HAL
open science

RAD-Seq Reveals Patterns of Additive Polygenic Variation Caused by Spatially-Varying Selection in the American Eel (Anguilla rostrata)

Charles Babin, Pierre-Alexandre Gagnaire, Scott Pavey, Louis Bernatchez

To cite this version:

Charles Babin, Pierre-Alexandre Gagnaire, Scott Pavey, Louis Bernatchez. RAD-Seq Reveals Patterns of Additive Polygenic Variation Caused by Spatially-Varying Selection in the American Eel (Anguilla rostrata). Genome Biology and Evolution, 2017, 9 (11), pp.2974-2986. 10.1093/gbe/evx226 . hal-02395860

HAL Id: hal-02395860

https://hal.science/hal-02395860v1

Submitted on 5 Dec 2019

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

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

RAD-Seq Reveals Patterns of Additive Polygenic Variation Caused by Spatially-Varying Selection in the American Eel (*Anguilla rostrata*)

Charles Babin[1,*], Pierre-Alexandre Gagnaire[2], Scott A. Pavey[3], and Louis Bernatchez[1]

[1]IBIS (Institut de biologie intégrative et des systèmes), Département de biologie, Université Laval, Québec, Canada

[2]Institut des Sciences de l'Évolution-Montpellier (ISEM-CNRS UMR5554), SMEL, Sète, France

[3]Department of Biological Sciences and Canadian Rivers Institute, University of New Brunswick, Saint-John, Canada

*Corresponding author: E-mail: charles.babin.1@ulaval.ca.

Accepted: November 9, 2017

Abstract

The American Eel (*Anguilla rostrata*) has an exceptional life cycle characterized 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 single nucleotide polymorphisms to re-evaluate the extent to which local selection affects the spatial distribution of genetic diversity in this species. We used environmental association methods to identify markers under spatially-varying selection, which indicated that selection affects ~1.5% of the genome. We then evaluated the extent to which candidate markers collectively vary with environmental factors 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.

Key words: panmixia, spatially-varying selection, environmental association, RAD-sequencing, genome scan, polygenic selection.

Introduction

Panmictic species challenge the classic view of local adaptation, whereby individuals from local populations are more fit to their own habitat compared with migrants from different populations (Kawecki and Ebert 2004). Indeed, local adaptation cannot develop in panmictic species because 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 (Avise et al. 1986; Wirth and Bernatchez 2003; Côté et al. 2013; Pavey et al. 2017) 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,

© The Author 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

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 (Béguer-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 and Koops 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 versus 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; Le Corre and Kremer 2012). 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. 2017). 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 Eels, and controlled for temporal effects by focusing 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 et al. (2012) were used for this study. Those individuals were sampled in 2008 at the mouth of 12 tributaries, soon after recruitment to

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

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 (fig. 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 2. 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 (2015) using the restriction enzyme, *Eco*RI 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 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, 2013). Reads were then mapped on the American Eel genome (Pavey et al. 2017) 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. 2017). SNPs were called using SAMtools mpileup and BCFtools view modules (Li et 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 minimizing 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 1,500 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; last accessed November 6, 2017). 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.

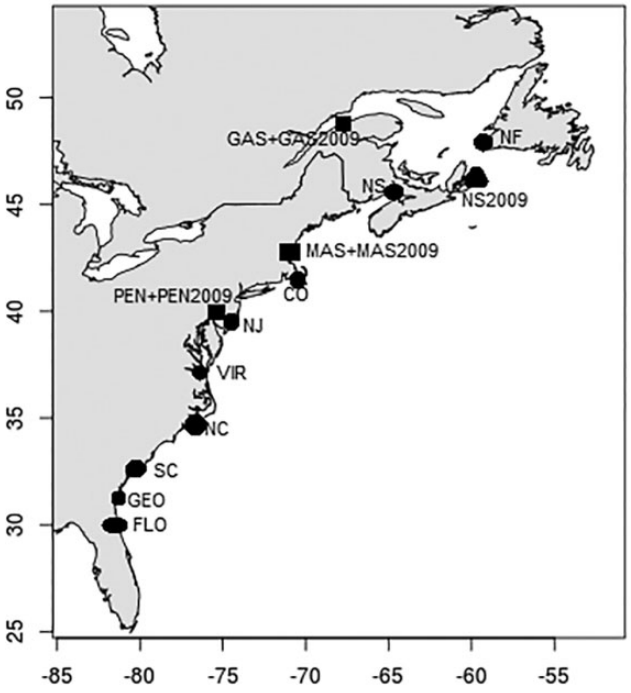

FIG. 1.—Location map of the 13 sampling sites where American glass eels were collected for this study. Labels correspond to the site codes provided in table 1. The sites identified by squares were sampled in 2008 and 2009, sites identified by circles were sampled in 2008, and the site identified by a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

Redundancy Analysis (RDA) to Detect Outlier Loci

A 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 data set 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 1,000 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

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

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
Gaspesie	GAS	47	48, 78	−67, 7	14, 21	2008
Gaspesie	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

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 neither 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 RDA 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 >3, and above 3 for at least 6 of the 12 runs. We used a threshold >3 following the recommendation of Jeffreys (1961) for the significance of Bayes factor. Then, a RDA 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 data sets. A total of 100 data sets 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 data sets. Then only one run of Bayenv2 was performed for each data set 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 data set for the different variables.

Detecting Variables with the Strongest Contribution to Outlier Variation

Conditioned RDAs (partial RDAs) were performed on a combined data set comprising all the outliers detected by the RDA and by the Bayenv2 approaches. These partial

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

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 favored in a given environment (Hancock et al. 2011; Aarnegard 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 favored 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; last accessed November 6, 2017; Huang et al. 2007). Zebrafish (*Danio rerio*) annotations were used as a reference genome for the GO term enrichment analysis, because this species has the most complete list of annotated genes with known functions among teleost fishes.

Results

Genome-Wide Polymorphism Data Set

Sequencing of RAD libraries resulted in a total of 3.2×10^9 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.

RDA to Detect Loci Associated with Explanatory Variables

The 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 (fig. 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, Supplementary Material online).

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 (fig. 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, Supplementary Material online).

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

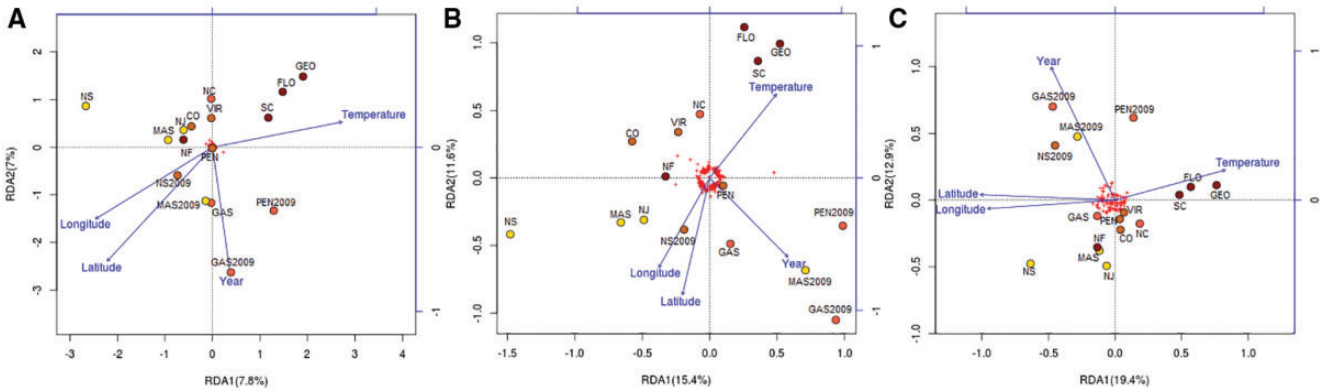

Fig. 2.—(*A*) Redundancy analysis (RDA) performed with 12,098 filtered SNPs, using latitude, longitude, temperature, and year as constraining variables. (*B*) Redundancy analysis performed using the 197 SNPs detected to be potentially affected by selection with the RDA outlier detection approach. (*C*) Redundancy analysis performed using the 90 markers significantly associated with environmental variables in Bayenv2 analysis. Only the first two 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 four 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.

Among these, 15 outlier markers were associated with the latitude of the sampling site, 16 to the longitude (among which four 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 100 bp 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 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 (fig. 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 data set compared with the simulated ones for the four variables (fig. 3*A*–*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 data set. 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 data set 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 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, Supplementary Material online.

Using Additive Scores to Evaluate Polygenic Variation Patterns

The correlations assessed between each variable and the corresponding additive scores calculated using outlier loci (fig. 4*A*–*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 ($R^2 = 0.154$, $P \leq 10E\text{-}15$), followed by the latitude ($R^2 = 0.152$, $P \leq E10\text{-}15$) and the temperature ($R^2 = 0.142$, $P \leq E10\text{-}15$). Correlations between the 10th and 90 centile of polygenic scores in a site and variables were significant for all three variables

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

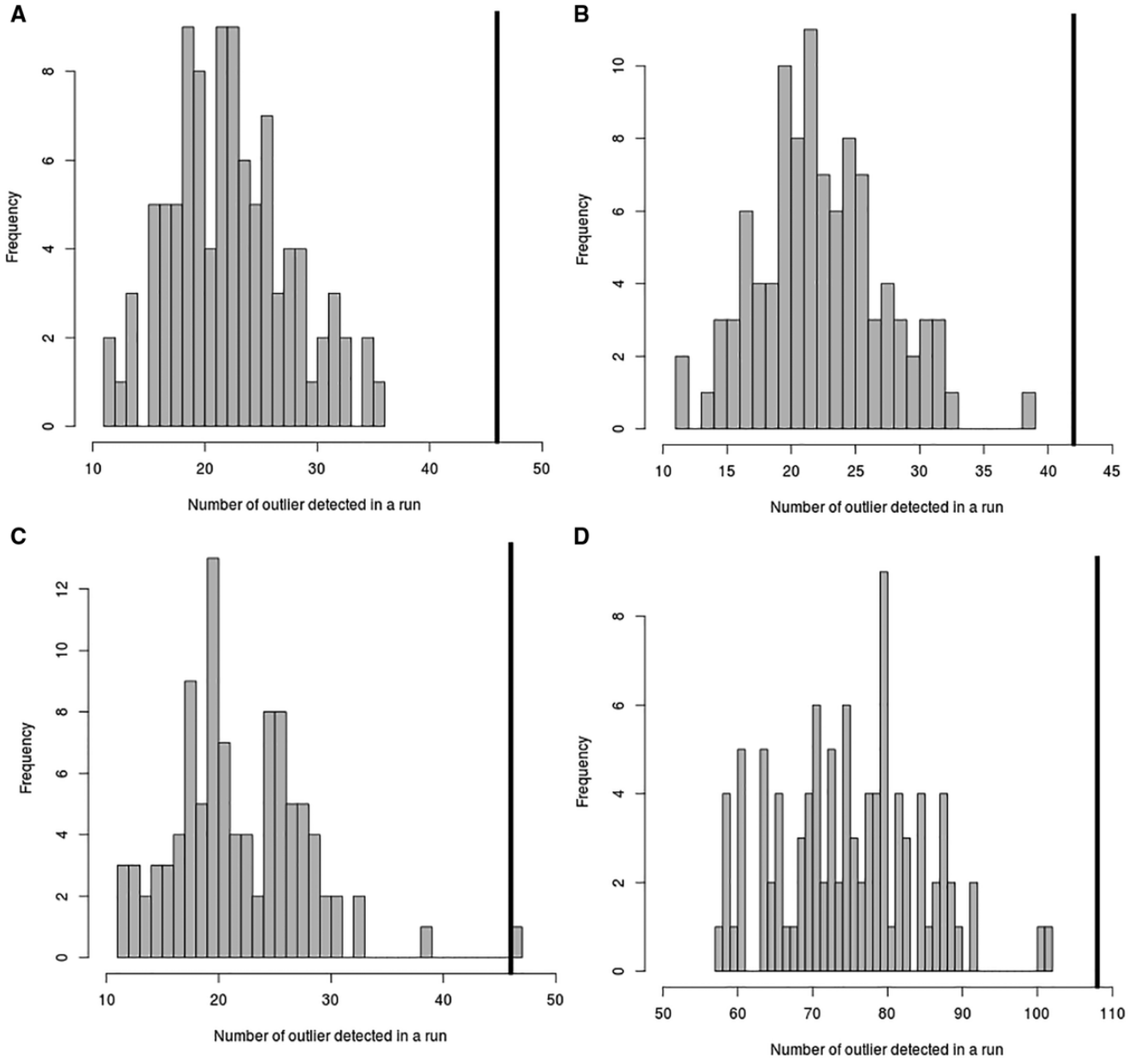

Fig. 3.—Histograms of the number of outliers detected in 100 simulations for the different environmental variables; latitude (*A*), longitude (*B*), temperature (*C*), and year (*D*). The *x* axis represents the number of outliers detected in a run. The *y* axis represents the number of simulations that detected this number of outlier. The vertical black bar represents the average number of outlier detected in the runs made with the real data set. 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.

($R^2 = 0.65$–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.

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 least one nearby gene enabling the identification of 241 genes associated with an outlier (supplementary table 3, Supplementary Material online). 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

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

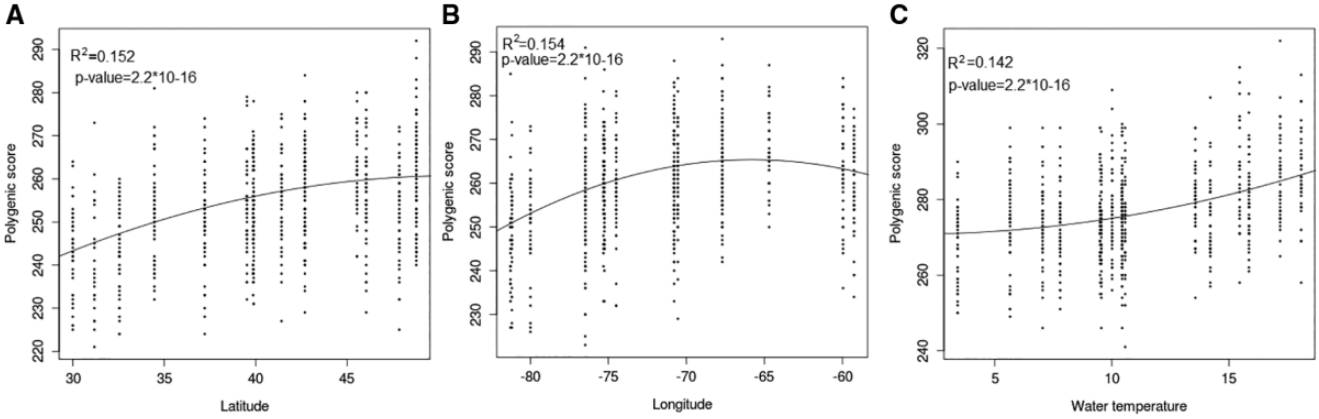

Fig. 4.—Correlations between additive individual polygenic scores based on all 259 outlier markers and each of three explanatory variables: Latitude (*A*), longitude (*B*), and temperature (*C*). Additive polygenic scores were obtained at the individual level by summing the number of locally favorable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

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

	Linear Model		Quadratic Model	
Variable	**10th Centile**	**90th Centile**	**10th Centile**	**90th 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	$R^2 = 0.87$ $P = 1 \times 10^{-7}$	$R^2 = 0.65$ $P = 0.00017$	$R^2 = 0.9$ $P = 3.43 \times 10^{-7}$	$R^2 = 0.65$ $P = 0.001$

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 data set.

No GO 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 data set 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 nonmodel 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 ~1.5% of the genome, which is higher but still close to previous comparable estimates (Pavey et al. 2015; Laporte et al.

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

2016; Pujolar et al. 2014). 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 signaling 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 data set 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. (2013) found yearly variation in the effective number of breeders (Nb) 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 >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 100 kb). 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.

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

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 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 favorable alleles displayed by individuals affects their chance of survival in a given location. Compared with 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 covariation with environmental and biotic parameters, which is consistent with the molecular footprints of selection along latitudinal gradients observed in many species, such as *Drosphila 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 with 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 analyzed 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

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

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 signaling 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 productions 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.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

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. 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].

Literature Cited

Aarnegard ME. 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. TIG 31(8):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. Mol Ecol. 15(7):1903–1916.

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Res. 25(22):4692–4693.

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. Proc Natl Acad Sci U S A. 83(12):4350–4354.

Barrett RDH, Hoekstra HE. 2011. Molecular spandrel: test of adaption at the genetic level. Nat Rev Genet. 12(11):767–780.

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. Nat Commun. 6:8705.

Benestan 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*). Mol Ecol. 24(13):3299–3315.

Bevan P. 2001. Insulin signaling. J Cell Sci. 114(Pt 8):1429–1430.

Blair LM, Granka JM, Feldman MW. 2014. On the stability of the Bayenv method in assessing human SNP-environment associations. Hum Genomics 8(1):1.

Bonhommeau S, Castonguay M, Rivot E, Sabatié R, Le Pape O. 2010. The duration of migration of Atlantic *Anguilla* larvae. Fish Fish. 11(3):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. Mol Ecol. 23(18):4444–4457.

Cairns DK, Secor DA, Morrison WE, Halett JA. 2009. Salinity-linked growth in anguillid eels and the paradox of temperate-zone catadromy. J Fish Biol. 74(9):2094–2114.

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

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. Mol Ecol. 22(11):3124–3140.

Cheng C, et al. 2012. Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. Genetics 190(4):1417–1432.

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

Cho HM, Kim JY, Kim H, Sun W. 2014. Phosphatase and actine regulator 4 is associated with intermediate filament in adult neural stem cells and their progenitor astrocytes. Histochem Cell Biol. 142(4):411–419.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185(4):1411–1425.

Côté CL, et al. 2013. Population genetics of the American Eel (*Anguilla rostrata*): Fst=0 and North Atlantic oscillation effects on demographic fluctuations of a panmictic species. Mol Ecol. 22(7):1763–1776.

Debieu M, et al. 2013. Co-variation between seed dormancy, growth rate and flowering time changes with latitude in *Arabidopsis thaliana*. PLoS ONE. 8(5):e61075.

De Jong G, Bochdanovitz Z. 2003. Latitudinal clines in *Drosophila melanogaster:* body size, allozyme frequencies inversion frequencies, and the insulin-signalling pathway. J Genet. 82(3):207–223.

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? Mol Ecol. 23(8):2006–2019.

Frichot E, Schoville SD, Bouchard G, François O. 2013. Testing for association between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 30(7):1687–1699.

Gagnaire PA, Albert V, Jónsson B, Bernatchez L. 2009. Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. Mol Ecol. 18(8):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(2):725–736.

Gagnaire PA, Gaggiotti OE. 2016. Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. Curr Zool. 62:603–616.

Gompert Z, et al. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Mol Ecol. 23(18):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. Spat Stat. 8:145–155.

Gunther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. Genetics 195(1):205–220.

Ha CM, et al. 2015. SNX14 is a bifunctional negative regulator for neuronal 5-HT6 receptor signalling. J Cell Sci. 128(9):1848–1861.

Häcker H, Karin M. 2006. Regulation and function of IKK and IKK-related kinases. Sci STKE. 2006(357):re13.

Hancock AM, et al. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome. Science 334(6052):83–86.

Hedgecock D, Pudovkin AI. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bull Mar Sci. 87(4):971–1002.

Hoban S, et al. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat. 188(4):379–397.

Huang DW, et al. 2007. The DAVID gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8(9):R183.

Hurrell JW, Kushnir Y, Visbeck M. 2001. Climate – the north Atlantic oscillation. Science 291(5504):603–605.

Jacobsen MW, et al. 2017. Assessing pre- and post-zygotic barriers between North Atlantic eels (*Anguilla anguilla* and *Anguilla rostrata*). Heredity 118(3):266–275.

Jeffreys H. (1961) Theory of probability. Oxford: Oxford University Press. 470 pp.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecol Lett. 7(12):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(3):624–637.

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. Hum Mol Gen. 24(14):4006–4023.

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? Ecol Evol. 2(5):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. Mol Ecol. 25(1):219–237.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol Ecol. 21(7):1548–1566.

Levene H. 1953. Genetic equilibrium when more than one ecological niche is available. Am Nat. 87(836):331–333. 10.1086/281792)

Li H, Durbin R. 2009. Fast and accurate short-read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760.

Li H, et al. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25(16):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(21):2987–2993.

Maes GE, Pujolar JM, Hellemans B, Volckaert FAM. 2006. Evidence for isolation by time in the European eel (*Anguila anguilla* L.). Mol Ecol. 15(8):2095–2107.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17(1):10–12.

McCleave JD, Kleckner RC. 1982. Selective tidal stream transport in the estuarine migration of glass eels of the American Eel (*Anguilla rostrata*). ICES J Mar Sci. 40(3):262–271.

Oksanen J, et al. 2017. vegan: Community Ecology Package. R package version 2.3-2. Available from: http://CRAN.R-project.org/package=vegan, last accessed November 6, 2017.

Palm DC, Rohwer JM, Hofmeyr JHS. 2013. Regulation of glycogen synthase from mammalian skeletal muscle – a unifying view of allosteric and covalent regulation. FEBS J. 280(1):2–27.

Pavey SA, et al. 2015. RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American Eel. Curr Biol. 25(12):1666–1671.

Pavey SA, et al. 2017. Draft genome of the American eel (*Anguilla rostrata*). Mol Ecol Res. 17(4):806–811.

Pritchard JK, Di Rienzo A. 2010. Adaptation-not by sweeps alone. Nat Rev Genet. 11(10):665–667.

Pujolar JM, Maes GE, Volckaert FAM. 2006. Genetic patchiness among recruits in the European eel *Anguilla anguilla*. Mar Ecol Prog Ser. 307:209–217.

Pujolar JM, et al. 2014. Genome-wide single generation signatures of local selection in the panmictic European eel. Mol Ecol. 23(10):2514–2528.

Pujolar JM, et al. 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: Springer. 493 pp.

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. Mol Ecol. 25(1):79–88.

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019

Ulrik MG, et al. 2014. Do North Atlantic eels show parallel patterns of spatially varying selection? BMC Evol Biol. 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. Rev Fish Biol Fish. 20(2):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. Mar Ecol Prog Ser. 168:35–43.

Wellenreuther M, Hansson B. 2016. Detecting polygenic evolution: problems, pitfalls and promises. TIG 32(3):155–164.

Williams GC, Koehn RK, Mitton JB. 1973. Genetic differentiation without isolation in the American Eel (*Anguilla rostrata*). Evolution 27(2):192–204.

Wirth T, Bernatchez L. 2003. Decline of north Atlantic eels: a fatal synergy? Proc R Soc Lond B. 270(1516):681–688.

Yeaman S, Otto S. 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution 65(7):2123–2129.

Yeaman S. 2015. Local adaptation by alleles of small effect. Am Nat. 186(S1):S74–S89.

Associate editor: Charles Baer

Downloaded from https://academic.oup.com/gbe/article-abstract/9/11/2974/4616584 by biuupm Bibliotheque Universitaire UPM user on 05 December 2019