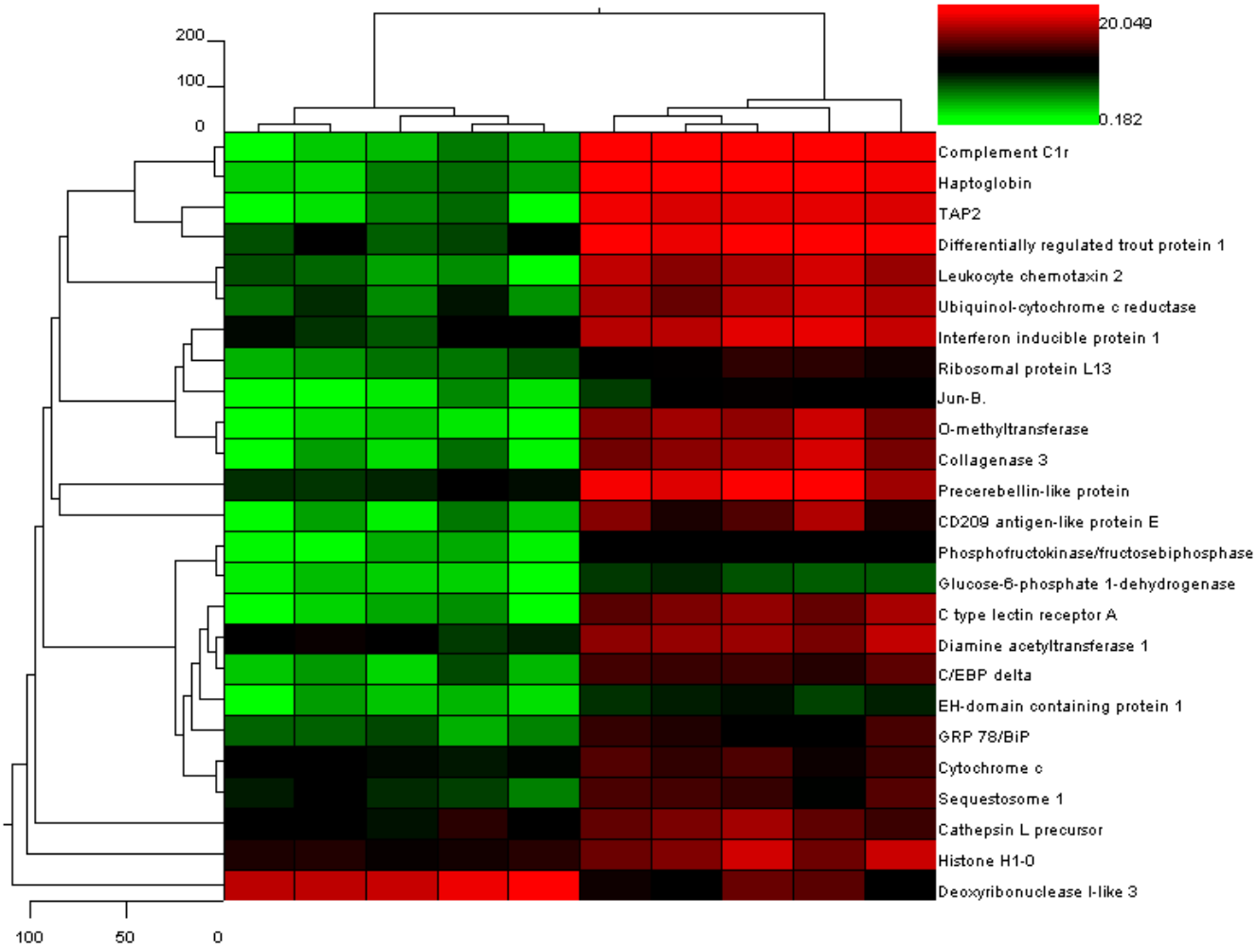


Figure 5.2 Graphical representation of the differences in transcript abundance between five saprolegniasis-affected (right) and five healthy (left) fish.

The normalized expression level is represented for the 25 most significantly differentially expressed genes by a coloured box. The palette of colours symbolizing low to high normalized expression levels is shown on the upper right part of the figure (green = low, red = high). “Unknown” genes where not considered and only the data from the most significant clone was considered in the case of genes represented by several significant cDNA clones. Hierarchical clustering of gene expression data by gene and by experiment is shown as a horizontal and a vertical tree, respectively. The trees represent relationships in expression patterns, with branch lengths indicative of the magnitude of the differences between gene expression patterns.

expressed in fish infections in previous studies were also over-expressed in saprolegniasis-affected fish (Table 5.2). Again, many genes were represented by more than one significant clone and most of these genes have well-characterized immune functions. Table 5.3 presents cDNA clones for genes previously reported as under-transcribed in diseased fish and whichalso showed under-expression in saprolegniasis-affected fish. These include 17 clones corresponding to the collagen alpha 1 coding gene. Table 5.4 presents a subset of genes, previously observed as differentially expressed in a given direction in other studies, but that were differentially expressed in a different direction in this study. Finally, an important corpus of cDNA clones corresponding to genes whose expression has not previously been reported to change in immuno-stimulated fish were differentially expressed in this study (Table 5.5). While some of these have characterized roles in immunity, several other functional groups not directly related to the immune system were represented.

5.6 Discussion

Various histological and physiological observations were made in previous studies following infection of different fish species with *Saprolegnia*. Histopathological changes beneath the superficial mycelia include dermal necrosis and oedema during early stages with deeper myofibrillar necrosis and extensive haemorrhage in the more progressive lesions (Beakes *et al.* 1994). Other studies reported signs of inflammation, lethargy and paralysis in severely infected fish (Hughes 1994). Also, data suggested damage to the liver (Richards and Pickering 1979) and thymus (Alvarez *et al.* 1995) . While some authors have suggested that *Saprolegnia* could exert immunosuppression on infected fish (Bly *et al.* 1992, Torto-Alalibo *et al.* 2005, Alvarez *et al.* 1995), Bly and co-workers (1993) observed an acute phase response in *Saprolegnia*-infected channel catfish.

Here, we investigated for the first time the genome-wide transcriptional response of fish to saprolegniasis. Over-transcription of several acute phase protein coding genes (Tab. 5.1) suggests that saprolegniasis-affected salmon undergo an acute phase response, providing molecular support to the results of Bly and co-workers on catfish (1993). Among

Table 5.2 Genes which were significantly over-transcribed in saprolegniasis-affected juvenile salmon and that were observed as over-transcribed in fish infected with other agents in previous studies.

Gene	P-value†	cDNA clone	Fold change	Reference*
Cathepsin L	9.58E-05	6	2.22	2,4
Cathepsin D	3.52E-03		1.37	2,4
Collagenase 3	3.09E-04	2	12.92	4,5
Matrix metalloproteinase	1.01E-03	1	4.27	1,3,4,6
CXCR4	4.80E-03	1	1.54	4
NADPH oxidase flavocytochrome b small subunit	4.45E-03	2	1.44	1,6
6-phosphogluconate dehydrogenase	3.39E-03	1	2.17	6
Interferon inducible protein 1	7.45E-05	3	5.80	7
Phosphofructokinase/fructosebiphosphatase 2	3.19E-05	1	3.79	1
Leukocyte chemotaxin 2	3.62E-04	2	9.45	1,2,3
Jun-B.	5.22E-04	3	4.27	1,5
CCAAT/enhancer binding protein delta	5.01E-04	2	5.22	4
Biotinidase	8.84E-04	1	3.90	2
O-methyltransferase	1.92E-04	1	15.52	1,3,6
Plasma glutathione peroxidase	4.36E-03	1	1.89	1,2,3,6
Agglutination-aggregation factor 18K-LAF	1.06E-03	2	3.36	1,3,6
TAP2	9.58E-05	2	13.45	3,6
S-100/ICaBP-like	6.71E-04	1	2.43	6
Ribosomal protein L13	4.16E-04	1	3.59	2,3
Ribosomal protein S27-like	4.50E-03	1	1.78	1
Vacuolar ATP synthase 16 kDa proteolipid subunit	2.34E-03	5	1.72	1,4,6
Thioredoxin interacting protein	6.29E-04	2	2.62	5
Thioredoxin	1.35E-03	2	2.20	1,3,4
Solute carrier family 31, member 1	7.99E-04	1	5.06	4
Solute carrier family 37, member 4	8.63E-04	1	1.94	4
Solute carrier family 25, member 5	2.72E-03	4	1.42	4
Sequestosome 1	5.22E-04	2	2.51	3
Thrombospondin-4	9.53E-04	1	4.12	4
Fibronectin	3.05E-03	1	2.20	4
26S proteasome regulatory subunit p27	1.02E-03	1	1.97	4
26S proteasome regulatory subunit p44.5	3.21E-03	1	1.60	4
Endoplasmin	1.26E-03	5	2.56	1
Plasma retinol-binding protein	1.42E-03	4	2.04	3
Polyposis locus protein 1 homolog	1.86E-03	2	1.99	1
Translation initiation factor 4E binding protein 3	2.53E-03	1	1.76	4
Probable RNA-dependent helicase p68	4.95E-03	1	1.79	6
Heat shock protein 90-beta	2.55E-03	1	1.39	5,6
CD63	3.84E-03	2	1.57	4
Placental thrombin inhibitor	3.61E-03	1	1.78	6
Insulin-like growth factor I	4.76E-03	1	1.96	6
Histone H1-0	2.45E-04	5	2.51	7††
Histone H1.3	2.63E-03	1	0.65	
Histone H2A.X.	3.72E-03	1	0.60	
H1fx-prov protein	2.38E-03	1	1.92	

P-values from the ANOVA permutation-based F-test, the number of significant cDNA clones, the average fold change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously observed in fish infected with other agents are presented. A supplementary online table (Supp. Table 5.2) presents this information in a more detailed form, with individual fold changes, P-values and Q-values for each cDNA clone as well as GeneBank accessions.

* 1: Tsoi *et al.* (2004), 2: Baynes (2001), 3: Ewart *et al.* (2005), 4: Meijer *et al.* (2005), 5: Kurobe *et al.* (2005), 6: Rise *et al.* (2004), 7: Byon *et al.* 2005.
†In the case of genes with multiple significant cDNA clones, only the smallest observed P-value is presented.
† † The observed histone H1 subtype was not specified in Byon *et al.* (2005). We presented all subtypes in this table for easier comparison.

Table 5.3 Genes which were significantly under-transcribed in saprolegniasis-affected fish and which were observed as under-transcribed in immuno-stimulated fish in previous studies.

Gene	P-value†	cDNA clone	Fold change	Reference*
Selenoprotein P, plasma, 1b	4.32E-03	1	0.68	3
Deoxyribonuclease I-like 3	2.13E-04	2	0.25	4
Collagen alpha 1(I)	1.09E-03	17	0.33	3,4
UDP-glucuronosyltransferase 2B5	1.80E-03	2	0.66	1
Ependymin	2.66E-03	3	0.63	3
Myosin regulatory light chain 2	4.98E-03	1	0.76	3
Troponin I	3.70E-03	1	0.58	4
Type II keratin E2	1.52E-03	2	0.38	4††
Keratin 13	1.60E-03	1	0.49	
Keratin, type II cytoskeletal 6A	1.83E-03	3	0.71	
Keratin 12	1.87E-03	2	0.51	
Keratin, type I cytoskeletal 14	2.47E-03	1	0.54	
Keratin, type I cytoskeletal 13	3.49E-03	1	0.45	

P-values from the ANOVA permutation-based F-test, the number of significant cDNA clones on the array, the average fold change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously are presented. A supplementary online table (supp. Table 5.3) presents this information in a more detailed form.

* 1: Tsoi *et al.* (2004), 2: Baynes (2001), 3: Ewart *et al.* (2005), 4: Meijer *et al.* (2005), 5: Kurobe *et al.* (2005), 6: Rise *et al.* (2004), 7: Byon *et al.* 2005.
†In the case of genes with multiple significant cDNA clones, only the smallest observed P-value is presented.
††The observed keratin type was not specified in Meijer *et al.* (2005).

Table 5.4 Genes observed as differentially expressed in different directions in this study and in previous studies.

Gene	P-value†	cDNA clone	Fold change	Reference*
Tob1 protein	2.28E-03	1	2.41	Under-transcribed in 3
B-cell translocation gene 1	2.28E-03	2	1.78	Under-transcribed in 3
Prostaglandine D synthase	1.26E-03	2	0.40	Over-transcribed in 2,3
Glutathione S-transferase kappa 1	2.57E-03	1	0.70	Over-transcribed in 3,6
Microsomal glutathione S-transferase 3	2.93E-03	2	0.80	Over-transcribed in 3,6
High mobility group protein 1	2.89E-03	1	0.70	Over-transcribed in 6
Putative steroid dehydrogenase KIK-I	4.85E-03	1	0.65	Over-transcribed in 6
ATP synthase alpha chain	1.32E-03	1	0.25	Over-transcribed in 6
Proteasome subunit alpha type 2	1.21E-03	1	0.25	Over-transcribed in 6

P-values from the ANOVA permutation-based F-test, the number of significant cDNA clones on the array, the average fold change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously observed in fish infected with other agents are presented. A supplementary online table (supp. Table 5.4) presents this information in a more detailed form.

* 1: Tsoi *et al.* (2004), 2: Baynes (2001), 3: Ewart *et al.* (2005), 4: Meijer *et al.* (2005), 5: Kurobe *et al.* (2005), 6: Rise *et al.* (2004), 7: Byon *et al.* 2005.

†In the case of genes with multiple significant cDNA clones, only the smallest observed P-value is presented.

these genes, C1r and C2 were not observed as over-transcribed in previous studies assessing genome-wide transcriptional responses of fish to various infectious agents. The very strong over-transcription of the C1r coding gene in this study might reflect activation of the classical complement pathway in infected fish. The over-transcription of the mannose-binding lectin (Tab. 5.1) suggests that the MB-lectin pathway is also vitalized. Considering the results for C3 and C6 (Tab. 5.1), all three complement activation pathways may be more active in saprolegniasis affected fish. The major positive APP serum amyloid A and C-reactive proteins were not represented on the microarray we used. Serum amyloid P was represented by four cDNA clones, but none was significantly differentially expressed in infected fish. Serum amyloid P is a major positive mammalian APP but was shown to be under-transcribed in *A. salmonicida* infected Artic charr (Jensen *et al*. 1997); however, this result was unclear and contested (Bayne and Gerwick 2001). Jensen and co-workers (Jensen *et al*. 1997) suggested that serum amyloid P transcriptional activation might be sex-specific; this could explain why it was not significantly over-expressed in our samples, which were all sexually immature.

The transcription level of most genes identified as differentially expressed both in this and in previous studies assessing the global gene expression response to immune challenges in fish changed in the same direction (Tab. 5.2). Among the genes represented in Tab. 5.2, several genes over-transcribed in infected fish encode proteases with a strong collagenolytic activity. Over-transcription of these genes may facilitate leucocyte transmigration. Other induced immunity-related genes include the ATP-binding cassette transporter TAP, which is essential to the MHC class I antigen presenting system, the cytokine receptor CXCR4, CD63, leucocyte chemotaxin 2, glutathione peroxidase, an agglutination and aggregation factor and the interferon inducible protein 1. It is noteworthy that the transcription factors JunB and CCAAT/enhancer binding protein coding genes, both over-transcribed in this study, transactivate APP-coding genes through cytokine and glucocorticoid response elements (Baumann *et al*. 1991). More surprising is the induction of some histone-coding genes, also observed by Byon *et al*. (2005). However, recent data suggest that histones may play a role in innate immunity in a wide range of animal species (Silphaduang *et al*. 2006). Here and in Meijer *et al*. (2005), under-expression of cyto-

structural keratin-coding genes was observed, as well as that of collagen alpha 1-coding gene. We also observed down-regulation of four other genes coding for collagens (Tab. 5.5).

In a few cases, the direction of the infection-driven expression changes did not correspond between this and previous studies (Tab. 5.4). Also, some expression changes detected in other studies (e.g. alpha-1-microglobulin down-regulation) were not detected here as significant. This could be due to many differences in the experimental protocols, including different infection protocols, microarray platforms, statistical treatment of the data, timing of sampling, water temperatures and the fact that different tissues were sampled and different infectious agents were involved. Different host fish species were also used in some of these studies. The tissue-specificity of the response and the timing of sampling likely explain most of the observed lacks of correspondence. Hence, even within a study, results between tissues are not always consistent (Rise *et al*. 2004a). Alternatively, some of these differences may reflect specificity of the response to saprolegniaceae infection. While saprolegnaceae are suspected to exert immunosuppressive actions (Alvarez *et al*. 1995), *Tob-1* and the *B-cell translocation gene 1* were both under-transcribed instead of over-transcribed in infected fish, as previously reported. *Tob-1* is known to maintain T cells in a quiescent state and has to be down-regulated for T cells to proliferate and release cytokins (Tzachanis *et al*. 2001), and the *B-cell translocation gene 1* also negatively regulates cell proliferation (Berthet *et al*. 2002).

Finally, expression differences for several genes between infected and healthy fish was evidenced here for the first time. The fact that all previous studies (except Meijer *et al.* 2005) merely used a fold-change criterion for identifying differentially expressed genes probably explains that many differences with a smaller than 2-fold change in Tab. 5.5 were not previously detected (Jin *et al*. 2001). As stated earlier, some of the changes listed in Tab. 5.5 might also reflect specificities in our experimental settings; others may represent infection stage-specific or saprolegniaceae-specific response genes. Nonetheless, many functional classes are represented in Tab. 5.5. Several genes encoding proteins implicated

Table 5.5 Genes which were significantly differently transcribed in whole saprolegniasis-affected juvenile salmon and that had not been observed as so in fish immuno-challenged with other agents in previous studies.

Gene and functional class	P-value†	cDNA clone	Fold-change
Immunity-related gene			
Prostaglandin-H2 D-isomerase precursor	2.97E-03	1	0.49
Acidic chitinase L	2.33E-03	1	0.55
Calreticulin	2.42E-03	2	2.45
G protein-coupled receptor 2	2.92E-03	1	1.76
FYN-binding protein	3.37E-03	1	1.51
Small inducible cytokine	9.69E-04	2	1.78
Myeloperoxidase	1.48E-03	1	0.45
Glucose-6-phosphate 1-dehydrogenase X	3.20E-04	1	1.98
Immunoglobulin heavy chain binding protein	3.52E-04	1	3.57
Immunophilin FKBP12.6	1.46E-03	1	1.64
Extracelluar matrix components			
Collagen alpha 1(X)	7.46E-04	3	0.29
Collagen alpha 2(I)	1.21E-03	7	0.31
Collagen a3(I)	3.75E-03	1	0.35
SPARC	3.96E-03	1	0.60
Electron transport chain			
Cytochrome c	4.58E-04	2	2.24
Ubiquinol-cytochrome c reductase complex 7.2 kDa protein	4.05E-04	1	8.47
Cytochrome c oxidase subunit IV isoform 1	2.62E-03	1	0.68
Cytochrome c oxidase polypeptide VIIa	4.95E-03	1	0.74
NADH-ubiquinone oxidoreductase 30 kDa subunit	2.00E-03	2	0.81
NADH-ubiquinone oxidoreductase B18 subunit	4.75E-03	1	0.91
ATP synthase e chain	1.32E-03	1	0.74
ATP synthase D chain	3.35E-03	1	1.95
ATP synthase coupling factor 6	4.18E-03	1	0.40
Ion transport			
Na+/K+ ATPase 1	1.88E-03	1	1.66
Na+/K+ ATPase 3	3.32E-03	1	1.75
Na+/K+ ATPase alpha subunit isoform 1b/ii	4.91E-03	2	1.08
Signal transduction			
Mcl-1a	1.11E-03	1	2.45
Tumor-associated calcium signal transducer	1.65E-03	1	2.14
Centd2	4.68E-03	1	1.50
Transcription/transcription activation			
40S ribosomal protein S16.	3.42E-03	2	2.78
RNA polymerase II transcriptional coactivator p15	7.14E-04	3	0.65
Zinc finger A20 domain containing 2, like	9.27E-04	1	3.63
Zinc finger protein 330	1.14E-03	1	2.06
XP8	1.36E-03	3	2.76
Aryl hydrocarbon receptor 2b	8.95E-04	2	2.38
RNA-binding region containing protein 2	3.16E-03	1	0.79
Similar to thyroid hormone receptor interactor 3	4.70E-03	1	0.55
Metal ion binding protein			

Selenoprotein W	1.04E-03	1	0.77
Metallothionein B	1.75E-03	1	1.75
Metallothionein-I	2.82E-03	1	2.05
Pyrimidine biosynthesis			
CTP synthase	1.12E-03	1	2.09
Enhancer of rudimentary homolog	1.15E-03	1	2.29
Protein degradation			
Similar to Ubiquitin binding enzyme	1.75E-03	1	1.90
Ubiquitin-conjugating enzyme E2 A	3.25E-03	1	2.09
Ubiquitin-conjugating enzyme E2 D2	4.00E-03	1	1.60
F-box only protein 2	2.86E-03	1	2.39
Calpain small subunit 1	3.63E-03	1	1.65
His-tagged cytosolic leucine aminopeptidase	4.13E-03	1	0.86
Meprin A alpha-subunit	3.66E-03	1	0.53
Protein localisation/folding			
Signal recognition particle 9 kDa protein	4.21E-03	1	0.65
SSR alpha subunit	3.89E-03	1	0.57
Protein disulfide-isomerase A3	1.77E-03	4	1.89
DNA replication/repair			
DNA polymerase delta subunit 4	1.81E-03	1	0.67
Proliferating cell nuclear antigen	4.05E-03	1	0.70
Cell structure and adhesion/cellular junction			
Actin-5C.	1.35E-03	2	2.03
Beta-actin	4.09E-03	1	1.30
Periostin	1.96E-03	1	0.39
Neural-cadherin	2.11E-03	1	2.13
Epithelial-cadherin	3.42E-03	1	1.77
Claudin-3	2.39E-03	1	2.21
Integrin, beta-like 1	2.64E-03	1	2.16
Nicotinamide riboside kinase 2	2.45E-03	2	1.79
Miscellanous			
Delta-6 fatty acyl desaturase	2.04E-03	1	0.79
Diamine acetyltransferase 1	3.20E-04	4	2.94
ALDH class 2	1.18E-03	1	0.64
Transposase	1.67E-03	1	2.45
Nucleophosmin	3.56E-03	1	2.58
EH-domain containing protein 1	4.69E-04	1	2.33
Uncoupling protein 2	1.49E-03	2	2.22
Fatty acid-binding protein, intestinal	2.09E-03	5	0.40
Glutamine synthetase	4.94E-03	2	1.82
Protein translation factor SUI1 homolog GC20	3.32E-03	3	1.77
Small nuclear ribonucleoprotein E	2.59E-03	1	2.57
Mid1 interacting protein 1	2.78E-03	1	0.61
Transport protein SEC61 gamma subunit	3.22E-03	1	1.89
Gelsolin	3.82E-03	1	0.62

P-values from the ANOVA permutation-based F-test, the number of significant cDNA clones on the array and the average fold change (expression level in infected fish over that in healthy fish) are presented. A supplementary online table (supp. Table 5.5) presents this information in a more detailed form, with

individual fold changes, P-values and Q-values for each cDNA clone, GeneBank accessions, as well as other significant genes with unknown function.

[†]In the case of genes with multiple significant cDNA clones, only the smallest observed P-value is presented.

in cellular structure and cellular junctions appear over-transcribed in infected fish. Also, G protein-coupled receptor 2, which is involved in T cell-mediated skin inflammation processes (Homey *et al*. 2002), was induced in infected fish. More surprisingly, the prostaglandin-H2 D-isomerase, acidic mammalian chitinase and myeloperoxidase-coding genes, three genes with known immune functions, were under-expressed in infected fish, which may reflect late infection-stage specific changes.

Potential caveats of this study include the fact that some of the observed differences may reflect the lower condition of the infected fish (lower social status, stress, other infections). Also, infection may not have been at identical stages in the different fish sampled, though it was always advanced and all infected samples displayed similar moribund behaviours. In some cases, observed changes might also reflect pathological or morbid changes instead of the immune response itself, since, as Meijer and co-workers (2005), we sampled heavily infected fish. Two inconsistencies were also observed in our data: two different cDNA clones representing small inducible cytokine and Na+/K+ ATPase alpha subunit isoform 1b/ii were observed as significantly differentially expressed in different directions. The same situation occurred in Rise et al. (2004a), which suggested that it might be due to a misidentified EST. Alternatively, one of the two clones of each pair may be a false positive, or the two clones may represent isoforms regulated in opposite directions.

Several authors argue that innate ("non-specific") immunity might be more important than adaptive ("specific") immunity in fish. In most animals, it would also be a more efficient response to fungus-like-pathogen driven infections (Ellis 2001). Indeed, many of the genes observed as differentially expressed in saprolegniasis-affected fish are associated with innate immunity. Yet evidence suggests that fish, as humans, show substantially distinct APR to viral, fungal and microbial agents (Ellis 2001, Bayne and Gerwick 2001). Here, we characterized for the first time the genome-wide transcriptional response of fish to saprolegniasis and identified several differentially expressed genes, many of which were not previously observed. Some of these might represent elements of a saprolegniaceae-specific immune response. Further studies comparing the immune response of fish to various infectious agents in controlled conditions could reveal whether the non-specific

immune response of fish to diverse pathogens differs. Such differences could be used for molecular diagnosis of fish condition using dendograms similar to Fig. 5.2 illustrating the expression of a set of genes with pathogen-specific patterns. In addition, several cDNA clones with unknown functions were differentially expressed in saprolegniasis-affected fish. Each of these represents a potential uncharacterized molecular actor of immunity in fish, and perhaps in other taxonomic groups.

5.7 Acknowledgments

This work was supported by a research grant from Science and Engineering Research Canada (NSERC strategic program) to L.B., H.G. and J.J.D., a scholarship from the *Fonds Québécois de la Recherche sur la Nature et les Technologies* (FQRNT) to C.R. This study is a contribution to the research programs of CIRSA (*Centre Inter-universitaire de Recherche sur le Saumon Atlantique*) and *Québec-0céan*. We thank Serge Higgins and the staff at the LAboratoire de Recherche en Sciences Aquatiques (LARSA), Université Laval, Québec for their assistance in rearing fish. We are also grateful to Richard Janvier (Pavillon Marchand, Université Laval) for help with electronic microscopy.

Chapitre 6 : Disturbance of social hierarchy by an invasive species: a gene transcription study

6.1 Résumé

De plus en plus d'études rapportent des perturbations écologiques ou évolutives chez des populations naturelles faisant face à des espèces envahissantes. Ainsi, on a récemment mis en évidence que la présence de truite arc-en-ciel (*Oncorhynchus mykiss*) introduite dans l'habitat naturel du saumon atlantique (*Salmo salar*) abolit chez ce dernier la hiérarchie sociale. Les acteurs moléculaires et génétiques à la base de la plasticité phénotypique sont peu connus; cette étude vise à identifier certains de ceux contribuant à la perte plastique de hiérarchie sociale chez le saumon en présence de truite ainsi qu'à identifier des gènes impliqués dans les comportements liés à la dominance. Pour ce faire, nous avons comparé, à l'aide de bio-puces, les profils de transcription génique cérébraux de saumons juvéniles dominants et subordonnés en présence ou en l'absence de truite. Les saumons dominants et subordonnés étaient plus semblables en présence de truite, et ce tant au niveau du comportement que du transcriptome cérébral. Les gènes pour lesquels le niveau de transcription différait entre saumons dominants et subordonnés en l'absence de truite étaient en général sur-exprimés chez les saumons dominants et incluaient des gènes impliqués dans la synthèse et la dégradation protéiques, la structure neuronale, de même que le transport d'oxygène. Cette étude est l'une des rares à avoir mis évidence une relation entre plasticité phénotypique et transcription, et contribue ainsi à une meilleure compréhension des mécanismes moléculaires de la plasticité dans un contexte écologique d'importance.

6.2 Abstract

Ecological and evolutionary changes in native populations facing invasion by exotic species are increasingly reported. Recently, it has been shown that competition with exotic rainbow trout (*Oncorhynchus mykiss*) disrupts dominance hierarchies within groups of native Atlantic salmon (*Salmo salar*). The genetic and molecular actors underlying phenotypic plasticity are poorly understood. Here, we aimed at identifying the genetic and molecular actors contributing to this plastic loss of dominance hierarchies as well as at identifying genes implicated in behaviours related to social dominance. By using microarrays, we compared the genome-wide gene transcription profiles in brains of dominant *versus* subordinate juvenile Atlantic salmon in presence or absence of a rainbow trout. Adding the trout competitor resulted in dominant and subordinate salmon being more similar, both behaviourally and at the level of brain gene transcription patterns. Genes for which transcription levels differed between dominant and subordinate salmon in the absence of exotic trout were mainly over-expressed in dominant salmon and included genes implicated in protein turnover, neuronal structural change and oxygen transport. Our study provides one of the few examples demonstrating a close interplay between behavioural plasticity and gene transcription, therefore contributing to the understanding of the molecular mechanisms underlying these processes in an ecologically relevant context.

6.3 Introduction

Biological invasion of species outside their native range is among the most important factors contributing to the ongoing biodiversity crisis (Clavero and Garcia-Berthou 2005, Lawler *et al*. 2006). Exotic species have strong ecological and evolutionary effects on invaded ecosystems (Mack *et al*. 2000, Strauss *et al*. 2006, Suarez and Tsutsui 2008). Notably, behavioural changes in native populations facing invasion by exotic species have been reported in several taxa (Lodge 1993). Rainbow trout (*Oncorhynchus mykiss*) are native to tributaries of the Pacific Ocean in Asia and North America, but have been introduced for food or sport in many locations throughout the world (Leprieur *et al*. 2008). Rainbow trout and Atlantic salmon display high microhabitat overlap in the wild, and it has recently been shown that competition imposed by the exotic rainbow trout strongly disrupted dominance hierarchies within groups of native Atlantic salmon (*Salmo salar*), as well as the phenotypic correlation between several behaviours (Blanchet *et al*. 2007). Juvenile salmonids are territorial and form distinct social hierarchies both in the wild and when reared in captivity (Blanchet *et al*. 2007, Overli *et al*. 1999). Identifying the genetic and molecular actors contributing to the plastic loss of dominance hierarchies previously reported in Atlantic salmon (Blanchet *et al*. 2007) is of fundamental interest in behavioural ecology (Fitzpatrick *et al*. 2005, Owens 2006), behavioural physiology and behavioural genetics (Hofmann 2003).

Research in behavioural physiology and genetics has allowed the identification of several candidate genes and endogenous molecules modulating aggressive, territorial or dominance-related behaviours (reviewed in Nelson and Chiavegatto (2001)). In salmonid fish, the physiological causes and consequences of social status have been the subject of considerable research (reviewed in Gilmour *et al.* (2005)). Modulation of brain monoaminergic activity (of neurons that secrete the monoamine neurotransmitters dopamine, norepinephrine and serotonin) by social interactions is generally seen as the basis for behavioural differences between fish of high and low social status (Winberg and Nilsson 1993). Chronically high plasmatic levels of the corticosteroid hormone cortisol were repeatedly observed in socially defeated animals (Overli *et al*. 1999) and constitutes

evidence of chronic stress in subordinate fish (Gilmour *et al*. 2005); this chronic stress could be related to many of the adverse physiological consequences of social subordination. Moreover, the so-called "challenge hypothesis" gives a central role to androgens (mainly testosterone and 11-ketosterone in fish) in the establishment of social hierarchy following contact among conspecifics (Oliveira *et al*. 2001, Fitzpatrick *et al*. 2005, Wingfield *et al*. 1990). Neuropeptides of the vasotocin family (in fish: arginine vasotocin) and oxytocin-like peptides (in fish: isotocin) also have a role in behavioural plasticity, including plasticity of aggressive behaviour (Goodson 2005). Other compounds associated with aggressive behaviour include nitric oxide (Chiavegatto and Nelson 2003), GABA (Miczek *et al*. 2003), somatostatin (Trainor et Hofmann 2007), histamine, noradrenaline as well as several growth factors (neurotrophins), signalling proteins and metabolic enzymes (for a more exhaustive list, consult Table 2 from Nelson and Chiavegatto (2001)).

Current knowledge may well represent only "the tip of the iceberg" of the complex architecture that controls aggressive behaviours (Edwards *et al*. 2006). Researchers recently used microarrays, which can track thousands of genes at once, to identify genes transcribed at different levels in the brains or whole bodies of animals from selected highly aggressive *versus* poorly aggressive strains or from dominant *versus* subordinate animals within strain (rainbow trout: Sneddon *et al*. 2005, *Drosophila*: Edwards *et al*. 2006, Dierick and Greenspan 2006, cichlid fish: Aubin-Horth *et al*. 2007). Three of these studies identified several hundred differentially transcribed candidates from which genes previously identified as implicated in aggressive behaviour were conspicuously missing but in which genes implicated in functions such as energy metabolism, protein synthesis and even muscular contraction were over-represented (Edwards *et al*. 2006, Sneddon *et al*. 2005, Dierick et Greenspan 2006). In contrast, four candidates, including arginine vasotocin, were identified in the study of cichlid fishes (Aubin-Horth *et al*. 2007).

The present study aimed at identifying genes regulating behaviours related to social dominance but also at understanding the association between gene expression and behavioural plasticity in the ecological context of species invasion. Hence, we compared, using a 16 006-gene salmonid microarray, the genome-wide gene transcription profiles of

dominant *versus* subordinate juvenile Atlantic salmon in the presence or absence of a rainbow trout (exotic competitor) to test whether gene expression differences would reflect the plastic loss of dominance hierarchies in juvenile Atlantic salmon competing with rainbow trout. Particularly, we tested the hypotheses that (i) social hierarchies within pairs of Atlantic salmon changed in the presence of rainbow trout and (ii) changes in gene expression correspondingly occurred.

6.4 Methods

6.4.1 Behavioural experiment and analysis

We used young-of-the-year (YOY) Atlantic salmon and rainbow trout caught by electrofishing in the Malbaie River (Québec, Canada, 47 ° 67 'N; 70° 16' W). In the sympatric section of the river, both species occupied similar macro-habitats and micro-habitat overlap increased as fish grew (see Blanchet *et al.* (2007) for more details). Atlantic salmon were sampled in locations where rainbow trout are not present (i.e., above a human-controlled fish ladder) to avoid potential effects of previous encounters with rainbow trout. We selected juvenile salmon and trout of similar size to avoid confounding the effects of size and species (Blanchet *et al*. 2007). In September 2005, Atlantic salmon and rainbow trout were transferred from Malbaie River to the laboratory. They were reared in separate holding tanks and fed *ad libitum* with commercial fish food pellets before experiments started.

Behavioural experiments were all performed simultaneously using 12 artificial channels made of transparent Plexiglas (Fig. 6.1). The channels and apparatus (i.e., water depth and velocity, water temperature, luminosity, etc.) are fully described in Blanchet *et al.* (2007). The only difference was the length of each channel (here, each was 0.60 m long, 0.30 m wide and 0.30 m deep, Fig. 6.1). Food rations (0.3 g artificial pellets) were manually dispensed each morning at a fixed food source, i.e. the upstream end of the channel (see Fig. 6.1). Twenty-four immature Atlantic salmon were visually selected from the holding tank to constitute twelve pairs of fish of similar size (mean fork length ± SD: 66.20 mm ±

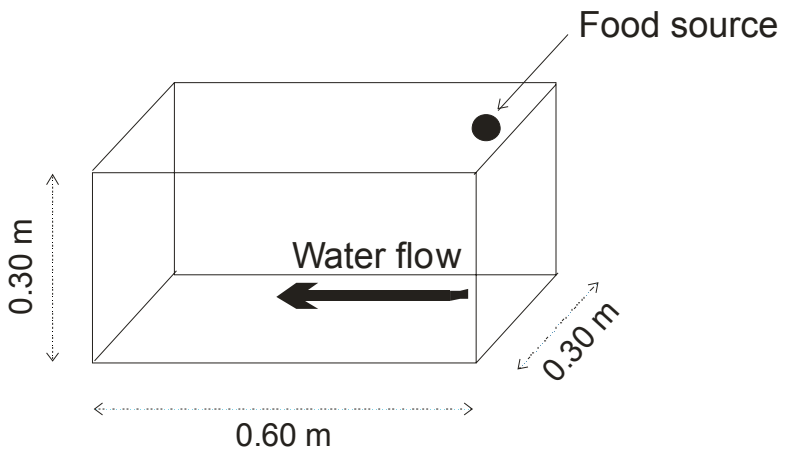


Figure 6.1 Experimental setting used to test the behaviour of Atlantic salmon raised in two competitive contexts.

In total, twelve artificial channels were used simultaneously for the behavioral tests. The food was dispensed at a single fixed point and we recorded: (1) the position of each fish relative to the food source and (2) the time each fish spent being active.

5.19 mm). No length differences were detected between treatments (mean fork length, ANOVA, $F(1,22) = 0.07$, $p = 0.78$). Each salmon was anaesthetized, measured and individually marked (Visible Implant Elastomer tags, Northwest Marine Technology, Shaw Island, Washington) before being released in the aquaria.

The behavioural experiment was performed in three steps. First, following introduction of the fish in the aquaria, the status of dominance (*i.e.*, subordinate or dominant) of individuals within each salmon pair was evaluated. Dominance was measured following the methodology described (Sloman *et al*. 2002). A mean behaviour score for position from the food source, food acquisition and social interaction was calculated for each fish during the first four days of the experiment (see Sloman *et al.* (2002) for more details). The fish with the highest score in a given pair was considered the dominant of that pair. At the end of these four days, the hierarchy was stable within each pair. Second, during the next three days we performed behavioural observations to characterise the behaviour of dominant and subordinate salmon in the absence of competing rainbow trout. Each channel was observed for 5 min each morning, directly after feeding the fish. We measured two behaviours: (1) the position of each fish relative to the food source and (2) the time each fish spent being active. A fish was considered as being active when it was out of a refuge, facing the current, and propped up on its pectoral fins. During these observations, aggressive acts were sparse and were not recorded. Third, after this seven-day period, one additional competitor was added to each of the aquarium. We added one rainbow trout in six channels (interspecific competition treatment) and one Atlantic salmon in the six other channels (intraspecific competition treatment). These supplementary fish were chosen haphazardly from the stock. This substitutive design allowed us to evaluate the effect of competition by the exotic species relative to an equivalent level of intraspecific competition and to maintain the same fish density in both treatments. The rainbow trout did not significantly differ in size from the Atlantic salmon we added (two-tail t-test, $t = -0.63$, $p = 0.538$). After a two day acclimation, we recorded the behaviour of members of each pair (previously identified as dominant and subordinate) in presence of a competing rainbow trout or of a third salmon (following the approach described above).

During the experiment, one subordinate fish died in the intraspecific competition treatment; the number of replicates was then five instead of six for this treatment. To evaluate whether the social hierarchies within pairs of Atlantic salmon changed after the addition of rainbow trout, we compared the behavioural repertory of dominant and subordinate Atlantic salmon before (second step of the experiment) and after the addition of a competitor (rainbow trout or Atlantic salmon, third step of the experiment) in the aquaria. Instead of analysing each behavioural variable independently, we used a multivariate analysis of variance with repeated measures (MANOVAR: Zar 1999, Crawley 2007) to test for behavioural changes between the second and the third step of the experiment. The dependent variables were the position of each fish relative to the food source (log transformed) and the time each fish spent being active (arcsine transformed). We used the "dominance rank" (dominant or subordinate) and the "competitive treatment" (interspecific or intraspecific competition) as independent variables. The "period of observation" (before or after the addition of a competitor) was the within-subject factor (*i.e.*, the repeated measure). All possible interaction terms were considered.

6.4.2 Transcriptomic experiment

6.4.2.1 RNA extraction, labelling and cDNA hybridisation

Following the 10-day behavioural experiment, all fish were anaesthetised and whole brains were taken from both salmon of each initial pair. Brains were immediately frozen in liquid nitrogen, and later homogenised individually in TRIZOL@Reagent (Invitrogen, San Diego) using a Diax 100 homogeniser (Heidolph instruments). Total RNA was extracted as previously described (Roberge *et al*. 2006, Roberge *et al*. 2007). For each sample, 5µg total RNA was retro-transcribed and labelled using Genisphere Array 350 3DNA array detection kits and the Superscript II retro-transcriptase (Invitrogen, San Diego) according to the manufacturer's instructions. Transcription profiles of six dominant and six subordinate salmon were contrasted on six microarrays. Three of the salmon pairs considered in the microarray experiment had faced competition by an exotic rainbow trout in the last part of the behavioural experiment while a third salmon had been added in the aquaria of the other

three pairs (see above). The cDNA microarrays used here were obtained through the consortium Genomic Research on All Salmon Project (cGRASP, available from Ben F. Koop, bkoop@uvic.ca), and contain 16,006 unique salmonid cDNA clones, some of which may yet represent the same transcript (von Schalburg *et al*. 2005).

6.4.2.2 Signal detection, data preparation and statistical analysis

Signal detection and data preparation was done as previously reported (Roberge *et al*. 2006). Spots with mean intensities for both the dominant or subordinate categories smaller than the mean intensity of control empty spots plus twice its standard deviation or with a coefficient of variation above one for either the dominant or subordinate categories were removed from the analysis, leaving 5142 and 5124 cDNA clones to be analysed for the interspecific and intraspecific competition experiments, respectively. Gene transcription data from the interspecific and intraspecific competition experiments were analyzed in two separate ANOVA using the MAANOVA R package (Kerr *et al*. 2000, Kerr *et al*. 2002). The ANOVA model included in each case the "array" term as a random term and the "social rank" (dominant or subordinate) and "dye" terms as fixed terms. A permutation-based F-test (Fs, with 1000 permutations) was then performed and restricted maximum likelihood was used to solve the mixed model equations. Specifically, R/MAANOVA recreates a null distribution of the data by randomly permuting the columns in the datasets in order to calculate the permutation-based p-value. Q-values were calculated from the permutation based p-values using the Q-value R package (Storey 2002). The Q-value of a test measures the proportion of false positives incurred (false discovery rate or FDR) when that particular test is called significant. Hierarchical clustering analysis between genes and between treatments was run using the GeneSight 3.5 software (BioDiscovery).

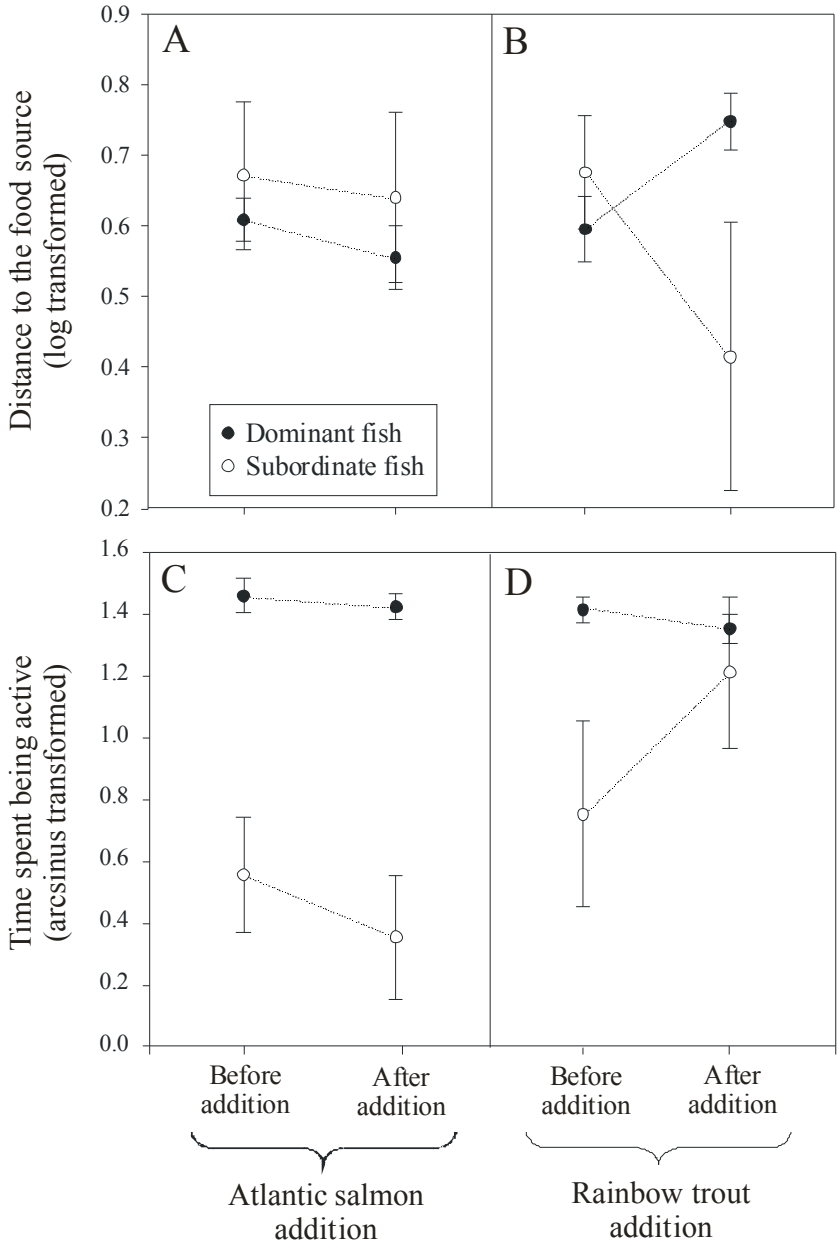


Figure 6.2 Behavioural characteristics of juvenile Atlantic salmon raised in two competitive contexts.

The behavioural characteristics of dominant (black dots) vs. subordinates (white dots) Atlantic salmon (Salmo salar) before and after the addition of either intraspecific (A and C), or interspecific (the exotic rainbow trout; B and D) competitors. The distance each fish was from the food source (upper panel), and the time each fish spent being active (lower panel) were used to characterize the behaviour of dominant and subordinate salmon.

Table 6.1 Results of a MANOVAR used to evaluate whether behavioural changes of dominant and subordinate Atlantic salmon occurred when rainbow trout or Atlantic salmon are added as competitors.

	Wilks'λ	**F-value**	**d.f.**	**P-value**
Independent variables				
Dominance rank	0.463	9.905	2,17	0.001
Competitive Treatment	0.888	1.065	2,17	0.366
Period of observation	0.892	1.044	2,17	0.373
Dominance*Treatment	0.709	3.488	2,17	0.053
Dominance*Treatment	0.607	5.485	2,17	0.014
Treatment*Period	0.581	6.154	2,17	0.009
Dominance*Treatment*Period	0.391	13.191	2,17	<0.0001

6.5 Results

6.5.1 Behavioural experiment

As previously reported, rainbow trout strongly disrupt the social hierarchy between subordinate and dominant juvenile Atlantic salmon (Table 6.1, Figures 6.1A-6.1D) (Blanchet *et al*. 2007). Indeed, in the absence of rainbow trout, dominant and subordinate salmon significantly differed in the behaviours they displayed, with dominant fish being closer to the feeding source (Figures 6.1A-6.1B) and also more active (Figures 6.1C-6.1D). After rainbow trout were added into the system, subordinate and dominant fish tended to be behaviourally more similar one to each other, particularly in time spent being active (Figure 6.1D). For the distance to the feeding source, the situation was almost reverse since subordinate salmon tended to be closer to the feeding source (Figure 6.1B). However, it is worth noting the huge variation observed for subordinate salmon, which may suggest a stronger interaction between individuals. When an Atlantic salmon was added into the aquaria instead of a rainbow trout, neither the dominant nor the subordinate fish were affected by this additional competitor (Figures 6.1A, 6.1C), thus supporting the idea that the effect of rainbow trout on the behaviour of Atlantic salmon was highly species-specific (see also Leprieur *et al.* 2008). Indeed, dominant fish remained closer to the feeding source than subordinate fish (Figure 6.1A) and were still more active (Figure 6.1C).

6.5.2 Transcriptomic experiment

Figure 6.3 shows that, for any given significance threshold, substantially more genes are differentially transcribed between dominant and subordinate Atlantic salmon in the purely intraspecific situation (Figure 6.3A) than in the interspecific competition experiment (Figure 6.3B). This result is not associated with increased experimental error in the interspecific competition experiment, since both experiments were carried out at the same time, by the same person using the same material. Moreover, the average coefficient of variation (CV) of the normalized hybridization signals for all genes was smaller for the interspecific than for the intraspecific competition situation (0.235 and 0.240, respectively).

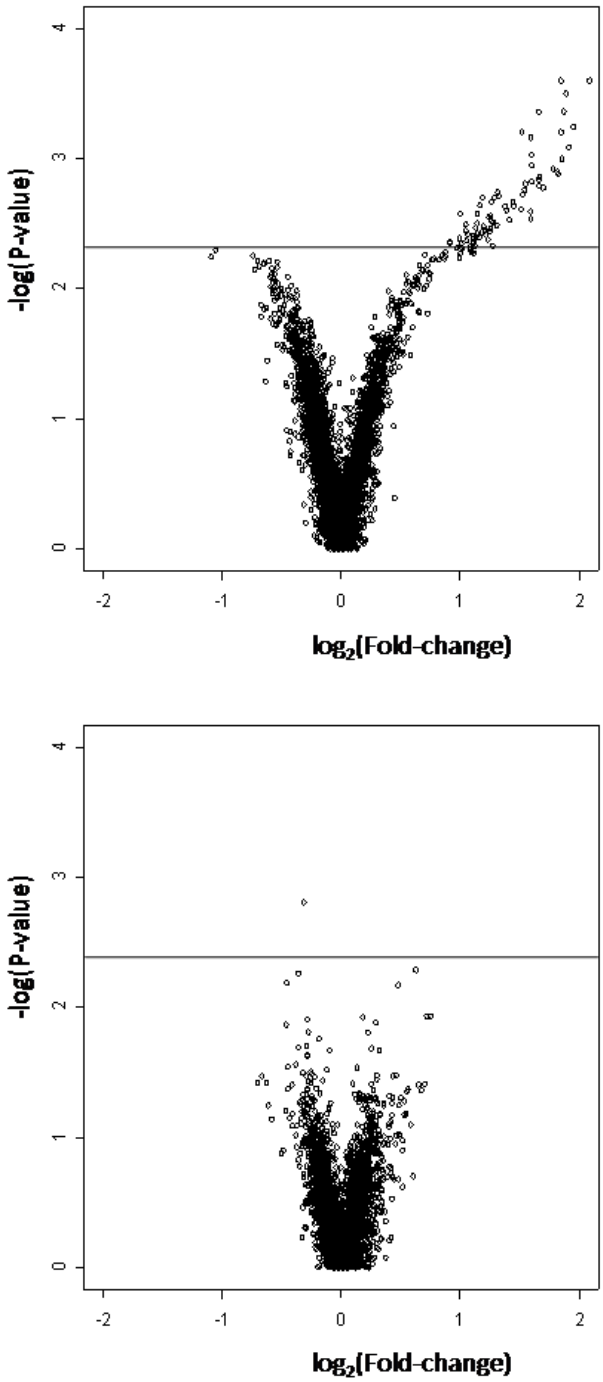


Figure 6.3 Genome-wide gene transcription profiles in brains of juvenile Atlantic salmon raised in two competitive contexts.

In an ANOVA comparing (A) dominant and subordinate salmon in the absence of rainbow trout and (B) dominant and subordinate salmon in presence of a rainbow trout, these volcano plots present the significance (-log(P-value), Y-axis) of the observed difference in transcription for each of the 5142 detected gene against the magnitude of this difference (log_2(average fold change), X-axis). Positive log_2(average fold change) values represent genes over-transcribed in the brain of dominant juvenile salmon while negative log_2(average fold change) values represent gene under-transcribed in dominant versus subordinate salmon.

Only one gene showed significant transcription level differences between dominant and subordinate salmon that had been exposed to a rainbow trout at the $P < 0.005$ significance threshold, which is less than expected by chance alone. This gene was hence considered as a false positive and therefore not interpreted any further. By contrast, the comparison of dominant and subordinate fish in a purely intraspecific competition context revealed more significant differences than expected by chance (73 significant cDNA clones, while 25 are expected by chance alone at the $P < 0.005$ significance threshold). This must however be interpreted cautiously since estimation of the expected number of false positives from the total number of spots analyzed might be misleading as a same gene can be represented by several spots (same or different ESTs) and the expression of many genes is expected to be correlated, which cannot be accounted for in the analysis. Interestingly, all of the differentially expressed transcripts at $P < 0.005$ appeared over- rather than under-transcribed in dominant individuals**,** which corroborates the results of previous transcriptomic experiments which have documented gene transcription differences implicated in aggressive behaviour (Edwards *et al*. 2006, Dierick and Greenspan 2006).

Table 6.2 presents the 27 different gene products corresponding to the 73 cDNA clones which showed significant transcription level differences in the brains of dominant and subordinate juvenile salmon ($P < 0.005$) in the intraspecific experiment. Given the small sample size and the inter-individual variability of the detected signal, these candidates still have non-negligible chances of being false positives (q-values between 0.159 and 0.196). Six candidates were marked as “unknown”, since the corresponding cDNA clone sequence did not generate any BLAST hits with e-values $< 1 \times 10^{-15}$ and an informative name during the array annotation process. The functions of the remaining candidates are discussed in the following section.

Figure 6.3 shows that the normalized hybridization signals for the 15 most significantly differentially expressed non-redundant genes can be used to accurately separate dominant and subordinate individuals in the intraspecific experiment. Such sorting was not possible for pairs exposed to the trout competitor (not shown).

Table 6.2 Gene products corresponding to the 73 cDNA clones which showed significant transcription level differences in the brain of dominant and subordinate juvenile salmon ($P < 0.005$) in the absence of rainbow trout.

Gene product or cDNA clone	P-value	Q-value	Fold change	cDNA clone
Protein degradation				
Ubiquitin-conjugating enzyme E2G 2	$2.5X10^{-04}$	0.159	4.0	1
Ubiquitin carboxyl-terminal hydrolase isozyme L1	$3.4X10^{-03}$	0.195	2.6	1
Proteasome subunit alpha type 1	$4.7X10^{-03}$	0.196	2.2	1
Oxygen transport				
Hemoglobin α	$2.5X10^{-04}$	0.159	4.4	22
Hemoglobin β	$4.4X10^{-04}$	0.174	4.1	18
Hemoglobin ε	$6.3X10^{-04}$	0.181	3.7	6
Immunity-related				
Peptidyl-prolyl cis-trans isomerase B	$8.2X10^{-04}$	0.193	4.0	2
MHC class I antigen pseudogene and proteosome subunit LMP7/PSMB8	$3.2X10^{-03}$	0.195	2.9	1
Apoptosis-related				
Caspase 8	$1.0X10^{-03}$	0.193	3.8	1
TGFB-inducible early growth response protein 2	$2.3X10^{-03}$	0.194	3.2	1
Signal transduction				
Tumor protein D53	$1.4X10^{-03}$	0.193	3.5	1
Guanine nucleotide-binding protein	$3.0X10^{-03}$	0.195	2.7	1
Transcription/protein synthesis				
DNA-directed RNA polymerase	$2.3X10^{-03}$	0.194	2.8	2
60S ribosomal protein L28	$3.8X10^{-03}$	0.196	2.4	2
Actin cytoskeleton organisation				
Actin-related protein 1 homolog B	$2.9X10^{-03}$	0.195	2.9	1
Kelch-like protein 1	$4.5X10^{-03}$	0.196	2.1	1
Miscellanous				
Midasin	$1.9X10^{-03}$	0.193	3.4	1
Myosin regulatory light chain 2	$4.2X10^{-03}$	0.196	2.4	1
Brain lipid-binding protein	$4.6X10^{-03}$	0.196	2.1	1
Collagen alpha 2(I)	$4.8X10^{-03}$	0.196	2.5	1
Biotinidase	$4.9X10^{-03}$	0.196	2.2	1
Unknown function				
CA053773 UNKNOWN	$1.3X10^{-03}$	0.193	3.6	1
CA060279 UNKNOWN	$1.3X10^{-03}$	0.193	3.6	1
CA037818 UNKNOWN	$1.6X10^{-03}$	0.193	3.6	1
CA061786 UNKNOWN	$2.0X10^{-03}$	0.193	3.5	1
CB501353 UNKNOWN	$4.3X10^{-03}$	0.196	2.3	1
CK991021 UNKNOWN	$4.9X10^{-03}$	0.196	2.1	1

Permutation-based P-values from the ANOVA are presented, as well as the corresponding Q-values, the average fold change in gene transcription level and the number of distinct significant cDNA clones corresponding to each gene product. In cases where a gene was represented by more than one significant cDNA clone, only data from the most significant cDNA clone is presented.

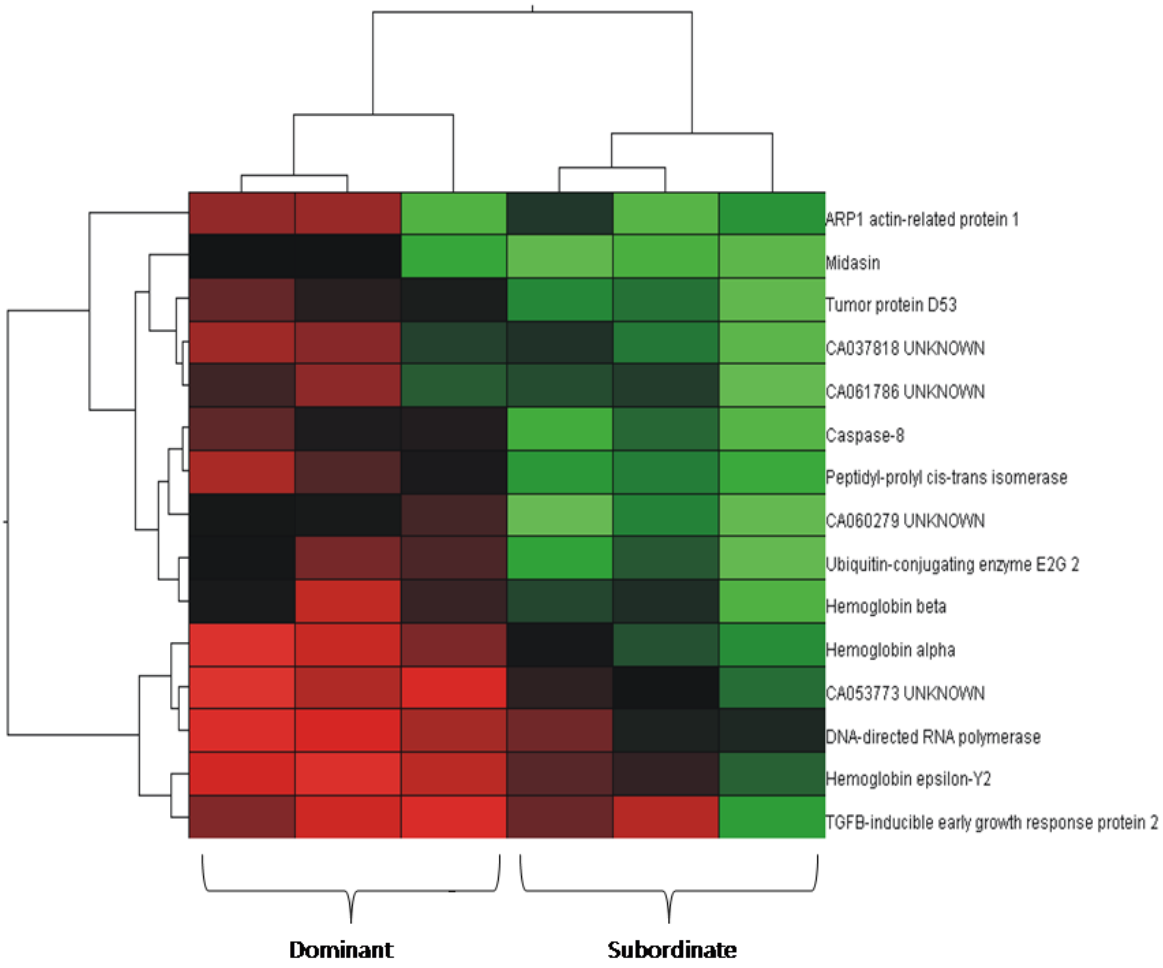


Figure 6.4 Differences in transcript abundance in the brains of subordinate and dominant juvenile Atlantic salmon.

Graphical representation of the differences in transcript abundance in the brains of three subordinate (right) and three dominant (left) juvenile salmon in the absence of rainbow trout. For each individual, the normalized transcription level is represented for the 15 most significantly differentially expressed genes by a coloured box (red: high expression, green: low expression). Only the data from the most significant cDNA clone were considered in the case of genes represented by several significant cDNA clones. Hierarchical clustering of gene expression data by gene and by experiment is shown as a horizontal and a vertical tree, respectively. The trees represent relationships between expression patterns, with branch length indicative of the magnitude of the differences between these patterns across genes or samples.

6.6 Discussion

This study identified genes implicated in behavioural differences related to social dominance, which contributes to the understanding of the relationship between gene expression and behavioural plasticity in the context of competitive interactions between native and invasive species. Namely, our results provide evidence for the differential transcription of 27 different genes between dominant and subordinate salmon. Additionally, the greater degree of similarity in the behaviour of subordinant and dominant salmon in presence of the exotic competitor (see also Blanchet *et al.* 2007) reflected the paucity of transcriptional differences observed between subordinate and dominant salmon after the introduction of an interspecific competitor in the gene transcription experiment. Thus, the presence of the exotic competitor (rainbow trout) apparently suppressed most of the transcriptional differences between dominant and subordinate salmon. Some of the suppressed differences might represent changes causing the loss of dominance hierarchy, whereas others might be a consequence of it; this study cannot disentangle such causal links. Yet, the identification of genes differentially regulated between dominant and subordinate salmon in absence but not in presence of trout is a first step towards clarifying the molecular mechanisms associated with the plastic breakdown of social hierarchies. In particular, co-regulated and functionally related candidate genes could help identifying molecular actors implicated in the differences at the behavioural and transcriptional levels. Yet hierarchical clustering by differentially expressed genes (Figure 6.3) and scanning of the literature for their potential regulatory relations did not reveal any conspicuous pattern that would point to one or a few key regulator genes in the present study.

Admittedly, the use of a microarray not specific for brain cDNA to analyse gene expression in the brain is not without limitations (Edwards *et al*. 2006, Dierick and Greenspan 2006, Aubin-Horth *et al*. 2005). For instance, the most obvious candidate genes for aggressive behaviour (see Introduction) were not represented on the microarray we used, which was not specific for brain tissue (4 of the 33 salmonid cDNA libraries used for constructing the

microarray were from brain tissue), and were therefore not among the candidates identified here. This can also generate results which seem puzzling at first glance. For example, we observed differential expression of several cDNA clones representing three different globin genes (Table 6.2). Differential haemoglobin chain expression in brain tissue between distinct phenotypes or populations, including salmonids, has been reported previously (Dierick and Greenspan 2006, Aubin-Horth *et al*. 2005, St-Cyr *et al*. 2008). Hence, higher brain expression of both alpha- and beta-globin mRNA was observed in Atlantic salmon reared in laboratory conditions compared to fish reared in natural streams (Aubin-Horth *et al*. 2005). Unlike their mammalian counterparts, mature fish erythrocytes are nucleated and can synthesise haemoglobin while circulating in the blood (Lund *et al*. 2000, Speckner *et al*. 1989). This raises the hypothesis that, in fish, increased transcription of haemoglobin genes could occur within the fish nucleated red blood cells and contribute to or be a consequence of a dominant social status.

Three genes implicated in protein degradation were over-transcribed in dominant *versus* subordinate salmon (Table 6.2). The expression of one of these, *ubiquitin carboxyl-terminal esterase L1*, is highly specific to neurons (and reproductive organs) in mouse and may insure ubiquitin stability within neurons (Osaka *et al*. 2003). This and the over-transcription of genes implicated in transcription and translation (Table 6.2) might suggest increased overall protein turnover in the brains of dominant fish. Differential transcription of several genes implicated in protein degradation as well as of ribosomal proteins has also been observed in *Drosophila* strains selected for aggressive behaviour (see Dierick *et al.* (2006), but they did not specify in which strain individual genes were over-transcribed). Protein degradation and synthesis were also mentioned among the main functional categories of genes showing contrasting transcription levels between dominant and subordinate rainbow trout (Sneddon *et al*. 2005). However, the authors did not specify the identity of the genes and direction of the over-transcription. A fourth gene over-transcribed in dominant fish and implicated in protein degradation (Table 6.2) encodes a proteasome subunit critical for class I antigen presentation in mouse, proteasome subunit LMP7 (Fehling *et al*. 1994). This gene was therefore classified in the “immunity-related” category

rather than in "protein degradation". Interestingly, the gene encoding kelch-like 1, a protein primarily expressed in brain where it is hypothesised to have a role in the organisation of the actin cytoskeleton (Nemes *et al*. 2000), was also over-expressed in dominant salmon (Table 6.2). While behavioural plasticity is expected to be initially based on changes in neuronal activity and excitability as well as endocrine responses, subsequent changes in brain and behaviour (e.g. memory formation) are expected to result from structural and physiological changes in neurons (Hofmann 2003). Also, it has recently been found that neuron proliferation was reduced in subordinate *versus* dominant rainbow trout (Sorensen *et al*. 2007). In this context, over-expression of *kelch-like 1* in dominant salmon could be implicated in increased structural changes in neurons or in organizing newly formed neurons. In the same way, brain lipid-binding protein, also over-expressed in dominant *versus* subordinate salmon (Table 6.2), is a fatty acid-binding protein that was suggested to play a role in neuronal and glial cell differentiation (Bennett *et al*. 1994).

To conclude, our study provides one of the few examples demonstrating a close interplay between behavioural plasticity and changes in gene expression in an ecologically relevant context. Behavioural plasticity is a key mechanism for animals facing rapid ecological changes such as species invasion (Pigliucci 2005), and molecular mechanisms underpinning this plasticity are actually not completely understood (Fitzpatrick *et al*. 2005). Our study therefore contributes substantially to this common effort of clarifying the molecular mechanisms of behavioural plasticity. Moreover, our results provide evidence for the influence of an introduced competitor on salmon intra-specific competitive interactions. Since such intra-specific interactions are known to play a role in the evolution of salmon reproductive strategies, this raises the hypothesis that the introduction of rainbow trout could impact on the evolution of salmon populations for such traits. Indeed, two major male reproductive strategies co-exist in Atlantic salmon (anadromous dominant males and sexually precocious sneakers) which appear to be partly heritable (Garant *et al*. 2003a) and are linked to the dominance status of individuals at the juvenile stage (Metcalfe *et al*. 1989, Hofmann *et al*. 1999). In the context of game theory (Gross 1996), the virtual suppression of dominance hierarchies in salmon by exotic rainbow trout may then disrupt the

evolutionarily stable strategy (ESS) of the two male reproductive strategies in salmon. Moreover, the identification of genes for which the transcription level is altered by intra- and inter-specific interactions provides candidates towards a better understanding of the molecular mechanisms that could be involved in the evolution of salmon reproductive strategies.

6.7 Acknowledgments

This work was supported by a research grant from Science and Engineering Research Canada (NSERC strategic program) to L.B., H.G. and J.J. Dodson, and by a scholarship from the *Fonds Québécois de la Recherche sur la Nature et les Technologies* (FQRNT) to C.R. We are also grateful to associate editor H. Ellegren and three anonymous referees for their constructive comments on an earlier version of the manuscript. This study is a contribution to the research programmes of the CIRSA (*Centre Inter-universitaire de Recherche sur le Saumon Atlantique*), the RAQ (*Réseau Aquaculture Québec*) as well as of Québec-Océan.

Chapitre 7 : Conclusion générale

7.1 Changements évolutifs récents chez le saumon d'élevage

L'un des principaux objectifs des travaux présentés dans cette thèse était d'étudier les mécanismes moléculaires de l'évolution contemporaine chez le saumon atlantique à l'aide de bio-puces. La création, par sélection artificielle maintenue sur plusieurs générations, de lignées de saumon d'élevage est un exemple d'évolution contemporaine dirigée par une activité humaine et son étude a fait l'objet des deux premiers chapitres.

Dans le premier chapitre, nous avons pu identifier des gènes présentant, entre saumons d'élevage et sauvages, des différences héritables du niveau de transcription. Plusieurs de ces gènes codaient pour des protéines impliquées, notamment, dans le métabolisme énergétique, la synthèse protéique et l'immunité; la modification du niveau de transcription de ces gènes pourrait être en lien avec la sélection pour une croissance plus rapide et une résistance accrue aux maladies. Toutefois, puisque des animaux entiers ont été utilisés, il nous impossible d'affirmer que ces différences sont le résultat de changements du niveau d'expression intracellulaire de ces gènes; certaines de ces différences de niveau de transcription génique pourraient aussi, par exemple, résulter d'une croissance différentielle des alevins et d'une différence dans les proportions de leurs divers tissus. La même expérience a été faite à la fois pour une souche d'élevage norvégienne et une souche d'élevage canadienne, chacune comparée à sa plus proche population sauvage. Étant donnée la similitude des pressions de sélection artificielle (voir chapitre 1) dans ces deux pays, ce système constituait également un modèle pour l'étude des mécanismes de l'évolution parallèle. Ainsi, bien qu'un très grand nombre de gènes puissent influencer plus ou moins directement le bilan énergétique et, par-là, le taux de croissance et la taille à maturité, 16% des changements observés l'ont été à la fois dans les souches d'élevage canadienne et norvégienne. De plus, pour plus du tiers des gènes dont le niveau de transcription différait entre saumons d'élevage et sauvages, le niveau de transcription différaient également entre les populations sauvages de Norvège et du Canada et, dans la plupart des cas, le changement observé dans la souche d'élevage semblait «contrebalancer» partiellement la différence observée entre pays (tableau 2.2). Ces résultats suggèrent, en

accord avec Haldane (1932), que certains gènes participent de façon disproportionnée à la production de variation génétique pour les caractères phénotypiques sélectionnés (taux de croissance et taille à maturité, notamment) chez des populations de saumon atlantique évoluant en parallèle, et ce tant en nature qu'en captivité. L'observation de ce que les changements entraînés par l'élevage semblent souvent contrebalancer partiellement des différences préexistantes entre pays suggère en outre l'existence de contraintes génétiques ou développementales limitant la magnitude de changement pouvant être sélectionnée pour un gène donné. Globalement, ces résultats nous renseignent sur les mécanismes génétiques de l'évolution parallèle et permettent d'évaluer l'ampleur des différences génétiques accumulées en 25 à 35 ans entre saumons d'élevage et sauvages.

L'étude présentée dans le deuxième chapitre de cette thèse visait à évaluer l'importance des différences héritables entre les transcriptomes de saumons sauvages et de saumons hybrides issus de deux générations de croisement de saumons d'élevage à des saumons sauvages. Cette étude a révélé que la plupart des différences observées au chapitre 1 entre saumons d'élevage et sauvages se diluaient, s'étiolaient suite à des rétrocroisements répétés entre saumons d'élevage et saumons sauvages (tableau 3.2), suggérant une base génétique additive des niveaux de transcription pour ces gènes. Cependant, nos résultats suggèrent que les différences sont plus nombreuses et présentent une étendue et une magnitude beaucoup plus importantes dans cette étude que dans celle présentée au chapitre 1. Il semble ainsi que la plupart des différences observées ici n'aient pas été détectables dans l'étude du chapitre 1 et soient le résultat de ce que le niveau de transcription de nombreux gènes serait sous contrôle génétique polygénique et non-additif. Ce résultat est compatible avec ceux de plusieurs autres études (e.g. Gibson et coll. 2004, Vuylsteke et coll. 2005, Auger et coll. 2005, Hedgecock et coll. 2007) qui concluaient que les niveaux de transcription géniques semblaient fréquemment être sous contrôle génétique polygénique et non-additif. Ainsi, les saumons d'élevage auraient accumulé encore plus de changements qu'il nous avait été possible d'en détecter au premier chapitre. Sous l'angle de la conservation, les résultats de ces deux premiers chapitres mettent en évidence que l'interaction génétique de saumons d'élevage et de saumons sauvages en nature peut

modifier de façon substantielle le contrôle génétique de la transcription et avoir des conséquences imprévisibles sur l'adaptation des saumons de populations naturelles à leur environnement. En conséquence, ces résultats appuient l'idée que des mesures limitant les échappées de saumons d'élevage et leur reproduction en nature doivent être prises rapidement.

7.2 Balayage transcriptomique et génétique quantitative de la transcription

L'identification de gènes transcrits à des niveaux moyens différents entre différentes populations naturelles ne suffit pas pour conclure que ces différences sont le résultat de la sélection naturelle, même lorsque l'effet de l'environnement est contrôlé. Dans le troisième chapitre de la présente thèse, nous avons proposé et appliqué une nouvelle méthode permettant d'identifier des gènes dont le niveau de transcription aurait été modifié par l'action de la sélection directionnelle entre deux ou plusieurs groupes ou populations. Ainsi, l'estimation, à l'aide d'un «modèle animal» (Kruuk 2004), de la variance génétique additive du niveau de transcription de chaque gène détecté nous a permis d'estimer le Q_{st} du niveau de transcription ainsi que son héritabilité. Nous avons alors obtenu, pour les 1044 gènes dont l'héritabilité du niveau de transcription était significative, la distribution des Q_{st} du niveau de transcription et considéré les 1.5% des gènes avec les Q_{st} les plus grands (16 gènes, tableau 4.1) comme ceux pour lesquels il est le plus probable que le niveau de transcription ait évolué en 6 générations sous l'effet de la sélection naturelle. Il est intéressant de noter que tous ces gènes étaient surexprimés dans la sous-population de l'amont de la passe migratoire. De plus, plusieurs catégories fonctionnelles reliées à la respiration cellulaire étaient sur-représentées parmi les gènes présentant une héritabilité significative du niveau de transcription. Ainsi, on peut émettre l'hypothèse que le niveau de transcription des gènes impliqués dans la respiration cellulaire serait plus souvent sous contrôle génétique additif que celui de l'ensemble des gènes exprimés chez les juvéniles de saumon.

Dans ce chapitre, nous avons également appliqué trois tests statistiques afin de vérifier, gène par gène, la normalité et l'unimodalité des données de transcription (ce qui,

étonnament, n'avait été entrepris qu'une seule fois, et seulement pour la normalité). Gibson et Weir (2005) ont récemment proposé que les données de transcription pourraient bien souvent adopter une distribution à quelques modes discrets plutôt qu'une distribution normale étant donnée l'observation fréquente de QTL d'expression à effets majeurs. Toutefois, nos résultats montrent que, pour la très grande majorité des gènes, la distribution des données de transcription pour 80 individus est compatible avec une distribution unimodale et normale. Ainsi, la forte prévalence de QTL d'expression ayant des effets importants dans la littérature scientifique pourrait être en partie basée sur un biais de détection des QTL à effets majeurs ou sur un biais dans la publication des résultats de ce type d'expérience.

7.3 Plasticité phénotypique et transcriptome

7.3.1 Réponse immunitaire à la saprolegniose

La forte densité de poissons en milieu d'aquaculture augmente grandement les risques de maladie et leur transmission; les poissons qui s'échappent des installations d'élevage peuvent aussi déclencher des épidémies dans les populations sauvages (Naylor et coll. 2005). Les saprolegniacées pathogènes causent la saprolegniose, l'une des plus importantes maladies en eau douce chez les poissons; elle entraîne chaque année d'importantes pertes économiques pour l'aquaculture partout à travers le monde, et les salmonidés y sont particulièrement vulnérables (Tampieri et coll. 2003, Hughes 1994). Le cinquième chapitre de ma thèse présente une étude dans laquelle on compare les niveaux de transcription de 16006 gènes chez des juvéniles de saumons sains ou atteints de saprolegniose. Les résultats obtenus contredisent l'hypothèse selon laquelle les oomycètes pathogènes responsables de cette maladie inhiberaient le système immunitaire de l'hôte poisson (Alvarez et coll. 1995, Bly et coll. 1992) en ce qu'ils montrent clairement une activation de la transcription de plusieurs gènes codant pour des protéines de la réponse immunitaire non-spécifique (tableau 5.1), notamment des protéines du complément. En outre, plusieurs autres gènes dont on sait que le niveau de transcription est modifié chez les poissons suite à des infections par d'autres agents pathogènes ont également vu leur niveau de transcription modifié dans cette étude (tableaux 5.2 et 5.3). Puisque l'augmentation du niveau de

transcription de plusieurs gènes de la réponse immunitaire est attendue chez des poissons malades, cette étude constitue le seul volet de la thèse comportant des témoins positifs; les résultats de cette partie valident ainsi la technique et le matériel utilisés pour les expériences des autres chapitres. De plus, plusieurs gènes pour lesquels un changement d'expression en réponse à un pathogène n'avait jamais été décrit auparavant ont montré un tel changement en réponse à la saprolegniose (tableau 5.5). Bien que plusieurs autres raisons puissent expliquer qu'une partie des résultats obtenus soient propres à notre étude (stade de l'infection, espèce hôte, stade de vie de l'hôte, procédure d'infection, analyse statistique), on peut supposer que ce fait s'explique aussi par une certaine spécificité de la réponse à chaque pathogène. Ainsi, et ce malgré son nom, la réponse immunitaire non-spécifique à des infections virales, fongiques ou bactériennes diffère chez les poissons comme chez l'humain (Ellis 2001, Bayne et Gerwick 2001).

Par ailleurs, ce projet cadre bien avec l'objectif principal du Réseau Aquaculture Québec (RAQ, dont je suis membre-étudiant), qui est d'amener des chercheurs à étudier des questions présentant un intérêt pour les aquaculteurs et à partager avec eux leurs résultats.

7.3.2 Transcription et plasticité phénotypique du comportement

Les activités humaines favorisent l'établissement d'espèces dans des habitats qu'elles n'occupaient pas au préalable et ces espèces «envahissantes» peuvent grandement perturber les écosystèmes où elles sont introduites, jusqu'à y entraîner l'extinction d'espèces indigènes (Clavero et Garcia-Berthou 2005). Afin d'identifier des acteurs moléculaires impliqués dans la perte plastique de dominance sociale récemment mise en évidence par Blanchet et collaborateurs (2007) chez de jeunes saumons en présence de truite, j'ai comparé le niveau de transcription de 16006 gènes dans les cerveaux de saumons juvéniles dominants ou subordonnés en présence ou en l'absence de truite arc-en-ciel. Les résultats montrent une différence significative du niveau de transcription de 27 gènes (73 clones d'ADN complémentaires) entre saumons dominants et subordonnés en l'absence de truite. Parmi ces gènes, plusieurs codent pour des protéines impliquées dans la dégradation protéique, le transport sanguin d'oxygène ainsi que la différentiation neuronale. En

présence de truite, il n'y a plus qu'un seul gène montrant une différence significative d'expression entre saumons dominants et subordonnés (figure 6.3). La présence de truite rend également les comportements sociaux des saumons subordonnés et dominants plus semblables (figure 6.2). Ces résultats suggèrent qu'il existe un lien entre le niveau de transcription des 27 gènes identifiés (tableau 6.2) et le comportement social des saumons, et aussi que la modification du niveau de transcription de certains de ces gènes pourrait être à la base du changement plastique de comportement social chez les saumons en présence de truite. Ce chapitre contribue ainsi à développer les connaissances sur les liens entre la transcription génique et la plasticité comportementale dans le contexte d'interactions compétitives entre espèces envahissantes et espèces indigènes. En outre, il a été mis en évidence que le statut de dominance d'un saumon juvénile aurait une forte influence sur sa stratégie reproductive future (Metcalfe et coll. 1989, Hofmann et coll. 1999). De ce fait, la présence de truite envahissante pourrait entraîner des changements dans les fréquences respectives des diverses stratégies reproductives possibles dans les populations naturelles de saumon.

7.4 Perspectives

La technique des bio-puces permet de rechercher, à l'échelle du génome entier, des différences du niveau de transcription. D'abord réservée à des études biomédicales, son application à des questions évolutives a permis d'identifier plusieurs gènes dont le contrôle génétique du niveau de transcription aurait évolué dans divers contextes et chez diverses espèces (voir chapitre 1 et 3). Toutefois, la sensibilité de cette technique aux variations expérimentales et le grand nombre de tests statistiques effectués dans l'analyse des résultats peuvent résulter en une proportion substantielle de faux-positifs, et la vérification par une technique alternative, le plus souvent la PCR quantitative, peut permettre d'appuyer les résultats obtenus à l'aide de bio-puces pour les gènes présentant le plus d'intérêt (Draghici 2003). D'autre part, le contrôle du niveau de transcription n'est qu'un des nombreux niveaux de contrôle existant entre les gènes, l'activité des protéines qu'ils encodent et le phénotype. Ainsi, il est fort possible qu'un gène soit transcrit à des niveaux différents sans que ni l'activité de la protéine qu'il encode ni le phénotype ne s'en trouvent changés. Dans ce contexte, si la PCR quantitative peut permettre de confirmer avec de plus grands

échantillons, pour les quelques gènes candidats les plus intéressants, des différences de niveaux de transcription d'abord observées à l'aide de bio-puces, elle n'apporte pas d'information nouvelle quand aux autres niveaux de contrôle de l'expression phénotypique. Ainsi, au chapitre 3, les données obtenues par PCR quantitative avec deux fois plus d'individus étaient équivalentes aux résultats de bio-puces (figure 3.2); Gibson et Weir (2005) écrivent d'ailleurs que, pour des expériences comportant trois répétitions ou plus par traitement, les résultats de bio-puces sont en général fiables et précis. En conséquence, il semble qu'il soit plus intéressant de confirmer les résultats de bio-puces par des techniques de protéomique ou de physiologie, ce qui peut permettre de savoir si la différence observée du niveau de transcription se traduit en une différence dans la quantité ou l'activité de la protéine correspondante. Ainsi, des protocoles permettant le dosage précis de plusieurs protéines existent et pourraient être appliqués afin de vérifier les résultats obtenus pour les principaux candidats identifiés dans les différentes parties de cette thèse. Ultimement, des études futures devraient tenter, pour les gènes candidats les plus prometteurs, d'évaluer comment la variation génétique observée du niveau de transcription influence le phénotype, la survie et la valeur adaptative. De même, le lien entre des niveaux transgressifs de transcription génique (chapitre 3) et le phénotype mériterait d'être exploré pour tenter de résoudre cet apparent paradoxe : les niveaux de transcription géniques semblent généralement sous contrôle génétique polygénique mais non-additif tandis que la plupart des caractères phénotypiques continus seraient sous contrôle génétique polygénique additif (Gibson et Weir 2005).

Il sera également intéressant de tenter d'identifier, dans l'ADN génomique, les variations à la base de différences héréditaires du niveau de transcription. L'identification de polymorphismes dans les séquences en 5' des gènes n'est évidemment pas suffisante pour conclure que ces différences influencent le niveau de transcription. L'identification des régions génomiques régulant en *cis* la transcription d'un gène est un travail de longue haleine (voir, par exemple, Trottier et coll. 1995) et la caractérisation de tous les polymorphismes pouvant entraîner, dans une population naturelle donnée, des différences du niveau de transcription pour un gène donné sera une tâche colossale et pour laquelle il n'existe pas, à ma connaissance, de stratégie expérimentale « à haut débit ». Des études de

liaison génétique entre marqueurs neutres et transcriptome (eQTL) pourraient toutefois faciliter l'identification de régions génomiques importantes dans le contrôle génétique des phénotypes transcriptomiques observés. L'analyse hiérarchique regroupant les gènes selon la similitude de leurs profils d'expression à travers divers échantillons (voir, par exemple, la figure 5.2) pourrait également, dans certains cas, permettre d'identifier des gènes régulateurs responsables en *trans* de plusieurs différences observées au niveau du transcriptome.

Enfin, des résultats récents suggèrent que la régulation épigénétique aurait un rôle central dans la variation et l'évolution des niveaux de transcription génique (Choi et Kim 2008). L'étude des liens entre le transcriptome et les patrons de méthylation de l'ADN et de méthylation et d'acétylation des histones permettrait d'approfondir les connaissances à ce sujet. L'étude de l'influence des micro-ARN sur l'évolution du transcriptome est également prometteuse. En effet, on sait que ces petits ARN régulateurs non-codants découverts dans les années 1990 répriment l'expression de nombreux gènes et sont impliqués dans la différentiation cellulaire, la croissance cellulaire et le métabolisme (Wienholds et coll. 2005, Ambros 2003), notamment.

8. Liste des références citées

Alderman DJ (1994) Control of oomycete pathogens in aquaculture. In: *Salmon saprolegniasis* (ed. Mueller GJ), pp. 111-130. U.S. Department of energy, Portland.

Alderman DJ, Polglase JL (1984) A comparative investigation of the effects of fungicides on *Saprolegnia parasitica* and *Aphanomyces Astaci*. *Transactions of the British Mycological Society*, **83**, 313-318.

Alvarez F, Razquin B, Villena A, Fierro PL, Zapata A (1988) Alterations in the peripheral lymphoid organs and differential leukocyte counts in *Saprolegnia*-infected brown trout, *Salmo trutta fario*. *Veterinary Immunology and Immunopathology*, **18**, 181-193.

Alvarez F, Villena A, Zapata A, Razquin B (1995) Histopathology of the thymus in *Saprolegnia*-infected wild brown trout, *Salmo trutta l*. *Veterinary Immunology and Immunopathology*, **47**, 163-172.

Ambros V (2003) MicroRNA pathways in flies and worms: Growth, death, fat, stress, and timing. *Cell*, **113**, 673-676.

Anderson E., *Introgressive hybridization*, New York : J. Wiley, 1949.

Arthur W.A., *The origin of animal body plans*, Cambridge : Cambridge university press,1997.

Aubin-Horth N, Bourque JF, Daigle G, Hedger R, Dodson JJ (2006) Longitudinal gradients in threshold sizes for alternative male life history tactics in a population of Atlantic salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2067-2075.

Aubin-Horth N, Desjardins JK, Martei YM, Balshine S, Hofmann HA (2007) Masculinized dominant females in a cooperatively breeding species. *Molecular Ecology*, **16**, 1349-1358.

Aubin-Horth N, Letcher BH, Hofmann HA (2005) Interaction of rearing environment and reproductive tactic on gene expression profiles in Atlantic salmon. *Journal of Heredity*, **96**, 261-278.

Auger DL, Gray AD, Ream TS *et al.* (2005) Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics*, **169**, 389-397.

Baumann H, Jahreis GP, Morella KK *et al.* (1991) Transcriptional regulation through cytokine and glucocorticoid response elements of rat acute phase plasma-protein genes by C/Ebp and Junb. *Journal of Biological Chemistry*, **266**, 20390-20399.

Bayne CJ, Gerwick L (2001) The acute phase response and innate immunity of fish. *Developmental and Comparative Immunology*, **25**, 725-743.

Bayne CJ, Gerwick L, Fujiki K, Nakao M, Yano T (2001) Immune-relevant (including acute phase) genes identified in the livers of rainbow trout, Oncorhynchus mykiss, by means of suppression subtractive hybridization. *Developmental and Comparative Immunology*, **25**, 205-217.

Beakes GW, Wood SE, Burr AW (1994) Features which characterize *Saprolegnia* isolates from salmonids fish lesions - A review. In: *Salmon saprolegniasis* (ed. Mueller GJ), pp. 33-66. U.S. Department of Energy, Portland, Oregon.

Beattie JH, Wood AM, Newman AM *et al.* (1998) Obesity and hyperleptinemia in metallothionein (-I and -II) null mice. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 358-363.

Beaumont MA (2005) Adaptation and speciation: what can F-st tell us? *Trends in Ecology & Evolution*, **20**, 435-440.

Bennett E, Stenvers KL, Lund PK, Popko B (1994) Cloning and characterization of a cDNA encoding a novel fatty-acid-binding protein from rat brain. *Journal of Neurochemistry*, **63**, 1616-1624.

Bentsen HB (1991) Quantitative genetics and management of wild populations. *Aquaculture*, **98**, 263-266.

Berthet C, Guehenneux F, Revol V *et al.* (2002) Interaction of PRMT1 with BTG/TOB proteins in cell signalling: molecular analysis and functional aspects. *Genes to Cells*, **7**, 29-39.

Blanchet S, Loot G, Bernatchez L, Dodson JJ (2007) The disruption of dominance hierarchies by a non-native species: an individual-based analysis. *Oecologia*, **152**, 569-581.

Bly JE, Lawson LA, Dale DJ *et al.* (1992) Winter saprolegniosis in channel catfish. *Diseases of Aquatic Organisms*, **13**, 155-164.

Bly JE, Lawson LA, Szalai AJ, Clem LW (1993) Environmental-factors affecting outbreaks of winter saprolegniosis in channel catfish, *Ictalurus punctatus* (rafinesque). *Journal of Fish Diseases*, **16**, 541-549.

Bochdanovits Z, van der Klis H, de Jong G (2003) Covariation of larval gene expression and adult body size in natural populations of Drosophila melanogaster. *Molecular Biology and Evolution*, **20**, 1760-1766.

Bowtell D., Joseph S., *DNA microarrays: a molecular cloning manual*, New York : Cold spring Harbor Laboratory Press, 2003.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752-755.

Byon JY, Ohira T, Hirono I, Aoki T (2005) Use of a cDNA microarray to study immunity against viral hemorrhagic septicemia (VHS) in Japanese flounder (*Paralichthys olivaceus*) following DNA vaccination. *Fish & Shellfish Immunology*, **18**, 135-147.

Carroll RL (2000) Towards a new evolutionary synthesis. *Trends in Ecology & Evolution*, **15**, 27-32.

Chesler EJ, Lu L, Shou SM *et al.* (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics*, **37**, 233-242.

Chiavegatto S, Nelson RJ (2003) Interaction of nitric oxide and serotonin in aggressive behavior. *Hormones and Behavior*, **44**, 233-241.

Choi JK, Kim YJ (2008) Epigenetic regulation and the variability of gene expression. *Nature Genetics*, **40**, 141-147.

Clavero M, Garcia-Berthou E (2005) Invasive species are a leading cause of animal extinctions. *Trends in Ecology & Evolution*, **20**, 110.

Coltman DW, O'Donoghue P, Jorgenson JT *et al.* (2003) Undesirable evolutionary consequences of trophy hunting. *Nature*, **426**, 655-658.

Conover DO, Munch SB (2002) Sustaining fisheries yields over evolutionary time scales. *Science*, **297**, 94-96.

Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in qene expression after 20,000 generations of evolution in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1072-1077.

Crawley J.M., *Statistics: an introduction using R*, Chichester : John Wiley and sons,2007.

Cui XQ, Affourtit J, Shockley KR, Woo Y, Churchill GA (2006) Inheritance patterns of transcript levels in F-1 hybrid mice. *Genetics*, **174**, 627-637.

Davis DR, Epp MD, Riordan HD (2004) Changes in USDA food composition data for 43 garden crops, 1950 to 1999. *Journal of the American College of Nutrition*, **23**, 669-682.

Dennis G, Sherman BT, Hosack DA *et al.* (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology*, **4**.

Denver DR, Morris K, Streelman JT *et al.* (2005) The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. *Nature Genetics*, **37**, 544-548.

Derome N, Duchesne P, Bernatchez L (2006) Parallelism in gene transcription among sympatric lake whitefish (Coregonus clupeaformis Mitchill) ecotypes. *Molecular Ecology*, **15**, 1239-1249.

Didier DK, Schiffenbauer J, Woulfe SL, Zacheis M, Schwartz BD (1988) Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class-Ii Y-box. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 7322-7326.

Dierick HA, Greenspan RJ (2006) Molecular analysis of flies selected for aggressive behavior. *Nature Genetics*, **38**, 1023-1031.

Dobzhansky T., *Evolution*, San Francisco : W.H. Freeman,1977.

Draghici S., *Data analysis tools for DNA microarrays*, London : CRC Press, 2003.

Dudoit S, Gendeman RC, Quackenbush J (2003) Open source software for the analysis of microarray data. *Biotechniques*, 45-51.

Edwards AC, Rollmann SM, Morgan TJ, Mackay TFC (2006) Quantitative genomics of aggressive behavior in Drosophila melanogaster. *Plos Genetics*, **2**, 1386-1395.

Ellis AE (2001) Innate host defense mechanisms of fish against viruses and bacteria. *Developmental and Comparative Immunology*, **25**, 827-839.

Enard W, Khaitovich P, Klose J *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science*, **296**, 340-343.

Ewart KV, Belanger JC, Williams J *et al.* (2005) Identification of genes differentially expressed in Atlantic salmon (*Salmo salar*) in response to infection by *Aeromonas salmonicida* using cDNA microarray technology. *Developmental and Comparative Immunology*, **29**, 333-347.

Falconer D.S., Mackay T.F.C., *Introduction to quantitative genetics*, Essex : Longman, 1996.

FAO, FISHSTAT Plus. 2004.

Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Reviews Genetics*, **4**, 651-657.

Fehling HJ, Swat W, Laplace C *et al.* (1994) MHC class-I expression in mice lacking the proteasome subunit lmp-7. *Science*, **265**, 1234-1237.

Ferea TL, Botstein D, Brown PO, Rosenzweig RF (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 9721-9726.

Fisher L.D., van Belle G., *Biostatistics: a methodology for health sciences*, New York, 1993.

Fitzpatrick MJ, Ben Shahar Y, Smid HM *et al.* (2005) Candidate genes for behavioural ecology. *Trends in Ecology & Evolution*, **20**, 96-104.

Fleming IA, Agustsson T, Finstad B, Johnsson JI, Bjornsson BT (2002) Effects of domestication on growth physiology and endocrinology of Atlantic salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 1323-1330.

Fleming IA, Einum S (1997) Experimental tests of genetic divergence of farmed from wild Atlantic salmon due to domestication. *Ices Journal of Marine Science*, **54**, 1051-1063.

Fleming IA, Hindar K, Mjolnerod IB *et al.* (2000) Lifetime success and interactions of farm salmon invading a native population. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267**, 1517-1523.

Friars GW, Bailey JK, Oflynn FM (1995) Applications of selection for multiple traits in cage-reared Atlantic salmon (Salmo salar). *Aquaculture*, **137**, 213-217.

Fujiki K, Bayne CJ, Shin DH, Nakao M, Yano T (2001) Molecular cloning of carp (*Cyprinus carpio*) C-type lectin and pentraxin by use of suppression subtractive hybridisation. *Fish & Shellfish Immunology*, **11**, 275-279.

Fujiki K, Gerwick L, Bayne CJ *et al.* (2003) Molecular cloning and characterization of rainbow trout (*Oncorhynchus mykiss*) CCAAT/enhancer binding protein beta. *Immunogenetics*, **55**, 253-261.

Futuyma D.J., *Evolutionary biology*, Sunderland : Sinauer Associates, 1986.

Garant D, Dodson JJ, Bernatchez L (2000) Ecological determinants and temporal stability of the within-river population structure in Atlantic salmon (Salmo salar L.). *Molecular Ecology*, **9**, 615-628.

Garant D, Dodson JJ, Bernatchez L (2003a) Differential reproductive success and heritability of alternative reproductive tactics in wild Atlantic salmon (Salmo salar L.). *Evolution*, **57**, 1133-1141.

Garant D, Fleming IA, Einum S, Bernatchez L (2003b) Alternative male life-history tactics as potential vehicles for speeding introgression of farm salmon traits into wild populations. *Ecology Letters*, **6**, 541-549.

Gharrett AJ, Smoker WW, Reisenbichler RR, Taylor SG (1999) Outbreeding depression in hybrids between odd- and even-broodyear pink salmon. *Aquaculture*, **173**, 117-129.

Gibson G, Riley-Berger R, Harshman L *et al.* (2004) Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. *Genetics*, **167**, 1791-1799.

Gibson G, Weir B (2005) The quantitative genetics of transcription. *Trends in Genetics*, **21**, 616-623.

Giger T, Excoffier L, Day PJR *et al.* (2006) Life history shapes gene expression in salmonids. *Current Biology*, **16**, R281-R282.

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, **440**, 242-245.

Giles PJ, Kipling D (2003) Normality of oligonucleotide microarray data and implications for parametric statistical analyses. *Bioinformatics*, **19**, 2254-2262.

Gilk SE, Wang IA, Hoover CL *et al.* (2004) Outbreeding depression in hybrids between spatially separated pink salmon, Oncorhynchus gorbuscha, populations: marine

survival, homing ability, and variability in family size. *Environmental Biology of Fishes*, **69**, 287-297.
Gilmour KM, DiBattista JD, Thomas JB (2005) Physiological causes and consequences of social status in salmonid fish. *Integrative and Comparative Biology*, **45**, 263-273.
Gjedrem T, Gjoen HM, Gjerde B (1991) Genetic origin of norwegian farmed Atlantic salmon. *Aquaculture*, **98**, 41-50.
Gjoen HM, Bentsen HB (1997) Past, present, and future of genetic improvement in salmon aquaculture. *Ices Journal of Marine Science*, **54**, 1009-1014.
Goodson JL (2005) The vertebrate social behavior network: Evolutionary themes and variations. *Hormones and Behavior*, **48**, 11-22.
Grant PR, Grant BR (1995) Predicting microevolutionary responses to directional selection on heritable variation. *Evolution*, **49**, 241-251.
Groeneveld E, Kovac M, Wang T (1990) PEST, a general purpose BLUP package for multivariate prediction and estimation. In: *Proceedings of the 4th world congress on genetics applied to livestock production* pp. 488-491. Edinburgh.
Gross MR (1996) Alternative reproductive strategies and tactics: Diversity within sexes. *Trends in Ecology & Evolution*, **11**, A92-A98.
Gross MR (1998) One species with two biologies: Atlantic salmon (Salmo salar) in the wild and in aquaculture. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 131-144.
Grozinger CM, Robinson GE (2002) Microarray analysis of pheromone-mediated gene expression in the honey bee brain. *Integrative and Comparative Biology*, **42**, 1237.
Haldane J.B.S., *The causes of evolution*, London : Longmans, 1932.
Hart P.J.B., Reynolds J.D., *Handbook of fish biology and fisheries* , Malden : Blackwell Pub.,2002.
Hartigan JA, Hartigan PM (1985) The Dip Test of Unimodality. *Annals of Statistics*, **13**, 70-84.
Harvey P.H., Pagel M.D., *The comparative method in evolutionary biology*, Oxford1991.
Hatai K., Hoshia G.-I. (1994) Pathogenicity of *Saprolegnia parasitica* Coker. In: *Salmon saprolegniasis* (ed. Mueller GJ), pp. 87-98. U.S. Department of Energy, Portland.
Hedgecock D, Lin JZ, DeCola S *et al.* (2007) Transcriptomic analysis of growth heterosis in larval Pacific oysters (Crassostrea gigas). *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2313-2318.
Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633-1638.
Hendry AP, Farrugia TJ, Kinnison MT (2008) Human influences on rates of phenotypic change in wild animal populations. *Molecular Ecology*, **17**, 20-29.
Hendry AP, Kinnison MT (2001) An introduction to microevolution: rate, pattern, process. *Genetica*, **112**, 1-8.
Hendry AP, Wenburg JK, Bentzen P, Volk EC, Quinn TP (2000) Rapid evolution of reproductive isolation in the wild: Evidence from introduced salmon. *Science*, **290**, 516-518.
Hindar K, Fleming IA, McGinnity P, Diserud A (2006) Genetic and ecological effects of salmon farming on wild salmon: modelling from experimental results. *Ices Journal of Marine Science*, **63**, 1234-1247.
Hofmann HA (2003) Functional genomics of neural and behavioral plasticity. *Journal of Neurobiology*, **54**, 272-282.

Hofmann HA, Benson ME, Fernald RD (1999) Social status regulates growth rate: Consequences for life-history strategies. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 14171-14176.

Homey B, Alenius H, Muller A *et al.* (2002) CCL27-CCR10 interactions regulate T cell-mediated skin inflammation. *Nature Medicine*, **8**, 157-165.

Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. *Genome Biology*, **4**.

Hsiao A, Ideker T, Olefsky JM, Subramaniam S (2005) VAMPIRE microarray suite: a web-based platform for the interpretation of gene expression data. *Nucleic Acids Research*, **33**, W627-W632.

Hubner N, Wallace CA, Zimdahl H *et al.* (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genetics*, **37**, 243-253.

Hughes G.C. (1994) Saprolegniasis: Then and now: A retrospective. In: *Salmon saprolegniasis* (ed. Mueller GJ), pp. 3-34. U.S. Department of Energy, Portland.

Iwama G., Nakanishi T., *The fish immune system*, San Diego : Academic Press, 1996.

Jensen LE, Hiney MP, Shields DC *et al.* (1997) Acute phase proteins in salmonids - Evolutionary analyses and acute phase response. *Journal of Immunology*, **158**, 384-392.

Jin W, Riley RM, Wolfinger RD *et al.* (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*, **29**, 389-395.

Kamoun S (2003) Molecular genetics of pathogenic Oomycetes. *Eukaryotic Cell*, **2**, 191-199.

Kerr MK, Afshari CA, Bennett L *et al.* (2002) Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinica*, **12**, 203-217.

Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, **7**, 819-837.

Kettlewell B., *The evolution of melanism*, Oxford : Clarendon Press, 1973.

Khaitovich P, Paabo S, Weiss G (2005) Toward a neutral evolutionary model of gene expression. *Genetics*, **170**, 929-939.

Khatri P, Draghici S, Ostermeier GC, Krawetz SA (2002) Profiling gene expression using onto-express. *Genomics*, **79**, 266-270.

King MC, Wilson AC (1975) Evolution at 2 levels in humans and chimpanzees. *Science*, **188**, 107-116.

Kirst M, Myburg AA, De Leon JPG *et al.* (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiology*, **135**, 2368-2378.

Knudsen S., *A biologist's guide to analysis of DNA microarray data*, New York : J. Wiley, 2002.

Kocabas AM, Li P, Cao DF *et al.* (2002) Expression profile of the channel catfish spleen: Analysis of genes involved in immune functions. *Marine Biotechnology*, **4**, 526-536.

Kono T, Ponpornpisit A, Sakai M (2004) The analysis of expressed genes in head kidney of common carp *Cyprinus carpio* L. stimulated with peptidoglycan. *Aquaculture*, **235**, 37-52.

Kono T, Sakai M (2001) The analysis of expressed genes in the kidney of Japanese flounder, *Paralichthys olivaceus*, injected with the immunostimulant peptidoglycan. *Fish & Shellfish Immunology*, **11**, 357-366.

Koskinen MT, Haugen TO, Primmer CR (2003) Contemporary fisherian life-history evolution in small salmonid populations. *Nature*, **421**, 656.

Kovac M, Groeneveld E, Garcia-Cortés LA (2002) VCE-5: A package for the optimization of dispersion parameters. In: *7th world congress on genetics applied to livestock production* Montpellier, France.

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 873-890.

Kurobe T, Yasuike M, Kimura T, Hirono K, Aoki T (2005) Expression profiling of immune-related genes from Japanese flounder *Paralichthys olivaceus* kidney cells using cDNA microarrays. *Developmental and Comparative Immunology*, **29**, 515-523.

Landry C, Bernatchez L (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (Salmo salar). *Molecular Ecology*, **10**, 2525-2539.

Lawler JJ, Aukema JE, Grant JB *et al.* (2006) Conservation science: a 20-year report card. *Frontiers in Ecology and the Environment*, **4**, 473-480.

Leary RF, Allendorf FW, Knudsen KL (1983) Consistently High Meristic Counts in Natural Hybrids Between Brook Trout and Bull Trout. *Systematic Zoology*, **32**, 369-376.

Lemos B, Meiklejohn CD, Caceres M, Hartl DL (2005) Rates of divergence in gene expression profiles of primates, mice, and flies: Stabilizing selection and variability among functional categories. *Evolution*, **59**, 126-137.

Leprieur F, Beauchard O, Blanchet S, Oberdorff T, Brosse S (2008) Fish invasions in the world's river systems: When natural processes are blurred by human activities. *PLoS Biology*, **In press**.

Leung YF, Cavalieri D (2003) Fundamentals of cDNA microarray data analysis. *Trends in Genetics*, **19**, 649-659.

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. *Trends in Genetics.*, **23**, 60-66.

Liu L, Hawkins DM, Ghosh S, Young SS (2003) Robust singular value decomposition analysis of microarray data. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 13167-13172.

Lodge DM (1993) Biological Invasions - Lessons for Ecology. *Trends in Ecology & Evolution*, **8**, 133-137.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981-994.

Lund SG, Phillips MCL, Moyes CD, Tufts BL (2000) The effects of cell ageing on protein synthesis in rainbow trout (Oncorhynchus mykiss) red blood cells. *Journal of Experimental Biology*, **203**, 2219-2228.

Mack RN, Simberloff D, Lonsdale WM *et al.* (2000) Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689-710.

Mackay TFC, Heinsohn SL, Lyman RF *et al.* (2005) Genetics and genomics of Drosophila mating behavior. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6622-6629.

Maecher M., Ringach D (2004). The diptest package.

McGinnity P, Prodohl P, Ferguson K *et al.* (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, Salmo salar, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 2443-2450.

Meijer AH, Verbeek FJ, Salas-Vidal E *et al.* (2005) Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to *Mycobacterium marinum* infection. *Molecular Immunology*, **42**, 1185-1203.

Meiklejohn CD, Townsend JP (2005) A Bayesian method for analysing spotted microarray data. *Briefings in Bioinformatics*, **6**, 318-330.

Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892-903.

Merila J, Sheldon BC, Kruuk LEB (2001) Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112**, 199-222.

Metcalfe NB, Huntingford FA, Graham WD, Thorpe JE (1989) Early Social-Status and the Development of Life-History Strategies in Atlantic Salmon. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **236**, 7-19.

Meyer FP, Jorgenson TA (1983) Teratological and other effects of malachite green on development of rainbow-trout and rabbits. *Transactions of the American Fisheries Society*, **112**, 818-824.

Miczek KA, Fish EW, De Bold JF (2003) Neurosteroids, GABA(A) receptors, and escalated aggressive behavior. *Hormones and Behavior*, **44**, 242-257.

Mills JC, Roth KA, Cagan RL, Gordon JI (2001) DNA microarrays and beyond: completing the journey from tissue to cell. *Nature Cell Biology*, **3**, E175-E178.

Monks SA, Leonardson A, Zhu H *et al.* (2004) Genetic inheritance of gene expression in human cell lines. *American Journal of Human Genetics*, **75**, 1094-1105.

Morley M, Molony CM, Weber TM *et al.* (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743-747.

Muller EE, Locatelli V, Cocchi D (1999) Neuroendocrine control of growth hormone secretion. *Physiological Reviews*, **79**, 511-607.

Nam BH, Yamamoto E, Hirono I, Aoki T (2000) A survey of expressed genes in leukocytes of Japanese flounder, *Paralichthys olivaceus* infected with Hirame rhabdovirus. *Developmental and Comparative Immunology*, **24**, 13-24.

Naylor R, Hindar K, Fleming IA *et al.* (2005) Fugitive salmon: Assessing the risks of escaped fish from net-pen aquaculture. *Bioscience*, **55**, 427-437.

Neish GA, Green BR (1977) Fractionation of *Saprolegnia diclina* (oomycetes) satellite dnas by $AgNO_3$-Cs_2SO_4 density gradient centrifugation. *Biochimica et Biophysica Acta*, **479**, 411-415.

Nelson RJ, Chiavegatto S (2001) Molecular basis of aggression. *Trends in Neurosciences*, **24**, 713-719.

Nemes JP, Benzow KA, Koob MD (2000) The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). *Human Molecular Genetics*, **9**, 1543-1551.

Norris AT, Bradley DG, Cunningham EP (1999) Microsatellite genetic variation between and within farmed and wild Atlantic salmon (Salmo salar) populations. *Aquaculture*, **180**, 247-264.

O'Flynn FM, Bailey JK, Friars GW (1999) Responses to two generations of index selection in Atlantic salmon (Salmo salar). *Aquaculture*, **173**, 143-147.

Ohno S., *Evolution by gene duplication*, New York : Springer-Verlag, 1970.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics*, **32**, 261-266.

Oleksiak MF, Kolell KJ, Crawford DL (2001) Utility of natural populations for microarray analyses: Isolation of genes necessary for functional genomic studies. *Marine Biotechnology*, **3**, S203-S211.

Oleksiak MF, Roach JL, Crawford DL (2005) Natural variation in cardiac metabolism and gene expression in Fundulus heteroclitus. *Nature Genetics*, **37**, 67-72.

Oliveira RF, Lopes M, Carneiro LA, Canario AVM (2001) Watching fights raises fish hormone levels - Cichlid fish wrestling for dominance induce an androgen surge in male spectators. *Nature*, **409**, 475.

Osaka H, Wang YL, Takada K *et al.* (2003) Ubiquitin carboxy-terminal hydrolase L1 binds to and stabilizes monoubiquitin in neuron. *Human Molecular Genetics*, **12**, 1945-1958.

Overli O, Harris CA, Winberg S (1999) Short-term effects of fights for social dominance and the establishment of dominant-subordinate relationships on brain monoamines and cortisol in rainbow trout. *Brain Behavior and Evolution*, **54**, 263-275.

Owens IPF (2006) Where is behavioural ecology going? *Trends in Ecology & Evolution*, **21**, 356-361.

Palumbi S.R., *The evolution explosion*, New York : Norton, 2001.

Perry GML, Audet C, Bernatchez L (2005) Maternal genetic effects on adaptive divergence between anadromous and resident brook charr during early life history. *Journal of Evolutionary Biology*, **18**, 1348-1361.

Phillips RB, Hartley SE (1988) Fluorescent Banding-Patterns of the Chromosomes of the Genus Salmo. *Genome*, **30**, 193-197.

Pickering AD (1994) Factors influencing the susceptibility of salmonid fish to saprolegniasis. In: *Salmon saprolegniasis* (ed. Mueller GJ), pp. 67-86. U.S. Department of energy, Portland.

Pigliucci M., *Phenotypic plasticity*, Baltimore : The Johns Hopkins University Press, 2001.

Pigliucci M (2005) Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution*, **20**, 481-486.

Posthuma L, Vanstraalen NM (1993) Heavy metal adaptation in terrestrial invertebrates - A review of occurrence, genetics, physiology and ecological consequences. *Comparative Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology*, **106**, 11-38.

Potokina E, Druka A, Luo ZW *et al.* (2008) Gene expression quantitative trait locus analysis of 16,000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant Journal*, **53**, 90-101.

Price TD, Qvarnstrom A, Irwin DE (2003) The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 1433-1440.

Quackenbush J (2002) Microarray data normalization and transformation. *Nature Genetics*, **32**, 496-501.

Quinn TP, Kinnison MT, Unwin MJ (2001) Evolution of chinook salmon (Oncorhynchus tshawytscha) populations in New Zealand: pattern, rate, and process. *Genetica*, **112**, 493-513.

Quinton CD, McMillan I, Glebe B (2004) Development of an Atlantic salmon genetic improvement program: Genetic parameters of harvest body weight and carcass quality traits estimated with animal models.

Randi E (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, **17**, 285-293.

Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ (1998) Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science*, **56**, 15-33.

Rauw WM, Luiting P, Bakken M *et al.* (2000) Behavioural differences in non-reproductive adult females in a long-term selection experiment for litter size in mice. *Applied Animal Behaviour Science*, **66**, 249-262.

Remington DL, Purugganan MD (2003) Candidate genes, quantitative trait loci, and functional trait evolution in plants. *International Journal of Plant Sciences*, **164**, S7-S20.

Resibois A, Cuvelier L, Svoboda M, Heizmann CW, Thony B (1999) Immunohistochemical localisation of pterin-4 alpha-carbinolamine dehydratase in rat peripheral organs. *Histochemistry and Cell Biology*, **111**, 381-390.

Richards RH, Pickering AD (1979) Changes in serum parameters of *Saprolegnia*-infected Brown Trout, *Salmo trutta l*. *Journal of Fish Diseases*, **2**, 197-206.

Ridley M., *Evolution*, Malden, 2004.

Rieseberg LH, Archer MA, Wayne RK (1999) Transgressive segregation, adaptation and speciation. *Heredity*, **83**, 363-372.

Rifkin SA, Houle D, Kim J, White KP (2005) A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature*, **438**, 220-223.

Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the Drosophila melanogaster subgroup. *Nature Genetics*, **33**, 138-144.

Rise ML, Jones SRM, Brown GD *et al.* (2004a) Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiological Genomics*, **20**, 21-35.

Rise ML, von Schalburg KR, Brown GD *et al.* (2004b) Development and application of a salmonid EST database and cDNA microarray: Data mining and interspecific hybridization characteristics. *Genome Research*, **14**, 478-490.

Roberge C, Einum S, Guderley H, Bernatchez L (2006) Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. *Molecular Ecology*, **15**, 9-20.

Roberge C, Paez DJ, Rossignol O *et al.* (2007) Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology*, **44**, 1374-1383.

Rye M, Gjerde B (1996) Phenotypic and genetic parameters of body composition traits and flesh colour in Atlantic salmon, *Salmo salar* L. *Aquaculture Research*, **27**, 121-133.
Savan R, Sakai M (2002) Analysis of expressed sequence tags (EST) obtained from common carp, *Cyprinus carpio* L., head kidney cells after stimulation by two mitogens, lipopolysaccharide and concanavalin-A. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, **131**, 71-82.
Schadt EE, Monks SA, Drake TA *et al.* (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature*, **422**, 297-302.
Schluter D (2000) Ecological character displacement in adaptive radiation. *American Naturalist*, **156**, S4-S16.
Schluter D, Clifford EA, Nemethy M, McKinnon JS (2004) Parallel evolution and inheritance of quantitative traits. *American Naturalist*, **163**, 809-822.
Schulze A, Downward J (2001) Navigating gene expression using microarrays - a technology review. *Nature Cell Biology*, **3**, E190-E195.
Shieh GS, Fan TH, Chu HP (2008) A Bayesian approach to assessing differential expression of microarray data. *Journal of Statistical Computation and Simulation*, **78**, 179-191.
Silphaduang U, Hincke MT, Nys Y, Mine Y (2006) Antimicrobial proteins in chicken reproductive system. *Biochemical and Biophysical Research Communications*, **340**, 648-655.
Sloman KA, Wilson L, Freel JA *et al.* (2002) The effects of increased flow rates on linear dominance hierarchies and physiological function in brown trout, Salmo trutta. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*, **80**, 1221-1227.
Smith TB, Bernatchez L (2008) Evolutionary change in human-altered environments. *Molecular Ecology*, **17**, 1-8.
Sneddon LU, Margareto J, Cossins AR (2005) The use of transcriptomics to address questions in behaviour: Production of a suppression subtractive hybridisation library from dominance hierarchies of rainbow trout. *Physiological and Biochemical Zoology*, **78**, 695-705.
Sorensen C, Overli O, Summers CH, Nilsson GE (2007) Social regulation of neurogenesis in teleosts. *Brain Behavior and Evolution*, **70**, 239-246.
Speckner W, Schindler JF, Albers C (1989) Age-dependent changes in volume and hemoglobin content of erythrocytes in the carp (Cyprinus carpio L). *Journal of Experimental Biology*, **141**, 133-149.
Speed T., *Statistical analysis of gene expression microarray data*, Boca Raton : Chapman & Hall, 2003.
Spitze K (1993) Population-Structure in Daphnia-Obtusa - Quantitative Genetic and Allozymic Variation. *Genetics*, **135**, 367-374.
St-Cyr J, Derome N, Bernatchez L (2008) The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Molecular Ecology*, **In press**.
Stahl G (1987) Genetic Population Structure of Atlantic Salmon. (eds. Ryman N, Utter F), University of Washington Press, Seattle.
Stamatoyannopoulos JA (2004) The genomics of gene expression. *Genomics*, **84**, 449-457.
Stern DL (2000) Perspective: Evolutionary developmental biology and the problem of variation. *Evolution*, **54**, 1079-1091.
Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology and Evolution*, **18**, 94-101.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **64**, 479-498.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440-9445.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671-688.

Strauss SY, Lau JA, Carroll SP (2006) Evolutionary responses of natives to introduced species: what do introductions tell us about natural communities? *Ecology Letters*, **9**, 354-371.

Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. *Molecular Ecology*, **17**, 351-360.

Swanson-Wagner RA, Jia Y, DeCook R *et al.* (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F-1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6805-6810.

Taberlet P, Valentini A, Rezaei HR *et al.* (2008) Are cattle, sheep, and goats endangered species? *Molecular Ecology*, **17**, 275-284.

Tampieri MP, Galuppi R, Carelle MS *et al.* (2003) Effect of selected essential oils and pure compounds on *Saprolegnia parasitica*. *Pharmaceutical Biology*, **41**, 584-591.

Thodesen J, Grisdale-Helland B, Helland SJ, Gjerde B (1999) Feed intake, growth and feed utilization of offspring from wild and selected Atlantic salmon (Salmo salar). *Aquaculture*, **180**, 237-246.

Thorstad EB, Heggberget TG, Okland F (1998) Migratory behaviour of adult wild and escaped farmed Atlantic salmon, Salmo salar L., before, during and after spawning in a Norwegian river. *Aquaculture Research*, **29**, 419-428.

Torto-Alalibo T, Tian MY, Gajendran K *et al.* (2005) Expressed sequence tags from the oomycete fish pathogen *Saprolegnia parasitica* reveal putative virulence factors. *Bmc Microbiology*, **5**.

Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. *Molecular Biology and Evolution*, **20**, 955-963.

Trainor BC, Hofmann HA (2007) Somatostatin and somatostatin receptor gene expression in dominant and subordinate males of an African cichlid fish. *Behavioural Brain Research*, **179**, 314-320.

Trottier E, Belzil A, Stoltz C, Anderson A (1995) Localization of A Phenobarbital-Responsive Element (Pbre) in the 5'-Flanking Region of the Rat Cyp2B2 Gene. *Gene*, **158**, 263-268.

Tsoi SCM, Ewart KV, Penny S *et al.* (2004) Identification of immune-relevant genes from Atlantic salmon using suppression subtractive hybridization. *Marine Biotechnology*, **6**, 199-214.

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5116-5121.

Tymchuk WE, Biagi C, Withler R, Devlin RH (2006) Growth and behavioral consequences of introgression of a domesticated aquaculture genotype into a native strain of Coho salmon. *Transactions of the American Fisheries Society*, **135**, 442-455.

Tymchuk WE, Devlin RH (2005) Growth differences among first and second generation hybrids of domesticated and wild rainbow trout (Oncorhynchus mykiss). *Aquaculture*, **245**, 295-300.

Tzachanis D, Freeman GJ, Hirano N *et al.* (2001) Tob is a negative regulator of activation that is expressed in anergic and quiescent T cells. *Nature Immunology*, **2**, 1174-1182.

Uddin M, Wildman DE, Liu GZ *et al.* (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 2957-2962.

Van der Waaij EH (2004) A resource allocation model describing consequences of artificial selection under metabolic stress. *Journal of Animal Science*, **82**, 973-981.

Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623-3642.

Verspoor E, McCarthy E (1997) Genetic divergence at the NAD+-dependent malic enzyme locus in Atlantic salmon from Europe and north America. *Journal of Fish Biology*, **51**, 155-163.

Voet D, Voet JG, *Biochemistry*, New York : J.Wiley & Sons, 2004.

Von Schalburg KR, Rise ML, Cooper GA *et al.* (2005) Fish and chips: Various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *Bmc Genomics*, **6**.

Vuylsteke M, van Eeuwijk F, Van Hummelen P, Kuiper M, Zabeau M (2005) Genetic analysis of variation in gene expression in Arabidopsis thaliana. *Genetics*, **171**, 1267-1275.

Waddington CH (1961) Genetic assimilation. *Advances in genetics*, **10**, 257-293.

West MAL, Kim K, Kliebenstein DJ *et al.* (2007) Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics*, **175**, 1441-1450.

West-Eberhard M.J., *Developmental plasticity and evolution*, New York : Oxford University Press, 2003.

West-Eberhard MJ (2005) Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6543-6549.

Whisler HC, *Identification of Saprolegnia ssp. pathogenic in chinook salmon,* Portland, U.S. Department of energy, 1996.

Whitehead A, Crawford DL (2006) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197-1211.

Whitfield CW, Band MR, Bonaldo MF *et al.* (2002) Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Research*, **12**, 555-566.

Wienholds E, Kloosterman WP, Miska E *et al.* (2005) MicroRNA expression in zebrafish embryonic development. *Science*, **309**, 310-311.

Wilson AC (1976) Gene regulation in evolution. In: *Molecular evolution* (ed. Ayala FJ), pp. 225-234. Sinauer associates, inc, Sunderland.

Winberg S, Nilsson GE (1993) Roles of Brain Monoamine Neurotransmitters in Agonistic Behavior and Stress Reactions, with Particular Reference to Fish. *Comparative*

Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology, **106**, 597-614.

Wingfield JC, Hegner RE, Dufty AM, Ball GF (1990) The Challenge Hypothesis - Theoretical Implications for Patterns of Testosterone Secretion, Mating Systems, and Breeding Strategies. *American Naturalist*, **136**, 829-846.

Wolfinger RD, Gibson G, Wolfinger ED *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, **8**, 625-637.

Yang HS, Henning D, Valdez BC (2005) Functional interaction between RNA helicase II Gu alpha and ribosomal protein L4. *Febs Journal*, **272**, 3788-3802.

Yano T (1996) The nonspecific immune system: humoral defense. In: *The immune fish system* (eds. Iwama G, Nakanishi T), pp. 106-159. Academic Press, San Diego.

Yvert G, Brem RB, Whittle J *et al.* (2003) Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nature Genetics*, **35**, 57-64.

Zar J.H., *Biostatistical analysis*, Upper Saddle River : Prentice Hall International, 1999.

9. Annexes

Supplementary Table 2.1 PNRS representing genes differentially expressed between the progeny of farmed and wild Canadian salmons. PNRS are classified in various biological/cellular functional groups. The average fold change in gene expression in farmed salmon compared to wild individuals (data from 13 arrays) and its standard deviation is given for each PNRS, as well as P-values obtained from the F-test on the mixed model of ANOVA. Genes for which several PNRS were significant are in bold. Colors identify genes or functional entities showing parallel changes in Canada and Norway. Clone identification follows notation from the GRASP. GeneBank accession numbers refer to the GeneBank sequence the most similar to the amplified cDNA portion spotted on the array.

Function		PNRS	Average fold change	Fold change SD	Tabulated P-value	Permutation corrected Pvalue
Energy metabolism						
	Glycolysis	rgb^523^238^[N] [AY024367] Salmo salar glycerol-3-phosphate dehydrogenase	0,86	0,11	$5,45X10^{-5}$	$2,16X10^{-4}$
		nwh^13^82^[P] [AAB82747] glyceraldehyde-3-phosphate dehydrogenase	0,86	0,12	$1,61X10^{-4}$	$4,41X10^{-4}$
		lna^17^81^[P] [AAK49985] glyceraldehyde phosphate dehydrogenase	0,89	0,13	$1,99X10^{-3}$	$2,84X10^{-3}$
		nwh^6^30^[P] [AAB82747] glyceraldehyde-3-phosphate dehydrogenase	0,83	0,08	$2,44X10^{-7}$	$1,22X10^{-5}$
		lna^2^51^[N] [AF360980] glyceraldehyde phosphate dehydrogenase	0,86	0,13	$4,15X10^{-4}$	$8,38X10^{-4}$
		nwh^19^56^[P] [NP_036627] (NM_012495) aldolase A, fructose-bisphosphate	0,86	0,13	$3,45X10^{-4}$	$7,38X10^{-4}$
	Citric acid cycle					
		rgb^532^199^[P] [Q04467] ISOCITRATE DEHYDROGENASE	0,89	0,12	$1,51X10^{-3}$	$2,29X10^{-3}$
		rgb^510^291^[P] [XP_034795] malate dehydrogenase 2, NAD (mitochondrial)	0,87	0,08	$1,60X10^{-4}$	$4,40X10^{-4}$
	Oxidative phosphorylation					
		lna^13^4^[P] [NP_031773] (NM_007747) cytochrome c oxidase, subunit Va	0,91	0,13	$8,12X10^{-3}$	$9,21X10^{-3}$
		nwh^19^38^[P] [3CYT_I] Chain I, Cytochrome c	0,84	0,13	$4,65X10^{-5}$	$1,92X10^{-4}$
		rgb^505^196^[P] [3CYT_I] Chain I, Cytochrome c	0,84	0,12	$2,20X10^{-6}$	$3,33X10^{-5}$
		rgb^529^51^[P] [3CYT_I] Chain I, Cytochrome c	0,90	0,12	$5,58X10^{-3}$	$6,69X10^{-3}$
		plnb^502^12^[P] [P56384] ATP synthase lipid-binding protein	0,87	0,10	$2,70X10^{-4}$	$6,30X10^{-4}$
		rgb^510^161^[P] [P05631] ATP SYNTHASE GAMMA CHAIN	0,90	0,07	$2,93X10^{-3}$	$3,90X10^{-3}$
	Fatty acid metabolism					
		lna^14^6^[P] [AAK58094] (AF380998) liver basic fatty acid binding protein	1,16	0,18	$3,39X10^{-3}$	$4,42X10^{-3}$
Metal ions sequestration/transport						
		pitl^505^8^[P] [P49947] FERRITIN, MIDDLE SUBUNIT (FERRITIN M)	1,14	0,23	$7,46X10^{-3}$	$8,54X10^{-3}$
		rtch^1^5^[P] [BAA13148] ferritin H-3 [Oncorhynchus mykiss]	1,17	0,22	$1,88X10^{-3}$	$2,72X10^{-3}$
		rtch^1^9^[P] [BAA13147] ferritin H-2 [Oncorhynchus mykiss]	1,16	0,28	$7,91X10^{-3}$	$9,00X10^{-3}$
		rtah^1^16^[N] [AF013800] Salvelinus alpinus metallothionein A (MT A)	0,72	0,14	$3,14X10^{-11}$	$1,18X10^{-6}$
		nwh^20^24^[N] [X97274] S.salar mRNA for metallothionein A	0,81	0,17	$1,88X10^{-5}$	$1,13X10^{-4}$
		pitl^505^292^[N] [X97274] S.salar mRNA for metallothionein A	0,81	0,19	$1,68X10^{-5}$	$1,06X10^{-4}$
		lna^11^94^[N] [X97274] S.salar mRNA for metallothionein A	0,85	0,10	$2,64X10^{-5}$	$1,37X10^{-4}$
		pitl^501^263^[N] [X97274] S.salar mRNA for metallothionein A	0,68	0,20	$7,22X10^{-9}$	$2,75X10^{-6}$
		lna^11^94^[N] [X97274] S.salar mRNA for metallothionein A	0,87	0,18	$1,15X10^{-3}$	$1,85X10^{-3}$

		nwh^16^40^[P] [AAG37428] (AY009128) ISCU2	1,24	0,28	$1,45X10^{-4}$	$4,14X10^{-4}$
		pha^501^73^[P] [AAG53688] (AF322071) selenoprotein Pa	0,89	0,16	$4,12X10^{-3}$	$5,21X10^{-3}$
Unknown						
		nwh^7^56^UNKNOWN	1,58	0,39	$3,99X10^{-12}$	$1,18X10^{-6}$
		nwh^6^76^UNKNOWN	0,84	0,11	$1,54X10^{-5}$	$1,02X10^{-4}$
		nwh^9^26^UNKNOWN	0,89	0,14	$4,31X10^{-3}$	$5,40X10^{-3}$
		lna^12^46^[P] [NP_079904] (NM_025628) RIKEN cDNA 2010000G05	0,90	0,10	$3,17X10^{-3}$	$4,17X10^{-3}$
		rbha^4^69^[N] [BI468155] EST00562	0,86	0,14	$4,65X10^{-4}$	$9,07X10^{-4}$
		nwh^8^69^[P] [AAK01372] (AF315730) putative membrane protein	0,91	0,22	$4,46X10^{-3}$	$5,56X10^{-3}$
		rgb^516^264^[N] [AU081031] AU081031 Oncorhynchus mykiss cDNA clone KA'3	0,90	0,13	$6,25X10^{-3}$	$7,34X10^{-3}$
		rblb^1^60^[P] [P87362] Bleomycin hydrolase	0,88	0,09	$8,06X10^{-4}$	$1,39X10^{-3}$
		srkc^8^24^[N] [BG934162] Salmo salar cDNA clone SK1-0455	1,16	0,22	$3,61X10^{-3}$	$4,65X10^{-3}$
		nwh^18^16^[N] [BG936201] Atlantic Salmon spleen cDNA clone SS1-0555	0,90	0,10	$8,36X10^{-3}$	$9,44X10^{-3}$
		srkh^1^61^UNKNOWN	0,89	0,20	$8,72X10^{-3}$	$9,81X10^{-3}$
Muscular function						
		nwh^10^59^[P] [Q91483] Parvalbumin beta 2 (Major allergen Sal s 1)	0,83	0,14	$3,57X10^{-5}$	$1,61X10^{-4}$
		nwh^19^62^[P] [Q91482] Parvalbumin beta 1 (Major allergen Sal s 1)	0,91	0,13	$8,39X10^{-3}$	$9,47X10^{-3}$
mRNA splicing						
		nwh^3^11^[P] [AAK60397] (AY034062) PTB-associated splicing factor	0,83	0,14	$1,22X10^{-5}$	$8,83X10^{-5}$
Transcription regulation						
		rgb^524^257^[P] [NP_031817] (NM_007791) cysteine rich protein	1,14	0,17	$6,84X10^{-3}$	$7,94X10^{-3}$
		plnb^512^38^[P] [NP_571962] (NM_131887) CCAAT/enhancer binding protein delta	0,84	0,14	$5,91X10^{-5}$	$2,28X10^{-4}$
		rgb^506^32^[N] [AJ012192] cellular nucleic acid binding protein	1,14	0,14	$3,12X10^{-3}$	$4,12X10^{-3}$
		rgb^533^191^[N] [AF266238] transducer of ERBB-2	0,91	0,08	$4,92X10^{-3}$	$6,02X10^{-3}$
		nwh^15^54^[P] [CAA68857] (Y07572) HES1	1,20	0,31	$4,40X10^{-3}$	$5,49X10^{-3}$
		nwh^15^54^[P] [CAA68857] (Y07572) HES1	1,17	0,32	$8,07X10^{-3}$	$9,16X10^{-3}$
Development/Growth regulation			1,19	0,31		
		nwh^12^52^[N] [AF403539] growth hormone receptor isoform 1	1,23	0,27	$1,81X10^{-3}$	$2,63X10^{-3}$
		nwh^15^44^[P] [NP_006399] (NM_006408) anterior gradient 2 homolog	1,18	0,22	$2,29X10^{-3}$	$3,18X10^{-3}$
Cell division and growth						
		plnb^506^341^[P] [NP_612557] (NM_138548) nucleoside diphosphate kinase	0,90	0,12	$5,57X10^{-3}$	$6,68X10^{-3}$
		nwh^18^49^[P] [AAC03020] nucleoside diphosphate kinase	0,86	0,19	$4,04X10^{-4}$	$8,26X10^{-4}$
		nwh^18^87^[P] [AAC03020] nucleoside diphosphate kinase	0,87	0,12	$6,92X10^{-4}$	$1,23X10^{-3}$
		pitl^504^374^[P] [AAB70839] (AF019980) ZipA	1,26	0,42	$4,94X10^{-3}$	$6,05X10^{-3}$
		rgb^513^2^[N] [BG934065] similar to Granulin (grn)	1,14	0,12	$1,95X10^{-3}$	$2,80X10^{-3}$
	Metabolism and Cell Cycle Regulation	srkc^2^69^[P] [P54613] Serine/threonine protein phosphatase 2A	1,16	0,21	$8,06X10^{-3}$	$9,16X10^{-3}$
Immunity						
		khc^4^57^[P] [AAG02511] MHC class I heavy chain	1,11	0,08	$6,07X10^{-3}$	$7,17X10^{-3}$
		nwh^18^62^[N] [BG935727] similar to Class II histocompatibility antigen	1,28	0,38	$1,34X10^{-3}$	$2,09X10^{-3}$
		pha^501^284^[P] [AAF14529] prostaglandin endoperoxide synthase-1	1,24	0,25	$1,68X10^{-4}$	$4,54X10^{-4}$

Protein synthesis						
		pitl^502^333^[N] [AJ427629] Salmo salar 18S rRNA gene	0,76	0,16	$2,90X10^{-8}$	$4,71X10^{-6}$
		rgb^526^214^[P] [NP_076425] (NM_023936) mitochondrial ribosomal protein S34	0,89	0,19	$1,44X10^{-3}$	$2,20X10^{-3}$
		lna^9^42^[P] [CAA10040] Ran protein [Salmo salar]	1,16	0,18	$3,31X10^{-3}$	$4,33X10^{-3}$
Digestive enzyme						
	Hatching	nwh^13^18^[P] [AAL40376] (AC087333) High choriolytic enzyme 1	1,20	0,21	$1,05X10^{-3}$	$1,71X10^{-3}$
Miscellanous						
	Apopotosis regulation	**nwh^3^63^[P] [NP_061289] (NM_018819) brain protein 44-like**	0,87	0,16	$1,33X10^{-3}$	$2,07X10^{-3}$
		rgb^544^332^[P] [NP_061289] (NM_018819) brain protein 44-like	0,89	0,12	$3,03X10^{-3}$	$4,01X10^{-3}$
	Brain Glycoprotein	lna^11^92^[N] [BG936120] SS1-0466 similar to Ependymin	1,23	0,31	$2,09X10^{-3}$	$2,96X10^{-3}$
		rbhb^3^85^[P] [BAB21488] (AB046524) ovomucin alpha-subunit	0,85	0,23	$1,42X10^{-3}$	$2,18X10^{-3}$
	Proteinase Inhibitor	lna^12^27^[P] [Q91195] Cystatin precursor	0,90	0,15	$5,71X10^{-3}$	$6,83X10^{-3}$
	Actin Polymerisation	plnb^504^342^[P] [AAH12596] Similar to actin related protein 2/3 complex, subunit 4	1,15	0,15	$2,24X10^{-3}$	$3,12X10^{-3}$
	Protein Translocation	plnb^503^138^[N] [BG936471] similar to Protein translocation complex beta	1,15	0,20	$3,74X10^{-3}$	$4,79X10^{-3}$

Supplementary Table 2.2 PNRS representing genes differentially expressed between the progeny of farmed and wild Norwegian salmons. PNRS are classified in various biological/cellular functional groups. The average fold change in gene expression in farmed salmon compared to wild individuals (data from 10 arrays) and its standard deviation is given for each PNRS, as well as P-values obtained from the F-test on the mixed model of ANOVA. Genes for which several PNRS were significant are in bold. Colors identify genes or functional entities showing parallel changes in Canada and Norway. Clone identification follows notation from the GRASP. GeneBank accession numbers refer to the GeneBank sequence the most similar to the amplified cDNA portion spotted on the array.

Function		PNRS	Average fold change	Fold change SD	Tabulated P-value	Permutation corrected Pvalue
Energy metabolism						
	Short term energy Reserves					
		rbha^4^86^[P] [S13164] creatine kinase (EC 2.7.3.2)	1,2	0,3	$7,1X10^{-3}$	$8,80X10^{-3}$
	Glycolysis					
		lna^17^81^[P] [AAK49985] glyceraldehyde phosphate dehydrogenase	0,8	0,3	$8,0X10^{-3}$	$9,76X10^{-3}$
	Oxidative phosphorylation					
		rbna^1^21^[P] [AAK54362] NADH dehydrogenase subunit 5	1,3	0,2	$3,6X10^{-5}$	$3,61X10^{-4}$
		rgb^533^365^[P] [NP_008455] NADH dehydrogenase subunit 5	1,3	0,3	$2,5X10^{-3}$	$4,08X10^{-3}$
		rgb^525^123^[P] [NP_008454] NADH dehydrogenase subunit 4	1,3	0,4	$1,9X10^{-3}$	$3,44X10^{-3}$
		pitl^505^178^[P] [BAA82837] (AB023582) ATP synthase beta-subunit	0,8	0,2	$2,1X10^{-3}$	$3,62X10^{-3}$
Metal ions sequetration/transport						
		lna^9^17^[P] [AAB34575] ferritin heavy subunit; ferritin H	1,2	0,3	$6,2X10^{-3}$	$7,96X10^{-3}$
		rtch^1^9^[P] [BAA13147] ferritin H-2	1,2	0,2	$4,8X10^{-3}$	$6,55X10^{-3}$
Unknown						
		nwh^7^56^UNKNOWN	1,4	0,5	$3,3X10^{-3}$	$4,94X10^{-3}$
		nwh^6^76^UNKNOWN	0,8	0,1	$1,8X10^{-5}$	$2,60X10^{-4}$
		lna^10^21^UNKNOWN	0,9	0,1	$3,7X10^{-3}$	$5,44X10^{-3}$
		srkb^1^91^UNKNOWN	0,9	0,2	$6,4X10^{-3}$	$8,18X10^{-3}$
		lna^9^95^UNKNOWN	0,9	0,2	$2,1X10^{-3}$	$3,70X10^{-3}$
		lna^3^77^UNKNOWN	0,9	0,1	$4,9X10^{-3}$	$6,72X10^{-3}$
		lna^11^64^UNKNOWN	1,2	0,2	$8,2X10^{-3}$	$9,92X10^{-3}$
		srkb^1^7^UNKNOWN	0,9	0,1	$4,4X10^{-3}$	$6,16X10^{-3}$
		rgb^511^345^[N] [BG935999] cDNA clone SS1-0335 similar to IL-1b intron	0,9	0,1	$1,3X10^{-3}$	$2,62X10^{-3}$
		lna^9^79^[P] [AAA72048] unnamed protein product	0,8	0,1	$1,1X10^{-3}$	$2,35X10^{-3}$
		lna^8^66^[N] [BF228689] EST00400	0,8	0,1	$1,2X10^{-4}$	$6,51X10^{-4}$
		srkh^2^87^[N] [BF228558] EST00269	1,2	0,1	$2,9X10^{-3}$	$4,51X10^{-3}$
		rgb^534^55^[N] [BG933940] cDNA clone SK1-0222 similar to beta-actin 3' UTR	1,2	0,3	$7,8X10^{-3}$	$9,56X10^{-3}$

	srkc^2^35^[N] [BG935820] cDNA clone SS1-0134	0,8	0,2	$1,7X10^{-3}$	$3,14X10^{-3}$
	nwh^18^16^[N] [BG936201] cDNA clone SS1-0555	0,8	0,1	$1,0X10^{-3}$	$2,29X10^{-3}$
	nwh^18^16^[N] [BG936201] cDNA clone SS1-0555	0,9	0,2	$8,1X10^{-3}$	$9,87X10^{-3}$
	shc^501^292^[N] [AL132660] similar to clone RP1-249H1	1,2	0,2	$5,6X10^{-3}$	$7,40X10^{-3}$
		0,8	0,2		
Molecular chaperone					
	plnb^513^197^[P] [S21175] dnaK-type molecular chaperone hsc71	1,5	0,7	$7,0X10^{-3}$	$8,74X10^{-3}$
	srkc^2^72^[P] [BAA89277] (AB027708) CCT (chaperonin containing T-complex polypeptide 1) epsilon subunit	1,3	0,3	$1,0X10^{-3}$	$2,30X10^{-3}$
	shc^503^266^[N] [BG934616] SK1-0931 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0931 similar to Heat shock protein 70 (hsp70)	1,2	0,2	$3,7X10^{-3}$	$5,38X10^{-3}$
Protein synthesis					
	rblb^2^86^[P] [NP_112362] (NM_031100) ribosomal protein L10 [Rattus norvegicus]	1,2	0,2	$5,2X10^{-3}$	$7,01X10^{-3}$
	nwh^8^8^[P] [AAG09233] (AF176302) brain acidic ribosomal phosphoprotein P0 [Rana sylvatica]	1,2	0,2	$1,9X10^{-3}$	$3,22X10^{-3}$
	nwh^8^8^[P] [AAG09233] (AF176302) brain acidic ribosomal phosphoprotein P0 [Rana sylvatica]	1,2	0,2	$1,7X10^{-3}$	$3,46X10^{-3}$
	rblb^2^79^[P] [CAA10040] Ran protein [Salmo salar]	1,2	0,3	$5,5X10^{-3}$	$7,25X10^{-3}$
	rga^501^22^[P] [AAG38613] elongation factor 1 alpha [Salmo salar]	1,2	0,2	$5,5X10^{-4}$	$1,57X10^{-3}$
	srkc^4^35^[P] [CAA40268] (X56953) protein synthesis initiation factor 4A [Mus musculus]	1,2	0,2	$1,1X10^{-3}$	$2,37X10^{-3}$
	srkc^7^86^[P] [CAA40268] (X56953) protein synthesis initiation factor 4A [Mus musculus]	1,2	0,3	$5,2X10^{-3}$	$6,98X10^{-3}$
Extra-cellular matrix protein					
	shc^501^353^[P] [BAA33381] alpha 3 type I collagen [Oncorhynchus mykiss]	1,6	0,7	$3,5X10^{-5}$	$3,56X10^{-4}$
	rgb^530^221^[P] [O93484] Collagen alpha 2(I) chain precursor	1,7	0,8	$4,2X10^{-4}$	$1,35X10^{-3}$
	rbha^4^88^[N] [AB021698] Oncorhynchus mykiss mRNA for matrix metalloproteinase-2, complete cds	1,3	0,4	$4,7X10^{-4}$	$1,43X10^{-3}$
	rblb^3^48^[P] [AAC99813] secreted protein, acidic, rich in cysteine [Oncorhynchus mykiss]	1,3	0,3	$1,3X10^{-3}$	$2,64X10^{-3}$
Immunity					
	nwh^18^29^[P] [AAG30026] C-type lectin 2-2 [Oncorhynchus mykiss]	0,7	0,5	$2,0X10^{-3}$	$3,53X10^{-3}$
	rgb^508^226^[P] [AAF63469] (AF227738) mannose binding-like lectin precursor [Danio rerio]	0,8	0,1	$3,6X10^{-3}$	$5,34X10^{-3}$
	rgb^515^244^[P] [BAA22040] (D87910) PA28 beta subunit [Mus musculus]	1,2	0,1	$3,9X10^{-3}$	$5,61X10^{-3}$
	rgb^542^83^[P] [CAA67765] pentraxin [Salmo salar]	0,9	0,1	$4,0X10^{-3}$	$5,70X10^{-3}$
	lna^16^11^[P] [P00719] LYSOZYME G (1,4-BETA-N-ACETYLMURAMIDASE) (GOOSE-TYPE LYSOZYME)	0,8	0,1	$9,5X10^{-4}$	$2,18X10^{-3}$
	pitl^510^180^[N] [AF180478] Salmo salar clone BA1 beta-2 microglobulin (B2m) mRNA, complete cds	0,8	0,2	$7,6X10^{-3}$	$9,33X10^{-3}$
Blood transport protein					
	rgb^509^64^[P] [CAA65955] alpha-globin [Salmo salar]	1,3	0,3	$2,5X10^{-3}$	$4,07X10^{-3}$
	rgb^514^194^[P] [S04589] hemoglobin alpha chain	1,2	0,2	$3,6X10^{-3}$	$5,36X10^{-3}$
	plnb^510^212^[P] [JH0472] apolipoprotein A-I precursor	1,9	1,0	$5,6X10^{-5}$	$4,51X10^{-4}$
	nwh^20^30^[P] [P24774] PLASMA RETINOL-BINDING PROTEIN I (PRBP-I)	0,8	0,2	$5,4X10^{-3}$	$7,21X10^{-3}$

Signal transduction						
		plnb^513^115^[P] [MCON] calmodulin	1,3	0,2	$6,8X10^{-5}$	$4,94X10^{-4}$
		plnb^510^56^[P] [MCON] calmodulin	1,2	0,2	$4,2X10^{-3}$	$5,98X10^{-3}$
		plnb^503^47^[P] [P02593] Calmodulin	1,2	0,1	$7,3X10^{-3}$	$9,04X10^{-3}$
		pitl^501^335^[P] [P02593] Calmodulin	1,3	0,3	$4,9X10^{-3}$	$6,71X10^{-3}$
		rbna^1^8^[P] [1703469D] protein phosphatase 1 delta	1,3	0,4	$3,5X10^{-3}$	$5,23X10^{-3}$
Transcription regulation						
		nwh^6^37^[N] [AF266238] Gillichthys mirabilis transducer of ERBB-2	0,8	0,2	$1,6X10^{-4}$	$7,66X10^{-4}$
		rbha^4^94^[P] [AAB50916] (S83517) tinman homeodomain homolog	1,2	0,1	$2,7X10^{-4}$	$1,04X10^{-3}$
		nwh^9^34^[P] [BAA36334] (AB014611) XFD-6	1,2	0,2	$1,8X10^{-3}$	$3,31X10^{-3}$
		rgb^530^222^[P] [NP_004261] (NM_004270) cofactor required for Sp1 transcriptional activation, subunit 9	0,9	0,1	$5,2X10^{-3}$	$7,00X10^{-3}$
Development/Growth regulation						
		rgb^517^56^[P] [AAK59991] (AY034136) embryonic ectoderm development protein	1,2	0,2	$3,0X10^{-3}$	$4,65X10^{-3}$
Muscular function						
		nwh^8^70^[N] [AF330142] actin	0,8	0,2	$4,2X10^{-3}$	$5,97X10^{-3}$
		nwh^7^65^[P] [AAF01275] (AF165215) Sk-tropomodulin	1,2	0,2	$4,7X10^{-3}$	$6,49X10^{-3}$
Digestive enzyme						
		rblb^1^20^[P] [BAA86981] (AB025009) novel member of chitinase family	1,3	0,4	$2,2X10^{-3}$	$3,81X10^{-3}$
		rgb^523^32^[P] [AAF31644] (AF154571) putative chitinase	0,9	0,1	$6,2X10^{-3}$	$7,98X10^{-3}$
Lysosomal Enzyme						
		plnb^504^355^[N] [U90321] Oncorhynchus mykiss cathepsin D	0,8	0,2	$2,5X10^{-3}$	$4,12X10^{-3}$
		rgb^509^158^[P] [AAK69706] procathepsin L	1,2	0,2	$4,4X10^{-4}$	$1,39X10^{-3}$
Miscellanous						
		rgb^533^32^[N] [BG936460] SS1-0835 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0835 similar to Ornithine decarboxylase antizyme	1,2	0,2	$1,2X10^{-3}$	$2,51X10^{-3}$
		tc^5^57^[P] [AAK00216] ovarian cysteine protease inhibitor	0,9	0,1	$2,5X10^{-3}$	$4,14X10^{-3}$
		pitl^505^105^[P] [P07746] HIGH MOBILITY GROUP-T PROTEIN (HMG-T)	1,3	0,2	$6,1X10^{-4}$	$1,67X10^{-3}$
		rblb^2^31^[P] [AAA83776] (U41499) arsenite-resistance protein	1,2	0,2	$4,8X10^{-3}$	$6,57X10^{-3}$
		srkc^5^91^[P] [NP_035845] (NM_011715) WD40 repeat protein 1	1,2	0,1	$4,8X10^{-3}$	$6,59X10^{-3}$
		rgb^525^383^[P] [NP_114019] (NM_031831) Nogo-A; foocen	1,3	0,3	$3,1X10^{-4}$	$1,12X10^{-3}$
	Cell Structure	nwh^20^46^[N] [BG933915] SK1-0197 Atlantic Salmon kidney cDNA clone SK1-0197 similar to Beta actin 1 (act1)	1,6	0,7	$1,2X10^{-3}$	$2,46X10^{-3}$
	Nucleotide Metabolism	rgb^508^361^[P] [1GTH_A] Chain A, Dihydropyrimidine Dehydrogenase (Dpd) From Pig, Ternary Complex With Nadph And 5-Iodouracil	1,2	0,2	$7,5X10^{-3}$	$9,21X10^{-3}$
		plnb^510^221^[P] [P21566] Cofilin	1,2	0,3	$7,8X10^{-3}$	$9,53X10^{-3}$

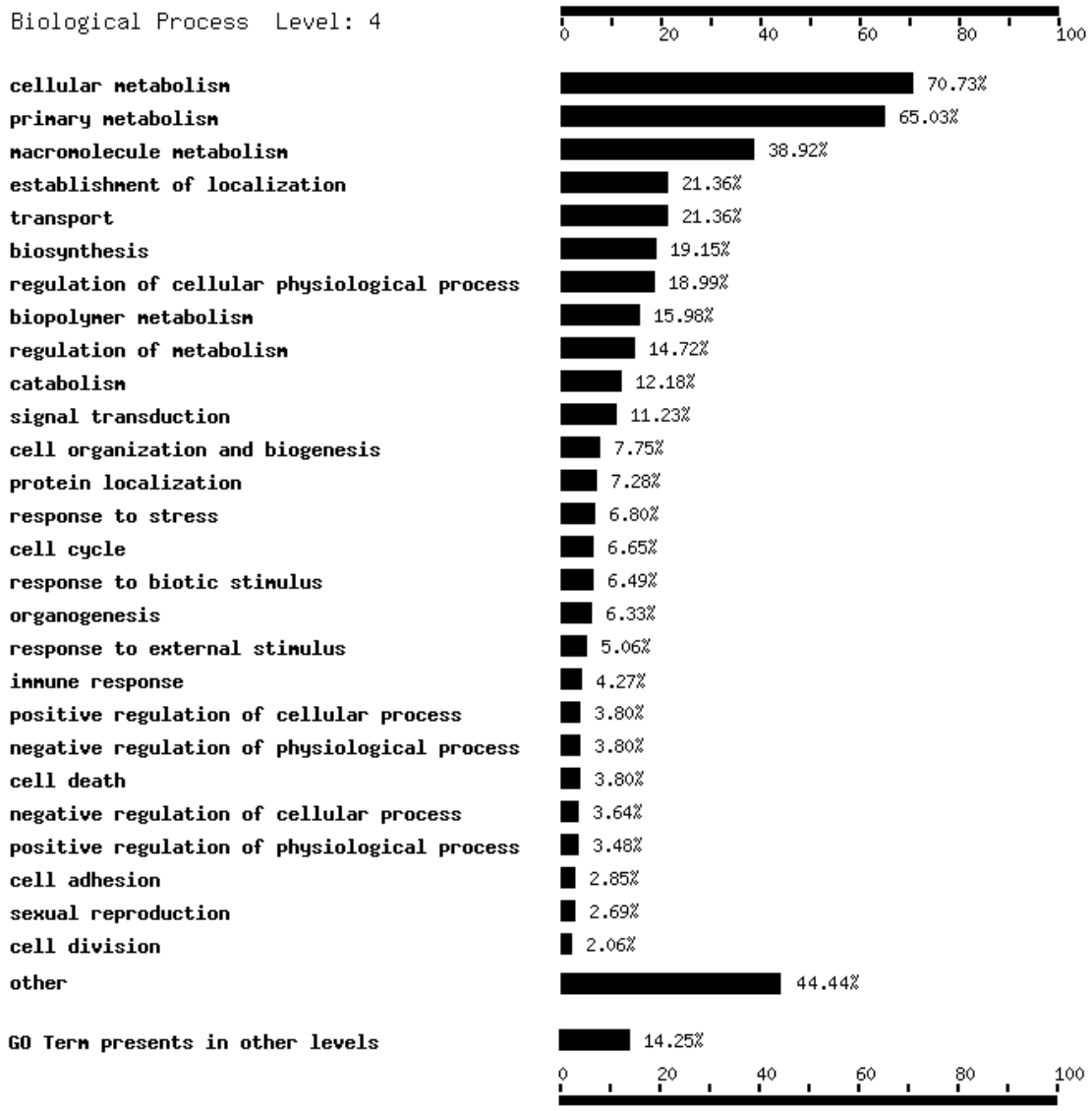
Supplementary Table 2.3 PNRS representing the 171 genes differentially expressed between the progeny of wild salmon from Norway (Namsen river) and Canada (St-John river) at the P < 0.01 threshold on permutation-based P-values. The average fold change in gene expression in Canadian compared to Norwegian wild individuals (data from 22 individuals) is given for each PNRS, as well as P-values obtained from the F-test on the ANOVA model. GeneBank accession numbers refer to the GeneBank sequence the most similar to the amplified cDNA portion spotted on the array. Clone identification follows notation from the GRASP.

PNRS	Tabulated P-value	Permutation-based P-value	Fold-change
rgb^508^382^[P] [AAA58771] HMG-1	$7{,}34X10^{-6}$	0	2,29
nwh^13^91^UNKNOWN	$5{,}02X10^{-6}$	0	1,76
rgb^510^291^[P] [XP_034795] malate dehydrogenase 2, NAD (mitochondrial)	$2{,}71X10^{-5}$	$6{,}35X10^{-6}$	1,85
pitl^508^308^[P] [AAK95210] (AF402836) 40S ribosomal protein S27-1	$3{,}60X10^{-5}$	$1{,}04X10^{-5}$	1,87
rtfh^1^79^[P] [AAK63073] (AF385081) 60S ribosomal protein L13	$6{,}57X10^{-5}$	$1{,}77X10^{-5}$	2,08
rgb^518^182^[P] [Q90YQ6] 40S ribosomal protein S17	$7{,}38X10^{-5}$	$2{,}09X10^{-5}$	2,31
pitl^504^62^[P] [AAK95124] (AF401552) ribosomal protein P1	$9{,}38X10^{-5}$	$3{,}18X10^{-5}$	1,71
nwh^11^86^[P] [Q91482] Parvalbumin beta 1 (Major allergen Sal s 1)	$9{,}81X10^{-5}$	$3{,}27X10^{-5}$	2,57
rgb^504^270^[P] [T11749] transferrin	$1{,}31X10^{-4}$	$4{,}54X10^{-5}$	1,54
srkc^2^30^UNKNOWN	$1{,}53X10^{-4}$	$5{,}63X10^{-5}$	1,57
plnb^510^163^[P] [P49947] Ferritin, middle subunit (Ferritin M)	$1{,}60X10^{-4}$	$5{,}86X10^{-5}$	1,99
rtfh^1^44^[P] [NP_033118] (NM_009092) ribosomal protein S17	$1{,}72X10^{-4}$	$6{,}31X10^{-5}$	1,86
pitl^512^264^[P] [AAF61384] cytochrome c oxidase subunit III	$1{,}79X10^{-4}$	$6{,}67X10^{-5}$	1,83
rtdh^1^39^[N] [BG936357] SS1-0729 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0729	$1{,}85X10^{-4}$	$7{,}13X10^{-5}$	2,07
srkc^8^24^[N] [BG934162] SK1-0455 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0455	$2{,}10X10^{-4}$	$8{,}13X10^{-5}$	1,69
rgb^509^335^[P] [AAG16767] (AF192528) integrin beta-1 subunit	$2{,}23X10^{-4}$	$8{,}76X10^{-5}$	1,95
rgb^527^353^[P] [NP_571645] (NM_131570) mannose binding-like lectin	$2{,}61X10^{-4}$	$1{,}02X10^{-4}$	0,57
pitl^510^45^[N] [BI468041] EST00451 Atlantic salmon Lambda Zap Express liver cDNA library Salmo salar cDNA clone LRR8-g03 5' similar to 60S Ribosomal Protein L13A	$2{,}94X10^{-4}$	$1{,}16X10^{-4}$	1,93
rtdl^2^8^UNKNOWN	$3{,}63X10^{-4}$	$1{,}53X10^{-4}$	1,91
nwh^17^41^UNKNOWN	$4{,}45X10^{-4}$	$2{,}03X10^{-4}$	0,60
rbha^2^57^[P] [AAH10583] (BC010583) Similar to mannosyl (alpha-1,6-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase	$4{,}54X10^{-4}$	$2{,}07X10^{-4}$	1,35
lna^13^26^UNKNOWN	$6{,}50X10^{-4}$	$3{,}23X10^{-4}$	1,80
pitl^508^357^[P] [XP_018182] (XM_018182) similar to ribosomal protein L29; 60S ribosomal protein L29; heparin/heparan sulfate-interacting protein; HP/HS-interacting protein	$6{,}91X10^{-4}$	$3{,}48X10^{-4}$	1,48
pitl^501^61^[P] [XP_015462] hypothetical protein XP_015462	$7{,}40X10^{-4}$	$3{,}74X10^{-4}$	1,47
pha^501^64^[P] [AAH06699] (BC006699) Unknown (protein for IMAGE:3498575)	$7{,}43X10^{-4}$	$3{,}77X10^{-4}$	1,47
srkb^1^70^[N] [BG936209] SS1-0563 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0563 similar to Poly A binding protein	$8{,}18X10^{-4}$	$4{,}22X10^{-4}$	1,48
nwh^18^48^UNKNOWN	$1{,}01X10^{-3}$	$5{,}31X10^{-4}$	0,57
plnb^503^185^[N] [AF151524] Oncorhynchus mykiss cytochrome P450 2K5 (CYP2K5)	$1{,}01X10^{-3}$	$5{,}38X10^{-4}$	0,67
plnb^503^306^[P] [CAC19682] (AL137790) dJ1003J2.1 (sodium and chloride-dependent transporter NTT4)	$1{,}05X10^{-3}$	$5{,}57X10^{-4}$	2,42
lna^12^5^[P] [P49207] 60S ribosomal protein L34	$1{,}06X10^{-3}$	$5{,}66X10^{-4}$	1,52
rgb^507^112^[N] [BG934180] SK1-0473 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0473 similar to Beta actin 1 (act1)	$1{,}10X10^{-3}$	$5{,}94X10^{-4}$	2,15
pitl^505^211^[P] [AAA57047] (M24507) ubiquitin	$1{,}12X10^{-3}$	$6{,}09X10^{-4}$	1,45
rtfl^1^62^[P] [CAC44627] (AJ290422) 60s ribosomal protein L44 (L36A)	$1{,}14X10^{-3}$	$6{,}21X10^{-4}$	1,90
rgb^507^217^[P] [CAC44159] putative ribosomal protein L27A protein	$1{,}14X10^{-3}$	$6{,}25X10^{-4}$	1,83
lna^1^50^UNKNOWN	$1{,}21X10^{-3}$	$6{,}74X10^{-4}$	1,57
srkc^5^15^[N] [BG935820] SS1-0134 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0134	$1{,}23X10^{-3}$	$6{,}82X10^{-4}$	1,73
rgb^544^175^[P] [P50990] T-complex protein 1, theta subunit (TCP-1-theta) (CCT-theta)	$1{,}27X10^{-3}$	$7{,}06X10^{-4}$	1,47

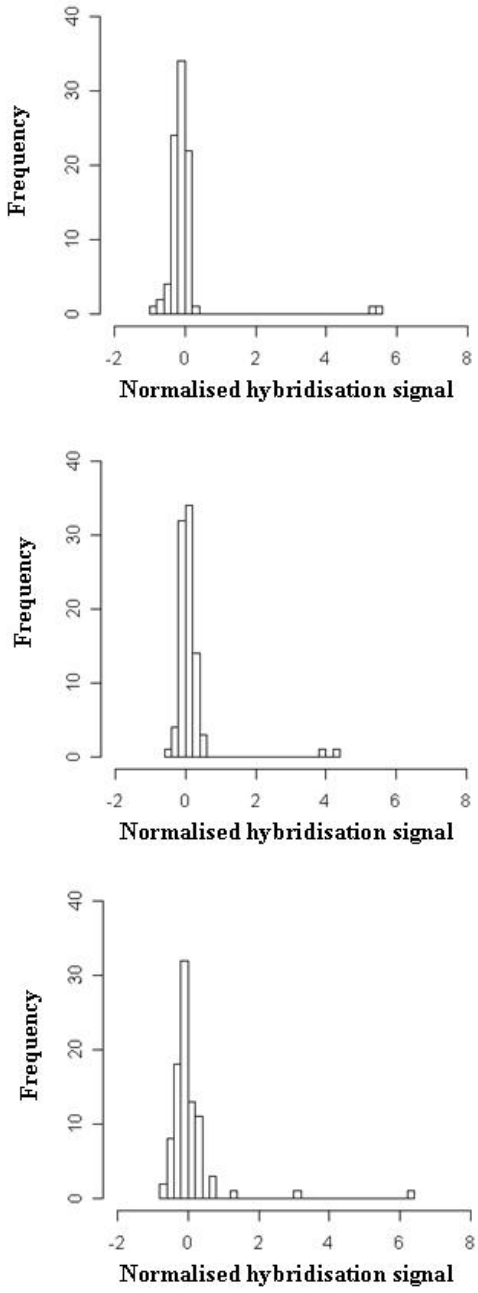
nwh^10^52^UNKNOWN	$1,27X10^{-3}$	$7,06X10^{-4}$	1,44
rteh^1^82^UNKNOWN	$1,28X10^{-3}$	$7,13X10^{-4}$	1,30
rgb^527^230^[N] [BG935768] SS1-0073 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0073 similar to Coronin	$1,37X10^{-3}$	$7,65X10^{-4}$	0,60
srkc^8^2^[P] [AAD30275] heat shock protein hsp90 beta	$1,40X10^{-3}$	$7,83X10^{-4}$	1,86
rgb^508^373^[P] [AAG22824] 40S ribosomal protein S11	$1,46X10^{-3}$	$8,31X10^{-4}$	1,93
lna^11^92^[N] [BG936120] SS1-0466 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0466 similar to Ependymin	$1,50X10^{-3}$	$8,61X10^{-4}$	0,68
lna^9^95^UNKNOWN	$1,57X10^{-3}$	$9,04X10^{-4}$	0,66
pitl^510^16^[P] [Q91487] 60S RIBOSOMAL PROTEIN L13A (TRANSPLANTATION ANTIGEN P198 HOMOLOG)	$1,58X10^{-3}$	$9,08X10^{-4}$	1,92
nwh^6^96^[P] [S66520] tropomyosin, fast muscle	$1,59X10^{-3}$	$9,11X10^{-4}$	2,42
lna^9^74^[P] [1909362A] ribosomal protein L27	$1,65X10^{-3}$	$9,52X10^{-4}$	1,58
rgb^501^164^[P] [CAA65945] beta-globin	$1,65X10^{-3}$	$9,53X10^{-4}$	1,82
rgb^531^78^[P] [NP_038791] (NM_013763) transducin (beta)-like 2	$1,69X10^{-3}$	$9,82X10^{-4}$	0,65
rgb^524^281^[P] [AAA91456] (U35728) transcription factor A	$1,70X10^{-3}$	$9,87X10^{-4}$	1,34
lna^16^7^[N] [BG935098] SL1-0407 Atlantic Salmon liver Salmo salar cDNA clone SL1-0407	$1,72X10^{-3}$	$1,00X10^{-3}$	0,59
lna^16^72^UNKNOWN	$1,73X10^{-3}$	$1,01X10^{-3}$	1,46
nwh^2^78^[P] [AAC23580] fast myotomal muscle troponin-I	$1,77X10^{-3}$	$1,03X10^{-3}$	2,04
rgb^532^29^[P] [T12505] hypothetical protein DKFZp434C192.1	$1,81X10^{-3}$	$1,06X10^{-3}$	0,64
rgb^509^261^[N] [BG934645] SK1-0962 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0962 similar to hematological and neurological expressed sequence 1	$1,82X10^{-3}$	$1,07X10^{-3}$	1,35
rgb^532^26^[N] [L77086] Salmo salar MHC class II alpha	$1,89X10^{-3}$	$1,12X10^{-3}$	0,52
pitl^505^275^[P] [BAB22488] (AK002973) putative	$1,94X10^{-3}$	$1,16X10^{-3}$	1,62
srkh^2^2^[N] [BG935920] SS1-0249 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0249 similar to Ferritin H	$1,95X10^{-3}$	$1,17X10^{-3}$	1,48
srkc^7^31^[N] [AY071855] Oncorhynchus mykiss clone GC41 MHC class I antigen (Onmy-U41p) pseudogene, partial sequence; and proteosome subunit LMP7/PSMB8, exons 3 through 6	$1,97X10^{-3}$	$1,19X10^{-3}$	1,31
nwh^18^49^[P] [AAC03020] nucleoside diphosphate kinase	$2,05X10^{-3}$	$1,25X10^{-3}$	1,67
nwh^10^26^[P] [Q91482] Parvalbumin beta 1 (Major allergen Sal s 1)	$2,09X10^{-3}$	$1,28X10^{-3}$	1,89
nwh^14^17^[P] [BAB68518] (AB071428) hatching enzyme EHE13	$2,11X10^{-3}$	$1,30X10^{-3}$	1,71
rgb^532^92^[P] [AAK27451] (AF338735) hypothetical PHD zinc finger protein XAP135	$2,15X10^{-3}$	$1,32X10^{-3}$	0,64
rgb^525^123^[P] [NP_008454] NADH dehydrogenase subunit 4	$2,26X10^{-3}$	$1,40X10^{-3}$	1,41
nwh^3^88^[N] [AF470044] Oncorhynchus mykiss microsatellite OMM1280 sequence	$2,37X10^{-3}$	$1,47X10^{-3}$	1,62
rbha^3^10^[P] [NP_033966] (NM_009836) chaperonin subunit 3 (gamma)	$2,44X10^{-3}$	$1,53X10^{-3}$	1,87
nwh^8^70^[N] [AF330142] Oncorhynchus mykiss actin	$2,48X10^{-3}$	$1,55X10^{-3}$	2,23
rgb^542^124^[N] [BG936034] SS1-0373 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0373 similar to Thymosin beta	$2,51X10^{-3}$	$1,58X10^{-3}$	1,64
rtah^1^76^[P] [AAH13499] (BC013499) Unknown (protein for MGC:19163)	$2,60X10^{-3}$	$1,64X10^{-3}$	0,61
rtfh^1^71^[P] [AAB53643] fatty acid binding protein H-FABP	$2,71X10^{-3}$	$1,73X10^{-3}$	1,60
rgb^542^374^[P] [BAB62189] glyceraldehyde 3-phosphate dehydrogenase	$2,94X10^{-3}$	$1,89X10^{-3}$	1,41
rgb^544^280^[P] [BAB55662] collagen a3(I)	$2,94X10^{-3}$	$1,89X10^{-3}$	1,69
pitl^508^339^[P] [BAB27102] (AK010667) data source:MGD, source key:MGI:1913725, evidence:ISS~putative~ribosomal protein S23	$3,00X10^{-3}$	$1,94X10^{-3}$	1,54
nwh^10^28^[P] [O57561] 60S RIBOSOMAL PROTEIN L18A	$3,02X10^{-3}$	$1,96X10^{-3}$	1,52
nwh^20^76^UNKNOWN	$3,19X10^{-3}$	$2,08X10^{-3}$	0,62
nwh^9^7^UNKNOWN	$3,26X10^{-3}$	$2,14X10^{-3}$	0,59
pitl^508^61^[P] [CAC28137] (AJ291832) ADP,ATP translocase	$3,28X10^{-3}$	$2,16X10^{-3}$	1,34
shc^502^330^[P] [AAK51460] polyubiquitin	$3,43X10^{-3}$	$2,29X10^{-3}$	1,98
lna^11^17^[P] [BAA92287] (AB025577) Orla C4	$3,43X10^{-3}$	$2,29X10^{-3}$	0,65
lna^3^31^[P] [NP_061219] (NM_018749) eukaryotic translation initiation factor 3, subunit 7	$3,50X10^{-3}$	$2,34X10^{-3}$	0,60
lna^9^23^[P] [BAB32661] (AB055884) 60S ribosomal protein L35	$3,72X10^{-3}$	$2,51X10^{-3}$	1,43
pha^501^321^[P] [P27952] 40S RIBOSOMAL PROTEIN S2	$3,82X10^{-3}$	$2,60X10^{-3}$	1,96
pitl^508^249^[P] [BAA88568] ubiquitin	$3,86X10^{-3}$	$2,63X10^{-3}$	1,29
pitl^508^100^[N] [BG934317] SK1-0616 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0616 similar to Peptidylprolyl isomerase (EC 5.2.1.8)	$3,89X10^{-3}$	$2,65X10^{-3}$	1,49
shc^501^251^[P] [CAC21554] (AJ298278) poly(A) binding protein	$4,19X10^{-3}$	$2,89X10^{-3}$	1,65

rgb^520^290^[P] [NP_033086] regucalcin	$4,20X10^{-3}$	$2,89X10^{-3}$	0,66
rgb^508^33^[P] [CAC28060] (AJ303069) putative transposase	$4,42X10^{-3}$	$3,08X10^{-3}$	1,32
rgb^530^184^[P] [CAC36098] (AJ310911) putative ribosomal protein L14	$4,55X10^{-3}$	$3,18X10^{-3}$	1,52
lna^6^17^UNKNOWN	$4,61X10^{-3}$	$3,22X10^{-3}$	1,41
rbhb^7^77^[N] [BG934546] SK1-0854 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0854 similar to Protoporhyrinogen oxidase	$4,69X10^{-3}$	$3,29X10^{-3}$	1,38
rgb^518^183^[P] [AAK95124] (AF401552) ribosomal protein P1	$4,99X10^{-3}$	$3,54X10^{-3}$	1,63
rgb^507^58^[N] [AF005026] Salvelinus fontinalis serine protease-like protein	$5,00X10^{-3}$	$3,54X10^{-3}$	1,40
rgb^520^83^[P] [NP_079862] RIKEN cDNA 2510008H07	$5,11X10^{-3}$	$3,64X10^{-3}$	1,64
nwh^14^73^[N] [BG936210] SS1-0564 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0564 similar to nitrogen fixation protein (NIFU)	$5,15X10^{-3}$	$3,68X10^{-3}$	0,58
plnb^506^149^[P] [BAA13413] (D87515) aminopeptidase-B	$5,18X10^{-3}$	$3,69X10^{-3}$	0,58
nwh^9^75^UNKNOWN	$5,19X10^{-3}$	$3,70X10^{-3}$	0,56
rbhb^8^18^[P] [NP_000292] (NM_000301) plasminogen	$5,22X10^{-3}$	$3,72X10^{-3}$	1,36
rtdl^2^38^[N] [BG936356] SS1-0728 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0728 similar to repeat	$5,32X10^{-3}$	$3,80X10^{-3}$	1,56
srkc^9^29^UNKNOWN	$5,39X10^{-3}$	$3,86X10^{-3}$	1,24
nwh^2^61^UNKNOWN	$5,41X10^{-3}$	$3,88X10^{-3}$	1,22
rblb^3^94^[P] [BAB27417] (AK011124) NADH dehydrogenase (ubiquinone) Fe-S protein 4~data source:MGD, source key:MGI:1343135, evidence:ISS~putative	$5,52X10^{-3}$	$3,96X10^{-3}$	0,54
plnb^503^283^[P] [AAF61383] ATPase subunit 6	$5,95X10^{-3}$	$4,32X10^{-3}$	1,96
lna^10^67^[P] [P00718] LYSOZYME G (1,4-BETA-N-ACETYLMURAMIDASE) (GOOSE-TYPE LYSOZYME)	$5,99X10^{-3}$	$4,36X10^{-3}$	0,48
rtdh^1^28^UNKNOWN	$6,07X10^{-3}$	$4,43X10^{-3}$	0,59
lna^8^66^[N] [BF228689] EST00400 Atlantic salmon Lambda Zap Express spleen cDNA library Salmo salar cDNA clone SNN-1095 5'	$6,18X10^{-3}$	$4,52X10^{-3}$	0,55
nwh^17^57^[P] [XP_001357] (XM_001357) c-myc binding protein	$6,24X10^{-3}$	$4,57X10^{-3}$	0,66
nwh^8^38^[P] [O57523] APOLIPOPROTEIN A-I-1 PRECURSOR (APOA-I-1)	$6,26X10^{-3}$	$4,58X10^{-3}$	1,45
plnb^510^135^[P] [CAC44156] putative 40S ribosomal protein 20S protein	$6,40X10^{-3}$	$4,70X10^{-3}$	0,69
nwh^13^20^[P] [A42112] mucin-like peptide MLP 2677	$6,41X10^{-3}$	$4,71X10^{-3}$	0,59
lna^1^23^[N] [AJ313031] Oncorhynchus mykiss iip2 gene for interferon inducible protein 2, exons 1-2	$6,46X10^{-3}$	$4,76X10^{-3}$	1,43
plnb^506^349^[P] [AAC24313] (AC004382) Unknown gene product	$6,63X10^{-3}$	$4,90X10^{-3}$	0,63
rbha^4^86^[P] [S13164] creatine kinase (EC 2.7.3.2)	$6,77X10^{-3}$	$5,02X10^{-3}$	1,93
lna^17^56^[P] [P80601] 14.5 KD TRANSLATIONAL INHIBITOR PROTEIN (UK114 ANTIGEN)	$6,79X10^{-3}$	$5,03X10^{-3}$	0,72
rgb^517^96^[P] [CAC42525] (AL136300) bA10G5.1 (similar to ZNF183: zinc finger protein 183 (RING finger, C3HC4 type))	$6,80X10^{-3}$	$5,05X10^{-3}$	0,63
rgb^526^138^[P] [NP_008447] cytochrome c oxidase subunit I	$6,83X10^{-3}$	$5,07X10^{-3}$	1,52
plnb^503^203^[P] [XP_065874] (XM_065874) similar to TPA: carboxypeptidase O; CPO	$6,87X10^{-3}$	$5,11X10^{-3}$	0,60
nwh^19^13^UNKNOWN	$6,87X10^{-3}$	$5,11X10^{-3}$	1,47
nwh^11^85^[P] [Q91483] Parvalbumin beta 2 (Major allergen Sal s 1)	$7,02X10^{-3}$	$5,24X10^{-3}$	1,52
plnb^505^52^[P] [P20135] GLUTATHIONE S-TRANSFERASE 1 (GST-CL1) (GST CLASS-THETA)	$7,06X10^{-3}$	$5,28X10^{-3}$	0,73
lna^6^24^[N] [AF045053] Oncorhynchus mykiss cytochrome P450 monooxygenase CYP2K1v3	$7,24X10^{-3}$	$5,43X10^{-3}$	0,60
rgb^503^345^[P] [2105196A] argininosuccinate lyase	$7,28X10^{-3}$	$5,46X10^{-3}$	1,40
rgb^520^207^[P] [AAF71749] (AF257743) phosphoribosylglycinamide formyltransferase	$7,34X10^{-3}$	$5,51X10^{-3}$	0,57
rgb^509^202^[P] [P21533] 60S RIBOSOMAL PROTEIN L6 (NEOPLASM-RELATED PROTEIN C140)	$7,53X10^{-3}$	$5,67X10^{-3}$	1,68
srkh^1^32^[N] [BG934513] SK1-0819 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0819 similar to other reported ESTs	$7,71X10^{-3}$	$5,82X10^{-3}$	0,64
rgb^524^303^[P] [NP_571908] (NM_131833) ubiquitin-conjugating enzyme E2I2	$7,76X10^{-3}$	$5,87X10^{-3}$	1,21
rgb^526^361^[P] [AAB34576] ferritin middle subunit; ferritin M	$7,82X10^{-3}$	$5,92X10^{-3}$	1,32
rgb^507^12^[N] [BI468086] EST00496 Atlantic salmon Lambda Zap Express liver cDNA library Salmo salar cDNA clone LRR4-f05 5' similar to Beta-Globin	$8,09X10^{-3}$	$6,14X10^{-3}$	1,46
nwh^10^66^UNKNOWN	$8,25X10^{-3}$	$6,26X10^{-3}$	1,60
pitl^505^34^[N] [BG936652] SS1-1038 Atlantic Salmon spleen Salmo salar cDNA clone SS1-1038 similar to T-cell antigen receptor 3' UTR	$8,54X10^{-3}$	$6,51X10^{-3}$	0,55
lna^5^3^[P] [BAB47142] (AB050632) phospholipase A2	$8,62X10^{-3}$	$6,58X10^{-3}$	0,59
lna^16^95^[N] [BG933901] SK1-0182 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0182 similar to MHCI 3' UTR	$8,64X10^{-3}$	$6,59X10^{-3}$	0,53

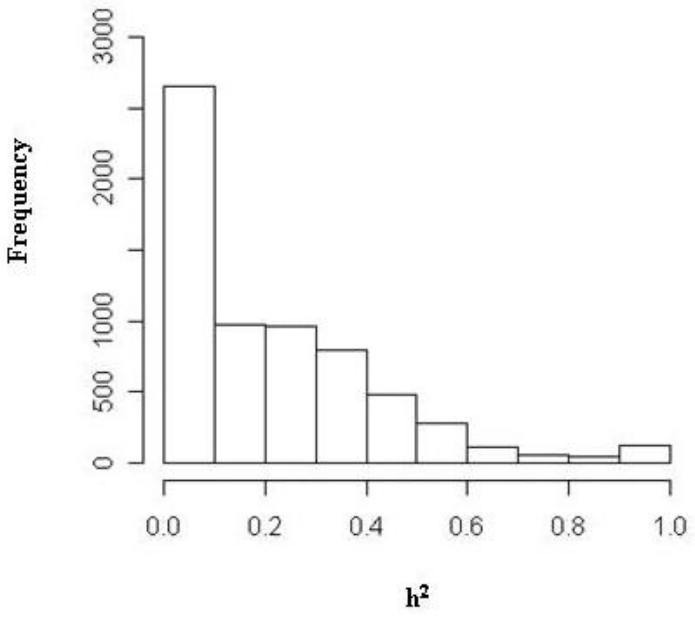
nwh^15^35^[P] [NP_000280] (NM_000289) phosphofructokinase, muscle; Phosphofructokinase, muscle type	$8,70X10^{-3}$	$6,65X10^{-3}$	0,65
rblb^2^39^[P] [NP_571655] (NM_131580) ribosomal protein, large, P0; acidic ribosomal phophoprotein	$8,71X10^{-3}$	$6,66X10^{-3}$	1,71
pha^501^154^[P] [AAD11573] aldolase B	$8,74X10^{-3}$	$6,69X10^{-3}$	1,64
rgb^533^231^[P] [1805343A] myosin:SUBUNIT=regulatory light chain	$8,84X10^{-3}$	$6,77X10^{-3}$	1,79
pitl^501^365^[P] [T50824] ribosomal protein S26	$8,86X10^{-3}$	$6,78X10^{-3}$	1,43
srkc^1^96^UNKNOWN	$9,08X10^{-3}$	$6,98X10^{-3}$	0,61
rgb^516^13^[P] [BAB22083] (AK002415) data source:MGD, source key:MGI:99500, evidence:ISS~epoxide hydrolase 2, cytoplasmic~putative	$9,16X10^{-3}$	$7,05X10^{-3}$	0,61
rgb^530^221^[P] [O93484] Collagen alpha 2(I) chain	$9,19X10^{-3}$	$7,07X10^{-3}$	1,93
nwh^8^40^[P] [NP_000597] (NM_000606) complement component 8, gamma polypeptide	$9,38X10^{-3}$	$7,22X10^{-3}$	0,55
rgb^519^25^[N] [BG934137] SK1-0428 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0428 similar to Ancient ubiquitious protein	$9,38X10^{-3}$	$7,23X10^{-3}$	0,57
pitl^510^129^[P] [AAD20460] (AF101385) ribosomal protein L11	$9,41X10^{-3}$	$7,25X10^{-3}$	1,37
rgb^533^117^[P] [AAH08146] (BC008146) Unknown (protein for MGC:18225)	$9,47X10^{-3}$	$7,31X10^{-3}$	1,11
rgb^525^241^[P] [NP_056629] (NM_015814) dickkopf homolog 3	$9,52X10^{-3}$	$7,35X10^{-3}$	0,68
pitl^512^151^[N] [BG934326] SK1-0625 Atlantic Salmon kidney Salmo salar cDNA clone SK1-0625 similar to polyposis protein	$1,00X10^{-2}$	$7,81X10^{-3}$	1,37
lna^16^73^[P] [AAC72281] (AF000167) 14.3 kDa perchloric acid soluble protein	$1,02X10^{-2}$	$7,91X10^{-3}$	0,75
rgb^507^211^[P] [AAD01429] S6 ribosomal protein	$1,02X10^{-2}$	$7,91X10^{-3}$	1,34
srkb^1^56^[N] [U25721] Oncorhynchus mykiss secreted protein, acidic, rich in cysteine (SPARC)	$1,02X10^{-2}$	$7,96X10^{-3}$	1,53
shc^503^284^[N] [BG936592] SS1-0974 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0974 similar to T-cell antigen receptor 3' UTR	$1,05X10^{-2}$	$8,20X10^{-3}$	0,59
pha^501^53^[P] [S68161] oligopeptide transport protein PepT1	$1,10X10^{-2}$	$8,72X10^{-3}$	1,18
rgb^508^40^[P] [AAG13347] (AF266227) serine-pyruvate aminotransferase	$1,12X10^{-2}$	$8,84X10^{-3}$	0,64
rgb^501^281^[N] [AF180488] Salmo salar clone BD3 beta-2 microglobulin (B2m)	$1,12X10^{-2}$	$8,84X10^{-3}$	1,23
rtfl^1^17^[P] [BAB28415] (AK012694) data source:MGD, source key:MGI:1929455, evidence:ISS~putative~ribosomal protein L23	$1,12X10^{-2}$	$8,85X10^{-3}$	1,48
lna^9^26^UNKNOWN	$1,12X10^{-2}$	$8,86X10^{-3}$	0,69
rtch^1^59^[P] [BAA13147] ferritin H-2	$1,12X10^{-2}$	$8,87X10^{-3}$	1,42
rgb^517^109^[P] [NP_031783] coproporphyrinogen oxidase; clone 560	$1,12X10^{-2}$	$8,92X10^{-3}$	0,63
rbla^1^45^[P] [NP_000962] (NM_000971) ribosomal protein L7; 60S ribosomal protein L7	$1,13X10^{-2}$	$8,95X10^{-3}$	0,70
rblb^1^48^[P] [NP_004441] (NM_004450) enhancer of rudimentary homolog; enhancer of rudimentary	$1,19X10^{-2}$	$9,48X10^{-3}$	1,70
nwh^17^63^UNKNOWN	$1,19X10^{-2}$	$9,49X10^{-3}$	0,68
rgb^531^16^[P] [T46285] hypothetical protein DKFZp566F194.1	$1,19X10^{-2}$	$9,49X10^{-3}$	0,68
plnb^506^20^[P] [AAL40376] (AC087333) High choriolytic enzyme 1	$1,19X10^{-2}$	$9,55X10^{-3}$	1,49
rteh^1^9^[P] [P14018] CLUSTERIN PRECURSOR (51.5 KD PROTEIN)	$1,20X10^{-2}$	$9,59X10^{-3}$	0,69
shc^504^71^[N] [AJ295231] Oncorhynchus mykiss iNOS/NOS2 gene for inducible nitric oxide synthase, exons 1-27	$1,20X10^{-2}$	$9,59X10^{-3}$	1,30
rgb^523^63^[P] [NP_004247] (NM_004256) organic cation transporter like 3	$1,20X10^{-2}$	$9,59X10^{-3}$	0,64
nwh^18^7^[N] [BG935831] SS1-0147 Atlantic Salmon spleen Salmo salar cDNA clone SS1-0147	$1,21X10^{-2}$	$9,74X10^{-3}$	0,67
nwh^14^21^UNKNOWN	$1,22X10^{-2}$	$9,75X10^{-3}$	0,77
srkc^3^42^[P] [BAB28364] (AK012624) data source:SPTR, source key:Q9P0Q1, evidence:ISS~homolog to HSPC227~putative	$1,23X10^{-2}$	$9,84X10^{-3}$	1,17
plnb^510^286^[P] [O60613] 15 kDa selenoprotein	$1,23X10^{-2}$	$9,89X10^{-3}$	1,28
plnb^505^247^[P] [AAG12207] lysozyme	$1,23X10^{-2}$	$9,90X10^{-3}$	0,56
nwh^16^61^[P] [Q64152] Transcription factor BTF3 (RNA polymerase B transcription factor 3)	$1,24X10^{-2}$	$9,94X10^{-3}$	1,57
rgb^514^364^[P] [AAG18369] immunoglobulin light chain	$1,24X10^{-2}$	$9,97X10^{-3}$	0,58



Suppl. Fig. 2.1 Proportions of genes on the array involved in several biological processes.


Frequency
0
10
20
30
40
-2
0
2
4
6
8
Normalised hybridisation signal
Frequency
0
10
20
30
40
-2
0
2
4
6
8
Normalised hybridisation signal
Frequency
0
10
20
30
40
-2
0
2
4
6
8
Normalised hybridisation signal

Suppl. Fig. 2.1



Suppl. Fig. 2.2

Supplementary Table 5.1 cDNA clones representing complement proteins as well as other acute phase protein coding genes that were significantly over-transcribed in saprolegniasis-affected fish. GeneBank accessions and names are given for each clone, as well as the P-value from the ANOVA permutation-based F-test, the corresponding Q-value, the estimated fold-change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously observed in fish infected with other agents.

cDNA clone	P-value	Q-value	Fold change	Reference*
CA039257 Complement C1r	$1.60X10^{-5}$	$2.70X10^{-2}$	40.04	
CB510638 Complement C1r	$3.19X10^{-5}$	$3.08X10^{-2}$	36.48	
CB510300 Complement C1r	$1.17X10^{-4}$	$3.36X10^{-2}$	19.42	
CA039055 Complement C2	$1.01X10^{-3}$	$3.60X10^{-2}$	1.79	
CB497763 Complement C3	$2.00X10^{-3}$	$3.64X10^{-2}$	2.12	1,2,4
CB493991 Complement C3	$2.67X10^{-3}$	$3.65X10^{-2}$	2.58	1,2,4
DN048269 Complement C3	$2.79X10^{-3}$	$3.65X10^{-2}$	2.53	1,2,4
CB497970 Complement C3	$3.14X10^{-3}$	$3.69X10^{-2}$	2.22	1,2,4
CB498425 Complement C3	$3.92X10^{-3}$	$3.77X10^{-2}$	2.12	1,2,4
CA038280 Complement C3	$4.37X10^{-3}$	$3.86X10^{-2}$	2.90	1,2,4
CA038163 Complement C3	$4.45X10^{-3}$	$3.86X10^{-2}$	2.19	1,2,4
CA052383 Complement C6	$2.68X10^{-3}$	$3.65X10^{-2}$	2.25	4
DN047719 Haptoglobin	$5.32X10^{-5}$	$3.36X10^{-2}$	37.96	1,2,3,6
CB511158 Precerebellin-like protein	$2.24X10^{-4}$	$3.60X10^{-2}$	12.89	1,2,3,6
CB497341 Precerebellin-like protein	$2.99X10^{-4}$	$3.60X10^{-2}$	9.95	1,2,3,6
CB502879 Differentially regulated trout protein 1	$1.60X10^{-4}$	$3.37X10^{-2}$	16.15	1,2,3,6
CB504496 Differentially regulated trout protein 1	$1.70X10^{-4}$	$3.39X10^{-2}$	19.04	1,2,3,6
CA046376 Differentially regulated trout protein 1	$1.81X10^{-4}$	$3.40X10^{-2}$	18.23	1,2,3,6
CK990321 Differentially regulated trout protein 1	$7.03X10^{-4}$	$3.60X10^{-2}$	3.54	1,2,3,6
CB492868 Hemopexin	$4.63X10^{-3}$	$3.92X10^{-2}$	2.80	2,6
CB511687 Ceruloplasmin	$4.33X10^{-3}$	$3.86X10^{-2}$	2.36	2,6
CA056544 Serotransferrin	$3.19X10^{-3}$	$3.69X10^{-2}$	2.70	2,6
CA040008 Serotransferrin	$3.69X10^{-3}$	$3.71X10^{-2}$	2.74	2,6
CA038612 Serotransferrin	$3.74X10^{-3}$	$3.72X10^{-2}$	3.25	2,6
CA051187 Mannose-binding protein C	$1.72X10^{-3}$	$3.62X10^{-2}$	2.15	1,2,3,6
CA056667 Mannose-binding protein C	$2.52X10^{-3}$	$3.64X10^{-2}$	2.23	1,2,3,6
CB511048 C type lectin receptor A	$7.45X10^{-5}$	$3.36X10^{-2}$	11.68	
CA056108 C type lectin receptor A	$1.17X10^{-4}$	$3.36X10^{-2}$	17.11	
CB496842 CD209 antigen-like protein E	$6.29X10^{-4}$	$3.60X10^{-2}$	8.31	

* 1: Tsoi et al. (2004), 2: Baynes (2001), 3: Ewart et al. (2005), 4: Meijer et al. (2005), 5: Kurobe et al. (2005), 6: Rise et al. (2004), 7: Byon et al. 2005.

Supplementary Table 5.2 cDNA clones corresponding to genes which were significantly over-transcribed in whole saprolegniasis-affected juvenile salmon and that were observed as over-transcribed in fish infected with other agents in previous studies. GeneBank accessions and names are given for each clone, as well as the P-value from the ANOVA permutation-based F-test, the corresponding Q-value, the estimated fold-change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously observed in fish infected with other agents.

cDNA clone	P-value	Q-value	Fold change	Reference*
CK990562 Cathepsin L	$9.58X10^{-5}$	$3.36X10^{-2}$	2.58	2,4
CB497026 Cathepsin L	$8.41X10^{-4}$	$3.60X10^{-2}$	1.94	2,4
CB511609 Cathepsin L	$1.04X10^{-3}$	$3.60X10^{-2}$	2.26	2,4
CA055974 Cathepsin L	$1.62X10^{-3}$	$3.61X10^{-2}$	2.31	2,4
CA051900 Cathepsin L	$1.79X10^{-3}$	$3.62X10^{-2}$	2.35	2,4
CA036990 Cathepsin L	$3.82X10^{-3}$	$3.75X10^{-2}$	1.88	2,4
CA767935 Cathepsin D	$3.52X10^{-3}$	$3.71X10^{-2}$	1.37	2,4
CB514083 Collagenase 3	$3.09X10^{-4}$	$3.60X10^{-2}$	13.15	4,5
CA064227 Collagenase 3	$3.84X10^{-4}$	$3.60X10^{-2}$	12.70	4,5
CK990871 Matrix metalloproteinase	$1.01X10^{-3}$	$3.60X10^{-2}$	4.27	1,3,4,6
CA054133 CXCR4	$4.80X10^{-3}$	$3.92X10^{-2}$	1.54	4
CB511689 NADPH oxidase flavocytochrome b small subunit	$4.45X10^{-3}$	$3.86X10^{-2}$	1.42	1,6
CA049007 NADPH oxidase flavocytochrome b small subunit	$4.79X10^{-3}$	$3.92X10^{-2}$	1.47	1,6
CA064428 6-phosphogluconate dehydrogenase	$3.39X10^{-3}$	$3.69X10^{-2}$	2.17	6
CK990447 Interferon inducible protein 1	$7.45X10^{-5}$	$3.36X10^{-2}$	7.67	7
CB500466 Interferon inducible protein 1	$1.49X10^{-4}$	$3.36X10^{-2}$	8.21	7
CB492283 Interferon inducible protein 1	$3.53X10^{-3}$	$3.71X10^{-2}$	1.53	7
CA049462 Phosphofructokinase/fructosebiphosphatase 2	$3.19X10^{-5}$	$3.08X10^{-2}$	3.79	1
CA037891 Leukocyte chemotaxin 2	$3.62X10^{-4}$	$3.60X10^{-2}$	12.18	1,2,3
CX984314 Leukocyte chemotaxin 2	$3.62X10^{-4}$	$3.60X10^{-2}$	6.73	1,2,3
CA056715 Jun-B.	$5.22X10^{-4}$	$3.60X10^{-2}$	3.73	1,5
CA054491 Jun-B.	$5.54X10^{-4}$	$3.60X10^{-2}$	4.47	1,5
CA056522 Jun-B.	$9.80X10^{-4}$	$3.60X10^{-2}$	4.63	1,5
CA041706 CCAAT/enhancer binding protein delta	$7.67X10^{-4}$	$3.60X10^{-2}$	4.94	4
CB497076 CCAAT/enhancer binding protein delta	$5.01X10^{-4}$	$3.60X10^{-2}$	5.51	4
CB492660 Biotinidase	$8.84X10^{-4}$	$3.60X10^{-2}$	3.90	2
CA039685 O-methyltransferase	$1.92X10^{-4}$	$3.41X10^{-2}$	15.52	1,3,6
CB509722 Plasma glutathione peroxidase	$4.36X10^{-3}$	$3.86X10^{-2}$	1.89	1,2,3,6

CB509470 Agglutination-aggregation factor 18K-LAF	$1.06X10^{-3}$	$3.60X10^{-2}$	3.49	1,3,6
CA044621 Agglutination-aggregation factor 18K-LAF	$1.14X10^{-3}$	$3.60X10^{-2}$	3.24	1,3,6
CB511230 TAP2	$9.58X10^{-5}$	$3.36X10^{-2}$	25.34	3,6
CA044190 TAP2	$4.57X10^{-3}$	$3.90X10^{-2}$	1.56	3,6
CB498320 S-100/ICaBP-like	$6.71X10^{-4}$	$3.60X10^{-2}$	2.43	6
CA061741 Ribosomal protein L13	$4.16X10^{-4}$	$3.60X10^{-2}$	3.59	2,3
CB505864 Ribosomal protein S27-like	$4.50X10^{-3}$	$3.87X10^{-2}$	1.78	1
CA058809 Vacuolar ATP synthase 16 kDa proteolipid subunit	$2.34X10^{-3}$	$3.64X10^{-2}$	1.90	1,4,6
CB493018 Vacuolar ATP synthase 16 kDa proteolipid subunit	$3.95X10^{-3}$	$3.78X10^{-2}$	1.70	1,4,6
CB514424 Vacuolar ATP synthase 16 kDa proteolipid subunit	$4.00X10^{-3}$	$3.78X10^{-2}$	1.95	1,4,6
CB497797 Vacuolar ATP synthase 16 kDa proteolipid subunit	$4.67X10^{-3}$	$3.92X10^{-2}$	1.67	1,4,6
CA052736 Vacuolar ATP synthase 16 kDa proteolipid subunit	$3.25X10^{-3}$	$3.69X10^{-2}$	1.38	1,4,6
CK990790 Thioredoxin interacting protein	$6.29X10^{-4}$	$3.60X10^{-2}$	2.71	5
CB494082 Thioredoxin interacting protein	$3.55X10^{-3}$	$3.71X10^{-2}$	2.52	5
CB498297 Thioredoxin	$1.35X10^{-3}$	$3.60X10^{-2}$	2.26	1,3,4
CA041451 Thioredoxin	$2.10X10^{-3}$	$3.64X10^{-2}$	2.14	1,3,4
CA042004 Solute carrier family 31, member 1	$7.99X10^{-4}$	$3.60X10^{-2}$	5.06	4
CA053629 Solute carrier family 37, member 4	$8.63X10^{-4}$	$3.60X10^{-2}$	1.94	4
CA058445 Solute carrier family 25, member 5	$2.72X10^{-3}$	$3.65X10^{-2}$	1.48	4
CA057185 Solute carrier family 25, member 5	$2.97X10^{-3}$	$3.67X10^{-2}$	1.30	4
CK990577 Solute carrier family 25, member 5	$3.38X10^{-3}$	$3.69X10^{-2}$	1.55	4
CA042906 Solute carrier family 25, member 5	$4.82X10^{-3}$	$3.92X10^{-2}$	1.38	4
CA056074 SQSTM1	$3.58X10^{-3}$	$3.71X10^{-2}$	2.17	3
CA046740 Sequestosome 1	$5.22X10^{-4}$	$3.60X10^{-2}$	2.85	3
CB509633 Thrombospondin-4	$9.53X10^{-4}$	$3.60X10^{-2}$	4.12	4
CA058127 Fibronectin	$3.05X10^{-3}$	$3.67X10^{-2}$	2.20	4
CB494160 26S proteasome regulatory subunit p27	$1.02X10^{-3}$	$3.60X10^{-2}$	1.97	4
CA050491 26S proteasome regulatory subunit p44.5	$3.21X10^{-3}$	$3.69X10^{-2}$	1.60	4
CB498208 Proteasome subunit alpha type 2	$2.75X10^{-3}$	$3.65X10^{-2}$	1.25	6
CA038358 Proteasome subunit alpha type 2	$1.21X10^{-3}$	$3.60X10^{-2}$	0.25	
CA769697 Endoplasmin	$1.26X10^{-3}$	$3.60X10^{-2}$	2.86	1
CB493461 Endoplasmin	$1.71X10^{-3}$	$3.62X10^{-2}$	2.89	1
CA059329 Endoplasmin	$2.08X10^{-3}$	$3.64X10^{-2}$	2.58	1
CB492804 Endoplasmin	$2.24X10^{-3}$	$3.64X10^{-2}$	2.61	1
CA057152 Endoplasmin	$3.16X10^{-3}$	$3.69X10^{-2}$	1.85	1

CA036673 Plasma retinol-binding protein	$1.42X10^{-3}$	$3.60X10^{-2}$	2.18	3
CB509509 Plasma retinol-binding protein	$1.84X10^{-3}$	$3.62X10^{-2}$	2.04	3
CB501248 Plasma retinol-binding protein	$2.36X10^{-3}$	$3.64X10^{-2}$	2.05	3
CB492550 Plasma retinol-binding protein	$2.87X10^{-3}$	$3.65X10^{-2}$	1.87	3
CB501051 Polyposis locus protein 1 homolog	$1.86X10^{-3}$	$3.62X10^{-2}$	1.94	1
CB492176 Polyposis locus protein 1 homolog	$2.83X10^{-3}$	$3.65X10^{-2}$	2.04	1
CB510945 Translation initiation factor 4E binding protein 3	$2.53X10^{-3}$	$3.64X10^{-2}$	1.76	4
CB488782 Probable RNA-dependent helicase p68	$4.95X10^{-3}$	$3.92X10^{-2}$	1.79	6
CB493619 Heat shock protein 90-beta	$2.55X10^{-3}$	$3.65X10^{-2}$	1.39	5,6
CA043166 CD63	$3.84X10^{-3}$	$3.75X10^{-2}$	1.56	4
CA041450 CD63	$4.08X10^{-3}$	$3.79X10^{-2}$	1.58	4
CA060339 Placental thrombin inhibitor	$3.61X10^{-3}$	$3.71X10^{-2}$	1.78	6
CB510709 Insulin-like growth factor I	$4.76X10^{-3}$	$3.92X10^{-2}$	1.96	6
CA769301 Histone H1-0	$2.45X10^{-4}$	$3.60X10^{-2}$	2.65	7
CB510848 Histone H1-0	$2.77X10^{-4}$	$3.60X10^{-2}$	2.34	7
CB497768 Histone H1-0	$4.05X10^{-4}$	$3.60X10^{-2}$	2.81	7
CA768242 Histone H1-0	$4.90X10^{-4}$	$3.60X10^{-2}$	2.44	7
CB491022 Histone H1-0	$2.22X10^{-3}$	$3.64X10^{-2}$	2.29	7
CA048620 Histone H1.3	$2.63X10^{-3}$	$3.65X10^{-2}$	0.65	
CB492849 Histone H2A.X.	$3.72X10^{-3}$	$3.72X10^{-2}$	0.60	
CB510682 H1fx-prov protein	$2.38X10^{-3}$	$3.64X10^{-2}$	1.92	

* 1: Tsoi et al. (2004), 2: Baynes (2001), 3: Ewart et al. (2005), 4: Meijer et al. (2005), 5: Kurobe et al. (2005), 6: Rise et al. (2004), 7: Byon et al. 2005.

Supplementary Table 5.3 cDNA clones corresponding to genes which were significantly over-transcribed in whole saprolegniasis-affected juvenile salmon and that were observed as over-transcribed in fish infected with other agents in previous studies. GeneBank accessions and names are given for each clone, as well as the P-value from the ANOVA permutation-based F-test, the corresponding Q-value, the estimated fold-change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously observed in fish infected with other agents.

cDNA clone	P-value	Q-value	Fold change	Reference*
CB509685 Selenoprotein P, plasma, 1b	$4.32X10^{-3}$	$3.86X10^{-2}$	0.68	3
CB509843 Deoxyribonuclease I-like 3	$2.13X10^{-4}$	$3.60X10^{-2}$	0.25	4
CK990723 Deoxyribonuclease I-like 3	$2.35X10^{-4}$	$3.60X10^{-2}$	0.25	4
CB488336 Collagen alpha 1(I)	$1.09X10^{-3}$	$3.60X10^{-2}$	0.31	3,4
CB505502 Collagen alpha 1(I)	$1.39X10^{-3}$	$3.60X10^{-2}$	0.36	3,4
CA058861 Collagen alpha 1(I)	$1.45X10^{-3}$	$3.60X10^{-2}$	0.34	3,4
CB498132 Collagen alpha 1(I)	$1.51X10^{-3}$	$3.60X10^{-2}$	0.28	3,4
CB497315 Collagen alpha 1(I)	$1.55X10^{-3}$	$3.60X10^{-2}$	0.30	3,4
CB500684 Collagen alpha 1(I)	$1.64X10^{-3}$	$3.61X10^{-2}$	0.32	3,4
CA047568 Collagen alpha 1(I)	$1.65X10^{-3}$	$3.61X10^{-2}$	0.36	3,4
CB494000 Collagen alpha 1(I)	$1.89X10^{-3}$	$3.62X10^{-2}$	0.29	3,4
CB494129 Collagen alpha 1(I)	$2.01X10^{-3}$	$3.64X10^{-2}$	0.33	3,4
CA061635 Collagen alpha 1(I)	$2.03X10^{-3}$	$3.64X10^{-2}$	0.30	3,4
CB494364 Collagen alpha 1(I)	$2.13X10^{-3}$	$3.64X10^{-2}$	0.44	3,4
CB500533 Collagen alpha 1(I)	$2.18X10^{-3}$	$3.64X10^{-2}$	0.29	3,4
CA064541 Collagen alpha 1(I)	$2.61X10^{-3}$	$3.65X10^{-2}$	0.32	3,4
CB493106 Collagen alpha 1(I)	$2.73X10^{-3}$	$3.65X10^{-2}$	0.35	3,4
CB498182 Collagen alpha 1(I)	$2.76X10^{-3}$	$3.65X10^{-2}$	0.36	3,4
CB494505 Collagen alpha 1(I)	$2.80X10^{-3}$	$3.65X10^{-2}$	0.35	3,4
CB512039 Collagen alpha 1(I)	$3.18X10^{-3}$	$3.69X10^{-2}$	0.39	3,4
CB503528 UDP-glucuronosyltransferase 2B5	$1.80X10^{-3}$	$3.62X10^{-2}$	0.65	1
CB497940 UDP-glucuronosyltransferase 2B5	$4.66X10^{-3}$	$3.92X10^{-2}$	0.67	1
CA059209 Ependymin	$2.66X10^{-3}$	$3.65X10^{-2}$	0.62	3
CA041709 Ependymin	$3.91X10^{-3}$	$3.77X10^{-2}$	0.64	3
CB509787 Ependymin	$3.95X10^{-3}$	$3.78X10^{-2}$	0.63	3
CB511383 Myosin regulatory light chain 2	$4.98X10^{-3}$	$3.92X10^{-2}$	0.76	3
CB506710 Troponin I	$3.70X10^{-3}$	$3.71X10^{-2}$	0.58	4
CB509935 Type II keratin E2	$1.52X10^{-3}$	$3.60X10^{-2}$	0.36	4†
CB493225 Type II keratin E2	$1.78X10^{-3}$	$3.62X10^{-2}$	0.39	
CB510328 Keratin 13	$1.60X10^{-3}$	$3.61X10^{-2}$	0.49	
CB496484 Keratin, type II cytoskeletal 6A	$1.83X10^{-3}$	$3.62X10^{-2}$	0.75	
CB494649 Keratin 12	$1.87X10^{-3}$	$3.62X10^{-2}$	0.47	
CK991350 Keratin, type I cytoskeletal 14	$2.47X10^{-3}$	$3.64X10^{-2}$	0.54	
CB493426 Keratin, type I cytoskeletal 13	$3.49X10^{-3}$	$3.71X10^{-2}$	0.45	
CB502500 Keratin, type II cytoskeletal 6A	$4.02X10^{-3}$	$3.78X10^{-2}$	0.68	

CB510882 Keratin, type II cytoskeletal 6A	$4.63X10^{-3}$	$3.92X10^{-2}$	0.71
CB493151 Keratin 12	$4.88X10^{-3}$	$3.92X10^{-2}$	0.55

* 1: Tsoi et al. (2004), 2: Baynes (2001), 3: Ewart et al. (2005), 4: Meijer et al. (2005), 5: Kurobe et al. (2005), 6: Rise et al. (2004), 7: Byon et al. 2005.

† The observed keratin type was not specified in Meijer et al. (2005).

Supplementary Table 5.4 cDNA clones corresponding to genes observed as differentially expressed in different directions in this study and in previous studies. GeneBank accessions and names are given for each clone, as well as the P-value from the ANOVA permutation-based F-test, the corresponding Q-value, the estimated fold-change (expression level in infected fish over that in healthy fish) and references in which the gene differential expression was previously observed in fish infected with other infectious agents.

cDNA clone	P-value	Q-value	Fold change	Reference*
CB492457 Tob1 protein	$2.28X10^{-3}$	$3.64X10^{-2}$	2.41	Under-transcribed in 3
CA052552 B-cell translocation gene 1	$2.28X10^{-3}$	$3.64X10^{-2}$	1.99	Under-transcribed in 3
CA056659 B-cell translocation gene 1	$4.89X10^{-3}$	$3.92X10^{-2}$	1.56	Under-transcribed in 3
CB509870 Prostaglandine D synthase	$1.26X10^{-3}$	$3.60X10^{-2}$	0.38	Over-transcribed in 2,3
CA045988 Prostaglandine D synthase	$2.43X10^{-3}$	$3.64X10^{-2}$	0.41	Over-transcribed in 2,3
CA061964 Glutathione S-transferase kappa 1	$2.57X10^{-3}$	$3.65X10^{-2}$	0.70	Over-transcribed in 3,6
CB496642 Microsomal glutathione S-transferase 3	$2.93X10^{-3}$	$3.67X10^{-2}$	0.82	Over-transcribed in 3,6
CB492698 Microsomal glutathione S-transferase 3	$4.25X10^{-3}$	$3.82X10^{-2}$	0.78	Over-transcribed in 3,6
CA051810 High mobility group protein 1	$2.89X10^{-3}$	$3.67X10^{-2}$	0.70	Over-transcribed in 6
CA044987 Putative steroid dehydrogenase KIK-I	$4.85X10^{-3}$	$3.92X10^{-2}$	0.65	Over-transcribed in 6
CB496691 ATP synthase alpha chain	$1.32X10^{-3}$	$3.60X10^{-2}$	0.25	Over-transcribed in 6
CA038358 Proteasome subunit alpha type 2	$1.21X10^{-3}$	$3.60X10^{-2}$	0.25	Over-transcribed in 6

* 1: Tsoi et al. (2004), 2: Baynes (2001), 3: Ewart et al. (2005), 4: Meijer et al. (2005), 5: Kurobe et al. (2005), 6: Rise et al. (2004), 7: Byon et al. 2005.

Supplementary Table 5.5 cDNA clones corresponding to genes which were significantly differently transcribed in whole saprolegniasis-affected juvenile salmon and that had not observed as so over-transcribed in fish infected with other agents in previous studies. Been GeneBank accessions and names are given for each clone, as well as the P-value from the ANOVA permutation-based F-test, the corresponding Q-value and the estimated fold- change (expression level in infected fish over that in healthy fish). Genes were classified in different functional categories using data from the literature.

	P-value	Q-value	Fold-change
Immunity-related gene			
CB497818 Prostaglandin-H2 D-isomerase precursor	$2.97X10^{-3}$	$3.67X10^{-2}$	0.49
CA043680 Acidic chitinase L	$2.33X10^{-3}$	$3.64X10^{-2}$	0.55
CK991302 Calreticulin	$2.42X10^{-3}$	$3.64X10^{-2}$	2.64
CK990225 Calreticulin	$2.68X10^{-3}$	$3.65X10^{-2}$	2.27
CK990254 G protein-coupled receptor 2	$2.92X10^{-3}$	$3.67X10^{-2}$	1.76
CB510754 FYN-binding protein	$3.37X10^{-3}$	$3.69X10^{-2}$	1.51
CB507502 Small inducible cytokine	$9.69X10^{-4}$	$3.60X10^{-2}$	3.11
CB510320 Small inducible cytokine	$2.33X10^{-3}$	$3.64X10^{-2}$	0.40
CB499401 Myeloperoxidase	$1.48X10^{-3}$	$3.60X10^{-2}$	0.45
CB498878 Glucose-6-phosphate 1-dehydrogenase X	$3.20X10^{-4}$	$3.60X10^{-2}$	1.98
X10xtracelluar matrix components			
CB498566 Collagen alpha 1(X)	$7.46X10^{-4}$	$3.60X10^{-2}$	0.21
CB498386 Collagen alpha 1(X)	$3.63X10^{-3}$	$3.71X10^{-2}$	0.39
CB486593 Collagen alpha 2(I)	$1.21X10^{-3}$	$3.60X10^{-2}$	0.33
CB507670 Collagen alpha 2(I)	$1.24X10^{-3}$	$3.60X10^{-2}$	0.30
CB492485 Collagen alpha 2(I)	$1.44X10^{-3}$	$3.60X10^{-2}$	0.32
CB493159 Collagen alpha 2(I)	$1.53X10^{-3}$	$3.60X10^{-2}$	0.29
CB509170 Collagen alpha 2(I)	$1.68X10^{-3}$	$3.61X10^{-2}$	0.30
CB505787 Collagen alpha 2(I)	$1.92X10^{-3}$	$3.62X10^{-2}$	0.32
CB498518 Collagen alpha 2(I)	$2.24X10^{-3}$	$3.64X10^{-2}$	0.34
CB492605 Collagen alpha 1(X)	$2.48X10^{-3}$	$3.64X10^{-2}$	0.28
CB494291 Collagen a3(I)	$3.75X10^{-3}$	$3.72X10^{-2}$	0.35
CB488287 SPARC	$3.96X10^{-3}$	$3.78X10^{-2}$	0.60
Miscellanous			
CA049957 Immunophilin FKBP12.6	$1.46X10^{-3}$	$3.60X10^{-2}$	1.64
CA037683 Delta-6 fatty acyl desaturase	$2.04X10^{-3}$	$3.64X10^{-2}$	0.79
CA042983 Diamine acetyltransferase 1	$3.20X10^{-4}$	$3.60X10^{-2}$	4.39

CA056199 Diamine acetyltransferase 1	$5.43X10^{-4}$	$3.60X10^{-2}$	3.74
CK990709 Diamine acetyltransferase 1	$9.90X10^{-4}$	$3.60X10^{-2}$	1.92
CB492733 Diamine acetyltransferase 1	$3.29X10^{-3}$	$3.69X10^{-2}$	1.70
CA064093 Immunoglobulin heavy chain binding protein	$3.52X10^{-4}$	$3.60X10^{-2}$	3.57
CA057213 ALDH class 2	$1.18X10^{-3}$	$3.60X10^{-2}$	0.64
CA046194 Transposase	$1.67X10^{-3}$	$3.61X10^{-2}$	2.45
CB488691 Nucleophosmin	$3.56X10^{-3}$	$3.71X10^{-2}$	2.58
CB502353 EH-domain containing protein 1	$4.69X10^{-4}$	$3.60X10^{-2}$	2.33
Electron transport chain			
CB494539 Cytochrome c	$4.58X10^{-4}$	$3.60X10^{-2}$	2.12
CB508397 Cytochrome c	$2.17X10^{-3}$	$3.64X10^{-2}$	2.36
CA038364 Ubiquinol-cytochrome c reductase complex 7.2 kDa protein	$4.05X10^{-4}$	$3.60X10^{-2}$	8.47
CB496924 Cytochrome c oxidase subunit IV isoform 1	$2.62X10^{-3}$	$3.65X10^{-2}$	0.68
CB511353 Cytochrome c oxidase polypeptide VIIa	$4.95X10^{-3}$	$3.92X10^{-2}$	0.74
CB497786 NADH-ubiquinone oxidoreductase 30 kDa subunit	$2.00X10^{-3}$	$3.64X10^{-2}$	0.78
CB497247 NADH-ubiquinone oxidoreductase 30 kDa subunit	$4.17X10^{-3}$	$3.81X10^{-2}$	0.84
CB493686 NADH-ubiquinone oxidoreductase B18 subunit	$4.75X10^{-3}$	$3.92X10^{-2}$	0.91
CA047666 ATP synthase e chain	$1.32X10^{-3}$	$3.60X10^{-2}$	0.74
CB488180 ATP synthase D chain	$3.35X10^{-3}$	$3.69X10^{-2}$	1.95
BU965693 ATP synthase coupling factor 6	$4.18X10^{-3}$	$3.81X10^{-2}$	0.40
Uncoupling protein			
CB517525 Uncoupling protein 2	$1.49X10^{-3}$	$3.60X10^{-2}$	2.46
CB492139 Uncoupling protein 2	$2.41X10^{-3}$	$3.64X10^{-2}$	1.99
Ion transport			
CB510517 Na+/K+ ATPase 1	$1.88X10^{-3}$	$3.62X10^{-2}$	1.66
CK990492 Na+/K+ ATPase 3	$3.32X10^{-3}$	$3.69X10^{-2}$	1.75
CA041920 Na+/K+ ATPase alpha subunit isoform 1b/ii	$4.91X10^{-3}$	$3.92X10^{-2}$	1.49
CB497552 Na+/K+ ATPase alpha subunit isoform 1b/ii	$4.65X10^{-3}$	$3.92X10^{-2}$	0.66
Signal transduction			
CB500095 Mcl-1a	$1.11X10^{-3}$	$3.60X10^{-2}$	2.45

CB515252 Tumor-associated calcium signal transducer	$1.65X10^{-3}$	$3.61X10^{-2}$	2.14
CB509663 Centd2	$4.68X10^{-3}$	$3.92X10^{-2}$	1.50
Transcription/transcription activation			
CB496720 40S ribosomal protein S16.	$3.42X10^{-3}$	$3.69X10^{-2}$	2.53
CB496771 40S ribosomal protein S16.	$3.68X10^{-3}$	$3.71X10^{-2}$	3.02
CB509763 RNA polymerase II transcriptional coactivator p15	$7.14X10^{-4}$	$3.60X10^{-2}$	0.58
CK990730 RNA polymerase II transcriptional coactivator p15	$9.69X10^{-4}$	$3.60X10^{-2}$	0.70
CA062504 RNA polymerase II transcriptional coactivator p15	$4.54X10^{-3}$	$3.88X10^{-2}$	0.69
CA037255 Zinc finger A20 domain containing 2, like	$9.27X10^{-4}$	$3.60X10^{-2}$	3.63
CA048582 Zinc finger protein 330	$1.14X10^{-3}$	$3.60X10^{-2}$	2.06
CA043387 XP8	$1.36X10^{-3}$	$3.60X10^{-2}$	2.83
CB497385 XP8	$2.56X10^{-3}$	$3.65X10^{-2}$	3.13
CB490586 XP8	$3.59X10^{-3}$	$3.71X10^{-2}$	2.33
CB507234 Aryl hydrocarbon receptor 2b	$8.95X10^{-4}$	$3.60X10^{-2}$	2.84
CB511953 Aryl hydrocarbon receptor 2b	$1.93X10^{-3}$	$3.62X10^{-2}$	1.92
CB514828 RNA-binding region containing protein 2	$3.16X10^{-3}$	$3.69X10^{-2}$	0.79
CB492815 Similar to thyroid hormone receptor interactor 3	$4.70X10^{-3}$	$3.92X10^{-2}$	0.55
Metal ion binding protein			
CB510789 Selenoprotein W	$1.04X10^{-3}$	$3.60X10^{-2}$	0.77
CB510653 Metallothionein B	$1.75X10^{-3}$	$3.62X10^{-2}$	1.75
CB492197 Metallothionein-I	$2.82X10^{-3}$	$3.65X10^{-2}$	2.05
Pyrimidine biosynthesis			
CB491425 CTP synthase	$1.12X10^{-3}$	$3.60X10^{-2}$	2.09
CB510645 Enhancer of rudimentary homolog	$1.15X10^{-3}$	$3.60X10^{-2}$	2.29
Protein degradation			
CB513466 Similar to Ubiquitin binding enzyme	$1.75X10^{-3}$	$3.62X10^{-2}$	1.90
CB516480 Ubiquitin-conjugating enzyme E2 A	$3.25X10^{-3}$	$3.69X10^{-2}$	2.09
CA057329 Ubiquitin-conjugating enzyme E2 D2	$4.00X10^{-3}$	$3.78X10^{-2}$	1.60
CA060138 F-box only protein 2	$2.86X10^{-3}$	$3.65X10^{-2}$	2.39

CB505442 Calpain small subunit 1	$3.63X10^{-3}$	$3.71X10^{-2}$	1.65
CB497353 His-tagged cytosolic leucine aminopeptidase	$4.13X10^{-3}$	$3.80X10^{-2}$	0.86
CB492942 Meprin A alpha-subunit	$3.66X10^{-3}$	$3.71X10^{-2}$	0.53
Protein localisation			
CB496732 Signal recognition particle 9 kDa protein	$4.21X10^{-3}$	$3.81X10^{-2}$	0.65
CB492591 SSR alpha subunit	$3.89X10^{-3}$	$3.77X10^{-2}$	0.57
Protein folding			
CK991247 Protein disulfide-isomerase A3	$1.77X10^{-3}$	$3.62X10^{-2}$	1.62
CB494398 Protein disulfide-isomerase A3	$4.49X10^{-3}$	$3.87X10^{-2}$	1.83
CA048973 Protein disulfide-isomerase A3	$2.46X10^{-3}$	$3.64X10^{-2}$	2.29
CA048996 Protein disulfide-isomerase A3	$2.99X10^{-3}$	$3.67X10^{-2}$	1.80
DNA replication/repair			
CK990880 DNA polymerase delta subunit 4	$1.81X10^{-3}$	$3.62X10^{-2}$	0.67
CB498769 Proliferating cell nuclear antigen	$4.05X10^{-3}$	$3.79X10^{-2}$	0.70
Cellular structure			
CA047595 Actin-5C.	$1.35X10^{-3}$	$3.60X10^{-2}$	1.94
CA042507 Actin-5C.	$4.26X10^{-3}$	$3.82X10^{-2}$	2.12
CA039603 Beta-actin	$4.09X10^{-3}$	$3.79X10^{-2}$	1.30
Cell adhesion/cellular junction			
CB497400 Periostin	$1.96X10^{-3}$	$3.64X10^{-2}$	0.39
CA767874 Neural-cadherin	$2.11X10^{-3}$	$3.64X10^{-2}$	2.13
CB496570 Epithelial-cadherin	$3.42X10^{-3}$	$3.69X10^{-2}$	1.77
CB510540 Claudin-3	$2.39X10^{-3}$	$3.64X10^{-2}$	2.21
CB511661 Integrin, beta-like 1	$2.64X10^{-3}$	$3.65X10^{-2}$	2.16
CB509802 Nicotinamide riboside kinase 2	$2.45X10^{-3}$	$3.64X10^{-2}$	1.70
CB492503 Nicotinamide riboside kinase 2	$4.91X10^{-3}$	$3.92X10^{-2}$	1.88
Fatty acid absorbtion			

CB503762 Fatty acid-binding protein, intestinal	$2.09X10^{-3}$	$3.64X10^{-2}$	0.33
CB517047 Fatty acid-binding protein, intestinal	$2.79X10^{-3}$	$3.65X10^{-2}$	0.38
CB498443 Fatty acid-binding protein, intestinal	$2.81X10^{-3}$	$3.65X10^{-2}$	0.30
CA044478 Fatty acid-binding protein, intestinal	$3.45X10^{-3}$	$3.69X10^{-2}$	0.35
CB498358 Fatty acid-binding protein, intestinal	$4.92X10^{-3}$	$3.92X10^{-2}$	0.64
Translation			
CA057937 Protein translation factor SUI1 homolog GC20	$3.32X10^{-3}$	$3.69X10^{-2}$	1.75
CB489126 Protein translation factor SUI1 homolog GC20	$3.27X10^{-3}$	$3.69X10^{-2}$	1.74
CK990998 [GO] [P48024] Protein translation factor SUI1 homolog	$3.69X10^{-3}$	$3.71X10^{-2}$	1.83
CA037863 Small nuclear ribonucleoprotein E	$2.59X10^{-3}$	$3.65X10^{-2}$	2.57
Sub-cellular organisation			
CB496719 Mid1 interacting protein 1	$2.78X10^{-3}$	$3.65X10^{-2}$	0.61
CB497886 Transport protein SEC61 gamma subunit	$3.22X10^{-3}$	$3.69X10^{-2}$	1.89
CB515714 Gelsolin	$3.82X10^{-3}$	$3.75X10^{-2}$	0.62
Neuronal function			
CA063277 Glutamine synthetase	$4.94X10^{-3}$	$3.92X10^{-2}$	1.90
CB514092 Glutamine synthetase	$4.96X10^{-3}$	$3.92X10^{-2}$	1.75
Clones with known homologs but no known function			
CB510827 Clone IMAGE:5600737	$2.42X10^{-3}$	$3.64X10^{-2}$	0.31
CB510327 FLJ13263	$3.75X10^{-3}$	$3.72X10^{-2}$	2.47
CA055675 Similar to dJ40E16.3	$4.88X10^{-3}$	$3.92X10^{-2}$	2.13
CA042416 Zonadhesin-like gene	$1.19X10^{-3}$	$3.60X10^{-2}$	0.56
BU965755 Gene model 631	$1.41X10^{-3}$	$3.60X10^{-2}$	0.30
CA040745 ReO_6	$1.41X10^{-3}$	$3.60X10^{-2}$	1.51
CB515475 ChaC-like protein	$2.70X10^{-3}$	$3.65X10^{-2}$	2.36
CB505255 ATP-binding cassette sub-family F member 2	$2.93X10^{-3}$	$3.67X10^{-2}$	1.75
CA043517 Ariadne-2 protein homolog	$2.95X10^{-3}$	$3.67X10^{-2}$	1.70
CB504548 NDRG3 protein	$3.10X10^{-3}$	$3.69X10^{-2}$	0.88
CA038241 Toxin-1	$3.46X10^{-3}$	$3.69X10^{-2}$	1.83
CB510652 NipSnap2 protein	$4.15X10^{-3}$	$3.81X10^{-2}$	0.85

Supplementary Table 5.6 cDNA clones marked as "unknown" on the microarray used in this study, meaning that they did not generate any BLAST hits with e-values less than 1X10-15 and an informative name, and which were significantly differentially transcribed in whole saprolegniasis-affected juvenile salmon. GeneBank accessions are given for each clone, as well as the P-value from the ANOVA permutation-based F-test, the corresponding Q-value and the estimated fold-change (expression level in infected fish over that in healthy fish).

cDNA clone	P-value	Q-value	Fold change
CA038202 UNKNOWN	$1.38X10^{-4}$	$3.36X10^{-2}$	18.38
CK990934 UNKNOWN	$2.56X10^{-4}$	$3.60X10^{-2}$	0.28
CB510363 UNKNOWN	$2.67X10^{-4}$	$3.60X10^{-2}$	2.75
CA043945 UNKNOWN	$3.41X10^{-4}$	$3.60X10^{-2}$	3.74
CA040785 UNKNOWN	$3.73X10^{-4}$	$3.60X10^{-2}$	4.35
CA037557 UNKNOWN	$4.37X10^{-4}$	$3.60X10^{-2}$	2.53
CB493980 UNKNOWN	$4.37X10^{-4}$	$3.60X10^{-2}$	0.41
CB512519 UNKNOWN	$4.80X10^{-4}$	$3.60X10^{-2}$	0.54
CA064590 UNKNOWN	$4.80X10^{-4}$	$3.60X10^{-2}$	2.56
CK990499 UNKNOWN	$5.54X10^{-4}$	$3.60X10^{-2}$	2.71
CB501749 UNKNOWN	$5.75X10^{-4}$	$3.60X10^{-2}$	2.32
CA038736 UNKNOWN	$5.75X10^{-4}$	$3.60X10^{-2}$	2.23
CA038688 UNKNOWN	$5.97X10^{-4}$	$3.60X10^{-2}$	2.95
CA063116 UNKNOWN	$5.97X10^{-4}$	$3.60X10^{-2}$	3.92
CB501135 UNKNOWN	$6.18X10^{-4}$	$3.60X10^{-2}$	2.34
CA063008 UNKNOWN	$6.50X10^{-4}$	$3.60X10^{-2}$	2.48
CA044682 UNKNOWN	$6.60X10^{-4}$	$3.60X10^{-2}$	2.97
CB496744 UNKNOWN	$6.71X10^{-4}$	$3.60X10^{-2}$	5.58
CB509739 UNKNOWN	$6.87X10^{-4}$	$3.60X10^{-2}$	6.94
CA061749 UNKNOWN	$7.03X10^{-4}$	$3.60X10^{-2}$	1.95
CA037636 UNKNOWN	$7.24X10^{-4}$	$3.60X10^{-2}$	4.90
CB500588 UNKNOWN	$7.46X10^{-4}$	$3.60X10^{-2}$	0.33
CK990495 UNKNOWN	$7.56X10^{-4}$	$3.60X10^{-2}$	1.91
CA038875 UNKNOWN	$7.78X10^{-4}$	$3.60X10^{-2}$	2.83
CA053760 UNKNOWN	$8.09X10^{-4}$	$3.60X10^{-2}$	1.89
CA770640 UNKNOWN	$8.20X10^{-4}$	$3.60X10^{-2}$	2.14
CB517483 UNKNOWN	$8.31X10^{-4}$	$3.60X10^{-2}$	2.75
CA060885 UNKNOWN	$8.31X10^{-4}$	$3.60X10^{-2}$	1.90
CA055146 UNKNOWN	$8.63X10^{-4}$	$3.60X10^{-2}$	2.58
CB492818 UNKNOWN	$8.73X10^{-4}$	$3.60X10^{-2}$	2.47
CA051732 UNKNOWN	$9.05X10^{-4}$	$3.60X10^{-2}$	3.54
CB505814 UNKNOWN	$9.37X10^{-4}$	$3.60X10^{-2}$	0.55
CK990753 UNKNOWN	$9.37X10^{-4}$	$3.60X10^{-2}$	2.65
CA044333 UNKNOWN	$1.05X10^{-3}$	$3.60X10^{-2}$	0.53
CK991328 UNKNOWN	$1.09X10^{-3}$	$3.60X10^{-2}$	4.77
CB497895 UNKNOWN	$1.11X10^{-3}$	$3.60X10^{-2}$	0.30
CA058990 UNKNOWN	$1.17X10^{-3}$	$3.60X10^{-2}$	2.48
CB510523 UNKNOWN	$1.19X10^{-3}$	$3.60X10^{-2}$	0.65
CB503047 UNKNOWN	$1.25X10^{-3}$	$3.60X10^{-2}$	0.57
CA049920 UNKNOWN	$1.27X10^{-3}$	$3.60X10^{-2}$	2.05
CB493886 UNKNOWN	$1.28X10^{-3}$	$3.60X10^{-2}$	0.56
CA057909 UNKNOWN	$1.29X10^{-3}$	$3.60X10^{-2}$	2.10
CA053790 UNKNOWN	$1.31X10^{-3}$	$3.60X10^{-2}$	1.65

CB511254 UNKNOWN	$1.33X10^{-3}$	$3.60X10^{-2}$	2.23
CB494446 UNKNOWN	$1.37X10^{-3}$	$3.60X10^{-2}$	2.08
CA056322 UNKNOWN	$1.45X10^{-3}$	$3.60X10^{-2}$	2.58
CB507066 UNKNOWN	$1.48X10^{-3}$	$3.60X10^{-2}$	0.29
CB502618 UNKNOWN	$1.54X10^{-3}$	$3.60X10^{-2}$	3.14
CK990820 UNKNOWN	$1.54X10^{-3}$	$3.60X10^{-2}$	1.76
CB510484 UNKNOWN	$1.57X10^{-3}$	$3.60X10^{-2}$	4.12
CK990939 UNKNOWN	$1.59X10^{-3}$	$3.61X10^{-2}$	2.36
CA058804 UNKNOWN	$1.60X10^{-3}$	$3.61X10^{-2}$	2.89
CA039238 UNKNOWN	$1.62X10^{-3}$	$3.61X10^{-2}$	3.70
CA063841 UNKNOWN	$1.68X10^{-3}$	$3.61X10^{-2}$	2.03
CB509762 UNKNOWN	$1.72X10^{-3}$	$3.62X10^{-2}$	0.50
CB505667 UNKNOWN	$1.79X10^{-3}$	$3.62X10^{-2}$	0.51
CB508532 UNKNOWN	$1.85X10^{-3}$	$3.62X10^{-2}$	3.01
CA063605 UNKNOWN	$1.86X10^{-3}$	$3.62X10^{-2}$	1.73
CB510598 UNKNOWN	$1.93X10^{-3}$	$3.62X10^{-2}$	3.64
CK991084 UNKNOWN	$1.94X10^{-3}$	$3.62X10^{-2}$	0.33
CB517822 UNKNOWN	$1.97X10^{-3}$	$3.64X10^{-2}$	0.59
CA057119 UNKNOWN	$2.03X10^{-3}$	$3.64X10^{-2}$	1.94
CA051500 UNKNOWN	$2.06X10^{-3}$	$3.64X10^{-2}$	2.84
CK990878 UNKNOWN	$2.07X10^{-3}$	$3.64X10^{-2}$	2.54
CB501817 UNKNOWN	$2.09X10^{-3}$	$3.64X10^{-2}$	1.88
CB510626 UNKNOWN	$2.13X10^{-3}$	$3.64X10^{-2}$	0.46
CA052769 UNKNOWN	$2.15X10^{-3}$	$3.64X10^{-2}$	3.48
CB510654 UNKNOWN	$2.18X10^{-3}$	$3.64X10^{-2}$	2.85
CB500863 UNKNOWN	$2.19X10^{-3}$	$3.64X10^{-2}$	1.99
CB510093 UNKNOWN	$2.20X10^{-3}$	$3.64X10^{-2}$	1.83
CA054628 UNKNOWN	$2.21X10^{-3}$	$3.64X10^{-2}$	2.17
CB514276 UNKNOWN	$2.26X10^{-3}$	$3.64X10^{-2}$	2.60
CB507979 UNKNOWN	$2.27X10^{-3}$	$3.64X10^{-2}$	0.48
CA040136 UNKNOWN	$2.29X10^{-3}$	$3.64X10^{-2}$	1.68
CA058034 UNKNOWN	$2.30X10^{-3}$	$3.64X10^{-2}$	2.28
CB507633 UNKNOWN	$2.35X10^{-3}$	$3.64X10^{-2}$	0.47
CB501151 UNKNOWN	$2.40X10^{-3}$	$3.64X10^{-2}$	2.66
CA044624 UNKNOWN	$2.47X10^{-3}$	$3.64X10^{-2}$	1.91
CA048593 UNKNOWN	$2.50X10^{-3}$	$3.64X10^{-2}$	1.48
CB516782 UNKNOWN	$2.52X10^{-3}$	$3.64X10^{-2}$	0.61
CA040167 UNKNOWN	$2.63X10^{-3}$	$3.65X10^{-2}$	1.52
CA046630 UNKNOWN	$2.69X10^{-3}$	$3.65X10^{-2}$	0.41
CB509397 UNKNOWN	$2.76X10^{-3}$	$3.65X10^{-2}$	2.21
CB516592 UNKNOWN	$2.84X10^{-3}$	$3.65X10^{-2}$	2.22
CA037858 UNKNOWN	$2.86X10^{-3}$	$3.65X10^{-2}$	2.47
CB516800 UNKNOWN	$2.95X10^{-3}$	$3.67X10^{-2}$	0.73
CA057768 UNKNOWN	$2.98X10^{-3}$	$3.67X10^{-2}$	1.54
CA048077 UNKNOWN	$3.01X10^{-3}$	$3.67X10^{-2}$	0.49
CB501053 UNKNOWN	$3.03X10^{-3}$	$3.67X10^{-2}$	0.58
BU965696 UNKNOWN	$3.04X10^{-3}$	$3.67X10^{-2}$	1.89
CK990528 UNKNOWN	$3.04X10^{-3}$	$3.67X10^{-2}$	1.92
CB510279 UNKNOWN	$3.07X10^{-3}$	$3.68X10^{-2}$	0.47

CA052620 UNKNOWN	$3.09X10^{-3}$	$3.69X10^{-2}$	2.18
CA051860 UNKNOWN	$3.10X10^{-3}$	$3.69X10^{-2}$	1.86
CA063043 UNKNOWN	$3.14X10^{-3}$	$3.69X10^{-2}$	1.83
CB493974 UNKNOWN	$3.19X10^{-3}$	$3.69X10^{-2}$	1.79
CB496791 UNKNOWN	$3.21X10^{-3}$	$3.69X10^{-2}$	0.73
CA051087 UNKNOWN	$3.24X10^{-3}$	$3.69X10^{-2}$	1.77
CB509250 UNKNOWN	$3.28X10^{-3}$	$3.69X10^{-2}$	2.36
CA058444 UNKNOWN	$3.30X10^{-3}$	$3.69X10^{-2}$	2.40
CB501863 UNKNOWN	$3.34X10^{-3}$	$3.69X10^{-2}$	1.50
CA041124 UNKNOWN	$3.35X10^{-3}$	$3.69X10^{-2}$	2.54
CB492842 UNKNOWN	$3.40X10^{-3}$	$3.69X10^{-2}$	0.66
CB509472 UNKNOWN	$3.40X10^{-3}$	$3.69X10^{-2}$	0.60
CK991319 UNKNOWN	$3.45X10^{-3}$	$3.69X10^{-2}$	1.38
CB515801 UNKNOWN	$3.51X10^{-3}$	$3.71X10^{-2}$	0.51
CA051186 UNKNOWN	$3.54X10^{-3}$	$3.71X10^{-2}$	1.75
CB505679 UNKNOWN	$3.62X10^{-3}$	$3.71X10^{-2}$	1.83
CB499820 UNKNOWN	$3.64X10^{-3}$	$3.71X10^{-2}$	2.09
CA059706 UNKNOWN	$3.66X10^{-3}$	$3.71X10^{-2}$	1.91
CK991267 UNKNOWN	$3.73X10^{-3}$	$3.72X10^{-2}$	1.92
CB494335 UNKNOWN	$3.80X10^{-3}$	$3.74X10^{-2}$	0.39
CA038351 UNKNOWN	$3.88X10^{-3}$	$3.77X10^{-2}$	2.21
CB494761 UNKNOWN	$3.88X10^{-3}$	$3.77X10^{-2}$	2.02
CA045102 UNKNOWN	$3.97X10^{-3}$	$3.78X10^{-2}$	1.92
CB497819 UNKNOWN	$4.01X10^{-3}$	$3.78X10^{-2}$	1.44
CK990518 UNKNOWN	$4.04X10^{-3}$	$3.79X10^{-2}$	1.66
CA043437 UNKNOWN	$4.09X10^{-3}$	$3.79X10^{-2}$	0.73
CB497624 UNKNOWN	$4.11X10^{-3}$	$3.79X10^{-2}$	1.70
CA041641 UNKNOWN	$4.13X10^{-3}$	$3.80X10^{-2}$	1.76
CA064261 UNKNOWN	$4.19X10^{-3}$	$3.81X10^{-2}$	1.65
CB507001 UNKNOWN	$4.20X10^{-3}$	$3.81X10^{-2}$	1.77
CB501113 UNKNOWN	$4.24X10^{-3}$	$3.82X10^{-2}$	0.52
CB509558 UNKNOWN	$4.28X10^{-3}$	$3.83X10^{-2}$	1.67
CB505946 UNKNOWN	$4.37X10^{-3}$	$3.86X10^{-2}$	0.57
CK990472 UNKNOWN	$4.41X10^{-3}$	$3.86X10^{-2}$	1.88
CA057111 UNKNOWN	$4.41X10^{-3}$	$3.86X10^{-2}$	1.55
CA058394 UNKNOWN	$4.43X10^{-3}$	$3.86X10^{-2}$	1.52
CB496980 UNKNOWN	$4.46X10^{-3}$	$3.86X10^{-2}$	2.83
CB514107 UNKNOWN	$4.49X10^{-3}$	$3.87X10^{-2}$	0.76
CK990406 UNKNOWN	$4.51X10^{-3}$	$3.87X10^{-2}$	0.67
CA058530 UNKNOWN	$4.57X10^{-3}$	$3.90X10^{-2}$	1.66
CA038646 UNKNOWN	$4.71X10^{-3}$	$3.92X10^{-2}$	0.46
CA039186 UNKNOWN	$4.73X10^{-3}$	$3.92X10^{-2}$	0.48
CB508776 UNKNOWN	$4.75X10^{-3}$	$3.92X10^{-2}$	0.89
CB494013 UNKNOWN	$4.77X10^{-3}$	$3.92X10^{-2}$	0.64
CB492837 UNKNOWN	$4.81X10^{-3}$	$3.92X10^{-2}$	1.84
CK990916 UNKNOWN	$4.81X10^{-3}$	$3.92X10^{-2}$	0.50
CB492165 UNKNOWN	$4.86X10^{-3}$	$3.92X10^{-2}$	1.71
CA052906 UNKNOWN	$4.86X10^{-3}$	$3.92X10^{-2}$	0.68
CK990777 UNKNOWN	$4.99X10^{-3}$	$3.92X10^{-2}$	1.45