



Analyse génomique de la sélection spatialement variable chez l'anguille d'Amérique (*Anguilla rostrata*)

Mémoire

Charles Babin

Maîtrise en biologie
Maître ès sciences (M. Sc.)

Québec, Canada

© Charles Babin, 2017

Analyse génomique de la sélection spatialement variable chez l'anguille d'Amérique (*Anguilla rostrata*)

Mémoire

Sous la direction de :

Louis Bernatchez, directeur de recherche

Résumé

L'anguille d'Amérique est un poisson avec un cycle de vie très particulier. En effet, elle occupe une aire de répartition qui s'étire du Groenland aux Caraïbes, mais tous les individus se reproduisent dans la mer des Sargasses. Après la reproduction, les larves sont dispersées de façon aléatoire jusqu'aux côtes. Ce lieu de reproduction unique fait en sorte que tous les individus de l'espèce appartiennent à la même population. Par contre, les conditions environnementales varient grandement au sein de l'aire de répartition, puisque celle-ci s'étend de régions subarctiques à des régions subtropicales, ce qui confronte les individus à des conditions différentes selon l'endroit jusqu'où ils dérivent et peut entraîner la sélection d'allèles différents selon les régions. Les objectifs de cette étude étaient d'identifier les régions du génome soumises au phénomène de sélection spatialement variable et quels mécanismes sont affectés par la sélection. Pour ce faire, 710 individus en provenance de 13 sites différents représentant une grande partie de l'aire de répartition de l'espèce ont été séquencés. Un total de 12 098 SNP a été obtenu. Des méthodes d'association environnementale et d'analyse de redondance ont été employées pour identifier des marqueurs potentiellement sous sélection spatialement variable. Un total de 183 marqueurs a été identifié comme étant sous sélection spatialement variable. L'interaction entre les différentes régions sous sélection a également été évaluée en utilisant des scores polygéniques additifs. Des corrélations significatives entre ces scores polygéniques et la latitude, la longitude et la température ont été identifiées. Finalement, nous avons identifié les gènes à proximité des marqueurs potentiellement sous sélection. Parmi ces gènes, le mécanisme de réponse à l'insuline était le seul mécanisme significativement enrichi. Cette étude a permis de mieux documenter l'étendue de la sélection spatialement variable chez l'anguille d'Amérique en montrant qu'il semble y avoir de la sélection dans de nombreuses régions du génome.

Abstract

The American eel is a fish with a complex life cycle. The eel occupy a wide species range from Greenland to the Caribbean, but all eels reproduce in the Sargasso Sea. After the reproduction, the larvea are advected randomly to the coast by ocean currents. Because of this reproduction mode, all the American Eel are in the same population. On the other hand, the range is extending from subarctic to subtropical regions and the eels occupying these different regions are facing really different environmental conditions. These differents conditions could result in the selection of different alleles. The objective of this study was to identify the different regions of the genome that are affected by this phenomenon of spatially-varying selection and which mecanisms are affected by selection. A total of 710 glass eels captured in 12 different sites representing an important part of the species range were sequenced to reach these objectives. After sequencing, 12 098 SNPs were conserved for further analysis. Using environmental association and redundancy analyses approaches, 183 of these markers were identified to be potentially under spatially-varying selection. The interaction between these differents regions was analyzed using additive polygenic scores. Significant correlations were identified between these polygenic scores and the latitude, longitude and temperature. Genes close to outliers were identified and gene ontology analyses were made. The only significantly enriched pathway was the insuline signalling pathway. With this study our understanding of the spatially-varying selection in the American Eel has been increased.

Table des matières

Résumé.....	iii
Abstract.....	iv
Table des matières.....	v
Liste des tableaux.....	vi
Liste des figures.....	vii
Remerciements.....	viii
Avant-propos.....	x
Introduction.....	1
Forces évolutives.....	1
Panmixie.....	2
Techniques de séquençage de nouvelle génération.....	3
Anguille d'Amérique.....	3
Cycle vital.....	3
Sélection spatialement variable.....	5
Déclin des populations d'anguilles.....	7
Objectif du projet.....	7
Chapter 1: RAD-seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American Eel (<i>Anguilla rostrata</i>).....	9
Résumé.....	10
Abstract.....	11
Conclusion.....	38
Identification de marqueurs potentiellement sous sélection spatialement variable.....	38
Effet additif des marqueurs potentiellement sous sélection.....	39
Fonctions enrichies chez les marqueurs potentiellement sous sélection.....	40
Bibliographie.....	42

Liste des tableaux

Table 1 Names and codes of sampling locations, number of individuals per site, latitude, longitude, river mouth temperature and sampling year. Temperatures for 2008 are the average sea-surface temperature near the river mouth during the 10 days prior to the sampling day during recruitment. Temperatures for 2009 are the average sea-surface temperature of the river mouth during the sampling month.....18

Table 2 Correlation coefficients and p-values of the relationship between the 10th and 90th centile of the polygenic scores of each of the sampling sites and the environmental variables. The correlation coefficient and p-values were calculated following a quadratic and a linear model.....27

Liste des figures

Figure 1 Location map of the 13 sampling sites where American glass eels were collected for this study. Labels correspond to the site codes provided in table 1. The sites identified by squares were sampled in 2008 and 2009, sites identified by circles were sampled in 2008 and the site identified by a triangle was sampled in 2009.....16

Figure 2 (A) Redundancy analysis (RDA) performed with 12 098 filtered SNPs, using latitude, longitude, temperature and year as constraining variables. (B) Redundancy analysis performed using the 197 SNPs detected to be potentially affected by selection with the RDA outlier detection approach (C) Redundancy analysis performed using the 90 markers significantly associated with environmental variables in Bayenv2 analysis. Only the first 2 axes are shown for each RDA. Sites are identified by their code (Table 1) and by colored points reflecting water temperature value, from yellow for colder sites to dark red for warmer sites. The direction of main variation for each of the 4 constraining variables (Year, Latitude, Longitude, Temperature) is indicated by a blue vectors. SNPs contributions appear as red '+' symbols. The proportion of total variance explained by each axis is indicated in percent.....24

Figure 3 Histograms of the number of outliers detected in 100 simulations for the different environmental variables; latitude (A), longitude (B), temperature (C) and year (D). The x axis represents the number of outliers detected in a run. The y axis represents the number of simulations that detected this number of outlier. The vertical black bar represents the average number of outlier detected in the runs made with the real dataset. Note that the number of outliers detected in each run was higher than the number of outliers used for further analyses as reported in the main text. This is because not always the same outliers were detected in each run and in order to remain conservative in downstream analyses, we retained only the markers that were significant in 6 runs or more out of 12.....26

Figure 4 Correlations between additive individual polygenic scores based on all 259 outlier markers and each of 3 explanatory variables: latitude (A), longitude (B) and temperature (C). Additive polygenic scores were obtained at the individual level by summing the number of locally favourable alleles. Correlation coefficients and p-values of the quadratic correlation model are shown at the top of each panel.....28

Remerciements

J'ai toujours été curieux et je me suis toujours intéressé aux sciences. Au secondaire, les cours de biologie étaient parmi mes favoris surtout lorsqu'on parlait de l'évolution et de la génétique mendélienne. Je trouvais fascinant que l'alternance de 4 molécules chimiques puisse résulter en des organismes aussi différents que les virus, les bactéries, les méduses ou les champignons. Je trouve encore mystérieux que toute cette diversité ait pu évoluer simplement par le jeu des mutations. Lorsque le temps est venu de m'inscrire à l'université, je me suis dit que je pourrais poursuivre mes études en biologie pour assouvir ma soif de connaissances et ma curiosité par rapport aux questions de génétique.

Au cours de mon baccalauréat, j'ai voulu voir à quoi ressemblait la recherche et j'ai donc contacté Louis pour voir si je pouvais faire un stage dans son laboratoire afin de le découvrir. Louis me connaissait déjà, car j'ai côtoyé deux de ses fils au sein des scouts pendant une dizaine d'années. Il m'a offert de travailler au laboratoire durant l'été, si j'obtenais une bourse du CRSNG. J'ai obtenu la bourse et j'ai pu découvrir les rudiments du travail de laboratoire en aidant Simon Bernatchez dans son projet de maîtrise portant sur le touladi. J'ai beaucoup aimé faire du laboratoire et j'ai voulu faire une initiation à la recherche dans le cadre de mon bac. Le projet qui m'avait été confié était toutefois de grande ampleur et a fini par devenir la maîtrise dont vous allez lire le mémoire dans les prochaines pages.

J'aimerais remercier mon directeur Louis Bernatchez de m'avoir permis de faire partie de son laboratoire formés d'étudiants de très haut niveau qui m'ont permis de progresser grandement. Je le remercie également d'avoir cru en moi en me confiant un projet de séquençage de nouvelle génération avec des centaines d'individus alors que je n'étais qu'un étudiant au bac.

J'aimerais remercier Scott Pavey, qui m'a montré comment me servir d'un terminal et comment me débrouiller dans les méandres de la bio-informatique. Il m'a également aidé à développer la méthode que nous avons utilisée pour identifier nos marqueurs. Le génome de l'anguille d'Amérique qu'il a assemblé a grandement facilité les analyses.

Mon collaborateur Pierre-Alexandre Gagnaire a contribué de façon importante aux analyses en me faisant des suggestions d'analyses à réaliser et en m'aidant à interpréter les résultats qui ont découlé de

ces analyses. Il m'a également aidé à mieux comprendre les analyses.

Pour la partie laboratoire, j'aimerais remercier Clément Rougeux qui m'a aidé lors de la préparation des librairies en vue du séquençage. J'aimerais également remercier tous les autres membres du laboratoire qui m'ont permis de mener mon projet de maîtrise à bon port en faisant des suggestions.

Finalement, j'aimerais remercier ma famille qui m'a toujours encouragé à poursuivre mes études et m'a permis de développer ma curiosité.

Avant-propos

L'article qui figure à l'intérieur de ce mémoire s'intitule « RAD-seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American Eel (*Anguilla rostrata*) ». Il a été publié par le journal *Genome Biology and Evolution* dans son volume 9 issue 11 (<https://doi.org/10.1093/gbe/evx226>).

Les coauteurs de cet article sont Pierre-Alexandre Gagnaire, Scott Pavey et mon directeur Louis Bernatchez. Je suis l'auteur principal de l'article.

Pierre-Alexandre Gagnaire a contribué à l'analyse et à l'interprétation des résultats. J'ai effectué le travail de laboratoire ainsi que l'analyse et l'interprétation des résultats. J'ai aussi écrit le manuscrit. Tous les auteurs ont contribué à l'amélioration et à la révision du manuscrit.

Introduction

Forces évolutives

L'évolution est contrôlée par 4 forces évolutives. Trois d'entre elles sont neutres, la mutation, la dérive génétique et la migration. Elles sont dites neutres, car elles n'agissent pas de manière adaptative. Les mutations apparaissent de façon aléatoire un peu partout dans le génome et sont le moteur de l'évolution. Pour sa part, la dérive génétique fait varier stochastiquement à chaque génération la fréquence allélique des marqueurs au sein d'une population de façon plus ou moins importante en fonction de la taille effective de celle-ci. Elle engendre généralement une différenciation entre les populations, car les variations de fréquence allélique diffèrent entre les populations, puisqu'elles sont aléatoires et dépendent de la taille effective qui varie entre les populations. Les populations avec une taille effective moins importante vont avoir des fréquences alléliques qui vont varier de façon plus importante que celles avec une plus grande taille effective. La migration permet quant à elle d'échanger des allèles entre les populations, ce qui réduit la différenciation entre celles-ci. Plus le flux génique est important entre 2 populations, moins elles seront différenciées l'une de l'autre. La migration a donc un effet inverse de celui de la dérive génétique en homogénéisant les populations entre elles. La quatrième force évolutive, la sélection naturelle, est quant à elle dite directionnelle, car elle agit seulement sur quelques régions du génome en faisant varier en fréquence les marqueurs avantageux ou désavantageux dans certaines conditions, tout en n'affectant pas les autres. Les marqueurs sous sélection varient en fonction des pressions exercées par l'environnement et ne sont souvent pas les mêmes dans les différentes populations d'une même espèce, puisqu'elles ne sont pas confrontées aux mêmes conditions environnementales. Les phénotypes qui sont avantageux dans certaines conditions peuvent ainsi être sélectionnés (Herron et Freeman, 2014).

La sélection naturelle peut s'exercer de différentes façons. Si elle favorise les phénotypes dans une direction précise, elle sera dite directionnelle. Si les phénotypes qui sont favorisés sont les plus extrêmes, elle sera dite divergente, car les phénotypes moyens seront sélectionnés de façon négative. Lorsque ce sont plutôt les phénotypes moyens qui sont favorisés, la sélection est stabilisatrice. La pression de sélection dépend également de l'avantage sélectif de l'allèle. Si la mutation est très avantageuse pour les individus qui la portent, sa fréquence allélique augmentera plus rapidement que

celle d'une mutation qui procure moins d'avantages à son porteur. Les mutations qui sont fortement désavantageuses pour leurs porteurs seront de leur côté très négativement sélectionnées, car leurs porteurs auront de la difficulté à se reproduire, ce qui diminuera leur fréquence allélique de façon importante (Mitchell-Olds et al., 2007).

Panmixie

La plupart des espèces sont composées de plusieurs populations avec plus ou moins de connectivité entre elles. Les populations sont plus ou moins différenciées les unes des autres selon leur degré de connectivité. Elles s'adaptent souvent aux conditions présentes dans leur environnement, ce qui peut résulter en l'apparition d'adaptations locales au sein de celles-ci, après plusieurs générations de sélection pour les mêmes allèles. Ces adaptations locales procurent aux individus locaux un meilleur fitness dans les conditions locales comparativement à celui d'individus en provenance d'autres habitats soumis à ces mêmes conditions (Kawecki et Ebert, 2004). Toutefois, certaines espèces ne sont formées que d'une seule population. Chez ces espèces, le flux génique est si important qu'il prévient la différenciation entre les individus et empêche l'apparition d'une structure de populations. Ces espèces sont dites panmictiques, car la reproduction se fait à l'échelle de l'espèce de façon aléatoire (Wapples et Gaggiotti, 2006). Le développement d'adaptations locales chez les espèces panmictiques est impossible à cause de l'absence de structure de populations qui fait en sorte que les allèles se propagent librement au sein de l'aire de répartition de l'espèce. Malgré la sélection qui s'opère sur certains marqueurs avantageux dans une région de l'aire de répartition, il n'y a aucune barrière qui empêche les allèles qui sont avantageux ailleurs de parvenir jusqu'à cette région et les allèles avantageux dans ce site de parvenir à d'autres sites où ils seront moins avantageux (Kawecki et Ebert, 2004). Les travaux de Levene (1953) ont toutefois démontré que la sélection balancée peut maintenir un polymorphisme chez certains marqueurs, lorsque différents allèles sont favorisés au sein de différentes niches écologiques. Le polymorphisme peut ainsi se maintenir au sein d'espèces panmictiques à cause de la mortalité différentielle des individus en fonction de leur génotype dans les différents habitats. Cette mortalité différentielle fait en sorte que les fréquences alléliques peuvent varier en fonction des conditions environnementales, malgré l'absence de structure, car certains allèles seront plus avantageux dans certaines conditions et leurs porteurs auront donc un meilleur taux de survie. On nomme ce phénomène sélection spatialement variable en panmixie. Les anguilles sont de bons exemples de ce phénomène (Gagnaire et al., 2012).

Techniques de séquençage de nouvelle génération

Depuis quelques années, la biologie moléculaire a été profondément transformée par l'arrivée des techniques de séquençage de nouvelle génération, qui ont permis une augmentation exponentielle des données à un coût relativement faible. En effet, jusqu'à récemment les études sur les organismes non modèles ne portaient que sur quelques dizaines de marqueurs microsatellites à la fois, car il fallait développer des amorces pour chacun des marqueurs, ce qui prenait énormément de temps et coûtait assez cher. Maintenant, les techniques comme le RAD-sequencing ou le GBS permettent de séquencer des milliers de marqueurs SNP répartis un peu partout dans le génome sans avoir à faire de développement au préalable. Elles consistent à uniquement séquencer les régions voisines de site de restriction. Ces techniques permettent de séquencer seulement une partie du génome, ce qui réduit grandement les coûts de séquençage (Metzker, 2010). Cette augmentation du nombre de marqueurs analysés permet de scruter une plus grande partie du génome et donc de faire des analyses de génétique des populations plus précises. Il est également plus facile d'identifier des marqueurs potentiellement sous sélection et les mécanismes physiologiques qui sont affectés par ces phénomènes avec les nouvelles techniques de séquençage, puisque beaucoup plus de régions du génome sont examinées qu'auparavant. Les régions du génome soumises à la sélection ont donc plus de chances d'être séquencées et l'utilisation de plus de marqueurs permet de mieux évaluer la structure des populations. La réduction des coûts de séquençage a également permis l'assemblage d'un grand nombre de génomes au cours des dernières années. La disponibilité d'un génome de l'espèce étudiée ou d'une espèce proche permet de faciliter les différentes analyses bioinformatiques et de mieux identifier les gènes impliqués dans les processus d'adaptation en permettant de localiser précisément les marqueurs par rapport aux gènes (Ellegren, 2014).

Anguille d'Amérique

Cycle vital

L'anguille d'Amérique (*Anguilla rostrata*) est une espèce de poisson catadrome présente en Amérique, du Groenland aux Antilles et tout le long de la côte atlantique. Le cycle de vie de l'anguille est très complexe. Tout d'abord, toutes les anguilles d'Amérique naissent dans la mer des Sargasses dans l'océan Atlantique. Peu après leur éclosion, les larves se métamorphosent en leptocéphales. Ces larves

sont plates comme des feuilles. Cette forme leur permet de se laisser dériver au gré des courants marins jusqu'à la côte américaine (Vélez-Espino et Koops, 2010). Une fois sur le plateau continental, les leptocéphales se transforment en civelles, qui sont des larves cylindriques et transparentes. Arrivées sur la côte, elles peuvent y rester pour vivre en eau salée, aller vivre en eau saumâtre dans les estuaires ou remonter les cours d'eau pour passer leur vie en eau douce. Au cours de leur croissance, les anguilles deviendront d'abord des anguillettes, puis finiront par se pigmenter pour devenir des anguilles jaunes. C'est sous cette forme qu'elles passeront la majorité de leur vie. Après quelques années, elles se métamorphoseront de nouveau pour devenir cette fois des anguilles argentées. C'est à ce stade que les anguilles deviennent matures sexuellement. Sous cette forme, elles feront le chemin inverse vers la mer des Sargasses pour s'y reproduire et y mourir (Jessop, 2010).

Le cycle de vie des anguilles est longtemps resté mystérieux. En effet, la reproduction de l'anguille d'Amérique n'a jamais été observée et aucune adulte n'a jamais été vue dans la mer des Sargasses. La migration des adultes vers la mer des Sargasses pour s'y reproduire n'a pu être confirmée que très récemment en suivant la trajectoire vers la mer des Sargasses d'une anguille argentée femelle qui avait été marquée (Béguer-Pon et al., 2015). La mer des Sargasses avait auparavant été identifiée comme lieu de reproduction potentiel de l'espèce, car c'est à cet endroit que les leptocéphales les plus petites ont été capturées (Hanel et al., 2014). L'anguille d'Amérique partage son aire de reproduction avec son espèce sœur, l'anguille européenne (*Anguilla anguilla*). L'anguille européenne a un mode de vie et un cycle de vie similaire à l'anguille d'Amérique, mais elle passe la partie continentale de son cycle de vie en Europe et en Afrique du nord. Des hybrides entre les 2 espèces ont d'ailleurs été identifiés en Islande (Albert et al., 2006).

La durée de vie des anguilles d'Amérique varie beaucoup selon la latitude à laquelle les individus vont vivre et selon leur sexe (Jessop, 2010). En effet, les mâles atteignent le stade d'anguille argentée à une plus petite taille que les femelles. Ils sont aussi moins âgés lorsqu'ils atteignent ce stade. Le taux de croissance annuel des anguilles diminue avec l'augmentation de la latitude, car la saison de croissance diminue également avec celle-ci à cause de la diminution de la température. L'âge à la maturité augmente avec la latitude, passant de 5 ans pour les régions les plus au sud à plus de 20 ans pour les régions les plus au nord chez les femelles. Chez les mâles, il passe de 5 ans au sud à 13 ans au nord. Les mâles atteignent la même longueur peu importe la latitude, tandis que la longueur des femelles augmente avec la latitude. Les femelles sont en moyenne 2 fois plus longues que les mâles lorsqu'elles

sont matures. Les individus les plus grands de l'espèce sont les femelles du fleuve Saint-Laurent et des Grands Lacs, qui grandissent lentement et durant très longtemps. Le taux de croissance varie également selon l'habitat des anguilles. En effet, les anguilles d'eau salée ou d'eau saumâtre ont un taux de croissance plus grand que des anguilles vivant dans les habitats d'eau douce à proximité et qui sont donc soumis à des températures semblables (Cairns et al., 2009). La répartition des sexes est également fort différente selon les sites. Certains sites ne comprennent que des mâles ou des femelles, alors que d'autres sites comprennent autant de mâles que femelles (Oliveira et al., 2001). De leur côté, les facteurs déterminant le sexe chez l'anguille ne sont pas clairement identifiés. Différents facteurs environnementaux ont été proposés, comme la densité de population, la température de l'eau, le taux de croissance initial des individus, l'habitat, le taux de salinité ou la latitude. Les études se contredisent toutefois entre elles à ce propos (Davey et Jellyman, 2005).

Sélection spatialement variable

Puisque que toutes les anguilles se reproduisent dans la mer des Sargasses des individus ayant passé leur vie adulte à des milliers de kilomètres se reproduisent ensemble. Ce mode de reproduction empêche la différenciation entre les individus ayant vécu dans différentes régions de l'aire de répartition et fait en sorte que tous les individus de l'espèce appartiennent à une seule et même population. L'anguille d'Amérique est donc une espèce panmictique, même si elle occupe une grande aire de répartition (Côté et al., 2013).

Malgré tout, au cours de leur vie, les individus vont être confrontés à des conditions environnementales assez différentes selon l'endroit où les courants marins vont les transporter. En effet, les Antilles et les régions subarctiques offrent des conditions climatiques assez différentes que ce soit au niveau de la température, de l'ensoleillement ou de la présence de glace. La distance à franchir pour atteindre la côte à partir de la mer de Sargasses varie également beaucoup. Elle varie de 1500 kilomètres pour atteindre la côte de la Géorgie à 5000 kilomètres pour atteindre des régions plus au nord comme la Gaspésie (Jessop, 2010). Ces distances différentes font en sorte que les individus atteignent la côte à un âge différent selon la latitude. En Floride, les civelles arrivent sur la côte en janvier, alors qu'à Terre-Neuve, elles n'arrivent qu'en juillet (Gagnaire et al., 2012). Il y a également des anguilles qui vivent en eau douce, en eau saumâtre et en eau salée. Ces différents milieux confrontent les individus à des conditions distinctes dans lesquelles certains allèles permettront à leurs porteurs de mieux performer,

car le phénotype qui en résulte leur procure un avantage dans ce milieu. Cette sélection différentielle fait en sorte que les fréquences alléliques de certains marqueurs favorables dans une région seront plus grandes. Cette sélection sera annulée à chaque nouvelle génération, car le mode de dispersion passif des larves fait en sorte que les individus ont de bonnes chances d'être confrontés à des conditions différentes de celles subies par leurs parents. La sélection opère donc indépendamment à chacune des générations, car un grand nombre d'individus ne seront pas adaptés aux conditions du site où ils se retrouveront et ne survivront donc pas assez longtemps pour aller se reproduire. Les anguilles d'Amérique sont donc probablement l'objet de sélection spatialement variable en panmixie (Pujolar et al., 2014).

La sélection spatialement variable est étudiée chez l'anguille d'Amérique depuis de nombreuses années. En effet, des études ont été réalisées dans les années 70 avec des marqueurs allozymes. Trois marqueurs allozymes avec une fréquence allélique qui était corrélée avec la latitude à laquelle les individus avaient été capturés ont été identifiés comme étant potentiellement sous sélection spatialement variable avec la latitude (Williams et al., 1973). Une autre étude a été effectuée sur une centaine de marqueurs situés dans des régions codantes du génome. Ces marqueurs avaient été préalablement identifiés comme étant les plus différenciés entre des individus en provenance de Floride et de Gaspésie. La fréquence allélique de 13 de ces marqueurs était corrélée avec la température moyenne de l'eau des sites où les anguilles avaient été capturées. Ces marqueurs sont donc potentiellement sous sélection avec la température de l'eau (Gagnaire et al., 2012). Une étude semblable a également été effectuée chez l'anguille européenne. Cette fois-ci, des technologies de séquençage de nouvelle génération ont été employées, ce qui a permis d'utiliser un nombre beaucoup plus important de marqueurs que dans l'étude précédente. Plus de 50 000 marqueurs ont été analysés pour trouver des traces de sélection spatialement variable. Parmi eux, 754 marqueurs ont été identifiés comme étant sous sélection spatialement variable. La fréquence allélique des marqueurs était corrélée avec la température de l'eau des sites d'échantillonnage, la latitude des sites ou leur longitude (Pujolar et al., 2014). Une étude portant l'effet sélectif de la salinité chez l'anguille d'Amérique a permis d'identifier 331 marqueurs parmi les 15 500 analysés, dont la fréquence allélique était lié à la salinité du site où les anguilles avaient été capturées. Plus de 55% de ces marqueurs était presque fixés dans un des 2 milieux malgré la panmixie (Pavey et al., 2015).

Déclin des populations d'anguilles

Au cours des dernières décennies, un déclin important des populations a été observé chez plusieurs espèces d'anguilles, comme l'anguille européenne (*Anguilla anguilla*), l'anguille japonaise (*Anguilla japonica*) et l'anguille d'Amérique (Hanel et al., 2014). Chez l'anguille d'Amérique, la diminution la mieux documentée a été observée dans le fleuve St-Laurent et dans les Grands Lacs. Un déclin de plus de 98% du nombre de juvéniles en montaison a été observée entre les années 1980 et 2009 au niveau du barrage Moses-Saunders sur le fleuve Saint-Laurent. Ce déclin très important des effectifs a mené à la classification de l'anguille d'Amérique comme espèce menacée au Canada. La pêche commerciale et sportive a été interdite en Ontario et a diminué de façon importante au Québec (COSEPAQ, 2012). De nombreuses causes ont été proposées pour expliquer ce déclin, comme la destruction des habitats potentiels, la surpêche, la présence de barrages nuisant à la migration ou la pollution de l'eau (Castonguay et al., 1994). Des variations climatiques au niveau océanique, comme l'oscillation nord-atlantique ont également été proposées pour expliquer le déclin du nombre d'individus (Côté et al., 2013). Les causes du déclin ne sont toutefois pas encore très bien comprises.

Objectif du projet

L'objectif de ce projet est d'utiliser les techniques de séquençage de nouvelle génération pour étudier la sélection spatialement variable chez l'anguille d'Amérique, afin d'étendre l'étude de Gagnaire et al. (2012) à une plus grande partie du génome pour mieux documenter le phénomène de sélection spatialement variable chez cette espèce. Étudier une plus grande du génome permet de voir à quelle échelle la sélection spatialement variable s'effectue. Pour ce faire, les mêmes variables environnementales, la latitude, la longitude et la température de l'eau ont été employées pour détecter si des marqueurs étaient sous sélection avec celles-ci. De plus, l'utilisation d'individus capturés au même endroit lors de 2 années successives a permis de quantifier l'effet temporel de la sélection, ce qui a permis de vérifier si le phénomène de sélection est constant au fil des années ou s'il varie en fonction des différentes conditions environnementales auxquelles les anguilles sont confrontés au cours de leur migration en direction de la côte. Des analyses de redondance et d'associations environnementales ont été employées pour identifier des marqueurs potentiellement sous sélection spatialement variable avec les différentes variables environnementales. L'assemblage et l'annotation du génome de l'anguille d'Amérique ont permis de positionner les différents marqueurs analysés par rapport aux gènes. Les

gènes situés à proximité des marqueurs potentiellement sous sélection ont donc pu être identifiés ainsi que les mécanismes potentiellement affectés par la sélection spatialement variable chez l'anguille d'Amérique. Une approche d'additivité a également été utilisée afin de vérifier la présence de sélection polygénique exercée par les variables environnementales chez cette espèce.

Chapter 1: RAD-seq reveals patterns of additive polygenic variation caused by spatially-varying selection in the American Eel (*Anguilla rostrata*)

Chapitre 1: Le RAD-seq révèle la présence de variations polygéniques additives causées par la sélection spatialement variable chez l'anguille d'Amérique (*Anguilla rostrata*)

Résumé

L'anguille d'Amérique possède un cycle de vie exceptionnel, qui inclut une reproduction panmictique à l'échelle de l'espèce, une dispersion aléatoire des individus et de la sélection dans un habitat hautement hétérogène qui s'étend de régions subtropicales à des régions subarctiques. Les conséquences génétiques des pressions de sélection spatialement variable sont étudiées depuis des décennies chez cette espèce, révélant de subtils clines dans les fréquences alléliques de quelques loci qui contrastent avec une panmixie complète dans le reste du génome. Puisque la reproduction homogénéise les fréquences alléliques à chaque génération, la taille de l'échantillon et la proportion du génome qui est couverte sont critiques dans ce contexte pour atteindre une puissance statistique suffisante pour détecter des loci sous sélection. Nous avons donc séquencé 12 098 marqueurs SNP chez 710 individus en provenance de 12 sites afin de réévaluer jusqu'à quel point la sélection locale affecte la distribution spatiale de la diversité génétique chez l'anguille d'Amérique. Nous avons employé des analyses d'association environnementale afin d'identifier des marqueurs sous sélection spatialement variable et nous avons découvert qu'environ 1% du génome semble affecté par la sélection. Nous avons ensuite évalué à quel point les marqueurs candidats reflètent collectivement les variations environnementales en utilisant des scores polygéniques additifs. Nos résultats sont compatibles avec la présence de sélection polygénique agissant sur les individus les plus maladaptés aux différents habitats occupés par les anguilles dans leur aire de distribution. Les gènes associés aux marqueurs sous sélection étaient significativement enrichis pour le mécanisme de réponse à l'insuline, ce qui indique que les compromis nécessaires afin d'occuper un habitat aussi varié impliquent une bonne régulation du métabolisme. Cette étude montre l'efficacité des approches utilisant les scores polygéniques additifs pour la détection des effets sélectifs dans un environnement complexe.

Abstract

The American Eel (*Anguilla rostrata*) has an exceptional life cycle characterised by panmictic reproduction at the species scale, random dispersal and selection in a highly heterogeneous habitat extending from subtropical to subarctic latitudes. The genetic consequences of spatially-varying selection in this species have been investigated for decades, revealing subtle clines in allele frequency at a few loci that contrast with complete panmixia on the vast majority of the genome. Because reproduction homogenizes allele frequencies every generation, sampling size and genomic coverage are critical to reach sufficient power to detect selected loci in this context. Here, we used a total of 710 individuals from 12 sites and 12 098 high-quality SNPs to re-evaluate the extent to which local selection affects the spatial distribution of genetic diversity in the American Eel. We used environmental association methods to identify markers under spatially-varying selection, which indicated that selection affects about 1,5% of the genome. We then evaluated the extent to which candidate markers collectively reflect environmental variation using additive polygenic scores. We found significant correlations between polygenic scores and latitude, longitude and temperature which are consistent with polygenic selection acting against maladapted genotypes in different habitats occupied by eels throughout their range of distribution. Gene functions associated with outlier markers were significantly enriched for the insulin signaling pathway, indicating that the trade-offs inherent to occupying such a large distribution range involve the regulation of metabolism. Overall, this study highlights the potential of the additive polygenic scores approach in detecting selective effects in a complex environment.

Introduction

Panmictic species challenge the classic view of local adaptation, whereby individuals from local populations are more fit to their own habitat compared to migrants from different populations (Kawecki and Ebert, 2004). Indeed, local adaptation cannot develop in panmictic species since random mating and random dispersal repeatedly bring individuals carrying “maladaptive” alleles every generation over the entire species range (Yeaman and Otto, 2011). Locally deleterious alleles should be eventually removed by selection, but this elimination can take several hundred generations, especially if selected alleles have variable effects across different environments (Gagnaire et al., 2012). In certain conditions, such as when species occupy different niches where different alleles are favored, spatially-varying selection could maintain polymorphism at selected loci (Levene, 1953). Most of the time, however, locally selected loci are only transiently polymorphic and eventually attain fixation if selection is not well balanced (Yeaman, 2015).

The American Eel (*Anguilla rostrata*) is a catadromous panmictic species (Avisé et al., 1986; Wirth and Bernatchez, 2003; Côté et al., 2013; Pavey et al. 2016) occupying a large geographic range that extends from the Caribbean to Greenland all along the North American Atlantic coast. Environmental conditions dramatically differ between the southern tropical and the northern subarctic part of the species range, translating into a strong latitudinal gradient of selection (Gagnaire et al., 2012; Williams et al., 1973). The species also occupies a wide variety of freshwater, brackish and saltwater habitats, as well as clean versus polluted sites, which represent additional selective challenges (Pavey et al., 2015; Laporte et al. 2016). Although eels spend most of their life in continental or coastal waters, maturing adult eels eventually return to the Sargasso Sea to mate and then die (Béguet-Pon et al., 2015). Young planktonic larvae (leptocephali) that hatch in the Sargasso Sea are then advected to the American coast by passive drift in the Antilles Current and the Gulf Stream (Bonhommeau et al., 2010; Vélez-Espino et

al., 2010). As they reach the continental shelf, the leptocephali undergo metamorphosis into glass eels that use a selective tidal stream transport to reach the coast (McCleave and Kleckner, 1982). Recruiting glass eels can either stay in saltwater along the coast, settle in estuarine brackish water or continue their inland migration upstream into freshwater (Cairns et al., 2009). In this complex life cycle, the panmictic allelic combinations that are produced every generation are randomly dispersed across the whole habitat, making it impossible for local adaptation to appear even if eels have to face strong environmental selective mortality rates due to varying environmental constraints across their range.

The genetic consequences of spatially-varying selection have been studied in the American Eel for decades, leading to the finding of temporally stable latitudinal clines at allozyme loci Williams et al., 1973; Koehn and Williams, 1978). The question has been revisited more recently with population genomic approaches, with the aim to screen for gene-associated SNPs exhibiting genetic differentiation between the extreme parts of the species range despite panmixia (Gagnaire et al., 2012). Candidate outlier SNPs scrutinized for association with environmental variables in a large cohort of individuals sampled throughout the species range allowed validating the influence of spatially-varying selection at 13 coding gene SNPs, most of which were partly associated with river mouth temperature when eel larvae approach the coast. Most of these SNPs were found to be transient polymorphisms contributing to genetic-by-environment interactions during progress toward allelic fixation (Gagnaire et al., 2012).

A less resolved issue concerns the proportion of the genome that is influenced by spatially-varying selection. In a genome scan study based on 50 000 RAD-seq markers, Pujolar et al. (2014) identified 710 loci probably influenced by spatially-varying selection in the European Eel (*Anguilla anguilla*), the American Eel sister species. Among these, only 74 loci displayed significant associations with environment variables, representing a small fraction (~0.15%) of the SNPs considered. Higher

proportions (0.8-1%) were found in the American Eel in two RAD-Seq studies that documented differential selection between freshwater and saltwater habitats (Pavey et al., 2015) and between clean versus polluted sites (Laporte et al., 2016). Both of these studies, however, applied a polygenic statistical framework (multi-locus random forest algorithm approach) that differs substantially from genome-scan statistical methods searching for single-locus selection based on allele frequencies. Previous studies comparing the efficiency of polygenic multi-locus analysis vs. single locus genome scan to detect selection in other fish species revealed that the former approach surpassed the latter (e.g. Bourret et al., 2014). More generally, population genomic studies face statistical challenge to detect individual outliers in panmictic species for two reasons (Gagnaire and Gaggiotti, 2016). First, moderate selection on single gene traits can only generate small allele frequency changes within a single generation. Second, a comparable intensity of selection acting on polygenic traits results in even smaller allele frequency changes, because the selective pressure is spread over the many loci contributing to the traits under selection (Pritchard and Di Rienzo, 2010). Because sample size determines the power of such tests to detect a selected locus with a given effect size, increasing sample size while controlling for confounding effects (e.g. temporal variation) remains the most effective way to detect loci influenced by selection.

The goal of this study was to re-evaluate the extent to which spatially-varying selection affects the spatial distribution of genetic diversity in the American Eel. We implemented a reference-guided RAD-Seq genome scan based on the published sequence and annotations of the American Eel genome (Pavey et al., 2016). In order to detect small allele frequency changes generated by selection, we more than doubled the sample size and localities used in previous RAD-Seq studies in Atlantic Eel, and controlled for temporal effects by focussing on glass eel cohorts from two consecutive years. We used environmental association methods and confirm the results of these analyses by simulations to reduce the false positive detection of markers influenced by selection, building on the principle that genetic-

by-environment associations are not confounded by neutral geographic differentiation given the panmictic nature of American Eel. We then used additive polygenic scores (Gagnaire and Gaggiotti, 2016) to evaluate how the cumulative signal displayed by outlier loci correlates with environmental factors, both at the individual and the sampling site levels. The genes associated with outlier markers and their functions were then analyzed to test for functional enrichments that could provide indications on the currently unknown phenotypes that are undergoing spatially-varying selection.

Materials and methods

Sampling, library preparation and genotyping

A total of 710 glass eels studied by Gagnaire (2012) were used for this study. Those individuals were sampled in 2008 at the mouth of 12 tributaries, soon after recruitment to estuaries. Among these, three geographically distant sites were sampled again in 2009 for temporal replication and a 13th site was sampled in Nova Scotia in 2009 only (Figure 1). Sampling sites cover the core of the American Eel's distribution range spanning from Newfoundland to Florida. Details on sampled individuals are provided in Table 1. Except for one site from Georgia (n=24), the mean number of samples was 45 by site.

Whole genomic DNA was extracted from tissues using a salt extraction protocol (Aljanabi and Martinez, 1997) with RNase A treatment. DNA quality was checked on agarose gel and DNA concentration was quantified using PicoGreen (Invitrogen). DNA libraries were prepared with the RAD-sequencing protocol used by Benestan et al. (2015) using the restriction enzyme, *EcoRI* which recognized a 6pb cutting site. The libraries were prepared in pools of 12 individuals. For the sequencing, four pools of 12 individuals with different sets of barcodes were put in the same

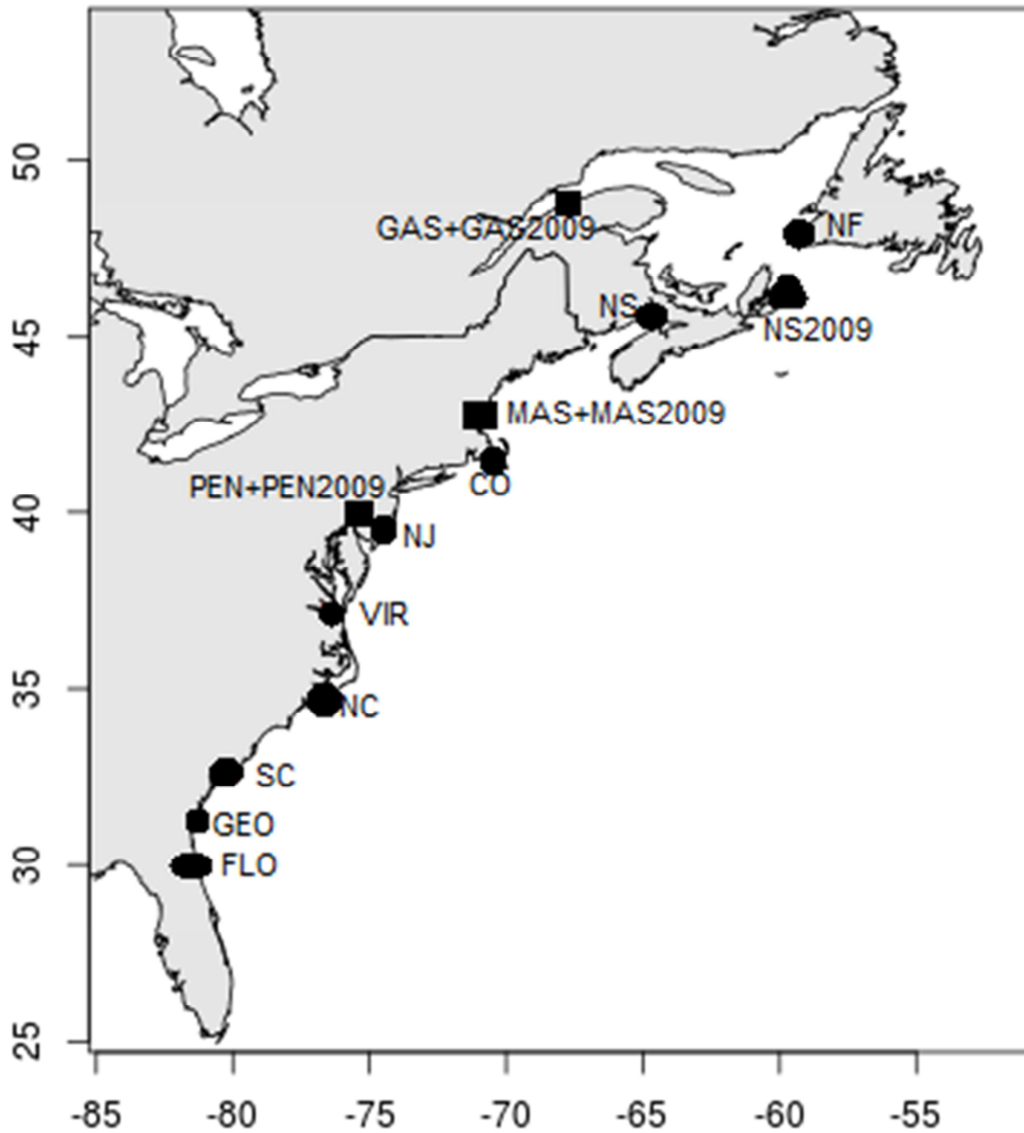


Fig. 1 Location map of the 13 sampling sites where American glass eels were collected for this study. Labels correspond to the site codes provided in table 1. The sites identified by squares were sampled in 2008 and 2009, sites identified by circles were sampled in 2008 and the site identified by a triangle was sampled in 2009.

sequencing lane for a total of 17 sequencing lanes. 82 individuals with a low sequencing success were sequenced twice to increase their coverage. Single-end 100 bp sequencing was performed on an Illumina HiSeq 2000 at the Genome Quebec Innovation Center (McGill University, Montreal, Canada).

Following sequencing, adapters were removed from the reads using Cutadapt (Martin, 2011). The reads were then demultiplexed and trimmed to 80 bp using the *process_radtags* module of Stacks version 1.32 (Catchen et al., 2011; Catchen et al., 2013). Reads were then mapped on the American Eel genome (Pavey et al., 2016) using BWA aln (Li and Durbin, 2009). The American Eel genome has a size of 1.41 Gb and is assembled on 79 000 scaffolds (Pavey et al., 2016). SNPs were called using SAMtools mpileup and BCFtools view modules (Li and al., 2009). Only markers present in at least 80% of the individuals and with a global minor allele frequency of at least 0.02 were retained. A maf of 0.02 was applied as a compromise between minimising false SNPs caused by sequencing errors and retaining rare alleles. The remaining markers were then filtered based on their sequencing depth to keep only SNPs with an overall coverage of at least 1500 reads for downstream analyses. The SNPs allele frequencies were then estimated using estpEM (Gompert et al., 2014) which uses the EM algorithm (Li, 2011) to take into account genotype likelihoods of the different individuals present in the population to deal with low coverage sequencing data. The program was run independently for each site and temporal replicate.

Explanatory variables

To identify markers influenced by spatially-varying selection, potential explanatory variables were considered, including the latitude and longitude of the sampling sites, the river mouth temperature and year at recruitment (Table 1). Temperature data were obtained from a National Oceanic and Atmospheric Administration database (SST14NA)

(https://www.class.ngdc.noaa.gov/saa/products/search?datatype_family=SST14NA). For the 2008 individuals, the river mouth temperature for each site was determined by the average sea-surface temperature near the river mouth during the 10 days prior to the sampling day during recruitment. That time period was chosen because it corresponded to the most highly correlated temperature with loci under selection in Gagnaire et al. (2012). For the 2009 individuals, the exact sampling date was not available, and therefore, the temperature used for those samples was the average sea-surface temperature of the river mouth during the sampling month.

Table 1 Names and codes of sampling locations, number of individuals per site, latitude, longitude, river mouth temperature and sampling year. Temperatures for 2008 are the average sea-surface temperature near the river mouth during the 10 days prior to the sampling day during recruitment. Temperatures for 2009 are the average sea-surface temperature of the river mouth during the sampling month.

Sites	Site code	Number of individuals	Latitude	Longitude	Temperature	Year
Connecticut	CO	48	41,41	-70,55	10,44	2008
Florida	FLO	40	30	-81,19	18,09	2008
Gaspésie	GAS	47	48,78	-67,7	14,21	2008
Gaspésie	GAS2009	42	48,78	-67,7	10,57	2009
Georgia	GEO	24	31,18	-81,28	15,46	2008
Massachusetts	MAS	48	42,69	-70,79	7,04	2008
Massachusetts	MAS2009	48	42,69	-70,79	5,65	2009
North Carolina	NC	48	34,46	-76,48	10,44	2008
Newfoundland	NF	47	47,85	-59,26	15,86	2008
New Jersey	NJ	46	39,52	-74,5	7,79	2008
Nova Scotia	NS	39	45,54	-64,7	3,41	2008
Nova Scotia	NS2009	48	46,04	-59,96	9,55	2009
Pennsylvania	PEN	45	39,89	-75,26	9,5	2008
Pennsylvania	PEN2009	48	39,86	-75,32	13,57	2009
South Carolina	SC	46	32,55	-80	17,19	2008
Virginia	VIR	46	37,21	-76,49	10,01	2008

Redundancy analysis to detect outlier loci

A redundancy analysis (RDA) was performed on all the markers retained after the filtering steps using the R package 'VEGAN' v. 2.3-2 (Oksanen et al., 2017). The RDA was used to estimate the proportion of genetic variance in the entire dataset that is explained by the environmental factors. Then, the significance of the RDA was tested with an ANOVA. . Both, the joint effects of the four constraining

variables as well as their marginal effects were tested using ANOVAs performed with 1000 permutations. Also, the effect of each variable was partitioned using conditioned RDA, in order to evaluate the percentage of genetic variance explained by each variable separately, after removing the variance due to partial correlation with other variables. Here again, an analysis of variance (ANOVA) was performed to verify the significance level of each of the four partial RDAs. Finally, the explicative importance of the different variables was represented as vectors in biplot graphs.

In order to identify subsets of loci showing extreme contributions (RDA outliers) to the main directions of constrained variation, we selected the markers with the 0.5% most positive and negative scores for the first and the second RDA axes. We chose a threshold of 0.5% as a compromise between not being too conservative nor too permissive. To minimize the effect of linkage disequilibrium, only one marker per RAD locus pair was retained. When more than one outlier was found on the same locus pair among the potential outliers, the marker with the highest score was retained. A locus pair is two reads that are separated by the same restriction site. A new redundancy analysis was then performed with the subset of markers identified as outliers in order to evaluate the extent to which genetic variation at outlier loci was explained by each explanatory variable.

Detection of outliers using Bayenv2

In order to detect outlier loci possibly influenced by spatially-varying selection, we used the program Bayenv2 (Coop et al., 2010; Gunther and Coop, 2013) which searches for exceptional correlations between explanatory variables and allele frequencies at individual locus. Although this method specifically takes into account the underlying population structure, it can also deal with the particular case of a panmixia. We used the markers retained after the different filtering steps to calculate the covariance matrix using all samples with 100,000 iterations. Then, 12 runs of 100,000 iterations each

were performed using the covariance matrix to identify the most strongly associated markers with each of the four explanatory variables. These 12 replicates aimed at ensuring high repeatability, in order to reduce the number of false positive detection due to the variance in the level of support observed among runs for some markers (Blair et al., 2014). Markers were considered to be outlier if their average Bayes factor across the 12 runs was higher than 3, and above 3 for at least 6 of the 12 runs. We used a threshold higher than 3 following the recommendation of Jeffreys (1961) for the significance of Bayes factor. Then, a redundancy analysis was performed with the subset of outlier identified with Bayenv2 in order to evaluate the proportion of their total variance explained by the constraining variables taken together and separately

The rate of false positives detected by Bayenv2 was determined using simulated datasets. A total of 100 datasets were simulated by randomly assigning the individuals to another site, to simulate a random dispersal without the effect of selection. A covariance matrix was calculated for each of these datasets. Then only one run of Bayenv2 was performed for each dataset given our goal of estimating the number of false positives. We then compared the number of outlier detected in each simulation for the different variables to the average number of outlier detected by the 12 runs made on the real dataset for the different variables.

Detecting variables with the strongest contribution to outlier variation

Conditioned redundancy analyses (partial RDAs) were performed on a combined dataset comprising all the outliers detected by the RDA and by the Bayenv2 approaches. These partial RDAs aimed at identifying which explanatory variable contributed the most to each outlier SNP variation. Marker scores were ranked for each partial RDA performed separately with each variable, and the variable for which a given marker had the highest rank was considered to be the most influencing variable.

Using additive scores to visualize patterns of polygenic variation

The cumulative signal of multiple outlier SNPs markers can be assessed using polygenic scores to evaluate how their joint effect mirrors environmental variation (Gagnaire and Gaggiotti, 2016). When the selective effects of individual allele are unknown, polygenic scores are obtained for each individual by summing over loci the number of alleles that are inferred to be favoured in a given environment (Hancock et al., 2011; Aarnegard et al., 2014). This relates directly to the concept of additive genetic variance within a theoretical quantitative genetics framework (Roff, 1997). Here, alleles that were associated with increasing values of temperature, latitude or longitude were identified based on the sign of their correlation with a given variable for each outlier. Polygenic scores were then computed at the individual eel level. Individual polygenic scores for each eel were obtained by summing over loci the number of favoured alleles in a given environmental condition. Correlation between individual and latitude, longitude and temperature were then independently assessed to evaluate how well the cumulated signal of locally putative adaptive alleles varies with each explanatory variable. For each variable, three models (a linear, a quadratic and a logistic) were tested to identify which one fits best the relationship between the polygenic scores and the environmental variable. For each variable, we selected the model with the lowest Akaike information criterion (AIC) value. To evaluate the extent to which extreme polygenic scores found in each locality were correlated with the environmental variables, we assessed the correlations between the 10th and 90th centile of the polygenic score of each site and the environmental variables values. Correlations were tested using a linear and a quadratic model.

Annotation of outlier loci and GO term enrichment analysis

Using the mapping of outlier markers onto the American Eel genome, functional annotations were used to identify the closest gene sequences for each outlier. Only the closest genes located within 100 kb from each outlier were considered. A gene ontology (GO) term enrichment analysis was then performed on the list of candidate genes potentially associated with outlier loci using DAVID web-server version 6.7 (<https://david.ncifcrf.gov/tools.jsp>) (Huang et al., 2007). Zebrafish (*Danio rerio*) annotations were used as a reference genome for the GO term enrichment analysis, since this species has the most complete list of annotated genes with known functions among teleost fishes.

Results

Genome-wide polymorphism dataset

Sequencing of RAD libraries resulted in a total of $3.2 \cdot 10^9$ reads. After the alignment, 7.5 millions of SNPs were identified by SAMtools. Following SNP calling, 656 244 SNPs that were present in at least 80% of the individuals were kept, among which 12,098 were retained after applying minor allele frequency and coverage depth thresholds. The 12,098 SNPs, which were used to detect candidate markers potentially affected by spatially-varying selection in downstream analyses, were present on 837 scaffolds.

Redundancy analysis to detect loci associated with explanatory variables

The redundancy analysis (RDA) conducted using all 12,098 retained markers did not reveal a significant amount of genetic variation associated with the four constraining variables ($F=1.04$,

$p=0.188$). Although the RDA model including the four explanatory variables explained 27.5% of the genetic variance observed among sites, this amount was lower than the cumulated variance explained by the first four unconstrained PC axes. The four variables explained similar proportions of genetic variance varying between 6.6% and 6.8%, but their marginal effects were not significant. Despite this lack of significance, the projection of sampling locations on the first two RDA axes closely reflected the geography and thermal characteristics of sampling sites along the directions indicated by the geographic and temperature vectors (figure 2A). The 0.5% of the markers with the most positive and negative scores for the first two axes that were considered as RDA outliers represented 197 markers after retaining only one marker per RAD locus pair (supplementary table 1).

The RDA performed using the subset of 197 outliers detected by the global RDA explained 35 % ($F=1.4759$, $p=0.018$) of the genetic variance present at these markers. The first two RDA axes respectively explained 15.4 and 11.6 % of genetic variance among sites (figure 2B). The temperature vector was in an opposite direction comparing to the vectors for latitude and longitude and the vector for the year of capture was orthogonal to the three other constraining variables.

Detection of loci associated with explanatory variables using Bayenv2

A total of 90 SNPs detected as candidate outliers with Bayenv2 were retained for further downstream analyses (Supplementary table 2). Among these, 15 outlier markers were associated with the latitude of the sampling site, 16 to the longitude (among which 4 were also latitude-related outliers), 16 markers to local temperature and 47 markers were related to the year of recruitment of the glass eels. Among the 90 outliers detected with Bayenv2, 19 were among the 197 RDA outliers, and 9 additional ones were located within 100bp of one of these RDA outliers. Thus, 31% percent of the Bayenv2 outliers were also detected by the RDA outlier detection approach. The RDA performed on this subset of 90

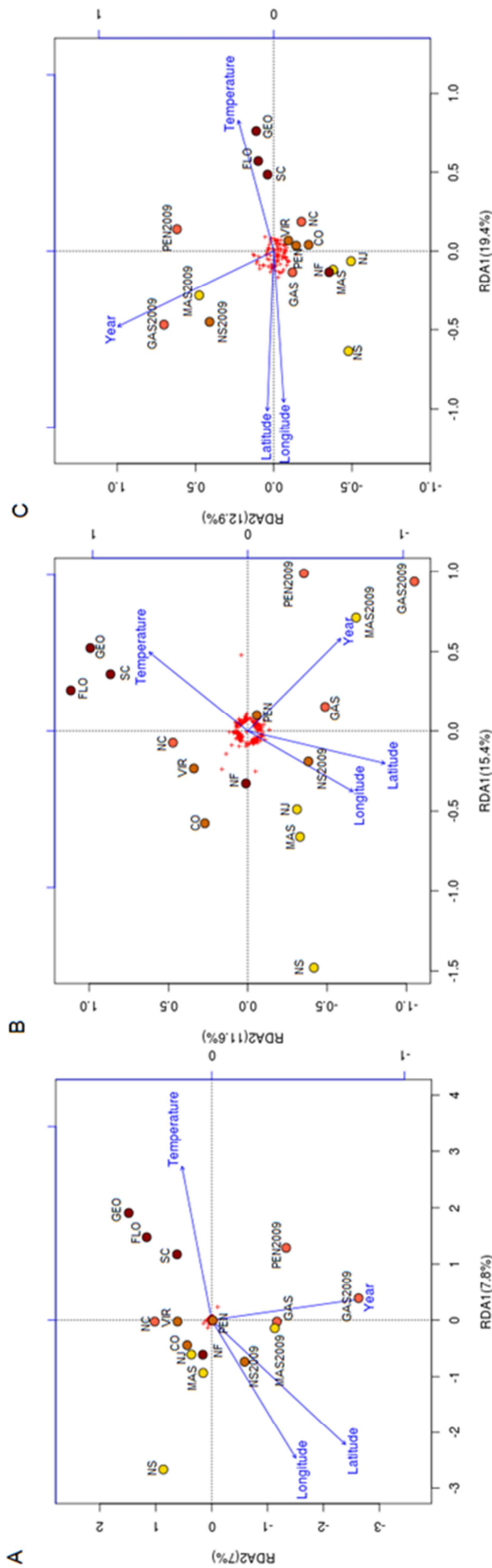


Fig. 2 (A) Redundancy analysis (RDA) performed with 12 098 filtered SNPs, using latitude, longitude, temperature and year as constraining variables. (B) Redundancy analysis performed using the 197 SNPs detected to be potentially affected by selection with the RDA outlier detection approach (C) Redundancy analysis performed using the 90 markers significantly associated with environmental variables in Bayenv2 analysis. Only the first 2 axes are shown for each RDA. Sites are identified by their code (Table 1) and by colored points reflecting water temperature value, from yellow for colder sites to dark red for warmer sites. The direction of main variation for each of the 4 constraining variables (Year, Latitude, Longitude, Temperature) is indicated by a blue vectors. SNPs contributions appear as red '+' symbols. The proportion of total variance explained by each axis is indicated in percent.

Bayenv2 outlier markers explained 45.6% ($F=2.3086$, $p<0.001$) of the genetic variance present at these markers. The first 2 axes respectively explained 19.4 and 12.9% of genetic variance among sites (figure 2C). The temperature was mostly related the first axis and in opposite directions to both longitude and latitude whereas the year of capture was mostly related to the second axis.

The simulations revealed that a significantly higher number of outliers were detected by Bayenv2 with the real dataset compared to the simulated ones for the 4 variables (Figure 3A-D). Thus, except for the temperature for which one simulation out of 100 returned a number of outliers higher than the real data set, no run detected a higher number of outlier than the average of the 12 runs made with the real dataset. These results therefore indicate that false positives are unlikely to have driven the observed patterns of association between allelic and environmental variation observed at some loci, which is therefore suggestive of selection driven by local environmental condition. Note that the number of outliers detected in each run was higher than the number of outliers used for further analyses. This is because, not always the same outliers were detected in each run and in order to remain conservative in downstream analyses, we retained only the markers that were significant in 6 runs or more out of 12.

Using conditioned RDA to control for confounding effects

Partial RDAs were performed on a dataset combining all 259 outliers that were detected by one or both approaches. The year of recruitment explained the most important proportion of the total variance (10.3%, $p=0.025$), followed by the temperature (8%, $p=0.12$) as well as the latitude and the longitude (6.1%, $p=0.4$). The variable with the highest influence on each marker was identified. The year of recruitment was related to 76 outliers, followed by the temperature with 70 markers. The latitude was

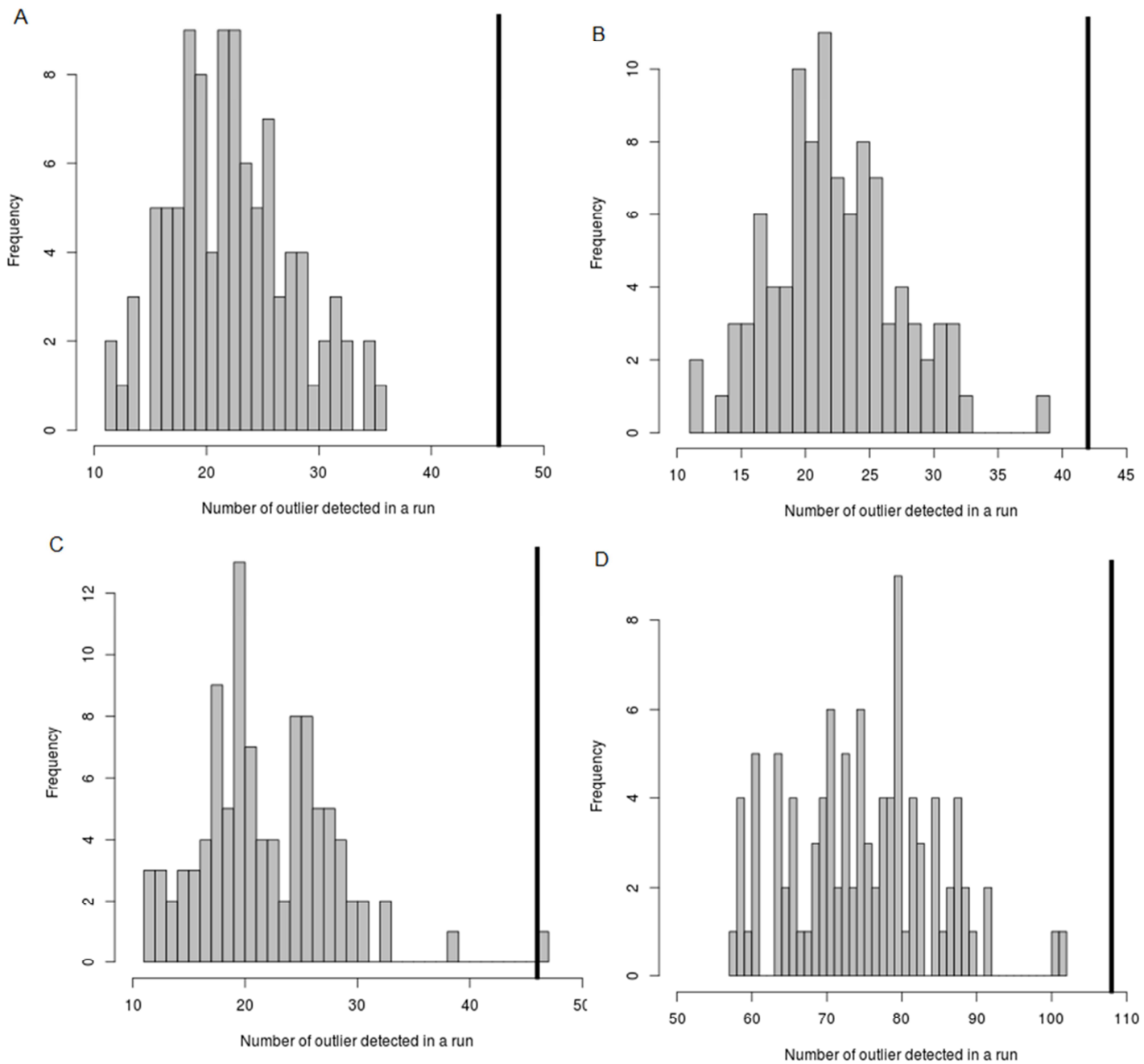


Fig. 3 Histograms of the number of outliers detected in 100 simulations for the different environmental variables; latitude (A), longitude (B), temperature (C) and year (D). The x axis represents the number of outliers detected in a run. The y axis represents the number of simulations that detected this number of outlier. The vertical black bar represents the average number of outlier detected in the runs made with the real dataset. Note that the number of outliers detected in each run was higher than the number of outliers used for further analyses as reported in the main text. This is because not always the same outliers were detected in each run and in order to remain conservative in downstream analyses, we retained only the markers that were significant in 6 runs or more out of 12.

related to 64 markers and the longitude to 49 markers. In some cases, the variable that was detected to be related to an outlier by Bayenv was not the same as for the partial RDA. The variable related to each outlier marker is indicated in the supplementary table 3.

Using additive scores to evaluate polygenic variation patterns

The correlations assessed between each variable and the corresponding additive scores calculated using outlier loci (Figure 4A-C) and were all significant. For the three variables, the quadratic model had the lowest AIC, and it was followed by linear model and then the logistic model. Longitude was the variable with the highest correlation coefficient ($R^2=0.154$, $p \leq 10E-15$), followed by the latitude ($R^2=0.152$, $p \leq 10E-15$) and the temperature ($R^2=0.142$, $p \leq 10E-15$). Correlations between the 10th and 90 centile of polygenic scores in a site and variables were significant for all three variables ($R^2=0.65-0.87$) (Table 2). Correlations for the latitude and the temperature were best explained by a linear model. For longitude, the quadratic model was the best one.

Table 2 Correlation coefficients and p-values of the relationship between the 10th and 90th centile of the polygenic scores of each of the sampling sites and the environmental variables. The correlation coefficient and p-values were calculated following a quadratic and a linear model.

Variable	Linear model		Quadratic model	
	10 th centile	90 th centile	10 th centile	90 th centile
Latitude	$R^2=0.65$ $p=0.00016$	$R^2=0.76$ $p=1.17 \times 10^{-5}$	$R^2=0.71$ $p=0.0003$	$R^2=0.76$ $p=8.5 \times 10^{-5}$
Longitude	$R^2=0.63$ $p=0.0002$	$R^2=0.53$ $p=0.00148$	$R^2=0.86$ $p=3.29 \times 10^{-6}$	$R^2=0.86$ $p=2.66 \times 10^{-6}$
Temperature	$R^2=0.87$ $p=1 \times 10^{-7}$	$R^2=0.65$ $p=0.00017$	$R^2=0.9$ $p=3.43 \times 10^{-7}$	$R^2=0.65$ $p=0.001$

Annotation of outlier loci and GO term enrichment analysis

The 259 outliers detected combining Bayenv2 and RDA analyses were distributed across 204 different scaffolds, 45 of which had more than one outlier. A total of 217 outliers were within 100 kb from at

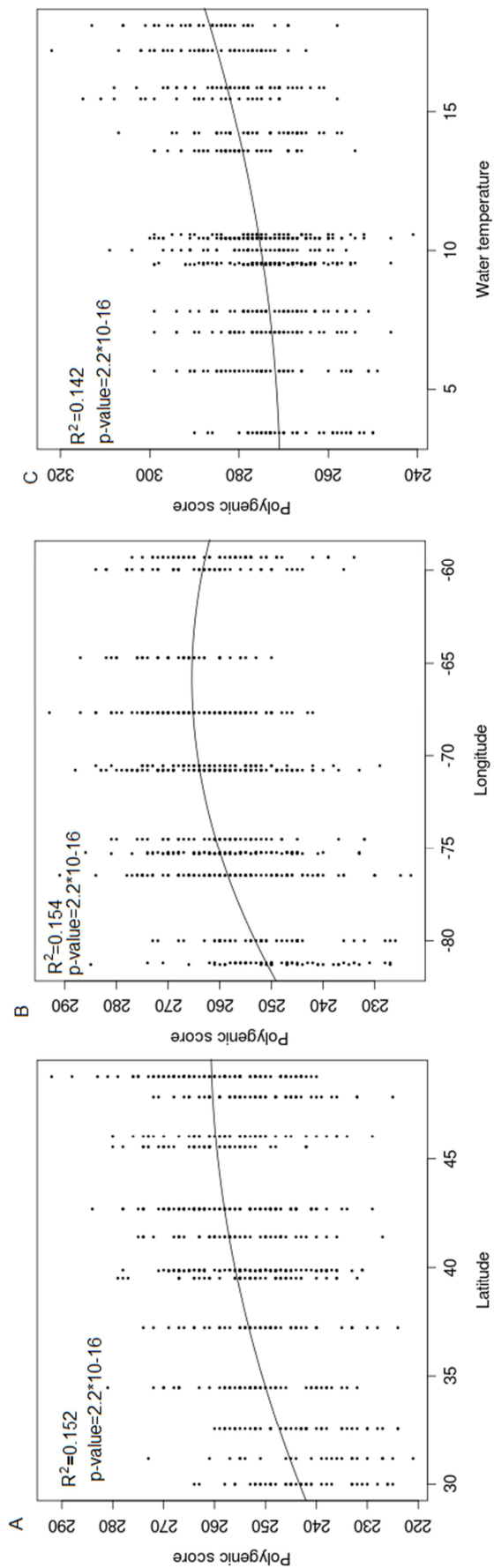


Fig. 4 Correlations between additive individual polygenic scores based on all 259 outlier markers and each of 3 explanatory variables: latitude (A), longitude (B) and temperature (C). Additive polygenic scores were obtained at the individual level by summing the number of locally favourable alleles. Correlation coefficients and p-values of the quadratic correlation model are shown at the top of each panel.

least one nearby gene enabling the identification of 241 genes associated with an outlier (Supplementary table 3). Among them, eight genes had an exonic outlier SNP and 50 genes had an intronic outlier, one gene had both exonic and intronic outliers and 56 genes were within a 10 kb distance from an outlier. For the 28 markers detected both by Bayenv2 and RDA, two markers were located in exons, 8 markers were in introns and 10 were within 10 kb from the nearest gene. The proportion of markers present in introns (20%) and exons (3%) was similar between the outliers and the whole dataset.

No Gene Ontology term was found significantly enriched in the analysis performed with DAVID. However, the KEGG pathway insulin signaling pathway showed a significant enrichment (p-value=0.04). Among the 171 genes implicated in this pathway, five (PRKACAA, GYS1, PTPN1, IKBKB, PHKG1A) were represented in our dataset of genes potentially affected by selection. The gene GYS1 has an outlier into one of its intron. This gene codes for the glycogen synthase muscle isoform which produce glycogen in the muscle (Palm et al., 2013). The gene IKBKB, which has an exonic outlier, codes for a subunit of the protein IKK involved in the NF-kappa-B pathway which regulates the expression of multiple genes (Häcker and Karin, 2006).

The list of outlier-associated genes also contained four genes coding for proteins implicated in the mitochondrial electron transport (NDUFA4, ATP5F1, NDUFA10, COX10) and five genes (SNX14, PHACTR4B, SHANK3, ESRP2 and NFATC3) that were previously detected to be potentially under spatially-varying selection between the glass eel and the yellow eel stages in the European Eel (Pujolar et al., 2015). Among them, two have intronic outliers (ESRP2 and NFATC3).

Discussion

The goal of this study was to implement a genome-scan approach to evaluate the extent to which spatially-varying selection shapes genetic variation across the American Eel genome. Environmental correlation methods combined with simulations were used to reduce the false positive detection rate of outlier loci, controlling for temporal effects to identify polymorphisms influenced by selection within a single generation. We also evaluated the extent to which candidate markers collectively reflect environmental variation using an approach based on additive polygenic scores which has been rarely applied thus far in studies of non-model species. After screening a total of 12,098 stringently filtered RAD SNP markers in 710 glass eels from the entire species range, we identified 183 loci potentially affected by spatially-varying selection after controlling for temporal outliers. This represents approximately 1.5% of the genome, which is higher but still close to previous comparable estimates (Pavey et al., 2015; Laporte et al. 2016; Pujolar et al., 2013). This result is thus consistent with polygenic selection acting on multiple mutations with spatially-varying effects on fitness. Below, we discuss why the list of candidate loci detected here represents only a fraction of the genomic regions affected by spatially-varying selection. To support the “many genes of small effects” hypothesis, we show that subsets of outlier loci individually associated to a similar environmental variable collectively display a highly significant association with environmental variation. Finally, we searched for functional enrichment in biological pathways that may provide cues for the unknown gene networks under polygenic selection. We found an overrepresentation of the insulin signalling pathway, which together with previous findings (Gagnaire et al., 2012), indicate that lipid and saccharide metabolism pathways are among the main biological functions under spatially-varying selection across one the most ecologically contrasted species range in the marine realm.

Detection of individual loci affected by selection

The small changes in allele frequency at individual loci generated by spatially-varying selection in panmixia challenge the power of detecting the molecular footprints of selection, especially when selection acts on polygenic traits (Gagnaire and Gaggiotti, 2016). Environmental association methods (e.g. Coop et al., 2010; Frichot et al., 2013; Guillot et al., 2014) offer interesting approaches to detect polygenic selection, although their relative power and error rate differ depending on the scenario considered (de Villemereuil et al., 2014; Wellenreuther and Hansson, 2016). Here, we were interested in detecting patterns of allele frequencies associated with a selection gradient, a scenario in which Bayenv usually shows a low error rate. Although Bayenv takes into account the covariance in allele frequency across samples to test for locus-specific association with each environmental variable, the absence of neutral population structure to account for in our dataset should at worst make outlier detection more conservative. The panmictic reproduction regime of the American Eel should thus eliminate the risk of detecting false positives due to spatial autocorrelation, a classical pitfall for genetic-environment association methods (Hoban et al., 2016). To complete this approach, we combined it with a constrained ordination method (RDA) that uses linear combinations of environmental variables to describe how environmental variations explain patterns of allele frequencies across samples. Because in that case outliers were identified as the strongest contributors to the constrained axes, the RDA approach seems appropriate to detect markers potentially influenced by multiple explanatory variables simultaneously. On the downside, this approach does not include a statistical framework to address the significance of candidate markers individually, which implies defining a somewhat arbitrary empirical threshold for outlier detection. Despite these differences between the two methods, 31% of the outliers that were detected by Bayenv were also found with the RDA approach, thus providing additional evidence for spatially-varying selection at these loci.

When combining all the 259 outliers detected by the Bayenv and RDA approaches, 29.3% of these markers were related to the year of recruitment which explained 10.3% of the total outlier genetic variance after accounting for confounding effects in a conditioned RDA. The importance of the recruitment year could be explained by the fact that environmental conditions vary among years potentially resulting in selection for different alleles. For example, the river mouth temperature at the time of larval recruitment to the coast was different between the two samplings years for the sites that were sampled both in 2008 and in 2009 at about the same time of the year. The genetic composition of the pool of reproducing adults in the Sargasso Sea may also vary from year to year. For instance, Côté et al. (2012) found yearly variation in the effective number of breeders (N_b) associated to the North Atlantic Oscillation (NAO), which could affect the gene pool composition of glass eels. The NAO has also a marked influence on North Atlantic surface winds which affects the Atlantic thermohaline circulation (Hurrell et al., 2001). A recent meta-analysis examining factors influencing selection in more than 160 species of plants and animals has found that the NAO was one of the factor explaining the biggest part of the temporal variation in natural selection observed in these species (Sipielsky et al., 2017). Finally, the sweepstake reproductive success of adults (Hedgecock and Pudovkin, 2011) may generate temporal variation in the genetic composition of yearly cohorts, as already observed in the European Eel (*A. anguilla*) (Maes et al., 2006; Pujolar et al., 2006). Therefore, the predominance of temporal outliers could be explained by a combination of temporally variable selective regimes acting on stochastically changing gene pools due to environmental forcing and sweepstake reproduction. The 183 remaining outlier loci were individually correlated with either temperature (38%), latitude (35%) or longitude (27%), and were therefore likely affected by spatially-varying selection acting during the early glass eel life stage (Gagnaire et al., 2012).

Inevitably, a genome scan approach based on environmental correlations may be underpowered to detect selected alleles with small individual effect, because subtle correlations between allele

frequencies and explanatory variables may not reach statistical significance. This effect is also amplified by the reduced capacity of detecting linkage disequilibrium between our RAD markers and the actual causative alleles under selection, due to limited marker density (<1 SNP per 100kb). Because there is a tradeoff between the power to detect true positive associations and the risk of false positive detections, our list of candidate loci detected here probably contains false positives, while only representing a fraction of the genomic regions affected by spatially-varying selection.

Additive effect of outliers on environmental correlations

Because of the above limitations, and in order to determine the extent to which individual outlier loci collectively mirror environmental variation across the American Eel distribution range, we also used a multilocus approach based on additive polygenic scores for different environmental factors. Ideally, genotypic variation at outlier loci detected in genome scans should be compared with individual phenotypic data to establish associations between genotype, phenotype and fitness (Barrett and Hoekstra, 2011). In the case of polygenic adaptation, additive polygenic scores can be used to assess the joint contribution of individual loci to a quantitative phenotype or a measure of fitness (Gagnaire and Gaggiotti, 2016; Hancock et al., 2011; Aarnegard et al., 2014). Because spatially-varying selection effects on phenotype remain unknown in the American Eel, we could not use this approach to validate the genotype-phenotype link at outlier loci. However, we used a similar approach to evaluate the additive signal of outlier loci against environmental variation. In order to circumvent the possible circularity of an analysis based on the pooled polygenic scores of whole samples, additive polygenic scores were calculated at the individual level. Outlier loci were detected using single locus association tests that do not consider individual genotype information separately, nor preferential allelic associations among loci. Given the panmictic nature of eel population structure, and in absence of

selective effects, genotypic similarity among individuals from a given sampling site would not be expected to be more important than among individuals from different sites. Here, correlations between individual polygenic scores and environmental variables were highly significant and best followed a quadratic model. Interestingly, these quadratic relationships closely mirror the one previously documented between the condition factor and the latitude of glass eels (Laflamme et al., 2012). This provides cues for a possible link between genotype and phenotype and suggests that the total number of locally favourable alleles displayed by individuals affects their chance of survival in a given location. Compared to previous studies on the same topic, our results thus provide new indications on the spatial patterns of polygenic variation generated by spatially-varying selection in eels. This should encourage further developments to specifically detect the molecular footprints of polygenic selection within a multilocus quantitative genetics framework (e.g. Stephan, 2016).

Although the selective factors associated with longitude are still not very clear, the finding of higher polygenic scores at intermediate longitudes as supported by the quadratic model suggest that selection favors different optimal genotypes between the core region of the species range and its periphery on a longitudinal axis. For latitude the correlation may be explained by its strong co-variation with environmental and biotic parameters, which is consistent with the molecular footprints of selection along latitudinal gradients observed in many species, such as *Drosophila melanogaster* (Adrion et al., 2015), *Arabidopsis thaliana* (Debieu et al., 2013) and *Anopheles gambiae* (Cheng et al., 2012).

Moreover, eels that were captured at higher latitude had to migrate for a longer time and distance compared to the ones that were captured at lower latitudes. During their oceanic dispersal, eel larvae face different environmental conditions depending on where they are brought by the ocean currents. This could result in the selection for different alleles during larval transport which mostly follows a latitudinal trajectory. The relationship between polygenic scores and latitude follows a quadratic model which revealed that polygenic scores increase more rapidly at the lowest latitude than at the highest

ones, which is selection gradient appears to be steeper in the southern part of the range.

The river mouth temperature during glass eel recruitment was the third strongest predictor of outlier polygenic scores. River mouth temperature, which can vary from 3 to 18°C at time of recruitment, has previously been identified to be an important selective agent responsible for spatially-varying selection in both the American Eel (Gagnaire et al., 2012) and the European Eel (Pujolar et al., 2014). As for the latitude, the relationship between the polygenic score and temperature follows a quadratic model into which the scores increase more rapidly in the warmer sites than in the colder ones, supporting that the strongest selection gradient is imposed by high temperatures. These observations, as well as the distribution of extreme polygenic scores across sampling sites, suggest that selection favors different optima throughout the species range by eliminating locally maladapted extreme genotypes.

GO term enrichment analysis of outlier loci

In order to find cues regarding the phenotypic effects of the selected loci, we analysed the functions of the 241 genes that could be associated with outliers. Among these, four genes coding for subunits of different mitochondrial electron transport chain complexes were found. This is a relevant result considering the role of the electron transport chain in ATP production, and the range of distances that glass eels have to travel to reach different parts of the species range during dispersal from the Sargasso Sea.

Another result of interest was the finding of five genes that were previously identified to be under selection between life stages in the European Eel (Pujolar et al., 2015). Because American and European eels are only partially reproductively isolated and exchange genes through introgressive hybridization (Albert et al., 2006; Gagnaire et al., 2009; Jacobsen et al., 2017), the sharing of locally

adaptive variants is expected to occur. However, previous studies have found only limited evidence for parallel patterns of spatially varying selection in Atlantic eels (Ulrik et al., 2014). Among the five candidate genes shared between species, three (*pactr4b*, *SNX14* and *SHANK3*) are implicated in brain development (Cho et al., 2014; Ha et al., 2015; Kozol et al., 2015), perhaps reflecting similar selective constraints on this biological pathway in the two species.

Finally, the insulin signaling pathway, which was represented by five outlier genes, was found to be significantly enriched among the list of outlier genes potentially affected by selection. Here again, this pathway was previously found to be significantly enriched among genes affected by selection between life stages in the European Eel (Pujolar et al., 2015), which strengthens its candidate functional role. The insulin pathway is activated in response to the binding of the hormone on its receptor and plays a central role in the regulation of usage and storage of cellular glucose (Bevan, 2001). Phenotypically, this pathway has also been related to body size in *Drosophila* (De Jong and Bochdanovitz, 2003). In American glass eels, measures of otolith daily growth increments have shown a latitudinal gradient in growth rate (Wang and Tzeng, 1998), with faster growing and earlier maturing glass eels at intermediate latitudes. The overrepresentation of the insulin signalling pathway playing a role in the use of energy reserve also corroborates previous findings indicating selection on lipid and saccharide metabolism pathways (Gagnaire et al., 2012). Altogether, this suggests that metabolic and energy production are among the more important ones pathways under spatially-varying selection in the American Eel. Because spatial variation in individual growth rates has been reported (Wang and Tzeng, 1998), future studies should aim at testing whether outlier polygenic scores partly account for differences in growth rate across the species range, or whether these are mainly caused by phenotypic plasticity in response to environmental conditions.

Finally, the admittedly small subset of annotated genes found here to be potentially under selection

represents a good first set of candidates for further functional studies on the mechanism implicated in the spatially varying selection in American Eel. For instance, the expression of these genes in individuals facing different environmental conditions (e.g. different water temperatures) could be analyzed in order to test for differential allelic expression and its association with growth related phenotypes in the different conditions.

Authors' contribution

C.B. and L.B. conceived the study. C.B. performed the lab work. C.B., P.A.G and S.A.P analyzed the data. C.B and P.A.G wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript.

Funding

This work was supported by a Natural Science and Engineering Research Council of Canada (NSERC) discovery grant and Canadian Research Chair in Genomics and Conservation of Aquatic Resources to L.B. [grant number STGP 493768].

Acknowledgements

We thank Clément Rougeux who helped for library preparation, Damien Boivin-Delisle who helped for DNA extraction and all the Bernatchez lab members for discussion and support. We are also thankful to two anonymous referees and Associate Editor C Baer for their constructive and useful comments on an earlier version of the manuscript.

Conclusion

Étant son cycle de vie assez particulier qui fait en sorte que tous les individus appartiennent à la même population, l'anguille d'Amérique est une espèce intéressante à étudier pour mieux comprendre les mécanismes de sélection. En effet, sa grande taille effective (Côté et al., 2013) et l'absence de structure de population font en sorte que la sélection naturelle est la principale force évolutive qui s'exerce sur cette espèce. La grande étendue de l'aire de répartition fait également en sorte que les différents individus sont confrontés à un grand éventail de conditions environnementales selon l'endroit où ils dérivent. Ces différentes conditions peuvent mener à la sélection de différents allèles, car ceux-ci permettent aux individus de mieux répondre aux pressions de sélection qui s'exercent sur eux. En identifiant des marqueurs potentiellement sous sélection, cette étude permet donc d'augmenter notre compréhension des phénomènes de sélection chez l'anguille d'Amérique. Des méthodes utilisant des scores additifs polygéniques ont également été utilisées, ce qui est relativement nouveau.

Identification de marqueurs potentiellement sous sélection spatialement variable

Pour étudier ce phénomène, des individus ont été capturés dans 12 sites répartis dans une grande partie de l'aire de répartition, de Terre-Neuve à la Floride. Ils ont ensuite été séquencés et après divers traitements bio-informatiques, 12 098 marqueurs présents chez 80 % des individus et suffisamment polymorphiques ont été conservés pour les analyses subséquentes. À l'aide de 2 approches, les analyses de redondance et d'association environnementale, 259 marqueurs dont les fréquences alléliques dans les différents sites étaient fortement corrélées aux valeurs des variables environnementales ont pu être identifiés. Une vingtaine de marqueurs ont été identifiés comme étant potentiellement sous sélection par les 2 méthodes de détection. Les analyses de redondance partielles ont permis d'identifier la proportion de la variance génétique expliquée par les différentes variables parmi les marqueurs potentiellement sous sélection. L'année de capture explique la plus grande proportion de la variance avec 10.3%, suivi de la température avec 8%. La latitude et la longitude expliquent quant à elles 6% de la variance génétique chacune. C'est également l'année de capture qui est la variable ayant la plus grande influence sur le plus grand nombre de marqueurs avec 76 marqueurs. La température affecte de façon plus importante 70 marqueurs. La latitude affecte 64 marqueurs et la longitude est liée à 49 marqueurs. Comme plus du quart de ces marqueurs (76 sur 259) sont principalement associés à l'année

de capture des individus, il semble y avoir une variation temporelle dans la sélection qui s'exerce sur les anguilles. Cette variation temporelle est probablement causée par les conditions qui varient entre les années ainsi que le pool génique qui est aussi différent selon les années. En effet, le nombre de reproducteurs présents dans la mer Sargasses varie de façon importante d'année en année et cette variation semble liée à l'oscillation nord-atlantique (Côté et al., 2013). L'oscillation nord-atlantique (NAO) est une variation des courants au niveau océanique et semble avoir un effet sélectif important chez de nombreuses espèces (Sipielsky et al. 2017).

Effet additif des marqueurs potentiellement sous sélection

Pour mieux visualiser l'effet cumulatif des divers marqueurs potentiellement sous sélection, une approche d'additivité a également été employée (Hancock et al, 2011; Arnegard et al, 2014; Gagnaire et Gaggiotti, 2016). Pour ce faire, des scores polygéniques ont été calculés pour chacun des individus. Le score polygénique correspond au nombre d'allèles favorables porté par chacun des individus. Les allèles favorables sont ceux qui sont liés à des valeurs croissantes des variables environnementales. Les scores polygéniques ont été calculés uniquement avec les marqueurs potentiellement sous sélection. Des corrélations ont ensuite été effectuées entre les scores polygéniques et les valeurs de température, de la latitude et de longitude des différents d'échantillonnage dans lesquels les individus ont été capturés. Les corrélations ont été significatives pour les trois variables et elles avaient des coefficients de corrélation qui sont comparables ($R^2=0.142-0.154$). Les corrélations entre les scores polygéniques et les variables environnementales suivent toutes un modèle quadratique. La température et la latitude sont les variables pour lesquelles l'effet sélectif est le plus évident. Il est logique que des températures plus ou moins froides entraînent la sélection d'allèles différents. La température avait d'ailleurs déjà été identifiée comme étant un agent sélectif à la fois chez l'anguille d'Amérique (Gagnaire et al., 2012) et l'anguille européenne (Pujolar et al., 2014). Comme le climat varie souvent avec la latitude, il est également logique que celle-ci ait un effet sélectif. Des clines latitudinaux ont été décrits chez de nombreuses espèces, que ce soit chez la mouche (Adrion et al., 2015) ou chez des plantes, comme *Arabidopsis thaliana* (Debieu et al., 2013). Il est plutôt surprenant que la longitude ait un coefficient de corrélation comparable aux 2 autres variables étant donné que son effet sélectif n'est pas très clair. Toutefois, comme la longitude des sites augmente avec celle de la latitude, les individus capturés plus à l'est l'ont été dans des régions plus froides de l'aire de répartition par rapport aux individus capturés plus à l'ouest, ce qui peut entraîner la sélection d'allèles différents au sein de ces sites.

Fonctions enrichies chez les marqueurs potentiellement sous sélection

Les gènes situés à moins de 100 kb de part et d'autres des marqueurs ont été identifiés, car ils sont potentiellement sujets à la sélection. Des tests d'enrichissement ont été effectués sur les fonctions de ces gènes et aucune fonction n'était significativement enrichie. Les gènes faisant partie du mécanisme de réponse à l'insuline étaient par contre significativement enrichies parmi les gènes potentiellement sous sélection. Il est impliqué dans le métabolisme des sucres au sein de la cellule. Ce mécanisme a été lié à la taille des individus chez la drosophile (De Jong et Bochdanovits, 2003). Ce mécanisme avait également été identifié comme étant enrichi parmi les marqueurs potentiellement sous sélection entre les étapes du cycle de vie chez l'anguille européenne (Pujolar et al., 2015). Cinq gènes identifiés dans notre étude comme étant potentiellement sous sélection avaient déjà été identifié comme tel dans cette même étude. De façon surprenante, aucun de ces 5 gènes ne fait partie du mécanisme de réponse à l'insuline, même si ce mécanisme est enrichi parmi les gènes identifiés comme étant potentiellement sous sélection dans les 2 études. Ce sont donc des gènes potentiellement sous sélection différents qui font partie du mécanisme au sein des 2 espèces. Le fait que ce mécanisme soit lié à la taille du corps pourrait peut-être expliquer certaines différences dans les taux de croissance qui sont observées dans les différentes régions de l'aire de répartition. En effet, il y a de grandes différences dans le taux de croissance des anguilles selon leur latitude, le taux de croissance diminuant avec l'augmentation de la latitude (Jessop et al., 2010). Ces différences étaient attribuées à la saison de croissance qui diminue avec l'augmentation de la latitude, mais il pourrait également y avoir une composante génétique à cette différence entre les taux de croissance.

Malgré le fait que les mêmes individus et les mêmes variables environnementales aient été utilisé, aucun des gènes identifiés par Gagnaire et al. (2012) comme étant potentiellement sous sélection spatialement variable n'a été trouvé à proximité des marqueurs potentiellement sous sélection. Les 13 gènes en question sont situés dans des régions du génome dans lesquelles aucun des 12098 marqueurs retenus pour cette étude ne figure et c'est pour cette raison que ces gènes ne sont pas présents parmi ceux qui sont liés aux marqueurs potentiellement sous sélection.

Pour conclure, la présence de plusieurs régions potentiellement sous sélection spatialement variable chez l'anguille d'Amérique a été confirmée. Une variation annuelle a également pu être identifiée au sein des sites qui ont été échantillonnées durant 2 années successives. Notre analyse utilisant des scores

polygéniques additifs semble montrer la présence de sélection polygénique chez l'anguille d'Amérique. Le mécanisme de réponse à l'insuline est enrichi parmi les gènes potentiellement sous sélection. Il serait intéressant d'examiner dans une future étude s'il y a des variations dans l'expression des gènes potentiellement sous sélection en fonction des conditions environnementales et quelles sont les effets phénotypiques des marqueurs potentiellement sous sélection. Il serait également intéressant de vérifier si certains marqueurs sont liés au taux de croissance des individus et d'identifier les gènes qui influencent la croissance des anguilles d'Amérique.

Bibliographie

- Aarnegard ME et al. (2014) Genetics of ecological divergence during speciation. *Nature*, **507**, 307-311.
- Adrion JR, Hahn MW, Cooper BS (2015) Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends in Genetics*, **31**, 434-444.
- Albert, V, Jónsson, B, Bernatchez, L (2006). Natural hybrids in Atlantic eels (*Anguilla anguilla*, *A. rostrata*): evidence for successful reproduction and fluctuating abundance in space and time. *Molecular Ecology*, **15**, 1903-1916.
- Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692-4693.
- Barrett RDH, Hoekstra HE (2011) Molecular spandrel: test of adaption at the genetic level. *Nature Reviews Genetics*, **12**, 767-780.
- Avise JC, Helfman GS, Saunders NC, Hales LS (1986) Mitochondrial DNA differentiation in North Atlantic eels: Population genetic consequences of a unusual life history pattern. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 4350-4354.
- Béguer-Pon M, Castonguay M, Shan S, Benchetrit J, Dodson JJ (2015) Direct observations of American eels migrating across the continental shelf to the Sargasso Sea. *Nature Communications*, **6**, 8705.
- Benestan L, Gosselin T, Perrier C, Sainte-Marie B, Rochette R, Bernatchez L (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology* **24**, 3299-3315.
- Bevan P (2001) Insulin signaling. *Journal of Cell Science*, **114**, 1429-1430.
- Blair LM, Granka JM, Feldman MW (2014) On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, **8**, 1.
- Bonhommeau S, Castonguay M, Rivot E, Sabatie R, Le Pape O (2010) The duration of migration of Atlantic *Anguilla* larvae. *Fish and Fisheries*, **11**, 289-306.
- Bourret V, Dionne M, Bernatchez L (2014) Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular Ecology*, **23**, 4444-4457.
- Cairns DK, Secor DA, Morrison WE, Hallett JA (2009) Salinity-linked growth in anguillid eels and the paradox of temperate-zone catadromy. *Journal of Fish Biology*, **74**, 2094-2114.
- Castonguay M, Hodson PV, Couillard CM, Eckersley MJ, Dutil JD, Verreault G (1994) Why is recruitment of the american eel, *Anguilla rostrata*, declining in the St. Lawrence river and gulf? *Canadian Journal of Fisheries and Aquatic sciences*, **51**, 479-488.

Catchen JM, Amores A, Hohenloe PA, Cresko WA, Postlethwait JH (2011) Stacks: Building and genotyping loci *de novo* for short read sequences. *G3*, **1**, 171-182.

Catchen JM, Hohenloe PA, Bassham S, Amores A, Cresko WA. 2013 Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124-3140. (doi:10.1111/mec.12354)

Cheng CD, White BJ, Kamdem C, Mokaitis K, Constantini C, Hahn MW (2012) Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resquencing approach. *Genetics*, **190**, 1417-1432.

Cho HM, Kim JY, Kim H, Sun W (2014) Phosphatase and actin regulator 4 is associated with intermediate filament in adult neural stem cells and their progenitor astrocytes. *Histochemistry and Cell Biology*, **142**, 411-419.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411-1425.

COSEPAC (Comité sur la situation des espèces en péril au Canada) (2012) Évaluation et rapport de situation du COSEPAC sur l'anguille d'Amérique *Anguilla rostrata* au Canada. COSEPAC, Ottawa.

Côté CL, Gagnaire PA, Bourret V, Verreault G, Castonguay M, Bernatchez L (2013) Population genetics of the American Eel (*Anguilla rostrata*): F_{st}=0 and North Atlantic oscillation effects on demographic fluctuations of a panmictic species. *Molecular Ecology*, **22**, 1763-1776.

Davey AJH, Jellyman DJ (2005) Sex determination of freshwater eels and management options for manipulation of sex. *Reviews in Fish Biology and Fisheries*, **15**, 37-52.

De Jong G, Bochdanovitz Z (2003) Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies inversion frequencies, and the insulin-signalling pathway. *Journal of Genetics*, **82**, 207-223.

Debieu M, Tang C, Stich B, Sikosek T, Effgen S, Joshephs E, Schmitt J, Nordborg M, Koornneef M, de Meaux J (2013) Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana*. *PLOS One*, **8**, e61075.

de Villemereuil P, Frichot E, Bazin E, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, **23**, 2006-2019.

Ellegren H (2014) Genome sequencing and population genomics in non-model organism. *Trends in Ecology and Evolution*, **29**, 51-63.

Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for association between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687-1699.

Gagnaire, PA, Albert, V, Jónsson, B, Bernatchez, L (2009). Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. *Molecular Ecology*, **18**, 1678-1691.

- Gagnaire PA, Normandeau E, Côté C, Hansen MM, Bernatchez L (2012) The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (*Anguilla rostrata*). *Genetics*, **190**, 725-736.
- Gagnaire PA, Gaggiotti OE (2016) Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, **62**, 603-616.
- Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**, 4555-4573.
- Guillot G, Vitalis R, le Rouzic A, Gautier M (2014) Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spatial Statistics*, **8**, 145-155.
- Gunther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205-220.
- Ha et al. (2015) SNX14 is a bifunctional negative regulator for neuronal 5-HT6 receptor signalling. *Journal of Cell Science*, **128**, 1848-1861.
- Häcker H, Karin M (2006) Regulation and function of IKK and IKK-related Kinases. *Science's STKE* **2006**, re13.
- Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, **334**, 83-86.
- Hanel R, Stepputtis D, Bonhommeau S, Castonguay M, Schaber M, Wysujack K, Vobach M, Miller MJ (2014) Low larval abundance in the Sargasso Sea: new evidence about reduced recruitment of the atlantic eels. *Naturwissenschaften*, **101**, 1041-1054.
- Hedgecock D, Pudovkin AI (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science*, **87**, 971-1002.
- Herron JC, Freeman S (2014) Evolutionary analysis. 5e édition. Prentice-Hall.
- Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Whitlock MC (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, **188**, 379-397.
- Horsley V, Pavlath GK (2002) NFAT: ubiquitous regulator of cell differentiation and adaptation. *The Journal of Cell Biology*, **156**, 771-774.
- Huang DW et al. (2007) The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biology*, **8**, R183.
- Hurrell JW, Kushnir Y, Visbeck M (2001) Climate-The north atlantic oscillation. *Science*, **291**, 603-605.

- Jacobsen MW, Smedegaard L, Sørensen SR, Pujolar JM, Munk P, Jónsson B, Magnussen E, Hansen MM (2017) Assessing pre- and post-zygotic barriers between North Atlantic eels (*Anguilla anguilla* and *Anguilla rostrata*). *Heredity*, **118**, 266-275.
- Jeffreys H. (1961) Theory of probability. Oxford University Press, Oxford, UK. 470 pp.
- Jessop BM (2010) Geographic effects on American eel (*Anguilla rostrata*) life history characteristics and strategies. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 326-346.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225-1241.
- Koehn RK, Williams GC (1978) Genetic differentiation without isolation in the American Eel (*Anguilla rostrata*). II. Temporal stability of geographic patterns. *Evolution*, **32**, 624-637.
- Laflamme S, Côté C, Gagnaire PA, Castonguay M, Bernatchez L (2012) RNA/DNA ratios in American glass eels (*Anguilla rostrata*): evidence for latitudinal variation in physiological status and constraints to oceanic migration? *Ecology and Evolution*, **2**, 875-884.
- Laporte M et al. (2016) RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. *Molecular Ecology*, **25**, 219-237.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548-1566.
- Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, **87**, 331-333.
- Li H, Durbin R (2009) Fast and accurate short-read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
- Li H, Handsaker B, Wysoker A, Fenell T, Ruan J, Homer N, Marth G, Abecassis G, Durbin R, 1000 Genomes Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987-2993.
- Kozol RA et al. (2015) Two knockdown models of the autism genes SYNGAP1 and SHANK3 in zebrafish produce behavioral phenotypes associated with embryonic disruptions of brain morphogenesis. *Human Molecular Genetics*, **24**, 4006-4023.
- McCleave JD, Kleckner RC (1982) Selective tidal stream transport in the estuarine migration of glass eels of the American Eel (*Anguilla rostrata*). *ICES Journal of Marine Science*, **40**, 262-271.
- Maes GE, Pujolar JM, Hellemans B, Volckaert FAM (2006) Evidence for isolation by time in the European eel (*Anguilla anguilla* L.). *Molecular Ecology*, **15**, 2095-2107.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.

Embnet.journal, **17**, 10-12.

Metzker ML (2010) Sequencing technologies-the next generation. *Nature Reviews Genetics*, **11**, 31-46.

Mitchell-Olds T, Willis JH, Golstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits. *Nature Reviews Genetics*, **8**, 845-856.

Oliveira K, McCleave JD, Wippelhauser JS (2001) Regional variation and the effect of lake: river area on sex distribution of American eel. *Journal of Fish Biology*, **58**, 943-952.

Palm DC, Rohwer JM, Hofmeyr JHS (2013) Regulation of glycogen synthase from mammalian skeletal muscle – a unifying view of allosteric and covalent regulation. *The FEBS Journal*, **280**, 2-27.

Pavey SA, Gaudin J, Normandeau E, Dionne M, Castonguay M, Audet C, Bernatchez L (2015) RAD Sequencing Highlights Polygenic Discrimination of Habitat Ecotypes in the Panmictic American Eel. *Current Biology*, **25**, 1666-1671.

Pavey SA, Laporte M, Normandeau E, Gaudin J, Letourneau L, Boivert S, Corbeil J, Audet C, Bernatchez L (2016) Draft genome of the American eel (*Anguilla rostrata*). *Molecular Ecology Resources*

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

Pujolar JM, Maes GE, Volckaert FAM (2006) Genetic patchiness among recruits in the European eel *Anguilla anguilla*. *Marine Ecology Progress Series*, **307**, 209-217.

Pujolar JM et al. 2014 Genome-wide single generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, **23**, 2514-2528.

Pujolar JM, Jacobsen MW, Bekkevold D, Lobón-Servià J, Jónsson B, Bernatchez L, Hanssen MM (2015) Signatures of natural selection between life cycle stages separated by metamorphosis in European eel. *BMC Genomics*, **16**, 600.

Roff DA 1997 Evolutionary quantitative genetics. New York, USA, Springer.

Sipielsky AM et al. (2017) Precipitation drives global variation in natural selection. *Science*, **355**, 959-962.

Stephan, W (2016) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology* **25**, 79-88.

Ulrik, MG et al. (2014) Do North Atlantic eels show parallel patterns of spatially varying selection? *BMC Evolutionary Biology*, **14**, 138.

Vélez-Espino LA, Koops MA (2010) A synthesis of the ecological processes influencing variation in the life history and movement patterns of American eel: towards a global assessment. *Reviews in Fish Biology and Fisheries*, **20**, 163-186.

- Wang, CH, Tzeng WN (1998) Interpretation of geographic variation in size of American eel *Anguilla rostrata* elvers on the Atlantic coast of North America using their life history and otolith ageing. *Marine Ecology Progress Series*, **168**, 35-43.
- Wapples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419-1439.
- Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu HZ, Shen SH, Guo W, Xing Y, Carstens RP (2010) An ESRP-regulated splicing program is abrogated during the epithelial-mesenchymal transition. *The EMBO Journal*, **29**, 3286-3300.
- Wellenreuther M, Hansson B (2016) Detecting polygenic evolution: problems, pitfalls and promises. *Trends in Genetics*, **32**, 155-164.
- Williams GC, Koehn RK, Mitton JB (1973) Genetic differentiation without isolation in the American Eel (*Anguilla rostrata*). *Evolution*, **27**, 192-204.
- Wirth T, Bernatchez L (2003) Decline of north atlantic eels: a fatal synergy? *Proceedings of the Royal Society B-Biological Sciences*, **270**, 681-688.
- Yeaman S, Otto S (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*, **65**, 2123-2129.
- Yeaman S (2015) Local adaptation by alleles of small effect. *The American Naturalist*, **186**, S74-S89.

Annexes

Supplementary table 1 Markers detected to potentially under spatially-varying selection by the redundancy analysis (RDA) among all the markers

Outlier marker position on the American Eel genome
scaffold_5:27546
scaffold_7:79224
scaffold_14:280565
scaffold_14:315211
scaffold_19:31101
scaffold_34:189861
scaffold_34:342814
scaffold_40:121905
scaffold_44:36535
scaffold_69:224656
scaffold_77:33637
scaffold_84:97072
scaffold_92:59240
scaffold_92:146867
scaffold_92:318412
scaffold_95:6269
scaffold_100:105509
scaffold_100:121840
scaffold_100:298357
scaffold_100:412876
scaffold_116:46684
scaffold_116:72149
scaffold_124:87079
scaffold_125:287879
scaffold_137:6254
scaffold_138:64209
scaffold_150:179380
scaffold_151:108509
scaffold_151:230113
scaffold_154:105066
scaffold_158:142877
scaffold_162:51022
scaffold_170:241467
scaffold_182:129578
scaffold_189:69414
scaffold_189:101214
scaffold_195:179856
scaffold_197:324412
scaffold_232:210075
scaffold_242:227809
scaffold_246:51244
scaffold_258:121241

scaffold_262:95247
scaffold_262:183263
scaffold_274:131955
scaffold_280:36149
scaffold_287:142419
scaffold_287:148865
scaffold_287:371343
scaffold_291:99350
scaffold_295:53696
scaffold_311:289784
scaffold_313:6691
scaffold_319:143343
scaffold_327:255876
scaffold_340:31425
scaffold_342:79861
scaffold_352:54413
scaffold_355:29087
scaffold_355:108622
scaffold_358:69712
scaffold_359:79078
scaffold_361:82578
scaffold_362:51167
scaffold_369:85889
scaffold_383:267527
scaffold_403:209810
scaffold_405:13633
scaffold_407:8558
scaffold_410:63300
scaffold_415:3651
scaffold_415:57072
scaffold_421:111812
scaffold_423:233807
scaffold_423:262444
scaffold_424:88862
scaffold_424:106443
scaffold_431:267768
scaffold_433:555
scaffold_448:122712
scaffold_451:58347
scaffold_459:129782
scaffold_470:24037
scaffold_487:248073
scaffold_491:30286
scaffold_494:28558
scaffold_496:85216
scaffold_498:184546
scaffold_498:303194

scaffold_518:201179
scaffold_527:23524
scaffold_537:25331
scaffold_544:27837
scaffold_578:58969
scaffold_585:33117
scaffold_591:286759
scaffold_591:374520
scaffold_592:101457
scaffold_596:91712
scaffold_597:5180
scaffold_606:50456
scaffold_616:91295
scaffold_618:454006
scaffold_624:204317
scaffold_634:60381
scaffold_634:91776
scaffold_634:158730
scaffold_639:40258
scaffold_639:67366
scaffold_639:304977
scaffold_643:71385
scaffold_663:152614
scaffold_670:21235
scaffold_683:1680
scaffold_683:72800
scaffold_688:165118
scaffold_691:254689
scaffold_698:186659
scaffold_701:146826
scaffold_713:110580
scaffold_716:123714
scaffold_717:36728
scaffold_719:29929
scaffold_732:21878
scaffold_754:38858
scaffold_760:149657
scaffold_777:90364
scaffold_781:17339
scaffold_788:24627
scaffold_795:18780
scaffold_795:27857
scaffold_804:25399
scaffold_805:41597
scaffold_809:16803
scaffold_822:384256
scaffold_826:174023

scaffold_833:112277
scaffold_856:35188
scaffold_861:89833
scaffold_863:251064
scaffold_867:12604
scaffold_867:76090
scaffold_877:153060
scaffold_881:54322
scaffold_895:89426
scaffold_901:150908
scaffold_909:16139
scaffold_911:54248
scaffold_911:57310
scaffold_911:121610
scaffold_912:109641
scaffold_914:4304
scaffold_921:257474
scaffold_928:97341
scaffold_932:11770
scaffold_932:13668
scaffold_932:122927
scaffold_932:161897
scaffold_935:97924
scaffold_938:48366
scaffold_942:65803
scaffold_945:53185
scaffold_956:189647
scaffold_960:260632
scaffold_965:72999
scaffold_970:57778
scaffold_985:98633
scaffold_997:98227
scaffold_1001:17312
scaffold_1001:318071
scaffold_1006:232868
scaffold_1011:66186
scaffold_1028:67251
scaffold_1028:71981
scaffold_1028:118728
scaffold_1032:132584
scaffold_1033:178594
scaffold_1038:154437
scaffold_1043:89889
scaffold_1047:16531
scaffold_1053:51773
scaffold_1070:509
scaffold_1074:9666

scaffold_1079:156034
scaffold_1103:32188
scaffold_1113:18930
scaffold_1114:51373
scaffold_1118:62217
scaffold_1122:8995
scaffold_1129:72565
scaffold_1149:87112
scaffold_1165:30033
scaffold_1169:240839
scaffold_1180:11663
scaffold_1197:18681
scaffold_1211:80067
scaffold_1214:324455

Supplementary table 2 Markers detected to potentially under spatially-varying selection by the Bayenv analysis among all the markers

Outlier marker position
scaffold_11:58687
scaffold_14:315187
scaffold_15:42098
scaffold_34:189861
scaffold_40:121878
scaffold_40:121905
scaffold_69:224656
scaffold_75:157410
scaffold_77:33637
scaffold_95:80441
scaffold_97:52452
scaffold_100:295966
scaffold_100:436400
scaffold_146:92244
scaffold_146:176964
scaffold_146:256515
scaffold_151:230113
scaffold_158:142877
scaffold_174:88308
scaffold_195:40811
scaffold_210:40685
scaffold_222:79698
scaffold_258:121241
scaffold_287:308175
scaffold_291:99450
scaffold_311:289784
scaffold_313:53700
scaffold_315:70452
scaffold_336:158271
scaffold_349:24120
scaffold_355:108622
scaffold_369:85927
scaffold_393:55797
scaffold_407:8595
scaffold_433:4532
scaffold_434:197643
scaffold_442:1356
scaffold_494:59598
scaffold_497:45977
scaffold_510:195183
scaffold_527:23551
scaffold_582:91959
scaffold_585:33117

scaffold_624:225641
scaffold_639:67366
scaffold_639:124790
scaffold_662:197223
scaffold_675:164885
scaffold_687:99154
scaffold_688:130826
scaffold_690:118084
scaffold_698:186652
scaffold_726:63476
scaffold_771:144456
scaffold_781:17367
scaffold_788:24627
scaffold_791:24873
scaffold_833:112277
scaffold_842:203554
scaffold_844:27437
scaffold_844:27504
scaffold_847:182126
scaffold_854:107194
scaffold_858:112204
scaffold_882:190659
scaffold_912:109641
scaffold_926:28753
scaffold_932:122900
scaffold_935:97980
scaffold_935:185435
scaffold_939:33287
scaffold_942:65803
scaffold_953:46205
scaffold_972:132011
scaffold_981:91365
scaffold_1003:68681
scaffold_1006:171827
scaffold_1009:89652
scaffold_1011:66186
scaffold_1033:45910
scaffold_1033:99093
scaffold_1047:133147
scaffold_1072:20575
scaffold_1113:18930
scaffold_1139:125085
scaffold_1149:46085
scaffold_1180:11663
scaffold_1182:28022
scaffold_1202:223120
scaffold_1202:223122

