Bases génomiques de la divergence adaptative et de la mortalité en mer chez le saumon atlantique (*Salmo salar*)

Thèse

Vincent Bourret

Doctorat en Biologie
Philosophiae Doctor (Ph.D.)

Québec, Canada

RÉSUMÉ

L'importance historique du saumon atlantique et son exploitation contemporaine en font une espèce prioritaire en conservation. Cette thèse propose l'atteinte de plusieurs objectifs liés aux différents enjeux de gestion et de conservation touchant l'espèce. De plus, en privilégiant une approche génomique, la mise en évidence des bases génétiques de la divergence adaptative était au cœur de la présente thèse. D'abord, nous avons cherché à évaluer les changements temporels dans la composition génétique d'une population sauvage de saumon atlantique suivant l'introgression de saumon d'élevage. Bien que les résultats n'aient pas montré de changement temporel en termes de richesse allélique ou de diversité génétique, nous avons démontré que cette introgression se traduit par une altération de l'intégrité génétique de la population indigène, incluant une perte possible d'adaptation. Ensuite, nous avons participé au développement et à l'essai d'une biopuce à SNP en réalisant l'étude de génétique des populations la plus détaillée jamais réalisée sur le saumon atlantique. Nos résultats ont révélé trois groupes génétiques régionaux en Europe et des zones de contact secondaire entre ces groupes. Ces zones seraient potentiellement associées à des barrières exogènes et endogènes, ce qui rend l'interprétation équivoque quant à l'influence de l'environnement sur la divergence adaptative. Dans ce contexte, l'objectif suivant de la thèse était d'améliorer notre compréhension des liens entre l'environnement et la divergence génétique des populations. Nos résultats amènent de nouvelles perspectives sur les liens entre la variation environnementale et la divergence génétique neutre et adaptative. Spécifiquement, nous avons montré que le climat et la géologie des rivières étaient significativement associés à la divergence potentiellement adaptative et neutre des populations. Finalement, nous avons cherché à explorer les déterminismes génomiques de la mortalité en mer des saumons atlantiques. Par une méthode novatrice multilocus, nous avons observé un patron de mortalité sélective en mer temporellement répété. Ces résultats supportent l'hypothèse voulant que la sélection cause principalement de petits changements de fréquences alléliques à plusieurs loci covariants plutôt qu'un petit nombre de changements à effet majeur. En somme, cette thèse contribue significativement à l'avancement des connaissances dans plusieurs contextes cruciaux liés à la gestion et la conservation de l'espèce.

ABSTRACT

The historical significance of Atlantic salmon and its contemporary exploitation have made this species a central focus in conservation biology. This thesis addresses a number of questions linked to important challenges for this species' conservation and management. Moreover, by emphasizing a genomic approach, we aimed to systematically disentangle neutral and adaptive genetic divergence. First, we documented temporal changes in the genetic make up of a wild Atlantic salmon population following introgression from farmed escapees. Although our results did not show any significant temporal changes in allelic richness and gene diversity, introgression has resulted in significant alterations of the genetic integrity of the native population, including a possible loss of adaptation to wild conditions. Then, we participated in the development and testing of a SNP-array and conducted the most extensive population genetic study on Atlantic salmon to date. We found three major regional genetic groups in Europe and secondary contact zones between those groups. These zones were associated with putative endogenous and exogenous barriers, rendering the interpretation of environmental influence on potentially adaptive divergence equivocal. In this context, the next objective was to improve our understanding of links between the environment and genetic divergence of Atlantic salmon populations. Our results provide valuable insight into the links between environmental variation and both neutral and potentially adaptive genetic divergence. In particular, we have shown that climate and geological characteristics were significantly associated with both potentially adaptive and neutral genetic divergence. Finally, we explored the genomic bases for sea mortality of Atlantic salmon. Using a novel multilocus approach, we observed a pattern of genetically-based selective mortality at sea, which was repeated over time. These results support the hypothesis that selection mainly causes small changes in allele frequencies among many co-varying loci rather than a small number of changes in loci with large effects. Overall, this thesis has significantly improved our knowledge of many critical aspects of Atlantic salmon population genetics, which are tightly linked to conservation and management.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

« *We know it well*
that none of us acting alone
can achieve success »
Nelson Mandela

REMERCIEMENTS

Une thèse de doctorat ne peut être signée que de la main d'une seule personne. Néanmoins, je suis tout à fait conscient que cette opportunité m'est offerte aujourd'hui grâce au soutien, aux encouragements et à la collaboration de gens précieux que j'ai côtoyés au cours des cinq dernières années. Bien sûr, un doctorat représente une formation académique de haut niveau, mais sa réalisation dépend principalement de la persévérance et de la détermination, soit l'expression d'une volonté nourrie par ma passion et celle de mon entourage. Si mon cheminement s'achève maintenant avec un profond sentiment de fierté et un brin de nostalgie, c'est que j'ai bénéficié d'un entourage exceptionnel pendant mon parcours. Par conséquent, je tiens à vous témoigner ici toute ma gratitude.

En premier lieu, je tiens d'abord à remercier mon directeur de thèse de m'avoir confié ce magnifique projet. Louis, merci de m'avoir fait confiance pour mener à bien les nombreux et ambitieux objectifs que comportait cette thèse. Ta passion pour la biologie évolutive est contagieuse et ton flair exceptionnel te permet toujours de fournir l'étincelle qui ravive la flamme affaiblie d'un étudiant. Il y a maintenant plus de dix ans que tu m'as donné une chance unique de m'épanouir pleinement en biologie. Dix ans au cours desquels j'ai pu profiter de ressources académiques et matérielles que peu de labos peuvent offrir à leurs étudiants, tout en bénéficiant d'un traitement humain malgré tes hautes exigences. Je retiendrai de toi ta grande ouverture, ta disponibilité et le fameux petit plus qui fait la différence. Aussi, je suis reconnaissant pour toutes les rencontres marquantes que tu m'as permis de faire dans la convivialité du laboratoire et au travers des nombreuses opportunités de stages et de congrès dont tu m'as permis de bénéficier. Si j'exerce aujourd'hui une profession que j'adore, c'est à toi que je le dois. Pour ça et tout le reste que je ne peux exprimer, merci !

À ma co-directrice, Mélanie, merci pour ton apport et surtout ton support dans la réalisation de ce projet. Une présence amicale et chaleureuse que j'ai particulièrement appréciée dans les moments difficiles. Je suis bien conscient que cette thèse s'appuie d'abord sur une collecte d'échantillons et de données que tu as mené de front et je suis très reconnaissant de la confiance que tu as fait preuve en me passant le flambeau. Merci également pour ton

professionnalisme, ta rigueur scientifique et ta passion pour ton ancien sujet de thèse. De collègue de bureau à co-directrice et collègue à nouveau, je suis heureux d'avoir pu profiter de ton enthousiasme et ton énergie.

Je tiens également à remercier les membres de mon comité d'encadrement, Nadia Aubin-Horth et Julian Dodson, qui ont également agi à titre de membre de mon comité d'examen doctoral et jury pour ma présentation de projet. Merci pour vos judicieux conseils et votre ouverture d'esprit. Dans un contexte similaire, j'aimerais remercier Julie Turgeon qui m'a offert un mentorat inattendu en pédagogie l'espace d'une session. Cela a été une expérience très formatrice et j'en suis très reconnaissant. J'ai également beaucoup apprécie te côtoyer au sein du comité de programme. Ces deux occasions m'ont permis de découvrir et de m'inspirer de ton dévouement pour la biologie, mais surtout pour les étudiants actuels et futurs.

Ce projet s'inscrit dans le cadre d'une collaboration provinciale, nationale et internationale. Je tiens donc à remercier les intervenants à ces différents niveaux pour leur apport substantiel à la réalisation de cette thèse. Merci donc à Benoît Landry et Vanessa Cauchon pour leur aide technique au Ministère du Développement Durable, de l'Environnement, de la Faune et des Parcs du Québec. Merci également à Patrick O'Reilly du Ministère des Pêches et Océans du Canada pour sa précieuse et généreuse collaboration au cours des cinq dernières années. Finalement, un merci à Sigbjorn Lien et Matthew Kent, de même qu'à toute l'équipe de CIGENE en Norvège pour leurs réalisations techniques d'abord, mais aussi pour m'avoir fait confiance dans la réalisation du chapitre 3 de la présente thèse.

Le laboratoire Bernatchez, comme une entité singulière, est très important dans la réalisation des projets de tous les étudiants qui y passent. Comme dans une fourmilière, chacun y joue un rôle déterminant. Ainsi, j'aimerais d'abord remercier Guillaume Côté pour son support technique et moral inconditionnel. Guillaume, tu as à cœur ce labo comme personne et ton dévouement pour tous les projets et étudiants est sans pareil. À la personne la plus généreuse au monde, merci Geneviève pour ta bonne humeur perpétuelle et ta générosité. Autant comme collègue étudiante que comme professionnelle au labo, merci

pour ton écoute, tes encouragements et les nombreux services de correction et d'aide en laboratoire, ce n'aurait pas été aussi souriant et agréable sans toi. Merci aussi à Éric Normandeau pour son aide bioinformatique et Lucie Papillon *aka* la mémoire du laboratoire pour ton aide et support ponctuel très apprécié. Merci aussi à Sébastien Bélanger-Deschênes pour son aide en laboratoire à ma première session de doctorant. J'ai également le cœur gros en pensant au plaisir et au privilège que j'ai eu de travailler, côtoyer et bénéficier de l'aide de Marie-Hèlène Perreault, une très grande petite personne. Un gros merci également à Melissa Evans et Anne Dalziel qui ont su être critiques et correctrices sans pitié pour mes manuscrits et mon anglais imparfait, mais toujours en me laissant croire que finalement ce n'était pas si pire que ça, yeah right ! J'ai une pensée pour mes collègues de bureau Antoine, Éliane et Simon avec qui ce fût un plaisir de discuter quotidiennement, mais une pensée toute spéciale pour Julien April avec qui j'ai partagé bien plus qu'un bureau. Ce fût un privilège pour moi que de te voir cheminer au travers de ton doctorat, tu es un biologiste d'exception, merci pour le soutien moral, les discussions enrichissantes et tous les bons moments vécus à la pêche, à la boxe et à la bière. Bière que je lève maintenant à tous mes collègues du labo et toutes les autres belles rencontres que j'ai fait dans ce labo au fil des cinq dernières années, sans vous nommez par peur d'en oublier (et par manque d'espace), vous savez qui vous êtes et vous savez que je vous aime. Santé !

Je veux aussi remercier mes nouveaux collègues de la protection de la faune pour leur soutien et leur compréhension des derniers mois. Micheline, Francine, Catherine et Vicky, vous êtes une équipe de travail formidable et des personnes aussi passionnées que passionnantes.

Évidemment, certaines rencontres au laboratoire ont été plus marquantes de que d'autres. Ce fût le cas pour une petite escouade de postdocs français qui m'ont tendu maintes embuscades à La Barberie et ailleurs. Fabien (et Carine), Christopher et Pierre-Alexandre. Les sorties au Mont-Albert, la pêche, dégustations de vins, de bières et de terroirs ou tout simplement les discussions franches que nous avons eues ensemble resteront autant de souvenirs solidement ancrés dans ma mémoire. Arne, Kristian, Kim et Melissa font aussi

parti des amis qui vivent loin maintenant mais qui garderont une place de choix dans mon cœur. Notre amitié dépasse maintenant les frontières.

Je tiens à remercier particulièrement chaleureusement Pierre-Alexandre Gagnaire pour son instinct scientifique hors du commun. Pierre, chaque discussion avec toi m'a inspiré une nouvelle idée ou un nouveau point de vue face à mes analyses ou interprétations. Peu de gens ont une passion aussi vive combinée à un talent inné et profond pour la science. Tu fais partie des rares qui y voient clair même quand le brouillard est épais et la beauté de tout ça, c'est que tu n'hésites jamais à guider les autres pour les aider à y voir aussi clair que toi. Comme si c'était possible… Autant comme ami que comme scientifique tu demeureras une grande inspiration. Au plaisir !

À mes amis de toujours. Merci de m'avoir permis de décrocher et de me rappeler qu'il y a plus et surtout plus important que le travail dans la vie. Pete, Val, Oli, Émilie, Pat, Guillaume et Marika, je sais que je peux toujours compter sur vous et je me considère privilégié de pouvoir partager vos petits et grands bonheurs.

À ma famille et ma belle-famille, autant que pour mes amis sinon plus, votre soutien et l'équilibre que vous m'avez offert auront été essentiels. Yan, Noa, Dan et Suzie, cela ne m'inquiète pas, mais surtout ne changez pas ! Des gens aux valeurs familiales aussi solides que les vôtres sont une inspiration. Mon frère Jean-Simon, Alice et mes filleuls adorés Henri et Émile, ce que vous me permettez de vivre est unique. Votre famille est littéralement l'extension de la mienne et j'ai chéri de nombreux moments de bonheur à vos côtés au cours des cinq dernières années. Tous ces moments m'ont aidé à avancer et à atteindre mes objectifs, je vous adore. Sylvie, merci aussi pour ton soutien inconditionnel et tous les bons moments que nous partageons ensemble en famille. Papa et Maman, évidemment, je ne serais pas là sans vous. S'il est vrai qu'un doctorat est une épreuve de détermination et de persévérance, il est aussi vrai que je vous dois la force et l'inspiration à ne jamais lâcher et surtout à toujours donner le meilleur de moi-même et un peu plus. C'est également votre support et vos encouragements qui m'ont permis de me rendre à ce point. Soyez certains que j'apprécie chaque moment, petits et grands, à vos côtés. Merci !

Finalement, les mots me manquent, mais qu'importe puisque les mots justes n'existent pas pour signifier toute l'étendue de ma gratitude envers celle que je considère comme la pierre angulaire de ma vie. Sonia, une grande part du mérite de ce doctorat te revient sans conteste. Au milieu de ce doctorat, nous y avons ajouté un élément de défi en accueillant Alexia. Depuis, mais déjà auparavant, ce sont ta force de caractère, ta générosité et ton amour qui nous ont permis de garder un équilibre de vie sain. Évidemment, Alexia et toi avez été aux premières loges pour vivre les hauts et les bas de ma vie de doctorant. Dans les bons comme dans les moins bons moments, vous m'avez toujours encouragé et insufflé la persévérance nécessaire et je vous en serai toujours reconnaissant. Je n'ai pas besoin d'un doctorat pour être comblé ou me sentir important, car je vous ai et vous me comblez au quotidien. Merci infiniment ! Je vous adore…

AVANT-PROPOS

Cette thèse est organisée en six chapitres incluant des chapitres pour l'introduction et la conclusion générale (chapitres 1 et 6). Les quatre autres chapitres sont sous formes d'articles scientifiques dont trois sont publiés à ce jour et le quatrième soumis.

Le chapitre 2 est publié sous la référence : Bourret V, O'Reilly PT, Carr JW, Berg PR & Bernatchez L (2011) Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity*, 106: 500–510. Cette publication apparaît dans un numéro spécial du journal *Heredity* intitulé : *Genetics of local adaptation in salmonid fishes*.

Le chapitre 3 est publié sous la référence : Bourret V, Kent MP, Primmer CR, Vasemägi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L & Lien S (2013) SNP-array reveals genome wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, 22: 532-551. Cette publication apparaît dans un numéro spécial du journal *Molecular Ecology* intitulé : *Evolutionary Ecological Genomics*.

Le chapitre 4 est publié sous la référence : Bourret V, Dionne M, Kent MP, Lien S & Bernatchez L (2013) Landscape Genomics in Atlantic Salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution*, 67: 3469-3487. Cette publication apparaît dans un numéro spécial du journal *Evolution* intitulé : *Evolutionary Landscape Genetics*.

Le chapitre 5 est soumis sous la référence : Bourret V, Dionne M & Bernatchez L (2014) Co-varying genotypic changes associated with selective mortality at sea in Atlantic salmon (*Salmo salar*).

L'analyse des résultats et la rédaction de chaque chapitre ont été réalisées par Vincent Bourret sous la supervision du directeur de la thèse, Louis Bernatchez. La co-directrice de la thèse, Mélanie Dionne, a contribué au raffinement des chapitres 4 et 5 par ses commentaires et suggestions lors des analyses et de la rédaction. Le responsable du laboratoire au *Centre for Integrative Genetics* (CIGENE), en Norvège, Matthew P. Kent, a réalisé le génotypage à l'aide de la biopuce à SNP pour les chapitres 3, 4 et 5.

De plus, en raison de son implication dans le développement de la biopuce chez CIGENE et de la rédaction des éléments relatifs au développement de cette dernière dans les sections « Méthodes » et « Résultats » du chapitre 3, Matthew P. Kent partage la qualité d'auteur principal pour ce dernier. Pour ce même chapitre, il est à noter que Sigbjorn Lien et Louis Bernatchez se partagent la qualité d'auteur sénior pour leur supervision de Matthew P. Kent et Vincent Bourret et leur leadership dans la réalisation de la publication. Également pour ce chapitre 3, notez que Craig R. Primmer a réalisé l'arbre génétique et Anti Vasemägi a réalisé les analyses d'identification des régions génomiques sous sélection. Tous les autres auteurs de ce chapitre ainsi que les susmentionnés ont commenté la version finale du manuscrit avant publication.

Pour le chapitre 2, Patrick T. O'Reilly a fourni les données brutes pour les marqueurs microsatellites en plus de participer activement au raffinement du manuscrit. Jonathan W. Carr et Patrick T. O'Reilly ont fourni de l'information essentielle quant aux méthodes d'échantillonnage et au contexte expérimental. Finalement, pour ce chapitre 2, Paul R. Berg était responsable du génotypage des 388 SNPs chez CIGENE.

Chapitre 1 Introduction Générale

1.1 Problématique

Depuis des siècles, le saumon atlantique (*Salmo salar*) est une espèce à l'importance culturelle et économique incontestée de part et d'autre de l'Atlantique Nord, tant pour son exploitation commerciale passée, sportive et industrielle actuelle. Malheureusement, on assiste depuis quelques décennies à un déclin marqué de l'espèce sur l'ensemble de sa répartition (ICES 2013). Ainsi, son importance historique et son exploitation contemporaine en font une espèce de premier plan en biologie évolutive et de la conservation. En effet, son comportement philopatrique et sa vaste répartition dans un environnement hétérogène offrent un fort potentiel pour l'établissement d'adaptations locales (Taylor 1991; Vasemägi et al. 2005). De nombreuses études ont démontré des évidences indirectes ou directes d'adaptations locales telles que des traits phénotypiques et d'histoire de vie corrélés à l'environnement (Taylor 1991), des clines génétiques le long de gradients environnementaux (Verspoor et al. 2005), la performance accrue de saumons sauvages en milieu naturel par rapport aux saumons d'élevage (McGinnity et al. 1997) et l'héritabilité de la résistance aux pathogènes (Landry & Bernatchez 2001). Néanmoins, les démonstrations d'adaptations locales chez ce poisson demeurent un sujet débattu en raison de leur nature indirecte, d'autant plus que leurs mise en évidence amèneraient des outils considérables pour la définition d'unités de conservation ou de gestion (van Tienderen et al. 2002; Conover et al. 2006; Primmer 2009). Démontrer l'existence de variations génétiques adaptatives et les bases génétiques de la divergence adaptative chez le saumon atlantique constitue donc toujours un défi de taille pour les généticiens des populations.

C'est justement dans cette optique que ce projet propose l'atteinte de plusieurs objectifs liés aux différents enjeux de gestion et de conservation touchant le saumon atlantique. Ainsi, en privilégiant une approche génomique permettant la caractérisation de la divergence génétique neutre et adaptative, il est question d'aborder les thèmes de : (1) l'impact de la production aquacole de l'espèce sur les populations sauvages, (2) la mise en évidence de marqueurs génétiques potentiellement sous l'effet de la sélection naturelle, (3) les associations gène-environnement qui soutiennent la divergence adaptative des populations naturelles et (4) la mortalité sélective lors de la première année de migration en mer. Ultimement, en utilisant le saumon atlantique et le contexte particulier qui caractérise ses

populations, nous espérons contribuer significativement à l'amélioration des connaissances des champs scientifiques fondamentaux qui seront brièvement revus dans le cadre conceptuel suivant.

1.2 Génétique des populations

Pour comprendre l'influence des forces évolutives qui structurent la génétique des populations, il faut d'abord décrire et quantifier la variation génétique dans une population, puis les patrons de variation entre les populations (Hedrick 2005). Avant l'émergence des nouvelles technologies de séquences et l'abondance de marqueurs de polymorphisme mononucléotidique (SNP ; « *Single Nucleotide Polymorphism* »), l'approche classique des 25 dernières années en génétique des populations mesurait principalement l'influence des forces évolutives neutres (flux génique, mutation et dérive génétique) sur la diversité génétique des populations en évitant les biais causés par la sélection naturelle (Nielsen et al. 2009). Par cette approche, il est possible de mesurer la part de responsabilité des forces antagonistes de la dérive génétique qui tend à favoriser la divergence entre populations via la variation aléatoire de la composition génétique entre populations isolées géographiquement *versus* le flux génique qui tend plutôt à homogénéiser la diversité génétique présente parmi les populations. Cependant, puisque c'est la variation génétique adaptative qui affecte et soutient la valeur adaptative des populations (Conover et al. 2006), sa caractérisation est aussi nécessaire et on assiste effectivement à un renouvellement de l'intérêt envers la détection de la sélection dans les populations naturelles (Nielsen et al. 2009).

La sélection divergente, en l'absence d'autres forces évolutives, entraîne une évolution indépendante de certains traits dans des populations isolées. Ces traits adaptatifs, qui confèrent aux individus locaux une valeur adaptative relative supérieure à celle d'individus immigrants, sont appelés « adaptations locales » (Kawecki & Ebert 2004). Bien que plusieurs méthodes pour détecter l'effet de la sélection au niveau des populations existent, celles qui permettent sa mise en évidence directe sont rares et difficilement applicables en milieu naturel (Endler 1986). Cet état de fait est caractéristique des espèces comme le saumon atlantique où les transplantations croisées, les expériences en jardins communs de

tous les stades de vie et ainsi les mesures de l'aptitude viagère sont pratiquement impossibles. Pour ces espèces, il est donc pertinent de s'intéresser aux différentes méthodes de détection de la divergence adaptative afin d'optimiser la rigueur des observations indirectes qu'il est possible de faire par une approche de génomique des populations, soit la caractérisation d'un grand nombre de marqueurs génétiques dans les populations naturelles.

1.3 Détection de la sélection

En effet, en ce qui concerne la détection de la sélection dans les populations naturelles, l'étude de la génétique des populations permet des approches indirectes certes, mais avec l'avantage important d'être non invasives. L'une d'entre elles est l'identification de loci potentiellement soumis à la sélection d'après la méthode mise de l'avant par Lewontin et Krakauer (1973), une approche appelée balayage génomique. Contrairement au contexte démographique et aux forces évolutives neutres qui influencent la composition génétique de façon similaire sur tout le génome, la sélection, quant à elle, agit localement sur un plus petit nombre de loci en leur conférant une composition génétique contrastante d'une population à l'autre (Bonin et al. 2006). L'approche par balayage génomique, où la diversité génétique d'un très grand nombre de marqueurs distribués sur le génome est examinée, permet justement de distinguer les marqueurs typiquement neutres des marqueurs présentant une différenciation génétique atypique entre les populations et donc potentiellement soumis à l'effet de la sélection (Lewontin & Krakauer 1973). Dans certains contextes, des marqueurs atypiques ont été associés à un locus de trait quantitatif (QTL) important dans le processus de la sélection divergente (Via & West 2008), mais il reste que l'identification de gènes précis responsables de traits adaptatifs est peu fréquente. Généralement, lorsque c'est le cas, les marqueurs au comportement atypique sont associés à une région génomique soumise à l'effet de la sélection agissant directement sur un gène ou indirectement par un phénomène aussi appelé l'auto-stop génétique (« *genetic hitchhiking* »).

De nombreuses méthodes de balayage génomique ont été développées pour l'identification de régions génomiques sous l'influence de la sélection (Beaumont & Balding 2004; Joost et al. 2007; Foll & Gaggiotti 2008; Excoffier et al. 2009). La principale difficulté de ces tests

est l'obtention d'une distribution de F_{st} de référence représentative de la divergence neutre (Beaumont 2005). Ainsi, la différence entre ces tests réside principalement dans le modèle démographique derrière la simulation qui permet d'obtenir la distribution neutre de F_{st}. À l'origine des tests les plus robustes se trouve celui présenté par Beaumont & Nichols (1996), FDIST, qui propose d'obtenir la distribution des F_{st} des loci examinés entre les populations en fonction de l'hétérozygotie par simulation d'un modèle infini en îles et d'identifier comme atypiques les loci présents dans les queues de distribution. Les méthodes de Beaumont & Balding (2004), Foll & Gaggiotti (2008), Excoffier et al. (2009) représentent des versions améliorées de FDIST. BAYESFST de Beaumont & Balding (2004) et BAYESCAN de Foll & Gaggiotti (2008) utilisent une approche Bayésienne et leur force est surtout dans la séparation des effets de chaque locus, de chaque population et des interactions locus-population; BAYESCAN se distingue de BAYESFST par un test postérieur sur la probabilité d'un locus d'être sous l'effet de la sélection. La méthode d'Excoffier et al. (2009), quant à elle, utilise plutôt un modèle hiérarchique en îles qui permet différents taux de migration entre des groupes régionaux définis *a priori* par rapport au taux à l'intérieur de ceux-ci, ce qui aurait pour effet de réduire la détection de faux positifs. Finalement, la méthode SAM (« *spatial analysis method* ») de Joost et al. (2007) utilise une approche totalement différente qui tente de vérifier s'il existe des associations entre les fréquences d'allèles et les variables environnementales, écologiques ou physiques pour des marqueurs. Pour cette méthode, il est nécessaire de construire des matrices de conditions environnementales à partir de systèmes d'informations géographiques (SIG).

Il existe de nombreux cas où les balayages génomiques ont été utilisés avec succès dans le but d'identifier des gènes associés à des traits phénotypiques sous sélection environnementale (Storz & Wheat 2010; Gagnaire et al. 2012b), pour identifier des régions génomiques de divergence adaptative (Rogers & Bernatchez 2007; Flori et al. 2009; Hohenlohe et al. 2010; Gagnaire et al. 2012a), ou mettre en évidence des marqueurs soumis à l'introgression (Hohenlohe et al. 2011; Karlsson et al. 2011; Lamaze et al. 2012). Cependant, malgré ces démonstrations, les méthodes de balayage génomique demeurent critiquées pour plusieurs raisons, notamment en raison des modèles démographiques sous-jacents qui sont relativement loin de la réalité des systèmes étudiés et des taux élevés de

faux positifs rapportés lors de tests simulés (Bierne et al. 2011; Narum et al. 2011; Vilas et al. 2012; De Mita et al. 2013). Une première façon d'éviter certains biais de la méthode consiste en une approche de fenêtre coulissante (« *sliding-window* ») sur le génome où la différenciation moyenne est estimée pour les marqueurs se retrouvant sur une distance (fenêtre) donnée le long du génome et évaluée à savoir si elle est significativement plus élevée que sur l'ensemble du génome (Flori et al. 2009; Gagnaire et al. 2013). Cette façon de faire permet de considérer des blocs de marqueurs contigus et donc de réduire la variance des observations individuelles, mais elle demande une densité de marqueurs importante et au minimum une cartographie génétique fine pour être efficace. De plus, cette approche ne nous affranchit pas du possible effet de la variation de taux de recombinaison dans le génome qui peut entrainer une variation de divergence apparente à l'échelle du génome entre populations.

Par ailleurs, les cibles potentielles de ces balayages génomiques basés sur la détection de loci ou régions génomiques au comportement atypique se limitent aux loci à effet majeur alors qu'en principe, puisque la sélection devrait générer du déséquilibre de liaison entre plusieurs loci pour des traits quantitatifs complexes, on s'attend à ce que des changements de fréquences d'allèle subtils entre loci covariant résultent en un effet combiné sur un phénotype plus grand qu'un simple locus (Latta 1998; Le Corre & Kremer 2012). Ce dernier modèle se fonde sur un principe d'adaptation polygénique (« *polygenic adaptation* »), ou plusieurs gènes seraient impliqués dans l'expression d'un phénotype, par opposition au balayage sélectif (« *selective sweep* »), impliquant un seul gène. Certains auteurs suggèrent donc que la recherche de balayage sélectif, modèle selon lequel un allèle avantageux peut rapidement monter en fréquence dans une population, est trop limitant dans la recherche de divergence adaptative en milieu naturel (Pritchard & Di Rienzo 2010). Or, il existe bien des exemples convaincants où des allèles à effet majeur ont été détectés par balayage génomique, notamment pour le gène EDA chez l'épinoche à trois épines (*Gasterosteus aculeatus*) qui a été mis en évidence dans de nombreuses études (Hohenlohe et al. 2010; Deagle et al. 2012; Roesti et al. 2012). En contrepartie, il semble que le modèle d'adaptation polygénique soit mieux adapté dans le cas de traits quantitatifs complexes. L'étude de Yang et al. (2010) soutient cette hypothèse puisqu'elle a montré pour la taille

humaine, un trait quantitatif complexe, qu'une certaine proportion de la variance était attribuable à des marqueurs ayant un faible effet individuel. Bien que les deux paradigmes qui s'opposent présentement dans la littérature ne soient probablement pas mutuellement exclusifs en nature, la mise en évidence de divergence adaptative en milieu naturel, particulièrement pour les espèces non modèles, se limitent souvent à la détection de marqueurs atypiques par balayage génomique. Il serait donc avisé de développer et tester des méthodes de détection mieux adaptées au modèle d'adaptation polygénique.

Alors que les nouvelles technologies en biologie moléculaire ont amené une augmentation importante des études examinant les bases génétiques de l'adaptation et que certaines études en milieu naturel ont permis de mieux comprendre les processus écologiques responsables de la divergence adaptative, rares sont celles ayant aussi permis d'identifier les agents de sélection de même que les cibles potentielles à l'échelle génomique et fonctionnelle. Devant cet état de fait, la mise en évidence de marqueurs sous sélection peut permettre, via l'information sur leurs fonctions et leurs associations aux conditions environnementales, de faire le pont entre la diversité génétique adaptative et les agents de sélection responsables de cette divergence, qu'elle soit historique ou contemporaine.

1.4 Génomique du paysage

L'hétérogénéité de l'environnement est sans contredit un facteur pouvant générer de la divergence adaptative à l'intérieur d'espèces entre populations relativement isolées ou réparties sur un vaste territoire, un phénomène pouvant ultimement mener à la spéciation (Coyne & Orr 2004). En utilisant des informations biologiques et environnementales, la génétique du paysage tente d'expliquer comment des variations biotiques et abiotiques des conditions environnementales du milieu contribuent aux processus évolutifs qui structurent la génétique des populations en milieu naturel (Manel et al. 2003; Holderegger & Wagner 2008). Ce domaine connexe à la génétique des populations est en émergence depuis une dizaine d'années (Manel et al. 2013). Dans la plupart des études réalisées avant l'émergence des marqueurs de type SNP, la génétique du paysage se limitait à faire le lien entre la divergence génétique neutre et les contraintes environnementales au flux génique (ex. : Petren et al. 2005; Leclerc et al. 2008) alors que les études pouvant tenir compte de

divergence adaptative étaient rares (mais voir : Bonin et al. 2006; Gaggiotti et al. 2009; Manel et al. 2010ab). La faible couverture génomique et le caractère neutre et anonyme des marqueurs utilisés limitaient le potentiel pour une interprétation fonctionnelle des associations rapportées.

Par contre, dans le contexte actuel où les méthodes de séquençage à haut débit facilitent la découverte de marqueurs génétiques en nombre de plus en plus impressionnant, la génétique du paysage évolue vers la génomique du paysage. En effet, en vertu de couvertures génomiques beaucoup plus grandes fournies par des marqueurs provenant autant des régions codantes que non codantes du génome, il est désormais envisageable de départir la divergence adaptative de la divergence génétique neutre en plus d'identifier les fonctions et processus biologiques qui sont ciblés par la sélection environnementale. De cette façon, cet amalgame entre la génomique des populations et l'étude du paysage écologique permet l'identification de facteurs environnementaux influençant la divergence génétique adaptative des populations de même que l'échelle géographique à laquelle ils agissent (Gaggiotti et al. 2009). L'intérêt d'une telle approche réside donc principalement dans la possibilité de lier des marqueurs génétiques sous sélection à une ou des variables environnementales précises (Joost et al. 2007; Gaggiotti et al. 2009) et ainsi de faire le pont entre les agents de sélection, la divergence adaptative observée et ultimement un contexte fonctionnel permettant d'expliquer les fonctions biologiques qui sont ciblées par la sélection.

En résumé, la génomique du paysage représente une avenue florissante permettant de différencier l'influence des différentes forces évolutives en plus d'identifier des agents biologiques, écologiques ou environnementaux potentiellement responsables de divergences génétiques. Cependant, ce type d'approche a jusqu'à maintenant été restreint à l'utilisation de marqueurs neutres et à certains gènes candidats (mais voir Gaggiotti et al. 2009) et le développement de cadre analytique efficace pour mettre en évidence les associations gène-environnement représente un des défis les plus important en génomique du paysage. D'autant plus que l'identification d'agent de sélection et l'échelle géographique à laquelle ils agissent constituent des éléments de connaissance clés en

matière de conservation génétique des populations, particulièrement pour une espèce comme le saumon atlantique.

1.5 Conservation génétique des populations

En conservation, l'objectif avoué est de protéger la biodiversité à tous ses niveaux, que ce soit la diversité globale des écosystèmes, la diversité en espèces pour chacun de ces écosystèmes ou encore la diversité génétique pour chacune des espèces (Primack 2006). Cette dernière est tout aussi importante que la première puisqu'elle permet de maintenir le potentiel évolutif d'une espèce, soit sa capacité à survivre et à s'adapter aux changements de son milieu (Frankham et al. 2002). Toutefois, avant de pouvoir s'affairer à la protection de la biodiversité à son niveau le plus fin, il faut d'abord la caractériser. C'est de cette façon que la génétique des populations vient servir la biologie de la conservation puisqu'elle contribue à l'identification de groupes distincts appelés « unités de conservation », une méthode au cœur des plans de conservation modernes. Plusieurs concepts ont été avancés pour permettre aux gestionnaires de définir de telles unités, mais celui ayant reçu le plus d'attention est sans doute le concept d'« unité évolutive significative » (ESU) présenté pour la première fois par Ryder (1986). Une ESU peut être définie comme une population ou un groupe de populations qui nécessite une gestion distincte et/ou qui devrait être considéré comme une priorité de conservation en raison de sa différence (Allendorf & Luikart 2007).

Les ESUs (et leurs variantes, e.g. unités désignables au Canada) ont maintenant un statut légal important dans plusieurs pays, mais le concept a été revisité de nombreuses fois depuis Ryder (Fraser & Bernatchez 2001). D'une part, Moritz (1994) a présenté les critères les plus simples et pratiques dans ce débat en proposant que les ESUs représentent des groupes monophylétiques réciproques sur la base de l'ADN mitochondrial et de différences de fréquences d'allèles pour des marqueurs nucléaires. Bien que pratique, cette définition allait légèrement à l'encontre de celle de Waples (1991) qui avait été le premier à présenter un cadre de travail précis basé sur 2 types d'information pour décrire les ESUs, à savoir (i) l'isolement reproductif à long terme, en quelque sorte similaire aux critères de Moritz, mais également (ii) les caractères adaptatifs uniques. Crandall et al. (2000), quant à eux,

rejoignaient plutôt Waples en mettant l'emphase sur les différences adaptatives tant aux niveaux génétique qu'écologique. Dans cet ordre d'idées, van Tienderen et al. (2002) ont proposé un nouveau concept d'« unités fonctionnelles significatives » (FSU) basé sur les différences de fréquences alléliques pour des gènes aux fonctions écologiques importantes. Les FSUs reflèteraient donc mieux la variation écologique et les différents caractères adaptatifs des populations comparativement aux informations fournies par la variation génétique neutre. En ce sens, Bonin et al. (2007) ont d'ailleurs suggéré l'utilisation d'un nouveau paramètre dans l'évaluation des populations à prioriser pour la conservation, soit l'indice d'adaptation des populations (PAI), un paramètre décrivant le niveau adaptif particulier d'une population par rapport aux autres.

Bien qu'aucun ne fasse l'unanimité, il apparaît évident que les différents concepts d'ESU et celui de FSU se chevauchent et se complètent, mais surtout que l'essentiel au moment de définir des unités de conservation est vraisemblablement d'intégrer le plus d'informations concordantes au niveau phénotypique, environnemental et génétique (Fraser & Bernatchez 2001; Allendorf & Luikart 2007). À cet effet, au niveau génétique, il est primordial de distinguer l'influence des différentes forces évolutives dans la structuration génétique des populations naturelles et d'identifier les variations génétiques adaptatives et fonctionnelles influencées par la sélection naturelle, une démarche malheureusement peu appliquée aux espèces non modèles en milieu naturel jusqu'à maintenant. D'autant plus que selon van Tienderen et al. (2002), les gènes les plus importants à considérer dans un plan de conservation sont ceux sous sélection directionnelle ou divergente puisque ces gènes rendent les populations distinctes génétiquement et leur permettent de s'adapter aux pressions de sélection du milieu. Une caractérisation spatio-temporelle de la variation génétique adaptative s'avère donc essentielle dans une vision moderne de gestion et de conservation des stocks exploités (Allendorf et al. 2008; Carlson & Seamons 2008).

1.6 Enjeux de gestion et de conservation du saumon atlantique

1.6.1 Espèce globalement structurée

Depuis sa divergence des autres salmonidés, l'histoire évolutive du saumon atlantique a subi de nombreux événements de divergence qui se reflètent aujourd'hui dans la structure

génétique hiérarchique qui caractérise ses populations. Au premier niveau, la divergence entre les populations européennes et nord-américaines remonterait approximativement à 600 000 ans (King et al. 2007). Au second niveau, autant en Europe qu'en Amérique du Nord, de nombreuses études ont montré la présence de structuration génétique au niveau régional (Verspoor 2005; Dionne et al. 2008; Tonteri et al. 2009). Par contre, il n'existe aucune étude singulière qui a permis d'examiner la structure génétique régionale en incluant des populations provenant de l'ensemble de l'aire de répartition de l'espèce. Néanmoins, prises dans leur ensemble, les études réalisées sur une échelle assez grande montrent systématiquement une structuration régionale. Dépendamment des populations incluses, des marqueurs génétiques utilisés, des analyses effectuées et des interprétations, le continent européen serait structuré en plus de trois groupes et l'Amérique du Nord en plus de quatre groupes régionaux génétiquement distincts (King 2007). Finalement, les niveaux de structuration les plus fins incluent les rivières et même les tributaires. Les rivières montrent effectivement une différentiation génétique suffisamment importante pour être considérées comme génétiquement distinctes (ex.: Verspoor 2005; Tonteri et al. 2009; Dionne et al. 2008). C'est cette structure qui justifie notamment en partie la gestion indépendante qui est appliquée aux différents cours d'eau de nos jours dans plusieurs régions du monde. Bien que l'on peut la considérer comme une « microstructure », des études ont montré qu'à l'intérieur même d'une rivière, certains tributaires abritent des populations génétiquement distinctes entre elles, poussant la structure génétique à une échelle extrêmement fine (exemples : Vähä et al. 2007; Dionne et al. 2009).

Au-delà de la divergence neutre, de nombreux événements de sélection naturelle et artificielle sont sans doute également responsables d'une part de la divergence génétique existante. En effet, ceux-ci incluent entre autres des changements d'habitat associés à plusieurs événements de glaciation qui ont eu pour effet de restreindre des populations à des habitats d'eau douce ou bien de façon plus contemporaine, des changements anthropogéniques de différentes sources (Verspoor et al. 2007). Malheureusement, ces derniers sont potentiellement à l'origine d'une situation globale très préoccupante pour l'espèce, justifiant notamment de grands efforts de gestion et de conservation. Selon King et al. (2007), plusieurs aspects de la structuration génétique des populations sont

fondamentaux pour la gestion et la conservation, soit le nombre de populations et leur occupation du territoire, l'impact du flux génique entre elles s'il en est un et l'organisation potentielle du saumon atlantique en métapopulation. Dans le contexte actuel, il va de soi d'ajouter que des connaissances supplémentaires au niveau de la divergence adaptative seraient également pertinentes et souhaitables pour une meilleure gestion de l'espèce.

1.6.2 Situation en Amérique du Nord

Bien qu'ayant subi une réduction importante de ses effectifs en Amérique du Nord, le saumon atlantique revêt une grande importance socio-économique pour les régions, en raison des importantes retombées amenées par la pêche sportive, évaluées à 56 millions de revenus annuels pour les économies locales et provinciales (MPO, 2009). L'espèce représente un élément important de l'alimentation de plusieurs peuples autochtones et l'impact de l'importante industrie piscicole dans le déclin de l'espèce est encore aujourd'hui méconnu. Pour toutes ces raisons, le saumon atlantique jouit d'une attention particulière en matière de gestion et de conservation au Canada. Une raison supplémentaire réside dans le fait que la grande diversité de traits, tant physiologiques, comportementaux que morphologiques des différentes populations locales de salmonidés est largement acceptée comme un indice d'adaptation aux environnements locaux (Taylor 1991; Adkison 1995; Garcia de Leaniz et al. 2007) alors que très peu d'information existe quant aux agents de sélection influençant directement cette adaptation.

Incidemment, le saumon atlantique couvre un vaste territoire en Amérique du Nord, allant historiquement des côtes de la Nouvelle-Angleterre à celles du Groenland, mais c'est particulièrement au Canada que se retrouve la majorité des populations de saumon de l'ouest de l'Atlantique Nord. En effet, plusieurs centaines de cours d'eau canadiens, répartis sur les côtes de cinq provinces, abritent une ou plusieurs populations de saumon atlantique. Ces dernières occupent donc des habitats aux caractéristiques hétérogènes et soumises à une grande variabilité de conditions environnementales. Plusieurs groupes ont étudié la génétique des populations de saumon dans l'est du Canada à l'aide de marqueurs microsatellites (O'Reilly et al. 1996; Fontaine et al. 1997; Palstra et al. 2008), mais plus récemment, une étude à plus large échelle a démontré une ségrégation régionale des

populations basée sur la différenciation génétique et la répartition géographique (Dionne et al. 2008). À partir de 51 populations majoritairement du Québec, Dionne et al. (2008) ont identifié 7 groupes régionaux malgré la grande connectivité observée entre ces groupes et relevé que le régime thermique serait corrélé avec la structure génétique observée. Afin de mieux comprendre et prédire la réponse adaptative des populations face à différentes conditions environnementales (Naish & Hard 2008), il serait donc approprié d'examiner l'influence de la grande diversité des conditions environnementales sur les patrons de diversité génétique adaptatifs.

C'est d'ailleurs dans ce contexte que Dionne et al. (2007 & 2008) ont utilisé des marqueurs microsatellites conjointement avec une caractérisation de la diversité génétique des gènes du complexe majeur d'histocomptabilité (CMH) pour respectivement 51 et 34 populations de saumons atlantiques, majoritairement au Québec (46/51 rivières). Cette étude a révélé une structure de population hiérarchique, c'est-à-dire qu'il a été démontré que les populations pouvaient être regroupées selon un appariement régional sur la base de similarité de la composition génétique neutre, mais également que chaque rivière et parfois même certains tributaires d'une même rivière abritaient des populations suffisamment différenciées pour être considérées comme distinctes (Dionne et al. 2008 & 2009a). Par ailleurs, la caractérisation du gène CMH classe *IIβ* a permis de mettre en évidence une variation de la diversité génétique de ce gène associée aux régimes thermique et bactérien, soit un patron de différenciation différent de celui observé pour les marqueurs neutres, soutenant ainsi l'hypothèse d'une sélection balancée maintenue par la pression de sélection des pathogènes (Dionne et al. 2007).

En somme, plusieurs facteurs potentiellement responsables de la divergence adaptative chez le saumon atlantique peuvent être identifiés en examinant la littérature. Premièrement, le régime de température est considéré comme le facteur environnemental le plus susceptible de promouvoir l'adaptation locale chez le saumon sur la base des différences morphologiques, physiologiques, de traits d'histoire de vie et génétiques (Taylor 1991; Jonsson et al. 2001; Garcia de Leaniz et al. 2007; Dionne et al. 2008). Un deuxième facteur est l'incidence de pathogènes (Dalgaard et al. 2003; Dionne et al. 2007 & 2009b), elle-

même fortement influencée par la température. En dernier lieu, les formations géologiques qui caractérisent les bassins versants des rivières ont une forte influence sur les propriétés physico-chimiques des rivières (Stallard & Edmond 1983; Johnson et al. 1997). Étant donné que ces caractéristiques ont précédemment été corrélées avec la structure génétique des populations en France (Perrier et al. 2011), elles pourraient avoir une influence sur la composition génétique de l'Atlantique Ouest, une hypothèse n'ayant pas encore été testée. Étant donné le statut précaire du saumon atlantique et l'abondance d'indices concernant le caractère distinctif de ses populations, il est impératif de se doter d'un cadre de gestion et de conservation fondé sur des connaissances approfondies de sa diversité génétique neutre, mais également adaptative et fonctionnelle.

1.6.3 Mortalité en mer

Le déclin des populations de saumon en Amérique du Nord est une situation indéniable depuis 1970 où le nombre estimé d'adultes retournant en rivière a diminué de plus de 60% (MPO 2009), une situation remarquée également de l'autre côté de l'Atlantique (ICES 2013). Ce constat serait principalement le résultat d'une importante réduction de la survie en mer au cours des trois dernières décennies (Friedland et al. 2003b). La plupart des études réalisées sur cette problématique ont associé à la mortalité en mer une cause environnementale directe ou indirecte telle que les fluctuations de la température des eaux marines (Todd et al. 2008; Hvidsen et al. 2009) et l'augmentation de la pollution (Rees et al. 2005). La croissance des juvéniles, un facteur indirect susceptible de réguler la survie en mer, est également influencé par les conditions environnementales (Friedland et al. 2000; Friedland et al. 2003ab). Le saumon atlantique fait donc face à une menace aux causes indéterminées dont les impacts se font surtout sentir par une diminution marquée du nombre d'adultes retournant en rivière suite à la migration en mer (MPO 2009; Dionne et al. 2013).

L'environnement marin comporte plusieurs défis de taille pour les saumons atlantiques, notamment en termes de prédation ou de dépense énergétique. En fait, la migration en mer serait la phase la plus couteuse du cycle de vie du saumon atlantique où les premiers moments en mer seraient critiques en terme de mortalité (Friedland et al. 2000; Thorstad et

al. 2012). De plus, des taux variables de retour en rivière ont été rapportés pour des populations provenant de différentes régions géographiques (Dionne et al. 2013). Devant cet état de fait et considérant que cette problématique de mortalité accrue en mer soit présente depuis seulement quelques décennies, il est probable que des pressions de sélection récente et variable dans l'espace affectent les populations de saumon. Néanmoins, bien que plusieurs hypothèses crédibles aient été avancées pour expliquer ce phénomène, aucune étude n'a utilisé une approche moléculaire pour documenter les changements de composition génétique pouvant résulter de telles pressions.

1.6.4 Aquaculture

L'industrie aquacole de saumon atlantique est une industrie mondialement florissante depuis quelques décennies. Malheureusement, malgré toutes les précautions de l'industrie, plusieurs menaces planent sur les populations naturelles de saumon en raison notamment des rejets de la production de masse et des échappées d'élevage. D'ailleurs, ces derniers menacent particulièrement l'intégrité génétique des populations naturelles et peuvent donc effriter les caractéristiques propres de ces populations telles que des adaptations locales. Les saumons domestiques sont élevés en milieu contrôlé où les pressions de sélections sont relâchées afin d'optimiser des traits tels que la croissance rapide (Theodorou & Couvet 2004). Ces saumons exhibent donc d'importantes différences par rapport aux individus sauvages à plusieurs traits phénotypiques comme par exemple le taux de croissance, l'agressivité, les comportements de fuite et la tolérance à l'acidité (Jonsson & Jonsson 2006; Fraser et al. 2008). Toutes ces différences s'ajoutent aux changements génétiques qui peuvent survenir sans lien avec l'adaptation à la captivité, entre autre en raison de l'effet fondateur et la dérive génétique (Skaala et al. 2004; Roberge et al. 2006).

Conséquemment, bien que ces différences permettent de supposer une maladaptation des individus domestiques et donc une incapacité à survivre en milieu sauvage, l'impact des échappées d'élevage et leur mélange avec les populations sauvages représente une menace réelle pour ces dernières (Fleming et al. 2000; Theodorou & Couvet 2004). En effet, des études portant sur la variation phénotypique ont montré que les croisements entre saumons domestiques et sauvages peuvent avoir un effet réducteur sur l'aptitude et les adaptations

locales, contribuant ultimement à de possibles extinctions de populations sauvages (McGinnity et al. 2003; Fraser et al. 2008). Malgré les conséquences phénotypiques et génétiques potentielles des échappées d'élevage, peu d'études ont évalué les flux géniques et la possible introgression des échappées d'élevage dans les populations sauvages de saumon atlantique (Skaala et al. 2006 ; Glover et al. 2013). Ces études offrent tout de même le potentiel de départager la part de responsabilité de l'industrie aquacole sur les changements récents qui semblent affecter les populations sauvages et particulièrement leur composition génétique neutre et adaptative.

1.6.5 Développements technologiques

Les récentes avancées technologiques en génomique ont permis de développer de nombreux outils pour les espèces prioritaires en conservation (Primmer 2009). Le saumon atlantique bénéficie également de ces technologies puisqu'un groupe norvégien du centre de génétique intégrative (CIGENE) a développé une biopuce qui permet le génotypage de plus de 5500 marqueurs de type SNP simultanément. Ces marqueurs ont également été utilisés par ce groupe pour construire une carte génétique de très haute résolution pour le saumon atlantique à partir de familles de 3000 rejetons et de certains traits quantitatifs (QTLs) (Brenna-Hansen et al. 2012). De plus, un consortium international a entrepris le séquençage du génome du saumon atlantique (*International Collaboration to Sequence the Atlantic Salmon Genome*), lequel sera complété en 2014. Ces nouvelles technologies offrent donc d'importantes ressources génomiques qui, conjointement avec le contexte écologique et environnemental du saumon atlantique, le place dans une position exceptionnelle pour l'étude de la base génétique de l'adaptation locale. Le potentiel d'identification de gènes fonctionnels influencés par les conditions climatiques et les perturbations de l'environnement est donc énorme (Kohn et al. 2006; Naish & Hard 2008). Les nouvelles ressources génomiques existantes aujourd'hui représentent donc autant d'outils pertinents qui devraient être mis à profit dans l'identification de divergence adaptative au cœur des différents enjeux de gestion et de conservation qui concernent le saumon atlantique.

1.7 Objectifs de la thèse

Cette thèse s'inscrit dans le cadre d'une subvention stratégique du Conseil de Recherches en Sciences Naturelles et en Génie du Canada ayant pour titre : « Pour une gestion durable du saumon atlantique: identification d'unités de gestion et impacts de l'exploitation par la pêche ». Les objectifs généraux de la thèse s'articulaient donc autour de grands enjeux de gestion et de conservation du saumon atlantique. De plus, ces objectifs se sont tous réalisés dans un contexte méthodologique offrant de nouvelles ressources génomiques, principalement une biopuce à plus de 5500 marqueurs génétiques de type SNP (polymorphisme mononucléotidique). Devant un tel avènement de nouvelles ressources génomiques, la mise en évidence des bases génétiques adaptatives était invariablement au cœur de tous les objectifs de la présente thèse.

Le deuxième chapitre a pour objectif principal de documenter les changements temporels dans la composition génétique de la population sauvage de saumon atlantique de la rivière Magaguadavic suivant la croissance rapide du nombre d'échappées d'élevage entrant dans la rivière. À l'aide de marqueurs microsatellites et SNPs, nous avons donc caractérisé les impacts des croisements entre les individus domestiques et sauvages en comparant des groupes d'individus domestiques et sauvages capturés dans la rivière au cours d'une série temporelle s'étalant sur plus de 25 ans. Étant donné que nous nous attendions à une diversité génétique plus faible dans le groupe d'individus domestiques, nous avions comme hypothèse de départ que l'introgression des échappées d'élevage dans la population sauvage serait accompagnée par une diminution de la diversité génétique de cette dernière. De plus, compte tenu du relâchement de la sélection naturelle et des conditions de sélection artificielle caractérisant l'élevage industriel, nous nous attendions à observer de fortes divergences génétiques à certains marqueurs potentiellement révélatrices de divergence adaptative entre les deux groupes d'individus. Finalement, une réduction de la divergence génétique neutre et adaptative aurait été attendue au sein de nos réplicas temporels en fonction de l'introgression de la population d'élevage.

Le troisième chapitre concerne principalement la mise à l'épreuve de la biopuce SNP. De ce fait, ce chapitre expose d'abord les étapes de développement de la biopuce, un processus

réalisé par l'équipe du Dr Sigbjorn Lien du *Centre for Integrative Genetics* (CIGENE) de l'université norvégienne des sciences de la vie à Às, Norvège. Dans le processus de contrôle de la qualité, nous avons réalisé l'étude de génétique des populations la plus détaillée jamais réalisée sur le saumon atlantique en génotypant 6176 SNPs chez 1430 individus provenant de 38 populations anadromes et résidentes (eau douce) couvrant l'ensemble de la distribution géographique naturelle de l'espèce. Les objectifs spécifiques de ce chapitre étaient donc d'une part de décrire la structure génétique des populations exposée par cette nouvelle ressource génomique et, d'autre part, de mettre en évidence des marqueurs potentiellement soumis à la sélection divergente. Ce dernier objectif a été réalisé en comparant des populations inter- et intra- groupes régionaux pour les populations anadromes et résidentes. Finalement, nous visions également à faire le lien entre les marqueurs potentiellement soumis à la sélection et leur association avec la variation environnementale et des fonctions ou processus biologiques.

L'objectif principal du quatrième chapitre est d'améliorer notre compréhension des liens entre l'environnement et la divergence génétique des populations de même que l'échelle géographique de l'adaptation locale. En se basant sur l'étude de Dionne et al. (2008) et les échantillons prélevés dans le cadre de cette étude, nous avons bonifié la couverture génomique et la caractérisation environnementale. Par conséquent, nous avons d'abord évalué la cohérence de la structure génétique observée avec les marqueurs SNPs avec la structure observée par Dionne et al. (2008). En marge de ce sous-objectif, il a été possible d'expérimenter une méthode de génotypage par groupe d'individus (« *bulk assays* »). Puis, nous avons testé s'il existe des associations significatives entre des marqueurs potentiellement soumis à la sélection divergente au niveau régional et 49 paramètres environnementaux propres aux 26 rivières étudiées. Ultimement, à l'aide de l'information de cartographie génétique (Brenna-Hansen et al. 2012) et d'annotations fonctionnelles, nous avons exploré la distribution génomique des marqueurs potentiellement sous sélection et les implications fonctionnelles de la divergence adaptative et environnementale des populations et groupes régionaux génétiquement distincts de saumon atlantique.

En dernier lieu, le cinquième chapitre concerne les déterminismes génomiques de la mortalité en mer lors de la première année de migration en mer des saumons atlantiques. Dans ce contexte, ce chapitre se propose de répondre à deux objectifs principaux. D'une part, en comparant la composition génomique de saumons juvéniles capturés au moment où ils quittent leur rivière natale (saumoneaux) et celle de saumons adultes de la même cohorte revenant après un an passé en mer (madeleineaux), nous avons cherché à identifier des marqueurs sélectionnés positivement lors de la première année de migration en mer. Le cadre expérimental s'appuyait sur un échantillonnage de saumoneaux et madeleineaux de même cohorte dans deux rivières et pour deux cohortes consécutives. D'autre part et dans une perspective plus fondamentale, nous avons comparé une méthode de détection de la sélection basée sur la différentiation des marqueurs individuels et une méthode considérant les marqueurs en groupe multilocus covariant, une méthode plutôt basée sur le principe de sélection polygénique. Ainsi, nous avons pu comparer l'efficacité de ces méthodes pour la détection de patrons de sélection en mer aux niveaux spatial et temporel.

Chapitre 2 Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees

Publié sous :

Bourret V, O'Reilly PT, Carr JW, Berg PR & Bernatchez L (2011) *Heredity* 106: 500–510.

2.1 Résumé

Chez certaines populations sauvages de saumon atlantique, le déclin rapide du nombre d'individus sauvages remontant leur rivière natale a été associé avec la prévalence de plus en plus importante de saumons domestiques échappés. Certaines études sur la variation phénotypique ont montré que les croisements entre saumons sauvages et saumons d'élevage pouvaient mener à une réduction de l'adaptation locale. Cependant, peu d'études ont tenté d'examiner les impacts d'un tel phénomène au niveau génomique, particulièrement chez les populations nord-américaines. Dans cette étude, nous avons documenté les changements temporels dans la composition génétique de la population sauvage de la rivière Magaguadavic (Baie de Fundy, Canada), une population de saumon atlantique sévèrement menacée qui a vraisemblablement été impactée par des croisements avec des individus domestiques depuis près de 20 ans. Nous avons génotypé des individus sauvages et domestiques remontant la rivière entre 1980 et 2005 à 112 SNPs et/ou 8 loci microsatellites afin d'identifier des changements dans la variation génétique neutre et adaptative. Malgré le déclin rapide de la population au cours des deux dernières décennies, aucun changement temporel significatif n'a été observé en termes de richesse allélique ou diversité génétique estimées à l'aide des marqueurs microsatellites. Ce dernier résultat pourrait être le reflet de l'introgression des saumons domestiques qui a été mise en évidence par un changement temporel significatif du déséquilibre de liaison. De plus, des balayages génomiques ont permis d'identifier une diminution chronologique des loci potentiellement soumis à la sélection directionnelle. Plus spécifiquement, nous avons identifié un SNP préalablement associé à un QTL important pour les taches présentes sur les juvéniles de saumon (tacons) qui retenait sa distinction génétique entre saumons sauvages et domestiques plus longtemps que les autres marqueurs potentiellement sous l'influence de la sélection divergente. En somme, ces résultats indiquent que les saumons domestiques introgressent dans la population de saumon sauvage de la rivière Magaguadavic. Cette situation se traduit par une altération significative de l'intégrité génétique de la population indigène, incluant une perte possible d'adaptation aux conditions naturelles.

2.2 Abstract

In some wild Atlantic salmon populations, rapid declines in numbers of wild returning adults has been associated with an increase in the prevalence of farmed salmon. Studies of phenotypic variation have shown that interbreeding between farmed and wild salmon may lead to loss of local adaptation. Yet, few studies have attempted to assess the impact of interbreeding at the genome level, especially among North American populations. Here, we document temporal changes in the genetic make up of the severely threatened Magaguadavic River salmon population (Bay of Fundy, Canada), a population that might have been impacted by interbreeding with farmed salmon for nearly 20 years. Wild and farmed individuals caught entering the river from 1980 to 2005 were genotyped at 112 SNPs, and/or eight microsatellite loci, to scan for potential shifts in adaptive genetic variation. No significant temporal change in microsatellite-based estimates of allele richness or gene diversity was detected in the wild population, despite its precipitous decline in numbers over the last two decades. This might reflect the effect of introgression from farmed salmon which was corroborated by temporal change in linkage-disequilibrium. Moreover, SNP genome scans identified a temporal decrease in candidate loci potentially under directional selection. Of particular interest was a SNP previously shown to be strongly associated with an important QTL for parr mark number, which retained its genetic distinctiveness between farmed and wild fish longer than other outliers. Overall, these results indicate that farmed escapees have introgressed with wild Magaguadavic salmon resulting in significant alteration of the genetic integrity of the native population, including possible loss of adaptation to wild conditions.

2.3 Introduction

Farmed fish are reared in environments where selective pressures are directed towards improving commercially important traits, namely rapid growth and older age at maturity (Theodorou & Couvet 2004). Consequently, farmed salmon are known to exhibit pronounced differences with wild counterparts at many phenotypic traits such as growth rate, aggressiveness, predator avoidance behavior, acid tolerance (Jonsson & Jonsson 2006; Fraser et al. 2008), which are likely associated with reduced reproductive success in the wild. Moreover, relaxed selection resulting from absence of predators, constant supply of food and medical treatment, as well as increased genetic drift caused by small effective population sizes may favour the accumulation of deleterious alleles. Furthermore, genetic changes unrelated to adaptation to captivity are bound to occur as a result of founder effects in the farmed populations (Roberge et al. 2006). Consequently, change in the genetic composition of captive populations has been observed in many studies on farmed Atlantic salmon of northern Europe (Skaala et al. 2004). Nevertheless, farmed salmon reproduce and hybridize with wild conspecifics, which raises major concerns for conservation programs aimed at protecting the genetic integrity of already diminished populations (Hutchings & Fraser, 2008).

Indeed, the impacts of farmed fish escaping and mixing with wild populations represent a serious threat to natural populations (Fleming et al. 2000; McGinnity et al. 2003; Theodorou & Couvet 2004). Potential alteration of the genetic composition of the wild population impacted by escapees may have significant consequences in terms of reduction of potential to respond to changing environments. Moreover, studies of phenotypic variation have shown that interbreeding between farmed and wild salmon may lead to a reduction in fitness and local adaptation, ultimately contributing to the possible extinction of wild populations (McGinnity et al. 2003; Fraser et al. 2008). Introgressive hybridization between farm and wild salmon also lead to misregulation of gene expression in the latter (Roberge et al. 2008; Normandeau et al. 2009). Yet, despite the potential genetic consequences arising from the introduction of farmed salmon, few studies have attempted to assess the impact of gene flow from farmed Atlantic salmon to wild Atlantic salmon

populations at the genetic level. Such studies require long-term monitoring of impacted populations, and are rare, especially in North America.

Many wild Atlantic salmon (*Salmo salar*) populations of the Bay of Fundy region (Canada) have collapsed since the 1980's and are now threatened with extirpation or have already been extirpated (DFO 2009). Causes for this rapid decline may include changes in oceanographic conditions, freshwater habitat degradation as well as impacts from the Bay of Fundy's commercial Atlantic salmon aquaculture industry (Carr et al. 2004). The Magaguadavic River, located in the Bay of Fundy region (Canada) has been monitored by local authorities since the establishment of the aquaculture industry over 30 years ago. The production of farmed salmon increased dramatically over the last 20 years, resulting in large numbers of fish escaping into the wild. Indeed, in the mid 90's, farmed salmon entering the Magaguadavic river far exceeded the number of wild salmon (Carr et al. 1997). Farmed females have also been shown to successfully spawn in the river. In 1993 alone, up to 55% of the redds were at least partially of farmed origin, despite a greater number of wild salmon entering the river compared to farmed salmon (Carr et al. 1997). At present, wild salmon represent a small proportion of salmon entering the river, whereas farmed fish are still abundant in the system. Clearly, the Magaguadavic River wild Atlantic salmon population may be facing two major challenges in regard to protecting its genetic integrity; a declining census population size and possible interbreeding with farmed salmon.

The main goal of this study was to document temporal change in the genetic composition of the Magaguadavic River wild Atlantic salmon population following the rapid increase in the number of farmed salmon entering the river from 1980 to the early 2000's. We used both microsatellite and SNP markers (comprising both putatively adaptive and neutral markers) to detect the possible effects of interbreeding between wild and farmed salmon at the genetic level. Under the assumption of admixture between the two groups, we first expected neutral markers to exhibit a lower diversity in the domesticated strain and consequently reduced diversity over time in wild introgressed individuals. Since the domestic strain originated from the nearby Saint John River (approximately 60 km from the Magaguadavic River), we also expected the genetic markers to naturally reveal weak

divergence at the majority of markers between farm and wild fish. However, we expected strong divergence at some markers as a result of the radically different selective pressures in captive and wild environments indicated by markers potentially affected (either directly or indirectly through linkage) by divergent or directional selection. Lastly, since evidence of farmed salmon spawning in this system has been shown earlier, if farm salmon have indeed been interbreeding with wild salmon through the 1990's, we expected genetic divergence between these two groups of salmon to be reduced over time as a result of introgression.

2.4 Methods

2.4.1 Sample collection

Samples were collected from wild and farmed salmon captured in the Magaguadavic River between 1980 and 2005. Wild adults were captured at the head of tide fishway trap after ascending the river (Figure 1). From mid-April to June 2000 smolts were either trapped as they descended the river using a fykenet in a small stream serving as the downstream fish bypass outlet for the head of tide hydroelectric dam, or they were collected 24 km upstream of the dam using a rotary screw trap (for example, Solutions Inc, Corvallis, Oregon, USA). Adult salmon used in this study were identified as wild, except in 1992 and 2000 when both wild and farmed individuals were sampled. Wild and farmed smolts were sampled in 2000. Adults and smolts were classified as wild or farmed origin using morphological features such as fin and gill cover erosion (associated with the farming environment), and by examining circuli patterns on scales as previously described by Carr et al. (1997) and Carr & Whoriskey (2006). Blood was collected and frozen for samples from 1992 and either air-dried scale and/or fin clips conserved in 95% ethanol were collected from individuals captured in 1980, and from 1996 to 2005. In summary, the sample collections (also called populations hereafter) used were the following: (1) WILD-1980, (2) WILD-1992, (3) FARMED-1992, (4) WILD-1996, (5) WILD-1998_99, (6) WILD-2000_adults, (7) WILD-2000_smolts, (8) FARMED-2000_adults, (9) FARMED-2000_smolts and (10) WILD-2002+.

2.4.2 DNA Genotyping

The microsatellite and SNP analyses carried out here involved two somewhat overlapping but different groups of samples, primarily because of the unsuitability of older samples for SNP analyses. Organization of samples for analyses is summarized in Table 1. A group composed exclusively of wild adults caught in 1980 (WILD-1980) (n = 31), 1992 (WILD-1992) (n = 60) and 2000 (WILD-2000_adults) (n = 34), farmed adults from 2000 (FARMED-2000_adults) (n = 53) as well as wild and farmed smolts sampled in 2000 (WILD-2000_smolts, FARMED-2000_smolts) (respectively n = 48 and n = 39) was genotyped for microsatellites only (referred to as GROUP A). A second group was genotyped for single nucleotidic polymorphisms (SNPs) (referred to as GROUP B). This

group was organized into different clusters based on aquaculture history in the Bay of Fundy as follows: farm escapees from early-aquaculture 1992 (FARMED-1992) (n = 20), wild adults from the early-aquaculture period in 1992 (WILD-1992) (n = 18), wild adults from the mid-aquaculture period in 1995-96 (WILD-1996) (n = 20), wild adults from the late-aquaculture period in 1998-99 (WILD-1998_99) (n = 22) and wild adults from the very late-aquaculture period in 2002+ (WILD-2002+) (n = 15). Note that all individuals analyzed for SNP variation were also genotyped at the same microsatellite loci as were the samples in GROUP A and the GROUP B WILD-1992 sample collection was a subset of the samples contained in the WILD-1992 collection.

DNA was extracted using Qiagen 96 well DNeasy plates, following the manufacturer's specifications (Qiagen, Valencia, CA, USA). We genotyped individuals at eight microsatellite loci. Polymerase Chain Reaction (PCR) amplifications were carried out in 10 µl volumes, containing between 1-100 ng of template DNA, 2 mM each dNTP, 0.5 µM labelled and unlabelled primers, 50 mM KCl, 0.5 units of Taq DNA polymerase supplied by MBI Fermentas (Burlington, ON, Canada) and 2.0 mM $MgCl_2$. Thermal cycling conditions were as follows: (94°C for 3 min)X1, (94°C for 1 min, 58°C for 30 sec, 72°C for 30 sec)X5, and (90°C for 30 sec, 58°C for 30 sec, and 72°C for 30 sec)X30, followed by a 15 min extension step at 72°C. Primer sequences for loci *Ssa* 197 and *Ssa* 202 are given in O'Reilly et al. (1996), *SSsp* 2201, *SSsp* 2210, *SSsp* 2215, *SSsp* 2216, *SSsp* 1G7 and *SSsp* 1605 are given in Paterson et al. (2004). PCR products were combined, and salt, unincorporated dNTPs, unincorporated labelled and non-labelled primers were removed using Qiagen's 96 MinElute 96 UF PCR purification kits, following the manufacturer's procedures. Fragments were size fractionated and detected using an Applied Biosystems 3130 XL (Carlsbad, CA, USA). One sample from each strip of eight tubes was duplicated to identify sample placement errors, strip inversions, and plate inversions. Two of 10 laboratory samples were analyzed in each group of 96 samples both to identify individual batches, and to ensure allele calls were standardized across batches of samples analysed over time.

SNP genotyping was done using the iPlex Gold® assay on the MassARRAY® platform (Sequenom, San Diego, CA) according to the manufacturer's instructions. In total, a panel of 388 Atlantic salmon SNPs were genotyped in 16 different multiplexes, with multiplexing levels ranging from 12 to 36. The 13 multiplexes with the highest number of SNPs from Moen et al. (2008) as well as the 3 multiplexes with the highest number of SNPs from Lorenz et al. (2010) were selected for genotyping. PCR and extension primers were designed with the software MassARRAY AssayDesign v3.1 (Sequenom). Allele separations were performed using the Sequenom MassARRAY Analyzer. Genotypes were assigned in real time on the basis of the mass peaks (Tang et al., 1999) using the MassARRAY SpectroTYPER RT v3.4 software (Sequenom). Manual inspection of all the results was carried out using the MassARRAY TyperAnalyzer v4.0 software (Sequenom). Individuals were genotyped only once for each multiplex since previous experiences with the SNPs revealed no discrepancies between replicate analyses (CIGENE, unpubl. information). Moreover, the genotyping was followed by highly stringent quality control criteria to keep only high quality markers for further analyses (see results).

2.4.3 Genetic variation and differentiation

For comparative purposes, we considered separate temporal replicates as individual populations in subsequent analyses. For microsatellite analysis, populations of GROUP B were considered with GROUP A with the exception of WILD-1992. Neutral microsatellite genetic variation within each population was quantified using standard descriptive statistics for each locus individually and globally: observed and expected heterozygosities (H_O and H_E) as well as Weir and Cockerham (1984) F_{IS} values were estimated using GENETIX (Belkhir et al. 2001), and the number of observed alleles as well as Nei's (1977) estimator of gene diversity, using FSTAT 2.9.3 (Goudet et al. 2002). We calculated allelic richness ($\hat{A}$), which is the number of different alleles corrected for variation in sample size, using the rarefaction method employed by FSTAT 2.9.3 to the smallest sample (n = 11). Locus *SSsp* 2201 was excluded from the calculation of allelic richness to increase the power of detecting differences in $\hat{A}$ (Leberg 2002) between populations since the sample size would have become too small had it included locus *SSsp* 2201 (n = 3). The distributions of within population gene diversity did not all conform to normality after Shapiro-Wilk tests (4 tests

with $P < 0.01$). On the other hand, allelic richness distribution all conformed to normality after Shapiro-Wilk tests (all $P >> 0.01$) and homogeneity of variance among populations was verified using a Bartlett test ($K^2 = 0.759$, df = 9, $P = 0.999$). Therefore, respectively, a Kruskal-Wallis analysis of variance (non-parametric) and an ANOVA (parametric) were used to determine if gene diversity and allelic richness changed among samples. Conformity of individual loci to Hardy-Weinberg equilibrium (HWE) expectations was also tested using FSTAT 2.9.3 and Fisher's combined probabilities across loci for each population was used to determine a global p-value for the set of Hardy-Weinberg equilibrium tests (Mosteller & Fisher 1948). Genic differentiation (*G*; Guo & Thompson 1992) at individual loci between all pairs of populations and significance values over all loci were obtained using Fisher's method (Ryman & Jorde, 2001) as implemented using GENEPOP 3.4 (Raymond & Rousset 1995). The sequential Bonferroni correction for multiple tests was applied to maintain the table-wide significance level at $\alpha = 0.05$ (Rice 1989) while testing for HWE and genic differentiation. After removal of putatively non-neutral loci identified in FDIST2 analyses (Beaumont & Nichols 1996) (see below), microsatellites and SNP markers were used separetely for calculations of pairwise genetic differentiation between populations using the F_{ST} of Weir and Cockerham (θ) as calculated with GENETIX after 1000 permutations for significance.

2.4.4 Linkage disequilibrium

We do not present any admixture analysis, for instance based on clustering methods (for example, Structure, Pritchard et al. 2000). Although such attempts were made (results not shown), the low level of differentiation observed for the vast majority of markers used (see Results) resulted in insufficient power to apply such methods adequately (Vaha & Primmer 2006). On the other hand, since natural outbred populations are expected to be in near genome wide linkage equilibrium relative to a situation of interbreeding with an invasive population which is expected to generate linkage disequilibrium in a recently introgressed population (Gaut & Long 2003; Allendorf & Luikart 2007), we used ARLEQUIN 3.5 (Excoffier & Lischer 2010) to test for pairwise LD within populations genotyped for SNPs. More specifically, we tested for the presence of significant associations between pairs of loci, based on a likelihood ratio test, where the likelihood of observing the sample

evaluated under the hypothesis of no association between loci (linkage equilibrium) is compared to the likelihood of observing the sample when association is allowed (Excoffier & Lischer 2010). When the likelihood of an association between two loci is significantly greater than no association, they are considered to be linked or under LD. The distributions of the numbers of linked loci within populations deviated significantly from normality (Shapiro-Wilk tests, all $P < 0.01$). Therefore, the median number of loci under LD per locus within each population was compared among populations with a Kruskal-Wallis analysis of variance to determine if LD changed across temporal samples of wild Magaguadavic Atlantic salmon. When significant differences among groups were observed, Wilcoxon signed-rank tests were used to compare median values between populations and determine which samples differed. Furthermore, we used MULTILOCUS (Agapow & Burt 2001) to calculate r_d, the multilocus LD within population, as a measure of genome wide LD over all SNP markers.

2.4.5 Genome scans

Numerous methods to detect loci potentially under the effects of natural selection have been developped over the last decade, most based on principles reported by Lewontin and Krakauer (1973). Some of these methods proposed to compare levels of genetic diversity and differentiation between populations as loci under directional selection should show larger differences, and loci under balancing selection should present less divergence between populations than neutrally evolving loci. FDIST, (Beaumont & Nichols 1996), was one of the first programs widely available to test for departures from neutrality, and the analyses carried out have become widely known as "genome scans". In the latest version of FDIST2, simulations under a finite island-model are performed to obtain a null distribution of F_{ST} values across loci as a function of heterozygosity, and loci with an unusually high or low F_{ST} value ("outliers") are generally considered to be potentially under the effect of natural selection (either directly or indirectly through linkage). Here we scanned genetic variation at SNP markers by comparing aquaculture samples to all wild temporal samples to detect changes in loci potentially under divergent selection. We used a confidence interval of 0.95 for the expected null differentiation meaning that loci over this interval had to be in the upper 0.025 tail of the distribution to be considered as potentially

under directional or divergent selection or in the lower 0.025 tail of the distribution to be considered as potentially under balancing selection.

2.5 Results

2.5.1 Genetic variation

Over all samples, all microsatellite loci were highly polymorphic, with the total number of alleles per locus ranging from 11 to 36 (mean = 24.38), and observed heterozygosity per locus across populations ranging from 0.5 to 1.0 (mean = 0.831) (Table S1). Mean gene diversity per locus per population ranged from 0.831 to 0.878 and allelic richness from 1 to 20 alleles per locus per population. Neither gene diversity nor allelic richness were significantly different among samples (Kruskal-Wallis $K^2 = 2.759$, df = 9 and $P = 0.973$; ANOVA F = 0.019, df = 1 and $P = 0.891$, respectively). The null hypothesis of HWE was not rejected for any locus and for any population after correcting for multiple tests ($\alpha = 0.000625$, k = 80). Fisher's combined probabilities across loci was not significant for seven out of eight populations ($P > 0.05$) with only FARM-2000_adults presenting significant departure from HWE ($P = 0.001$) and with small but positive F_{IS} indicative of a slight heterozygote deficiency (Table S1).

Out of the 388 SNP (Table S2) assayed, 348 markers yielded at least one genotype but only 267 markers were polymorphic. Consequently, 81 were discarded for further analyses. Markers with call rate (CR) inferior to 95% (54) and minor allele frequencies (MAF) less than 0.05 (101) were also excluded to ensure only high quality informative markers were utilized. Thus the final group of SNPs used in all subsequent analyses numbered 112, and average expected and observed heterozygosities are given in Table 1.

2.5.2 Differentiation

Pairwise genic tests of population differentiation at individual microsatellite loci yielded only 26 significant comparisons out of 360 after correction for multiple comparisons ($\alpha = 0.000139$, k = 360). This translated into a non-significant ($P = 0.999$) global F_{ST} value of 0.00096, reflecting an overall very weak level of differentiation among populations. Furthermore, microsatellite pairwise F_{ST} values (θ) were significant in only 20 out of 45 comparisons (Table 2) and ranged from 0.005 between WILD-2000_smolts and FARM-2000_adults and 0.019 between WILD-1980 and FARM-2000_adults and between WILD-1996 and WILD-2002+ when significant. The WILD-1980 population was associated with

the greatest number of significant pairwise F_{ST} estimates (8 of 9), and the largest F_{ST} values, ranging from 0.009 to 0.018 when compared to other wild population temporal replicates, and from 0.011 to 0.019 when compared to aquaculture sample collections. The sample collection WILD-2002+ was associated with the next largest F_{ST} values, ranging from 0.010 to 0.019 when compared against other wild temporal replicates, of which four of five were significant (WILD-1992, WILD-1996, WILD-2000_adults and WILD-2000_smolts). The lowest significant F_{ST} values based on microsatellite data were observed for pairwise comparisons between wild replicates post-1980 and aquaculture samples (range between 0.005 and 0.009). For SNP markers, pairwise F_{ST} values varied between 0.000 (in 5/10 comparisons) and 0.006 (between WILD-1996 and FARM-1992) (results not shown). None of these pairwise F_{ST} values were significant after removal of 6 SNPs identified by FDIST2 as potentially under divergent selection in a global test over all samples at a significant level of 0.025 (results not shown).

2.5.3 Linkage disequilibrium

SNP markers yielded variable patterns of linkage disequilibrium in wild temporal replicates and the 1992 farm salmon samples (Figure 2). The Kruskal-Wallis test of equality of medians among populations was significant (K^2 = 139.470, df = 4 and $P << 0.001$). Pairwise comparisons showed that FARM-1992 and WILD-1996 as well as WILD-1998_99 and WILD-2002+ exhibited similar numbers of linked loci per locus (respectively W = 6686 and 6753 with P = 0.394 and 0.321) while all 8 other comparisons were significantly different after corrections for multiple comparisons ($P << 0.001$). Multilocus LD (r_d) values within each population are given in Figure 2 and corroborate the detailed locus-by-locus LD analyzes within populations in the sense that all post-1992 wild temporal samples had higher r_d values than the 1992 wild sample, and the aquaculture sample (FARM-1992) presented the highest multilocus LD.

2.5.4 Fdist outlier detection test

Genome scan plots of marker F_{ST} as a function of heterozygosity comparing each of the four wild samples with the farmed salmon (FARM-1992) are shown in Figure 3, with outliers potentially under divergent selection over the 95% confidence level. Single locus

F_{ST} values for pairwise comparisons were approximately 0.01, varying slightly from locus to locus (between 0.006 for WILD-1998_99 and 0.011 for WILD-1996) whereas outlier F_{ST} values were all above 0.10. Overall, we found a two-fold reduction in the number of observed markers potentially under the effect of divergent selection after 1996, with the observed number of outliers, in chronological sequence, of 10 (WILD-1992), 9 (WILD-1996), 4 (WILD-1998_99) and 5 (WILD-2002+). Table 3 presents a summary of these markers, including their names, known linkage group used from Boulding et al. (2008) and Lorenz et al. (2010), as well as gene annotation when available. Three outliers (Contig14899_0107, Contig15610_550 and Contig16686_0431) identified in the genome scan involving WILD-1992, were also identified in comparisons with WILD-1995_96 and/or WILD-1998_99. Of these markers, Contig14899_0107 presented a nearly significant reduction in F_{ST} through time ($r^2 = 0.88$ $P = 0.062$), and was no longer an outlier in the WILD-2002+ genome scan (Figure 4). Loci showing less differentiation than that represented by the lower 95% confidence interval would normally be considered as potentially under balancing selection, but since this confidence level always fell below the $F_{ST} = 0$, we rejected the hypothesis of these markers being under the effect of balancing selection.

2.6 Discussion

This study is the first to examine the genetic impacts of introgression from farmed Atlantic salmon into a wild population using a wide panel of single nucleotide polymorphism (SNP), in addition to known neutral microsatellite markers. The long term monitoring of the Magaguadavic River of the Bay of Fundy in Eastern Canada permitted a comparison of the genetic composition of wild samples before and after the establishment and growth of the salmon farming industry in the area. More specifically, we assessed whether introgression affected (i) levels of neutral genetic diversity in wild Magaguadavic River Atlantic Salmon over time, (ii) the extent of differentiation between wild and farmed replicates at both types of markers, (iii) the signature of divergent selection possibly operating at some SNP loci and (iv) the initial degree of distinctiveness or conversely, the homogenization of wild and farmed salmon. Our results strongly suggest that introgression did change the genetic signature of the wild population, by, for example, reducing the divergence at neutral and non-neutral markers. Although levels of within population genetic diversity in the wild population did not appear to change, levels of differentiation between wild and farmed salmon were reduced through time, likely in response to the increasing presence and influence of farmed salmon in the Magaguadavic River. Perhaps more importantly, the overall number of loci potentially under divergent selection was reduced when wild temporal samples were contrasted with a reference farmed sample. Furthermore, a specific marker putatively involved in local adaptation (see Discussion below) epitomized the homogenizing effect of introgression by showing a reduction of differentiation in wild populations compared to farmed samples over time, until patterns of locus variability and population differentiation at this locus were indistinguishable from that expected due to neutral evolutionary forces.

2.6.1 Neutral genetic changes

The unexpected stability of levels of gene diversity and allelic richness of wild temporal replicates might have been foreseen in the light of introgression by a similarly diverse group of individuals and consistent input from this group. Previous studies of European domesticated fish strains showed reduced allelic variation compared to their European wild relatives (Mjølnerød et al. 1997; Norris et al. 1999; Skaala et al. 2004). Furthermore, since

farmed salmon typically involve a restricted number of effective breeder, selection for performance traits and low or no gene flow, we expected farmed Atlantic salmon to show this kind of reduced genetic variation. However, we found that farmed salmon caught entering or descending the river as smolts did not exhibit lower levels of allelic richness or gene diversity compared to wild salmon. Although this remains to be rigorously tested, it is possible that some of the numerous freshwater aquaculture facilities each has lost and retained different alleles over time, and that collectively, they contain a large standing pool of genetic variation. These stocks would have acted partly as a metapopulation in neutralizing any possible effect of an isolated group of founders or of genetic drift.

Admixture from farm escapees in the wild may threaten natural populations through "maladaptive" genetic changes (Skaala et al. 2006). Starting between 1992 and 1996, there were indications that introgression occurred in the wild Magaguadavic population. This was illustrated by the number of linked loci per locus in the wild replicate sample collections that went from a relatively stable state corresponding to genome wide equilibrium to numerous linkage desequilibrium strongly suggestive of a blending between two somewhat different populations. Indeed, the number of linked loci per locus in the wild sample from 1996 was similar to that observed in the farmed sample, which was possibly attributable to some variability in the origins or allele frequency distributions among the farmed population within the Magaguadavic River system. Moreover, this time period corresponded with the expansion of the farming industry in the Magaguadavic River and the observed turning point in relative abundance of wild versus farmed Atlantic salmon returning to the river as adults (Carr et al. 2004). Population bottlenecks also have the potential to increase LD for a short period of time (Gaut & Long 2003) but we would argue that in the present study, although the rapid decline of the wild population might have contributed to the increase LD and cannot be ruled out, the sudden genome wide increase in LD is more consistent with a scenario involving recombination imposed by admixture. Consequently, while after 1994 the number of farmed fish returning to the river outnumbered the rapidly declining wild Atlantic salmon population, our results suggest that the reduction of gene diversity expected in this declining population was counteracted by

the outbreeding effect of introgression from a slightly divergent and genetically variable population of aquaculture salmon.

To our knowledge, the most comparable study to ours involving wild Atlantic salmon was carried out by Skaala et al. (2006) and showed a reduction of differentiation and genetic distances among wild populations affected by farm escapees for 20-30 years despite no genetic diversity change in wild populations. Using microsatellites, we found a modest albeit significant genetic difference between pre-aquaculture wild individuals from 1980 and all other wild replicates except for the very late-aquaculture period samples. Moreover, comparisons of wild temporal samples also yielded significant levels of differentiation. Of the 5 of 7 significant comparisons involving pre- or early-aquaculture period samples and farmed fish, the highest value was reached between the oldest wild samples (WILD-1980 and WILD-1992) and most recent aquaculture adults (FARM-2000_adults). These results, along with the finding that the majority of non-significant genetic differences observed among the most recent (mid-aquaculture period and later) wild samples and aquaculture samples are consistent with a homogenization effect occuring between the farmed and wild population of the Magaguadavic River due to introgression. In similar studies, Koskinen et al. (2002) and Hansen et al. (2010) reached similar conclusions when they contrasted historical and contemporary replicates of, respectively, wild grayling and wild brown trout populations admixed with hatchery fishes; both observed a significant reduction in levels of differentiation between wild and hatchery strains reflecting recent admixture. In a recent study on brook charr (*Salvelinus fontinalis*), Marie et al. (2010) also documented a homogenization of populations structure among wild populations following stocking events with domestic fish, accompanied with a greater genetic similarity between wild and domestic fish. Our study yielded similar findings, that admixture did reduce genetic differentiation between the wild and farm salmon despite no overall effect detected on the level of genetic diversity in the wild population.

2.6.2 Possible effects on local adaptation

Although the genome scans performed here can be criticized for the lack of fit to reality with respect to the underlying models leading to false positives (for example, inclusion of

populations that have experienced recent and marked population bottlenecks (Foll & Gaggiotti 2008), we nevertheless consider the use of this approach as useful in at least obtaining a qualitative measure of the number of loci highly divergent between farmed salmon and wild temporal replicates as a good indication of the proportions of genomic regions potentially under the effect of divergent selection in each comparison. Here, outliers were approximately one order of magnitude more differentiated than the average SNP marker analyzed in the two sample groups compared. Moreover, our results suggest a 50% reduction in the number of loci potentially under divergent selection between wild and farmed salmon after only 3-5 generations of significant admixture occuring in the wild population. Also, the locus with the most pronounced differentiation (F_{ST} = 0.18) in the comparison between farmed and wild salmon in 1992 (early aquaculture period) was still identified as an outlier when compared in the mid- and late-aquaculture periods with a reduction of its differentiation value until it faded away into the neutral distribution in the most recent comparison. Admittedly, the finding of a single locus, Contig14899_0107, exhibiting this pattern over four time points is not unlikely to occur by chance, even given the high regression coefficient observed. However, this marker was associated with a quantitative trait locus (QTL) for parr mark number in a previous study (Boulding et al. 2008). Parr marks have been suggested to be a variable trait important for predator avoidance in the wild, allowing juvenile Atlantic salmon to blend in with the local substrate of the river (e.g. Donnelly & Dill, 1984). In summary, the large decrease in F_{ST} over four time points, the association of the locus with a previously identified QTL, and the fact that it presented the highest differentiation in the original genome scan represents compelling evidence that the trend observed is biologically meaningful and did not occur purely by chance. Therefore, this particular case of homogenization between wild and aquaculture fish in addition to the previously demonstrated reduction for markers potentially under divergent selection, suggests that the wild population of Atlantic salmon in the Magaguadavic River likely suffer from a loss of local adaptation exacerbated by introgression from farmed salmon.

On the other hand, this reduced number of outliers might also reflect the fading of selected traits in farmed Atlantic salmon as there is no indication of the directionality for the

divergent selection pressure underlying the high differentiation of these markers. However, why selection (both intentional and unintentional) in a captive environment would have caused an increase in genetic similarity with a given wild population is unclear. In any case, it could still be argued that regardless of whether the selected trait(s) represent both local adaptations to native river conditions in the wild population and the hatchery environment in the captive population, the homogenization most probably resulted in reduced adaptation to both environments. The admixture between both groups probably resulted in a breakdown of linkage disequilibrium around divergently selected regions of the genome (Via & West 2008) and incidently caused the observed reduction of detected outliers regardless to the actual selection occuring in the wild, which would have to be very strong to counteract the recombination occuring following introgression. Of course, further studies (such as that carried out by Fraser et al. (2008)), comparing fitness of wild and farmed-wild hybrid salmon, should be undertaken to measure fitness related traits of differently introgressed individuals in association with their actual allelic composition. However, in an obligate non-invasive field study such as this one, we argue that until more important genomic resources are available, or another significant decrease in the cost of whole genome sequencing is realized, our results make a strong case for reduced local adaptation in a wild population impacted by farm escapees.

2.6.3 Perspectives for aquaculture management

Hutchings and Fraser (2008) recently debated the pros and cons of using aquaculture strains derived from local versus non-local wild populations for salmon farming for a given region. Adding to this debate, our study suggests that although we observed introgression from farm escapees in the wild population with significant loss of neutral differentiation and a reduction of the potential number of loci under divergent selection, neutral genetic variability still prevailed in a rapidly declining population. Since there is evidence of captive-reared salmon establishing themselves outside of their native ranges (Volpe et al. 2000; Soto et al. 2001), despite the inferior fitness of farmed Atlantic salmon in natural environments (McGinnity et al. 2003), one could argue that a foreign strain might have had stronger impacts on the wild salmon in terms of changing the genetic composition of such a vulnerable population, impacting both neutral and selective divergence. Nevertheless, given

the geographic and likely phylogenetic proximity of the wild and putative source population of the aquaculture strain analyzed here, detecting outlier loci was already evidence of differential selective pressures acting on farmed and wild salmon. In addition, their fading through time, especially when associated with an important QTL for juvenile Atlantic salmon, certainly should be regarded as a warning for the potential loss of local adaptation in any kind of farming industries unintentionally releasing individuals in the natural environment

In conclusion, despite its relatively limited genomic coverage, this study emphasizes the potential importance of long term monitoring of wild populations impacted by farmed escapees. Furthermore, in this context and with increasingly important genomic resources soon to become available for non-model species such as Atlantic salmon, more in depth studies could further disentangle the relative impacts of farmed escapees on the neutral and adaptive diversity of affected populations on a genome wide basis.

2.7 Acknowledgements

We are grateful to M. Best and E. Merrill for their assistance in the field; to L. Hamilton, M. Cassista-da Ros, and L. de Mestral Bezanson (DFO) for DNA extractions and microsatellite genotyping analysis. The SNP genotyping was performed by CIGENE at the national technology platform, supported by the functional genomics programme (FUGE) in the Research Council of Norway. We also thank P.A. Gagnaire, J. Prunier, S. Renaut, M. Evans, C.R. Primmer and two anonymous reviewers for their very constructive inputs. The New Brunswick Environmental Trust Fund provided Field funding for this project. Research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to VB, from the Canadian Research Chair in genomics and conservation of aquatic resources to LB, and from funding provided by the Genomics Research and Development Initiative, Department of fisheries and Oceans (DFO), Canada to PTO.

2.8 Tables

Table 2.1 Summary of sample collection organization, number of sucessfully genotyped markers and principal genetic diversity parameters per population: allelic richness (Â) for microsatellites (excluding SSsp 2201), Nei's gene diversity, observed heterozygosity (HO), value of heterozygote deficit (FIS), average SNP expected (SNPs HE) and observed (SNPs HO) heterozygosities.

GROUP	Population	Years Sampled	N Samples	N Microsatellites	N SNPs	$\hat{A}$	Nei's Gene Diversity	H_O	F_{IS}	SNPs H_E	SNPs H_O
A	WILD-1980	1980	31	8		9.019	0.862	0.791	0.071		
	WILD-1992	1992	60	8		8.865	0.864	0.842	0.016		
	WILD-2000_adults	2000	34	8		8.766	0.864	0.817	0.054		
	WILD-2000_smolts	2000	48	8		8.733	0.862	0.847	0.024		
	FARM-2000_adults	2000	53	8		9.130	0.867	0.827	0.049		
	FARM-2000_smolts	2000	39	8		8.960	0.865	0.836	0.038		
B	WILD-1992*	1992	18	8	112	9.183	0.871	0.872	0.002	0.246	0.263
	WILD-1996	1996	20	8	112	8.609	0.878	0.780	0.114	0.242	0.269
	WILD-1998_1999	1998	2	8	112	8.764	0.858	0.820	0.038	0.252	0.247
		1999	20								
	WILD-2002+	2002	7	8	112	8.677	0.831	0.841	-0.011	0.249	0.257
		2003	4								
		2004	2								
		2005	2								
	FARM-1992	1992	20	8	112	8.847	0.878	0.874	0.009	0.279	0.247

* WILD-1992 of GROUP B is a subsample within WILD-1992 of GROUP A

Table 2.2 Pairwise measures of genetic differentiation based on allelic identity (θ_{ST}) at microsatellites. * asterisks indicate significant comparison ($P < 0.05$).

	WILD 1992	WILD 1996	WILD 1998-1999	WILD 2000_adults	WILD 2000_smolts	WILD 2002+	FARM 1992	FARM 2000_adults	FARM 2000_smolts
WILD-1980	0.009*	0.018*	0.015*	0.011*	0.012*	0.008	0.014*	0.019*	0.011*
WILD-1992		0.011*	0.002	0.004	0.001	0.012*	0.003	0.008*	0.008*
WILD-1996			0.000	0.005	0.004	0.019*	0.003	0.003	0.009*
WILD-1998_1999				0.000	0.000	0.011	0.000	0.000	0.005
WILD-2000_adults					0.000	0.012*	0.000	0.002	0.003
WILD-2000_smolts						0.010*	0.003	0.005*	0.008*
WILD-2002+							0.007	0.008*	0.010
FARM-1992								0.000	0.001
FARM-2000_adults									0.006*

Table 2.3 Summary information on outliers detected in genome scans comparing FARM-1992 to wild temporal samples at 112 SNPs. For each comparison, outlier's ID is given with its known linkage group and gene annotation. Nomenclature used for linkage group corresponds to that used in the ASalBASE (powered by cGRASP) Atlantic salmon linkage map (http://www.asalbase.org/sal-bin/map/index). N/A: non-available.

FARM-1992 *vs*	Outliers'ID	Linkage group	Gene annotation
WILD-1992	BASS119-B7-A09_382SNP	AS15	UNKNOWN
	BASS19-B7-F09_342SNP	AS14	UNKNOWN
	Contig14899_0107	AS22	NADH dehydrogenase subunit 5
	Contig17364_0264	N/A	Elongation factor
	Contig16686_0431	N/A	Small ubiquitin-related modifier
	Contig16628_1277	N/A	Antithrombin-III precursor
	Contig16686_0312	N/A	Small ubiquitin-related modifier
	Contig17081_268	AS04	Serine incorporator
	Contig15610_550	N/A	Glutamyl aminopeptidase
	Contig13137_0137	AS18	Protein MON2 homolog
WILD-1996	Contig16856_0321	N/A	Renin receptor precursor
	Contig16378_0529	ASA	Cell division protein kinase
	Contig16053_552	N/A	Mitogen-activated protein kinase
	Contig15610_504	N/A	Glutamyl aminopeptidase
	Contig15610_550	N/A	Glutamyl aminopeptidase
	Contig16207_0498	AS10	UNKNOWN
	Contig14579_490	AS02	Eukaryotic initiation factor
	Contig14782_0767	AS18	Very long-chain acyl-CoA synthetase
	Contig14899_0107	AS22	NADH dehydrogenase subunit 5
WILD-1998_99	Contig16221_0769	N/A	26S proteasome non-ATPase
	BASS133-B7-H09_429SNP	AS12	Retinoic acid receptor gamma b
	Contig16686_0431	N/A	Small ubiquitin-related modifier
	Contig14899_0107	AS22	NADH dehydrogenase subunit 5
WILD-2002+	BASS111-B7-D03_200SNP	AS05	MHC class I antigen
	Contig14638_214	AS22	Mitochondrial 28S ribosomal protein
	BASS113-B6A-F03_685SNP	AS27	UNKNOWN
	Contig15118_153	AS17	Mitochondrial 28S ribosomal protein
	Contig16207_0498	AS10	UNKNOWN

2.9 Figures

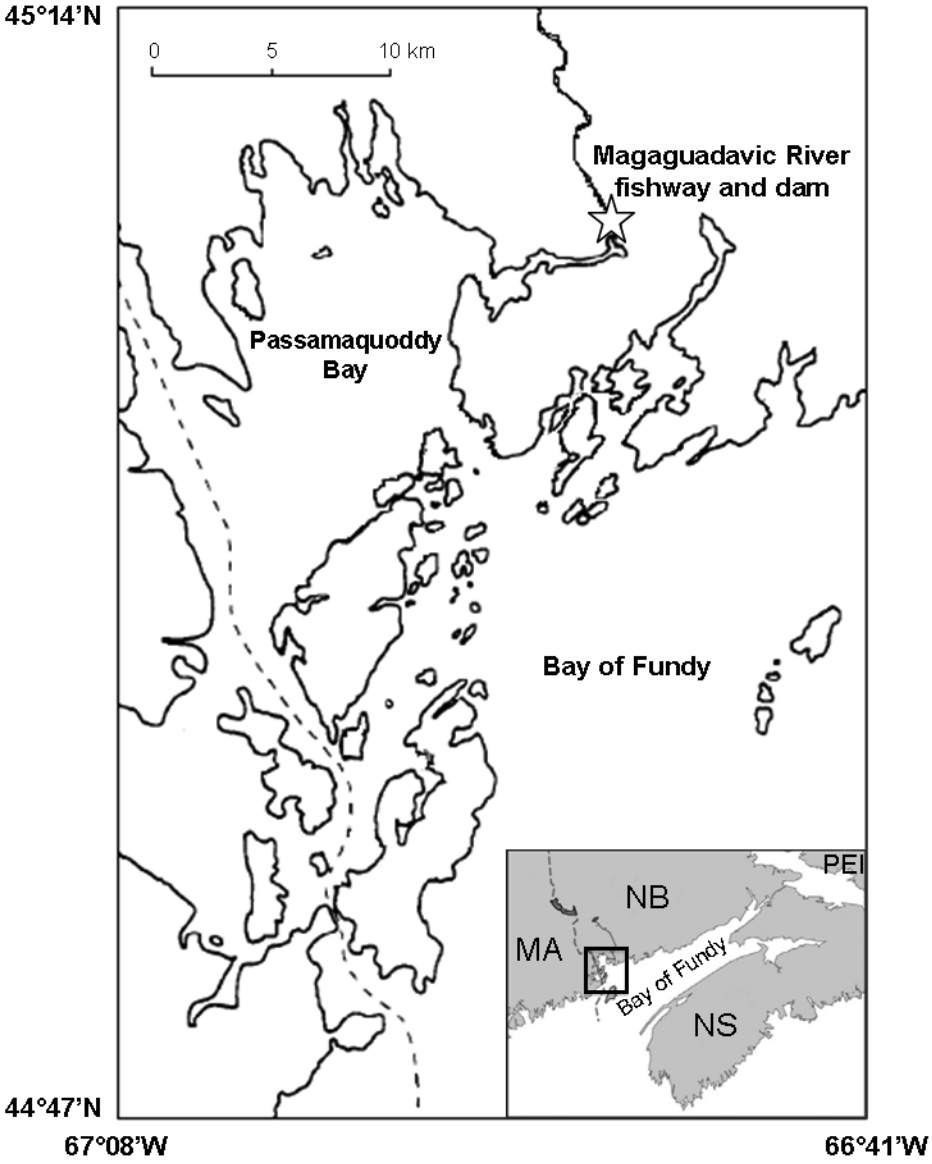


Figure 2.1 Map showing the Magaguadavic River mouth and the fishway located close to the Passamaquoddy Bay where most of New Brunswick commercial salmon sea-cage sites are in the Bay of Fundy. Modified with permission from Carr et al. (1997) and Carr et al. (2004). (NB = New Brunswick, NS = Nova Scotia, PEI = Prince Edward Island and MA = Maine)

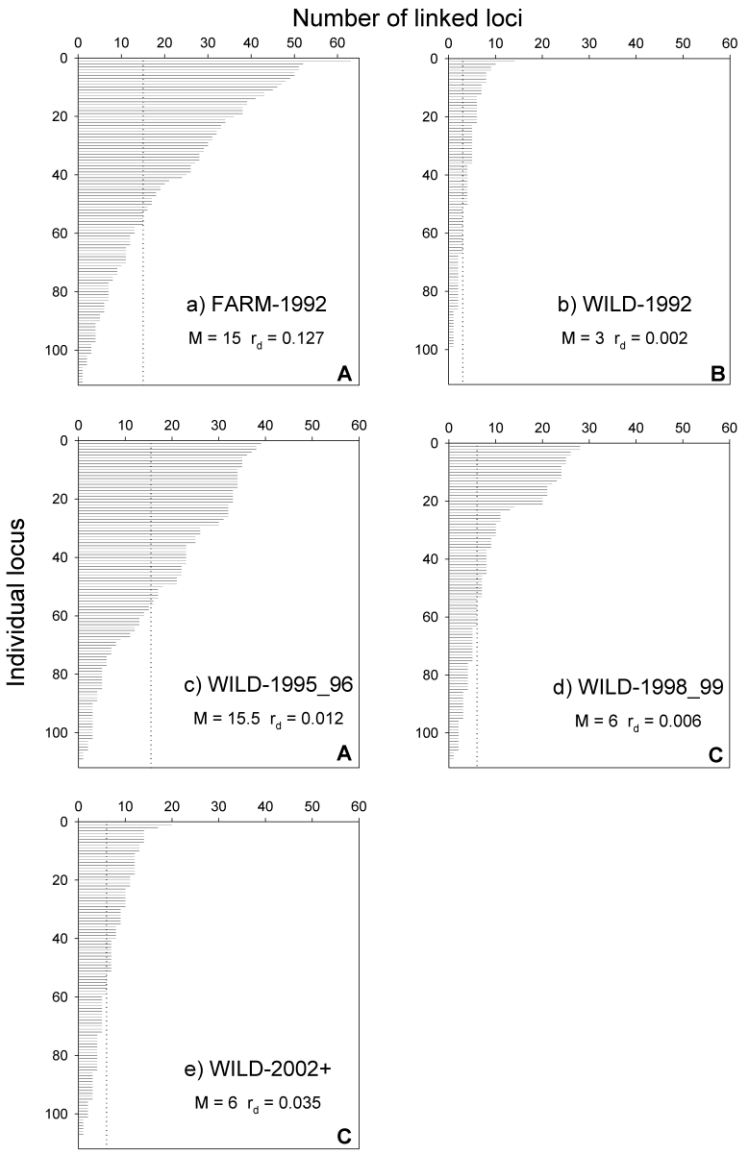


Figure 2.2 Number of SNP linked loci per individual SNP locus within population for: a) FARM-1992, b) WILD-1992, c) WILD-1995_96 d) WILD-1998_99 and e) WILD-2002+. Loci are not arranged in the same order among populations but by decreasing order according to the x-axis. Median number of linked loci per locus per population (M) and multilocus rd per population are indicated on each panel. Within population M are not different for populations sharing the same capital letter in the bottom-right corner of panels after Wilcoxon tests (see Results).

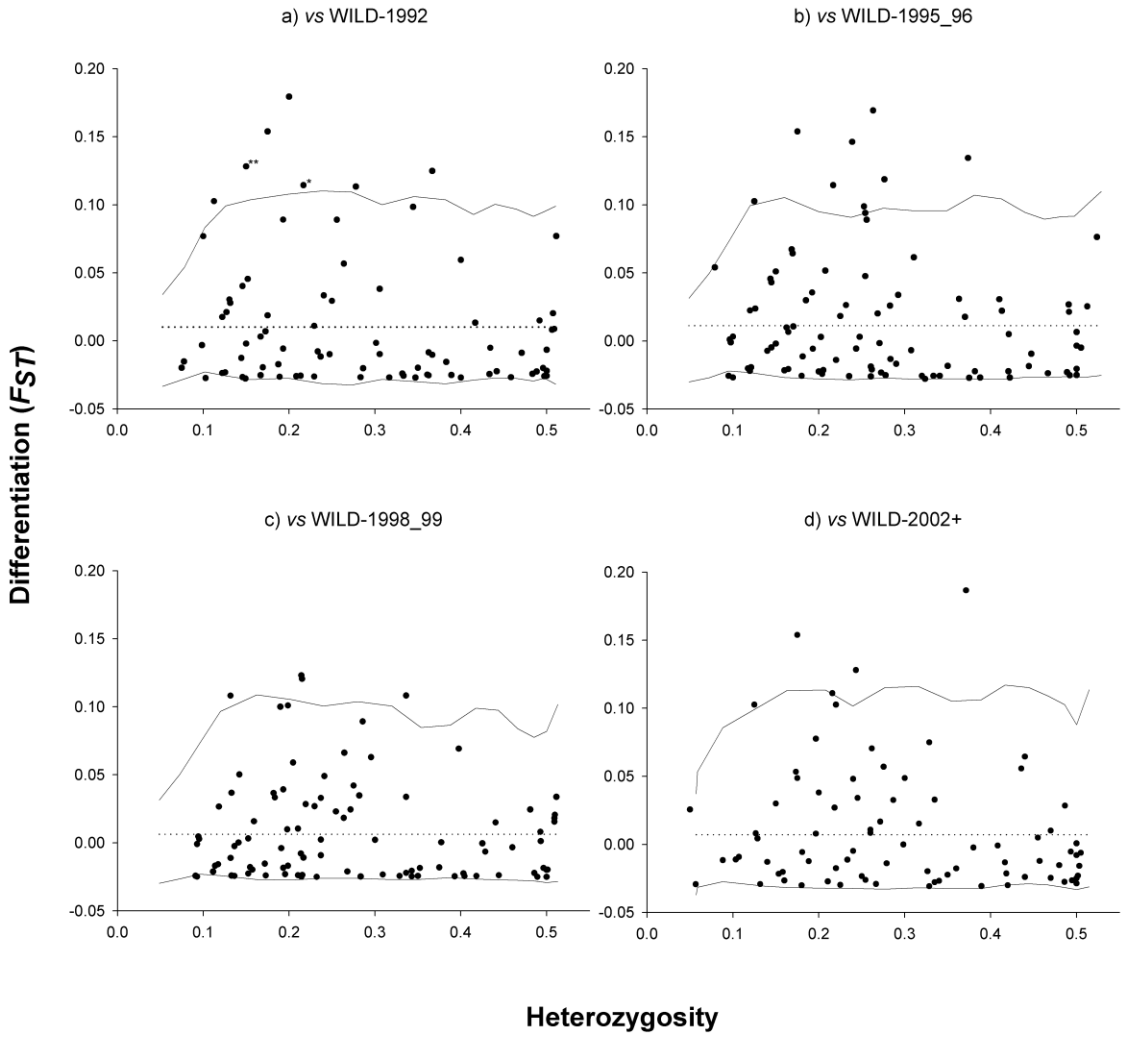


Figure 2.3 Differentiation (F_{ST}) as a function of heterozygosity as calculated by FDIST2 when comparing FARM-1992 with: a) WILD-1992, b) WILD-1995_96 c) WILD-1998_99 and d) WILD-2002+. On each panel, solid line represent upper and lower 95% confidence level and dotted line indicates the average F_{ST} across loci. Astericks (*) on outliers indicate two values represented by the same dot on the graphic (double astericks (**) three values). Only SNP markers were used.

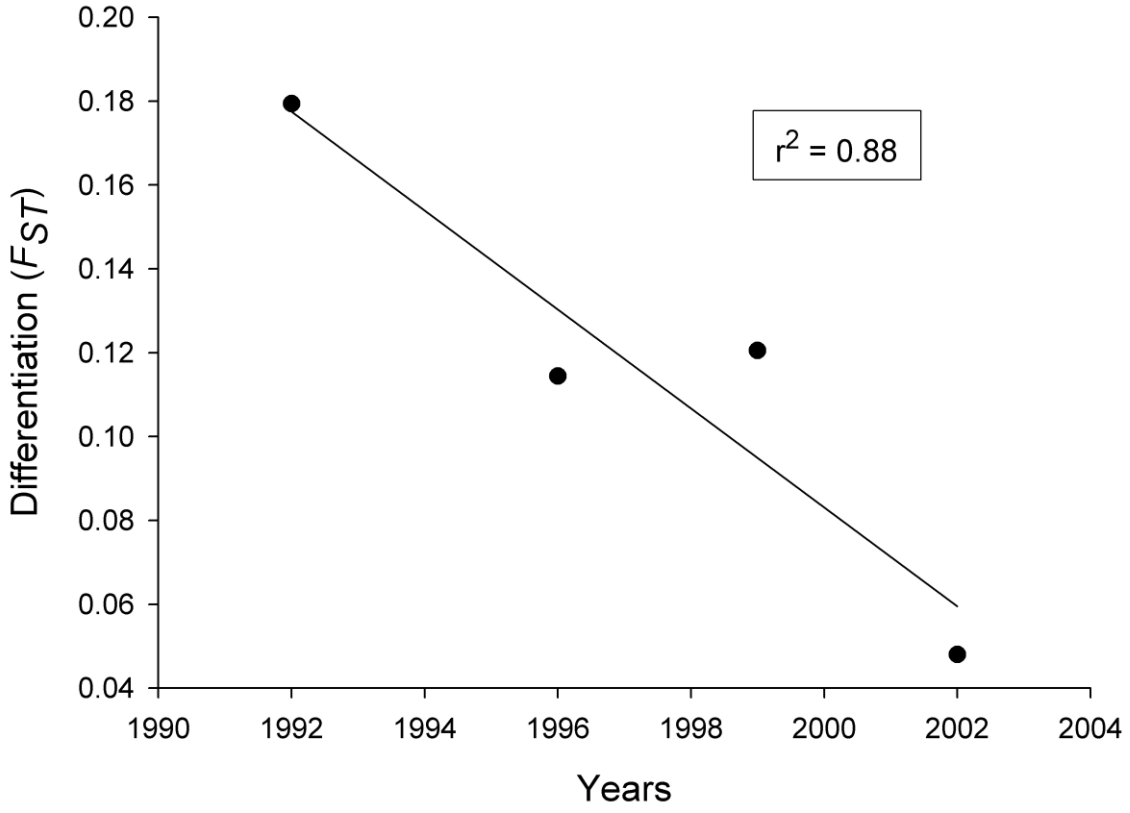


Figure 2.4 Differentiation (F_{ST}) of Contig14899_0107 across time as estimated by FDIST2 in genome scans (Figure 3). Wild temporal samples were considered as single year sample based on the sample year predominant in the population (WILD-1996 = 1996, WILD - 1998_99 = 1999 and WILD-2002+ = 2002). Regression value is indicated in the top-right corner ($P =$ 0.062).

2.10 Supplementary material

Hereafter listed supplementary tables are available online or on demand:

Table 2.S1 Summary of genetic diversity at 8 microsatellite loci among wild and aquaculture samples collections (population) obtained from the Magaguadavic River. Per locus in each population: *N* the number of individuals sucessfully genotyped, number of observed alleles (*A*), allelic richness***** ($\hat{A}$), expected (H_E) and observed (H_O) heterozygosities and correlation value of heterozygote deficit (F_{IS}). *****Allelic richness ($\hat{A}$*;* excluding *SSsp* 2201) corrected to $n = 11$ using the rarefaction method of FSTATS 2.9.3 (Goudet, 2002).

Table 2.S2 Summary of 388 single nucleotidic polymorphism (SNP) markers genotyped at CIGENE. Per marker information: alternative identification (ID), call rate, accession number in dbSNP (NCBI; http://www.ncbi.nlm.nih.gov/projects/SNP/) and sequence flanking the SNP. Per sucessfully genotyped markers: number of alleles, minor allele frequency, observed and expected heterozygosities. For each marker used in further analysis are identified with a YES in the column "112 used in analysis".

Chapitre 3 SNP-array reveals genome wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*)

Publié sous :

Bourret V, Kent MP, Primmer CR, Vasemägi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L & Lien S (2013) *Molecular Ecology* 22: 532-551.

3.1 Résumé

En raison de son importance économique et des efforts de gestion et de conservation dévoués à l'espèce pour la protéger d'un déclin majeur de ses populations, le saumon atlantique est l'une des espèces de poissons les plus étudiées dans le monde. Néanmoins, des ressources génomiques limitées ont restreint notre compréhension de l'architecture génétique de l'espèce et des bases génétiques adaptatives lui permettant d'occuper un environnement naturel et artificiel vaste et hétérogène. Dans cet article, nous décrivons le développement d'une biopuce à SNP de densité moyenne basée sur des ESTs (*Expressed Sequence Tags)* et la découverte de nouveaux marqueurs par séquençage génomique *de novo*. La biopuce a été utilisée pour réaliser la plus importante étude de structure génétique des populations jamais réalisée jusqu'à présent chez le saumon atlantique. Un total de 6176 marqueurs SNP ont été génotypés pour 38 populations naturelles anadromes et résidentes d'eau douce distribuées sur l'ensemble de l'aire de répartition naturelle de l'espèce. L'analyse en composantes principales différencie clairement les populations d'Europe et d'Amérique du Nord alors qu'en Europe, trois groupes régionaux principaux ont été identifiés pour la première fois dans une même étude. Nous avons évalué le potentiel des fréquences alléliques quantifiées à l'aide de la biopuce pour distinguer la divergence putative neutre et la divergence potentiellement adaptative des populations et groupes régionaux. En Europe, des zones de contact secondaire ont été identifiées entre les groupes régionaux. Ces zones sont potentiellement associées à des barrières exogènes et endogènes, ce qui rend l'interprétation équivoque quant à l'influence de l'environnement sur la divergence adaptative. Nous discutons finalement des applications potentielles de la biopuce en gestion, conservation et aquaculture.

3.2 Abstract

Atlantic salmon (*Salmo salar*) is one of the most extensively studied fish species in the world due to its significance in aquaculture, fisheries and ongoing conservation efforts to protect declining populations. Yet, limited genomic resources have hampered our understanding of genetic architecture in the species and the genetic basis of adaptation to the wide range of natural and artificial environments it occupies. In this paper, we describe the development of a medium density Atlantic salmon SNP-array based on Expressed Sequence Tags (ESTs) and genomic sequencing. The array was used in the most extensive assessment of population genetic structure performed to date in this species. A total of 6176 informative SNPs were successfully genotyped in 38 anadromous and freshwater wild populations distributed across the species natural range. Principal component analysis clearly differentiated European and North American populations, and within Europe, three major regional genetic groups were identified for the first time in a single analysis. We assessed the potential for the array to disentangle neutral and putative adaptive divergence of SNP allele frequencies across populations and among regional groups. In Europe, secondary contact zones were identified between major clusters where endogenous and exogenous barriers could be associated, rendering the interpretation of environmental influence on potentially adaptive divergence equivocal. A small number of markers highly divergent in allele frequencies (outliers) were observed between (multiple) freshwater and anadromous populations, between northern and southern latitudes, and when comparing Baltic populations to all others. We also discuss the potential future applications of the SNP-array for conservation, management and aquaculture.

3.3 Introduction

Since their divergence from other salmonids, Atlantic salmon (*Salmo salar* L.) populations have been subjected to strong natural and artificial selection throughout their evolutionary history. These included pronounced habitat shifts associated with glaciations, "landlocking" of fjords or lakes previously open to the sea, as well as various sources of anthropogenic changes (Verspoor *et al.* 2007). The greatest divergence between Atlantic salmon populations can be seen when comparing the European and North American lineages, which diverged approximately 600,000 years before present (ybp) (King *et al.* 2007). A more recent fundamental division is seen in populations of non-anadromous fish on both continents, which have evolved independently following "landlocking" of their rivers or fjords (Tessier & Bernatchez 2000; King *et al.* 2007). In addition, Atlantic salmon inhabit a wide range of habitats with different temperatures, from Spain to the high Arctic (Aas *et al.* 2011). Atlantic salmon thus represents an ideal species to disentangle the genomic basis of potential parallel evolution at both historic and recent scales.

More recently, there has been a dramatic decline of many wild Atlantic salmon populations due to overfishing, habitat destruction and possibly other indirect causes such as impacts from aquaculture or climatic change (Friedland *et al.* 2003; Carr *et al.* 2004; Ferguson *et al.* 2007; Ford & Myers 2008; Todd *et al.* 2008). Consequently, wild Atlantic salmon populations have been the focus of international conservation efforts, where population genetic studies have played an important role by genetically inferring phylogeographic patterns and defining Evolutionary Significant Units (ESU) as well as Management Units (e.g. Vasemägi *et al.* 2005; Palstra *et al.* 2007; Dionne *et al.* 2008; Tonteri *et al.* 2009). However, the limited resolution of the available genetic tools has meant that the genomic basis of local adaptation shaped by natural selection is still largely unknown for Atlantic salmon (Fraser *et al.* 2011). SNP-arrays represent excellent tools for studying population structure and the effect of natural and artificial selection at the genome level (e.g. The Bovine HapMap Consortium, Gibbs *et al.* 2009; Willing *et al.* 2010, many others). For example, for Atlantic cod (*Gadus morhua*), Bradbury *et al.* (2010) were able to show parallel clinal association between SNP allele frequencies and water temperatures across the species range, both in North America and Europe. A SNP-array for salmon would add

substantially to previous efforts by providing managers with an enhanced tool for definition of ESUs on the basis of neutral and adaptive divergence (Kohn *et al.* 2006; Primmer 2009). Moreover, it could also improve the resolution of differentiation of genetic stocks and the efficiency of mixed-stock analysis (Gauthier-Ouellet *et al.* 2009; Griffiths *et al.* 2010).

Aquaculture of Atlantic salmon is practiced on a large scale in both the Northern and Southern hemispheres, and today it is the most cultivated fish in the western world and the most farmed salmonid world-wide. This expanding industry is recognizing the potential benefits arising from development and application of high density SNP technologies (Dominik *et al.* 2010). Indeed, similar resources have been developed for other production species and are proving invaluable for the detection of quantitative trait loci (QTL) underlying economically important traits (Dekkers & Hospital 2002; Daetwyler *et al.* 2008), strain identification (Suekawa *et al.* 2010) and to implement marker-assisted selection and genomic selection (Soller 1990; Meuwissen *et al.* 2001; Dekkers & van der Werf, 2007). Moreover, accidental release and/or the escape of farmed fish into the surrounding wild environment has raised substantial concerns (McGinnity *et al.* 2003) over the potential genetic impact these fish can have on local wild populations at the genome level (e.g.: Hindar *et al.* 1991; Roberge *et al.* 2006; Normandeau *et al.* 2009; Bourret *et al.* 2011). High density SNP-arrays represent a powerful tool to assess the impacts of escapes on wild populations (i.e. introgression), to trace the source of escapees and to characterize the current status of at-risk wild populations (Karlsson *et al.* 2011).

The first goal of this study is to describe the development and application of a dense SNP-array for Atlantic salmon. Secondly, we present the most genetically detailed population study of Atlantic salmon whereby 6176 SNPs were genotyped in 1430 individuals from 38 naturally occurring anadromous and landlocked populations collected across the species range. While the main emphasis concerns a description of the population genetic structure of European populations, several North American populations are also included in the assessment. Thirdly, we aim to identify markers and chromosomal regions potentially under divergent selection among populations between and within lineages as well as between anadromous and landlocked populations. Finally, we consider whether potential

adaptive divergence is associated with key biological pathways, and environmental patterns of variation.

3.4 Materials and Methods

3.4.1 Detection of SNPs in EST databases and Genome Complexity Reduction (GCR)

EST mining was performed according to Hayes *et al*., (2007), and using the EST libraries described in this paper, a total of 9240 putative SNPs were identified. Limited availability of DNA excluded the possibility of preparing a Reduced Representation Library (RRL; Altshuler *et al.* 2000); instead we describe an approach called Genome Complexity Reduction (GCR) which includes a PCR step to isolate a distributed sub-fraction of the genome. Genomic DNA was extracted from adipose-fin clips collected from *Salmo salar* using a commercial extraction kit. Samples were digested to completion using Xba1 which, compared to other 6-cutter restriction enzymes (RE), produced the most uniform smear of DNA fragments. Synthetic double-stranded DNA adapters containing a PCR binding site were ligated to the resulting sticky-ended fragments using T4DNA ligase. The resulting product was amplified under standard PCR reaction conditions using a 15 seconds extension time, which promotes amplification of fragments up to 1200bp. The combination RE cut frequency and amplicon length means that a random fraction of the genome is represented in the final product. Gel analysis of the PCR product revealed an expected smear of fragments and a single region of high intensity, most likely resulting from over-amplification of a repeated DNA element. Since repeated DNA can produce false-positive SNPs and consume sequencing capacity we developed an optional step (reduced GCR; rGCR) whereby PCR product is separated on a 1% agarose gel and a fragment range (for example 500-700bp), which avoids brighter (repeat) regions being isolated. The typically low DNA recovery requires an additional PCR amplification, which is performed as described above. For SNP discovery GCR libraries were prepared for individual haploid (n=2) and diploid (n=6) samples, in addition rGCR libraries were prepared for the haploid individuals (n=2) and a single pool sample containing all diploid samples (n=1). Haploid fish were produced by fertilizing eggs with irradiated sperm according to Refstie (1983). All biological material originated from Norwegian commercial aquaculture strains (Aqua Gen).

3.4.2 SNP discovery

The diploid and haploid GCR libraries were each sequenced twice, while the rGCR diploid pool and haploid samples were sequenced just once using a Genome Sequencer FLX system (Roche). A total of 6,927,968 reads were generated across 19 runs with an average read length of 182bp. Reads were filtered for repeats and adapter sequence using Lucy (http://lucy.ceh.uvic.ca/repeatmasker/cbr_repeatmasker.py), approximately 30% of reads were eliminated from the data set because they were identified as repeat sequences. Atlas overlapper (Havlak *et al.* 2004) was then used to group 1,404,933 reads into 53,769 bins (minimum bin size = 5 reads; average = 26, overlap criteria: 92% identity, minimum read length = 40bp, min overlap = 36bp, MaxOverlapSeed = 50). Each set of binned reads was processed by Phrap (Gordon *et al.* 1998) to generate 77,858 contigs with an N50 contig length of 349bp. Putative SNP discovery was done by aligning individual binned reads back to their matching contigs using Cross_Match (Gordon *et al.* 1998). Criteria for calling a SNP were a minimum read coverage of 2, and minimum SNP coverage/total coverage ≥ 0.2. The possibility of these SNPs (n = 17,844) being located within unknown (and therefore unmasked) repeats was addressed by blasting the 40bp flanking each SNP against the contig database. Sequences found to match with 100% sequence identity at any position other than their source were discarded. SNPs were then ranked into categories (n=3) according to the genotypes of the individual diploid samples sequenced. Category-1 corresponded to SNPs for which there was evidence of homozygous and heterozygous (i.e., AA, BB and AB) allelotypes, category-2 to those with occurrence of both homozygote allelotypes (AA and BB), and category-3 to those with heterozygote and one homozygote (AB and AA or BB).

Since category-1 SNPs and SNPs with higher read-depth displayed the greatest proportion of polymorphic loci (data not shown), these SNPs were preferentially selected for the array. SNP-assay designability was assessed using Illumina's online Assay Design Tool (ADT; www.illumina.com) with a required minimum ADT score of 0.7. The final array order included 7021 SNPs originating from GCR, 9240 SNPs from alignment of expressed sequence tag (EST) reads from both European and Canadian material, 58 SNPs from public databases, 169 SNPs detected through BAC-end sequencing (Lorenz *et al.* 2010) and 63

SNPs detected by mitochondrial DNA resequencing (Karlsson *et al.* 2010). More details regarding the array development strategy and SNP discovery can be found in supplementary methods (Appendix S1).

3.4.3 DNA samples

A total of 1431 extracted DNA samples sampled between 1977 and 2008 from 38 sample sites (31 from Europe and 7 from North America, between 20 to 72 samples per location) were included in this study (Figure 1; Table 1), thus covering the natural geographical distribution of Atlantic salmon. Five locations in Europe and one in North America represented landlocked populations while other populations were anadromous.

3.4.4 Genotyping and quality control

Genotyping was performed according to the manufacturer's instructions using the Illumina infinium assay (Illumina, San Diego, CA, USA). The assay conversion rate was 92% with the final array containing 15,225 SNP assays. Using Illumina's Genotyping Module software, it was possible to examine each SNPs cluster pattern using data from a set of pedigree samples (n=3297) provided by a Norwegian aquaculture program (Aqua Gen AS, Norway) and data from this study (n=1430). Visual inspection allowed for the classification of SNPs into different categories (i.e., single locus SNPs, and the more complex paralogous sequence variants; PSVs, and multisite variants; MSVs, arising from genome duplication). Classification was objectively supported using the program described by Gidskehaug *et al.* (2011) and is reported in supplementary methods of Lien *et al.* (2011). Following this (see also Results), a total of 6176 SNP markers were retained in the subsequent analyses of the 1431 population samples, 55% of these were derived from EST, 43% from GCR, and the remaining 2% from the other SNP sources. Since European fish were used in GCR and rGCR, we compared mean observed heterozygosity in both EST and GCR SNPs for European and North American populations to assess potential ascertainment bias. We also compared pairwise population measure of differentiation (F_{ST}) for both types of markers. Moreover, within each identified regional genetic groups (see below), we pooled the per population minor allele frequency distribution across loci and compared each regional pooled distribution to each other using the Kolmogorov-Smirnov test in order to further

evaluate potential bias in minor allele distribution. Note that we systematically excluded within population monomorphic markers from the distributions.

3.4.5 Population structure and differentiation

We measured global and per SNP observed and expected heterozygosity within each population. Pair-wise genetic differentiation between populations was estimated by θ (Weir & Cockerham 1984) using ARLEQUIN 3.5 (Excoffier & Lischer 2010) with 10,000 permutations to determine statistical significance. Based on previous mtDNA analysis (King *et al.* 2007) and intrinsic regional life-history characteristics (e.g. migration, winter feeding areas, etc.), we determined the regional clustering of populations to avoid confounding hierarchical stratification in subsequent analyses. Thus, regional grouping among European populations was assessed first by calculating pairwise genetic distances using the D_A distance (Nei 1977). The resulting genetic distance matrix was used to construct a Neighbor-Joining (N-J) phylogram and confidence estimates on tree topology were obtained by re-sampling over loci with 1000 bootstrap replicates. The genetic distance estimation and bootstrapping procedures were carried out using POWERMARKER (Liu & Muse 2005). Based on the N-J tree (Figure 2), European wild populations were clustered in three regional groups: Atlantic, Baltic Sea and Barents + White Seas (hereafter referred as: Atlantic, Baltic and Barents-White). The composition of each sub-continental group is presented in Table 1. To confirm the relevance of hierarchical groupings (intercontinental and between regions within Europe), analyses of molecular variance (AMOVAs) were performed using ARLEQUIN 3.5. Seven North American populations were included mainly for estimating the overall divergence between continents, these populations have been chosen to represent most of the seven regional genetic groups defined by Dionne *et al.* (2008) in Québec and extend geographic coverage. A principal component approach was also used to identify the axes of greatest genetic differentiation between all populations. The principal components were constructed by decomposing the genomic relationship matrix among all individuals where the genomic relationship matrix was defined as in Yang *et al.* (2010). Principal components were fitted using R (R core development team).

3.4.6 Outlier markers detection

To identify the most divergent markers among populations, which may potentially comprise loci under divergent selection (but see Bierne *et al.* 2011), we used hierarchical Fdist (Excoffier *et al.* 2009), a genome scan analysis implemented in ARLEQUIN 3.5 (Excoffier & Lischer 2010). The finite island model that was used in Fdist has been shown to lead to large fraction of false positives in some systems, therefore ARLEQUIN 3.5 proposes to use a hierarchical island model where migration rates among groups are different than migration rates among populations within groups. Thus, this hierarchical method can detect outlier loci among groups of populations (F_{CT}). In cases where no hierarchical structure was present, outliers were detected using F_{ST} and therefore the method was identical to the Fdist test of Beaumont & Nichols (1996).

Although regional groups were represented in North American samples, a regular Fdist was used since only one population per group was genotyped. Overall, five hierarchical tests were carried out, including an intercontinental comparison between anadromous European and North American populations, and one intra-Europe test contrasting the three regional groups defined (see Results section). Three additional hierarchical tests were performed to contrast landlocked and anadromous populations within each group (respectively: Atlantic anadromous: NUM, GAU and LAR; and landlocked: LL_NAM and LL_BYG; Baltic anadromous: KUN, VIN and TOR; and landlocked: LL_SYS and LL_PYA; and Barents anadromous: YAP, LEB, PON, EMT and SUM; and landlocked: LL_PIS) (See Table 1 for abbreviations definition). Hierarchical and non-hierarchical Fdist were run three times each in ARLEQUIN 3.5. Only loci detected as outliers at the significance level of 0.01 in all three runs were reported as potentially under the effect of selection. Outliers from each test were compared to examine possible parallelism between continents or between landlocked *vs* anadromous populations from different regions. Evidence for parallelism would strengthen support for interpreting putative outliers as being under the effect of divergent selection.

3.4.7 Candidate genomic regions affected by selection

As an alternative to single-locus outlier tests, for European populations, we combined Atlantic salmon genetic map information (Lien *et al.* 2011) with a kernel-smoothing moving average approach (Hohenlohe *et al.* 2010) to generate genome-wide distributions of the divergence estimates (measured F_{CT}) from hierarchical and non-hierarchical Fdist (Excoffier *et al.* 2009). Only SNPs mapped by Lien et al. (2011) were used in this analysis. To identify genomic regions with an unexpectedly high proportion of SNPs showing elevated or decreased divergence indicative of divergent and balancing selection, respectively, we performed 30,000-1,000,000 permutations to estimate local p-values. We tested multiple smoothing parameter values and selected 2cM bandwidth that identified relatively narrow genomic regions of interest while being large enough to reduce sampling variance. For linkage groups (LGs) with relatively low SNP coverage, larger bandwidth (SSA08 = 5cM; SSA26 = 4cM and SSA29 = 3cM) was used. Despite the fact that the choice of bandwidth can have a strong effect on kernel density estimation, different smoothing parameters did not change the position of the major peaks (data not shown).

3.4.8 Gene ontology and SNP annotation

Blast2go (Gotz *et al.* 2008) was used to associate gene ontology (GO) annotation terms to all 6176 SNPs. Homology searching was first realized through a BLAST search of the available flanking sequences for each SNP on the NCBI nr public database with the *e*-value threshold set to $1x10^{-10}$. Blast2go then retrieved GO terms associated with the obtained BLAST hits. The output GO annotation was then classified in multilevel biological processes, molecular functions and cellular components from the most general (level 2) terms of each category to more specific (upper levels) terms. In order to determine if the biological processes, molecular functions or cellular components of the outliers herein identified were over-, equally or under-represented when compared to the 6176 analysed SNPs, we performed an enrichment analysis using Fisher's Exact Test corrected for multiple tests by applying a false discovery rate of 0.05 (FDR) (Benjamini & Hochberg 1995).

3.4.9 Clinal variation among outliers

Since latitude is integrative of many correlated environmental variables along a South-North coastal line, we tested for the presence of latitudinal clines in allele frequency at outlier SNPs among European populations. Given that the allele frequency of many outlier markers (34 out of 52) showed significant correlations with latitude (see Results), we visually compared the cline pattern at those markers with the 18 non-significant ones. When excluding the Barents-White populations from the regression analysis, we observed that the latter markers presented more significant cline patterns driven by shifts in Baltic populations. Therefore, instead of using latitude, we tested for clinal allelic frequency distribution using generalized linear models of coastal distances along two different continua: 1) from northeastern populations to southern populations excluding the Baltic populations and 2) using one of the Baltic population as the point of origin and measuring the distance away from this population excluding populations from the Barents-White group. For each continuum, respectively called the White-Barents and Baltic continua hereafter, we compared the outliers' patterns with 52 randomly selected markers among the assumedly neutral (non-outlier) markers (excluding European monomorphic markers). Finally, we used CFIT-6 (Gay *et al.* 2008) to test whether the observed clines presented a common center or width. We thus compared four models: 1) no constraint, 2) center constrained, slope not constrained 3) center not constrained, slope constrained and 4) center and slope constrained. Finally, outliers were positioned using the genetic map information from Lien *et al*. (2011).

3.5 Results

3.5.1 Genotyping and quality control

After initial quality control and classification of genotypes obtained from 1431 samples (Table 1; Figure 1 for site locations) we classified i) 5436 markers out of 15,225 SNP features on the array as single locus and polymorphic SNPs (i.e. diploid SNPs), ii) 1725 markers as being represented at two homologous loci and polymorphic (i.e. multisite variants, MSV), iii) 930 markers being represented at two homologous loci but displaying no polymorphism (i.e. paralogous site variants, PSV), iv) 1853 failed assays and v) 5281 non-polymorphic markers (i to iv detailed in Lien *et al.* 2011). Mitochondrial SNPs were excluded from further analysis. Among diploid and MSV SNPs, 6112 markers showed an overall minor allele frequency (MAF) over 0.01. These, along with 64 other markers with an overall MAF < 0.01, but with a MAF > 0.05 in at least one population, made for a total of 6176 SNPs used in all analysis except stated otherwise. From the initial 1431 samples, 95 were discarded as they generated an average call rate < 0.85 (proportion of SNPs genotyped), which was chosen as a threshold differentiating acceptable and unacceptable genotyping data. Table 1 shows summary data for call rates (*CR*), observed heterozygosity (H_O) and expected heterozygosity (H_E) across populations. A significantly higher overall observed heterozygosity (Wilcoxon rank sum test P < 0.001) for European compared to North American populations (0.241 and 0.148 respectively) indicated an ascertainment bias towards European populations or alternatively a naturally lower diversity in North American populations. This was anticipated as a large proportion of the SNPs on the array were detected by sequencing European aquaculture strains. GCR derived markers were 2.4 times more heterozygous among European populations than North American (average observed heterozygosity = 0.290 vs 0.122). However, it should be noted that a few GCR markers were monomorphic across European populations. EST derived markers, which should be less affected by ascertainment bias, were also more heterozygous in European samples albeit with a smaller difference (0.203 vs. 0.168) (Table S1). Pairwise F_{ST} values were not significantly different between GCR and EST derived markers for comparisons of populations within Europe or within North American (respective Wilcoxon tests P: 0.640 and 0.881), but significantly different for intercontinental comparisons (mean F_{ST}: GCR = 0.463 and EST = 0.603; Wilcoxon P < 0.001). According to the Kolmogorov-Smirnov

tests, within region pooled population distribution of MAF showed that the Atlantic group had a significantly different distribution compared to all other groups ($P < 0.005$). The Baltic, North American and landlocked groups were similar to each other in MAF distribution. The Barents-White group presented a significantly different distribution from all other groups while intermediate between the Atlantic groups and the others (Figure 3).

3.5.2 Genetic diversity and population differentiation

Within population observed and expected heterozygosities varied across populations and markers (Tables 1 and S2). North American populations showed lower observed heterozygosity (range: 0.123 NAR to 0.167 MAP) than European populations (range: 0.198 KUN to 0.303 TAN), while landlocked populations demonstrated the lowest values (0.104 LL_BYG to 0.170 LL_SYS). All samples showed similar observed and expected heterozygosities except for the Spanish Narcea population (NAC), which had higher observed heterozygosity than expected, probably due to stocking of genetically divergent non-native fish from northern European rivers (Ayllon *et al.* 2006). All pair-wise comparisons of genetic differentiation between populations were highly significant ($P < 0.001$, Table S3). Overall, F_{ST} varied between 0.011 (NES *vs* DIO and NES *vs* TWE) and 0.758 (NAR *vs* LL_BYG). The within Europe comparison varied between 0.011 and 0.347 (CAR *vs* KUN) while within North America, F_{ST} varied between 0.048 (CHA *vs* MAP) and 0.151 (NAR *vs* STP). Intercontinental differentiations along with intra-Europe landlocked comparisons showed the highest F_{ST} with respective ranges of 0.439 (CHA *vs* TUL and MAP *vs* TUL) to 0.636 (CAR *vs* NAR) and 0.320 (LL_PYA *vs* LL_SYS) to 0.630 (LL_BYG *vs* LL_NAM).

An AMOVA considering both continents as regional groups and another carried out at the intra-Europe level, considering the clusters identified previously with the phylogram as regional groups (Figure 2) showed significant genetic variation among groups (Table 2). In the intercontinental analysis, 41% of variation was found among continental groupings (Europe and North America) while within Europe, 7.55% of variation was found among regional groupings (Atlantic, Baltic and Barents-White). In a principal component analysis (PC) of the genomic relationships among individuals, seven PC factors individually

determined at least 1% of variation, and together explained 55.1% of the total genetic variation among populations. Principal components 1 to 4 accounted for 38.4%, 9%, 2.1% and 1.7% respectively. PC1, PC2 and PC3 differentiated the European populations from the North American populations (Figure 4a) while PC3 separated the populations within Europe almost along a Northeast-Southwest axis, with Spanish populations at one end and most Russian populations and Baltic populations at the other end (Figure 4b).

3.5.3 Outlier markers detection

Six genome scans (Figure 5) were performed using 6176 SNPs with a varying number of markers potentially under divergent (range between 26 and 139) and balancing selection (range between 20 and 101) at the 0.01 significance level (Table S4). The hierarchical genome scan performed over all anadromous populations revealed the highest number of loci potentially under divergent selection (139) while also yielding the lowest number of markers potentially under balancing selection (20) in spite of an average F_{CT} across loci of 0.311 (ranging from -0.055 to 0.997). European (mean F_{CT} = 0.063; ranging from -0.052 to 0.732) and North American (mean F_{ST} = 0.081; ranging from -0.024 to 0.780) genome scans showed similar numbers of potentially selected loci with respectively 52 and 51 divergent outliers and 61 and 46 markers under balancing selection. Four outliers were common to both European and North American genome scans. Three showed similar trends as two were detected as potentially under divergent selection (ESTNV_28701_207 and ESTNV_23580_687) and one under balancing selection (ESTNV_28516_389). However, one showed opposite trends (ESTNV_32552_113) as it was potentially under divergent selection in the European scan and under balancing selection in the North American scan.

All three hierarchical genome scans also revealed SNPs under potential divergent and balancing selection when anadromous populations were compared with landlocked populations from the same geographic region (respectively N divergent = 40, 49 and 40; N balancing = 31, 41 and 38; mean F_{CT} = 0.057, 0.060 and 0.100; maximal values 0.876, 0.883 and 0.983). Three outliers were common to landlocked *vs* anadromous genome scans. Marker ESTNV_33891_846 was found to be under balancing selection among Barents-White and Baltic populations, GCR_cBin31530_Ctg1_84 under balancing selection among

Baltic populations while under divergent selection in the Atlantic populations and GCR_cBin14325_Ctg1_429 was under divergent selection in Barents-White and Atlantic populations. None were common to all three genome scans.

3.5.4 Candidate genomic regions affected by selection

To further gain insight into the chromosomal patterns of divergence (measured as F_{CT}) along the mapped linkage groups, we identified genomic regions showing unexpectedly high or low divergence between different European regional groups. Altogether, 19 genomic regions showed elevated differentiation ($P < 0.01$) in European comparisons (Figure 6a; Table S5). A slightly larger number of genomic regions (n=25) exhibited reduced levels of differentiation ($P < 0.01$) (Figure 6a; Table S5). When compared with the single-locus outlier tests, both analyses often revealed similar candidate regions potentially under divergent selection (Figures S1-S29 e.g.: LG: SSA01, SSA05, SSA10, SSA12, SSA15-SSA17, SSA18, SSA20). The strongest evidence ($P < 10^{-5}$) for elevated or reduced differentiation was found in LG SSA08 and SSA16 (Figure 6bc).

3.5.5 Annotation of outlier SNPs

The BLAST (Altschul *et al.* 1990) yielded 2691 SNPs with significant hits (*e*-value < $1x10^{-10}$, Table S6). From these results, a total of 17,701 gene ontology (GO) terms were associated with SNPs. GO terms segregated into many levels of biological processes, molecular functions and cellular components with a mean level of 5.819. Overall, 19 level 2 biological pathways associated with cellular process, metabolic process and biological regulation were highlighted, and together these represented 48% of GO annotations. After correction for multiple testing, an enrichment analysis across both categories of outliers (divergent and balancing) did not indicate significant over- or under-representation of any biological pathway in the outliers identified. Among common outliers in European *vs* North American comparison, ESTNV_28701_207 (potentially under divergent selection) was associated with a b-cell receptor (CD22-like), which prevents over-activation of the immune system (Hatta *et al.* 1999), while marker ESTNV_32552_113 (divergent in Europe, balancing in North America) has a close relationship with a calcium ion binding protein. Common outliers among anadromous and landlocked populations revealed one

marker, ESTNV_33891_846 (under balancing selection in Barents-White and Baltic), which has an association with coagulation factor v.

3.5.6 Clinal variation among outliers

Among the 52 outliers potentially under divergent selection among all anadromous European populations, the population allelic frequencies of 34 markers were significantly correlated with latitude (R^2 ranging from 0.177 to 0.498, mean = 0.308). The generalized linear models applied to each of two different continuums, one starting in White Sea (excluding the Baltic) and the other in the Baltic Sea, revealed that 23 and 18 outliers present only in the White-Barents and Baltic continua respectively, and 11 outliers showing clinal patterns common to both continua. Therefore, all 52 markers showed a clinal pattern for one or both continua. The congruence of outlier markers was striking when contrasted with randomly selected neutral markers (Figure 7). In both continua, the model with no constraint was always the best one (Table S7), meaning that there was no common center or width associated with observed clines. However, the average localizations of the centers were 1092 km and 1009 km from the point of origin for the White Sea and Baltic continua respectively. These averages closely coincide with the boundaries between the Barents Sea *vs* Atlantic, and Baltic Sea *vs* Atlantic. This also closely matches the genetic boundaries inferred from the three genetic groupings defined earlier. Finally, the outliers showed a random distribution on the genetic map with no apparent clustering on specific linkage groups (Figure 8).

3.6 Discussion

This study represents the single most extensive population genetic study on Atlantic salmon performed to date. With 1360 individuals successfully genotyped for over 6000 SNP markers, it also stands as one of the most comprehensive population genetics studies of wild populations in a non-model species. An important outcome of our study, particularly for population management, is the robust confirmation of three major regional genetic groups occurring in Europe. Moreover, based on strong clinal patterns observed for identified outlier markers, we propose that these groups are bordered by secondary contact zones where highly divergent markers are associated with endogenous (intrinsic genetic incompatibilities) and possibly exogenous (environmental or ecological) barriers (Bierne *et al.* 2011).

3.6.1 Ascertainment bias

Among the 6176 SNPs retained on the SNP-array, 3383 were detected from available EST reads from both European and Canadian material. On the other hand, the GCR markers were discovered using only Norwegian commercial aquaculture fish meaning that these markers are potentially more prone to an eastern Atlantic European bias when used for genotyping more geographically dispersed samples. The sampling bias towards the Atlantic populations and to a lesser extent in the Barents-White groups could potentially bias our differentiation estimates. Indeed, Albrechtsen *et al.* (2010) observed a small upward bias in F_{ST} estimates using ascertained populations. They also showed that the intensity of the bias depends on the genetic distance of the populations being compared relative to the population within which the markers were originally developed, suggesting that F_{ST}, when two populations outside of the ascertained regional group are compared to each other, may be less affected by ascertainment bias.

A significantly higher genetic diversity (H_O) in European samples was observed for both EST and GCR based markers (Table S1). Contrasting estimates of genetic diversity between the continents are partly explained by the ascertainment bias, and are of concern for estimating and interpreting other genetic parameters. Indeed, we found that inter-continental population differentiations are significantly higher when measured with EST

based markers than with GCR markers. Therefore, this bias should be taken in consideration in any inter-continental comparisons using the array. That said, it is noteworthy that in a survey of microsatellite diversity, King *et al.* (2001) also found that genetic diversity was reduced in North American populations compared with European populations using markers developed with material from both origins. It is thus possible that different demographic history partly explains the overall differences in genetic diversity between continents.

In contrast, no significant difference in intra-continental population differentiation was observed when estimates were calculated with EST and GCR markers separately (data not shown). However, within Europe, genetic diversity is also regionally heterogeneous with lower genetic diversity associated with Baltic and landlocked populations. Here again, this regional pattern of variation was observed in previous studies using microsatellites and was attributed to the phylogeographic history of these populations (Säisä *et al.* 2005; Tonteri *et al.* 2007). Overall then, it appears that variation in genetic diversity between regions results from both historical contingency and ascertainment bias. We thus advise that absolute values in inter-continental comparisons should be interpreted carefully. However, the bias effect is less of a concern for within continental comparisons, especially for interpreting broad scale patterns of differentiation.

3.6.2 Population structure

While Atlantic salmon population structure has been extensively investigated (Ståhl 1987; King *et al.* 2001; Verspoor *et al.* 2005; Dionne *et al.* 2008), this study is the first to examine the structure of population over its entire range with extensive genomic coverage in a single original study. Overall, the results provided by the SNP markers show similar patterns as those previously observed, but the level of differentiation is amplified with F_{ST} values up to twice as high as previously reported, maybe due to the lower level of polymorphism in SNPs compared to microsatellites (Hedrick 1999). As an example, a within continent comparison of pairwise F_{ST} , Dionne *et al.* (2008) reported a F_{ST} value of 0.048 based on 12 microsatellites, compared to the value of 0.112 we observed with SNPs for the same comparisons. This enhanced structure definition is observed on every scale

examined. Indeed, the range-wide analysis suggested that 41% of the genetic variation occurred between the continents, which contrasts with the 21.9% reported by King *et al.* (2001). Although we did not observe any continent-specific alleles as previously observed for microsatellite markers (King *et al.* 2001; Wennevik *et al.* 2004), we did identify over 100 SNP markers with F_{CT} greater than 0.95, i.e. nearly diagnostic between continents.

Following the pioneering study of Ståhl (1987) who first identified two distinct genetic groups of Atlantic salmon within Europe, namely Eastern Atlantic and Baltic Sea, many studies subsequently focused on the colonization history of northern Europe following the last ice age and its impact on regional genetic groupings of Atlantic salmon populations (Kazakov & Titov 1991; Skaala *et al.* 1998; Koljonen 2001; Wennevik *et al.* 2004; Makhrov *et al.* 2005; Tonteri *et al.* 2005; Säisä et al. 2005). Most of these studies were congruent in differentiating the Baltic Sea populations and regrouping other populations into an Atlantic + Barents Sea + White Sea group. Verspoor *et al.* (2005) reported that more groups could be defined in Europe, especially within the previously named Atlantic group. More recently, Tonteri *et al.* (2009) demonstrated the genetic distinctiveness of the Eastern Barents Sea and White Sea populations from the Western Barents and Eastern Atlantic ones. Our results confirm this partitioning of northern populations and further suggest the contribution of at least three glacial refugia in colonizing European populations. Essentially, we observed three major clusters corresponding to an Atlantic group, a Baltic Sea group and a Barents-White Seas group. A relatively high proportion (7.55%) of the genetic variation occurred among these groups although a higher proportion (10.4%) was also attributed to differences among populations within each of them. These three groups are consistent with the main postglacial colonization routes previously identified as the West Atlantic, the Baltic Ice Lake and Eastern Barents Sea (Tonteri *et al.* 2007; 2009; and references therein).

Landlocked populations exhibited a structure also consistent with previous studies. However, this study allowed a finer understanding regarding the origins of these populations from anadromous ancestors. Tonteri *et al.* (2005) suggested that freshwater populations from the Baltic and White Sea basins originated from different populations,

explaining their segregation into two different clusters based on 14 microsatellite markers. Here, we observed roughly 1.5 times more differentiation when contrasting landlocked populations from these two basins with any anadromous populations of their respective basins. Additionally, freshwater Baltic populations clearly clustered with anadromous Baltic populations while the non-anadromous population from the White Sea basin clustered with the Barents-White group. Based on their geographical location, Norwegian populations were expected to cluster as Atlantic populations, but showed a high degree of divergence from all three regional genetic groups. Since their closest neighbors on the tree are Norwegian or Western European populations, it is highly possible that their differentiation from the Atlantic group stems from a very pronounced drift effect resulting from a founder event, which could have occurred during land upheaval following the last deglaciation in Norway (Berg 1985).

3.6.3 Signatures of selection

The detection of loci under selection using different genetic differentiation methods share several known caveats, particularly the potential for false positives with such a high degree of multiple testing (Foll & Gaggiotti 2008). Nonetheless, these methods have been successfully used for identifying candidate genes and QTL for local adaptation to environmental conditions (Schmidt *et al.* 2008; Storz & Wheat 2010; Gagnaire *et al.* 2012b), localizing genomic regions under adaptive divergence (Rogers & Bernatchez 2007; Flori *et al.* 2009; Hohenlohe *et al.* 2010; Gagnaire *et al.* 2012a), or identifying markers to investigate introgression (Karlsson *et al.* 2011; Hohenlohe *et al.* 2011; Lamaze *et al*. 2012). While a main objective of this study was to assess the extent of potential adaptive divergence among populations throughout the native range of Atlantic salmon, we were not able, in some comparisons, to distinguish the effect of selection from that of pronounced drift. In other comparisons, using multiple populations subject to the same putative selection reduced the chance that the outliers identified were mainly the result of drift.

Markers identified as outliers when contrasting genetic variation on both continents genetic compositions were almost differentially fixed within each continent. Given the estimated time of 600,000 years since continental inter-divergence (King *et al.* 2007) and very modest

contemporary migration rates between the two continental lineages, we were expecting many more strictly differentially fixed markers than were observed. Thus, the absence of fixed markers could mean that the migration rate between both lineages was previously underestimated. However, we cannot exclude that ascertainment bias, the method of SNP selection and the unbalanced coverage of genetic variation towards Europe might also contribute to this pattern.

Although environmental gradients occur along the geographical distribution of Atlantic salmon, contrasting the three major clusters identified failed to isolate markers that could be strictly under the influence of environmental selection. Instead, our results revealed that the majority of outlier markers showed allele frequencies that correlated with geographical clines (compared to non-outlier markers) in both the Baltic *vs* Atlantic and Barents-White *vs* Atlantic comparisons. This could be a result of historical colonization and persisting tension zones following secondary contact between once geographically isolated lineages. We did not identify a common center or width for these clinal outliers (Table S7). However, all markers showing clinal variation were strictly outliers, they showed an overall congruence in geographic patterns of variation (Figure 7), and that they were randomly distributed in the genome (Figure 6 & 8). This strongly suggests that these markers are likely to reveal endogenous genetic barriers (genetic incompatibilities) occurring in zones of secondary contact between distinct evolutionary lineages, rather than genetic-environment associations (exogenous barriers). As demonstrated by Bierne *et al.* (2011), endogenous genetic barriers can easily be coupled with exogenous genetic barriers associated with environmental gradients. Here, environmentally driven selection could be responsible for some of the observed clines since environmental heterogeneity, mainly in terms of temperature and salinity also occurs along these zones. However, our data and analysis do not allow us to tease apart markers potentially under the influence of environmental selection *vs* others. For instance, the Öresund and Danish Belts delimitate an important environmental discontinuity between the Baltic Sea and the Atlantic Coast characterized by an abrupt change in salinity (among other parameters) that have been identified as a selective agent in many species (Johannesson & Andre 2006; Gaggiotti *et al.* 2009; Limborg *et al.* 2009). Here, we found that this geographical region is associated with

numerous outlier markers showing clinal variation. We do not challenge the environmental selection acting in this area and certainly acknowledge previous studies identifying this possible hotspot for genetic-environment associations. However, our results suggest that careful interpretation of outlier markers is needed, especially in known secondary contact zones. Namely, as it is likely the case for the Baltic, hybrid zone theory predicts that tension zones will be trapped by natural barriers (Barton 1979; Barton & Hewitt 1985; Hewitt 1988; Bierne *et al.* 2011). Should this be the case in the Barents Sea as well, characterization of the putative environmental barrier operating there might reveal possible functional targets of selection. Overall, we emphasize that identifying genomic regions underlying a true genetic-environment association first needs functional support and/or convincing rejection of historical contingency on top of detailed environmental characterization. Therefore, in the case of Atlantic salmon, we suggest that future research should focus on disentangling the interplay of historical and selective forces.

As for the North American populations analyzed in this study, only cautious interpretations can be drawn given that results are still equivocal. Previous studies of Western Atlantic salmon provided evidence for a hierarchical genetic structure. However, depending on markers used and geographical coverage, defined regional groups differed among studies (King *et al.* 2007 and references therein; Dionne *et al.* 2008). Interestingly, there is evidence that North America was colonized by at least two major refuges given the presence of European mtDNA haplotypes in many landlocked populations (Knox *et al.* 2002; King *et al.* 2007). Nonetheless, the regional structure of anadromous populations was never associated with post-glacial colonization from multiple source populations, as in Europe. Instead, recent landscape genetics studies have shown that regional differences are likely driven by distinct environmental conditions, which would imply a regional scale of local adaptation (Dionne *et al.* 2007; 2008), which may be more common in anadromous salmonids than previously thought (Fraser *et al.* 2011). In this context, we argue that identified outliers could represent or be associated with actual targets of environmental selection. However the limited coverage of North American populations hampered the possibility to pinpoint any specific biological function or process among annotated outliers that could have established potentially functional targets of selection and thus, useful

candidate genes. Clearly, further investigation of adaptive divergence among North American populations will necessitate deeper sampling coverage within and among regional groups along with a detailed landscape genomics approach.

Contrasting landlocked versus anadromous populations in Europe identified markers that were strikingly differentiated between the two life history strategies. Such differences may indicate ecological selection and/or random genetic difference due to genetic drift. None of the outliers that were detected showed parallel patterns of divergence among the three regional genome scans we performed. However, we found three markers identified as outliers in two out of three scans. Although all three showed different patterns (divergent selection in one group, balancing selection in opposite directions in the other groups), and only one was successfully annotated, these exploratory scans suggest that there is sufficient divergence between landlocked and anadromous populations to justify further examination. We propose three main steps to undertake in a follow-up study that would aim specifically at elucidating the adaptive divergence between these populations. First, comparing regional replicates holds the key to distinguishing selection from drift. Second, a less stringent significance threshold could be used to detect outliers in order to allow more markers to be further examined and screened using ecological and functional context. Finally, particular attention should be paid to candidate genes involved in osmoregulation and immune related functions that are possible targets for selection given the contrasting environmental conditions encountered by anadromous and landlocked populations. The upcoming genome sequence (Davidson *et al.* 2010) of Atlantic salmon promises to improve the functional context and further contribute to disentangling historical from adaptive divergence.

3.6.4 Implications and perspectives

In addition to the recent demographic decline of most anadromous populations, related concerns such as the relative genetic contribution of individual populations to fisheries or bycatch, translocation, reintroduction strategies and assessments of farmed escapees impacts are all management issues that could benefit from the improved genetic information derived from the SNP-array. Information from the array allows finer, more precise definition of populations, management and evolutionary significant units. In this

study, using the SNP-array certainly enhanced resolution of such units by revealing increased level of differentiation estimates at every geographical scale relative to previous studies. Our study suggests that escapees from aquaculture (the majority of which stem from Norwegian aquaculture populations of the Atlantic group in Europe) could lead to different genetic consequences whether introgressing into wild populations of the same (Atlantic) or different (Barents-White) phylogeographical groupings. The array should contribute in resolving the continuing controversy surrounding the potential impacts and level of introgression of captive bred fish in wild populations. For instance, the SNP-array developed here has recently been used by Karlsson *et al.* (2011) to develop a sub-panel of markers that can discriminate between wild populations and the major strains of domestic Atlantic salmon used for farming in Europe, enabling the assessment of the impacts of farmed escapees on wild populations. Furthermore, linking genetic and environmental divergence could highlight biological processes that evolve under the effect of natural selection and identify the actual selective agents. Although further investigation is needed to delineate the relative contribution of adaptive divergence in observed patterns of genetic differentiation, the ubiquitous occurrence of highly divergent markers representing a diversity of biological functions, holds the potential for adding information about the adaptive nature of divergence in defining significant units of management and conservation. Being aware of historical contingency effect on observed patterns of differentiation, we are now better suited to grasp the real contribution of environmental selection in shaping population divergence in Atlantic salmon. Overall then, this SNP-array and subsequent versions of it should bring considerable benefits to Atlantic salmon management and conservation community as well as for aquaculture applications.

3.7 Acknowledgements

We are grateful to Eva Garcia-Vasquez, Jamie Stevens, Sigurdur Gudjonsson, Mélanie Dionne, Alexei Veselov, Jaakko Lumme, Jan Nilsson and Jaakko Erkinaro for providing samples. The SNP discovery, array development and genotyping were performed by CIGENE at the national technology platform, supported by the functional genomics programme (FUGE) in the Research Council of Norway. We also thank P.A. Gagnaire, Gonzalo Machado-Schiaffino and M. Bruneaux for their very constructive inputs in the signature of selection analysis and interpretation. A special thanks to Ben J. Hayes for his regular inputs to the manuscript and analyses. Research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to VB, from the Canadian Research Chair in genomics and conservation of aquatic resources to LB, the Academy of Finland to CRP and AV, Estonian Science Foundation to AV and the Beaufort Marine Research Award in Fish Population Genetics funded by the Irish Government under the Sea Change Programme to PMcG.

3.8 Tables

Table 3.1 Description of regional groupings and parameters associated with sample sites composing the groups: latitude and longitude, number of individuals genotyped (NGEN), number of individuals with call rate superior to 0.85 (N>85), average call rate per population (CR) and average expected (*HE*) and observed (*HO*) heterozygosities per population.

Regional Groups			Population ID	Code	Number	Years	Latitude	Longitude	N_{GEN}	$N_{>85}$	CR	H_E	H_O
Anadromous	*Europe*	*Barents-White*	Tuloma	TUL	1	2000	68,88	33,00	40	40	0.980	0.290	0.286
			Tana	TAN	2	1989 & 1995	70,50	28,38	55	29	0.929	0.296	0.303
			Emtsa	EMT	3	2001	64,57	40,48	40	40	0.986	0.255	0.249
			Lebyazhya	LEB	4	2001	66,95	41,28	40	40	0.982	0.275	0.275
			Yapoma	YAP	5	2000	66,27	36,85	40	40	0.985	0.263	0.261
			Pongoma	PON	6	1999	65,00	34,63	40	40	0.984	0.253	0.257
			Suma	SUM	7	2001	64,28	35,40	40	40	0.985	0.219	0.230
		Atlantic	Numedalslågen	NUM	8	1989	59,04	10,05	50	43	0.934	0.294	0.302
			Gaula	GAU	9	1989-90	63,34	10,23	43	43	0.964	0.303	0.302
			Lærdalselva	LAR	10	1977-97	61,10	7,47	72	25	0.968	0.297	0.298
			Dart	DAR	11	2005	50,40	-3,63	40	35	0.978	0.285	0.284
			Blackwater	BLW	12	2006	51,94	-7,84	40	40	0.989	0.287	0.289
			Dionard	DIO	13	2006	58,52	-4,80	40	40	0.984	0.287	0.287
			Foyle	FOY	14	2006	55,05	-7,25	40	40	0.990	0.276	0.273
			Moy	MOY	15	2004	54,10	-9,15	40	40	0.988	0.276	0.274
			North Esk	NES	16	2005	56,75	-2,42	40	40	0.989	0.288	0.289
			Tweed	TWE	17	1996	55,75	-1,98	40	37	0.981	0.284	0.283
			Loire	LOI	18	2006-07	47,27	-2,18	40	39	0.985	0.238	0.239
			Cares	CAR	19	2004	43,39	-4,51	40	40	0.982	0.199	0.268
			Piguena	PIG	20	2004	43,57	-6,08	20	20	0.983	0.238	0.241
			Narcea	NAC	21	2004	43,57	-6,08	20	20	0.970	0.219	0.265
			Selá	SEL	22	2004-05	65,83	-14,81	32	32	0.988	0.236	0.238
			Ölfusá	OLF	23	2004-05	63,93	-21,23	32	32	0.991	0.226	0.228
		Baltic	Tornionjoki	TOR	24	1997	65,81	24,15	40	40	0.988	0.215	0.213
			Vindelälven	VIN	25	2005	63,75	20,32	40	40	0.984	0.212	0.211
			Kunda	KUN	26	2005	59,52	26,53	40	40	0.990	0.195	0.198
	North America	*Anticosti*	Chaloupe	CHA	27	2004	49,14	-62,54	23	23	0.971	0.166	0.165
		Southern Québec	Matapédia	MAP	28	2004	47,97	-66,93	25	25	0.967	0.170	0.167
		Labrador	Saint-Paul	STP	29	2004	51,45	-57,70	25	25	0.982	0.155	0.151
		Lower North Shore	Gros Mécatina	MEC	30	2004	50,77	-59,09	25	25	0.981	0.139	0.138
		Québec city	Du Gouffre	DGO	31	2004	47,43	-70,48	25	25	0.982	0.148	0.148
		USA	Narraguagus	NAR	32	2001-04	44,60	-67,92	25	25	0.978	0.123	0.123
Landlocked	*Europe*	*Baltic*	Sysky	LL_SYS	33	1999	61,65	31,27	32	32	0.992	0.165	0.170
			Pyalma	LL_PYA	34	2004	62,40	35,87	40	40	0.983	0.139	0.141
		Barents-White	Pisto	LL_PIS	35	1999	65,26	30,56	40	40	0.988	0.144	0.148
		Atlantic	Namsen-Småblank	LL_NAM	36	2005-08	64,46	11,52	46	40	0.985	0.119	0.113
			Otra-Byglandsbleka	LL_BYG	37	2004	58,14	8,01	40	40	0.988	0.103	0.104
	North America	*Newfoundland*	Bristol Cove	LL_BCR	38	1997	46,63	-53,19	41	40	0.975	0.145	0.146

Table 3.2 Analysis of Molecular Variance (AMOVA) in two hierarchical groupings: a) groups defined as North America and Europe and b) groups defined as regional groups intra-Europe as identified in Table 1. *P-value < 0.001

a) Continental groups		
Source of variation	**df**	**Percentage of variation**
Among Groups	1	41.01*
Among populations within groups	30	8.84*
Within populations	2174	50.15*
b) European groups		
Source of variation	**df**	**Percentage of variation**
Among Groups	2	7.55*
Among populations within groups	23	10.39*
Within populations	1884	82.06*

3.9 Figures

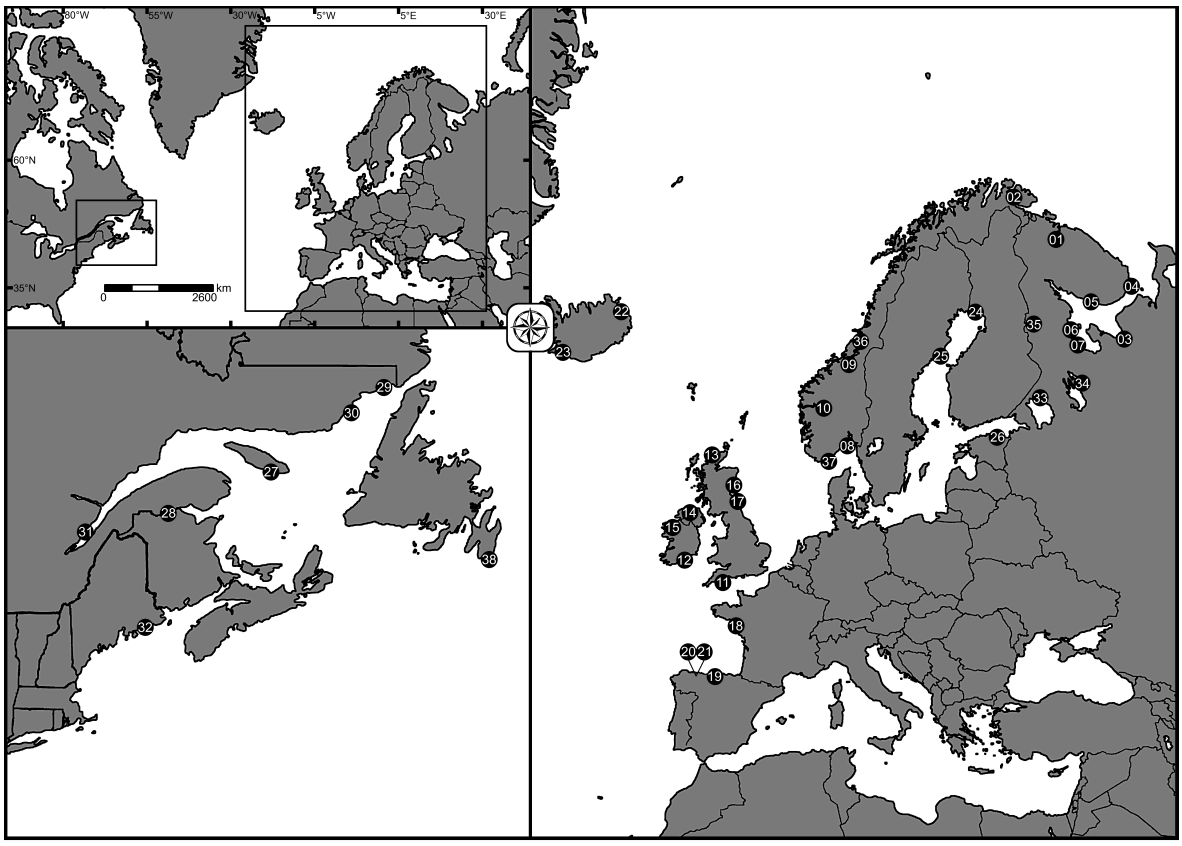


Figure 3.1 Map showing sample sites in Europe and North America. Populations are linked to the numbers in Table 1.

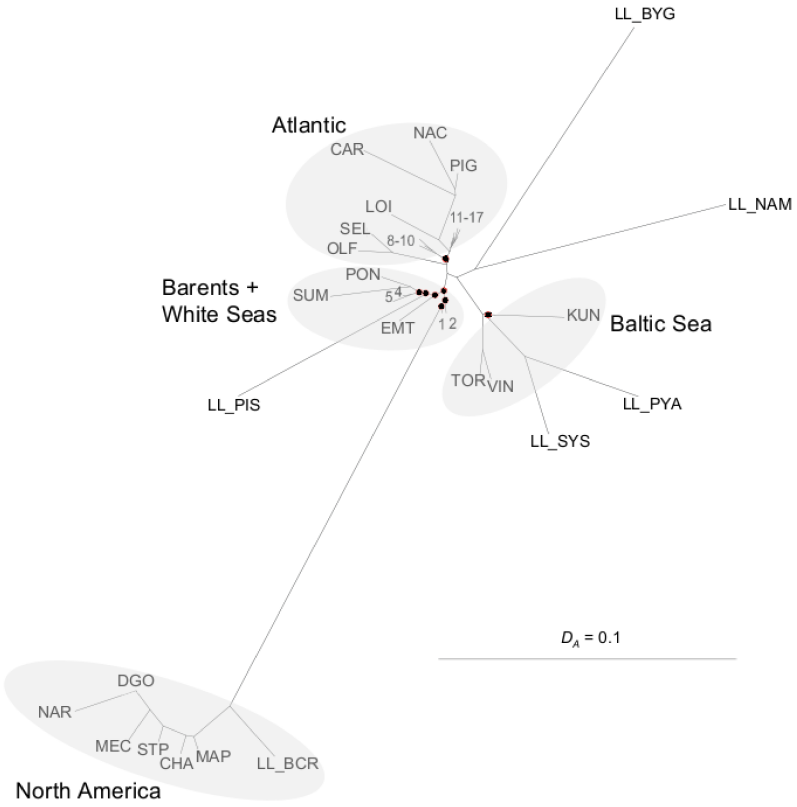


Figure 3.2 Genetic relationships in Atlantic salmon as resolved by Neighbor-Joining tree constructed using *DA* distance. Population code and numbers are as in Table 1 (populations with LL are landlocked). Nodes marked with a dot were supported by bootstrap support of less than 70% of 1000 replicates and others are supported by more than 70% of 1000 replicates.

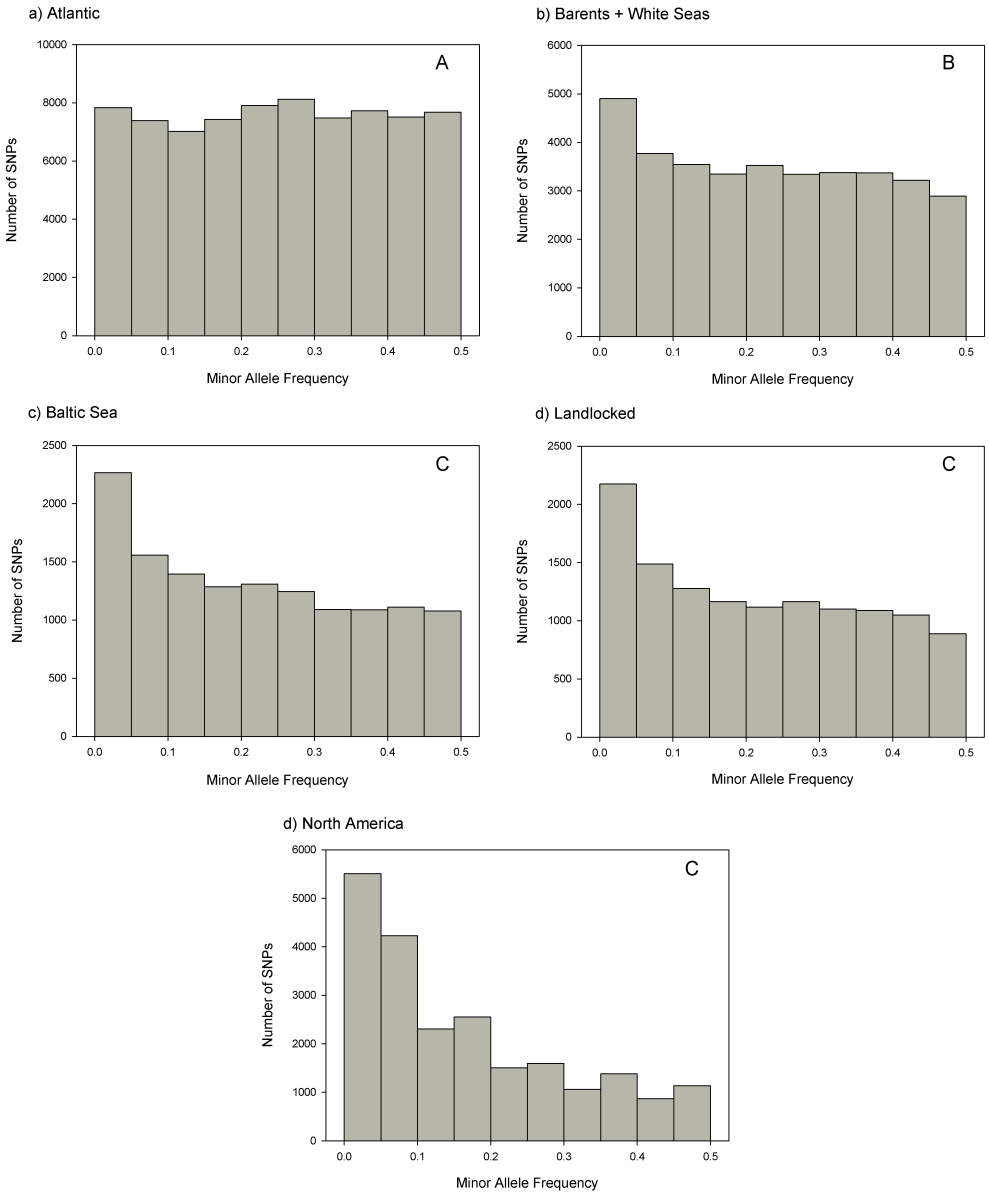


Figure 3.3 Within region pooled population distribution of minor allele frequency (MAF) for: a) Atlantic populations, b) Barents-White populations, c) Baltic populations, d) North America populations and e) landlocked populations. Panels sharing letters have distributions not significantly different according to the Kolmogorov-Smirnov test. Markers with MAF = 0 are not shown.

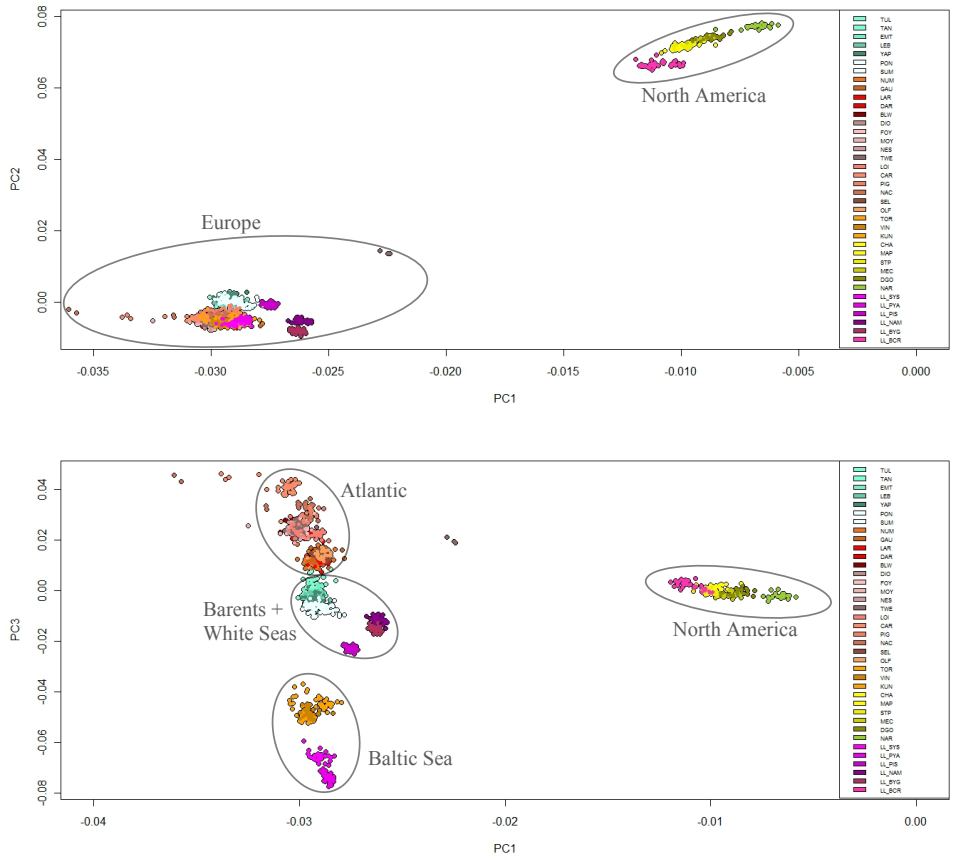


Figure 3.4 Principal components analysis of genetic differentiation among populations based on 6176 SNP markers (each point represents one individual) with: a) principal component 1 (PC1: 38.4% of variance) against PC2 (9.0% of variance); b) PC1 against PC3 (2.1% of variance). Each population shows a different color and color grading relates in part to geographical similarity.

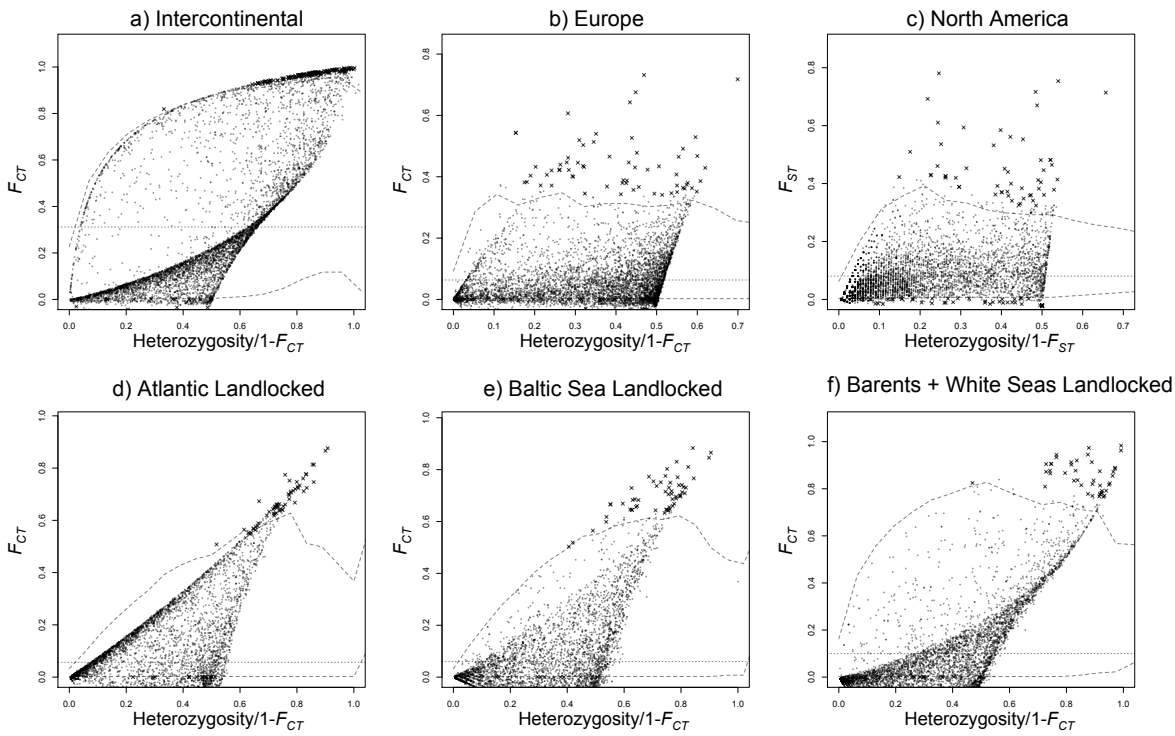
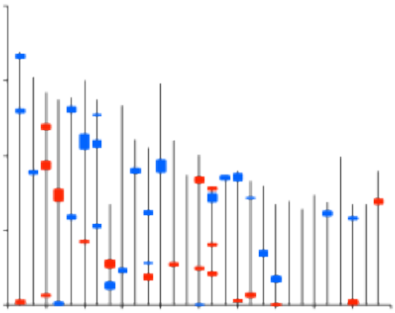
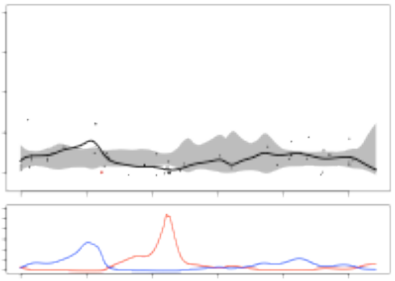


Figure 3.5 Differentiation (F_{CT} or F_{ST}) as a function of *heterozygosity/1-differentiation* as calculated by ARLEQUIN 3.5 for 6 genome scans : a) comparison between European and North American anadromous populations (hierarchical test), b) comparison between 3 regional groups (hierarchical test), c) comparison among North American anadromous populations (non-hierarchical test), d-f) comparison of landlocked *vs* anadromous populations within each of the three European groupings (Altantic, Baltic, Barents-White) (hierarchical test). In each panel, outliers markers ($P < 0.01$) are marked by X, dashed lines represent upper and lower 99% confidence level and dotted line indicates the average F_{CT} or F_{ST} across loci.

a)



b)



c)

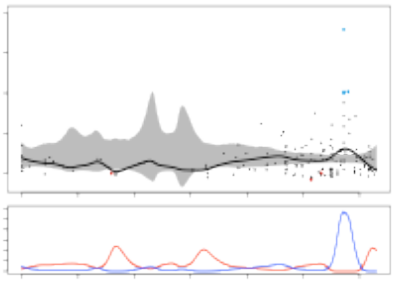


Figure 3.6 Genetic linkage map showing the distribution of genomic regions showing elevated or decreased divergence ($P < 0.01$) marked as blue and red bars, respectively in a) European populations. The distribution of F_{CT} along the linkage groups: b) SSA08 and c) SSA16. Grey areas correspond to 95% confidence interval of the smoothed kernel curve (black line). Individual outlier loci showing elevated or decreased divergence are marked as blue and red bars. Estimated significance levels (–log10 transformed p-value) for region of elevated or decreased divergence are shown as blue and red curves.

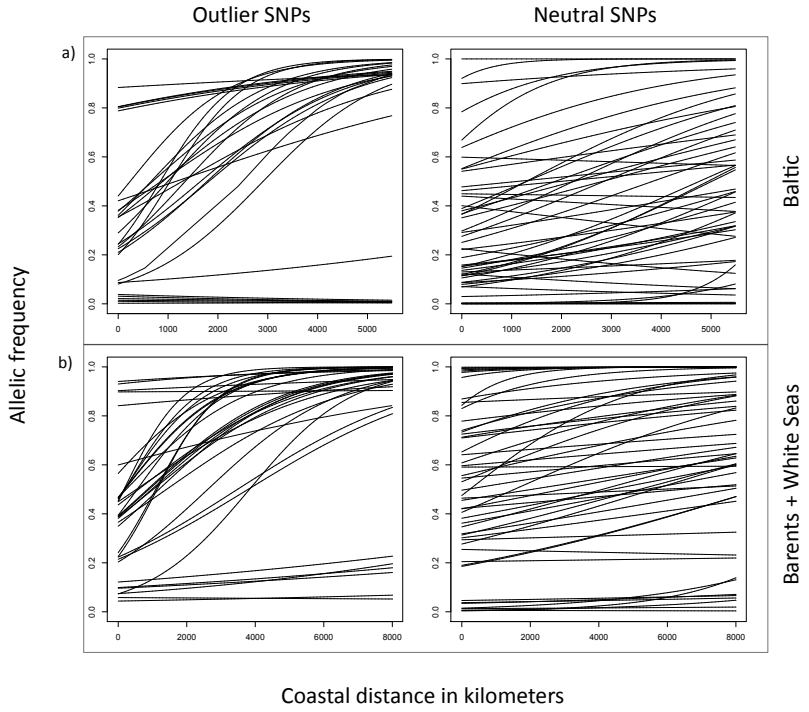


Figure 3.7 Generalized linear models illustrating geographical clines (or lack thereof) for 52 European outliers (left panels) and 52 randomly selected neutral (non-outlier) markers (right panels) shown along two different coastal distance continuum: a) the Baltic and b) the White-Barents.

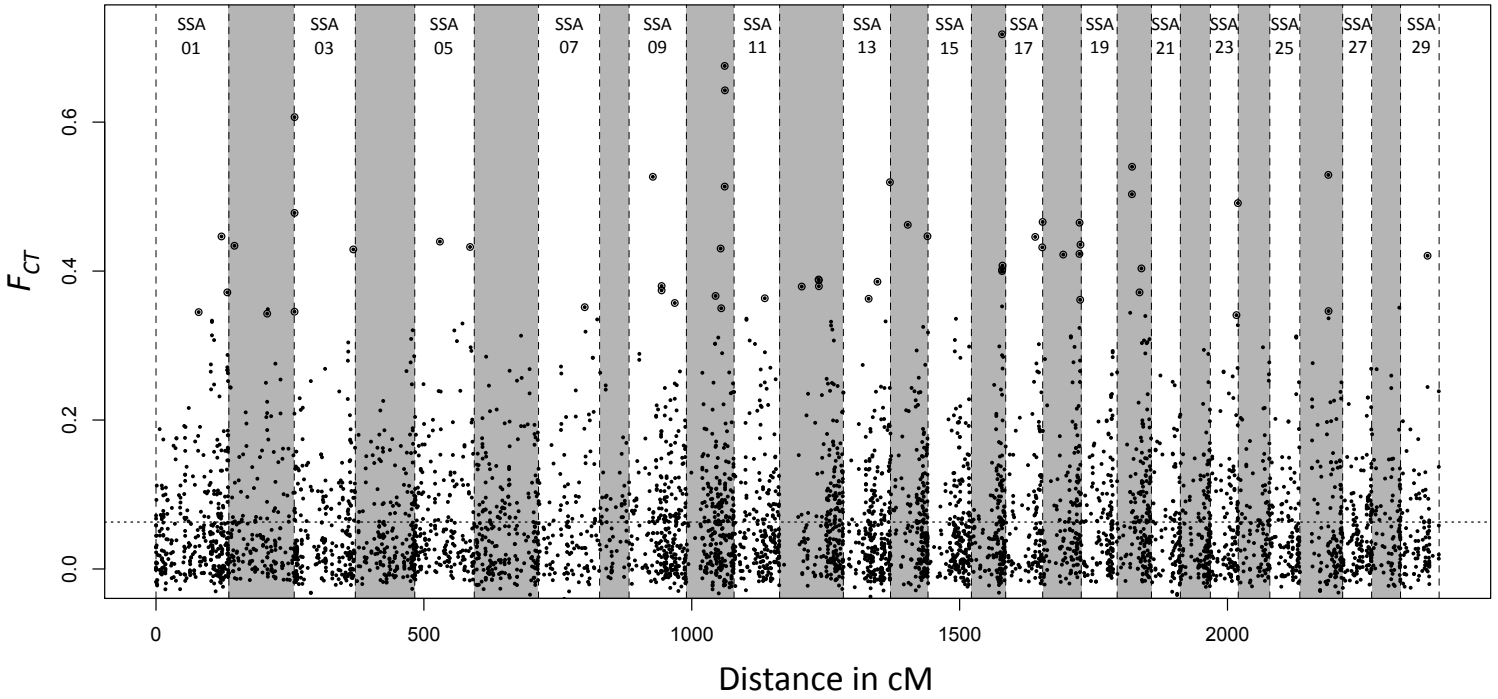


Figure 3.8 European regional differentiation (F_{CT}) of each SNP marker mapped on the European Atlantic salmon genetic map. Gray and white rectangles separated by vertical dashed lines represent separate linkage groups (named SSA--). Large black dots indicate outlier markers and the horizontal dotted line indicates the average F_{CT} among markers (0.063).

3.10 Supplementary material

Hereafter listed supplementary methods, tables & figures are available online or on demand:

Methods: *Strategy for producing a SNP array in Atlantic Salmon*
Chip development and SNP discovery

Table 3.S1 Observed heterozygosity (H_O) among GCR and EST based SNPs in both continents.

Table 3.S2 Single nucleotide polymorphism (SNPs) markers observed (H_O) and expected (H_E) heterozygosities per population.

Table 3.S3 Pairwise measures of genetic differentiation (F_{ST}).

Table 3.S4 Summary of the detection of markers potentially under selection following Fdist (North America) and hierarchical Fdist (Intercontinental, Europe, Russian Landlocked and Norwegian landlocked) genome scans implemented in arlequin 3.5 (Excoffier & Lischer 2010).

Table 3.S5 Summary of the detection of genomic regions showing elevated or decreased divergence in European (measured as FCT) salmon populations.

Table 3.S6 Blast results from BLAST2GO with blast e-value threshold of 1×10^{-3} and gene ontology (GO) terms annotation for blast of e-value inferior to 1×10^{-10}.

Table 3.S7 Likelihood of models tested to best explain the geographical positions of clines in (a) the Baltic continuum and (b) the Barents continuum.

Figures 3.S1 to 3.S29 The distribution of European F_{CT} along the linkage groups (number of Figure corresponds to linkage group number (e.g.: Figure S1 corresponds to linkage group SSA01). Grey areas correspond to 95% confidence interval of the smoothed kernel curve. Individual outlier loci showing elevated or decreased divergence are marked as blue and red bars. Estimated significance levels (–log10 transformed p-value) for region of elevated or decreased divergence are shown as blue and red curves.

Chapitre 4 Landscape Genomics in Atlantic Salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation

Publié sous :

Bourret V, Dionne M, Kent MP, Lien S & Bernatchez L (2013) *Evolution* 67: 3469-3487.

4.1 Résumé

Un nombre grandissant d'études examine maintenant les facteurs historiques et contemporains influençant l'évolution des populations naturelles. En combinant des données caractérisant la composition génomique des populations et les conditions environnementales de leur milieu, de telles études peuvent aspirer à améliorer notre compréhension de l'échelle génomique et géographique de l'adaptation locale en milieu naturel. Dans ce contexte, nous avons utilisé une approche de génomique du paysage à grande échelle afin d'examiner la différenciation adaptative et neutre caractérisant 54 populations de saumon atlantique d'Amérique du Nord représentant sept groupes régionaux génétiquement distincts préalablement définis. Plus de 5500 marqueurs SNP ont été génotypés pour 641 individus et 28 échantillons représentant 25 individus chacun (provenant d'une même population). Des analyses de balayage génomique, cartographie génétique et la caractérisation de 49 variables environnementales ont été combinées afin de réaliser une analyse de génomique du paysage novatrice. Nos résultats amènent de nouvelles perspectives sur les liens entre la variation environnementale et la divergence génétique neutre et potentiellement adaptative. Spécifiquement, nous avons identifié des marqueurs potentiellement sous l'effet de la sélection divergente de même que des associations entre cette divergence adaptative, des facteurs de sélection environnementale et des fonctions biologiques. Les analyses statistiques multivariées ont révélé de fortes corrélations entre les structures génétiques et environnementales des populations étudiées. Parmi les marqueurs potentiellement sous sélection, nous avons trouvé un enrichissement de fonctions liées à la croissance. Les caractéristiques climatiques (température et précipitations) et géologiques des rivières étaient significativement associées à la divergence potentiellement adaptative et neutre des populations y habitant. Ces associations devraient donc être considérées comme candidats importants impliqués dans l'adaptation à l'échelle régionale pour le saumon atlantique. Ainsi, cette étude contribue significativement à l'amélioration des outils utilisés dans les approches modernes de gestion et de conservation des populations naturelles de saumon atlantique.

4.2 Abstract

A growing number of studies are examining the factors driving historical and contemporary evolution in wild populations. By combining surveys of genomic variation with a comprehensive assessment of environmental parameters, such studies can increase our understanding of the genomic and geographical extent of local adaptation in wild populations. We utilized a large-scale landscape genomics approach to examine adaptive and neutral differentiation across 54 North American populations of Atlantic salmon representing seven previously defined genetically distinct regional groups. Over 5500 genome-wide SNPs were genotyped in 641 individuals and 28 bulk assays of 25 pooled individuals each. Genome scans, linkage map and 49 environmental variables were combined to conduct an innovative landscape genomic analysis. Our results provide valuable insight into the links between environmental variation and both neutral and potentially adaptive genetic divergence. In particular, we identified markers potentially under divergent selection, as well as associated selective environmental factors and biological functions with the observed adaptive divergence. Multivariate landscape genetic analysis revealed strong associations of both genetic and environmental structures. We found an enrichment of growth related functions among outlier markers. Climate (temperature-precipitation) and geological characteristics were significantly associated with both potentially adaptive and neutral genetic divergence and should be considered as candidate loci involved in adaptation at the regional scale in Atlantic salmon. Hence, this study significantly contributes to the improvement of tools used in modern conservation and management schemes of Atlantic salmon wild populations.

4.3 Introduction

The environment can influence evolutionary trajectories of living organisms by imposing selective pressures and limiting migration. The last decade has witnessed the birth of landscape genetics, a field devoted to understanding the contribution of environmental conditions on the evolutionary processes shaping population genetic structure in the wild (Manel et al. 2003). Although the emerging phase of landscape genetics mainly focused on linking neutral genetic divergence with ecological constraints to gene flow (e.g.: Petren et al. 2005; Leclerc et al. 2008), a limited number of studies have also accounted for adaptive divergence (Bonin et al. 2006; Gaggiotti et al. 2009; Manel et al. 2010ab). The use of relatively low genomic coverage and anonymous markers often limited the potential for functional inferences. Today, the availability of high throughput genomic tools combined with genome scan facilitate the investigation of adaptive genomic divergence. Ecological and landscape genomics now have greater power to disentangle adaptive from neutral genetic divergence and identify the environmental factors driving divergent selection (Bonin 2008). Furthermore, the possible links between newly developed single nucleotide polymorphism (SNP) markers and functional genes can reveal key biological processes or functions targeted by environmental selective pressures (Bonin 2008; Parisod and Holderegger 2012).

In widely distributed species, local populations often experience heterogeneous environmental conditions, and may evolve in these different environments for thousands of years. In such cases, these environmental conditions are suspected to have shaped distinct genetic composition among populations, and local adaptation of genetically based phenotypic traits. In general, local adaptation results in the superior fitness of indigenous individuals compared to emigrants (Kawecki and Ebert 2004). Nevertheless, the geographic extent of local adaptation can vary depending on the nature of the adaptation, the strength of the selection, the extent of gene flow between populations, and their effective population size (Lacy 1997; Hansen et al. 2002). On the other hand, the genomic extent of local adaptation can also differ among populations depending on the degree of genetic isolation (Feder and Nosil 2010) or the complexity of the trait under selection (e.g. single vs. multi-locus, pleiotropy, epistasis). Even in light of recent technical advances, investigating the

adaptive divergence of wild populations occupying vast heterogeneous environments is a daunting task, especially on non-model species.

While Atlantic salmon is not a classical model organism, the population genetic structure of this species has been extensively studied (e.g.: Vasemägi et al. 2005; King et al. 2007; Palstra et al. 2007; Tonteri et al. 2009; Bourret et al. 2013). In North America, the most extensive study on this species was performed by Dionne et al. (2008) and involved a comprehensive landscape genetics approach aiming to elucidate the environmental parameters influencing the genetic structure of 51 populations. Dionne et al. (2008) found that temperature regime and coastal distance from a southern reference influenced neutral divergence among populations, which suggested a regional component to local adaptation. A hierarchical structure analysis grouped populations into seven regional groups based upon differentiation at microsatellite markers. The recent development of a large panel of SNPs largely discovered from coding regions, offers a more powerful way to gain insights into the possible role of environmentally induced selective pressures in shaping patterns of adaptive divergence and further assess the geographic scale of local adaptation in Atlantic salmon (Bourret et al. 2013).

Local adaptation in salmonids has been recognized as a key evolutionary process driving phenotypic and genetic divergence among populations (Taylor 1991; Garcia de Leaniz et al. 2007; Fraser et al. 2011). However, local adaptation for these species is still equivocal in front of the difficulties of performing common garden and reciprocal transplant experiments, especially on large geographic scales. In Atlantic salmon, different populations exhibit divergence in morphological traits, migratory tactics and reproductive strategies that could be associated with local adaptation since they have been shown to be heritable (Claytor et al.1991; Vähä et al. 2007, Palstra et al. 2007, Paez et al. 2010). Moreover, recent studies have proposed that genetic diversity at MHC class-IIB genes could represent local adaptation to cope with pathogen diversity in rivers with different thermal regimes (Dionne et al. 2007, 2009).

The main objective of this study was to further investigate the environmental factors shaping patterns of genetic divergence and the scale of local adaptation in Atlantic salmon. We accomplished this by increasing the number of genetic markers used by 150 fold and including a larger set of environmental variables. More specifically, we first revisited the population genetic structure of Atlantic salmon to assess the congruence of a wider genomic coverage with previously shown genetic structure. We then tested for significant associations between variation in 49 climate, geological and river specific characteristics and regional genetic structure at markers identified as being potentially under divergent selection. Thirdly, using mapping information, we documented the genomic distribution of these outlier markers. Finally, using available gene annotations, we examine the functional implications of adaptive and environmental divergence among populations and regional groups of Atlantic salmon.

4.4 Materials and Methods

4.4.1 Samples

Samples of adult anadromous Atlantic salmon were collected in the summer of 2004. Methods for tissue collection, storage, DNA extraction and microsatellite analyses for samples collected from 51 rivers were previously detailed in Dionne et al. (2008). In this study, we also added samples from three new rivers. DNA was extracted from fin clips as described by Dionne et al. (2008) from a total of 1341 individuals from 54 rivers in Eastern Canada (Table 1 and 3; Figure 1). SNP genotyping (described below) was conducted using two different approaches. We first performed a single individual genotyping approach for 641 fish from 26 rivers representative of the seven previously identified regional groups, with an average of 25 individuals per river. Secondly, in order to increase the number of populations analyzed while limiting the cost of genotyping, we performed a bulk genotyping approach for 700 individuals from 28 rivers. Previous studies have shown that reliable SNP allele frequency estimates could be obtained by this method (Macgregor et al. 2008; Craig et al. 2009). To prepare bulk assays, DNA from 25 individuals per river was quantified in triplicate using Quant-iT PicoGreen dsDNA Assays (Life technologies, USA). The DNA concentrations for all 25 individuals were standardized to the concentration of the individual with the lowest DNA concentration. For each river, those 25 individuals were then pooled into a single 50µl bulk assay and used as individual for the genotyping steps.

4.4.2 Genotyping Quality Control

From the 26 populations genotyped individually, four were genotyped using version one (V1) of the SNP-array (described in Bourret et al. 2013) as they were used in the first assessment study of the array. The remaining populations (22 in individual genotyping and 28 in bulk assays) were genotyped using version two (V2) of the SNP-array developed by the Centre for Integrative Genetics (CIGENE, Norway). A total of 5349 SNP markers on V2 were selected from V1 for their high quality and 219 SNPs were added from new sequence data. These additional markers were assessed in the same fashion as markers on V1 and discovery and quality control methods for all 5568 SNPs on V2 can be found in

Bourret et al. (2013). Genotyping was performed according to the manufacturer's instructions using the Illumina infinium assay (Illumina, San Diego, CA, USA).

Samples with greater than 85% call rate (*CR*; proportion of SNPs successfully genotyped) were retained for future analyses. Markers absent from V2, but genotyped on V1 in four populations were excluded and subsequent quality control steps were then performed on the remaining V2's 5568 SNPs. Using Illumina's Genotyping Module software, we assessed each SNP's cluster pattern using all individual of North American available populations (n = 900 individuals). Visual inspection allowed for the classification of SNPs into different categories: i) single SNP ii) failed, iii) monomorphic and iv) paralogous sequence variants (PSVs), and v) multisite variants (MSVs) (Table S1). Markers falling in categories other than "single SNP" were excluded from further analyses as well as markers with minor allele frequency less than 1 percent (MAF < 0.01). Ascertainment bias was assessed in Bourret et al. (2013) and was suggested to be minimal in North American populations based on an L-shape distribution of MAF (high proportion of low frequency markers rapidly decreasing towards low proportion of high frequency markers).

4.4.3 Population structure on individual samples

Using individually genotyped samples (26 rivers), we measured global and per SNP observed and expected heterozygosity (H_O and H_E) within each population. In order to exclude markers potentially under divergent or balancing selection from the basic population structure analyses, we then used hierarchical Fdist (Excoffier *et al.* 2009), a genome scan analysis implemented in ARLEQUIN 3.5 (Excoffier & Lischer 2010). This hierarchical method can detect outlier loci among groups of populations (F_{CT}) and represents the most appropriate method for detection of selected markers in our system because previous studies found a hierarchical structure among North American populations (Dionne et al. 2008; Dionne et al. 2009). To this end, we classified rivers in seven regional groups previously identified by Dionne et al. (2008) for the detection of selection analysis. Markers with F_{CT} p-values > 0.01 were then used as the neutral basis for population structure analyses.

Pair-wise genetic differentiation between populations was estimated by the F_{ST} estimate of Weir and Cockerham (1984) using ARLEQUIN 3.5 with 10,000 permutations to determine statistical significance. To confirm the regional classification proposed by Dionne et al. (2008), a principal component analysis on individual genotypes was carried out using the SmartPCA program implemented in the EIGENSOFT package (Patterson et al. 2006) and fitted using R (R core development team). Furthermore, using the confirmed regional structure, two analyses of molecular variance (AMOVAs) were performed using ARLEQUIN 3.5, one on markers identified as potentially under divergent selection and a second on neutral markers.

4.4.4 Bulk assays population structure

In the Illumina's Genotyping Module software used to call genotypes, red and green signals (representing homozygotes AA and BB) are normalized to theta (θ) values of 0 and 1. The expected θ value for a heterozygote is 0.5, but generally, clusters are biased towards 0 or 1. To accurately predict allele frequencies from bulk assays, we calculated a k correction factor for each SNP, which is the average θ value for heterozygotes with $k = \theta$ hets/(1-θ hets). We selected the heterozygotes from genotyped individuals with $CR > 0.99$ in order to estimate the correction factor as accurately as possible. We then calculated the B allele frequency (BAF) on markers retained following the previously described filtering procedure while also excluding markers without heterozygotes among individuals ($CR >$ 0.99) with BAF = θ bulkassay/(θ bulkassay + k (1-θ bulkassay). In order to test the robustness of this genotyping method, bulk assays from 5 rivers that were previously genotyped using the single individual genotyping approach were also genotyped using the same individuals pooled in a single bulk assay per population (two replicates). A population's allele frequencies from both bulk assays and individual genotypes for each SNP were then contrasted using a simple regression and fitted using R. To further confirm the Dionne et al. (2008) regional classification and classify the additional three populations genotyped as bulk assays, we used a discriminant analysis implemented in SAS 9.1 (proc discrim; SAS Institute Inc.), which acts as a custom population assignment method. In the first step of this analysis, the population's allele frequencies for the 26 individually genotyped rivers are used to build a discriminant rule that classifies population in their

regional group based on statistical distances estimated by the procedure. Then, using this discriminant rule, bulk assays are given a probability of belonging to either one of the regional groups based on estimated population allele frequencies (BAF).

4.4.5 Environmental structure

Rivers (n = 26) used in individual genotyping were also characterized for 49 environmental parameters distributed among 3 main categories, namely climate, river properties and geological variables. A detailed list of environmental parameters is available in Table 4. Climate variables (n = 19) were extracted from 35 meteorological stations where values of numerous temperature and precipitation parameters were collected over a period of approximately 30 years (1971-2000, Environment Canada, http://climate.weatheroffice.gc.ca/Welcome_e.html). We used data directly from a station to characterize a river when it was located within 50km of the mouth of the river. Otherwise, we estimated values for a subset or all climate parameters using a Kriging interpolation with the Geostatistical Analyst in ArcGIS 9.2 (ESRI Inc.), a common geostatistical technique to predict values at unmeasured locations. River properties (n = 11) were obtained from the Ministère des Ressources naturelles et de la Faune du Québec (MRNF) and the Canada3D database available through the GéoGratis web site of the Ministry of Natural Resources Canada (NRC; http://geogratis.cgdi.gc.ca) as described in Dionne et al. (2008). When a river property parameter was missing for a given river, it was replaced by the average across all populations for that parameter (total of 23/437 cases for 6/19 parameters). Six categorical geological parameters were then considered and divided in 19 subcategories. The dominant subcategory (scored as 1) in a river's watershed for each parameter was determined by estimating the area covered by polygons of the geological layer associated with each of the subcategories and identifying the subcategory with the highest coverage. This was realized by examining the intercept between a watershed layer provided by the MRNF and a geological layer from the GeoScape Canada database available on the Natural Research Council of Canada (NRC) web site (www.nrcan.gc.ca/earth-sciences/products-services/mapping-product/geoscape/6032). Each geological parameter was then transformed into presence/absence scores for each subcategory where only the dominant subcategory of a given river was scored as 1. This

transformation was performed to include these categorical parameters along with continuous variables (climate and river properties) in further analyses. In order to minimize the co-linearity among all 49 environmental parameters, we performed a principal component analysis on populations using SAS 9.1 (proc factor; rotation = varimax). As a surrogate for environmental parameters, we then used a number of principal component factors (PC factors) equal to the number of eigenvalues greater than one, a widely used statistical rule known as the Kaiser–Guttman criterion (Yeomans and Golder 1982).

4.4.6 Genetic-environment associations

Association between population genetic structure and environmental parameters was assessed *via* a redundancy analysis (RDA), which is a special case of canonical correlation analysis (CCA). The CCA is a statistical test used to relate information from two different data tables. Here, using rivers as subjects, we specifically tested if the independent parameters (environmental PC factors) could predict the dependent parameters (allele frequencies). An analysis of variance (ANOVA; 1000 permutations) was then performed to assess the global significance of the RDA and a marginal ANOVA (1000 permutations) was also run to determine if environmental PC factors were significantly correlated with allele frequencies. We also estimated the effect size of the relationship between the two datasets using Wilks' Lambda, which is a parameter analogous to the correlation coefficient (R^2). We then computed Pearson's correlation coefficients for all markers used in the RDA and environmental PC factors. Moreover, in order to localize potential genomic regions under environmentally divergent selection, we used a linkage map for the North American Atlantic salmon (Brenna-Hansen et al. 2012) built using the same SNP-array to map markers potentially under divergent selection. We mapped each SNP according to the results from the hierarchical Fdist as well as the Pearson's correlation coefficient values for each significantly associated environmental factor. Analyses from this section were all performed in R.

4.4.7 Gene ontology and SNP annotation

Blast2go (Gotz *et al.* 2008) was used to associate gene ontology (GO) annotation terms to all SNPs retained for genomic analyses (3118 SNPs; see Results). A homology search was

first completed by performing a BLAST (Altschul *et al.* 1990) search of the available flanking sequences for each SNP on the NCBI nr public database with the *e*-value threshold set to $1x10^{-10}$. Blast2go then retrieved GO terms associated with the obtained BLAST hits. In order to determine if the biological processes, molecular functions or cellular components of the markers potentially under divergent selection were over-, equally or under-represented among outlier markers when compared to the entire retained SNP dataset, we performed an enrichment analysis using Fisher's Exact Test corrected for multiple tests by applying a false discovery rate of 0.05 (FDR; Benjamini and Hochberg 1995).

4.5 Results

4.5.1 Genotyping and quality control

One individual sample (SU-14) with a *CR* < 0.85 (0.77) was excluded from the dataset. After initial quality control and classification of genotypes obtained from 900 samples, we classified 3974 markers out of 5568 SNPs featured on the V2 array as single locus and polymorphic SNPs (i.e. diploid SNPs) for North American Atlantic salmon. Among the 3974 "good" SNPs, 3118 markers showed an overall MAF > 0.01. Therefore, besides the individuals genotyped in the bulk assays, 3118 SNPs and 640 individuals were kept for further analyses. Table 1 shows summary data for call rates (*CR*), observed (H_O) and expected heterozygosity (H_E) across populations (Table S2 across markers).

4.5.2 Population structure on individual samples

When we followed the regional grouping proposed by Dionne et al. (2008), the average F_{CT} across 3118 loci was 0.057 (ranging from -0.026 to 0.535). At the 0.01 and 0.05 significance level respectively, 68 and 179 markers were identified as potentially under divergent selection and 34 and 208 markers were identified as potentially under balancing selection (Table S3). Removing outlier markers at the 0.01 significance level (68 divergent and 34 balanced), 3016 markers were used as the basis for the neutral pair-wise differentiation estimates, PCA and AMOVA. All pair-wise comparisons of genetic differentiation between populations were highly significant (P < 0.001; Table S4). In a principal component analysis (PCA) on individual genotypes, five principal component (PC) factors determined at least 1% of the variation each, and together explained 10.5% of the total genetic variation among individuals. Principal components 1 to 3 accounted for 4.0%, 2.3% and 1.6% respectively. Principal component 1, PC2 and PC3 differentiated populations into the seven regional groups that Dionne et al. (2008) previously defined (Figure 2). Both AMOVAs on 3016 neutral SNPs and 68 divergent SNPs showed significant genetic variation among groups (Table 2). For neutral markers, the genetic variation attributed to differences among groups accounted for 6.42% while this percentage increased to 28.06% for divergent markers. Differentiation among populations within groups was similar with 3.04% and 3.00% for neutral and divergent SNPs respectively.

Thus, intergroup differentiation was about 4.5 times more pronounced at outlier markers than neutral loci while inter-population differentiation within groups remained the same.

4.5.3 Bulk assays population structure

Of the available individuals genotyped (n = 900), 728 had *CR* > 0.99 and were thus used to estimate the marker's *k* correction factors. Of the 3118 markers used in previous analyses, 287 markers did not have heterozygotes to estimate the correction factor and were excluded from bulk assays analyses. Correlations of population allele frequencies estimated from bulk assays to those measured with actual individual genotypes yielded correlation coefficients (R^2) ranging from 0.898 (MAP) to 0.921 (DGO) with all p-values < 0.001. Globally, population allele frequencies estimated from bulk assays were highly correlated with those estimated from individual genotypes with an overall R^2 = 0.909 (p-value < 0.001; Figure S1). Based on the allele frequencies of the 26 populations genotyped on an individual basis and on the regional group they belonged to, we were able to build a powerful discriminant rule. Regional distances estimated *via* the discriminant analysis allowed us to assign all 28 bulk-assayed populations to one of the seven regional groups with a probability of 1.00 (Table 3). Out of the 26 previously classified populations (Dionne et al. 2008), only the Ouelle and Laval rivers were both "missassigned" to the Southern Québec region instead of the Québec City region for the former and the Higher North Shore region for the latter population.

4.5.4 Environmental structure

In the remote Ungava region, compiled climate data was available for only a single meteorological station and the next closest stations were located over 1000 km away from the mouth of the rivers. Therefore, for these rivers, the interpolation yielded values outside a reasonable confidence interval (data not shown). For this reason and considering that these populations encounter extreme climatic conditions, the three rivers from Ungava (AF, GE and KO) were left out of environmental analyses to avoid confounding factors due to biased outlier values. Of the 23 rivers left, 11 were located within close distance to a meteorological station and did not require interpolated climate data. We then used the interpolated climate data partially for three rivers (for degree-days) and completely for the

remaining nine rivers. A PCA on environmental parameters presented ten PC factors with eigenvalues greater than 1, and together explained 92.1% of the total environmental variation among rivers. Principal components factors 1 to 4 accounted for 33.8%, 12.8%, 11.3% and 8.3% respectively. PC factor 1, PC factor 3 and PC factor 4 differentiated the populations along the spatial axes while PC factor 2 offered a similar dispersion of populations as PC factor 1 (not shown). In general, populations tended to group according to previously identified regional groups but several populations showed less clear patterns (Figure 3). Different PC factors were dominated by different categories of environmental parameters. As indicated by the loadings of parameters on the ten PC factors retained (Table 4), PC factor 1 was primarily loaded with climate data related to temperature (10/16), PC factor 2 by climate data related to precipitation (7/8), PC factor 3 by river properties (6/6), PC factor 4 and 7 by geological data (3/3 and 3/3 respectively).

4.5.5 Genetic-environment associations

Using the 23 rivers as subjects and the ten environmental PC factors as explanatory variables, Figure 4 shows a RDA performed on the 179 SNPs potentially under divergent selection (0.05 significance level) as the response variables. We used this outlier threshold to be consistent with correlation and annotation analyses where a less stringent threshold for outliers was used. Moreover, RDA results using more or fewer markers were similar, and adding more markers increased variance and p-values (data not shown). Globally, the RDA was highly significant with a P-value < 0.001 (ANOVA, F = 5.025). The first 10 RDA axes accounted for 80.7% of the variation, while the RDA axis 1 and 2 represented 49.7% and 12.4%, respectively. The marginal ANOVA showed that PC factors 1, 2, 4 and 7 were significant predictors of the populations' allele frequencies with P-values < 0.001 (respective F = 20.617, 3.379, 10.023 and 4.536; Figure 4). Principal component factors 3, 9 and 10 presented significant relationships but to a lesser degree (P-values between 0.015 and 0.038). The correlation between the multivariate environmental and genetic structure was highly significant (P-value < 0.001) with an effect size of 0.980 (analogous to R^2). The distribution of environmental PC factors 1, 2, 4 and 7 Pearson's correlation coefficients for the 179 outlier markers used are shown in Figure 5. A higher frequency of correlation coefficients greater than 0.6 is found for PC factor 1, whereas an increasing frequency of

lower coefficients generally characterized the PC factors with decreasing F-statistics values. The genomic distribution of F_{CT} values and PC factor 1 Pearson's correlation coefficients among all 3118 SNPs are shown in Figure 6. Overall, outlier markers are widely distributed among all linkage groups. Moreover, markers highly correlated with PC factor 1 are present on all linkage groups, whereas identified genetic outliers are often those with the highest correlation coefficient to PC factors on a given linkage group.

4.5.6 Gene ontology and SNP annotation

The BLAST and annotation steps in Blast2go yielded 1119 SNPs with annotations (Table S5). Using a FDR of 0.05, an enrichment analysis did not indicate significant over- or under-representation of any biological pathway among the 208 markers potentially under balancing selection. However, 12 Gene ontology terms (GO-terms) were over-represented among the 179 markers potentially under divergent selection, which were associated with 12 SNPs (Table S6). Molecular functions, biological processes and cell compartments associated with identified GO-terms suggested that these markers were associated with growth. The particular categories highlighted by the enrichment analysis include: positive regulation of JNK cascade, ephrin receptor binding, syndecan binding, frizzled binding. It should be noted that these categories were highlighted from a common hit of four markers for the syntenin-1 protein of Atlantic salmon (GenBank: ACI33400.1). We then plotted the relationship between population allele frequencies of these four SNPs against the environmental parameter with the highest loading on PC factor 1 (average temperature between May and September). All four models (generalized linear models) showed a similar significant regression (Figure 7 shows one example regression). For PC factor 1 and 4, we then divided outliers into quartiles according to their correlation coefficients and used the fourth quartile (highly correlated markers) to perform enrichment analyses on the set of 179 outliers as reference. The test for temperature correlated markers identified similar functions and the same four markers reported above, which all had correlation coefficients greater than 0.70 with PC factor 1. However, no terms were over-represented for geological-correlated markers for FDR = 0.05, while the enrichment was significant with a less stringent 0.05 p-value threshold.

4.6 Discussion

By surveying more than 3000 SNP markers widely distributed across the Atlantic salmon genome from 54 populations in combination with a thorough examination of 49 environmental factors in 23 North American rivers, we have been able to complete one of the most extensive landscape genomics analyses reported to date. The innovative statistical framework presented here demonstrated a very strong correlation between genetic and environmental structure characterized by significant associations between potentially adaptive divergence and climate. Geological parameters were also found to be important factors associated with potentially adaptive divergence. Our results suggest a regional component to local adaptation in Atlantic salmon that is associated with both climatic and geological factors. Furthermore, among markers potentially under divergent selection, we observed an enrichment of gene ontology terms associated with growth related functions, suggesting a role for these biological functions in the adaptive divergence among populations and regional groups.

4.6.1 Genetic divergence

The first objective of this study was to revisit the population genetic structure of Atlantic salmon in this system with a new set of SNP markers and confirm whether or not the regional structure revealed with microsatellites was supported with SNP-array genotypes. Two results strongly suggest that the neutral genetic structure supported by SNP markers is similar to that of the microsatellite markers. First, pairwise F_{ST} estimates obtained for both types of markers are highly correlated (data not shown). Second, the distribution of populations along the first three principal components of the PCA shows a regional organization identical to the one reported in Dionne et al. (2008). Regional differences were such that by using the population allele frequencies of 2831 markers, we were able to build a powerful discriminant rule to classify bulk-assayed populations to their regional group. Out of 26 bulk assays for which populations had already been associated with a regional group in Dionne et al. (2008), 92% were concordantly assigned to the same regional group using our discriminant rule and SNPs. In the only two discordant populations, the Ouelle River population was originally classified as part of the Québec City region but included in the Southern Québec region with SNP markers. This river is actually located on the western

limit of the Southern Québec group, which is bordered by the Québec City group. It thus seems plausible that this discordant regional assignment is at least partially due to an improved assignment power resulting from the increased number of markers. The Laval river population, however, was previously classified in the Higher North Shore regional group and is now assigned to the Southern Québec group, which is not geographically congruent. The only reasonable explanation we can offer at this stage is that the Laval, which is a small population (N < 100 individuals), might have been seriously affected by drift effects and occasional straying, causing this mixed signal in the population. On the other hand, two of the new populations analyzed with the bulk assays, Kecarpui and Des Escoumins, were assigned to regional groups concordant with geographical location. Overall, given the convincing regional assignment probabilities provided by the discriminant analysis, we argue that further interpretations concerning the regional genetic structure in the system could be extrapolated to bulk-assayed populations.

Apart from generally confirming the regional structure observed in Dionne et al. (2008), a second contribution of the new genomic dataset was the detection of 68 outliers potentially under divergent selection among populations. For this relatively small set of markers, the genetic variation attributed to differences among regions was more than four times the observed proportion for neutral SNPs (28.04% and 6.42% respectively). Comparing indirect estimates of migration rates between and within regional groups, Dionne et al. (2008) previously hypothesized that local adaptation at the regional scale was driving higher population differentiations for inter- rather than intra-group population comparisons even when the distance between populations was similar in either comparisons. This agrees with theoretical expectations that local selection in subdivided populations enhances between-deme genetic diversity (Charlesworth et al. 1997). It is also in agreement with our observation of an increased level of divergence between regions at outliers relative to neutral SNP markers but not between populations within a given regional group. In summary, our results provide further evidence for a possible role of selection in shaping regional population structure in Atlantic salmon.

4.6.2 Genetic-environment associations

Once the putative targets of divergent selection were identified from the genome scan, the second step was to identify the particular environmental variables acting as potential selective agents driving adaptive genetic divergence among regional groups. We found a strong association between the overall regional genetic groups and environmental structure. This indicated that these regional genetic groups also differed in ecological settings and that several environmental factors could represent selective agents leading to regional local adaptation. In particular, four PC factors were found to be strong predictors of genetic divergence. Temperature related PC factor 1 and precipitation related PC factor 2 were both climate factors correlated with genetic divergence, but displayed opposite direction vectors on the RDA axes, while geological related PC factors 4 and 7 oriented in relatively similar directions to each other. Owing to the orthogonal relationship between climate and geological vectors on the RDA, we chose to discuss environmental factors as two main axes correlated with potentially adaptive divergence, namely climate and geology. Furthermore, since PC factor 2 was the fourth factor in decreasing order of statistical significance and negatively correlated with temperature on the RDA axes, we hereafter refer to temperature as a proxy for climate conditions.

Temperature regime was identified as the most important selective agent in the system, as this climatic factor has the highest F-statistic for PC factor 1 in the RDA and the highest correlation coefficient with outlier markers. Numerous studies have suggested that temperature regime is an important variable influencing local adaptation in Atlantic salmon (reviewed in Taylor 1991 and Garcia de Leaniz et al. 2007) and that growth (Claytor et al. 1991; Nicieza et al. 1994ab; Paez et al. 2010) and immune related functions (Dionne et al. 2007, 2009) could be important targets of local selection. Southern Atlantic salmon populations live in warmer conditions and are known to grow faster and migrate to sea at younger age (Power 1981; Metcalfe and Thorpe 1990), which could be linked to living in a more productive environment. However, Nicieza et al. (1994ab) observed a significantly higher digestion and growth rate for salmon from higher latitude (Scotland) compared to salmon from southern latitudes (Spain) when reared at the same temperature. These results argue that countergradient selection has resulted in selection for more efficient growth in

northern latitudes to compensate for a shorter growing season. Therefore, rather than local adaptation for high growth rate in the southern latitudes, the countergradient theory (reviewed in Conover and Schultz 1995) suggests that cold temperatures, a proxy for short growing season length, might actually be selecting for more efficient growth in higher latitudes. Furthermore, temperature regime was found to be closely related to bacterial diversity in the wild, which in turn was associated with genetic diversity of an immune-competence gene, the major histocompatibility complex class-IIB gene (Dionne et al. 2007, 2009). Genetic diversity at this locus is suggested to be involved in local adaptation of Atlantic salmon to different pathogen communities associated with different thermal regimes.

Geological parameters were, after climate, the environmental factors with the strongest links to genetic divergence. These were primarily loaded by the geological provinces category, geological periods associated with rock formations and some specific rock or substrate types. All of these geological categories emphasized a division between populations from rivers draining either on the north or south shore of the St. Lawrence River, and supports the distinctiveness of Anticosti Island populations. Thus, rivers south of the St. Lawrence River and Anticosti belonged to the *Plate-forme du St-Laurent* and *Appalachian Orogen* geological provinces that are characterized by sedimentary rock formations dating to Silurian and Devonian periods of the Paleozoic era. In contrast, rivers draining on the north shore of the St. Lawrence River belong to the Grenville Province, which is mostly characterized by rock formation of gneiss type dating to the Mesoproterozoic period of the Precambrian era. Many factors can influence water chemistry, but aside from anthropogenic impacts, geology has been shown to be a dominant factor (Stallard and Edmond 1983; Johnson et al. 1997). Accordingly, the Southern Québec and Anticosti rivers are characterized by alkaline water. Since fish are surrounded by their environment, constant osmotic, ionic and pH regulation are required to maintain homeostasis. While water pH outside neutrality represents a stress for most fish, many species have adapted to alkaline or acidic waters (Pritchard 2003). We thus propose that regional specificity of geological parameters may interact with water chemistry of rivers to represent potential selective agents driving local adaptation in Atlantic salmon populations.

Furthermore, as reported in Perrier et al. (2011), which found geological areas to significantly correlate with Atlantic salmon population genetic structure in France, geological substrate is suspected to be instrumental in the propensity of salmon to return to their natal river to spawn (Stabell 1984; Dittman et al. 1996). Geological characteristics of rivers may be an important factor influencing the accuracy of homing behavior and consequently reduce straying among regional groups, which would further contribute to maintain a regional component of local adaptation.

4.6.3 Extent of local adaptation

By studying climatic and geological factors, we now have evidence to argue for environmental selection driving adaptive divergence at the regional level. In fact, the strong regional genetic differentiation associated with distinct environmental features allowed us to use bulks assays to assign an additional 28 populations to regional groups without requiring us to perform individual genotyping. We emphasize that bulk assays provided additional indications for potential regional local adaptation and better defined the boundaries of this component in our system. The significant associations of temperature and geology with regional population genetic structure thus suggests these parameters are among the most important environmental selective agents delineating the geographical scale of local adaptation.

We observed a genome-wide distribution of divergent outliers and environmentally correlated markers, which is not unexpected given the diversity of putative environmental selective agents identified. In fact, as discussed above, climate and geological characteristics can be used as proxies for many indirect ecological differences among populations from different regional groups, such as growing season length, pathogen diversity, and water chemistry. Thus, numerous targets of selection spread across the genome are more likely to emerge than a small number of targets in localized islands of adaptive divergence (Nosil et al. 2009). Theory predicts that local selection acting in a species with hierarchical population structure can lead to increased differentiation between demes. Furthermore, Charlesworth et al. (1997) found that under such conditions, local selection produced very high differentiation values for loci close to targets of selection but

also high values for distant neutral loci with no selection. This is precisely what we observed when exploring the environmental correlation with either all 3118 SNPs, only neutral SNPs or with only highly divergent markers. The RDA systematically indicated an overall association between genetic and environmental divergence. Since the association was strongest when considering only the fewest outliers markers (rather than diminishing the value of the observed association), we argue that it reinforces the environmental selection arguments. Albeit weaker than those of putative targets of local adaptation, concordant differentiation patterns across the genome might be indicative of what Thibert-Plante and Hendry (2010) referred to as a generalized barrier to gene flow. Other recent studies, for instance Cooke et al. (2012), also reported a concordant but amplified genetic divergence when comparing potentially selected markers correlated with contrasting environmental conditions against the neutral genetic divergence of characin fish (*Triportheus albus*) in Amazonia. However, in some cases contrasting patterns of adaptive *vs.* neutral divergence were detected due to a strong association between adaptive divergence and environmental condition not reflected in the neutral differentiation patterns (e.g.: Gaggiotti et al. 2009; Bradbury et al. 2010; Lee and Mitchell-Olds 2011). In such cases, some outlier loci could also be linked to genetic incompatibilities revealing ancestral divergence rather than actual markers associated with selected genes to exogenous (environmental) factors (Bierne et al. 2011).

4.6.4 Functional implications

Although interpretation of available sequence annotation should be made with caution, especially in an ecological context involving non-model species (Pavlidis et al. 2012; Pavey et al. 2012), they remain a useful tool for determining possible functional targets of selection, identifying candidate genes and framing hypotheses to link environmental selective agents and adaptive divergence at the genome level. The importance of temperature related selection was predominant among the markers potentially associated with adaptive divergence. Indeed, the GO categories over-represented among divergent outlier markers when compared to the 3118 markers were the same than those over-represented among the outlier markers highly correlated with temperature (fourth quartile) when compared to the complete set of divergent outliers ($p = 0.05$). Enriched annotations

were primarily associated with biological processes and functions linked to growth (e.g.: regulation of JNK cascade, ephrin and frizzled binding). As reported above, the sequences for markers revealing the most enriched GO categories blasted to a syntenin-1 sequence from Atlantic salmon. This protein functions as a binding protein for syndecan, a transmembrane domain protein with an important growth-factor-receptor activation function for one of its four forms (Carey et al. 1997). Given the regional differentiation at these SNP markers located in growth related genes, we propose that they might bear a signature of thermal local adaptation linked to countergradient selection imposed by growth season length. Thus, thermal regimes may act as a selective agent driving local adaptation on growth potential. This hypothesis should be tested by measuring the impact of outlier allelic variants on enzyme function and growth rate.

4.6.5 Conclusion

In summary, this study offers new insights on the geographic and genomic extent of local adaptation in Atlantic salmon by combining population genomics with landscape genetics. As in any landscape genetics study, the relationships between environmental conditions and genetic divergence presented are correlational and not necessarily causal, but nevertheless provide testable hypotheses for future studies, such as reciprocal transplants or specific genotypes impacts on enzyme function, growth or range-wide fitness measurements in wild Atlantic salmon. Owing to a wide geographic and genome coverage, we were able to rigorously confirm a hierarchical genetic structure, and found a strong regional component of both neutral and potentially adaptive divergence among the 54 populations we studied. We also found that this regional genetic structure was significantly correlated with an ecological structure described by a set of 49 environmental parameters. We found specific associations between environmental factors related to climate (temperature) and geology with markers potentially under divergent selection. This allowed us to propose putative environmental selective agents and candidate genes potentially involved in the process of local adaptation in Atlantic salmon. We also found that markers potentially under divergent selection were distributed throughout the genome. Finally, we were able to use annotations to infer a plausible causal link for environmental selection associated with growth related functions.

Although we reported very strong support for a regional component of local adaptation at the geographic level, we lacked the genomic coverage to investigate the extent of adaptive hitchhiking surrounding our genome-wide potential targets of selection at the genomic level. An enhanced genomic coverage could allow the investigation of precise genomic location potentially under the effect of divergent selection and document more precisely the location and size of islands of adaptive divergence (Feder and Nosil 2010). Finally, while our results call for experimental confirmation of the adaptive hypotheses that they propose, this study illustrates how landscape population genomics contribute to improve our knowledge of the evolutionary processes affecting populations and may help develop conservation tools integrating both genetic and environmental parameters and their interactions (Funk et al. 2012).

4.7 Acknowledgements

We are grateful to Gaétan Daigle for incredible support and insights on statistical analyses, Benoît Landry and technical assistance from the Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs du Québec (MDDEFP) in the identification of available environmental information, Anne C Dalziel for language editing and three anonymous reviewers for their comments on the original manuscript. The SNP discovery, array development and genotyping were performed by CIGENE at the national technology platform, supported by the functional genomics programme (FUGE) in the Research Council of Norway. This research was supported by a strategic grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) led by LB in close collaboration with the MRNF represented by MD and the Department of Fisheries and Oceans Canada (DFO). VB was supported by an Alexander Graham-Bell scholarship from NSERC and LB. is the Canadian Research Chair in genomics and conservation of aquatic resources. The authors declare no conflict of interest.

4.8 Tables

Table 4.1 Description of regional groupings and parameters associated with sample sites composing the groups: latitude and longitude, number of individuals genotyped (N), average call rate per population (CR) and average expected (*HE*) and observed (*HO*) heterozygosities per population.

Regional Groups	Population ID	Code	Latitude	Longitude	N	CR	H_O	H_E
Southern Québec	Miramichi	MIR	47,09	-65,32	25	0.99	0.181	0.179
	Matapédia	MAP	47,97	-66,93	25	0.97	0.204	0.202
	Grande Cascapédia	CS	48,21	-65,90	25	0.98	0.195	0.191
	St-Jean Gaspésie	SJQG	48,77	-64,43	25	0.99	0.205	0.202
	Sainte-Anne	SA	49,12	-66,50	25	0.99	0.199	0.196
	Matane	MAT	48,85	-67,53	25	0.98	0.203	0.202
Québec City	Malbaie	ML	47,65	-70,13	25	0.99	0.182	0.176
	Du Gouffre	DGO	47,43	-70,49	25	0.99	0.181	0.175
	Sainte-Marguerite	SM	48,25	-69,93	25	0.99	0.178	0.173
Higher North Shore	Trinité	TRI	49,42	-67,30	25	0.99	0.179	0.179
	Moisie	MOI	50,20	-66,07	25	0.99	0.181	0.176
	St-Jean Côte-Nord	SJQC	50,28	-64,33	25	0.99	0.185	0.181
	Natashquan	NAT	50,12	-61,80	25	0.99	0.170	0.174
Lower North Shore	Musquaro	MUS	50,22	-61,07	25	0.99	0.163	0.168
	Etamamiou	ET	50,27	-59,97	25	0.99	0.166	0.169
	Gros Mécatina	MEC	50,77	-59,08	25	0.98	0.162	0.162
Anticosti	Jupiter	JU	49,47	-63,58	25	0.99	0.202	0.197
	Aux Sumons	SU	49,42	-62,23	25	0.98	0.196	0.191
	Chaloupe	CHA	49,13	-62,53	23	0.97	0.201	0.198
Labrador	Napetipi	NAP	51,31	-58,06	25	0.99	0.185	0.185
	Saint-Paul	STP	51,47	-57,70	25	0.98	0.180	0.182
	Vieux-Fort	VF	51,32	-58,02	25	0.99	0.190	0.188
	Southwest Brook	SW	53,42	-57,23	25	0.99	0.177	0.174
Ungava	George	GE	58,82	-66,17	18	0.99	0.160	0.156
	Koksoak	KO	58,53	-68,17	25	0.99	0.165	0.163
	Aux Feuilles	AF	58,77	-70,07	25	0.99	0.159	0.157

Table 4.2 Analysis of Molecular Variance (AMOVA) using neutral markers (n = 3016) and divergent outlier markers (n = 68). *P-value < 0.001

Source of variation	df	Neutral SNPs	Divergent SNPs
Among Groups	6	6.42*	28.04*
Among populations within groups	19	3.40*	3.00*
Within populations	1254	90.18*	68.95*

Table 4.3 Discriminant analysis population assignment results for bulks assays in contrast to previous classification of Dionne et al. (2008). *NA* refers to population not previously classified

Number	Population ID	Previous classification	Bulk Assigned to
1	Ouelle	*Québec City*	*Southern Québec*
2	Laval	*Higher North Shore*	*Southern Québec*
3	Patapédia	*Southern Québec*	*Southern Québec*
4	Bonaventure	*Southern Québec*	*Southern Québec*
5	Petite Cascapédia	*Southern Québec*	*Southern Québec*
6	Cap-Chat	*Southern Québec*	*Southern Québec*
7	York	*Southern Québec*	*Southern Québec*
8	Madeleine	*Southern Québec*	*Southern Québec*
9	Grand Pabos	*Southern Québec*	*Southern Québec*
10	Darmouth	*Southern Québec*	*Southern Québec*
11	Causapscal	*Southern Québec*	*Southern Québec*
12	Mitis	*Southern Québec*	*Southern Québec*
13	Upsalquitch	*Southern Québec*	*Southern Québec*
14	Little Main	*Southern Québec*	*Southern Québec*
15	Kegwick	*Southern Québec*	*Southern Québec*
16	Des Escoumins	*NA*	*Québec City*
17	Jacques Cartier	*Québec City*	*Québec City*
18	Petit Saguenay	*Québec City*	*Québec City*
19	Kecarpui	*NA*	*Lower North Shore*
20	Corneille	*Lower North Shore*	*Lower North Shore*
21	Muddy Bay	*Labrador*	*Labrador*
22	Sand Hill	*Labrador*	*Labrador*
23	Eagle	*Labrador*	*Labrador*
24	Godbout	*Higher North Shore*	*Higher North Shore*
25	Aux Anglais	*Higher North Shore*	*Higher North Shore*
26	Watshishou	*Higher North Shore*	*Higher North Shore*
27	Aganus	*Higher North Shore*	*Higher North Shore*
28	Aux Rochers	*Higher North Shore*	*Higher North Shore*

Table 4.4 Description and summary of environmental parameters loadings on ten retained principal component (PC) factors after PCA on 49 parameters. Parameters are ordered according to their primary PC factor loading. Gray and white areas refer to alternance between PC factors row associations.

Parameter Descriptions	PC Factor 1	PC Factor 2	PC Factor 3	PC Factor 4	PC Factor 5	PC Factor 6	PC Factor 7	PC Factor 8	PC Factor 9	PC Factor 10
Climate - Average temperature May-September in °C	**0.98**	-0.12	0.02	0.03	0.00	0.03	-0.02	0.03	0.09	0.00
Climate - Degree-days >10°C from Apri-October	**0.96**	-0.14	0.01	0.06	0.16	0.05	0.12	0.05	-0.01	-0.01
Climate - Degree-days >5°C from Apri-October	**0.95**	-0.08	0.07	0.11	0.14	0.04	0.15	0.10	0.02	-0.03
Climate - Yearly Average temperature	**0.94**	-0.04	-0.03	0.18	0.11	-0.11	0.08	0.17	0.01	-0.01
Climate - Degree-days >15°C from Apri-October	**0.93**	-0.21	-0.06	-0.01	0.24	0.06	0.11	-0.02	-0.06	-0.01
Climate - Degree-days >0°C from Apri-October	**0.91**	-0.06	0.14	0.13	0.15	0.04	0.21	0.15	0.02	-0.05
Climate - Days with temperature >20°C in a year	**0.89**	-0.32	-0.07	-0.05	0.16	0.07	0.17	-0.13	-0.02	-0.01
Climate - Degree-days >18°C from Apri-October	**0.88**	-0.18	-0.06	-0.06	0.38	0.06	0.10	-0.05	-0.09	-0.01
River property - Longitude in decimal	**0.80**	-0.27	0.12	-0.01	-0.26	0.21	-0.20	-0.01	0.28	0.09
Climate - Average temperature December-March in °C	**0.72**	0.07	-0.08	0.33	0.24	-0.23	0.20	0.33	-0.07	0.04
Climate - Average temperature October-November in °C	**0.67**	0.10	-0.19	0.18	0.22	-0.39	0.18	0.34	-0.18	-0.04
Geological - RockType - Sedimentary	**0.59**	-0.03	-0.04	0.58	0.08	0.22	0.45	0.21	-0.03	-0.07
Geological - Era - Paleozoic	**0.59**	-0.03	-0.04	0.58	0.08	0.22	0.45	0.21	-0.03	-0.07
Climate - Yearly snowfall in cm	**-0.59**	-0.08	0.12	-0.04	0.23	0.30	0.14	-0.44	-0.36	-0.04
River property - Coastal distance from Miramichi River	**-0.77**	0.01	-0.04	-0.37	-0.08	-0.15	-0.35	-0.06	-0.16	-0.04
River property - Latitude in decimal	**-0.96**	-0.02	-0.02	-0.09	0.06	-0.07	0.04	-0.01	-0.07	-0.08
Climate - Rainfall from May to October in mm	-0.17	**0.95**	0.01	0.08	0.02	0.13	-0.02	-0.10	0.08	0.05
Climate - Days with rainfall >10mm	-0.16	**0.94**	-0.08	-0.03	-0.06	0.07	-0.04	-0.14	0.02	0.09
Climate - Yearly rainfall in mm	0.12	**0.92**	0.08	0.20	0.13	-0.16	-0.04	-0.07	0.07	-0.10
Climate - Days with rainfall >25mm	0.19	**0.83**	0.24	0.30	-0.07	-0.07	-0.07	-0.17	0.11	-0.08
Climate - Days with rainfall >5mm	-0.44	**0.76**	-0.20	-0.18	-0.07	0.13	-0.05	0.00	0.09	0.18
Climate - Rainfall from July to September in mm	-0.40	**0.75**	0.03	-0.03	0.07	0.37	0.11	-0.11	-0.02	0.11
Climate - Yearly precipitation (rain+snow) in mm	-0.32	**0.70**	0.17	0.12	0.29	0.10	0.06	-0.37	-0.21	-0.06
Geological - SubRockType - Paragneiss	-0.13	**0.59**	-0.22	-0.33	-0.13	-0.14	-0.02	0.16	-0.15	-0.19
River property - River length	-0.14	0.12	**0.94**	-0.12	0.05	0.05	0.02	-0.02	0.04	-0.15
River property - Migration difficulty (slope)	-0.07	0.03	**0.92**	-0.10	0.03	0.08	0.04	0.03	-0.04	-0.26
River property - Exploitation capacity	-0.08	0.03	**0.92**	-0.08	0.03	-0.01	-0.10	-0.16	0.04	0.07
River property - Maximum altitude	0.13	-0.16	**0.64**	-0.16	-0.04	0.23	0.48	0.21	0.23	-0.16
River property - Percentage of MSW	0.38	-0.12	**0.62**	0.03	-0.32	-0.05	0.07	0.03	0.46	0.28
River property - Percentage of Grilse	-0.38	0.12	**-0.62**	-0.03	0.32	0.05	-0.07	-0.03	-0.46	-0.28
Geological - Province - Plate-forme du Saint-Laurent	0.16	0.06	-0.18	**0.91**	-0.11	-0.10	-0.21	0.05	-0.06	-0.06
Geological - Period - Silurian	-0.06	0.15	-0.15	**0.90**	-0.03	-0.07	-0.05	-0.02	0.09	-0.01
Geological - SubRockType - Undivided sedimentary	0.41	0.04	-0.05	**0.70**	-0.29	-0.18	0.42	-0.13	-0.06	-0.07
Geological - SubRockType - Non-marine sedimentary	0.43	0.06	-0.01	-0.10	**0.88**	-0.05	-0.03	-0.04	-0.02	-0.02
Geological - Period - Carboniferous	0.43	0.06	-0.01	-0.10	**0.88**	-0.05	-0.03	-0.04	-0.02	-0.02
Climate - Days with rainfall >30mm	0.64	-0.07	-0.04	-0.03	**0.72**	-0.08	0.05	-0.11	-0.08	-0.05
Geological - Cambrian-Ordovician	0.05	0.10	0.01	-0.01	-0.02	**0.98**	0.04	-0.02	-0.01	0.00
Geological - SubRockType - Oceanic domain miogeoclinal	0.05	0.10	0.01	-0.01	-0.02	**0.98**	0.04	-0.02	-0.01	0.00
River property - Index described in Dionne et al. (2008)	0.32	-0.04	0.41	-0.14	-0.18	**0.61**	0.09	0.06	-0.24	-0.15
Geological - Period - Devonian	0.38	-0.01	0.14	-0.08	-0.27	-0.14	**0.81**	-0.24	-0.02	-0.03
Geological - Province - Apalachian Orogen	0.53	-0.08	0.11	-0.14	0.19	0.33	**0.69**	0.19	0.02	-0.03
Geological - Period - Mesoproterozoic	-0.06	0.06	0.27	-0.44	-0.25	-0.15	**-0.58**	-0.20	0.38	-0.18
Geological - SubRockType - Mix	0.02	-0.32	0.02	-0.06	-0.01	0.03	0.13	**0.87**	0.10	0.02
Geological - Period - Ordovician	0.31	-0.33	-0.05	0.12	-0.12	-0.03	-0.12	**0.72**	-0.11	-0.05
Geological - RockType - Intruded	-0.14	0.21	0.10	-0.07	0.06	-0.06	-0.04	0.08	**0.78**	0.03
River property - Level of exploitation	0.39	-0.13	0.12	0.34	-0.17	-0.03	0.07	-0.13	**0.49**	-0.15
Geological - SubRockType - Undivided gneiss	-0.21	-0.30	-0.06	-0.32	-0.10	-0.10	-0.35	-0.27	-0.08	**0.69**
Geological - Period - Paleoproterozoic-Mesoproterozoic	-0.46	0.22	-0.24	-0.19	0.10	-0.13	0.07	0.08	-0.42	**0.56**
Geological - SubRockType - Orthogneiss	-0.39	-0.18	0.27	-0.10	0.08	-0.04	-0.14	-0.12	-0.20	**-0.71**

4.9 Figures

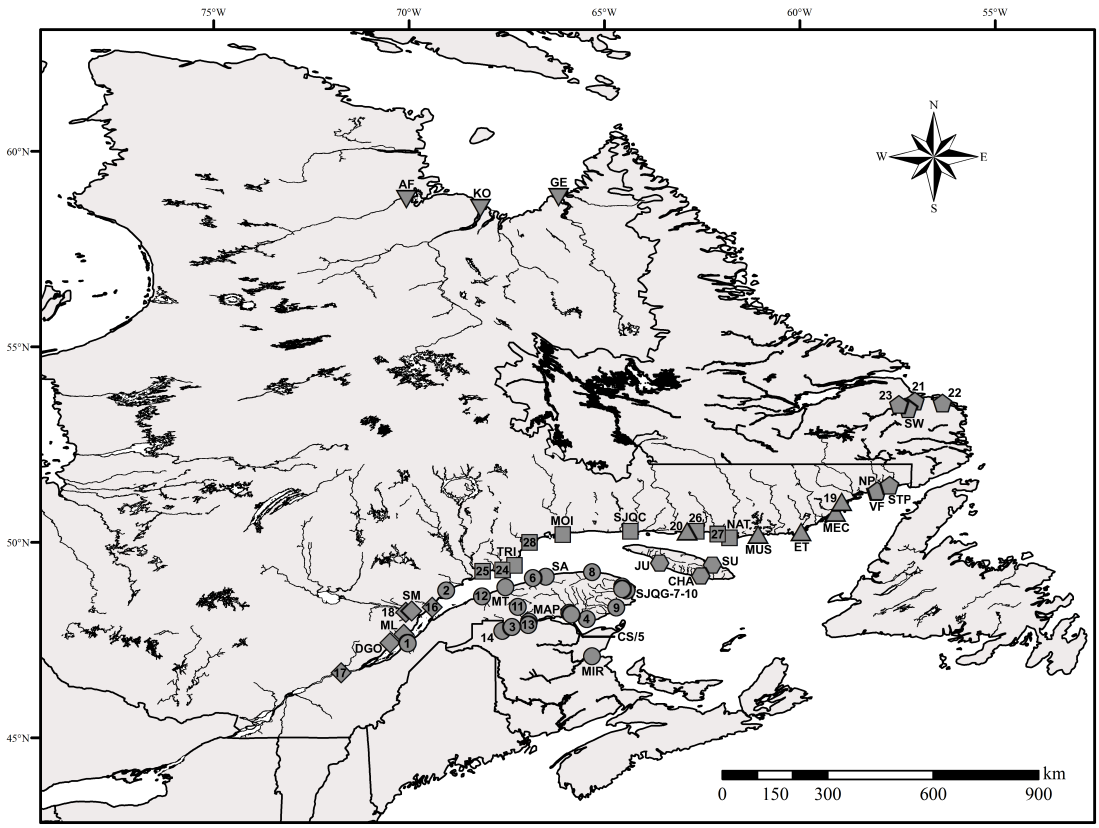


Figure 4.1 Map showing sample sites. Populations are linked to the river codes in Table 1 for individually genotyped populations (n = 26) and numbers in Table 3 for bulk assays. Tick forms relate to regional groups as reported in Dionne et al. (2008).

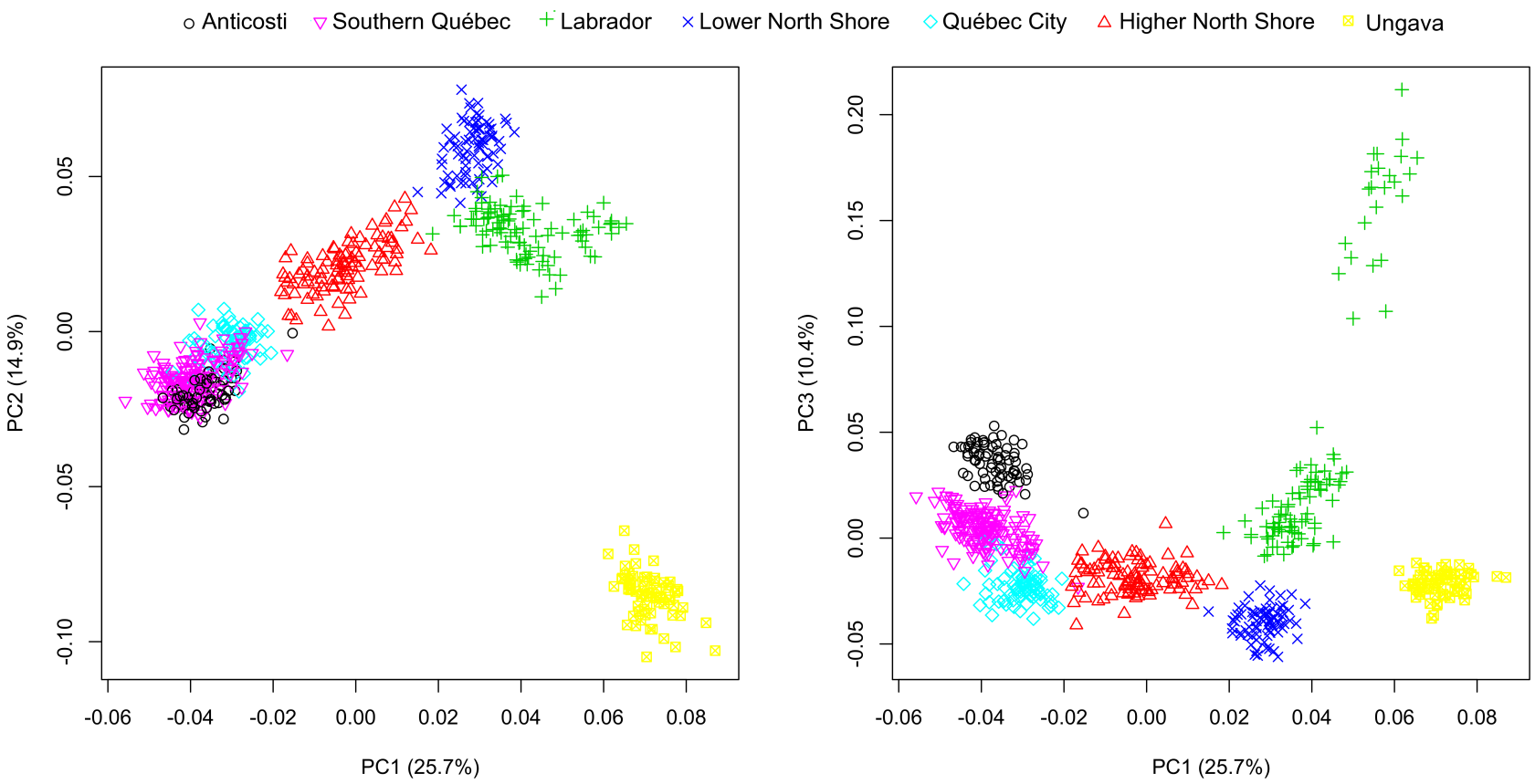


Figure 4.2 Principal components analysis of genetic differentiation among individuals based on 3016 SNP markers (each point represents one individual) with principal component 1 (PC1: 25.7% of variance) against PC2 (14.9% of variance) on the left panel and PC1 against PC3 (10.4% of variance) on the right panel. Color and tick reflect population's regional groups as reported in Table 1.

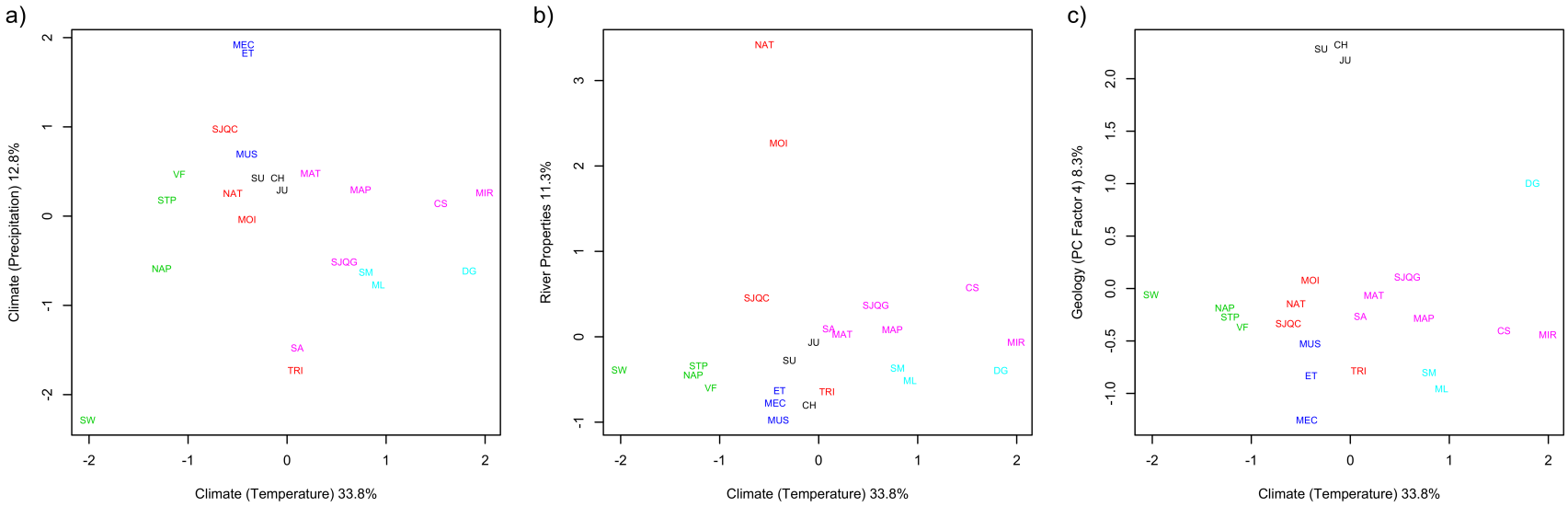


Figure 4.3 Principal components analysis of environmental parameters (n = 49) among populations (n = 23) with: a) Climate related PC factor 1 (temperature; 33.8% of variance) against Climate related PC factor 2 (precipitation; 12.8% of variance); b) PC factor 1 against PC factor 3 (river properties; 11.3% of variance); c) PC factor 1 against PC factor 4 (geology; 8.3% of variance). Population locations on the spatial axes are marked by their code name and colors reflect population's regional groups as reported in Table 1.

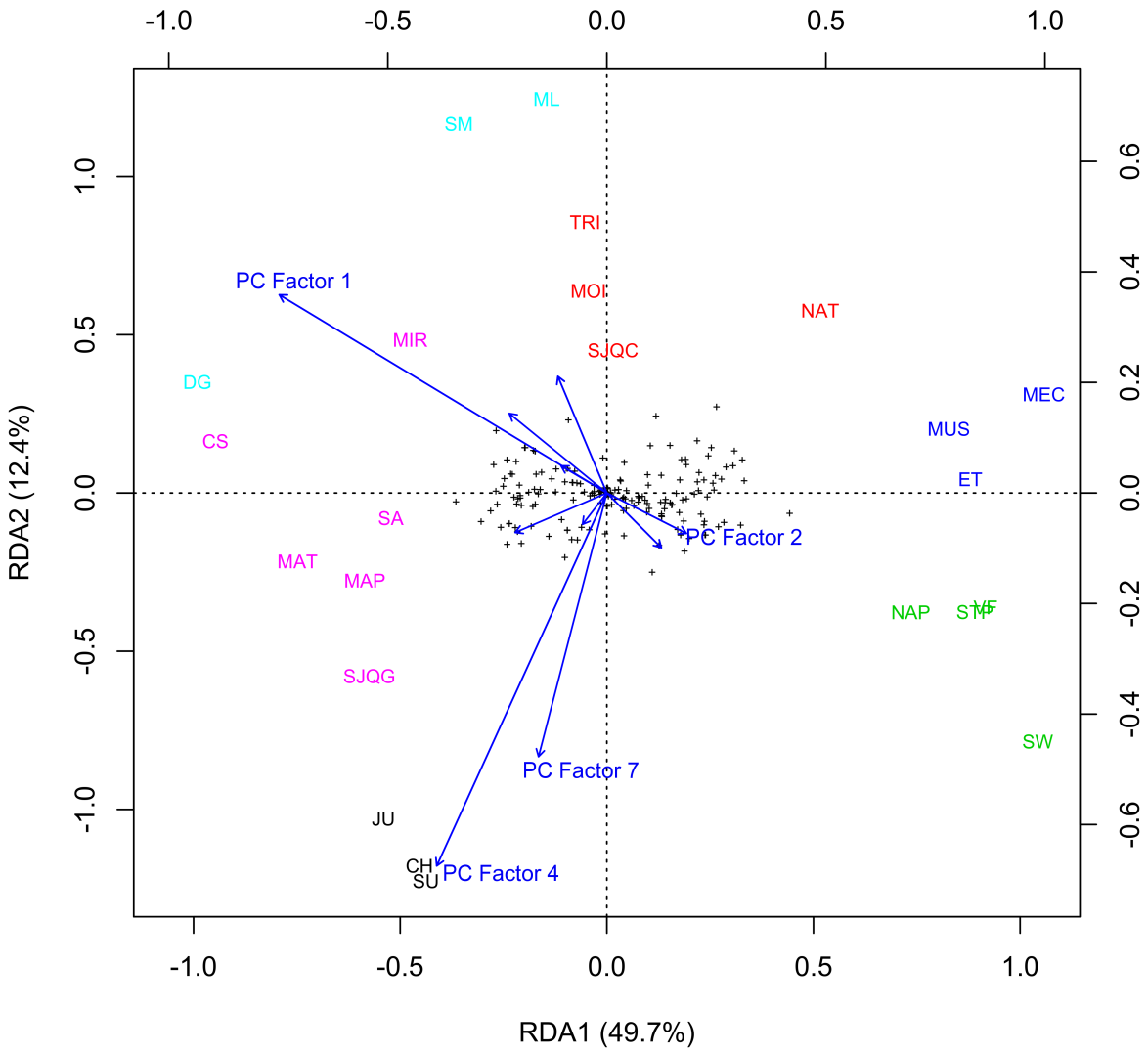


Figure 4.4 Redundancy analysis axes 1 (49.7% of variance) and 2 (12.4% of variance) showing the position of allele frequency vectors for the 179 SNP markers potentially under divergent selection at the 0.05 significance level (plus marks) and related environmental PC factors as blue arrows. Only environmental PC factors significantly associated with genetic markers are identified (p-values < 0.001). Markers' positions relate to scales on the bottom and left axes; Environmental PC factors positions relate to scales on top and right axes. Population locations on the spatial axes are marked by their code name and colors reflect population's regional groups as reported in Table 1.

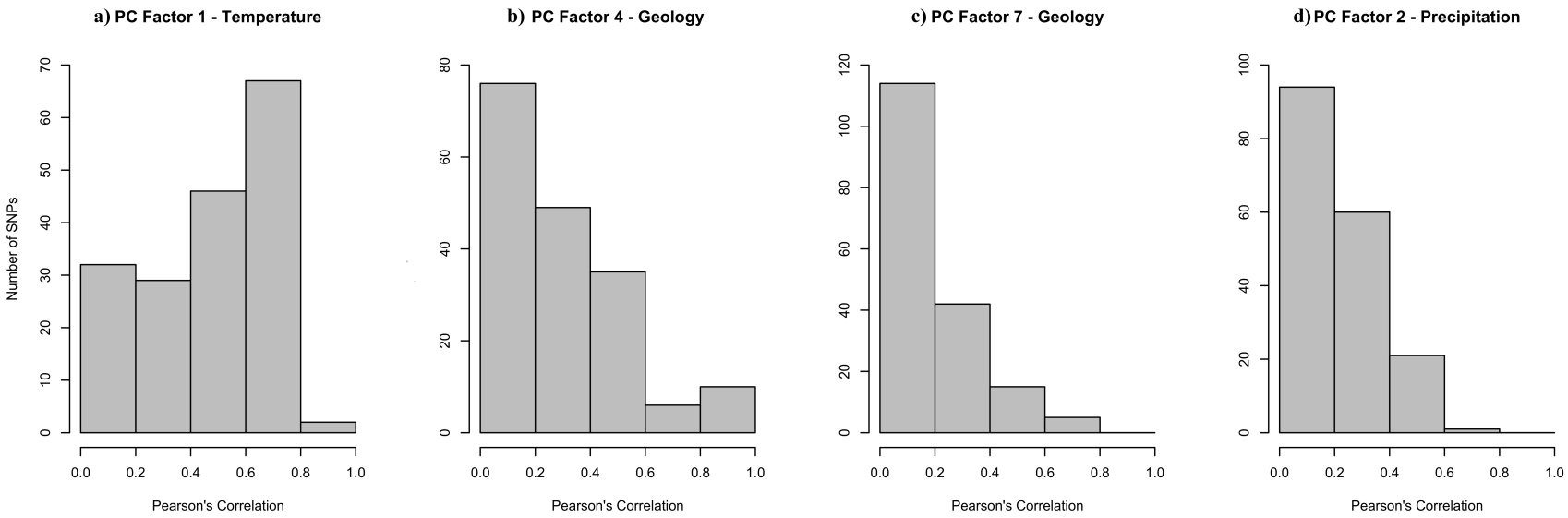


Figure 4.5 Pearson's correlation coefficient distribution for the 179 SNP markers potentially under divergent selection at the 0.05 significance level when correlated with environmental PC factors significantly associated with genetic markers (p-values < 0.001). Presented in order of decreasing F-static value of ANOVA from left to right where: a) PC factor 1 (climate-temperature), b) PC factor 4 (geology), c) PC factor 7 (geology) and d) PC factor 2 (climate-precipitation).

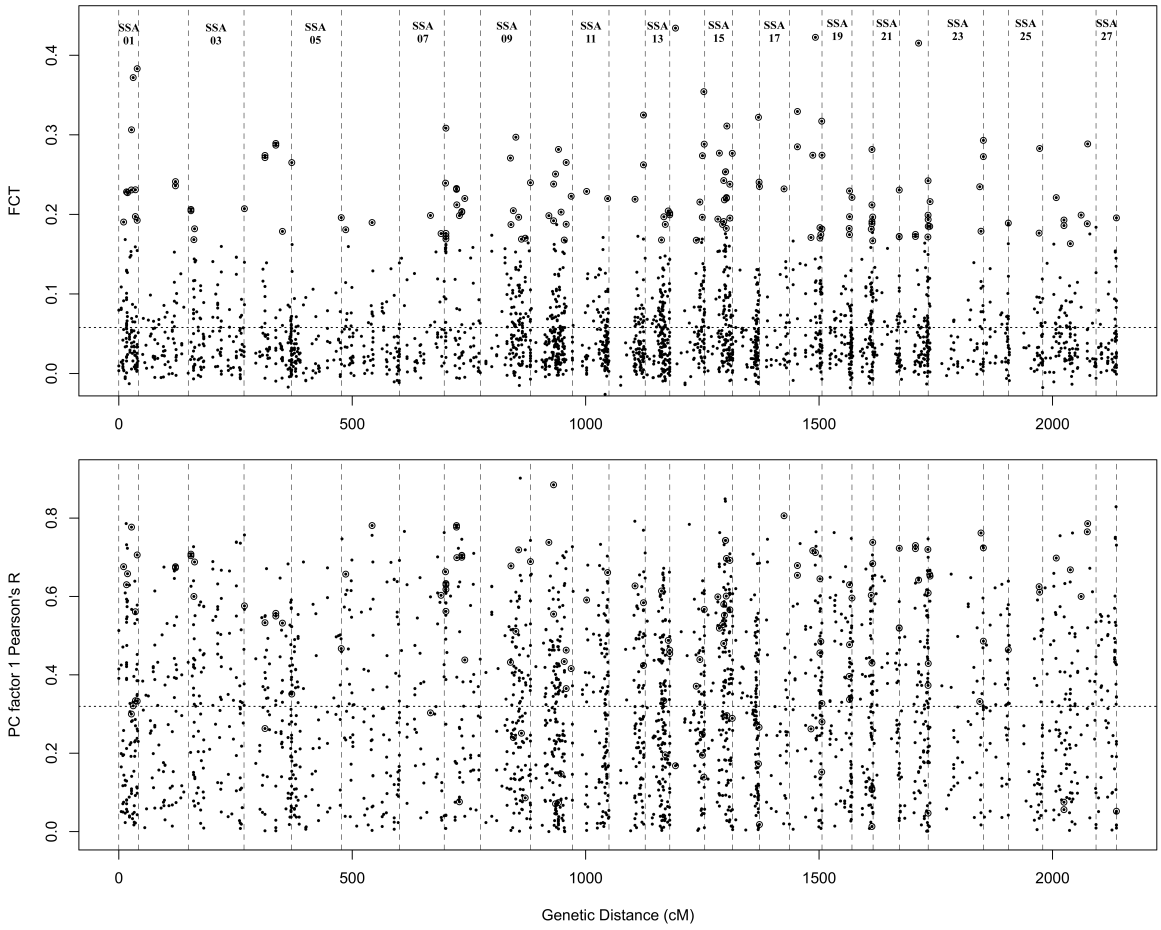


Figure 4.6 Genetic linkage map showing the distribution of regional differentiation (F_{CT}) of each SNP marker (n = 3118) on the top panel and Pearson's correlation coefficients (R) related to PC factor 1 (climate-temperature) on the bottom panel. Gray and white rectangles separated by vertical dashed lines represent linkage groups (named SSA--). On the top panel, circled dots indicate outlier markers (significance level p = 0.05) and the horizontal dotted line indicates the average F_{CT} among markers (0.058). On the bottom panel, circled dots indicate outlier markers (significance level p = 0.05) and the horizontal dotted line indicates the average Pearson's correlation coefficient (R*)* among markers (0.320).

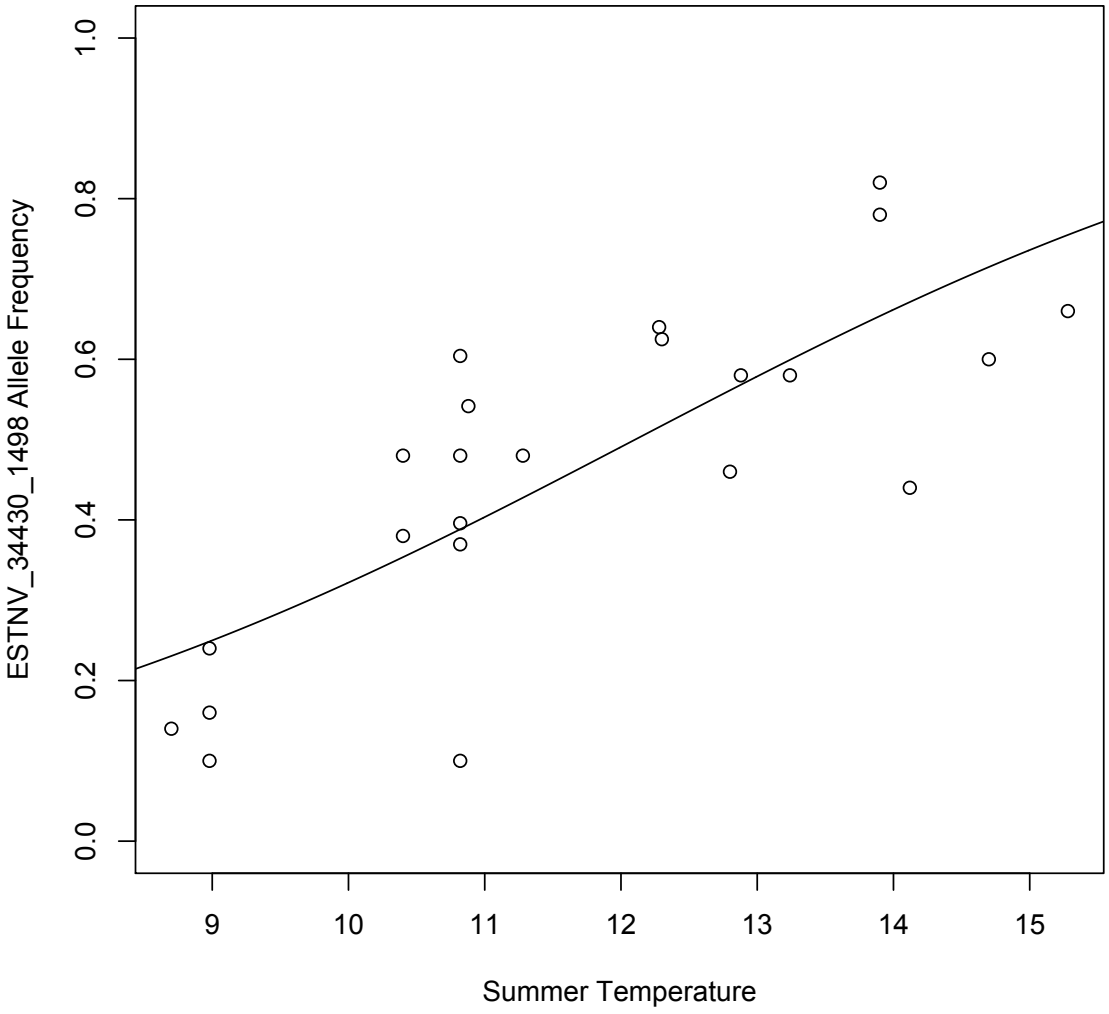


Figure 4.7 Generalized linear models illustrating the relation between marker population allele frequency and average seasonal temperatures between May and September.

4.10 Supplementary material

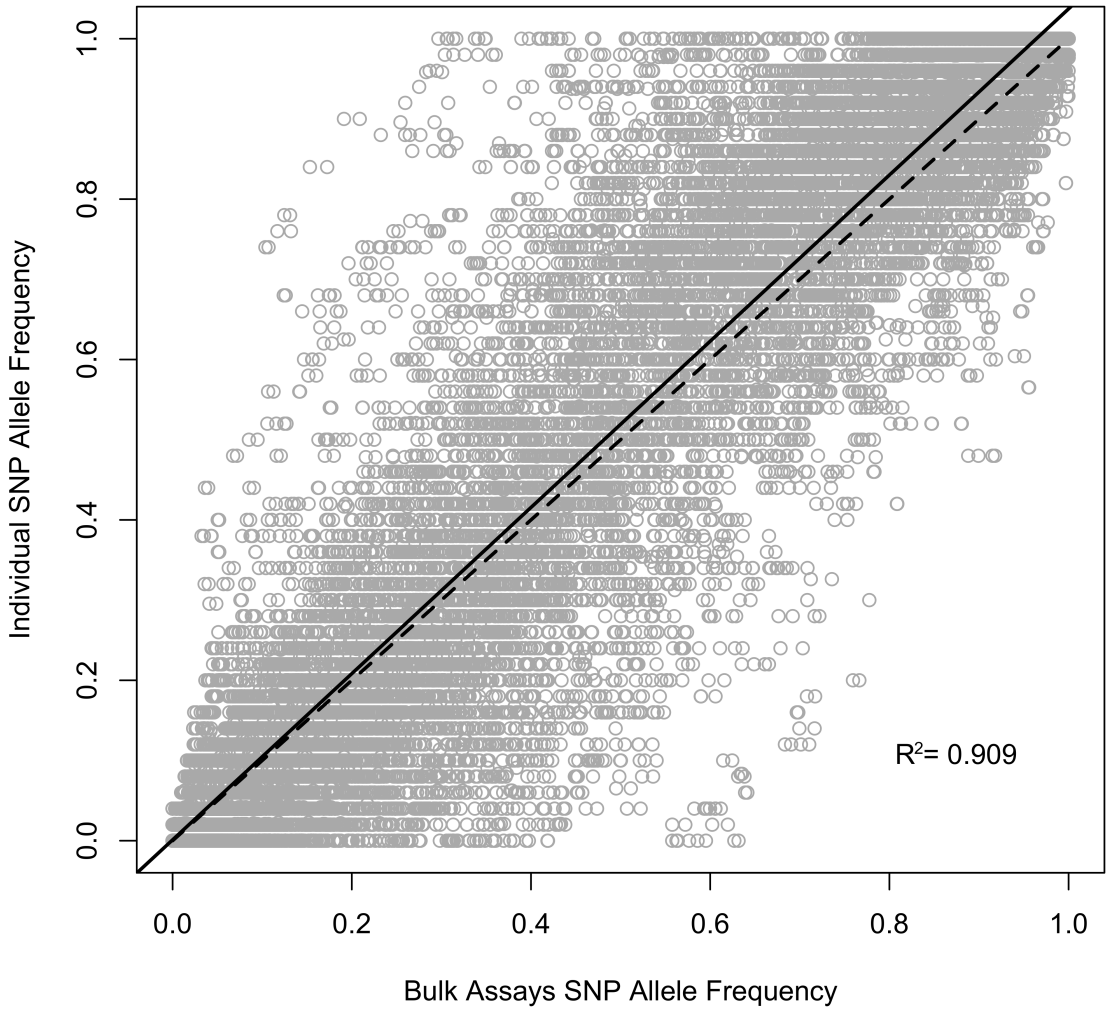


Figure 4.S1 Within population allele frequencies estimated from bulk assays for 2831 SNPs on the x-axis and corresponding allele frequencies for the same populations (n = 5) as measured by individual genotyping of the same individual (n = 25 individuals per population). The solid line indicates the fit of the regression (p-value < 0.001) and the dash line shows the position of a hypothetical perfect correspondence of both estimates (x = y).

Hereafter listed supplementary tables are available online or on demand:

Table 4.S1 Single nucleotide polymorphism (SNP) markers' categories.

Table 4.S2 Single nucleotide polymorphism (SNP) markers observed (H_O) and expected (H_E) heterozygosities per population.

Table 4.S3 Summary of the detection of markers potentially under selection using hierarchical F_{dist} genome scans implemented in ARLEQUIN 3.5 (Excoffier and Lischer 2009).

Table 4.S4 Pairwise measures of genetic differentiation (F_{ST}).

Table 4.S5 Blast results from BLAST2GO and gene ontology (GO) terms annotation.

Table 4.S6 Enrichment analysis results testing over- or under-representation of gene ontology (GO) terms.

Chapitre 5 Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multi-locus analysis surpasses genome scan

Soumis sous :
Bourret V, Dionne M & Bernatchez L (2014) *Molecular Ecology*

5.1 Résumé

Au cours des 30 dernières années, les populations de saumon atlantique ont subi un déclin global généralisé. Les causes de ce déclin sont vraisemblablement nombreuses et complexes, mais l'un des principaux facteurs identifié est l'augmentation de la mortalité en mer. Cependant, la phase de vie en mer du saumon atlantique demeure méconnue. L'examen des changements potentiels au niveau de la composition génomique des populations pendant la migration en mer détient le potentiel de mettre en évidence certains facteurs influençant la mortalité en mer. Dans ce contexte, nous avons génotypé 5568 SNPs chez deux populations de saumon atlantique représentant deux régions génétiques distinctes pour deux cohortes. Pour ces dernières, nous avons cherché à identifier des différences de fréquences alléliques et génotypiques entre les juvéniles (saumoneaux) quittant leur rivière natale et les adultes (madeleineaux) revenant en eau douce après une année en mer. En raison de la complexité des traits potentiellement associés à la mortalité en mer, nous avons comparé les résultats obtenus à l'aide d'une méthode de balayage génomique basée sur la différenciation génétique de marqueurs individuels et une nouvelle méthode multilocus afin de mettre en évidence de la mortalité différentielle ayant une base génétique. En ce qui concerne la méthode de balayage génomique aux loci individuels, bien que de nombreux marqueurs aient été identifiés comme étant potentiellement sous l'effet de la sélection naturelle, aucune évidence de sélection en parallèle ou répétée au niveau temporel n'a été trouvée. Au contraire, l'approche multilocus a permis de détecter un patron de sélection répété sur deux cohortes pour un groupe de 34 marqueurs SNPs dans une des deux populations. Aucun patron de mortalité sélective n'a été détecté dans l'autre population, suggérant différentes causes de mortalité en mer parmi les populations. Globalement, ces résultats supportent l'hypothèse que la sélection cause principalement de petits changements de fréquences alléliques pour plusieurs loci covariant plutôt qu'un petit nombre de changements à effet majeur. Conséquemment, une approche génomique multilocus offre possiblement un cadre méthodologique mieux adapté que le paradigme du balayage sélectif pour l'étude des signatures génomiques de la sélection naturelle et particulièrement pour des traits complexes dans des populations naturelles.

5.2 Abstract

Wild populations of Atlantic salmon have declined worldwide. While the causes for this decline may be complex and numerous, increased mortality at sea is predicted to be one of the major contributing factors. Examining the potential changes occurring in the genome-wide composition of populations during this migration has the potential to tease apart some of the factors influencing marine mortality. Here, we genotyped 5568 SNPs in Atlantic salmon populations representing two distinct regional genetic groups and across two cohorts to test for differential allelic and genotypic frequencies between juveniles (smolts) migrating to sea and adults (grilses) returning to freshwater after one year at sea. Given the complexity of the traits potentially associated with sea mortality, we contrasted the outcomes of a single-locus F_{ST} based genome scan method with a new multi-locus framework to test for genetically-based differential mortality at sea. While numerous outliers were identified by the single-locus analysis, no evidence for parallel, temporally repeated selection was found. In contrast, the multi-locus approach detected repeated patterns of selection for a multi-locus group of 34 co-varying SNPs in one of the two populations. No significant pattern of selective mortality was detected in the other population, suggesting different causes of mortality among populations. These results first support the hypothesis that selection mainly causes small changes in allele frequencies among many co-varying loci rather than a small number of changes in loci with large effects. They also point out that moving away from the a strict "selective sweep paradigm" towards a multi-locus genetics framework may be a more useful approach for studying the genomic signatures of natural selection on complex traits in wild populations.

5.3 Introduction

In the past decades we have witnessed a range-wide decline in wild Atlantic salmon populations. Despite the ban on commercial fishing, the majority of natural stocks are still declining. In North America, this decline has been characterized by a remarkable reduction in marine survival (Gibson et al. 2011; ICES 2013). Habitat degradation, the impacts of the salmon farming industry and climatic changes are commonly invoked to explain much of the decline in the species as a whole (Hansen & Quinn, 1998). The Atlantic salmon is an anadromous species that usually spends one or more years in freshwater before the juveniles physiologically prepare for seaward migration via the process of smoltification and then migrate towards their marine feeding grounds off the coast of Newfoundland-Labrador (for North American populations) and Greenland (for both North American and European populations). After one or more years at sea, adults return to reproduce in their natal river (Stabell 1984). Studying Atlantic salmon during the marine phase is critical for determining the reasons for population declines, but has proven very difficult. Therefore, we still know very little about this species' life in the marine environment. However, multiple monitoring studies have shown a sharp decline in sea survival indexes, suggesting that mortality at this life-history stage is contributing to the observed population decline and needs to be better studied (Dionne et al. 2013; ICES 2013).

The marine environment is challenging to salmon for a number of reasons, such as a higher risk of predation and the need to undertake a long distance migration. In fact, the seaward migration of Atlantic salmon is one of the most deadly phases of this species' life cycle as the bulk of mortality is thought to occur when entering the marine environment (Friedland et al. 2000; Thorstad et al. 2012). Differential survival at sea is most likely to involve a complex variety of traits associated with predator avoidance, osmoregulation, growth and even disease resistance as reported in previous studies on migratory salmon (Miller et al. 2011). Moreover, the accentuated decline over the last 30 years suggests that unusual or new selective forces may be acting more strongly on Atlantic salmon relative to the past. In addition, variable return rates vary among salmon from different geographic regions suggests differential mortality at sea among distinct populations (Dionne et al. 2013). Investigating genomic changes occurring during the marine phase of salmon may help to

determine whether reduced survival at sea is occurring randomly or is due to specific selective pressures.

From a fundamental perspective, disentangling neutral and selective changes occurring at the genome level is a crucial yet challenging goal in evolutionary biology. When it comes to identifying the specific targets of selection, some methods have received much criticism. Nevertheless, despite the numerous pitfalls and maladapted models extensively discussed in recent years (Bierne et al. 2011; Narum & Hess 2011; Vilas et al. 2012; De Mita et al. 2013; Lotterhos & Withlock 2014), methods based on the identification of markers with unusually high levels of genetic differentiation (e.g. genome scans) have proved to be successful in many situations. On the other hand, as reported in a growing number of quantitative genetic studies (Mackay et al. 2009; Hancock et al. 2010; Yang et al. 2010; Nadeau et al. 2011; Daub et al. 2013), outlier detection is essentially restricted to the investigation of single loci or genomic regions of large effects. However, since selection is expected to generate linkage disequilibrium between loci for complex quantitative traits, relatively low and subtle allelic changes among co-varying loci is expected to yield a combined effect greater than individual loci on phenotypes (Latta 1998; Le Corre & Kremer 2003). Therefore, it might be time for a paradigm shift from the quest for a strong selective sweep on individual loci to the consideration of polygenic soft sweeps when investigating potential targets of selection at the genomic level (Le Corre & Kremer 2012).

Although theory predicts that a beneficial new allele or standing genetic variation can be swept by a rapid shift in frequency caused by selection, Pritchard and Di Rienzo (2010) have argued that reducing adaptation to "selective sweeps at key loci is too limited". Obviously, there is a giant leap between models based upon selective sweeps and those which model polygenic adaptation, and fortunately, they are not mutually exclusive as both scenarios are probably occurring in nature. However, regardless of the traits investigated, recent studies have relied on single-locus genome scans to find selective sweeps, even if that means surfacing with a large number of false positive targets. While selective sweep models might be effective for simple traits, quantitative genetic theory makes a strong case

for polygenic adaptation in complex traits, as best exemplified by the genetic study of human height (Yang et al. 2010).

In this study our main goal was to test expectations of single-locus and multi-locus models on a complex trait such as marine survival that is predicted to be under recently occurring, novel selective pressures caused by recent environmental changes in the Atlantic Ocean (Mills et al. 2013). Over the past 30 years, two rivers in the Canadian province of Québec have been annually monitored to evaluate the evolution of intrinsic characteristics, abundance and rates of freshwater survival and adult returns in Atlantic salmon (Dionne et al. 2013). Every year, young Atlantic salmon leaving the rivers for the first time (called smolts) and adults returning after one (called grilse) or more years at sea (called multi-sea winter) are captured. This system offers a unique opportunity to follow spatially and temporally replicated cohorts of Atlantic salmon and to examine the genetic changes occurring during the marine phase of their life cycle. Here, we take advantage of this unique system to investigate the genomic changes occurring due to natural selection during the first year at sea in Atlantic salmon. This was achieved by contrasting single-locus F_{ST} based methods and a customized polygenic multi-locus approach to identifying genomic targets of selection and test for spatial and temporal variation in patterns of selection.

5.4 Materials and Methods

5.4.1 Samples

Fin clips were collected and preserved in 95% ethanol for 25 Atlantic salmon smolts leaving freshwater for the sea. Smolts were sampled across the peak of outmigration from freshwater to the marine environment in May and June of 2004 and 2005 in the Saint-Jean River, in Gaspésie, Québec, Canada and de la Trinité River, in the North Shore region of Québec, Canada. In addition to these 100 smolts, 100 fins clips were then collected from 25 adult Atlantic salmon returning after one year at sea (hereafter named grilses) in 2005 and 2006 from both rivers (Figure 1). These two rivers represent two previously identified regional genetic groups of salmon (Dionne et al. 2008) shown to be potentially locally adapted to distinct environmental conditions (Bourret et al. 2013b). Also, different return rates from the sea have been reported between salmon from these two regions, indicating that they suffer differential mortality at sea, which may indicate differences in the strength or identity of selective forces (Dionne et al. 2013). Smolts were captured using rotary traps at 8 km and 9 km upstream of the river mouth in the Saint-Jean and Trinité rivers, respectively. Adults were collected over the entire return migration period, from local anglers in the Saint-Jean River and from a fish ladder located at the river mouth in the Trinité River. For all samples total length, fork length and weight were measured, and scales were also collected to determine age. DNA was extracted from fin clips using the QIAGEN DNeasy Tissue Kit following the manufacturer's guidelines (Qiagen, Valencia, CA). Samples were sexed using a molecular sexing technique (Yano et al. 2012).

5.4.2 Genotyping Quality Control

Genotyping for all 200 DNA samples was performed on 5568 SNPs with the SNP-array developed by the Centre for Integrative Genetics (CIGENE, Norway) and reagents from the Illumina infinium assay (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Detailed methods for SNP discovery and quality control can be found in Bourret et al. (2013ab). The quality of individual samples was assured by only using individuals genotyped at a greater than 0.85 call rate (*CR*: proportion of genotyped SNPs*)*. Markers with minor allele frequencies less than 1% (MAF < 0.01) and markers missing in more than 1% of individuals were excluded from our analyses. Ascertainment bias was

assessed by Bourret et al. (2013a) and was shown to be minimal when comparing North American populations.

5.4.3 Signatures of selection

5.4.3.1 Single locus genome scan

To contrast single-locus and multi-locus approaches for detecting selection during life at sea in Atlantic salmon grilses, we first performed a single-locus F_{ST} outlier approach using LOSITAN (Antao et al. 2008). LOSITAN is a selection detection workbench based on the Fdist method of Beaumont and Nichols (1996). For each cohort from each river, we identified F_{ST} outliers between smolts and grilses for a total of four runs of the program. Each run was carried out using the recommended calculation of "Neutral mean F_{ST}", which removes potentially selected loci from the calculation of the mean F_{ST} after a first run of simulation and using the "Forcing mean F_{ST}" option. We used 50 000 simulations, at a significance level of 0.01. Markers identified as potentially under divergent selection and with a reported heterozygosity > 0.1 were then compared between cohorts in a given river, between rivers in a given cohort and overall to detect any significant parallel patterns of selection between cohorts and/or rivers for non-random selective mortality at sea.

5.4.3.2 Polygenic multi-locus approach

Here, the number of individual markers was reduced into sets of independently, co-varying groups of markers based on their genotypic distribution. For each marker, we coded individual genotypes as 0, 1 and 2, which referred to the count of one allele for each SNP. A principal component analysis (PCA) was first performed on the retained set of markers (see Results) for all 200 individuals using the PROC FACTOR and VARIMAX rotation procedure implemented in SAS 9.3 (SAS Institute Inc). The set of markers was then reduced to a number of factors identified using the parallel analysis (PA) criterion (Horn 1965). The PA criterion is a Monte-Carlo based simulation method that compares the observed eigenvalues with those obtained from uncorrelated normal variables. It is based on the rationale that eigenvectors from the dataset should have larger eigenvalues than parallel eigenvectors from random datasets with the same sample size and number of variables (Ford et al. 1986; Lautenschlager 1989). A factor or component was retained if

the associated eigenvalue was bigger than the 99th of the distribution of eigenvalues derived from a random data. The PA was then performed on 1000 random datasets (bootstraps) and a significance level of 0.01 to determine the number of factors to be retained in subsequent analyses. We could then determine the specific markers significantly correlated with a given principal component using a conservative significance level of 0.001, which corresponds to an absolute loading weight superior to 0.23 on a given factor.

In the second step, a canonical discriminant analysis based on the retained factor's scores was performed to examine the power to distinguish individuals from rivers, cohorts and life-stages using the PROC CAN procedure implemented in SAS 9.3. For statistically significant canonical axes, we then performed a generalized linear model to identify the sources of discrimination using the PROC GLM procedure in SAS 9.3. The model considered rivers, cohorts, life-stages, interactions among rivers and cohorts, and interactions among rivers and life-stages as possible sources of correlation with the canonical axis. We then determined the PCA factors significantly correlated with either discriminant canonical axis using a conservative significance level of 0.001, which corresponds to an absolute loading weight superior to 0.23 on a given factor, as mentioned above. These factors thus represented independent multi-locus factors potentially under selection according to the source of variation.

5.4.4 Covariance of allelic effects

Based on quantitative genetics theory, we predicted positive covariance (i.e. linkage disequilibrium) to have built up across loci under selection and associated with survival at sea (Latta 1998; Le Corre & Kremer 2003). We therefore estimated the covariance of allelic effects across populations and within each population for all pairwise combinations of SNPs. The covariance of allelic effects was calculated based on Equation 2 of Ma et al. (2010): $4a_i a_j D_{ij}$; where a_i and a_j are the additive effects of loci i and j, respectively and D_{ij} is the linkage disequilibrium (LD) between the two loci. As a surrogate of the scale of additive effects for each locus on survival at sea, we used the absolute loading weight on identified PC factors correlated with the canonical axis, which presented life-stage as a source of discrimination among individuals (see Results). LD was calculated using the

locus pairwise multi-allelic *D* in POWERMARKER (Liu & Muse 2005). We then compared the distribution of allelic effect covariance across the two populations for loci loading on significantly correlated PC factors potentially under selection to the distribution of mean covariance of allelic effect for a similar number of marker randomly chosen among all markers 1000 times. All of these analyses were performed in R (R Development Core Team).

5.4.5 Strength of selection

As a surrogate for fitness estimates, we thus used delta p (Δp), which is the absolute difference in frequency of the major allele between smolts and grilses. The allele frequency change between life-stages can be a direct response to differential survival, and in such as case, measuring Δp for a given marker is related to estimating its fitness values (Barrett et al. 2008). Therefore, to measure the strength of selection on a targeted group of markers we compared the distribution and mean of Δp for loci identified with the multi-locus approach to those loci identified as outliers in the single-locus outlier detection and the distribution for retained markers. Those comparisons were made by considering both rivers together and separately as well as cohorts pooled within each river. To localize potential genomic regions associated with mortality at sea, we used a linkage map of North American Atlantic salmon (Brenna-Hansen et al. 2012) to position the distribution of Δp as well as the loading weights of all markers according to their chromosomal distances on the 27 Atlantic salmon linkage groups. These analyses were all performed in R (R Development Core Team).

5.4.6 Gene ontology and SNP annotation

Blast2go (Gotz *et al.* 2008) was used to associate gene ontology (GO) annotation terms to all SNPs. Homology search was first conducted with a BLAST (Altschul et al. 1990) search of the available flanking sequences for each SNP on the NCBI nr public database with the *e*-value threshold set to $1x10^{-10}$. To determine if the biological processes, molecular functions or cellular components were over-, equally or under-represented among outlier markers or SNPs reported on a multi-locus group potentially under selection, we performed enrichment analyses using Fisher's Exact Test corrected for multiple tests by applying a false discovery rate of 0.05 (FDR; Benjamini & Hochberg 1995).

5.5 Results

5.5.1 Genotyping and quality control

All individuals had call rates > 95% so no fish were excluded from our analyses. After initial quality control of genotypes, we found that 3380 SNPs were polymorphic and reliable. Of these 3380, 457 markers were discarded either because they showed an overall MAF < 0.01 or were missing genotypes at a proportion greater than 0.01. Therefore, from 5568 SNPs on the SNP-array, 2923 SNPs from all 200 individuals were used in our analyses.

5.5.2 Signatures of selection

5.5.2.1 Single-locus genome scan

Many outliers were identified in each of the pairwise comparisons between smolts and grilse for each cohort in each river (Table S1). However, we found no evidence for parallelism among them. In the Saint-Jean River, 34 SNPs were identified as outliers in the 2004 cohort (F_{ST} range: 0.082 - 0.231) and 42 outliers were identified in the 2005 cohort (F_{ST} range: 0.093 - 0.198). Among those outliers potentially under divergent selection, only one (GCR_cBin9358_Ctg1_120) was found in both 2004 and 2005 cohorts. In the Trinité River, 34 (2004; F_{ST} range: 0.085 - 0.180) and 20 (2005; F_{ST} range: 0.096 - 0.184) SNPs were identified as outliers and none were the same in both cohorts. Two markers (ESTNV_22435_210, ESTV_16844_203) out of 64 were found to be parallel divergent outliers in the 2004 cohort between the Saint-Jean and Trinité populations, whereas one SNP (ESTNV_34235_752) out of 62 was parallel between both populations for the 2005 cohort when comparing smolts and grilses. No outliers were common to more than two comparisons. A summary of all four F_{ST} based genome scans is presented in Figure 2.

5.5.2.2 Polygenic multi-locus approach

Based on the parallel analysis (PA) criterion, 48 principal component factors (PC factors) were kept for subsequent analyses (Table S2). These PC factors accounted for 41% of the variation observed among genetic markers and the top three factors each accounted for at least 1% (PC1 = 5.06%; PC2 = 1.05%; PC3 = 1.00%). The first two axes of the canonical discriminant analysis displayed significant canonical correlations for populations, cohorts

or life-stages ($p < 0.001$; Table 1). Figure 3 illustrates the spatial distribution of individuals along these two axes. Generalized linear models indicated that population of origin (Saint-Jean *vs*. Trinité) was the only source of variation discriminated on the first axis ($F = 16\,543$; $P < 0.001$; Table 2a). Among the 48 PC factors initially retained, only PC factor 1 significantly associated with the first axis of canonical correlation with a loading weight of 0.418.

The second axis contained three significant (Table 2b) sources of variation. In particular, the interaction term between populations and life-stages ($F = 135.33$; $P < 0.0001$), life-stages ($F = 98.52$; $P < 0.0001$), and more marginally the cohort (2004 *vs*. 2005; $F = 12.24$; $P = 0.0006$) were correlated with PC factors loading on this second axis. Therefore, we found genetic differences between life-stages which are repeated in two independent cohorts but the interaction term indicates that this effect differs between populations. Indeed, the second axis clearly discriminates smolts and grilses from the Saint-Jean population but not from the Trinité population, a pattern repeated for both cohorts. One PC factor (PC factor 40) was also found to be significantly associated with the second axis of canonical correlation. PC factor 40 presented a canonical loading weight of 0.320 on which 34 markers have PC loading weights > 0.23 (Table S3). Therefore, these markers are significantly associated with the discriminant axis differentiating life-stages in the Saint-Jean population. Among these 34 markers, only two were also divergent outliers in the single-locus genome scans (ESTNV_22435_210 and ESTV_16844_203).

5.5.3 Allelic effects and strength of selection

The quantile-quantile plot clearly demonstrates that the distribution of allelic effect covariance among markers significantly correlated with PC factor 40 is skewed towards higher values compared to a random set of markers (Figure 4). Thus, except for an outlier point at the end of the distribution, the upper range of the distribution of values for multi-locus markers was more than three times higher than that obtained for random markers. Consequently, the markers associated with genetic differences between life-stages in the Saint-Jean River displayed significantly greater covariance in allelic effect than expected by chance, further confirming their association with different life-stages repeated for two

independent cohorts. Also, the distributions of Δp among all 2923 markers was skewed towards low values with a very high proportion of values between 0 and 0.02 and exponentially reducing in proportion for higher values up to 0.36 in cohort 2 (2005) of the Saint-Jean River and identical medians of 0.030 for both rivers (Figure 5a). We then found that markers identified with the multi-locus approach had Δp distributions, which were more uniform with medians of 0.045 and 0.030 for the Saint-Jean and Trinité rivers, respectively, translating into modest change in frequencies for these markers (Figure 5b). These small changes on a set of co-varying SNPs are consistent with the polygenic selection hypothesis acting on many genomic targets and causing small changes of allele frequencies. In contrast, the single-locus genome scan identified markers with medians of 0.090 and 0.075 for the Saint-Jean and Trinité rivers, respectively (Figure 5c), which is consistent with single large effect markers or genes putatively being highlighted by single locus methods. The genomic distribution of Δp and loading weight presented a widely spread distribution of values along the different linkage groups (Figure 6). The 34 markers identified using the multi-locus approach were distributed among 14 different linkage groups with a maximum of 4 SNPs being observed on a single linkage group (Figure 6).

5.5.4 Gene ontology and SNP annotation

The BLAST and annotation steps in Blast2go yielded annotations for 1119 SNPs. Using a FDR of 0.05, an enrichment analysis at the significance level of 0.05 did not indicate the significant over- or under-representation of any biological pathway among the markers under divergent selection compared to all other markers in any of the individual genome scans or for the 130 divergent outliers from all genome scans taken as a test group. Similarly, the enrichment analysis did not reveal any over- or under-representation of biological functions or processes for the 34 markers identified using the multi-locus approach compared to all other markers. Nevertheless many annotations were associated with mitochondrial functions. Annotations and GO-terms identified for the 34 markers significantly correlated with PC factor 40 are reported in Table S4.

5.6 Discussion

5.6.1 Detecting selective mortality – single-locus genome scan versus polygenic multi-locus approach

We assessed for the first time whether non-random, genetically-based selection could contribute to differential mortality during the first year at sea of salmon. We used a population genomic approach and sampled two genetically distinct Atlantic salmon populations by comparing outmigrating smolts and returning adults of the same cohorts. In doing so, we also compared the outcomes of a single locus outlier approach vs. polygenic multi-variable statistical method in detecting repeated (parallel) signals of selection at the genome level.

Both the single and multi-locus approaches identified markers potentially under divergent selection during the first year at sea, but only the multi-locus approach identified a clear pattern of non-random differential mortality that was repeated across two independent cohorts. Specifically, the single locus F_{ST} based genome scans found 0.7% to 1.4% of SNP markers to be potentially under the effect of divergent selection in any of the four comparisons of smolts and grilses. While these markers were highly differentiated between life-stages in independent comparisons, of the 130 markers identified as outliers by the single locus method, no more than two were significant in more than one analysis and none were parallel in more than two. Overall, the single locus genome scan approach was not able to detect any convincing pattern of genetically-based differential mortality between smolts and grilses. We cannot entirely refute the possibility that non-parallelism in these outliers translate different selective causes of mortality that varied in time and space. Yet, it would seem exceptional that selective factors would differ to the point that essentially no markers would be a common target of selection in either cohorts or populations. Instead, we propose that such pronounced non-parallelism in outliers detected mainly translate the random effect of factors other than selection (Le Corre & Kremer 2012).

In contrast, the multi-locus approach identified 34 co-varying SNPs that loaded on a multivariate axis of differentiation that could distinguish between smolts and grilses based on the multi-locus genotype composition in the Saint-Jean River. The same pattern of

differentiation was also repeated for both cohorts ruling out the possibility that the observed pattern between smolts and grilse was random or sample biased. These co-varying markers also displayed small changes in allele frequencies between life-stages rather than large changes translating into high levels of genetic differentiation. Thus, the distribution of allele frequency change was offset towards low values, especially when compared to genome scan outliers from our single-locus comparisons. Moreover, allele frequencies at these 34 SNP co-varied more than expected by chance. Based on the contrasting results between single locus genome scans *vs.* polygenic mutlivariate analysis, we argue that the complexity of a phenomenon such as genetically-based differential mortality at sea is best explored via a multi-locus genetic statistical framework, as predicted from quantitative genetics theory and in agreement with Le Corre & Kremer (2012). Indeed, since many selective agents are probably acting to cause mortality at sea and the phenotypic traits potentially under selection are most likely polygenic, it is plausible that many genes across the genome, associated with different biological functions are individually under the effect of relatively mild selection rather than a few genes of large effects. This is supported by the annotation analysis, which revealed that many biological functions were represented among the co-varying SNP set although no specific biological function was over-represented. Instead of detecting potential selective mortality as highlighted by the multi-locus approach, single locus genome scans revealed outlier markers randomly changing in time and space. Given the small number of markers common to both methods, which are suggestive of false-positives, our results are consistent with the high rate of false positives that F_{ST} based methods can yield in experiments lacking other ecological and functional support. The conditions needed to detect parallelism in outliers might be restricted to simple phenotypic traits encoded by genes with very large effects (e.g. Cresko et al. 2004), which emphasizes the need for a multivariate approach to identifying targets of selection of small effects, especially for complex traits and/or when multiple agents of selection are thought to be involved.

The use of genome-wide screens to detect selection in natural populations has become a very popular pursuit, particularly relying on the detection of outlier loci to uncover signatures of selection. Yet, these signatures are generally interpreted without questioning

basic model assumptions (Lotterhos & Whitlock 2014). Evidence is accumulating that this may lead to erroneous conclusions due to false positives (through recombination hotspots, population stratification, endogenous incompatibilities) or false negatives (e.g. weak selection relative to migration or drift) (Bierne et al. 2011; Le Corre and Kremer 2012). Somewhat surprisingly, studies to date have generally ignore the fact that most selected loci may be involved in polygenic adaptation in which case theory would predict small changes in allelic frequencies (Pritchard and Di Rienzo 2010). To our knowledge, only a handful of studies have addressed this is issue, all in humans, and using a different analytical framework than the one we propose here. For instance, Hancock et al. (2010) contrasted the results of approaches based on haplotype structure and differentiation of allele frequencies to those from a method for identifying SNP strongly correlated with environmental variables. Their results suggested that the first group of approaches tended to identify only variants with relatively strong phenotypic effects, whereas the environmental correlation methods can detect variants that make smaller contributions to an adaptive trait. More recently, Daub et al. (2013) recently proposed to evidence polygenic selection in humans by detecting signals of adaptation at the pathway or gene set level instead of analyzing single independent genes. Using a gene-set enrichment test to identify genome-wide signals of adaptation among human populations, the authors found that most pathways globally enriched for signals of positive selection are either directly or indirectly involved in immune response. Although not directly related to the detection of selection but instead in finding genotype-phenotype association, other studies have shown the potential superiority of a multivariate analytical framework over traditional genome wide association studies (GWAS) in explaining phenotypic variation. Yang *et al*. (2010) pointed out that SNPs discovered by genome-wide association studies (GWASs) account for only a small fraction of the genetic variation of complex traits in human populations, raising the question: "Where is the remaining heritability?" They thus estimated the proportion of variance for human height explained by SNP variation using a multivariate linear model analysis. They showed that 45% of variance could be explained by considering all SNPs simultaneously, and concluded that most of the heritability in human height is not missing but has not previously been detected because the individual effects are too small to pass stringent significance tests. Finally, in perhaps the only such studies in non-human, wild populations

performed to date, Ma *et al*. (2010) examined genetic variation in genes from the photoperiodic pathway in a tree species (*Populus tremula*) for signatures of diversifying selection in response to varying light regimes across a latitudinal gradient. While they failed to identify any obvious outlier SNPs using a genome scan approach, they observed a pronounced covariance in allelic effects across populations for growth cessation. Their results thus suggested that spatially variable selection could be affecting genes from the photoperiod pathway even if selection is not strong enough to cause individual loci to be identified as outliers. Overall, then studies in humans and that of Ma et al. (2010) corroborate the results of our study in pointing the potential superiority of a polygenic multi-locus analytical framework in detecting signal of selection acting on complex traits.

5.6.2 Differential mortality at sea and evolutionary changes

Many factors have been suggested to contribute to the large declines in Atlantic salmon populations across the globe. The majority of the identified factors focus on the freshwater and coastal phase of the species' life history where monitoring is more accessible (e.g. the impacts of fish farming, freshwater habitat degradation and warming, barrier to migration, etc.; reviewed in Aas et al. 2011). However, little is known about the factors contributing to mortality at sea, despite the observed declines in sea survival and many hypotheses for what is contributing to these declines. For example, climate change, changes in trophic levels and increased predation have all been proposed as hypothesis for increased mortality at sea (Beaugrand & Reid 2012; Sheehan et al. 2012). It is also unclear if such factors act on populations randomly or in a more deterministic manner. Given the repeated pattern of genotypic changes detected between smolts and grilse in both cohorts of the Saint-Jean River by the canonical discriminant analysis, our results strongly support the hypothesis of genetically-based selective mortality at sea in Atlantic salmon from the Saint-Jean River population. In fact, the temporal replication of similar genetically-based selective mortality for that population strengthens the hypothesis of non-random evolutionary changes occurring during the marine phase of this population. However, these sustained changes are also of concern because North American populations of Atlantic salmon have been established for at least 11,000 years (Verspoor et al. 2007) and it is likely that these populations have been adapting to marine migration since this time. Therefore, given the

time elapsed and their large population sizes, one could argue that any loci increasing marine survival in Atlantic salmon will have been fixed long ago, so observing current selective mortality is unlikely. However, given the fluctuating conditions occurring in the marine environment attributable to natural events such as the North Atlantic Oscillation (NAO) (Stenseth et al. 2003), we argue for a type of balancing selection maintaining diversity among traits underlying marine survival. Nevertheless, we did observe subtle yet significant and repeatable changes in allele frequencies for one of the two populations investigated. The presence of recently selected "marine" loci suggests that genetic variation for alleles contributing to marine survival has been maintained in these populations. This may be due to selective factors of recent origin, perhaps associated with recent changes in environmental conditions. For instance, important rise in sea temperature was shown as tightly linked to growth and survival of migrating Atlantic salmon (Beaugrand & Reid 2012; Mills et al. 2013).

On the other hand, results obtained for the Trinité River suggest that the same selective pressures did not affect this population, as observed genomic changes between life-stages in this population were apparently random, with no evidence for repeatable genetically-based selective mortality. Under the hypothesis of genetically-based selective mortality at sea, the contrasting patterns of selection between the Saint-Jean and Trinité rivers populations could possibly be explained by different migratory routes to the feeding grounds and/or migration to different feeding grounds off the coast of Newfoundland and Labrador. Indeed Atlantic salmon from populations of the North shore of the Gulf of St. Lawrence (where the Trinité River population is located) likely migrate out to sea through the Strait of Belle Isle at the northern tip of Newfoundland (Dutil & Coutu 1988). On the other hand, rivers further south from the Gaspésie Peninsula (where the Saint-Jean River population is located) most likely use both out migrating routes of the Gulf of St. Lawrence, namely the northern Strait of Belle Isle and the southern passage through the Cabot Strait (Belding 1940; Lefèvre et al. 2012). Moreover, it has been reported that Atlantic salmon from different regions in the United Kingdom segregate in distinct feeding grounds (MacKenzie et al. 2012). Populations from Saint-Jean and Trinité rivers represent two regional genetic groups of salmon previously identified (Dionne et al. 2008) and shown

to be potentially locally adapted to distinct environmental settings (Bourret et al. 2013b). In particular early post-smolt life-stage definitely occurs in different estuarine areas of the Gulf of St-Lawrence River for these populations. Also, there is evidence for high mortality rate during the early marine migration , especially in the estuaries and the river mouths, reaching up to 71% (reviewed in Thorstad et al. 2012). Also, Plantalech manel-la et al. (2011) reported differences in early (estuarine) migratory behavior of distinct populations of Atlantic salmon that could have fitness consequences. Other studies have also found a negative correlation between marine water temperatures and post-smolt survival of migrating Atlantic salmon (Friedland et al. 2000, Todd et al. 2008; Hvidsten et al. 2009). The Saint-Jean River estuary is located in the southern part of Québec and experiences higher temperatures than the Trinité River (Dionne et al. 2013). The estuary of the Saint-Jean River is also much larger than that of the Trinité River (Dionne et al. 2013) and as a result, smolts from the Saint-Jean River would need to spend more time crossing this environment which is typically associated with high mortality and selective pressures (e.g. predation, high temperature exposure; Thorstad et al. 2012). Overall, these observations raise the hypothesis that any recent change in the environmental conditions encountered along their two migratory routes may have differentially impacted the Saint-Jean and Trinité populations. More information on local early marine migration and monitoring of oceanic environmental conditions is needed to further test this hypothesis.

Alternatively, one could argue that the observed pattern of differentiation between smolts and grilse may be attributable to a sex-biased sampling between life stages since it is known that the majority of Atlantic salmon returning after one winter at sea in both the Saint-Jean and Trinité rivers are males whereas most multi-sea-winter (MSW) fish are females (Dionne et al. 2013). Thus, the observed genetically-based differentiation might reveal divergence between sexes rather than a result of differential mortality at sea. For this to explain the repeated differences observed between both populations would require that they differ in the proportion of the two sexes in returning grisles. Yet, grisles were composed of a majority of males in both populations (100% in SJ and 86% in TR), while the proportion of both sexes in smolts was relatively similar in both rivers (average of 74%

males in SJ and 72% males in TR). Therefore, it is unlikely that the differences observed between populations reflect a global effect of sex.

5.6.3 Concluding remarks

This study is the first attempt to use population genomics to study the evolutionary changes associated with the global declines in adult Atlantic salmon returns. Detecting a specific causative factor for mortality was beyond the scope of this study. Indeed, much more information about the physical and environmental conditions encountered by Atlantic salmon during their marine phase and the genomic composition of other life-stages will be required to reach this goal. Furthermore, given the complexity of the trait under selection and the diversity of potential selective forces in the marine environment, we acknowledge that increasing the number of individuals genotyped from each population as well as the geographic and genomic coverage will be a necessary step in future studies. Namely, a more comprehensive study would allow a finer characterization of markers and interactions underlying selective mortality across the genome. Indeed, genome wide association studies (GWAS) aiming at detecting association between variation at the genome and phenotypic level for complex traits typically involve the use of many more markers than what we could use here. Thus, even if we reach a genome coverage of about 100 markers per chromosome, which is much more than typical "saturated" genetic maps published until very recently, it is very likely that the relatively small number of markers we used led to missing many of the associations between genotyped markers and the actual target of selection in outbred salmon populations with limited linkage disequilibrium. Nevertheless, we were able to detect non-random, temporally repeatable, genetically-based selective mortality occurring between the smolts and grilse life-stages in one of two rivers. This represents a major step towards elucidating the dynamics of differential mortality at sea between salmon from different population origin. This pattern could only be detected with a multivariate genetic framework inspired from quantitative genetics theory and designed to consider markers as co-varying multi-locus entities. In contrast, the FST based single locus approach was unable to reveal any pattern of allele-frequency change putatively associated with differential mortality at sea, further supporting arguments in favor of a paradigm shift in the analysis of selection that has been largely based thus far on the detection of strong selective

effects causing large allele frequency on single markers to searching for polygenic selection of smaller effects causing modest by co-varied changes in allele frequencies at many genomic regions, which is predicted to be more commonly acting on complex, fitness-related traits in natural populations

5.7 Acknowledgements

We warmly acknowledge the contribution of Dr. Matthew P.Kent in providing his expertise for the genotyping of analysed individuals at CIGENE. We thank the Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs du Québec (MDDEFP) for sampling Atlantic salmon. We are grateful to Marie-Hélène Perreault and Geneviève Ouellet-Cauchon for technical assistance in the laboratory, Gaétan Daigle for support in statistical analyses, and Pierre-Alexandre Gagnaire and Anne C. Dalziel for constructive insights on the manuscript. The SNP discovery, array development and genotyping were performed by CIGENE at the national technology platform, supported by the functional genomics programme (FUGE) in the Research Council of Norway. This research was supported by a strategic grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) led by LB in close collaboration with the MDDEFP represented by MD and the Department of Fisheries and Oceans Canada (DFO). VB was supported by an Alexander Graham-Bell scholarship from NSERC and LB is the Canadian Research Chair in genomics and conservation of aquatic resources. The authors declare no conflict of interest.

5.8 Tables

Table 5.1 Summary of the canonical discriminant analysis. For each axis, eigenvalue, proportion of variance and cumulative variance accounted for, degrees of freedom, F statistic and P-value is given.

Axis	Eigenvalue	Proportion	Cumulative	DF	F-stat	P-value
1	86.7251	0.9652	0.9652	336	5.03	<0.0001
2	1.3629	0.0152	0.9804	282	1.47	<0.0001
3	0.7548	0.0084	0.9888	230	1.08	0.228
4	0.3852	0.0043	0.9931	180	0.82	0.948
5	0.2960	0.0033	0.9964	132	0.69	0.994
6	0.2291	0.0025	0.9989	86	0.56	0.999
7	0.0966	0.0011	1.0000	42	0.35	0.999

Table 5.2 Summary of the generalized linear models. For each of both significant canonical discriminant axis a) axis 1, and b) axis 2; the degrees of freedom, sum of squares, F statistics and P-values are given for each source of variation tested in the model. Significant values are bold and italicized.

a)

Source	DF	Sum Square	F-stat	P-value
Rivers	***1***	***16641***	***16543***	***<0.0001***
Life-stages	1	2.436	2.42	0.121
Cohorts	1	3.613	3.59	0.060
Life-stages*Cohorts	1	0.426	0.42	0.516
Rivers*Life-stages	1	0.270	0.27	0.605

b)

Source	DF	Sum Square	F-stat	P-value
Rivers	1	0.038	0.04	0.848
Life-stages	***1***	***102***	***98.52***	***<0.0001***
Cohorts	1	12.60	12.24	0.0006
Life-stages*Cohorts	1	0.385	0.37	0.542
Rivers*Life-stages	***1***	***139***	***135.33***	***<0.0001***

5.9 Figures

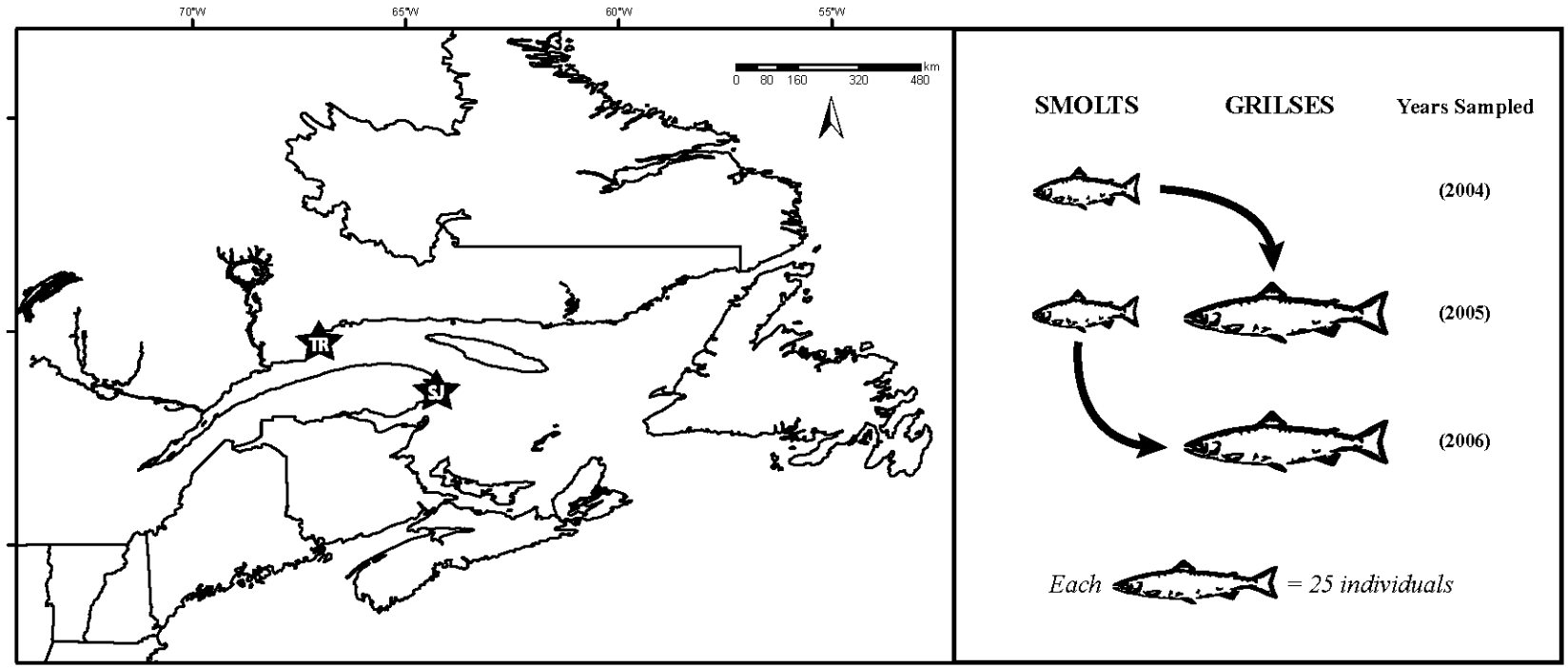


Figure 5.1 Map showing sample sites on the left panel: Trinité River (TR) and Saint-Jean River (SJ). The right panel is a schematic representation of the experimental design in each river. 25 individuals were sampled for each life-stages in each cohort in both rivers for a total of 200 individuals.

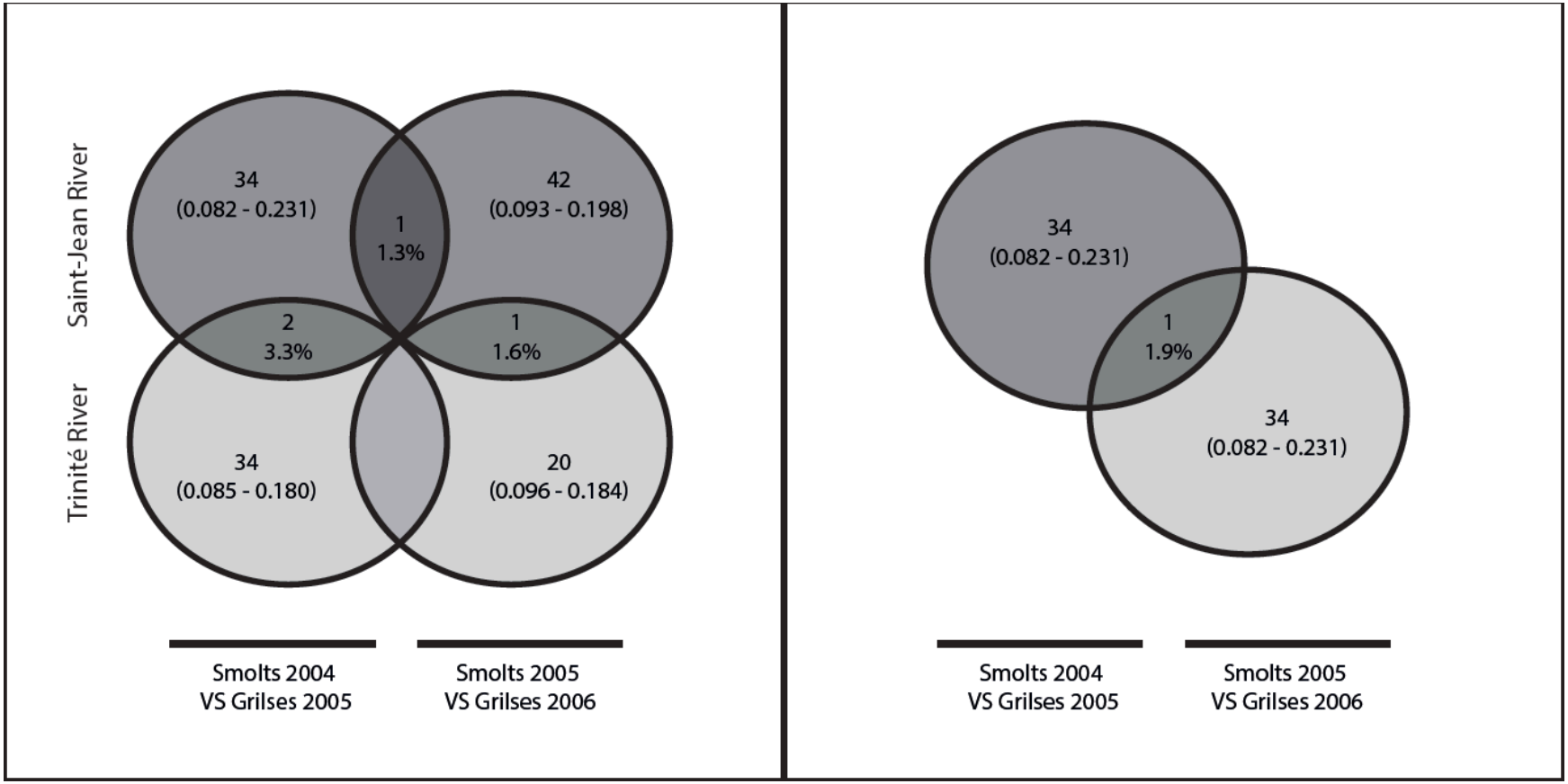


Figure 5.2 Summary of single locus genome scans results. Each circle represents a genome scan comparing juveniles (smolts) and adults (grilses) of a given population and cohort. Numbers in circles correspond to the number of loci under potential divergent selection in each genome scans with their given F_{ST} range in parenthesis. Numbers in overlapping regions correspond to common outliers found in two genome scans with the proportion of markers this number represent on the total number of outliers for both scans. Where there are no overlapping regions, common outliers were not found. Genome scan plots are presented on Figure S1.

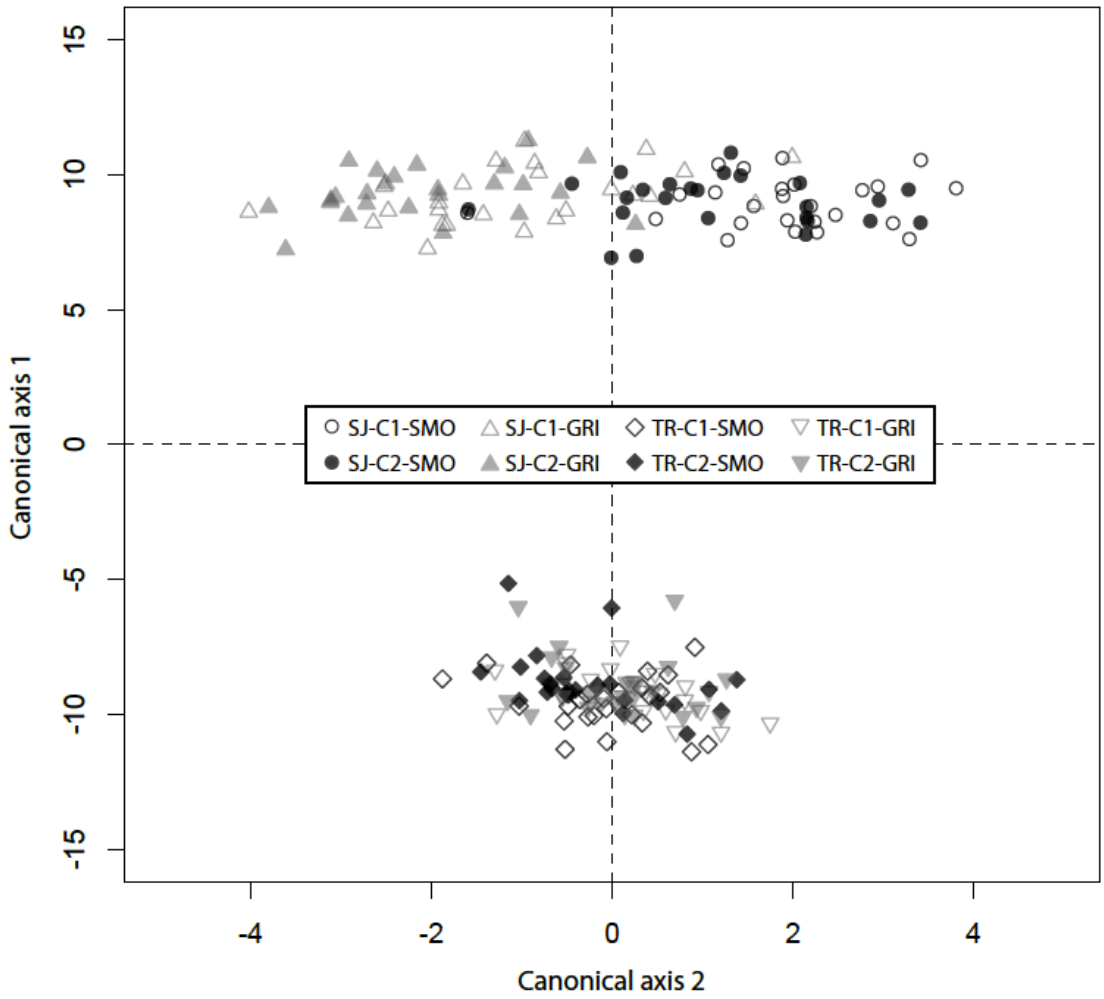


Figure 5.3 Canonical discriminant analysis of genetic differentiation among individuals based on 2923 markers grouped in 48 principal components. Canonical axis 1 is significantly correlated with genetic variation among rivers, and Canonical axis 2 is significantly correlated with life-stages and cohorts. (SJ = Saint-Jean River, TR = Trinité River, C1 = cohort 2004, C2 = cohort 2005, SMO = smolts and GRI = grilses)

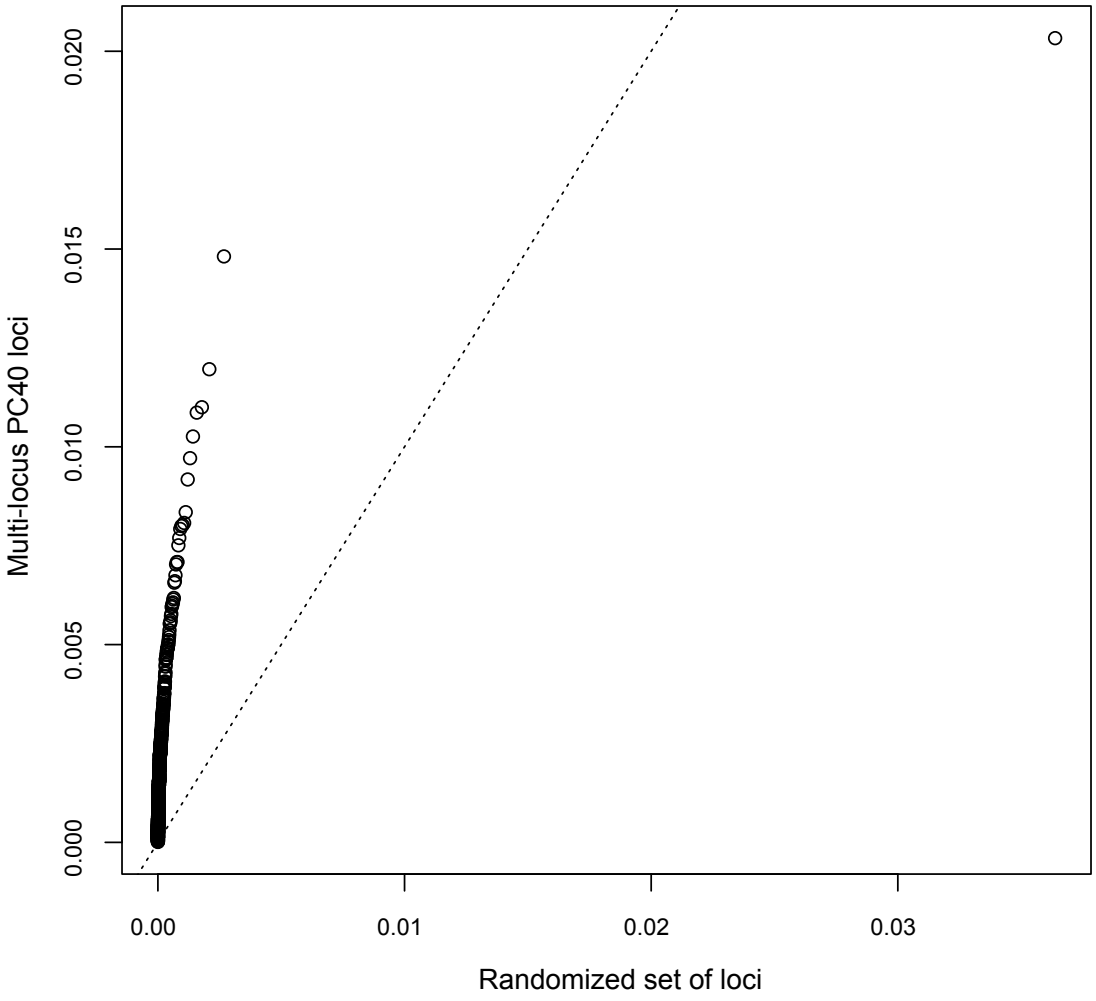


Figure 5.4 Quantile-quantile plot of the distribution of allelic effect covariance among 34 markers significantly correlated with PC factor 40 on the Y-axis and a 34 marker randomly chosen among all markers 1000 times on the X-axis. The dotted line represents X = Y relation.

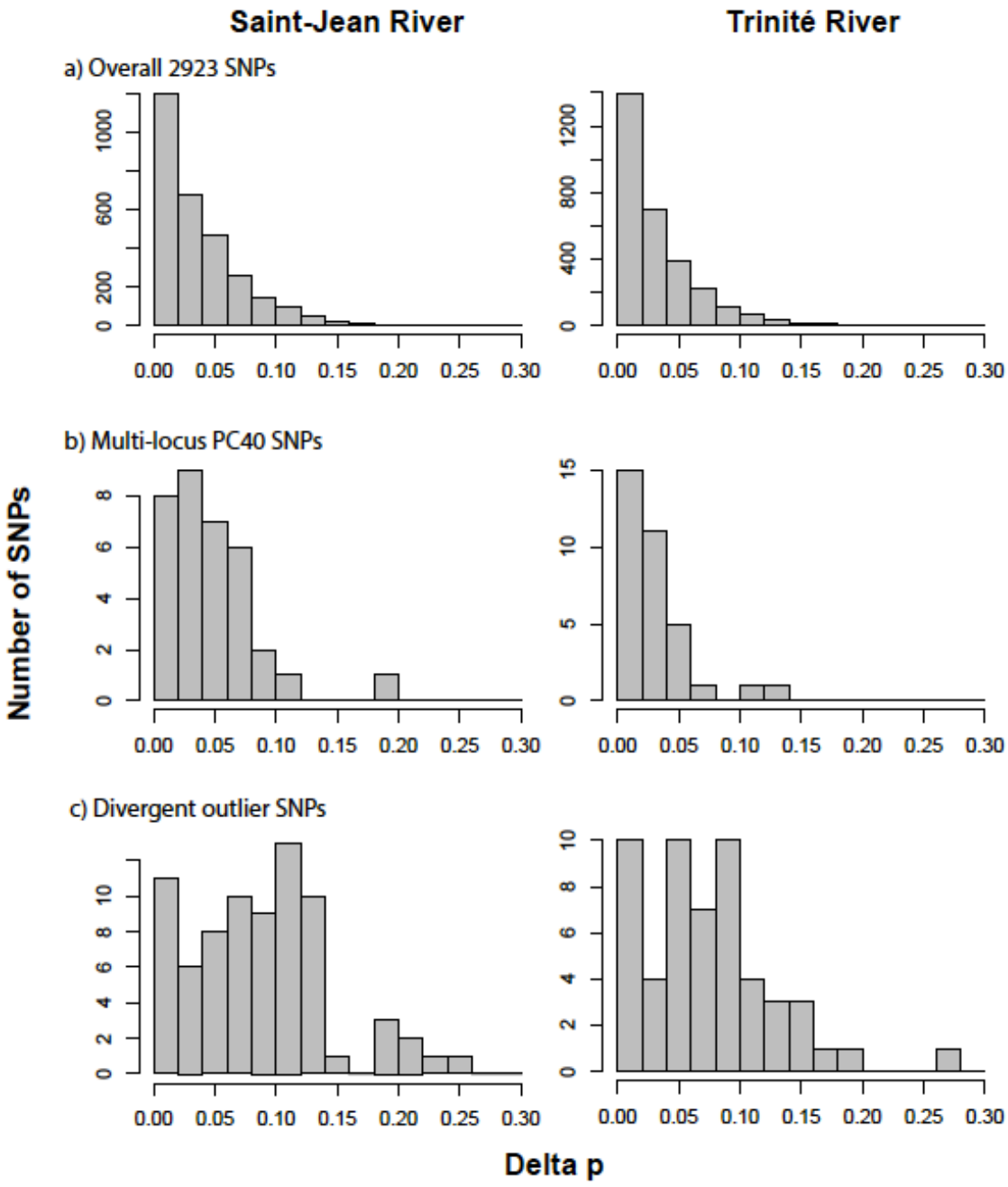


Figure 5.5 Distribution of allelic frequency changes (Δp) between life-stages in each population with the two cohorts grouped together: a) over all 2923 markers, b) among 34 markers significantly correlated with PC factor 40, and c) among single-locus genome scan divergent outliers.

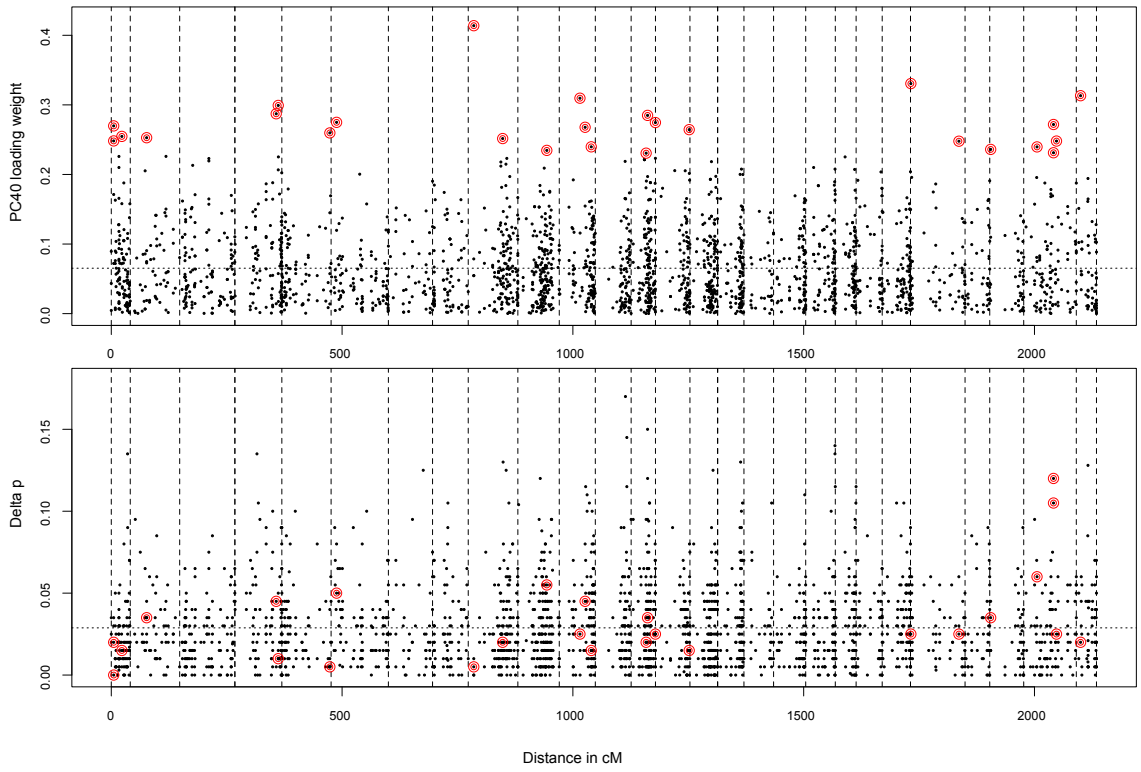


Figure 5.6 Genetic linkage map showing the distribution of principal component 40 (PC40) loading weight of each of the 2923 SNP markers on the top panel and overall allelic frequency changes between life-stages on the bottom panel. Vertical dashed lines distinguish linkage groups. On both panels, circled dots indicate the 34 SNPs significantly correlated with PC factor 40 and the horizontal dotted line indicates the average loading weight (top) and delta p (bottom) among all markers.

5.10 Supplementary material

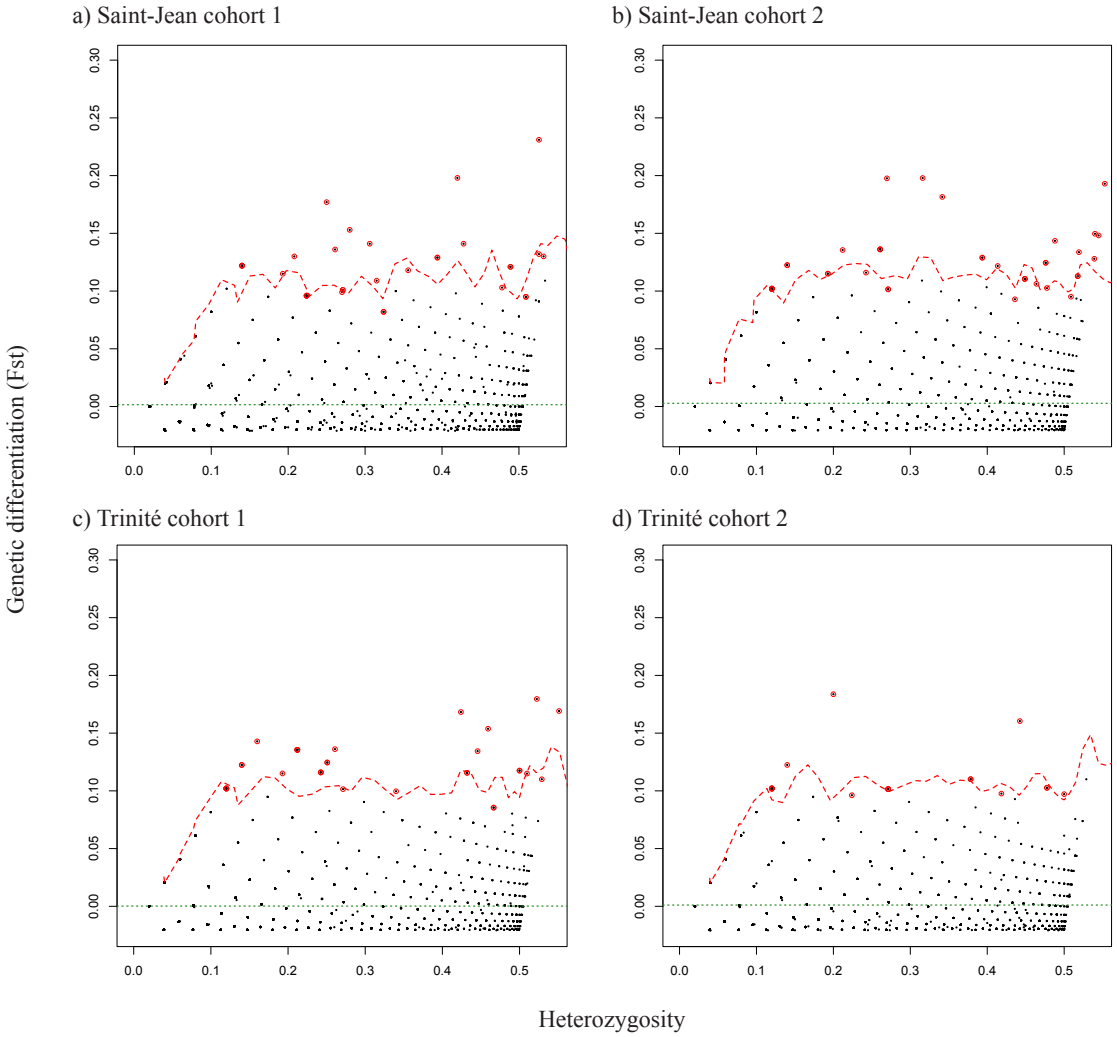


Figure 5.S1 Differentiation (F_{ST}) as a function of heterozygosity as estimated by LOSITAN (Antao et al. 2008) for four comparison of smolts and grilses in: a) Saint-Jean River cohort 1, b) Saint-Jean River cohort 2, c) Trinité River cohort 1, and d) Trinité River cohort 2. In each panel, outliers markers ($P < 0.01$) are marked by circled dots, dashed lines represent upper and lower 99% confidence levels and dotted lines indicates the average F_{ST} across all loci.

Hereafter listed supplementary methods, tables & figures are available online or on demand:

Table 5.S1 Summary of the detection of markers potentially under selection following genome scans implemented in LOSITAN (Antao et al. 2008). For each different detection schemes, observed heterozygosity (H_O), differentiation (F_{ST}), test's P-value and outlier status is given for each locus. In the outlier status column, number 1 indicates loci potentially under divergent selection, number 2 indicates loci potentially under balancing selection and NA indicates loci within neutral expectations (P-value > 0.005).

Table 5.S2 Summary of principal component analysis performed on 2923 SNP markers for 200 samples. Results are reported for the 48 PC factors retained based on Horn's parallel analysis criterion

Table 5.S3 Summary of markers' absolute loading weight on PC factor 40 in the *Abs(F40)* column. The *Selected* column indicates if a marker is considered part of the multilocus group under selection (1) or not (NA).

Table 5.S4 Blast results from BLAST2GO with blast *e*-value treshold of 1x10-3 and gene ontology (GO) terms annotation for blast of *e*-value inferior to 1x10-10. In "Sequence Description", "NA" is given to unsuccessful blast. In "Gene Ontologies", "P" is for biological process, "F" is for molecular function and "C" is for cellular component.

Chapitre 6 Conclusion Générale

Les objectifs de cette thèse s'inscrivaient dans un effort d'amélioration des connaissances liées à de nombreux volets touchant de près la gestion et la conservation du saumon atlantique. Dans ce contexte, cette thèse visait principalement à départager la part de la divergence génétique neutre de la divergence génétique adaptative. En effet, en vertu des connaissances acquises par le passé et des nouvelles opportunités offertes par l'avancement des ressources génomiques, il était devenu opportun d'explorer les bases génomiques de la divergence adaptative. Cette démarche visait ultimement à fournir un éclairage nouveau sur les mécanismes génomiques de la sélection et de l'adaptation locale pour le saumon atlantique en plus de permettre d'identifier des agents de sélection environnementaux et leurs cibles fonctionnelles. Globalement, l'atteinte de nos objectifs a donc permis de souligner l'importance de la divergence adaptative dans différents contextes de gestion et de conservation.

6.1 Sommaire des principaux résultats

En premier lieu, l'étude des impacts des échappées d'élevage sur la population sauvage de la rivière Magaguadavic du chapitre deux a permis d'identifier des changements relativement cryptiques par l'exploration de la divergence potentiellement adaptative. En effet, malgré le déclin rapide de la population au cours des deux dernières décennies et l'augmentation importante des échappées d'élevage se retrouvant dans la rivière, aucun changement significatif n'a été observé en termes de richesse allélique ou diversité génétique estimées à l'aide des marqueurs microsatellites pour la population sauvage. En théorie, dans une population en déclin, nous nous attendions à observer une diminution de la diversité génétique. Par conséquent, nos résultats suggèrent que l'introgression des saumons domestiques, mise en évidence par un changement temporel significatif du déséquilibre de liaison, a possiblement contré les effets négatifs de la baisse d'effectif de la population sauvage. Au niveau de la divergence potentiellement adaptative, c'est en l'isolant que nous avons identifié des impacts importants de l'introgression sur l'intégrité génétique fonctionnelle de la population de la rivière Magaguadavic. En général, nos résultats montraient une diminution chronologique des loci potentiellement soumis à la sélection directionnelle. Plus spécifiquement, nous avons identifié un SNP préalablement associé à un QTL important pour les taches présentes sur les juvéniles de saumon (tacons)

qui retenait sa distinction génétique entre saumons sauvages et domestiques plus longtemps que les autres marqueurs potentiellement sous l'influence de la sélection divergente. Cette situation se traduit donc par une altération significative de l'intégrité génétique de la population indigène, incluant une perte possible d'adaptation aux conditions naturelles.

Devant les bénéfices importants fournis par l'ajout de marqueurs génétiques potentiellement sous sélection dans le chapitre deux et le chapitre trois, la validation de la biopuce à plus de 6000 marqueurs de type SNP était attendue. Cette dernière a d'abord permis d'observer une grande cohérence entre les paramètres génétiques observés à l'aide des marqueurs SNP et des marqueurs génétiques précédemment utilisés. De plus, la structure génétique révélée par la biopuce est venue confirmer les structures génétiques précédemment observées, c'est-à-dire la grande divergence entre les populations d'Europe et d'Amérique du Nord et la structure hiérarchique des populations de saumon atlantique en Europe. Par ailleurs, pour la première fois, des populations de l'ensemble de la répartition européenne de l'espèce ont été regroupées dans une même analyse. Ceci a permis l'identification de trois groupes régionaux principaux qui tireraient leur origine de trois événements de colonisation indépendants. La distribution des fréquences alléliques des marqueurs potentiellement sous l'effet de la sélection divergente a révélé des patrons clinaux significativement corrélés avec la distance géographique le long de continuums linéaires établis entre les différents groupes régionaux. Ces clines identifiaient ainsi des zones de contact secondaire entre les groupes. Bien que ces zones coïncident avec la présence de gradients environnementaux, la possibilité que les clines observées soient imputables à des barrières endogènes, c'est-à-dire issues de contingence historiques, rend l'interprétation équivoque quant à l'influence de l'environnement sur la divergence adaptative.

Par une approche de génomique du paysage appliquée à des populations de saumon atlantique nord-américaines, le chapitre quatre a justement permis de mettre en lumière des liens étroits entre l'environnement et la divergence génétique adaptative en plus d'améliorer notre compréhension de l'échelle génomique et géographique de l'adaptation locale en milieu naturel. Spécifiquement, nous avons révélé de fortes corrélations entre les

structures génétique et environnementale des populations étudiées. Notre cadre méthodologique a permis de mettre en évidence deux grands axes de sélection environnementale corrélés avec la divergence génétique. En effet, parmi toutes les variables environnementales étudiées, les caractéristiques climatiques (température et précipitations) et géologiques des rivières sont celles qui ont montré une corrélation significative avec la divergence génétique des populations. De plus, les marqueurs potentiellement sous sélection les plus fortement corrélés avec les variables climatiques montraient un enrichissement de fonctions liées à la croissance. Ce dernier résultat nous a permis l'élaboration d'une explication fonctionnelle pour la divergence observée à ces marqueurs. Selon cette hypothèse, une part de la divergence adaptative observée serait imputable à un phénomène de sélection à contre-courant pour la croissance en fonction de la durée de la saison de croissance. En somme, puisque la structure génétique des populations de saumon atlantique étudiées ici comporte une composante régionale importante, nos résultats supportent fortement une composante régionale de l'adaptation locale chez le saumon atlantique nord-américain.

Finalement, les objectifs du chapitre cinq concernaient deux aspects fondamentaux complémentaires, mais relativement distants l'un de l'autre. D'une part, d'un point de vue orienté sur une problématique liée au saumon atlantique, nous voulions mettre en évidence certains facteurs influençant la mortalité accrue en mer par un examen des changements potentiels au niveau de la composition génomique des populations au cours de la migration. D'autre part, puisque nous examinions la divergence adaptative liée à un trait précis mais complexe, nous voulions confronter deux modèles de sélection qui sont débattus à la lumière des nombreuses études s'attardant maintenant à la détection de la sélection en milieu naturel. Dans un premier temps, bien que de nombreux marqueurs aient été identifiés comme étant potentiellement sous l'effet de la sélection naturelle, aucune évidence de sélection en parallèle ou répétée au niveau temporel n'a été trouvée par une méthode basée sur la différentiation individuelle des marqueurs. Par ailleurs, une approche multilocus novatrice a permis de détecter un patron de sélection répété sur deux cohortes dans une des deux populations à l'étude pour ce chapitre. Le fait qu'aucun patron de mortalité sélective n'ait pas été détecté dans l'autre population suggère que les causes de

cette mortalité sélective soient différentes d'une population à l'autre. L'explication la plus plausible pour cette divergence quant aux patrons observés entre les deux populations réside dans l'utilisation potentielle de différentes routes de migration entre la rivière d'origine et les zones d'alimentation en mer. Par conséquent, une certaine hétérogénéité dans les changements potentiels de conditions environnementales ou d'autres types le long de ces routes pourrait engendrer une réponse évolutive contrastée entre les populations.

6.2 Contributions

Globalement, il convient de souligner l'étroite relation entre les défis méthodologiques que représentait l'accomplissement des objectifs de cette thèse et les bénéfices qu'il est permis d'en tirer. Effectivement, c'est grâce à la haute résolution avec laquelle la structure génétique des populations a été examinée dans les chapitres de cette thèse que nous avons pu jeter un regard nouveau sur la divergence potentiellement adaptative du saumon atlantique tout en raffinant nos connaissances quant aux paramètres génétiques structurant ses populations. C'est ce raffinement qui a d'abord permis de mettre en évidence des impacts majeurs de l'industrie aquacole du saumon atlantique sur une population sauvage. De plus, c'est la caractérisation fine de la génomique des populations en relation avec les conditions environnementales qui aura permis d'identifier les liens étroits entre ces deux paramètres et des cibles potentielles quant aux mécanismes sous-jacents impliqués dans la divergence adaptative. Notre approche révèle également des changements évolutifs significatifs apparaissant lors de la migration en mer, une phase de vie méconnue à plusieurs égards pour laquelle la recherche n'a que peu de réponses à apporter. En raison de l'apport incontestable de l'étude des bases génomiques de la divergence adaptative du saumon atlantique, les objectifs atteints par cette thèse fournissent de nouvelles connaissances sur lesquelles les gestionnaires pourront s'appuyer dans l'élaboration de stratégies modernes de gestion et de conservation.

Plus spécifiquement, en ce qui concerne les impacts des échappées d'élevage sur les populations sauvages, bien que certaines études avaient identifié des impacts potentiellement néfastes (McGinnity et al. 2003; Skaala et al. 2006), peu d'études avaient mesuré ces impacts sur des populations naturelles, particulièrement en Amérique du Nord

où, à notre connaissance, aucune ne s'y était attardée. Nos résultats pour ce chapitre représentent donc non seulement une contribution majeure en ce sens, mais également en vertu de la mise en évidence d'une dégradation marquée de la divergence adaptative suivant l'introgression des saumons domestiques dans la population sauvage. Depuis la publication de nos résultats, des travaux importants ont également mis en évidence différents degrés d'introgression subis par des populations de saumon atlantique sauvage par des échappées d'élevage (Karlsson et al. 2011; Glover et al. 2013). Ces études n'ont pas examiné l'impact sur les bases génétiques adaptatives des populations sauvages, mais en vertu de nos résultats, il y a lieu de spéculer qu'en fonction du degré d'introgression et de l'apport plus ou moins constant des intrants domestiques, un effritement de la valeur adaptative et donc de la divergence adaptative des populations sauvages est à prévoir. De plus, il est à noter qu'en l'absence d'études comme la nôtre, l'industrie aquacole ne peut mesurer concrètement son impact sur l'intégrité génétique des populations. Devant cet état de fait, l'industrie a beau jeu de rejeter sa responsabilité sur les autres causes potentielles du déclin des populations comme celle de la rivière Magaguadavic. Sans vouloir minimiser l'importance de ces causes et sans prétendre que notre étude établit un tel lien, elle permet néanmoins de présenter un impact non équivoque de cette industrie sur les populations locales.

Bien que la responsabilité du développement de la biopuce à SNPs revienne totalement à l'équipe du *Centre for Integrative Genetics* (CIGENE), en Norvège, la validation de cette dernière représente également une base solide sur laquelle de nombreux travaux pourront s'appuyer. L'avènement de ce type d'outil génomique, applicable à l'étude de la génomique des populations en milieu naturel pour des espèces non modèles, s'est accompagné d'une prise de conscience importante liée aux biais de développement. Ce type de biais associé au déséquilibre des fréquences des marqueurs découverts entre les populations sources (du développement) et les autres populations peut être une source d'erreur importante dans l'estimation de paramètres génétiques (Albrechtsen et al. 2010). Dans le cas de la biopuce développée par CIGENE, la caractérisation du biais de développement était essentielle pour permettre aux futurs utilisateurs d'en faire un usage adéquat et surtout d'en planifier une utilisation avertie. De plus, notre validation pave la voie à l'utilisation de la biopuce ou de

nouvelles versions de celle-ci dans une grande variété d'applications, notamment pour résoudre des problématiques de pointe en gestion et conservation des populations (présente thèse), mieux circonscrire l'impact des échappées d'élevage dans les populations sauvages (Glover et al. 2013) et possiblement dans des programmes de sélection assistée pour l'élevage commercial ou pour de futures réintroductions (Dominik et al. 2010). Finalement, cette validation s'est effectuée en réalisant une analyse exhaustive de la structure génétique des populations naturelles de saumon atlantique, consolidant ainsi des connaissances cruciales quant aux patrons globaux de différentiation génétique dans une seule et même étude. La structure régionale en Europe démontre bien les différentes routes de colonisation préalablement suggérées dans plusieurs études indépendantes (Tonteri et al. 2007, 2009 et références incluses) et valide une fois de plus les bénéfices d'une approche à grande échelle spatiale pour ces espèces caractérisées par une vaste répartition.

Cette étude a également permis de souligner l'importance d'une caractérisation fine des paramètres environnementaux dans la mise en évidence de l'influence des conditions environnementales sur la composition et la divergence génétique des populations naturelles. D'autant plus que la génétique et maintenant la génomique du paysage sont des approches de plus en plus interpellées dans le contexte où les effets des changements climatiques sont au cœur de plusieurs programmes de recherche (Manel & Holderegger 2013; Petren 2013). Notre étude met également de l'avant un cadre méthodologique novateur très bien adapté au contexte exploratoire dans lequel s'inscrivent la plupart du temps les études de génétique des populations. Notre approche statistique évite notamment les pièges de la colinéarité des données tout en permettant d'avoir une vue d'ensemble des interactions entre toutes les variables. À partir d'une caractérisation fine de la composition génétique et du contexte environnemental des populations, notre approche démontre son efficacité dans l'identification d'agents de sélection environnementaux de même que les cibles fonctionnelles de cette sélection. Par conséquent, ce type d'approche s'inscrit parfaitement dans l'élaboration de nouvelles stratégies de gestion et de conservation visant l'intégration de paramètres génétiques et environnementaux, de même que leurs interactions (Primmer 2009; Funk et al. 2012).

La comparaison entre la méthode de détection de la sélection par la méthode de balayage génomique basée sur la différenciation individuelle des loci et notre approche multilocus représente un des faits marquants de cette thèse. Elle aura permis d'établir un exemple saillant quant au bien-fondé du concept de génétique quantitative selon lequel la sélection peut agir de façon subtile sur un grand nombre de marqueurs ou gènes lorsqu'elle agit sur un trait complexe (Latta 1998; Le Corre & Kremer 2012). Ce postulat s'oppose donc au concept de balayage sélectif ciblé par les méthodes de détection à loci individuels. Notre étude propose donc un changement de paradigme quant à la recherche de la sélection en populations naturelles, particulièrement dans le cas de traits complexes potentiellement soumis à l'effet de la sélection. Au-delà des considérations théoriques sur la détection de la sélection, la mise en évidence d'un patron de mortalité sélective en mer pour une population de saumon atlantique du Québec constitue une avancée importante vers une meilleure compréhension de ce phénomène affectant les populations naturelles. Malheureusement, nous n'étions pas en mesure d'identifier les facteurs à l'origine de cette mortalité. Par ailleurs, à mesure que de nouvelles connaissances surgiront à propos de la phase de vie marine du saumon atlantique et de la dynamique des facteurs physiques et environnementaux qui la caractérise, nos résultats pourront être mis en lumière.

6.3 Perspectives

La détection de la divergence adaptative est sans contredit l'élément théorique central autour duquel s'est articulée cette thèse. Avec le recul, prise dans son ensemble, cette thèse illustre bien le cheminement que certains chercheurs ont probablement emprunté face au défi que représentent l'utilisation des méthodes disponibles et l'interprétation des résultats de telles détections. D'abord, pour les chapitres deux et trois, les considérations méthodologiques concernaient essentiellement le choix de la bonne méthode analytique compte tenu du nombre de populations et de la structure globale des populations. Puis, tel que ressenti dans la réalisation des chapitres trois et quatre, les craintes concernant la détection de faux positifs et la maladaptation des modèles sous-jacents aux méthodes de détection se faisaient grandissantes ou à tout le moins mieux définies dans la littérature (De Mita et al. 2013). Par contre, ces craintes étaient contrebalancées par deux éléments clés liés au contexte expérimental des chapitres en question. D'une part, la nature hiérarchique

de la structure génétique des populations qui caractérise le saumon atlantique ne permet réellement qu'une méthode parmi les quelques rares développées jusqu'à maintenant, soit la version de FDIST dite hiérarchique d'Excoffier & Lischer (2010), de loin la plus performante dans notre contexte (De Mita et al. 2013). D'autre part, la nature exploratoire et descriptive de nos analyses permettait d'envisager une certaine marge d'erreur en fonction de l'interprétation que nous pouvions faire de nos résultats. Dans le cas du chapitre trois, les résultats étaient équivoques quant à la part de la sélection et de la contingence historique, mais nous ne disposions pas de données supplémentaires pour explorer d'autres avenues, maintenant ainsi un certain suspense. Par contre, dans le cas du chapitre quatre, les marqueurs identifiés comme potentiellement sous sélection ne le sont vraisemblablement pas tous, mais nous sommes confiants que notre démarche exploratoire était soutenue par une approche statistique suffisamment robuste et un contexte environnemental suffisamment bien défini pour avoir pleinement confiance en nos interprétations et conclusions. Finalement, il est survenu une dernière prise de conscience dans le cadre de la réalisation du chapitre cinq. La démarche visait ici à mettre en évidence la sélection sur un trait précis, de surcroît complexe et en parallèle avec peu d'échantillons dans un design expérimental. Aucune méthode existante ne semblait adaptée à un tel contexte. À la lumière des travaux de Le Corre & Kremer (2012), nous avons mis de l'avant une méthode personnalisée permettant de respecter le mécanisme de sélection du concept de génétique quantitative.

Avec le recul, ce cheminement permet un certain nombre de constats qui doivent être pris en compte dans les projets actuels et futurs qui visent la mise en évidence de bases génétiques adaptatives. En fonction de l'expérience fournie par la réalisation des différents chapitres de cette thèse, deux constats semblent particulièrement importants. Dans un premier temps, la nature du trait pour lequel on cherche à détecter la sélection est déterminante et doit servir à ajuster nos attentes et nos interprétations. Deuxièmement, les méthodes de balayage génomique permettent une exploration efficace des données génomiques, mais leur interprétation doit faire l'objet d'analyses robustes et d'investigation par d'autres moyens en aval pour en valider l'information.

En ce qui concerne le saumon atlantique et les aspects développés dans le cadre des chapitres de cette thèse, il reste beaucoup d'éléments à explorer. Dans le contexte des échappées d'élevage, pour les populations européennes où les effectifs demeurent importants par rapport aux populations extirpées de la Baie de Fundy, il serait particulièrement intéressant d'examiner l'introgression de traits sélectionnés en captivité (ou inversement en milieu naturel) en se servant de leurs bases génétiques et vérifier à quelle vitesse ces traits introgressent par rapport à la base neutre. Par exemple, il serait intéressant de voir comment se comporte les fréquences de gènes liés à la croissance en fonction des taux d'introgression variables rapportés par Glover et al. (2013).

Toujours en Europe, compte tenu des résultats du chapitre trois sur la structure régionale des populations et les clines observées, une approche de génomique du paysage à l'instar de celle du chapitre quatre serait appropriée. Un effort concerté permettrait certainement l'acquisition de données environnementales pour l'ensemble des rivières pour lesquelles nous disposons des données génétiques. Ces données pourraient résoudre une partie du questionnement lié à l'influence de l'environnement dans l'établissement des barrières au flux génique que les patrons clinaux semblent avoir révélées. De plus, comme plusieurs des marqueurs associés à ces clines se comportaient de la même manière entre les trois groupes, il serait également intéressant d'établir leur rôle ou leur fonction biologique, le cas échéant.

En Amérique du nord, l'hypothèse d'une sélection à contre-gradient pour une croissance plus rapide en fonction d'une saison de croissance plus courte reste à tester de façon empirique. À ce sujet, trop peu d'études ont fait le pont entre des résultats suggérant des mécanismes moléculaires régissant la sélection et une véritable démonstration de ces mécanismes (mais voir Storz & Wheat 2010). Il serait également intéressant de revisiter un contexte expérimental similaire à celui du chapitre cinq, mais pour lequel nous disposerions de plus de populations, plus de cohortes et surtout davantage de stades de vie impliqués ou non dans la migration en mer. Cela permettrait d'acquérir de meilleures connaissances quant aux changements de la composition génétique des populations en fonction des différents défis que représentent les différentes étapes de leur stade de vie et ultimement

d'identifier les facteurs responsables de ces changements et possiblement les fonctions ciblées par ces derniers.

Pour conclure, il est inévitable de ne pas mentionner la nécessité d'améliorer et de consolider de manière structurée nos connaissances des processus et fonctions biologiques associées aux différents marqueurs que nous caractérisons dans les populations naturelles. L'annotation des gènes demeure un sujet chaud, particulièrement quand vient le temps d'interpréter des annotations de gènes ou de marqueurs potentiellement sous sélection. De plus, dans le cas du saumon atlantique, le séquençage du génome qui sera complété officiellement en 2014 (Davidson et al. 2010)) permettra possiblement d'améliorer ces connaissances, mais aussi de développer de nouveaux outils génomiques encore mieux adaptés aux réalités de gestion et de conservation du saumon atlantique. Dans tous les cas, il sera intéressant de voir quels outils serviront adéquatement la gestion et la conservation des espèces et quelle forme prendra l'étude de la génétique des populations dans la prochaine décennie.

Bibliographie

Aas O, Einum S, Klemetsen A, Skurdal J (2011) *Atlantic Salmon Ecology* Wiley-Blackwell, Oxford.

Adkison MD (1995) Population differentiation in Pacific salmon: Local adaptation, genetic drift, or the environment? *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 2762-2777.

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* **1**, 101-102.

Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular Biology and Evolution* **27**, 2534-2547.

Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution* **23**, 327-337.

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations* Blackwell Publishing UK.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

Altshuler D, Pollara VJ, Cowles CR*, et al.* (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**, 513-516.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F(st)-outlier method. *Bmc Bioinformatics* **9**.

Ayllon F, Martinez JL, Garcia-Vazquez E (2006) Loss of regional population structure in Atlantic salmon, *Salmo salar* L., following stocking. *Ices Journal of Marine Science* **63**, 1269-1273.

Barrett RDH, Rogers SM, Schluter D (2008) Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255-257.

Barton NH (1979) Dynamics of hybrid zones. *Heredity* **43**, 341-359.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Beaugrand G, Reid PC (2012) Relationships between North Atlantic salmon, plankton, and

hydroclimatic change in the Northeast Atlantic. *Ices Journal of Marine Science* **69**, 1549-1562.

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

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences* **263**, 1619-1626.

Belding DL (1940) Migration of the Atlantic salmon (*Salmo salar*) in the Gulf of St. *Lawrence as Determined by Tagging Experiments Trans Amer Fish Soc* **69**, 290-295.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001) GENETIX 4.02, logiciel sous windows TM pour la génétique des populations. Laboratoire Génome, Population, Interactions: CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

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

Berg O (1985) The formation of non-anadromous populations of Atlantic salmon, *Salmo salar* L., in Europe. *Journal of Fish Biology* **27**, 805-815.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology* **20**, 2044-2072.

Bonin A (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology* **17**, 3583-3584.

Bonin A, Nicole F, Pompanon F, Miaud C, Taberlet P (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology* **21**, 697-708.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana*

temporaria). *Molecular Biology and Evolution* **23**, 773-783.

Boulding EG, Culling M, Glebe B, Berg PR, Lien S, Moen T (2008) Conservation genomics of Atlantic salmon: SNPs associated with QTLs for adaptive traits in parr from four trans-Atlantic backcrosses. *Heredity* **101**, 381-391.

Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L (2013) Landscape genomics in Atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation. *Evolution*, DOI: 10.1111/evo.12139.

Bourret V, Kent MP, Primmer CR*, et al.* (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **22**, 532-551.

Bourret V, O'Reilly PT, Carr JW, Berg PR, Bernatchez L (2011) Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity* **106**, 500-510.

Bradbury IR, Hubert S, Higgins B*, et al.* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B-Biological Sciences* **277**, 3725-3734.

Brenna-Hansen S, Li J, Kent MP*, et al.* (2012) Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence in situ hybridization analysis. *Bmc Genomics* **13**, doi:10.1186/1471-2164-1113-1432.

Carey DJ (1997) Syndecans: Multifunctional cell-surface co-receptors. *Biochemical Journal* **327**, 1-16.

Carlson SM, Seamons TR (2008) A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. *Evolutionary Applications* **1**, 222-238.

Carr JW, Anderson JM, Whoriskey FG, Dilworth T (1997) The occurrence and spawning of cultured Atlantic salmon (*Salmo salar*) in a Canadian river. *Ices Journal of Marine Science* **54**, 1064-1073.

Carr JW, Whoriskey F, O'Reilly P (2004) Efficacy of releasing captive reared broodstock into an imperilled wild Atlantic salmon population as a recovery strategy. *Journal*

of Fish Biology **65**, 38-54.

Carr JW, Whoriskey FG (2006) The escape of juvenile farmed Atlantic salmon from hatcheries into freshwater streams in New Brunswick, Canada. *Ices Journal of Marine Science* **63**, 1263-1268.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* **70**, 155-174.

Claytor RR, Maccrimmon HR, Gots BL (1991) Continental and ecological variance-components of European and North-American Atlantic salmon (*Salmo salar*) phenotypes. *Biological Journal of the Linnean Society* **44**, 203-229.

Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology* **69**, 21-47.

Conover DO, Schultz ET (1995) Phenotypic similarity and the evolutionary significance of countergradient variation *Trends in Ecology & Evolution* **10**, 248-252.

Cooke GM, Chao NL, Beheregaray LB (2012) Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. *Molecular Ecology* **21**, 2410-2427.

Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates, Sunderland, USA.

Craig JE, Hewitt AW, McMellon AE*, et al.* (2009) Rapid inexpensive genome-wide association using pooled whole blood. *Genome Research* **19**, 2075-2080.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* **15**, 290-295.

Cresko W, Amores A, Wilson C, *et al.* (2004) Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences USA*. **101**, 6050-6055.

Daetwyler HD, Schenkel FS, Sargolzaei M, Robinson JAB (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy*

Science **91**, 3225-3236.

Dalgaard MB, Nielsen CV, Buchmann K (2003) Comparative susceptibility of two races of *Salmo salar* (Baltic Lule river and Atlantic Conon river strains) to infection with *Gyrodactylus salaris*. *Diseases of Aquatic Organisms* **53**, 173-176.

Daub JT, Hofer T, Cutivet E*, et al.* (2013) Evidence for Polygenic Adaptation to Pathogens in the Human Genome. *Molecular Biology and Evolution* **30**, 1544-1558.

Davidson WS, Koop BF, Jones SJM*, et al.* (2010) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* **11**, 403-409.

De Mita S, Thuillet A-C, Gay L*, et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* **22**, 1383-1399.

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**, 22-32.

Dekkers JCM, Van der Werf JHJ (2007) Chapter 10: Strategies, limitations, and opportunities for marker-assisted selection in livestock. In: *Marker-assisted selection - current status and future perspectives in crops, livestock, forestry and fish* (eds. Guimarães EP, Ruane J, Scherf BD, Sonnino A, Dargie JD), pp. 167-184. FAO, Rome.

DFO (2009) Canada's Policy for Conservation of Wild Atlantic Salmon (ed. Canada DFO).

Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology* **17**, 2382-2396.

Dionne M, Caron F, Dodson JJ, Bernatchez L (2009) Comparative survey of within-river genetic structure in Atlantic salmon; relevance for management and conservation. *Conservation Genetics* **10**, 869-879.

Dionne M, Cauchon V, Harnois N, Fournier D (2013) Écologie et évolution des populations témoins de saumon atlantique au Québec : Rapport de recherche 2012, p. 80. Ministère des Ressources naturelles et de la Faune, Québec.

Dionne M, Miller KM, Dodson JJ, Bernatchez L (2009) MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society B-Biological Sciences* **364**, 1555-1565.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (2007) Clinal variation in mhc diversity with temperature: Evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution* **61**, 2154-2164.

Dittman AH, Quinn TP, Nevitt GA (1996) Timing of imprinting to natural and artificial odors by coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 434-442.

Dominik S, Henshall JM, Kube PD*, et al.* (2010) Evaluation of an Atlantic salmon SNP chip as a genomic tool for the application in a Tasmanian Atlantic salmon (*Salmo salar*) breeding population. *Aquaculture* **308**, S56-S61.

Donnelly WA, Dill LM (1984) Evidence for Crypsis in Coho Salmon, *Oncorhynchus-Kisutch* (Walbaum), Parr - Substrate Color Preference and Achromatic Reflectance. *Journal of Fish Biology* **25**, 183-195.

Dutil JD, Coutu JM (1988) Early marine life of Atlantic salmon, *Salmo salar*, postsmolts in the northern Gulf of St-Lawrence. *Fishery Bulletin* **86**, 197-212.

Endler JA (1986) *Natural Selection in the Wild* Princeton University Press, Princeton.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**, 1729-1747.

Ferguson A, Fleming IA, Hindar K*, et al.* (2007) Farm Escapes. In: *The Atlantic Salmon - Genetics, Conservation and Management* (eds. Verspoor E, Stradmeyer L, Nielsen JL), pp. 357-398. Blackwell Publishing, Oxford.

Fleming IA, Hindar K, Mjolnerod IB, Jonsson B, Balstad T, Lamberg A (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.

Flori L, Fritz S, Jaffrezic F*, et al.* (2009) The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. *Plos One* **4**, e6595.

Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate

for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977-993.

Fontaine PM, Dodson JJ, Bernatchez L, Slettan A (1997) A genetic test of metapopulation structure in Atlantic salmon (*Salmo salar*) using microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 2434-2442.

Ford JK, Maccallum RC, Tait M (1986) The application of exploratory factor analysis in applied psychology - A critical review and analysis. *Personnel Psychology* **39**, 291-314.

Ford JS, Myers RA (2008) A global assessment of salmon aquaculture impacts on wild salmonids. *PLoS Biology* **6**, 0411-0417.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics* Cambridge University Press, UK.

Fraser DJ, Cook AM, Eddington JD, Bentzen P, Hutchings JA (2008) Mixed evidence for reduced local adaptation in wild salmon resulting from interbreeding with escaped farmed salmon: complexities in hybrid fitness. *Evolutionary Applications* **1**, 501-512.

Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity* **106**, 404-420.

Friedland KD, Hansen LP, Dunkley DA, MacLean JC (2000) Linkage between ocean climate, post-smolt growth, and survival of Atlantic salmon (*Salmo salar* L.) in the North Sea area. *Ices Journal of Marine Science* **57**, 419-429.

Friedland KD, Reddin DG, Castonguay M (2003) Ocean thermal conditions in the post-smolt nursery of North American Atlantic salmon. *Ices Journal of Marine Science* **60**, 343-355.

Friedland KD, Reddin DG, McMenemy JR, Drinkwater KF (2003) Multidecadal trends in North American Atlantic salmon (*Salmo salar*) stocks and climate trends relevant to juvenile survival. *Canadian Journal of Fisheries and Aquatic Sciences* **60**, 563-583.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27**, 489-496.

Gaggiotti OE, Bekkevold D, Jorgensen HBH*, et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution* **63**, 2939-2951.

Gagnaire PA, Normandeau E, Bernatchez L (2012) Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. *Molecular Biology and Evolution* **29**, 2909-2919.

Gagnaire PA, Normandeau E, Côté C, Hansen MM, Bernatchez L (2012) The effect of spatially varying selection on locally adaptive alleles in the panmictic American eel (*Anguilla rostrata*). *Genetics* **190**, 725-736.

Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* **67**, 2483-2497.

Garcia de Leaniz C, Fleming IA, Einum S*, et al.* (2007) A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews* **82**, 173-211.

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. *Plant Cell* **15**, 1502-1506.

Gauthier-Ouellet M, Dionne M, Caron F, King TL, Bernatchez L (2009) Spatiotemporal dynamics of the Atlantic salmon (*Salmo salar*) Greenland fishery inferred from mixed-stock analysis. *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 2040-2051.

Gay L, Crochet PA, Bell DA, Lenormand T (2008) Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution* **62**, 2789-2806.

Gibbs RA, Taylor JF, Van Tassell CP*, et al.* (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528-532.

Gibson AJF, Bowlby HD, Hardie DC, O'Reilly PT (2011) Populations on the brink: low abundance of southern upland Atlantic salmon in Nova Scotia, Canada. *North American Journal of Fisheries Management* **31**, 733-741.

Gidskehaug L, Kent M, Hayes BJ, Lien S (2011) Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics* **27**,

303-310.

Glover KA, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala O (2013) Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *Bmc Genetics* **14**, 19p.

Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. *Genome Research* **8**, 195-202.

Gotz S, Garcia-Gomez JM, Terol J*, et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420-3435.

Goudet J, Perrin N, Waser P (2002) FSTAT, a Programm to Estimate and Test Gene Diversities and Fixation Indices. Institut d'Écologie, Université de Lausanne, Lausanne.

Griffiths AM, Machado-Schiaffino G, Dillane E*, et al.* (2010) Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in the southern part of the European range. *Bmc Genetics* **11**, 31.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372.

Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A (2010) Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 2459-2468.

Hansen LP, Quinn TR (1998) The marine phase of the Atlantic salmon (*Salmo salar*) life cycle, with comparisons to Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 104-118.

Hansen MM, Meier K, Mensberg KLD (2010) Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Molecular Ecology* **19**, 1787-1800.

Hansen MM, Ruzzante DE, Nielsen EE, Bekkevold D, Mensberg KLD (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Molecular Ecology* **11**, 2523-2535.

Hatta Y, Tsuchiya N, Matsushita M, Shiota M, Hagiwara K, Tokunaga K (1999)

Identification of the gene variations in human CD22. *Immunogenetics* **49**, 280-286.

Havlak P, Chen R, Durbin KJ*, et al.* (2004) The atlas genome assembly system. *Genome Research* **14**, 721-732.

Hayes B, Laerdahl JK, Lien S*, et al.* (2007) An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar)* expressed sequences. *Aquaculture* **265**, 82-90.

Hedrick PW (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**, 313-318.

Hedrick PW (2005) *Genetics of Population*, 3rd edition edn. Jones & Bartlett Publishing, Sudbury, USA.

Hewitt GM (1988) Hybrid zones - Natural laboratories for evolutionary studies. *Trends in Ecology & Evolution* **3**, 158-167.

Hindar K, Ryman N, Utter F (1991) Genetic effects of cultured fish on natural fish populations. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 945-957.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* **11**, 117-122.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics* **6**, e1000862.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* **58**, 199-207.

Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika* **30**, 179-185.

Hutchings JA, Fraser DJ (2008) The nature of fisheries- and farming-induced evolution. *Molecular Ecology* **17**, 294-313.

Hvidsten NA, Jensen AJ, Rikardsen AH*, et al.* (2009) Influence of sea temperature and initial marine feeding on survival of Atlantic salmon *Salmo salar* post-smolts from the Rivers Orkla and Hals, Norway. *Journal of Fish Biology* **74**, 1532-1548.

ICES (2013) Report of the working group on North Atlantic salmon (WGNAS), p. 378, Copenhagen, Denmark.

Johannesson K, Andre C (2006) Life on the margin: genetic isolation and diversity loss in

a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology* **15**, 2013-2029.

Johnson LB, Richards C, Host GE, Arthur JW (1997) Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology* **37**, 193-+.

Jonsson B, Forseth T, Jensen AJ, Naesje TF (2001) Thermal performance of juvenile Atlantic Salmon, *Salmo salar* L. *Functional Ecology* **15**, 701-711.

Jonsson B, Jonsson N (2006) Cultured Atlantic salmon in nature: a review of their ecology and interaction with wild fish. *Ices Journal of Marine Science* **63**, 1162-1181.

Joost S, Bonin A, Bruford MW*, et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955-3969.

Karlsson S, Moen T, Hindar K (2010) An extended panel of single nucleotide polymorphisms in Atlantic salmon (*Salmo salar* L.) mitochondrial DNA. *Conservation Genetics* **11**, 1171-1175.

Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011) Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular Ecology Resources* **11**, 247-253.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.

Kazakov RV, Titov SF (1991) Geographical patterns in the population-genetics of Atlantic salmon, *Salmo-salar* L, on USSR territory, as evidence for colonization routes. *Journal of Fish Biology* **39**, 1-6.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology* **10**, 807-821.

King TL, Verspoor E, Spidle AP*, et al.* (2007) Biodiversity and population structure. In: *The Atlantic Salmon - Genetics, Conservation and Management* (eds. Verspoor E, Stradmeyer L, Nielsen JL), pp. 117-166. Blackwell Publishing, Oxford.

Knox D, Lehman K, Reddin DG, Verspoor E (2002) Genotyping of archival Atlantic salmon scales from northern Québec & West Greenland using novel PCR primers for degraded DNA. *Journal of Fish Biology* **60**, 266-270.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation

genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Koljonen ML (2001) Conservation goals and fisheries management units for Atlantic salmon in the Baltic Sea area. *Journal of Fish Biology* **59**, 269-288.

Koskinen MT, Sundell P, Piironen J, Primmer CR (2002) Genetic assessment of spatiotemporal evolutionary relationships and stocking effects in grayling (*Thymallus thymallus*, Salmonidae). *Ecology Letters* **5**, 193-205.

Lacy RC (1997) Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy* **78**, 320-335.

Lamaze FC, Sauvage C, Marie A, Garant D, Bernatchez L (2012) Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*). *Molecular Ecology* **21**, 2877-2895.

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.

Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *American Naturalist* **151**, 283-292.

Lautenschlager GJ (1989) A comparison of alternatives to conducting Monte-Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research* **24**, 365-395.

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics* **164**, 1205-1219.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* **21**, 1548-1566.

Leberg PL (2002) Estimating allelic richness: Effects of sample size and bottlenecks. *Molecular Ecology* **11**, 2445-2449.

Leclerc E, Mailhot Y, Mingelbier M, Bernatchez L (2008) The landscape genetics of yellow perch (*Perca flavescens*) in a large fluvial ecosystem. *Molecular Ecology* **17**, 1702-1717.

Lee C-R, Mitchell-Olds T (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* **20**, 4631-4642.

Lefèvre MA, Stokesbury MJW, Whoriskey FG, Dadswell MJ (2012) Atlantic salmon post-smolt migration routes in the Gulf of St. Lawrence. *Ices Journal of Marine Science* **69**, 981-990.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* **74**, 175-195.

Lien S, Gidskehaug L, Moen T*, et al.* (2011) A dense SNP-based linkage map of Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *Bmc Genomics* **12**, 615.

Limborg MT, Pedersen JS, Hemmer-Hansen J, Tomkiewicz J, Bekkevold D (2009) Genetic population structure of European sprat *Sprattus sprattus*: differentiation across a steep environmental gradient in a small pelagic fish. *Marine Ecology-Progress Series* **379**, 213-224.

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128-2129.

Lorenz S, Brenna-Hansen S, Moen T*, et al.* (2010) BAC-based upgrading and physical integration of a genetic SNP map in Atlantic salmon. *Animal Genetics* **41**, 48-54.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, Article published online: 11 APR 2014 | DOI: 10.1111/mec.12725

Ma X-F, Hall D, St Onge KR, Jansson S, Ingvarsson PK (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics* **186**, 1033-1044.

Macgregor S, Zhao ZZ, Henders A, Martin NG, Montgomery GW, Visscher PM (2008) Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. *Nucleic Acids Research* **36**.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**, 565-577.

MacKenzie KM, Trueman CN, Palmer MR*, et al.* (2012) Stable isotopes reveal age-dependent trophic level and spatial segregation during adult marine feeding in populations of salmon. *Ices Journal of Marine Science* **69**, 1637-1645.

Makhrov AA, Verspoor E, Artamonova VS, O'Sullivan M (2005) Atlantic salmon

colonization of the Russian Arctic coast: pioneers from North America. *Journal of Fish Biology* **67**, 68-79.

Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution* **28**, 614-621.

Manel S, Joost S, Epperson BK*, et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**, 3760-3772.

Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology* **19**, 3824-3835.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Marie AD, Bernatchez L, Garant D (2010) Loss of genetic integrity correlates with stocking intensity in brook charr (*Salvelinus fontinalis*). *Molecular Ecology* **19**, 2025-2037.

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.

McGinnity P, Stone C, Taggart JB*, et al.* (1997) Genetic impact of escaped farmed Atlantic salmon (*Salmo salar* L.) on native populations: use of DNA profiling to assess freshwater performance of wild, farmed, and hybrid progeny in a natural river environment. *Ices Journal of Marine Science* **54**, 998-1008.

Metcalfe NB, Thorpe JE (1990) Determinants of geographical variation in the age of seaward migrating salmon, *Salmo salar*. *Journal of Animal Ecology* **59**, 135-145.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-1829.

Miller K, Li S, Kaukinen KH, *et al.* (2011) Genomic signatures predict migration and spawning failure in wild Canadian salmon. *Science* **331**, 214-217.

Mills KE, Pershing AJ, Sheehan TF, Mountain D (2013) Climate and ecosystem linkages explain widespread declines in North American Atlantic salmon populations.

Global Change Biology **19**, 3046-3061.

Mjolnerod IB, Refseth UH, Karlsen E, Balstad T, Jakobsen KS, Hindar K (1997) Genetic differences between two wild and one farmed population of Atlantic salmon (*Salmo salar*) revealed by three classes of genetic markers. *Hereditas* **127**, 239-248.

Moen T, Hayes B, Baranski M*, et al.* (2008) A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *Bmc Genomics* **9**.

Moritz C (1994) Defining Evolutionary-Significant-Units for conservation. *Trends in Ecology & Evolution* **9**, 373-375.

Mosteller F, Fisher RA (1948) Questions and Answers. *The American Statistician* **2**, 30-31.

MPO (2009) Politique du Canada pour la conservation du saumon atlantique sauvage, http://www.dfo-mpo.gc.ca/fm-gp/policies-politiques/wasp-pss/wasp-psas-2009-fra.htm.

Nadeau NJ, Jiggins CD (2011) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics* **26**, 484-492.

Naish KA, Hard JJ (2008) Bridging the gap between the genotype and the phenotype: linking genetic variation, selection and adaptation in fishes. *Fish and Fisheries* **9**, 396-422.

Narum SR, Hess JE (2011) Comparison of *FST* outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184-194.

Nei M (1977) F-Statistics and analysis of gene diversity in subdivided populations. *Annals of Human Genetics* **41**, 225-233.

Nicieza AG, Reiriz L, Brana F (1994) Variation in digestive performance between geographically disjunct populations of Atlantic salmon - countergradient in passage time and digestion rate. *Oecologia* **99**, 243-251.

Nicieza AG, Reyesgavilan FG, Brana F (1994) Differentiation in juvenile growth and bimodality patterns between northern and southern-populations of Atlantic salmon (*Salmo salar* L). *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **72**, 1603-1610.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**, 3128-3150.

Normandeau E, Hutchings JA, Fraser DJ, Bernatchez L (2009) Population-specific gene expression responses to hybridization between farm and wild Atlantic salmon. *Evolutionary Applications* **2**, 489-503.

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.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**, 375-402.

O'Reilly PT, Hamilton LC, McConnell SK, Wright JM (1996) Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 2292-2298.

Paez DJ, Morrissey M, Bernatchez L, Dodson JJ (2010) The genetic basis of early-life morphological traits and their relation to alternative male reproductive tactics in Atlantic salmon. *Journal of Evolutionary Biology* **23**, 757-768.

Palstra FP, O'Connell MF, Ruzzante DE (2007) Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Molecular Ecology* **16**, 4504-4522.

Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology* **17**, 3428-3447.

Parisod C, Holderegger R (2012) Adaptive landscape genetics: pitfalls and benefits. *Molecular Ecology* **21**, 3644-3646.

Paterson S, Piertney SB, Knox D, Gilbey J, Verspoor E (2004) Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes* **4**, 160-162.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *Plos Genetics* **2**, 2074-2093.

Pavey SA, Dodson JJ (2012) At the interface of behaviour, ecology and evolution: Insights from the world of fishes. *Current Zoology* **58**.

Pavlidis P, Jensen JD, Wolfgang S, Stamatakis A (2012) A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution* **X**, doi:10.1093/molbev/mss1136.

Perrier C, Guyomard R, Bagliniere J-L, Evanno G (2011) Determinants of hierarchical genetic structure in Atlantic salmon populations: environmental factors vs. anthropogenic influences. *Molecular Ecology* **20**, 4231-4245.

Petren K (2013) The evolution of landscape genetics. *Evolution* **67**, 3383-3385.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. *Molecular Ecology* **14**, 2943-2957.

Plantalech manel-la N, Chittenden CM, Okland F*, et al.* (2011) Does river of origin influence the early marine migratory performance of *Salmo salar*? *Journal of Fish Biology* **78**, 624-634.

Power G (1981) Stock characteristics and catches of Atlantic salmon (*Salmo salar*) in Québec, and Newfoundland and Labrador in relation to environmental variables. *Canadian Journal of Fisheries and Aquatic Sciences* **38**, 1601-1611.

Primack RB (2006) *Essentials of Conservation Biology*, 4th edition edn. Sinauer Associates, Sunderland, USA.

Primmer CR (2009) From Conservation Genetics to Conservation Genomics. In: *Year in Ecology and Conservation Biology 2009* (eds. Ostfeld RS, Schlesinger WH), pp. 357-368. Blackwell Publishing, Oxford.

Pritchard JB (2003) The gill and homeostasis: transport under stress. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology* **285**, R1269-R1271.

Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics* **11**, 665-667.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution* **49**, 1280-1283.

Rees CB, McCormick SD, Li WM (2005) A non-lethal method to estimate CYP1A

expression in laboratory and wild Atlantic salmon (*Salmo salar*). *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology* **141**, 217-224.

Refstie T (1983) Induction of diploid gynogenesis in Atlantic salmon and rainbow-trout using irradiated sperm and heat-shock. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **61**, 2411-2416.

Rice WR (1989) Analyzing Tables of Statistical Tests. *Evolution* **43**, 223-225.

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, Normandeau E, Einum S, Guderley H, Bernatchez L (2008) Genetic consequences of interbreeding between farmed and wild Atlantic salmon: insights from the transcriptome. *Molecular Ecology* **17**, 314-324.

Rogers SM, Bernatchez L (2007) The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonas* sp Salmonidae) species pairs. *Molecular Biology and Evolution* **24**, 1423-1438.

Ryder OA (1986) Species Conservation and Systematics - the Dilemma of Subspecies. *Trends in Ecology & Evolution* **1**, 9-10.

Ryman N, Jorde PE (2001) Statistical power when testing for genetic differentiation. *Molecular Ecology* **10**, 2361-2373.

Säisä M, Koljonen ML, Gross R*, et al.* (2005) Population genetic structure and postglacial colonization of Atlantic salmon (*Salmo salar*) in the Baltic Sea area based on microsatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 1887-1904.

Schmidt PS, Zhu CT, Das J, Batavia M, Yang L, Eanes WF (2008) An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16207-16211.

Sheehan TF, Reddin DG, Chaput G, Renkawitz MD (2012) SALSEA North America: a pelagic ecosystem survey targeting Atlantic salmon in the Northwest Atlantic. *Ices Journal of Marine Science* **69**, 1580-1588.

Skaala O, Hoyheim B, Glover K, Dahle G (2004) Microsatellite analysis in domesticated

and wild Atlantic salmon (*Salmo salar* L.): allelic diversity and identification of individuals. *Aquaculture* **240**, 131-143.

Skaala O, Makhrov AA, Karlsen T*, et al.* (1998) Genetic comparison of salmon from the White Sea and north-western Atlantic Ocean. *Journal of Fish Biology* **53**, 569-580.

Skaala O, Wennevik V, Glover KA (2006) Evidence of temporal genetic change in wild Atlantic salmon, *Salmo salar* L., populations affected by farm escapees. *Ices Journal of Marine Science* **63**, 1224-1233.

Soller M (1990) Genetic-mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. *Journal of Dairy Science* **73**, 2628-2646.

Soto D, Jara F, Moreno C (2001) Escaped salmon in the inner seas, southern Chile: Facing ecological and social conflicts. *Ecological Applications* **11**, 1750-1762.

Stabell OB (1984) Homing and olfaction in salmonids - A critical review with special reference to the Atlantic salmon. *Biological Reviews of the Cambridge Philosophical Society* **59**, 333-388.

Ståhl G (1987) Genetic population structure of Atlantic salmon. In: *Population genetics and fisheries management* (eds. Ryman N, Utter F), pp. 212-240. University of Washington Press, Seattle.

Stallard RF, Edmond JM (1983) Geochemistry of the Amazon 2. The influence of geology and weathering environment on the dissolved-load. *Journal of Geophysical Research-Oceans and Atmospheres* **88**, 9671-9688.

Stenseth NC, Ottersen G, Hurrell JW*, et al.* (2003) Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Nino Southern Oscillation and beyond. *Proceedings of the Royal Society B-Biological Sciences* **270**, 2087-2096.

Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* **64**, 2489-2509.

Suekawa Y, Aihara H, Araki M, Hosokawa D, Mannen H, Sasazaki S (2010) Development of breed identification markers based on a bovine 50K SNP array. *Meat Science* **85**, 285-288.

Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Koster H (1999) Chip-based genotyping by

mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 10016-10020.

Taylor EB (1991) A review of local adaptation in salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* **98**, 185-207.

Tessier N, Bernatchez L (2000) A mtDNA and microsatellite assessment of single *vs* multiple origins of landlocked Atlantic salmon (*Salmo salar* L.) from Lake Saint-Jean, Québec, Canada. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 797-804.

Theodorou K, Couvet D (2004) Introduction of captive breeders to the wild: Harmful or beneficial? *Conservation Genetics* **5**, 1-12.

Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology* **19**, 2301-2314.

Thorstad EB, Whoriskey F, Uglem I, Moore A, Rikardsen AH, Finstad B (2012) A critical life stage of the Atlantic salmon *Salmo salar*: behaviour and survival during the smolt and initial post-smolt migration. *Journal of Fish Biology* **81**, 500-542.

Todd CD, Hughes SL, Marshall CT, Maclean JC, Lonergan ME, Biuw EM (2008) Detrimental effects of recent ocean surface warming on growth condition of Atlantic salmon. *Global Change Biology* **14**, 958-970.

Tonteri A, Titov S, Veselov A*, et al.* (2005) Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from northern Europe. *Annales Zoologici Fennici* **42**, 1-22.

Tonteri A, Veselov AJ, Titov S, Lumme J, Primmer CR (2007) The effect of migratory behaviour on genetic diversity and population divergence: a comparison of anadromous and freshwater Atlantic salmon *Salmo salar*. *Journal of Fish Biology* **70**, 381-398.

Tonteri A, Veselov AJ, Zubchenko AV, Lumme J, Primmer CR (2009) Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo salar*) populations from the Barents and White seas, northwest Russia. *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 717-735.

Vähä J-P, Erkinaro J, Niemela E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*

16, 2638-2654.

Vähä JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* **15**, 63-72.

van Tienderen PH, de Haan AA, van der Linden CG, Vosman B (2002) Biodiversity assessment using markers for ecologically important traits. *Trends in Ecology & Evolution* **17**, 577-582.

Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution* **22**, 1067-1076.

Vasemägi 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, Beardmore JA, Consuegra S*, et al.* (2005) Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *Journal of Fish Biology* **67**, 3-54.

Verspoor E, Stradmeyer L, Nielsen JL (2007) *The Atlantic Salmon: Genetics, Conservation, and Management* Blackwell Publishing, Oxford.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* **17**, 4334-4345.

Vilas A, Perez-Figueroa A, Caballero A (2012) A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology* **25**, 1364-1376.

Volpe JP, Taylor EB, Rimmer DW, Glickman BW (2000) Evidence of natural reproduction of aquaculture-escaped Atlantic salmon in a coastal British Columbia river. *Conservation Biology* **14**, 899-903.

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of species under the endangered species act. *Marine Fisheries Review* **53**, 11-22.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* **38**, 1358-1370.

Wennevik V, Skaala O, Titov SF, Studyonov I, Naevdal G (2004) Microsatellite variation in populations of Atlantic salmon from North Europe. *Environmental Biology of Fishes* **69**, 143-152.

Willing EM, Bentzen P, van Oosterhout C*, et al.* (2010) Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology* **19**, 968-984.

Yang JA, Benyamin B, McEvoy BP*, et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-U131.

Yano A, Guyomard R, Nicol B*, et al.* (2012) An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Current Biology* **22**, 1423-1428.

Yeomans KA, Golder PA (1982) The Guttman-Kaiser criterion as a predictor of the number of common factors. *Statistician* **31**, 221-229.