Marine Genomics 62 (2022) 100933

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Comparative genomics and signatures of selection in North Atlantic eels

Jose Martin Pujolar [a], Magnus Wulff Jacobsen [b], Francesca Bertolini [c,*]

[a] Centre for Gelatinous Plankton Ecology and Evolution, National Institute of Aquatic Resources, Technical University of Denmark, Kongens Lyngby, Denmark
[b] Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Silkeborg, Denmark
[c] Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Kongens Lyngby, Denmark

ARTICLE INFO

Keywords:
Anguilla anguilla
Anguilla rostrata
Selection
F_{ST}
XP-EHH
iHS

ABSTRACT

Comparative genomic approaches can identify putative private and shared signatures of selection. We performed a comparative genomic study of North Atlantic eels, European eel (*Anguilla Anguilla*) and American eel (*A. rostrata*). The two sister species are nearly undistinguishable at the phenotypic level and despite a wide non-overlapping continental distribution, they spawn in partial sympatry in the Sargasso Sea. Taking advantage of the newly assembled and annotated genome, we used genome wide RAD sequencing data of 359 individuals retrieved from Sequence Nucleotide Archive and state-of-the-art statistic tests to identify putative genomic signatures of selection in North Atlantic eels. First, using the F_{ST} and XP-EHH methods, we detected apparent islands of divergence on a total of 7 chromosomes, particularly on chromosomes 6 and 10. Gene ontology analyses suggested candidate genes mainly related to energy production, development and regulation, which could reflect strong selection on traits related to eel migration and larval duration time. Gene effect prediction using SNPeff showed a high number of SNPs in noncoding regions, pointing to a possible regulatory role. Second, using the iHS method we detected shared regions under selection on a total of 11 chromosomes. Several hypotheses might account for the detection of shared islands of selection in North Atlantic eels, including parallel evolution due to adaptation to similar environments and introgression. Future comparative genomic studies will be needed to further clarify the causes and consequences of introgression, including the directionality of these introgression events.

1. Introduction

Natural selection plays a key role in shaping the genome of living organisms. Favorable mutations that increase the chance of individuals to survive and reproduce will increase in frequency over generations through positive selection, which may lead to allelic fixation over time. The spread and fixation of adaptive mutations result in a reduction of genetic variability in the region nearby the favorable allele due to the effect of genetic hitchhiking (Storz, 2005). In extreme cases, positive selection may eventually lead to speciation through the building of intrinsic or extrinsic reproductive barriers to gene flow (Nosil et al., 2005; Schluter, 2009; Nosil and Feder, 2013; Seehausen et al., 2014). Selective sweeps are defined as a reduction, elimination or change of genetic variation in genomic regions that are adjacent to causative variants that may occur in response to selective pressure within one or more populations (Kreitman, 2000). In contrast, "islands of genomic divergence" are regions where differentiation among groups is the highest and sometimes may also represent "islands of speciation" (Via and West, 2008; Marques et al., 2016). The appearance of these regions may vary depending on the number of copies of mutations or haplotypes that are beneficial and pass through generations. For this, selective sweeps have been classified as hard sweeps, when a single haplotype harboring a selectively advantageous allele rises in frequency (Smith and Haigh, 1974), or soft sweeps, when beneficial mutations exist on multiple haplotypes that rise in frequency simultaneously through positive selection (Hermisson and Pennings, 2005). In closely related species that share similar environments, signatures of selection may be shared, indicating that those regions may be relevant to environmental adaptations of both species and be the basis for phenotypic similarities (Kim et al., 2016).

With the advent of next generation sequencing methods, the ability of detecting selection at the genome-wide level has made a major breakthrough in both model and non-model organisms (Allendorf et al., 2010). In recent years, the number of high-quality genomes available has doubled, also for teleost fish (Rhie et al., 2021). Concurrently, a series of statistic tests have been adapted to identify signatures of

* Corresponding author.
E-mail address: franb@aqua.dtu.dk (F. Bertolini).

https://doi.org/10.1016/j.margen.2022.100933
Received 11 November 2021; Received in revised form 4 February 2022; Accepted 4 February 2022
Available online 16 February 2022

selection, including methods based on higher population differentiation than under neutral expectation (F_{ST} value), shifts in allele frequency spectrum (Tajima's D and Fay and Wu's H) and measures of linkage disequilibrium (iHS, XP-EHH), among others (see Ma et al., 2015).

North Atlantic eels, the European eel (*A. anguilla*) and the American eel (*A. rostrata*), provide an excellent opportunity to test for selection at homologous loci within and between species. They are sister-species (Minegishi et al., 2005) and are morphologically nearly undistinguishable and only differ in vertebral counts (Boëtius, 1980). The two species show a wide distribution range, with the European eel occurring in the eastern Atlantic from Morocco to Iceland including the Mediterranean Sea, while the American eel occurs in the western Atlantic from the Caribbean to Greenland (Tesch, 2003). Within species, molecular studies have demonstrated convincingly that both species are panmictic (Als et al., 2011; Côté et al., 2013; Pujolar et al., 2014a, Pujolar et al., 2014b; Enbody et al., 2021). North Atlantic eels are catadromous species and spawn in partial sympatry in the Sargasso Sea (Miller et al., 2015). After spawning, larvae are transported by the Gulf Stream to the shores of Europe and North America, respectively. Upon reaching the continental shelf, larvae metamorphose into glass eels that complete the migration into fresh, brackish and coastal waters as yellow eels. After a highly variable period of feeding (up to 30 years), they metamorphose into partially mature silver eels that migrate back to the spawning ground in the Sargasso Sea, where they reproduce once and die (Schmidt, 1923; Miller et al., 2019). Using mitogenome data, divergence time between the two species was estimated to ca. 3.38 Mya, possibly linked to changes in ocean currents (Jacobsen et al., 2014a). However, a recent study based on the joint allele frequency spectrum (JAFS) and Pairwise Sequentially Markovian Coalescent (PSMC) plots suggests a more recent divergence at between 1.3 and 2.4 Mya (Nikolic et al., 2020). Remarkably, only a moderate level of genetic differentiation is observed between species using nuclear genomic SNP markers (Jacobsen et al., 2014b) with comparable levels measured using AFLP and microsatellites (Mank and Avise, 2003; Wirth and Bernatchez, 2003; Gagnaire et al., 2009; Als et al., 2011). This low genetic differentiation is a likely result of gene flow between the species, which is supported by the observation of hybrids in larvae, glass eels and yellow eels (Gagnaire et al., 2009; Als et al., 2011; Pujolar et al., 2014a, 2014b; Jacobsen et al., 2017).

Previous genomic studies searching for signatures of selection within European eel were hampered by the lack of a high-quality reference genome and annotation. While islands of selection could not be properly tested, several genes related to metabolism, circadian rhythms, growth and development and defense responses were detected to be under selection (Pujolar et al., 2014a, 2015; Ulrik et al., 2014). When testing for positive selection between European and American eel, Jacobsen et al. (2014b), found that candidate SNPs were located within genes related to development and phosphorylation, consistent with the hypothesis that larval phase duration and migration loops play a key role in eel speciation. However, due to the assembly level of the eel genome at the time, composed by a large number of contigs and scaffolds (Henkel et al., 2012), it was not possible to assess if SNPs putatively under selection clustered in regions or islands. The recent availability of a high-quality reference genome of the European eel, assembled at chromosome level and including 30,761 annotated genes (Ncbi, GCA_013347855.1; Rhie et al., 2021) allows for a higher resolution investigation of signatures of selection and differentiations within and between North Atlantic eels. The newly assembled genome has been used to search for genetic footprints of differentiation within European eels using whole genome resequencing data (Enbody et al., 2021). In detail, the study compared Baltic vs. Mid-Atlantic samples in European eel and reported only a small region under selection located on chromosome 1 (covering roughly 6 kb around 81.2 Mbp) and two other regions on chromosomes 13 and 15. In agreement with previous studies showing that signatures of selection within the panmictic European eel were restricted to a single-generation and not heritable, the authors suggested that adaptation and survival of glass eels during the post-larval stages is based on phenotypic plasticity.

Taking advantage of the new assembled and annotated genome, the aim of the present study is to use a comparative genomics approach to search for private and shared signatures of selection between American and European eel. We used RAD sequencing data and applied three statistical approaches to identify genomic regions putatively under selection: (1) F_{ST}, (2) XP-EHH and (3) iHS methods. The location and effects of coding genetic variants on genes were predicted using SNPeff. Finally, the genes under selection were associated with biological functions by performing enrichment analysis. We were particularly interested in characterizing the signatures of selection shared by the two species, which will contribute to a better understanding of the adaptative differences between North Atlantic eels.

2. Material and methods

2.1. Data retrievement, alignment and variant calling

A total of 359 samples retrieved from the Sequence Nucleotide Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) and summarized in Table 1 were included in the analyses (Pujolar et al., 2014a; Jacobsen et al., 2014b; Pavey et al., 2015). All samples were obtained through RAD-sequencing and details for each of the samples are available in Table S1. Samples analyzed included 254 samples corresponding to European eel *A. anguilla* (AA) and 105 corresponding to American eel *A. rostrata* (AR). European eel individuals were collected at five geographical regions, two in the Mediterranean Sea, Valencia (VAL, Spain, $N = 52$) and Canet (CAN, France, $N = 18$), and three in the Northeast Atlantic, Ringhals (SWE, Sweden, $N = 21$), Gironde (GIR, France, $N = 115$) and Burrishoole/Lough Erne (IRE, Ireland, $N = 48$). Most American eel individuals were collected in Quebec ($N = 202$), while three individuals were collected in Nova Scotia. European eel individuals included glass eels ($N = 128$), silver eels ($N = 39$) and yellow eels ($N = 87$). Most American eel individuals were yellow eels ($N = 89$) with few glass eels ($N = 8$) and silver eels ($N = 8$).

Read quality control was initially assessed for each individual using FastQC (Andrews, 2015) and processed with Trimmomatic v0.36 (Bolger et al., 2014). Trimming was performed in order to remove Illumina adapters, low quality read ends and retaining only reads longer than 30 bases after trimming (ILLUMINACLIP:TruSeq2-PE.fa:2:30:10, HEADCROP:6, SLIDINGWINDOW:4:15, MINLEN:30). Filtered reads, both from European eel and from American eel were then mapped to the latest version available of the European eel reference genome fAngAng1.pri (GCF_013347855.1; https://www.ncbi.nlm.nih.gov/) using the mem algorithm in BWA v0.7.12 (Li and Durbin, 2009) with default parameters. Filtering of the mapped reads was performed with SAMtools v1.10

Table 1
Number of samples of American eel *Anguilla rostrata* (AR) and European eel *Anguilla Anguilla* (AA) included in the analyses, divided by developmental stage (i.e. Glass eel, Yellow eel, silver eel). For European eels, acronyms for the collected regions are the following: Valencia (Spain) = VAL, Canet (France) = CAN, Ringhals (Sweden) = SWE, Gironde (France) = GIR, and Burrishoole/Lough Erne (Ireland) = IRE.

AA	Glass eel	Yellow eel	Silver eel	Total
GIR	28	87		115
CAN	18			18
IRE	39		9	48
VAL	22		30	52
SWE	21			21
Total AA	**128**	**87**	**39**	**254**
AR				
Nova Scotia	3			3
Quebec	5	89	8	102
Total AR	**8**	**89**	**8**	**105**
Total (AA + AR)	**136**	**176**	**47**	**359**

(Li et al., 2009), retaining mapped reads with quality >20. Despite the species specificity of the genome and the fact that same alignment conditions were applied, percentage of high-quality mapped reads between the two species were similar, with 84.82% ± 0.97 for the European eel and 84.4% ± 1.32 for the American eel (Table S1).

Variant calling was performed in parallel with the mpileup and call function in BCFtools v.1.10 (Li et al., 2019) using the -m method and excluding sites in the proximity of indels. The resulting VCF output was then filtered retaining only biallelic variants with SNP quality ≥900 and sites for which average depth was more than 8 and less than 100. A further filtering was performed with PLINK 1.9 (Purcell et al., 2007) removing SNPs with call rate < 95% and minimum allele frequency MAF < 0.01 for a total of 655,497 retained SNPs. All individuals had a genotype rate > 80%.

2.2. Genetic diversity and population structure

Prior to all calculations, linkage disequilibrium-based variant pruning of all VCF files was conducted in PLINK v1.9 (Purcell et al., 2007) in order to obtain a pruned subset of markers in approximate linkage equilibrium with each other. The indep-pairwise option was used with a window size in variant count of 100, a variant count of 5 to switch window at the end of each step and a r^2 threshold of 0.5, retaining 595,846 SNPs.

Standardized genetic differentiation (F_{ST}) statistics between all European and American eel samples were calculated using VCFtools v0.1.14 (Danecek et al., 2011) following Weir and Cockerham (1984). F_{ST} values between life stages (glass eels, yellow eels and silver eels) and between population pairs were also calculated. In order to assess the significance of pairwise F_{ST} values, we carried out bootstrapping over loci to generate a confidence interval around the observed F_{ST}. *P*-values were calculated using a one-sample *t*-test for each pairwise F_{ST}.Population structure was initially explored with two-dimension reduction routines: Principal components analysis (PCA) based on the variance-standardized relationship matrix using the program smartPCA from the Eigensoft package (Patterson et al., 2006) and multidimensional scaling (MDS) based on raw Hamming distances using PLINK v1.9 (Purcell et al., 2007).

Population admixture was investigated with ADMIXTURE software (Alexander and Lange, 2011). Following the software guidelines, the cross-validation (CV) procedure was applied to choose the best *K* for the model, where *K* is the number of (sub) populations that was assumed for the analysis. Population structure on European eel was further investigated using the Bayesian assignment approach implemented in STRUCTURE (Pritchard et al., 2000), a model-based clustering algorithm that infers the most likely number of groups (K) in the data. The analysis was performed only considering glass eels ($N = 128$) to reduce computational time, with different K values, assuming an admixture model, correlated allele frequencies and without population priors. A burn-in of 10,000 steps followed by 100,000 additional Markov Chain Monte Carlo iterations were performed. For each K, 10 independent runs were conducted to check the consistency of results. The most likely K was inferred using the method of Evanno et al. (2005), which measures the steepest increase of the ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values.

2.3. Test for putative selection

Since more than one approach is usually required to capture the signatures of selection in the genome (Ma et al., 2015), three methods using the unpruned dataset were applied to detect selection signatures: one F_{ST}-based method and two haplotype-based methods (iHS and XP-EHH).

2.3.1. Fixation index (F_{ST})

The first method used in this study is based on differences in allele frequencies between populations or species by estimating the Fixation index (F_{ST}). Genetic differentiation between European and American eel was estimated by calculating sliding window F_{ST} using VCFtools v0.1.14 (Danecek et al., 2011) with the unpruned dataset. A 100-kb window was applied as recommended by Hohenlohe et al. (2010) for RAD sequencing data. Windows with window-F_{ST} values above the 99th percentile of the empirical distribution and with more than 10 SNPs were retained for further analyses. Regions above the 98th percentile were considered only when overlapping with the other type of analyses performed. Sliding window F_{ST} along chromosomes was plotted in R.

2.3.2. Integrated haplotype homozygosity score (iHS) and cross population extended haplotype homozygosity (XP-EHH)

The two haplotype methods used in this study are based on extended haplotype homozygosity (EHH). The intra-population standardized integrated haplotype score (iHS) statistic compares the integrated EHH profiles between two alleles at a given SNP in the same population (Voight et al., 2006), while the XP-EHH (Cross Population Extended Haplotype Homozygosity) statistic compares the integrated EHH profiles between two populations at the same SNP (Sabeti et al., 2007).

Phasing of the animals was performed with SHAPEIT v2 (O'Connell et al., 2014) and iHS and XP-EHH were calculated with the REHH package (Gautier and Vitalis, 2012). iHS was performed on European eel and American eel separately, while XP-EHH was performed between American and European eels. As ancestral alleles are unknown, for both iHS and XP-EHH analyses the absolute value of each SNP was averaged in 50% overlapping windows of 100Kb each. Then, only windows with more than 10 SNPs were retained and windows with values above the 99th percentile of the empirical distribution were considered for further analyses.

2.4. Candidate gene annotation and functional analyses

Gene detection was performed (i) on concordant overlapping regions between the two methods comparing AA vs AR (namely, F_{ST} and XP-EHH) and (ii) on shared regions within AA and AR detected by the iHS method. Gene predictions for the European eel genome available at the NCBI website were used to establish the genomic position of the candidate SNPs for local selection. SNPs were considered for further analyses when found in exonic or intronic regions or within 10 kilobases of a gene, which was based on the estimated decay of linkage disequilibrium in European eel (Hemmer-Hansen et al., 2014).

Gene descriptions were obtained from the Zebrafish Information Network (ZFIN) database, together with human orthologs, which were used in the GeneCards human gene database to retrieve the Entrez gene summary, the UniProtKB/Swiss-Prot summary as well as the Ensembl Gene IDs.

Functional interpretation of the set of candidate genes was obtained using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway approach implemented in DAVID v6.8 (Dennis Jr et al., 2003). Zebrafish was used as reference genome for annotation. Standard settings of gene count and ease in DAVID were used.

Functional Gene Onthology terms of the candidate genes were also analyzed with the ClueGo plug-in of Cytoscape v3.8.2 (Bindea et al., 2009), which visualizes the non-redundant biological terms for large clusters of genes in a functionally grouped network. Analyses were conducted using the Ensembl gene IDs, only considering GO terms with corrected *p*-values <0.05 and selecting the major significant GO terms as the representation of the group.

Functional effects of each SNP candidate to be under selection on the genome was predicted using SNPEff v4.3 (Cingolani et al., 2012). The proportion of synonymous and nonsynonymous changes found in the regions putative under selection were compared to the proportion found in the whole genome using a Chi-square Test.

3. Results

3.1. Population structure analysis

Despite the species specificity of the genome and the fact that same alignment conditions were applied, percentage of high-quality mapped reads between the two species were similar, with 84.82% ± 0.97 for the European eel and 84.4% ± 1.32 for the American eel (Table S1).

Genetic differentiation between the two North Atlantic eel species was moderately high, $F_{ST} = 0.087$. Within European eel, no differences were found between life stages (GE-YE, $F_{ST} = 0.0021$; GE-SE, $F_{ST} = 0.0002$; YE-SE, $F_{ST} = 0.0030$). Similarly, no differences were found between geographical locations within European eel, with F_{ST} values ranging from 0.0002 to 0.002 (Table 2). All F_{ST} values were statistically not significant.

A Multidimensional Scaling (MDS) including all individuals clearly separated the two species, European and American eel (Fig. 1a). A dedicated MDS plots for European and American eels showed no clustering of samples by geographic location, with all individuals appearing mixed in accordance with panmixia (Fig. 1b and c).

The admixture analysis suggested a cross-validation error = 2 for the full dataset, suggestive of the existence of two separate clusters for American eel and European eel respectively (Fig. S1a). The same comparison considering only European eel suggested a K = 1 scenario (data not shown). When conducting STRUCTURE analysis for the glass eels only (Fig. S1b), K = 2 was the most likely scenario; no differences in admixture proportions were found among locations (France-Atlantic: 8.18%, France-Mediterranean: 8.00%, Sweden: 8.48%, Spain: 8.78%, Ireland: 9.52%; Fig. S1b). The results are therefore in agreement with previous genetic studies showing that despite widespread geographical distributions, both AA and AR are panmictic (Côté et al., 2013; Pujolar et al., 2014a; Enbody et al., 2021).

3.2. Tests of selection between North Atlantic eels

The F_{ST} analyses comparing American vs European eels detected the presence of regions of high genetic differentiation (above the 99th percentile) on chromosomes 1, 3, 4, 5, 6, 7, 9, 10 and 18 (Fig. 2 and Table S2). The largest number of regions under selection were detected on chromosome 6, with a total of 11 regions between 30.1 and 37.6 Mbp, ranging from 100 Kb to 1.5 Mb (31.5–33.0 Mbp).

The XP-EHH analysis revealed genomic regions under selection on chromosomes 1 to 18, ranging from 0.1 Mb to 1.9 Mb (Fig. 3 and Table S3). A total of four regions larger than 1 Mb were detected on chromosome 6 (33.30–34.70 Mbp, 1,40 Mb), 9 (38.85–39.90 Mbp, 1.05 Mb), 10 (28.80–30.30 Mbp, 1.50 Mb) and 18 (28.30–30.25 Mbp, 1,95 Mb).

Overlapping regions between the two methods were detected on chromosomes 2, 4, 5, 6, 7, 9 and 10. The largest shared regions were a 3.65 Mbp region on chromosome 6 (from 31.45 to 35.10 Mbp) and a 950 kb region on chromosome 10 (28.80–29.75 Mbp) (Fig. 4). When investigating the genomic position of the regions putatively under selection (shared between the F_{ST} and XP-EHH methods), a hit was obtained for a total of 64 genes (Table 3). Out of all genes with a hit, 38 genes were fully characterized in zebrafish, while 35 orthologs were found in humans. Functional annotation analysis with DAVID showed two enriched clusters, namely ATP-binding and transmembrane (data not shown). ClueGo analysis identified the following enriched networks: SRP-dependent co-transitional protein targeting to transmembrane, farnesyltransferase activity, swim bladder development, nucleotide-

Table 2
F_{ST} output for the genetic differentiation between geographical locations of the European eel samples. Valencia (Spain) = VAL, Canet (France) = CAN, Ringhals (Sweden) = SWE, Gironde (France) = GIR, and Burrishoole/Lough Erne (Ireland) = IRE.

	GIR	CAN	VAL	IRE	SWE
GIR	*				
CAN	0.00000157	*			
VAL	0.0000206	0.00003	*		
IRE	0.0000138	0.0000518	0.0000202	*	
SWE	0.000246	0.00024666	0.0002463	0.00024652	*

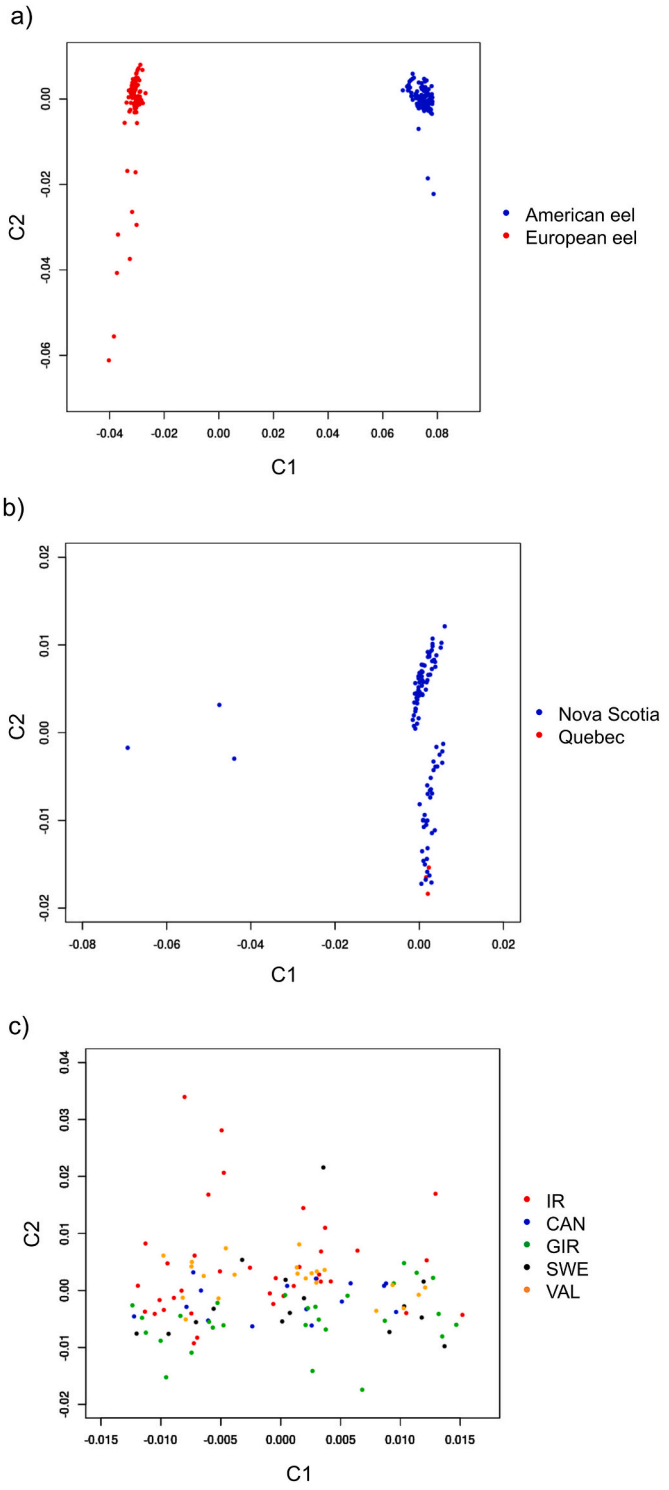

Fig. 1. Visualization of population structure through Multidimensional Scaling (MDS) including (a) clusterization of all European eel and American eel individuals b) American eels (b) European glass eels, where Valencia (Spain) = VAL, Canet (France) = CAN, Ringhals (Sweden) = SWE, Gironde (France) = GIR, and Burrishoole/Lough Erne (Ireland) = IRE.

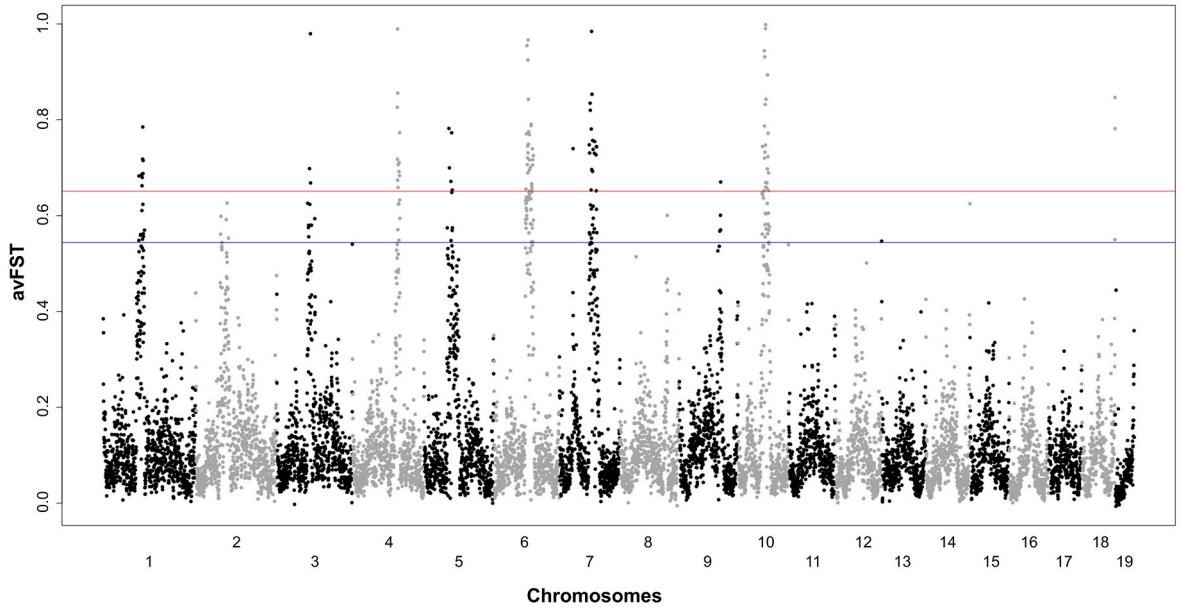

Fig. 2. Sliding-window FST (avFst) between European and American eels. 98th percentile (blue) and 99th percentile (red) thresholds are reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

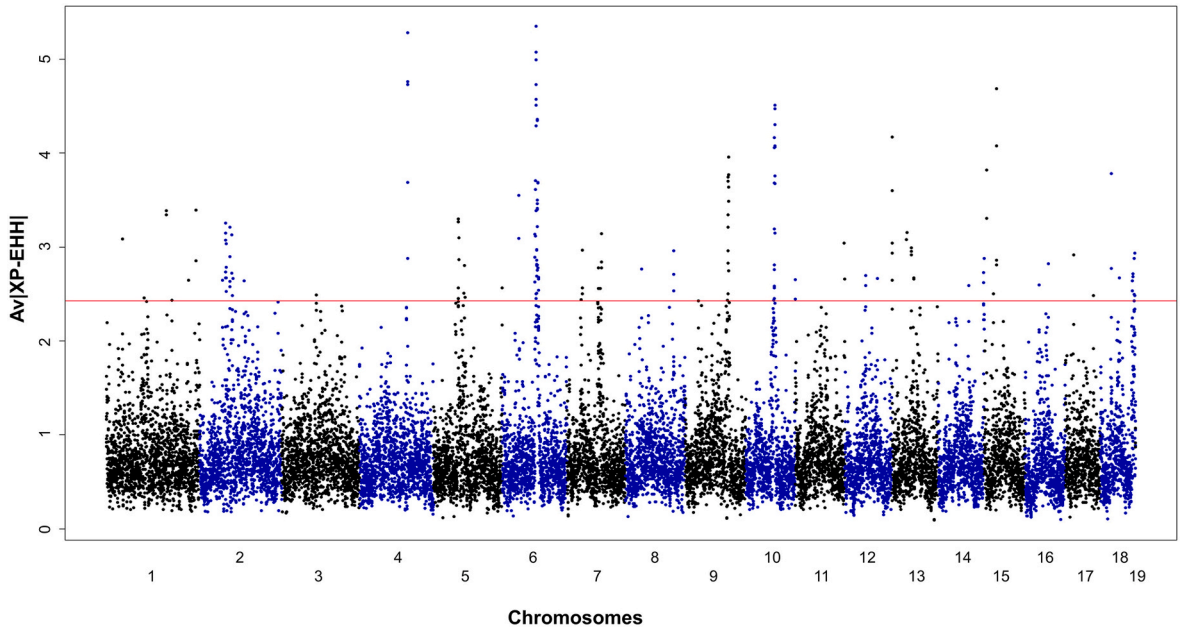

Fig. 3. Window based |XP-EHH| (av|XP-EHH|) between European and American eel. 99th percentile threshold (red) is reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

excision repair, negative regulation of osteoblast differentiation, DNA replication initiation and phosphorylase kinase regulator activity, among others (Fig. 5a).

Based on the SNPeff results, 98.5% of the SNPs within these shared regions were non-coding variants, including variant hits on introns (66.6%), and intergenic regions (25.6%) (Table S4). Only 1.5% of SNPs were in exons. Of those, none had a high (disruptive) effect, 2 (0.3%) had a moderate effect (all were missense variants causing a codon that produced a different amino acid) and 9 (1.2%) had a low effect (all were synonymous variants causing a codon that produces the same amino-acid). The two non-synonymous variants were found in gene extl3 on chromosome 6 and gene rbm19 on chromosome 10. All the synonymous variants were also found on chromosome 6 (on genes LOC118229874, efhc1, micu2, ahctf1, extl3 and kif13bb) and chromosome 10 (on genes rbm19 and LOC118207110). The proportion of nonsynonymous vs. synonymous substitutions was comparable to the proportion found in the whole genome (p-value = 0.2637, χ^2, df = 2; Table S5).

3.3. Tests of selection within species

Shared signatures of selection within European eel and within American eel were detected using iHS at a total of 11 chromosomes (Table 4 and Fig. S2). Most regions ranged in size between 100 and 200 Kb, except 2 regions of 250 Kb each and a region of 300 Kb on chromosome 2 and the two largest regions detected: a 800 kb region on chromosome 8 (from position 52.35 to 52.75 Mbp) and a 850 Kb region

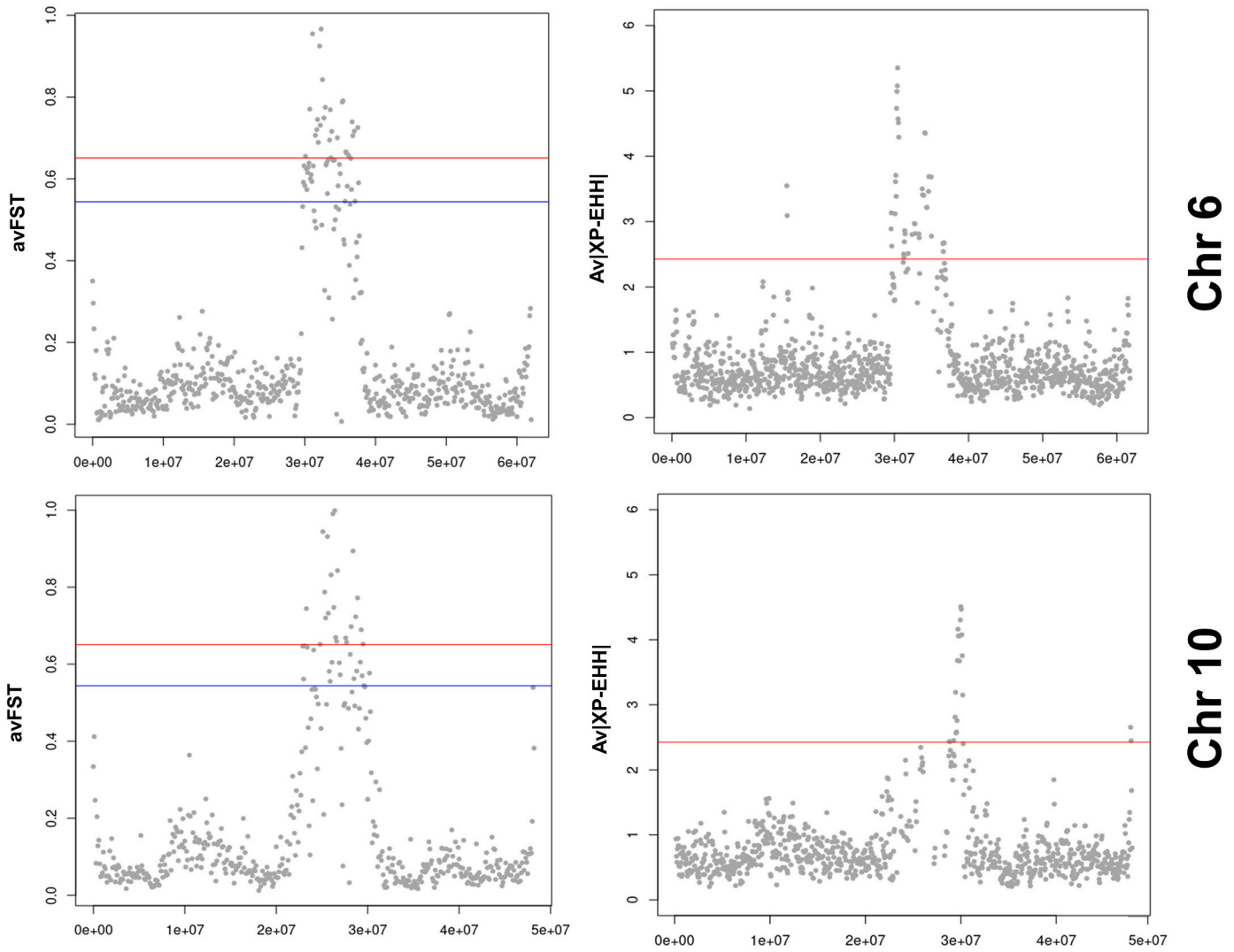

Fig. 4. Regions putatively under selection identified by the F_{ST} approach (left panel) and the XP-EHH approach (right panel) for chromosome 6 (above) and chromosome 10 (below). 98th percentile (blue) and 99th percentile (red) thresholds are reported. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on chromosome 16 (from position 2.45 to 3.30 Mbp).

As the analyses were conducted separately, SNP trend (i.e., identification of minor and major allele) based on the allele frequency of the common window for both American and European eel was considered to account for possible divergences between the two species (e.g., same signals but direction of the allele frequencies not concordant within the same window). Here, the SNP trend was consistent between the two species, with an average of 81.5% ±12.0 of SNPs with the same shared minor allele, meaning that the putative selective sweeps are going in the same direction in the two Atlantic eel species. When investigating the regions putatively under selection identified by the iHS test, a hit was obtained for a total of 127 genes (Table 4), the majority of which were not characterized. A total of 48 genes were fully characterized in zebrafish, while 45 orthologs were found in humans. Functional annotation analysis in DAVID showed two enriched clusters, namely zinc binding and transmembrane (data not shown). ClueGo analysis identified the following enriched networks: negative regulation of cytokine production, proteasome core complex, D-aspartate-oxidase activity, among others (Fig. 5b).

SNPeff results showed that 95.7% of the SNPs were non-coding variants, including variant hits on introns (55.1%) and intergenic regions (35.7%) (Tables S4 and S5). Only 4.3% of SNPs were in exons. Regarding effect impact, none had a high (disruptive) effect but 42 (2.4%) had a moderate effect and all were missense variants causing a codon that produced a different amino acid. These affected a total of 14 different genes located in chromosomes 1 (LOC118233928, si:dkey-238d184), 2 (si:ch211–156,118.8, agpat5, sox9a), 8 (nagpa, LOC118232976, psmb13a, tap2a), 11 (LOC118207914, LOC118214297) and 15 (LOC118214031, chaf1b, parp9). A total of 34 variants (1.9%) had a low effect and all were synonymous variants causing a codon that produces the same amino acid, including 12 variants in which the change occurred within the region of the splice site. The proportion of nonsynonymous vs. synonymous substitutions was significantly different from the proportion found in the whole genome (p-value = 0.0001, χ^2, df = 2; Table S5). This difference was explained by a higher rate of nonsynonymous mutations (app. 55.3% vs 32.1%;), which translates into a higher nonsynonymous/synonymous substitutions (dN/dS) ratio in the regions found using the iHS approach.

4. Discussion

4.1. Comparison across selection methods

Taking into consideration the high historical effective population

Table 3
Regions under selection detected by the F_{ST}/XP-HH approach, including chromosome (chr), start and end position (in Kbp), total length (in Kb) and genes within the region.

Chr	Start (Kbp)	End (Kbp)	Length (Kb)	Genes
2	29.70	29.85	150	phkg2, ccdc189, LOC118219765, stx4
4	44.90	45.10	200	acot9.1, c8g, LOC118225150
5	26.75	26.90	150	gtf2h1, tmem17
				prkg3, LOC118229874, ctsba, fdft1,
				gata4, xkr5a, tram2, efhc1,
6	30.10	30.70	600	LOC118229955, mrpl19, itsn2b
				LOC118229832, micu2, LOC118229847,
				nrbp1, si:ch211-57i17.5, LOC118229842,
				LOC118229849, mcm3, ahctf1, pth2,
				LOC118229114, elp3, LOC118229398,
				extl3, LOC118229497, ints9, kif13bb,
6	31.45	35.10	3650	cpsf2, LOC118230524, riox1, hmbox1b
				LOC118230112, LOC118230180,
6	36.65	36.80	150	LOC118230784
7	34.70	34.80	100	Uvssa
				LOC118235584, ckma, nccrp1,
9	38.85	39.10	250	LOC118235605, kptn
				zgc:158398, LOC118206160,
				LOC118237347, morc2, rbm19, thoc5,
				LOC118237225, LOC118207110,
				LOC118207109, LOC118207111,
				LOC118206604, LOC118237415,
10	28.80	29.75	950	eif4enif1

size estimated for North Atlantic eels (ranging from 100,000 to >1 million individuals) based on point-based approaches (Pujolar et al., 2014a) and analyses of demographic history through time (Nikolic et al., 2020), random drift is expected to be negligible. Hence, the major evolutionary force determining allele frequency differences is expected to be natural selection (together with gene flow), in a pattern that is expected to vary from locus to locus. Accordingly, most regions in our study showed low genetic differentiation, while a small set of regions showed a significantly high genetic differentiation together with a reduction in haplotype diversity that is consistent with the action of natural selection.

The identification of many shared regions putatively under selection using the F_{ST} and the XP-EHH methods is not surprising, as both methods capture similar signals in the genome and may correspond to similar types of selection events. In this sense, the F_{ST} method uses the divergence of allele frequencies among populations and is particularly sensitive in identifying fixed alleles with complete differentiation ($F_{ST} = 1$) (Ma et al., 2015). The XP-EHH approach is also based on population differentiation and compares the amount of extended haplotype homozygosity at each locus between two populations. This method is best at revealing selection signatures when haplotypes are close to or fully fixed (Sabeti et al., 2007).

The two between-population methods F_{ST} and XP-EHH are best at detecting complete or nearly complete signatures of selection. Due to the long time required to reach fixation, F_{ST} and XP-EHH are expected to identify older selection events between populations, and in particular the F_{ST} method is best suited for detecting evidence of selection occurring in the more distant past (Cadzow et al., 2014). In this sense, both FST and XP-EHH are powerful tools to detect "hard selective sweeps", in which a new mutation arises and spreads quickly to fixation due to natural selection (Smith and Haigh, 1974). Other scenarios might be more difficult to detect, especially when selection leads to shifts in allele frequencies rather than fixation.

When comparing all three methods, no regions were shared between iHS and the two between-population methods F_{ST} and XP-EHH. This is expected since all methods capture different signatures in the genome. In the case of iHS, the approach is based on the frequency of extended haplotypes within a population and is known to be sensitive when the frequency of the allele under selection is intermediate. In the study of Ma et al. (2015) using simulations, iHS is the most powerful statistic to detect selection when the target allele has a frequency between 0.4 and 0.8 but has limited power when the target allele is fixed. This contrasts with the performance of F_{ST} and XP-EHH, which are most powerful at fixation. Therefore, the iHS approach is more appropriate in scenarios where selection leads to shifts in allele frequencies rather than fixation. Those scenarios include "soft selective sweeps", in which multiple haplotypes harboring advantageous mutations are all favoured (Hermisson and Pennings, 2005). Since the iHS test has higher statistical power when selected alleles are at intermediate frequencies that have not yet reached fixation, it can detect signals of very recent or ongoing selective sweeps (Voight et al., 2006). The approach has been used to detect recent signatures of selection associated with human domestication in cattle (Maiorano et al., 2018) or farmed Nile tilapia (Cádiz et al., 2020).

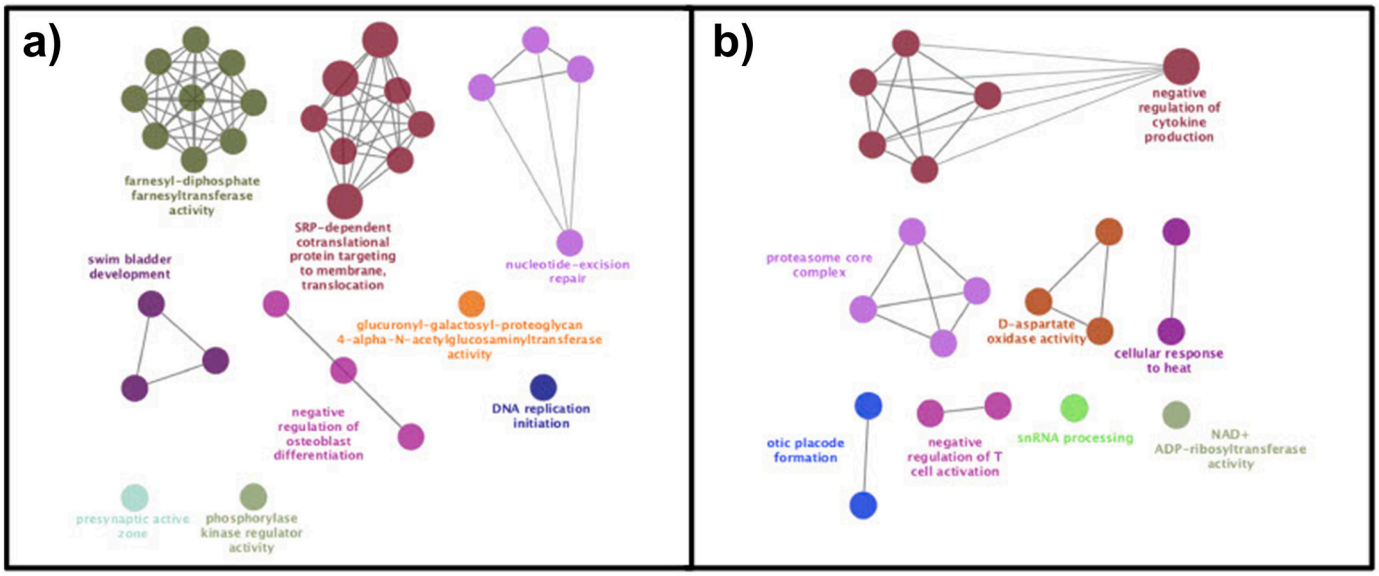

Fig. 5. ClueGo analyses of enriched pathways for the genes putatively under selection identified by (a) the F_{ST}/XP-EHH approach and (b) the iHS approach (b). Each node represent a gene; connection among nodes indicate clusterization for the enriched pathways indicated with the same colour.

Table 4
Shared regions under selection detected by the iHS approach, including chromosome (chr), start and end position (in Kbp), total length (in Kb), trend similarity (percentage of shared minor alleles) and genes within the region.

Chr	Start (Kbp)	End (Kbp)	Length (Kb)	% Trend similarity	Genes
1	31.55	31.70	150	89.04	ccdc25, LOC118233950, ddo, LOC118233928, zbtb24, LOC118234097, si:ch73-242 m19.1
1	36.20	36,35	150	76.19	–
1	68,30	68,45	150	80.36	LOC118215436, LOC118215476, LOC118215379, LOC118215359, LOC118229683, LOC118215459, LOC118215469, LOC118215263, LOC118215255
1	85,00	85,15	150	89.80	kmt2bb, psenen, hspb6, proser3, lin37, LOC118234383, hsc70, si:dkey-238d18.4, LOC118233131, LOC118235720, LOC118235737
2	28,40	28,65	250	61.18	si:ch211-156 l18.8, LOC118221812, mrpl45, LOC118220741, zbtb45
2	28,85	29,10	250	60.56	agpat5, lgsn, fbxo9, LOC118220569, col9a1a
2	30,05	30,35	300	70.67	srcap, zmp:0000001082, LOC118220353, LOC118219834
2	30,85	31,00	150	76.74	LOC118219834
2	73,30	73,40	100	90.74	sox9a
5	31,50	31,65	150	66.04	lgr4, heatr3, LOC118227994, LOC118227441, si:ch1073-291c23.2
7	30,85	31,00	150	76.92	LOC118232396, LOC118232397 LOC118232754, LOC118234710, LOC118232864, LOC118232902, LOC118233163, nagpa, ints8, LOC118234575,
8	34,40	34,55	150	95.10	LOC118233492 LOC118232960, LOC118232962, LOC118232976, lgals17, vps52, rps18, LOC118232970, slc39a7, LOC118232964, col11a2, hsd17b8, brd2a, psmb8a, psmb13a, psmb12, psmb9a, tap2a, LOC118232969, LOC118232966,
8	52,35	52,75	400	96.65	LOC118232971 LOC118235604, LOC118235593,
9	38,20	38,35	150	64.71	LOC118235594
10	22,65	22,85	200	86.84	LOC118206226 LOC118208037, ppil1, sars1, araf, ttll9, LOC118207914, LOC118208038,
11	36,45	36,60	150	89.61	LOC118207918 cwc22, LOC118214297, LOC118214294, LOC118214295,
15	0.30	0.45	150	90.38	LOC118214035 LOC118214663, LOC118214576, LOC118213829, LOC118213828,
15	3,05	3,20	150	88.71	LOC118213831
15	17,85	18,05	200	77.23	gja5a, LOC118214031, LOC118214530, LOC118214449, chaf1b, morc3a, LOC118214375, LOC118213770, LOC118213769, LOC118214651
15	22,35	22,45	100	81.03	cep70, faima, parp9, parp14rs1
16	1,25	1,40	150	100.00	LOC118215100 LOC118215105, LOC118215099, LOC118215112, LOC118215118, LOC118215114, LOC118215121,
16	2,45	3,30	850	97.83	LOC118215098 LOC118218367, LOC118217775, LOC118218368,
18	27,55	27,65	100	69.23	LOC118218369

4.2. Putative signatures of selection between North Atlantic eels

Footprints of selection found when comparing European eel vs American eel were apparent when interrogating the genome. Here, larger regions or islands were observed in a total of seven chromosomes, most recognizable in chromosomes 6 and 10, as suggested by the FST and XP-EHH methods. Genes containing candidate SNPs for positive selection showed significant enrichment for terms related mainly to ATP phosphorylation and development. When comparing the genes enriched in our study to Jacobsen et al. (2014b), which used an F_{ST} approach, but a low-assembly genome, a total of 25 genes were shared across studies, including kinases (phkg2, prkg3, ckma), mitochondrial-related and/or calcium activity genes (micu2, mrpl19, itsn2b), transcription/translation factors (ahctf1, cpsf2, eif4enif1, elp3, gtf2h1, nrbp1, riox1, thoc5, tram2) and genes relayed to cell growth and development (extl3, nccrp1, stx4).

This is in accordance with previous studies reporting selection on genes related to energetics, development and regulation (Gagnaire et al., 2012; Jacobsen et al., 2014a, 2014b, 2015, 2017) and points to selection in North Atlantic eels being driven by differences in larval development and duration time, as well as different energetic requirements due to the much longer migration loop in European eel (ca. 5000 km) in comparison with American eel (ca. 2000 km) (Tesch, 2003; Jacobsen et al., 2014b).

4.3. Shared signatures of selection in North Atlantic eels

We were also able to identify shared footprints of selection, namely regions of the genome that were identified under selection within both European eel and American eel, as suggested by the iHS method. Shared regions under selection were found in a total of 11 chromosomes and were generally small (150 kb on average with a maximum of 850 kb) in comparison with the regions found under selection using the F_{ST}/XP-EHH approach. The small dimensions of sweeps are in agreement with the length of the sweeps detected by Enbody et al. (2021) using whole genome resequencing. Overall, the identified regions showed a significant higher proportion of nonsynonymous mutations with moderate effect. Similar results have been observed in other studies and may relate to the effect of genetic hitchhiking during selective sweeps. Here mildly deleterious mutations may hitchhike along with the associated advantageous allele directly targeted by positive selection (Chun and Fay, 2009). This may lead to increased proportions of deleterious mutations as found in poplar (Zhang et al., 2016) and higher dN/dS as shown in

humans (Chun and Fay, 2009), dogs (Cruz et al., 2008; Marsden et al., 2016) and rice (Renaut and Rieseberg, 2015). While we cannot rule out the possibility that some of the associated mutations are advantageous it seems likely that the observed pattern reflects an increase in mildly deleterious mutations.

Genes containing candidate SNPs for positive selection were mainly involved in metal binding and transmembrane signaling receptor activity, although genes involved in ATPase activity, development and proteosome core complex were also identified. When comparing our results to previous studies looking at signatures of selection within species, Pujolar et al. (2014a) reported several genes putatively under selection related to calcium signaling (atp2a2, cacna1s, adra1d, tacr3, grin2a), neuroactive ligand-receptor interaction (drd1, gabra1, adora1, p2ry4, gria), focal adhesion (itga5, ptenb, tln1, capn2l, crf, braf), wnt signaling (ctbp2, srpf1, smad2, wnt7a, wnt8a), and circadian rhythms (per). In the present study, none of these genes were detected, however this is anticipated since local adaptation in eels is not expected due to panmixia and random larval dispersal (Gagnaire et al., 2012; Pujolar et al., 2014a; Pavey et al., 2015) and hence the genes putatively under selection in the study represent non-heritable single-generation signatures of selection.

In American eel, Gagnaire et al. (2012) tested a panel of 80 coding-gene SNPs and found evidence for spatially variant selection at 8 genes (acp, anx2, gpx4, hsp90a, mdh, nrap, prp40, ugp2), showing significant correlations between allele frequencies and environmental variables across the entire distribution range of American eel, particularly determined by temperature regimes. Interestingly, Ulrik et al. (2014) used the same panel of SNPs to genotype European eel individuals and showed significant associations between allele frequencies and environmental variables at 10 genes (aldh, aldr, cst, gapdh, krt, nex, pgk, psa4, trim35, ubia52). None of the genes found to be under selection in European eel showed any correlations with environmental factors in the previous study in American eel. Pujolar et al. (2015) used once more the same panel of SNPs when comparing European eel individuals from different life stages (glass eels vs silver eels) and found some of the same genes as in Ulrik et al. (2014), in particular gapdh and aldr, when testing correlations with environmental variables. Other genes were found under selection (clic5, ldhb, cst, csde1) when using an F_{ST} approach.

None of the genes putatively under selection in the studies of Gagnaire et al. (2012), Ulrik et al. (2014) and Pujolar et al. (2015) were picked up in the present study. However, those genes under selection in the previous studies were mainly identified using the software Bayenv that looks at correlations between allele frequencies and environmental variables (i.e., temperature). Those genes detected by Bayenv. likely reflect within generation-selection and not recent population-wide selection as expected under a selective sweep. Moreover, it is possible that the present study did not identify selection at any of those genes simply because RAD sequencing did not target those regions, as this method generates a reduced representation of the genome.

Genes putatively under selection in our study were mainly involved in metal binding and transmembrane signaling receptor activity as well as ATPase activity, unfolded protein binding activity, negative regulation of cytokine production and proteosome core complex among others. However, an enrichment test might be defective since most of the genes coded for proteins with highly similarity to other proteins but still uncharacterized. For instance, this included a gapdh-like protein (gapdh was under selection in Ulrik et al., 2014 and Pujolar et al., 2015).

Given that most SNPs putatively under selection in our study were found outside exons and coding regions (as revealed by SNPeff), it is possible that selection is mainly acting on regulation and expression rather than functional changes. The high number of candidate SNPs in noncoding regions of the genome likely reflects regulatory differences between North Atlantic eels. Similarly, in the case of sticklebacks Jones et al. (2012) found up to 83% of SNPs under selection located in noncoding regions with an assumed regulatory role. It is worth noting that RAD-sequencing only allows detection of a reduced representation of the genome wide variation, so this may represent a limiting factor when searching for missense mutations. Additionally, it should be taken into consideration that most mutations within an island of selection are likely to be hitchhiking and hence not the direct targets of selection making it difficult to differentiate between advantageous and neutral or deleterious mutations.

4.4. Shared selection resulting from introgression?

Soft selective sweeps should be more easily detected using the iHS approach, which looks at footprints of selection within species. The identification of regions under selection shared by both European and American eel in our study is compatible with several scenarios, including parallel evolution and introgression. Soft selective sweeps can occur in parallel in multiple separate locations as a response to similar environmental cues, as in the case of freshwater adaptation in sticklebacks (Jones et al., 2012). However, it seems unlikely that the large number of regions under selection detected by iHS in our study all result from parallel evolution. As mentioned above, Ulrik et al. (2014) found no evidence for parallelism between European and American eel when testing a panel 80 coding-gene SNPs with key metabolic functions and linked to environmental variables. A more plausible scenario is that the high number of regions under selection shared by North Atlantic eels is the result of introgression. Despite separate adult distributions, both European and American eel spawn in partial sympatry in the Sargasso Sea, which gives ample opportunity for hybridization (e.g. Als et al., 2011; Jacobsen et al., 2017). The occurrence of hybrids on the mainland is low (ca. 3% for European eel and 6.6% for American eel; Pujolar et al., 2014a; Jacobsen et al., 2017) except for Iceland where there is an important contribution of individuals of admixed ancestry (10.7%; Pujolar et al., 2014b). Moreover, JAFS analysis show support for ongoing gene flow between the species (Nikolic et al., 2020). Introgressive alleles that behave neutrally are only expected to be found sporadically and are likely to be lost via genetic drift. On the other hand, introgressed alleles that confer a selective advantage after introgression may reach higher frequencies even in the case of low levels of admixture (Evans et al., 2006). Introgressive alleles are more likely to persist if selected at the time of hybridization but even alleles that initially evolve neutrally can become adaptive at a later time (Mendez et al., 2012). Studies in humans have convincingly demonstrated that introgression of ancestral DNA from Neanderthals or Denisovans explains the evolution of immunity (Mendez et al., 2012), altitude adaptation (Huerta-Sánchez et al., 2014) and skin pigmentation genes (Vernot and Akey, 2014).

5. Conclusions

In summary, our study identifies putative genomic signatures of selection in North Atlantic eels. First, we found apparent islands of divergence on a total of seven chromosomes, particularly on chromosomes 6 and 10. Potential candidate genes were mainly related to energy production, which would reflect strong selection on traits related to eel migration. Moreover, the high number of candidate SNPs in noncoding regions of the genome might point to a regulatory role. Second, shared regions under selection were detectable on a total of 11 chromosomes. The detection of shared islands of selection in North Atlantic eels opens the door to several hypotheses, including parallel evolution due to similar environmental conditions and introgression. Future comparative genomic studies will be needed to further clarify the causes and consequences of introgression, including the directionality of these introgression events.

Declaration of Competing Interest

None.

Acknowledgments

We thank Computerome for assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2022.100933.

References

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinforma. 12, 246. https://doi.org/10.1186/1471-2105-12-246.

Allendorf, F.W., Hohenlohe, P.A., Luikart, G., 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11, 697–709. https://doi.org/10.1038/nrg2844.

Als, T.D., Hansen, M.M., Maes, G.E., Castonguay, M., Riemann, L., Aarestrup, K., Munk, P., Sparholt, H., Hanel, R., Bernatchez, L., 2011. All roads lead to home: panmixia of European eel in the Sargasso Sea. Mol. Ecol. 20, 1333–1346. https://doi.org/10.1111/j.1365-294X.2011.05011.x.

Andrews, S., 2015. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091–1093. https://doi.org/10.1093/bioinformatics/btp101.

Boëtius, J., 1980. Atlantic *Anguilla*. A presentation of old and new data of total numbers of vertebrae with special reference to the occurrence of *Anguilla rostrata* in Europe. Dana 1, 93–112.

Bolger, A.M., Lohse, M., Usabel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Cádiz, M.I., López, M.E., Díaz-Domínguez, D., Cáceres, G., Yoshida, G.M., Gomez-Uchida, D., Yáñez, J.M., 2020. Whole genome re-sequencing reveals recent signatures of selection in three strains of farmed Nile tilapia (*Oreochromis niloticus*). Sci. Rep. 10, 11514. https://doi.org/10.1038/s41598-020-68064-5.

Cadzow, M., Boocock, J., Nguyen, H.T., Wilcox, P., Merriman, T.R., Black, M.A., 2014. A bioinformatics workflow for detecting signatures of selection in genomic data. Front. Genet. 5, 293. https://doi.org/10.3389/fgene.2014.00293.

Chun, S., Fay, J.C., 2009. Identification of deleterious mutations within three human genomes. Genome Res. 19, 1553–1561. https://doi.org/10.1101/gr.092619.109.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNP in the genome of Drosophila melanogaster strain w118; iso-2; iso-3. Fly 6, 80–92. https://doi.org/10.4161/fly.19695.

Côté, C., Gagnaire, P.A., Bourret, V., Verrault, 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. Mol. Ecol. 22, 1763–1776. https://doi.org/10.1111/mec.12142.

Cruz, F., Vilà, C., Webster, M.T., 2008. The legacy of domestication: accumulation of deleterious mutations in the dog genome. Mol. Biol. Evol. 25, 2331–2336. https://doi.org/10.1093/molbev/msn177.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.B., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Dennis Jr., G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 4, R60. https://doi.org/10.1186/gb-2003-4-9-r60.

Enbody, E.D., Petterson, M.E., Sprehn, C.G., Palm, S., Wickstrom, H., Andersson, L., 2021. Ecological adaptation in European eels is based on phenotypic plasticity. Proc. Natl. Acad. Sci. U. S. A. 118 https://doi.org/10.1073/pnas.2022620118 e2022620118.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Evans, P.D., Mekel-Bobrov, N., Vallender, E.J., Hudson, R.R., Lahn, B.T., 2006. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into Homo sapiens from an archaic Homo lineage. Proc. Natl. Acad. Sci. U. S. A. 103, 18178–18183. https://doi.org/10.1073/pnas.0606966103.

Gagnaire, P.A., Albert, V., Jónsson, B., Bernatchez, L., 2009. Natural selection influences AFLP intraspecific genetic variability and introgression patterns in Atlantic eels. Mol. Ecol. 18, 1678–1691. https://doi.org/10.1111/j.1365-294X.2009.04142.x.

Gagnaire, P.A., Normandeau, E., Côté, C., Hansen, M.M., Bernatchez, L., 2012. The genetic consequences of spatially varying selection in the panmictic American eel. Genetics 190, 726–736. https://doi.org/10.1534/genetics.111.134825.

Gautier, M., Vitalis, R., 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176–1177. https://doi.org/10.1093/bioinformatics/bts115.

Hemmer-Hansen, J., Therkildsen, N.O., Pujolar, J.M., 2014. Population genomics of marine fishes: next generation prospects and challenges. Biol. Bull. 227, 117–132. https://doi.org/10.1086/BBLv227n2p117.

Henkel, C.V., Burgerhout, E., de Wijze, D.L., Dirks, R.P., Minegishi, Y., Jansen, H.J., Spaink, H.P., Dufour, S., Weltzien, F.A., Tsukamoto, K., van den Thillart, G.E.E.J.M., 2012. PLoS One 7. https://doi.org/10.1371/journal.pone.0032231 e32231.

Hermisson, J., Pennings, P.S., 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169, 2335–2352. https://doi.org/10.1534/genetics.104.036947.

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6 https://doi.org/10.1371/journal.pgen.1000862 e1000862.

Huerta-Sánchez, E., et al., 2014. Altitude adaptation in Tibet caused by introgression of Denisovan-like DNA. Nature 514, 194–197. https://doi.org/10.1038/nature13408.

Jacobsen, M.W., Pujolar, J.M., Gilbert, M.T.P., Moreno-Mayar, J.V., Bernatchez, L., Als, T.D., Lobón-Cervià, J., Hansen, M.M., 2014a. Speciation and demographic history of Atlantic eels (*Anguilla anguilla* and *A. rostrata*) revealed by mitogenome sequencing. Heredity 113, 432–442. https://doi.org/10.1038/hdy.2014.44.

Jacobsen, M.W., Pujolar, J.M., Bernatchez, L., Munch, K., Jian, J., Niu, Y., Hansen, M.M., 2014b. Genomic footprints of speciation in Atlantic eels *Anguilla anguilla* and *A. rostrata*. Mol. Ecol. 23, 4785–4798. https://doi.org/10.1111/mec.12896.

Jacobsen, M.W., Pujolar, J.M., Hansen, M.M., 2015. Relationship between amino acid changes in mitochondrial ATP6 and life-history variation in anguillid eels. Biol. Lett. 11, 20150014. https://doi.org/10.1098/rsbl.2015.0014.

Jacobsen, M.W., Smedegaard, L., Sørensen, S.R., Pujolar, J.M., Munk, P., Jónsson, B., Magnussen, E., Hansen, M.M., 2017. Assessing pre- and post-zygotic barriers between North Atlantic eels (*Anguilla anguilla* and *A. rostrata*). Heredity 118, 266–275. https://doi.org/10.1038/hdy.2016.96.

Jones, F.C., et al., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484, 55–61. https://doi.org/10.1038/nature10944.

Kim, E.S., Elbeltagy, A.R., Aboul-Naga, A.M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., Rothschild, M.F., 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot environment. Heredity 116, 255–264. https://doi.org/10.1038/hdy.2015.94.

Kreitman, M., 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. 1, 539–559. https://doi.org/10.1146/annurev.genom.1.1.539.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., Simianer, H., 2015. Properties of different selection signature statistics and a new strategy to combine them. Heredity 115, 426–436. https://doi.org/10.1038/hdy.2015.42.

Maiorano, A.M., et al., 2018. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One 13. https://doi.org/10.1371/journal.pone.0200694 e0200694.

Mank, J.E., Avise, J.C., 2003. Microsatellite variation and differentiation in North Atlantic eels. J. Hered. 94, 310–314. https://doi.org/10.1093/jhered/esg062.

Marques, D.A., Lucek, K., Meier, J.I., Mwaiko, S., Wagner, C.E., Excoffier, L., Seehausen, O., 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. PLoS Genet. 12 https://doi.org/10.1371/journal.pgen.1005887 e1005887.

Marsden, C.D., Ortega-Del Vecchyo, D., O'Brien, D.P., Taylor, J.F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R.D., Wayne, R.K., Lohmueller, K.E., 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc. Natl. Acad. Sci. U. S. A. 113, 152–157. https://doi.org/10.1073/pnas.1512501113.

Mendez, F.L., Watkins, J.C., Hammer, M.F., 2012. A haplotype at STAT2 introgressed from Neanderthals and serves as a candidate of positive selection in Papua New Guinea. Am. J. Hum. Genet. 91, 265–274. https://doi.org/10.1016/j.ajhg.2012.06.015.

Miller, M.J., Bonhommeau, S., Munk, P., Castonguay, M., Hanel, R., McCleave, J.D., 2015. A century of research on the larval distributions of the Atlantic eels: a re-examination of the data. Biol. Rev. Camb. Philos. Soc. 90, 1035–1064. https://doi.org/10.1111/brv.12144.

Miller, M.J., et al., 2019. Spawning by the European eel across 2000 km of the Sargasso Sea. Biol. Lett. 15, 20180835. https://doi.org/10.1098/rsbl.2018.0835.

Minegishi, Y., Aoyama, J., Inoue, J.G., Miya, M., Nishida, M., Tsukamoto, K., 2005. Molecular phylogeny and evolution of the freshwater eels genus Anguilla based on the whole mitochondrial genome sequences. Mol. Phylogenet. Evol. 34, 134–146. https://doi.org/10.1016/j.ympev.2004.09.003.

Nikolic, N., Liu, S., Jacobsen, M.W., Jónsson, B., Bernatchez, L., Gagnaire, P.A., Hansen, M.M., 2020. Speciation history of European (*Anguilla Anguilla*) and American eel (*A. rostrata*), analysed using genomic data. Mol. Ecol. 29, 565–577. https://doi.org/10.1111/mec.15342.

Nosil, P., Feder, J.L., 2013. Genome evolution and speciation: towards quantitative descriptions of pattern and process. Evolution 67, 2461–2467. https://doi.org/10.1111/evo.12191.

Nosil, P., Vines, T.H., Funk, D.J., 2005. Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59, 705–719.

O'Connell, J., et al., 2014. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet. 10 https://doi.org/10.1371/journal.pgen.1004234 e1004234.

Patterson, N., Price, A.L., Reich, D., 2006. Population structure and eigenanalysis. PLoS Genet. 2 https://doi.org/10.1371/journal.pgen.0020190 e190.

Pavey, S.A., 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. Curr. Biol. 25, 1666–1671. https://doi.org/10.1016/j.cub.2015.04.062.

Pritchard, J.K., Stephens, M., Donelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959. https://doi.org/10.1093/genetics/155.2.945.

Pujolar, J.M., Jacobsen, M.W., Als, T.D., Frydenberg, J., Munch, K., Jónsson, B., Jian, J. B., Cheng, L., Maes, G.E., Bernatchez, L., Hansen, M.M., 2014a. Genome-wide signatures of within-generation local selection in the panmictic European eel. Mol. Ecol. 23, 2514–2528. https://doi.org/10.1111/mec.12753.

Pujolar, J.M., Jacobsen, M.W., Als, T.D., Frydenberg, J., Magnussen, E, Jónsson, B., Jiang, X., Cheng, L., Bekkevold, D., Maes, G.E., Bernatchez, L., Hansen, M.M., 2014b. Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. Heredity 112, 627–637. https://doi.org/10.1038/hdy.2013.145.

Pujolar, J.M., Jacobsen, M.W., Bekkevold, D., Lobón-Cervià, J., Jónsson, B., Bernatchez, L., Hansen, M.M., 2015. Signatures of natural selection between life cycle stages separated by metamorphosis in European eel. BMC Genomics 16, 600. https://doi.org/10.1186/s12864-015-1754-3.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.

Renaut, S., Rieseberg, L.H., 2015. The accumulation of deleterious mutations as a consequence of domestication and improvement in sunflowers and other Compositae crops. Mol. Biol. Evol. 32, 2273–2283. https://doi.org/10.1093/molbev/msv106.

Rhie, A., et al., 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592, 737–746. https://doi.org/10.1038/s41586-021-03451-0.

Sabeti, P.C., et al., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913–918. https://doi.org/10.1038/nature06250.

Schluter, D., 2009. Evidence for ecological speciation and its alternative. Science 323, 737–741. https://doi.org/10.1126/science.1160006.

Schmidt, J., 1923. The breeding places of the eel. Philos. Trans. R. Soc. Lond. B 211, 179–208. https://doi.org/10.1098/rstb.1923.0004.

Seehausen, O., et al., 2014. Genomics and the origin of species. Nat. Rev. Genet. 15, 176–192. https://doi.org/10.1038/nrg3644.

Smith, J.M., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23, 23–35. https://doi.org/10.1017/S0016672308009579.

Storz, J.F., 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. 14, 671–688. https://doi.org/10.1111/j.1365-294X.2005.02437.x.

Tesch, F., 2003. The Eel. Blackwell Science Lt, Oxford, UK.

Ulrik, M.G., Pujolar, J.M., Ferchaud, A.L., Jacobsen, M.W., Als, T.D., Gagnaire, P.A., Frydenberg, J., Bocher, P.K., Jonsson, B., Bernatchez, L., Hansen, M.M., 2014. Do North Atlantic eels show parallel patterns of spatially varying selection? BMC Evol. Biol. 14, 138. https://doi.org/10.1186/1471-2148-14-138.

Vernot, V., Akey, J.M., 2014. Resurrecting surviving Neandertal lineages from modern human genomes. Science 343, 1013–1017. https://doi.org/10.1126/science.1245938.

Via, S., West, J., 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17, 4334–4345. https://doi.org/10.1111/j.1365-294X.2008.03921.x.

Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biol. 4 https://doi.org/10.1371/journal.pbio.0040072 e72.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.

Wirth, T., Bernatchez, L., 2003. Decline of North Atlantic eels: a fatal synergy? Proc. Biol. Sci. 270, 681–688. https://doi.org/10.1098/rspb.2002.2301.

Zhang, M., Zhou, L., Bawa, R., Suren, H., Holliday, J.A., 2016. Recombination rate variation, hitchhiking and demographic history shape deleterious load in poplar. Mol. Biol. Evol. 33, 2899–2910. https://doi.org/10.1093/molbev/msw169.