ORIGINAL ARTICLE

Evolutionary Applications

WILEY

Seascape genomics reveals limited dispersal and suggests spatially varying selection among European populations of sea lamprey (*Petromyzon marinus*)

Miguel Baltazar-Soares[1,2,3] ⓘ | J. Robert Britton[1] ⓘ | Adrian Pinder[1] |
Andrew J. Harrison[1] | Andrew D. Nunn[4] | Bernardo R. Quintella[5,6] |
Catarina S. Mateus[5] | Jonathan D. Bolland[4] | Jamie R. Dodd[4] | Pedro R. Almeida[5,7] |
Victoria Dominguez Almela[1] ⓘ | Demetra Andreou[1]

[1]Department of Life and Environmental Sciences, Faculty of Science and Technology, Bournemouth University, Dorset, UK

[2]MARE – Marine and Environmental Sciences Centre, ISPA – Instituto Universitário, Lisbon, Portugal

[3]Department of Biology, University of Turku, Turku, Finland

[4]University of Hull, Hull International Fisheries Institute, Hull, UK

[5]MARE—Marine and Environmental Sciences Centre, University of Évora, Évora, Portugal

[6]Department of Animal Biology, Faculty of Sciences, University of Lisbon, Lisbon, Portugal

[7]Department of Biology, School of Sciences and Technology, University of Évora, Évora, Portugal

Correspondence
Miguel Baltazar-Soares, Department of Life and Environmental Sciences, Faculty of Science and Technology, Bournemouth University, Dorset, UK.
Email: miguelalexsoares@gmail.com

Abstract

Sea lamprey *Petromyzon marinus* is an anadromous and semelparous fish without homing behaviors. Despite being a freshwater, free-living organism for a large part of their life cycle, its adulthood is spent as a parasite of marine vertebrates. In their native European range, while it is well-established that sea lampreys comprise a single nearly-panmictic population, few studies have further explored the evolutionary history of natural populations. Here, we performed the first genome-wide characterization of sea lamprey's genetic diversity in their European natural range. The objectives were to investigate the connectivity among river basins and explore evolutionary processes mediating dispersal during the marine phase, with the sequencing of 186 individuals from 8 locations spanning the North Eastern Atlantic coast and the North Sea with double-digest RAD-sequencing, obtaining a total of 30,910 bi-allelic SNPs. Population genetic analyses reinforced the existence of a single metapopulation encompassing freshwater spawning sites within the North Eastern Atlantic and the North Sea, though the prevalence of private alleles at northern latitudes suggested some limits to the species' dispersal. Seascape genomics suggested a scenario where oxygen concentration and river runoffs impose spatially varying selection across their distribution range. Exploring associations with the abundance of potential hosts further suggested that hake and cod could also impose selective pressures, although the nature of such putative biotic interactions was unresolved. Overall, the identification of adaptive seascapes in a panmictic anadromous species could contribute to conservation practices by providing information for restoration activities to mitigate local extinctions on freshwater sites.

KEYWORDS
freshwater biodiversity, limited dispersal, *Petromyzon marinus*, seascape genomics

1 | INTRODUCTION

Freshwater systems have played a key role in the evolution of human society by providing reliable sources of potable water and food, as well as communication and trading routes (Ormerod et al., 2010). To facilitate these services, rivers have been heavily modified in order to facilitate navigation and provide more efficient hydropower generation through the building of, for example, weirs and dams (Dudgeon et al., 2006; Grill et al., 2019). For anadromous fishes, including some species of salmonids and lampreys, these modifications can place considerable pressure on their freshwater life phases through blocking access to environmentally stable reproductive and nursery habitats (e.g., Branco et al., 2012; Mateus et al., 2012; Moser et al., 2020). Maintaining sustainable diadromous fish populations is vital both from a social-economic perspective—as many of these species are economically relevant—and from a biological standpoint, as spawning and foraging migrations ensure energy and biomass transfer between marine and freshwater environments (Limburg & Waldman, 2009).

Despite local freshwater extinctions, anadromous populations can still be maintained across large spatial scales through their use of the marine habitat, while using available rivers for spawning and nursery habitats (Moser et al., 2020). In this context, characterizing gene flow and inferring whether populations are locally adapted is essential for the sustainable replenishment of extant populations. Genetic and genomic resources have greatly contributed to understanding connectivity patterns (Asaduzzaman et al., 2020; Perrier et al., 2014), defining conservation units (Mateus et al., 2011; Turbek et al., 2023; Waples & Lindley, 2018), and pinpointing evolutionary signatures related to migration, maturation and/or local adaptation in several anadromous species, such as in the Atlantic salmon, brown trout or shad (Antognazza et al., 2021; Kapralova et al., 2011; Leitwein et al., 2017; Prince et al., 2017). Genome-wide information is nevertheless vital to ensure the effectiveness of conservation-driven translocation activities, as it would permit to devise genetic rescue strategies to mitigate the species' biodiversity losses and increase population resilience without compromising local adaptation (Whiteley et al., 2015).

The sea lamprey *Petromyzon marinus* is an anadromous, ancient jawless vertebrate whose native distribution is limited to the Western and Eastern coasts of the North Atlantic and inside the Mediterranean, where individual abundances fluctuate regionally and locally from one freshwater system to another (Guo & Britton, 2017; Boulêtreau et al., 2020). The sea lamprey has a complex lifecycle consisting of several developmental stages, where the major morphological changes are intrinsically associated with feeding behavior (Youson & Potter, 1979). Posthatching larvae, or ammocoetes, are free-living organisms that feed on detritus deposited on river sediment beds or through filtration of the water column (Guo et al., 2016). The metamorphosis into young adults marks the onset of the species' parasitic behavior, as sea-dwelling lampreys mandatorily feed on blood of other vertebrates (Quintella et al., 2021; Silva et al., 2013). Adult sea lampreys spend their life at sea, a stage that is expected to last between 10 and 28 months preceding sexual maturation (Beamish, 1980; Renaud & Cochran, 2019; Silva et al., 2013). Spawning migration towards freshwater, however, is not philopatric regarding spawning locations; contrary to other anadromous fishes, sea lampreys do not return to the natal rivers to spawn (Almada et al., 2008; Bergstedt & Seelye, 1995; Spice et al., 2012). Instead, it has been suggested that sea lampreys are steered by a migratory pheromone released by stream-resident larvae (Sorensen et al., 2005; Vrieze et al., 2011), although natural recolonization of freshwater locations without ammocoete presence has been recently noted (Reid & Goodman, 2020). Less understood is how selection acts across the sea lamprey life cycle, despite its numerous life-history events (such as the transition to parasitism linked to adulthood, onset of anadromous and catadromous migrations) providing substantial opportunities for selection to act. For instance, it has been recently shown that the parasitic adult stage imposes a selective pressure on the host, as suggested by evidence showing the transcriptional response of genes involved in the regulation of inflammation and cellular damage in parasitized lake trout (Goetz et al., 2016). Indeed, recent work investigating the molecular pathway of the renin-angiotensin system on lampreys has suggested that its evolution is linked with mimicry of the teleost-like angiotensin while parasitizing, in order to recreate host's environment and avoid rejection—referred to as endocrine mimicry (Goetz et al., 2016). Though plausibly expected, a putative relationship between the distribution of potential hosts and signatures of selection on the sea lamprey genome—which could indicate adaptation to the host community—has not been investigated at any spatial scale.

Transitions from fresh to salt water (and vice-versa) do impose an energetic cost associated with physiological and molecular re-arrangements to cope with the chemical composition of the different environments (Ferreira-Martins et al., 2016). Reduced tolerance to high salinities has been reported in sea lamprey adults performing spawning migrations; failure to locate suitable spawning locations in freshwater due to connectivity loss by dams and weirs requires re-entering estuaries and the marine environment to access adjacent river systems (Sunga et al., 2020). Assuming that tolerance to those transitions would be an adaptive trait, it could be hypothesized that ocean salinity would play a role in limiting adult sea lamprey dispersal. Within a certain geographic area, individuals could be adapted to fresh-saltwater gradients to minimize failure of reproduction by spending time searching for suitable freshwater spawning locations. Given the regional and local extinctions of sea lamprey populations in the European range (Hansen et al., 2016; Mateus et al., 2012), the absence of homing behavior might facilitate conservation practices based on, for example, the translocation of ammocoetes to maintain local biodiversity levels, as performed on the Columbia river basin with Pacific lampreys (Ward et al., 2012). Undertaking similar conservation measures to mitigate the local extinctions of European lamprey would thus demand an upscale of extant knowledge on the species´ connectivity and the identification of signatures associated with adaptive processes underlying their evolutionary potential

(Eizaguirre & Baltazar-Soares, 2014; McMahon et al., 2014). While traditional molecular tools have been extensively applied to investigate genetic structure or historical demography (Genner et al., 2012; Mateus et al., 2012), next-generation sequencing has been surprisingly lacking among the repertoire of studies dedicated to natural populations of sea lamprey. By harnessing the power of genome-wide information, we intend to revisit genetic structuring among populations of sea lamprey across the North Eastern Atlantic coast—hypothesizing panmixia—while exploring molecular signatures of the selective pressures proposed above. Inferring signatures of selection on a highly migratory species, such as sea lamprey, would help the understanding of how micro-evolutionary processes may act on complex life-history strategies. Concomitantly, characterizing the connectivity among spawning grounds and the evolutionary pressures that could be delimiting the marine dispersal of eastern Atlantic Sea lampreys is fundamental in establishing conservation measures befitting the species' wide distribution range.

2 | MATERIALS AND METHODS

2.1 | Sampling strategy across the North Eastern Atlantic

To explore the evolutionary history of sea lamprey populations across the Eastern Atlantic coast, 186 individuals were sampled from eight locations across their Northeast Atlantic distribution range (Figure 1). Fin tissue samples were available from the rivers Tagus ($n = 20$) and Mondego ($n = 20$) in Portugal, from the Ulla ($n = 22$) in northern Spain, from the Gironde estuary ($n = 18$) in France, from the rivers Severn ($n = 22$), Frome ($n = 34$) and Yorkshire Ouse ($n = 20$) in Great Britain, and from the river Roflsan ($n = 22$) in Sweden.

Migrating sea lamprey were captured using unbaited two-funnel eel pots. Captured individuals were held in water-filled containers (100 L) prior to general anesthesia (MS-222) that enabled a fin biopsy to be taken. All the sampled individuals were adults, except for those from the Frome, where, due to the unavailability of adults during the sampling period, ammocoetes were sampled. Ammocoete sampling was performed by dredging, with a dipnet, soft sediment areas of the river Frome. All the tissue samples were preserved in ethanol (96%) prior to extraction.

2.2 | DNA extraction, library preparation, and double-digest restriction-associated DNA tag sequencing

Genomic DNA was extracted from the fin or muscle tissues with Qiagen DNeasy Blood and Tissue kit (Hilden, Germany) following the manufacturer's instructions. Library preparation, sequencing, and bioinformatic processing of raw reads (demultiplexing
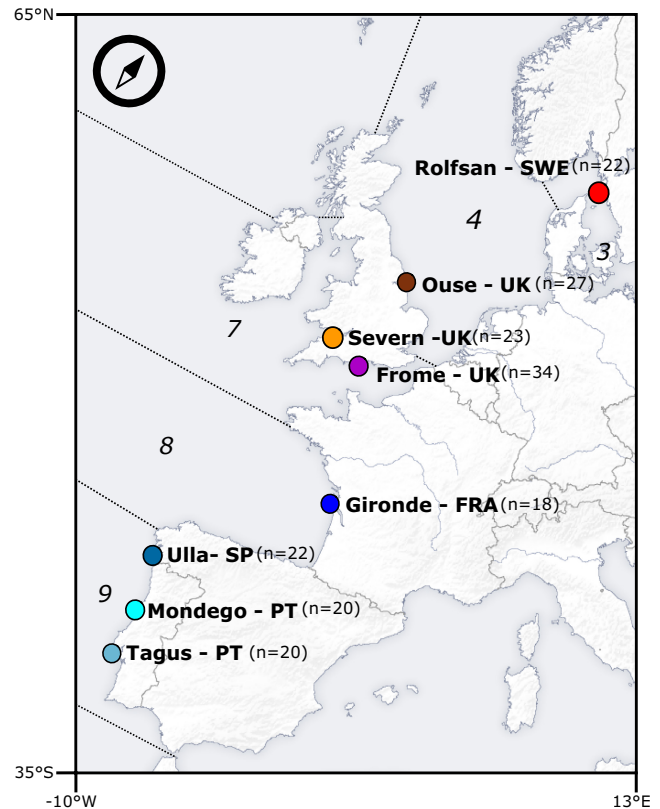
FIGURE 1 Geographic information and distribution of sampled locations. Here represented are the sampled freshwater locations (dots) and respective correspondence to FAO's major fishing areas in the Northeast Atlantic (italicized numbers), delineated by dashed lines. Specifically, *3* corresponds to Skagerrak, Kattegat, Sound, Belt Sea, and Baltic Sea, *4* to the North Sea, *7* to Irish Sea, West of Ireland, Porcupine Bank, Eastern and Western English Channel, *8* to the Bay of Biscay and *9* to the Portuguese waters. In parenthesis, the number of individuals collected for each location was utilized in this study.

of individual barcodes, removal of adaptors and barcodes, and trimming) were performed at the GENEWIZ© in Takeley, United Kingdom. Library preparation followed the double digest protocol of Peterson et al. (2012). Libraries were constructed by pooling 48 samples after individual barcoding. Genomic DNA was digested with EcoR1/Msp1 restriction enzymes. Two 96-well plates were sequenced on an Illumina HiSeq2500 with a 2×125bp paired-end (PE) configuration. Final library was size-selected between 300 and 400 bp inserts. Only reads with $Q > 30$ were kept for downstream analysis.

2.3 | Processing paired-end reads

Reads were aligned against the available sea lamprey genome (Smith et al., 2013) using the *bwa-mem* algorithm (Li, 2013) implemented in *bwa* (Li & Durbin, 2009). Bam files were imported into Stacks, version 2.2 (Catchen et al., 2013) for further analyses and variant calling. Only bi-allelic loci that were present in 80% of the

populations were kept, as well as one SNP per tag (Rochette & Catchen, 2017). Because the sampling spanned a wide geographic area, a minor allelic frequency of 0.01 (overall dataset) was filtered out to avoid losing unique information. We also kept only loci with the maximum observed heterozygosity of 0.5 (Roesti et al., 2012). To mitigate a potential effect of the high repeatability found on this species' genome, all the reads whose coverage was above 100× were also filtered out, along with loci that mapped to more than one catalog.

2.4 | Genetic diversity indices across sampled locations

All statistical analyses were computed in R v 3.4.1. (R Core Team, 2013). For each location, we estimated the heterozygosity (*Hs*), and inbreeding coefficient (F) with the *adegenet* R package (Jombart, 2008) and the number of private alleles (*pA*), nucleotide diversity (pi) with *Stacks*. Absolute values of *pA* were log-transformed to normalize the distribution prior to statistical analyses. The distribution of genome-wide diversity was first compared across sampled locations, where we investigated the relationship of each index with latitude and longitude. Relationships were tested between diversity indices and geographic distance using linear models. Because we collected ammocoetes from the Frome, we utilized PLINk2 (Chang et al., 2015) to perform a relatedness analyses (KING) across the overall dataset and utilizing a cut-off of 0.1 following the convention to cut off first-degree relationships as it could be expected if collected ammocoetes would be originated from the same breeding pairs (Chang et al., 2015).

2.5 | Population structure and connectivity among freshwater locations

Multiple approaches were used to investigate the population connectivity between the sampled locations. A PCA was performed in the R package *pcadapt* v 4 (Privé et al., 2020) in a first attempt to verify if sampling locations harbored genetically similar individuals. Pairwise F_{ST} was estimated in Arlequin v3.5 (10,000 permutations) (Excoffier & Lischer, 2010) to understand overall genome-wide diversity. Bayesian clustering methodology explored the distribution of genome-wide variation across the sampled sites, running *fast-structure* (Raj, Stephens, et al., 2014) with three independent iterations for *K* values of 1–8. The most likely *K* number was assessed using *chooseK* (Raj, Rothamel, et al., 2014). Admixture proportions were visualized with *compoplot* function in the *adegenet*. Lastly, discriminant analysis of principal component (DAPC) was used to assess individual clustering based on genetic similarities in *adegenet*. We performed DAPCs assuming both the original sampling location, that is, predefined populations, as well as exploring the configuration of clusters with the function *find. clusters*. Putative clusters were identified with Bayesian Information Criteria (BIC), and the assignment of individuals was compared with that of sampling locations.

2.6 | Investigating the relationship between genomic variation and environmental variables

Environmental and genetic variation tend to be interpreted together under the assumption that co-variation suggests signatures of evolutionary responses to environmental conditions (Günther & Coop, 2013). The wide range of environments experienced by sea lampreys during their life cycle renders environmental correlations a logical approach to investigate putative signatures of adaptation across spatial scales. Specifically, we investigated possible adaptations to the transition freshwater-saltwater-freshwater by investigating the relationship between genomic variation versus sea surface salinity. Adaptation to the nearby marine environment was further investigated by inspecting the relationships between genomic variation with surface temperature, a surrogate for water temperature, and dissolved oxygen concentration, both critical for ectotherm thermal regulation and metabolic processes in the marine environment. We also explored whether coastal anthropogenic activities could be a mediating factor to sea lamprey's marine distribution. We considered phosphate and nitrate concentrations as they are commonly associated with sewage effluent discharges and the consequent eutrophication of coastal systems at various spatial resolutions (Van Drecht et al., 2009).

All these variables were extracted from Bio-ORACLE—spatial resolution of 5 arcmin (c. 0.088 or 9.2 km at the equator (Assis et al., 2018; Tyberghein et al., 2012)) with customized R scripts. A total of 6 environmental parameters, that is, *sea surface salinity*, *sea surface temperature*, *dissolved oxygen*, *phosphate concentration*, and *nitrate concentration* were collated, including the minimum and maximum of the variables described in Table S1.

The obligatory parasitic stage of sea lamprey's life cycle suggests the species could have evolved adaptations towards the availability of certain host species or to the diversity of host communities. We tested that hypothesis by comparing the variation in the abundance of fishes that have been referenced in the literature as bearing sea lamprey sucking marks or adult individuals attached to them, to the genome-wide loci variance among sampled locations. The list of potential hosts included the Atlantic mackerel (*Scomber scombrus*), the Atlantic cod (*Gadus morhua*), the basking shark (*Cetorhinus maximus*), the European hake (*Merluccius merluccius*), the thick lip grey mullet (*Chelon labrosus*), the Atlantic salmon (*Salmo salar*), the twaite shad (*Alosa fallax*) and the allis shad (*Alosa alosa*) (Silva et al., 2014). Abundances were extracted from the Food and Agriculture Organization (FAO)'s FishStatJ—Software (FAO, 2016) as capture data (in tonnes) (Figure S1). We focused on areas within the Northeast Atlantic (Major Fishing Area 27), which corresponded to marine regions adjacent to sampled sites. More specifically, we collected data from area 3, corresponding to Skagerrak, Kattegat, Sound, Belt Sea, and Baltic Sea; area 4, the North Sea; area 7, corresponding to Irish Sea, West of Ireland, Porcupine Bank, Eastern and Western English Channel; area 8, the Bay of Biscay; and area 9, Portuguese waters. Captures on FishStatJ spanned a period of 10 years from 2006 to 2016, thus we calculated the 10-year harmonic average for each

species, per area. To reduce the variance associated with annual or decadal means, all abiotic and biotic variables were standardized for their z-scores prior to computation analyses. Standardized values were further used throughout the remaining analyses.

Relationships between the above-mentioned variables and genetic variation were first tested for the full SNP dataset in Bayenv2 (Günther & Coop, 2013). To increase the robustness of any observed gene–environment association, we imposed a double filter for statistical significance. Specifically, only the gene–environment relationships with a Bayes Factor > 10 (Jeffreys, 1998) and an Spearman's *rho* statistic above 0.2 (Günther & Coop, 2013) were considered as significant. In addition, we utilized latent factor mixed models (LFMM) to further inspect for environment correlations with the genetic data (Frichot et al., 2013). Contrary to the island model of neutral allelic frequency distribution to identify outliers assumed in Bayenv2, LFMMs require a priori information on population structure to infer the background noise of neutral differentiation (Frichot et al., 2013). We ran LFMM on the R package *LEA* (Frichot & François, 2015). Because the methodology is sensitive to missing data, we performed the imputation of null alleles based on estimated ancestry coefficients and ancestral genotype frequencies, as suggested by Frichot and François (2015) and considering the "mode" of each allele. We set the K to assume all possible populations, that is, 1–8, we replicated each run 10 times and applied a random set of 3000 SNPs to initialize the cross-entropy algorithm. Outliers were identified considering a z-score distribution of correlation values after p-value adjustment, as suggested by Frichot et al. (2013). This approach has been shown to be potentially complementary to other correlation-based strategies, though with a higher-than-average number of positives (Nielsen et al., 2020). As such, we further controlled for false discovery rate (FDR) according to the Benjamini–Hochberg procedure with a q = 0.1 (Benjamini & Hochberg, 1995). To control for the spatial heterogeneity of both abiotic and biotic factors, we added latitude and longitude as variables in Bayenv2 and LFMM. Loci associated with latitude and longitude were removed from the candidate's list, independently of being associated with any other factor.

Lastly, we utilized *pcadapt* to screen for outlier loci. The number of PCs to retain was chosen accordingly to the scree plot curve and p-values were subject to Bonferroni correction with $\alpha = 0.001$.

Only loci identified as outliers in all detection methods were classified as candidates under selection and were used in further analyses. We calculated diversity estimates and inferred population structure with the same methodology applied to the full loci dataset.

2.7 | Candidate loci annotation

Genomic sequences flanking candidate loci were retrieved from the sea lamprey genome (P.marinus_7.0) at Ensembl's database by extending the size of the RADtag 1000bp upstream and downstream the coordinates to which they mapped against. Extending the RADtag alongside the reference genome was performed with the

intent of capturing regions potentially linked to the candidate loci. Fragments were blasted against the NCBI database.

3 | RESULTS

3.1 | Sequencing statistics and filtering outputs

Overall, mapping resulted in a mean percentage of 36% (SE = 0.3%) reads kept, representing a mean of 1.94×10^6 (SE = 596.66) reads per individual. Mean coverage per individual was 20.25× (SE = 0.79), with the mean number of genotyped sites per locus being 269.83 (SE = 0.24); a total number of SNPs 30,910 loci were retained for downstream analysis. The complete statistics of the filtering steps associated with the variant calling pipeline can be found in Text S1–S3.

3.2 | Genetic diversity estimates

Genetic diversity estimates were similar among sampled locations when considering all sequenced sites, with *Hs* varying between 0.10 for the Frome 0.12 for other 6 locations (Table 1). The number of private alleles had higher variation, with the highest number found in the Rolfsan (228) and the lowest in Gironde (77), with the difference being significant according to geographic distance ($R^2 = 0.18$, $p = 0.01$) (Figure 2a), where the higher the distance between sampling areas, the greater the differential of private allele numbers between the sampling areas. There was also a significant and positive linear relationship between the number of private alleles and latitude ($R^2 = 0.72$, $p = 0.04$) (Figure 2b), suggesting an overall pattern of increased unique genomic diversity at higher latitudes. A total of 2 individuals, one from Rolfsan and other from Gironde, did not pass the 0.1 kinship cut-off but were still maintained in the analyses as those do not belong to the same population.

3.3 | Population structure and connectivity among freshwater locations

A principal component analysis revealed some individuals associated with the locations of Frome and Rolfsan to be separated from the cloud of points representing all other samples (Figure S1). For the DAPC, 62 PCs were retained following Jombart (2008) with a mean a-score = 0.11. Exploring the number of clusters (1–8) with *find. cluster* function and subsequent ordinate plotting and considering function-defined groups, we observed that the assignment did not match sampling locations (Figure S2); when we defined grouping concordant with sampling locations, we observed a distinct separation between Rolfsan and all other locations (Figure 3). Pairwise F_{ST} values were also generally low and nonsignificant, except between the Rivers Frome and Rolfsan ($F_{ST} = 0.010$, $p = 0.001$), and Frome and Ouse ($F_{ST} = 0.006$, $p = 0.001$) (Table S2).

TABLE 1 Geographic information and diversity indices of sampled locations.

Location	FAO fishing area	Latitude	Longitude	N	Hs	pA
Tagus – PT	9 (9.a)	38.66	−9.30	20	0.12	98
Mondego – PT	9 (9.a)	40.14	−8.89	20	0.11	84
Ulla – SP	9 (9.a)	42.52	−8.95	22	0.12	102
Gironde – FRA	8 (8.b)	44.62	−1.31	18	0.12	77
Frome – UK	7 (7.e)	50.71	−2.06	34	0.10	178
Severn- UK	7 (7.f)	51.37	−4.10	23	0.12	186
Ouse – UK	4 (4.b)	53.62	−0.04	27	0.12	94
Rolfsan – Sweden	3 (3.a)	57.42	12.07	22	0.12	228

Note: Sampled sites are coded for the names used in the manuscript. FAO Fishing area codes relate to the following Northeast Atlantic zones: 9.a – Portuguese Waters East, 8.b – Bay of Biscay Central, 7.e – Western English Channel, 7.f – Bristol Channel, 4.b – Central North Sea, 3.a – Skagerrak.

Abbreviations: *Hs*, heterozygosity estimate; *n*, number of samples; *pA*, private alleles.

3.4 | Gene–environment associations and gene ontology of putative candidate loci

There were 118 loci with a Bayes factor > 10; among these, 42 also had an absolute value of $\rho > 0.20$. The *pcadapt* approach, which accounts for population differentiation while identifying outliers, detected 629 candidate loci across variables. LFMMs detected candidate loci, per variable, in the range of those detected by *pcadapt*. However, decreasing the adopted q (from the suggested $q=0.1$ to $q=10^{-3}$) to adjust for the false discovery rate did not significantly reduce putative outliers. Instead, it placed their number in the range of what was observed by Nielsen et al. (2020), that is, 1/5th of SNPs in the dataset were deemed outliers. Altogether a total of 11 unique candidate loci was shared among approaches, which the LFMM analyses distributed as follows: *maximum sea surface temperature, minimum nitrogen concentration, maximum dissolved oxygen, cod abundance*, and *twaite shad abundance* correlated with 1 locus each; *maximum phosphorus concentration, minimum phosphorus concentration* and *basking shark abundance* correlated with 2 loci while 3 loci *allis shad abundance* (Figure 4). Heterozygosity estimates of candidate loci within populations were substantially above those considering the full panel of SNPs and ranged between 0.11 in Ulla and 0.20 in Mondego (*t*-test: Hs_{mean} All = 0.12; Hs_{mean} candidate = 0.16; p = 0.002). We did not detect population structure associated with candidate loci allelic frequencies: overall F_{ST} of −0.001 and the nonidentification of clusters either with or without *a prior* population information inserted into DAPCs (Figure S4).

Blasting the 11 candidate loci against NCBI database revealed interesting patterns in light of the selective pressure against which their frequencies relate. For instance, candidate loci associated with biotic variables *allis shad abundance* and *Atlantic cod abundance* (1838 and 13,235) were found to be located in the genomic regions corresponding to the *jVLR* gene encoding for variable lymphocyte receptors, which allow lymphocytes to detect antigens in their environment (Li et al., 2013). Locus 1838 also aligned against the

SAFB-like transcription modulator, the *NF-kB* gene, and the Hox3 homeobox protein, features involved in transcription regulation and RNA binding, and cellular response to stress. Candidate 13,235 was also identified in a region of the sea lamprey genome's Zinc finger OZF-like protein. Lastly, locus 13,283 were identified as a candidate to *twaite shad abundance* to a membrane-associated RING-CH type 3 finger ubiquitin ligase and critical regulators of immune responses (Lin et al., 2019) (Figure 5).

In relation to the putative abiotic pressures, candidate locus 26,013 was associated with both *maximum phosphate concentration* and *maximum dissolved oxygen concentration* and mapped to a genomic region coding identified as the par-3 family cell polarity regulator beta (*pardb3* gene), a highly conserved feature across metazoan genomes involved in cell polarity and adhesion (Thompson, 2022); loci 13,513 and 14,785, which associated with *maximum phosphate concentration* and *maximum temperature* respectively, mapped against *P. marinus´* inositol-3-phosphate synthase 1 (*isyna-1* gene) a molecule involved in the production of myo-inositol in a wide range of organisms including fish (Agranoff, 2009) and to *P. marinus´* homeodomain *HOXW10A-like* gene and Tcl-1 like transposon, both previously identified to be associated with long-term plasticity of spinal cord locomotor network in sea lamprey (Bevan et al., 2008). Loci 18,413 and 27,929 were mapped against highly repetitive microsatellite regions (Table 2).

4 | DISCUSSION

Despite being a key species in the understanding of the evolution of vertebrates (Smith et al., 2013; Sower, 2018), there remains a paucity of information on how populations of sea lamprey evolved in natural habitats. With their abundance declining in freshwater locations alongside the eastern Atlantic shore (Hansen et al., 2016), understanding connectivity among spawning locations, as well as the evolutionary potential of the species, is a pressing issue towards

FIGURE 2 Relationships between location-specific alleles and geography. Panel a represents the relationship between the absolute number of private allele differences (x-axis) as a function of geographic distance (y-axis) between population pairs. Panel b investigates the relationship between latitude (x-axis) and the log-scaled number of private alleles (y-axis) with sampled locations marked in the dots. Smooth bands of the linear model log(pa) ~ latitude were applied to aid the visualization of the trends.
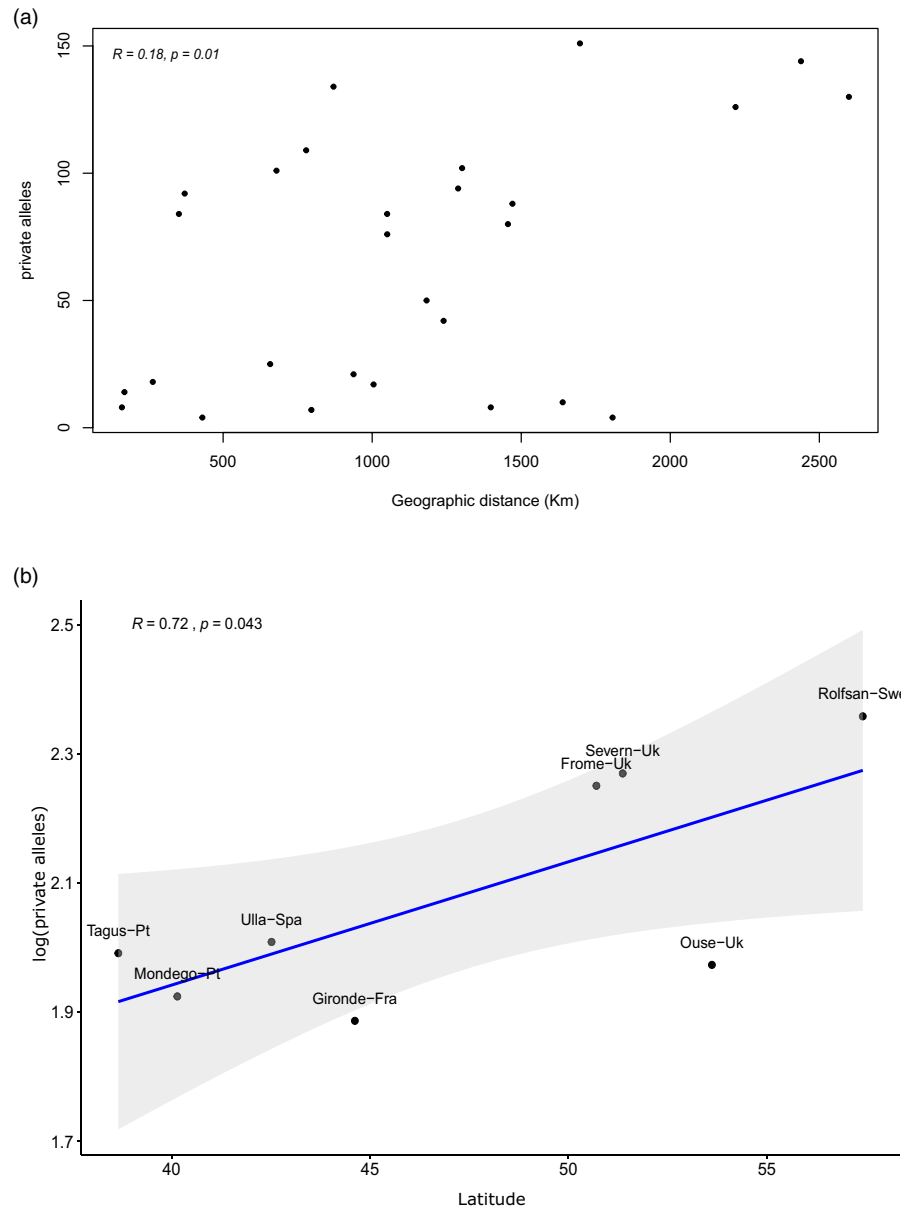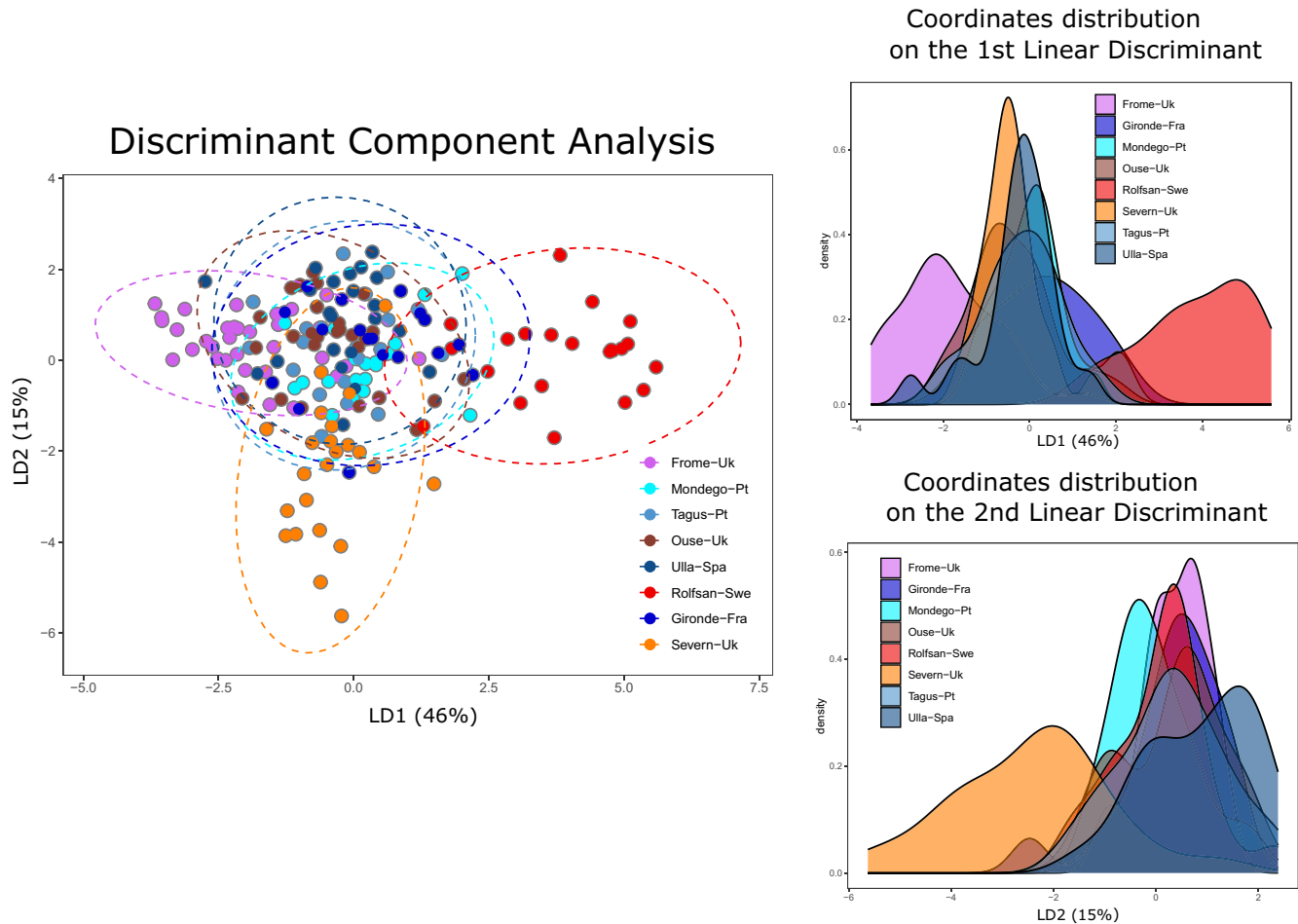

the conservation of this ancient organism. Here, we present the first genome-wide screen of sea lampreys´ genetic diversity, highlighting the microevolutionary processes governing the species' limited dispersal and to shed light on early stages of vertebrate evolution.

4.1 | Sea lampreys constitute a large eastern Atlantic metapopulation limited by northward dispersal

The absence of homing behavior towards natal spawning grounds renders sea lamprey an exception among diadromous species. Thus, the migratory behavior associated with the mandatory marine phase of sea lamprey would suggest a homogenization of genetic variance across the species' distribution range. It is not surprising then that patterns of panmixia have been reported among populations of sea lamprey on the western Atlantic coast (Waldman et al., 2008), on the

eastern Atlantic coast (Almada et al., 2008; Lança et al., 2014; Mateus et al., 2012; Rodríguez-Muñoz et al., 2004), as well as among populations of Pacific lamprey *Entosphenus tridentatus* (Spice et al., 2012). Despite the upscale to a genome-wide representation of molecular signatures, we observed a strikingly similar pattern of panmixia supported both by homogenized genetic diversity and almost complete absence of differentiation among sampled locations. However, our results did suggest that Atlantic sea lampreys may have limits to their dispersal, as also identified in their Pacific counterpart (Spice et al., 2012). Support for a limited dispersal scenario was provided by both the discriminant analyses on the whole SNP dataset, which pinpointed the north easternmost location of Rolfsan with some degree of differentiation in relation to all other locations, as well as a significant and positive correlation between the number of private alleles and latitude, which may be interpreted as a signature of limited connectivity with southern populations. The barrier imposed by the salinity gradient to other marine species across the Kattegat-Skagerrak

Discriminant Component Analysis

Coordinates distribution on the 1st Linear Discriminant

Coordinates distribution on the 2nd Linear Discriminant

FIGURE 3 Discriminant analyses of principal components (DAPC) with predefined population groups. Abstract distribution of individuals according to linear discriminant functions (1 and 2) that together captured 66% of genome-wide variation. Ellipses in dashed lines represent the 95% confidence interval of a normal distribution of PCs.

region, which defines the transition zone between the Northeast Atlantic and the Baltic Sea (Johannesson et al., 2020; Wennerström et al., 2013), is perhaps the most parsimonious explanation for the subtle differentiation and the high number of private alleles in the Rolfsan. In this context, the Rolfsan's lampreys might be part of a sub-population more likely to spawn within the Baltic Sea area, a scenario supported by scarce modern-era observations of sea lamprey spawning aggregations in the Vistula river, Bay of Gdansk, and Szczecin lagoon but not in German rivers (Thiel et al., 2009).

4.2 | Investigating dispersal limitations under the magnifying glass of molecular data

With generally low genome-wide differentiation over large spatial areas, allelic frequencies deviating from the neutral spectrum in marine populations have been increasingly associated with environmental cues characteristic of a heterogenous oceanic environment (Gagnaire & Gaggiotti, 2016). For example, seascape genomics have revealed that despite high levels of gene flow, temperature could explain distribution patterns in abalone (*Haliotis laevigata*), American

lobster (*Homarus americanus*) (Benestan et al., 2016) and cod (Clucas et al., 2019; Therkildsen et al., 2013), while salinity extended associations to herring (*Clupea harengus*) and cod (Berg et al., 2015; Guo et al., 2016). Sea lampreys have a very specialized set of life-history traits whereby the freshwater phase hinders a full-scale open ocean dispersal, as shown by the stark differentiation between Eastern and Western Atlantic coasts—$AMOVA_{FCT}=79.5\%$ between regions, $p<0.001$ (Genner et al., 2012), and their obligate parasitism could render their dispersal to be passively dependent on that of the hosts. Thus, it has been argued that limits to marine dispersal might be imposed both by abiotic and biotic factors (Guo et al., 2016).

By utilizing three different approaches to identify candidate loci under selection, we expected to mitigate issues related to high frequencies of false positives (type II errors), (Capblancq et al., 2018; Lotterhos & Whitlock, 2015; Narum & Hess, 2011; Nielsen et al., 2020), with the logical trade-off of increasing the rate of false negatives (Dalongeville et al., 2018).

Here, we report a relationship between the abundance of confirmed host species' such as the Atlantic cod, and the twaite shad, and allelic frequencies of loci located within the vicinity (maximum 1000 bp) of genomic regions annotated to immune-related

FIGURE 4 Bayesian clustering to explore the number of K in relation to sampling locations. Bayesian clustering generally suggests a high homogeny of admixture probabilities (y-axis) among sampling locations (x-axis), a pattern reproducible independently of the explored number of Ks.
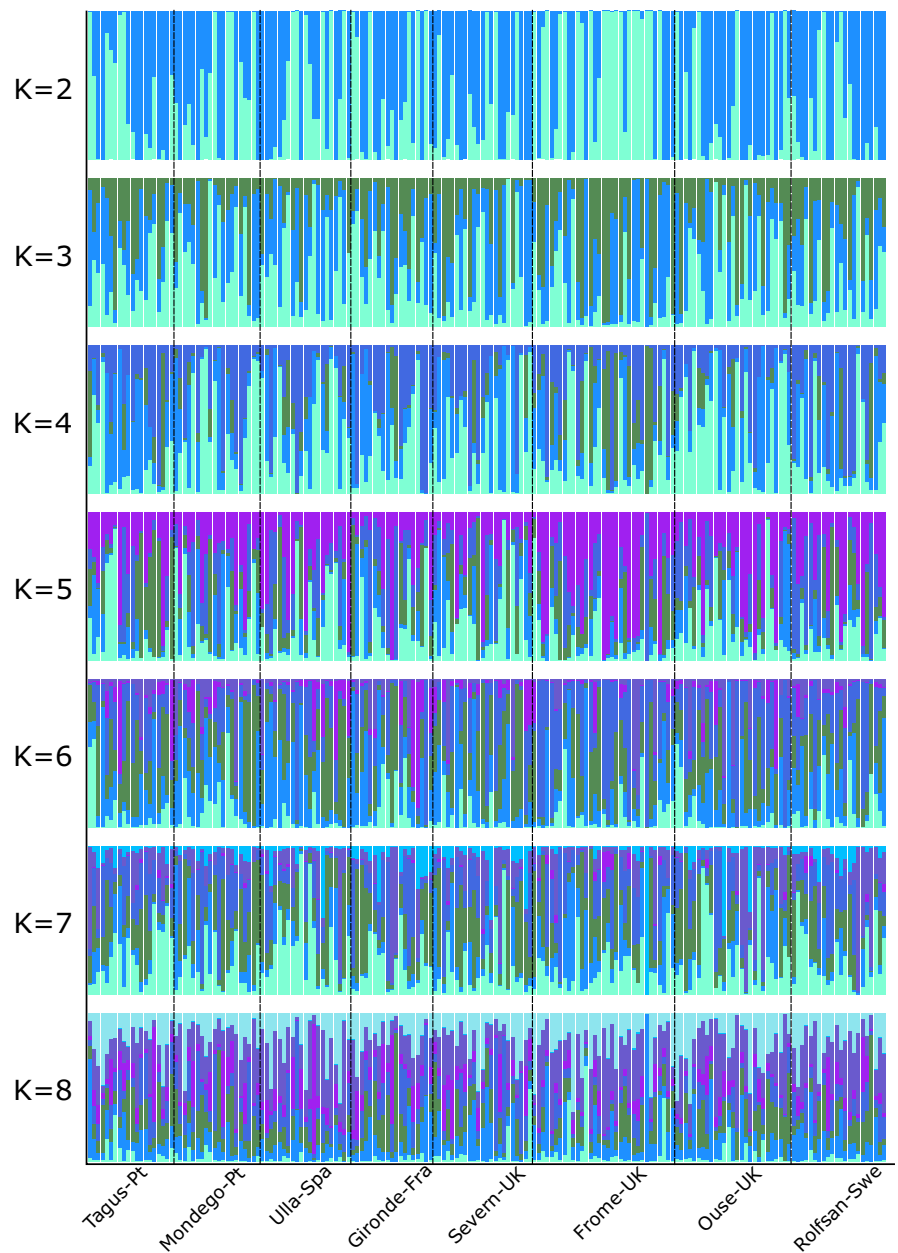
functions. Notably, variable lymphocyte receptors are an immune tool specific to jawless vertebrates, and RING-CH-type finger (MARCH) protein, is part of a family of E3 ubiquitin ligases also involved in the regulation of immune responses (Lin et al., 2019). Other genomic regions identified either on these loci or in other loci associated with these same biotic variables further support regulatory functions to be under selection in our target species. These results are suggestive of a hypothetical scenario of local adaptation where sea lampreys have evolved immune responses against the antigens of the most likely hosts that they will encounter at sea. It is known that sea lampreys elicit immune response from their hosts (Bullingham et al., 2022; Goetz et al., 2016), and thus it is plausible, at the light of host–parasite co-evolution, to expect adaptive signatures on the sea lamprey genome (Ebert & Fields, 2020; Eizaguirre et al., 2012). To the best of our knowledge, this is the first time that

genomic signatures of putative adaptation to host species have been shown in sea lamprey and consequently, at such an early stage of vertebrate adaptive immune system evolution (Boehm et al., 2012; Finstad & Good, 1964; Li et al., 2013). A plausible caveat to these observations relies on the accuracy of host abundances collected for this study. Food and Agricultural Organization fisheries datasets are based on capture and aquaculture data, and despite being extensively utilized to understand trends in fish abundance in the context of sustainable fisheries (Froese et al., 2012; Garibaldi, 2012; Naylor et al., 2021), their interpretation has historically experienced criticism (Daan et al., 2011; Pauly & Zeller, 2016). Alternatively, it could be the case that putative adaptive signatures reflect host–parasite co-evolution where the sea lamprey is the host. However, such scenario would imply the frequency distribution of a hypothetical sea lamprey parasite to co-vary with an abundance of sea

Candidate loci identified across environmental variables

Abiotic

Biotic
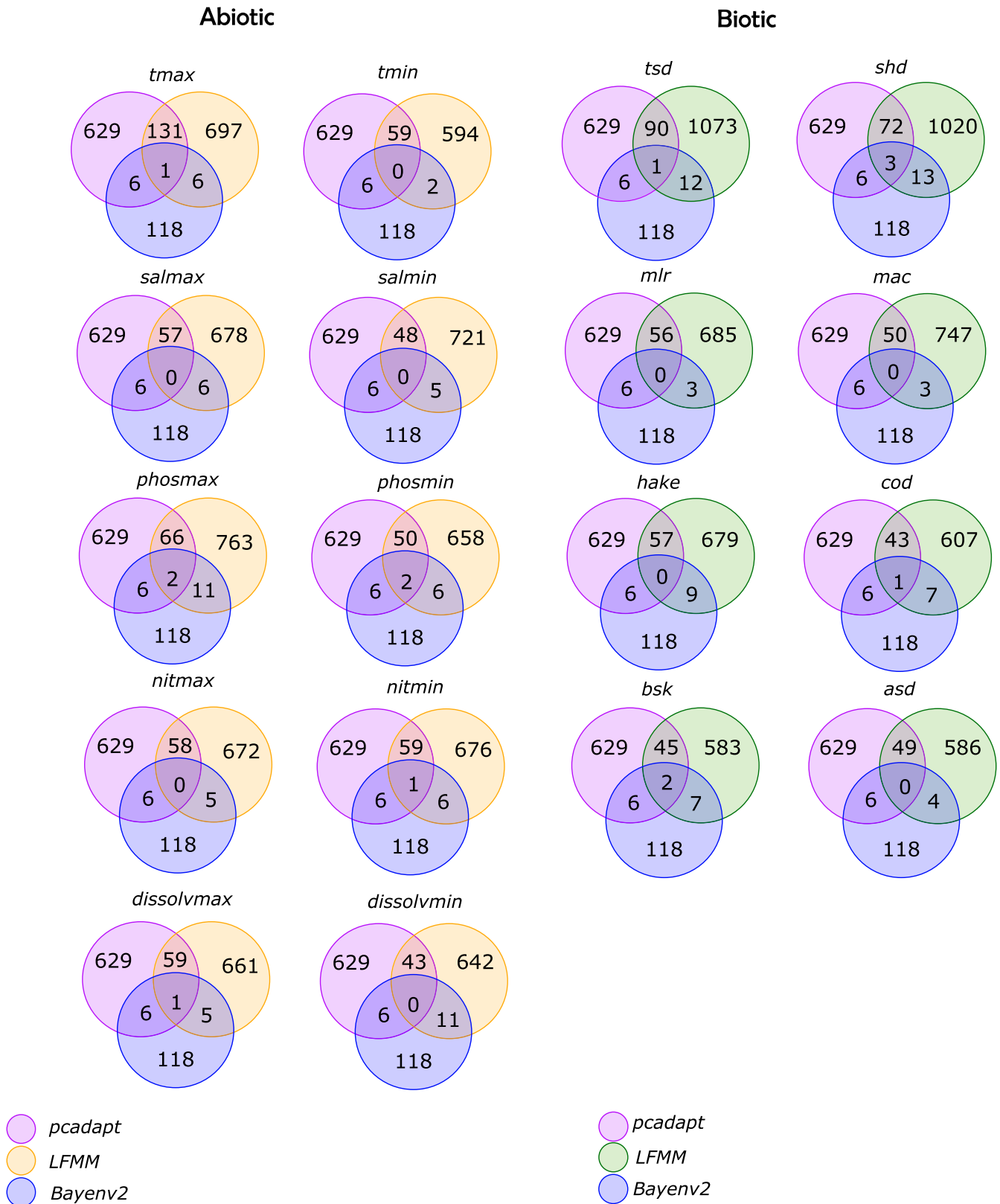
FIGURE 5 Venn diagram of candidate loci identified by multiple outlier detection techniques. Shown here is the number of candidate loci identified by each software. LFMM approach implemented in the R package *lea* is colored in yellow for the abiotic variables and in green for the biotic variables.

TABLE 2 Annotations of candidate loci against NCBI database.

Candidate locus	Environmental variable	Description
1838	*Allis shad abundance*	*L. camtschaticum jVLR* gene for variable lymphocyte receptor
		P. marinus SAFB-like transcription modulator
		Homeobox protein Hox-delta3
		E. morii NF-KB gene, promoter region
		P. marinus clone gVLR_Contig variable lymphocyte receptor (*VLR*)
3235	*Minimum phosphorus concentration*	*P. marinus* serine/threonine-protein phosphatase 4 regulatory subunit 3A-like
13,235	*Atlantic cod abundance*	*P. marinus* zinc finger protein OZF-like
		L. planeri variable lymphocyte receptor B
13,283	*Twaite shad abundance*	*P. marinus* membrane-associated ring-CH-type finger 3
13,513	*Maximum phosphorus concentration*	*P. marinus* inositol-3-phosphate synthase 1
		P. marinus homeodomain *HOXW10A-like* gene and Tcl-1-like transposon
14,785	*Maximum sea surface temperature*	*P. marinus* inositol-3-phosphate synthase 1
		P. marinus homeodomain *HOXW10A-like* gene and Tcl-1-like transposon, complete sequence
15,374	*Allis shad abundance*	*L. camtschaticum* clone BAC 37B15 homeobox protein Hox-eta4 (Hox-eta4) and wingless-type MMTV integration site family, member eta (Wnt-eta) genes
		P. marinus lamprin L-0.9 (*L-0.9*) gene
		L. camtschaticum microsatellite
18,413	*Basking shark abundance/ minimum phosphorus concentration*	*P. marinus* microsatellite
		L. camtschaticum microsatellite
23,021	*Allis shad abundance/minimum nitrogen concentration*	*P. marinus* DEAD-box helicase 23
26,013	*Maximum phosphorus concentration/maximum dissolved oxygen concentration*	*P. marinus* par-3 family cell polarity regulator beta
		P. marinus par-3 family cell polarity regulator beta
27,929	*Basking shark*	*P. marinus* microsatellite
		L. camtschaticum microsatellite

Note: Identified candidate loci blasted against NCBI database. The Full details of alignment length, *e*-value, and percentage similarity can be found in Table S4.

lamprey hosts (here investigated) across spatial scales. As in the case of sea lamprey being the parasite, data on sea lamprey's parasites is scarce and the functional link to fitness remains elusive (Shavalier et al., 2021).

From an abiotic perspective, we find it interesting that loci associated with *maximum phosphate concentration* and *maximum temperature* relate to a genomic region coding for the *isyna-1* gene, involved in the myo-inositol (formerly referred to as vitamin B_8) pathway. The biosynthesis of myo-inositol is an evolutionary conserved pathway (Majumder et al., 2003), known to modulate several vital physiological functions in living organisms, such as growth, immune efficiency, osmoregulation, and exposure to heavy metal concentrations (Cui et al., 2022). Interestingly, this locus has been identified as a candidate under selection also in the sand goby (*Pomatoschistus minutus*) in the context of adaptations to less saline conditions after postglacial expansion (Leder et al., 2021). Because the same candidate loci also identified other genomic features such as the *HOXW10A-like* gene and the Tcl-1-like transposon (both involved in highly

conserved mechanisms of homeostasis and plasticity of spinal cord locomotor network, respectively) it is suggested that the selection for organismal balance is plausible in the context of the heterogenous environment that sea lamprey experiences during its life cycle. While the fresh- to saltwater transition is generally important, our results suggest also the importance of the abiotic environment and putative anthropogenic impacts, such as phosphate concentration (Farmer, 2018), which along with dissolved oxygen concentration, is critical for the thermal regulation and metabolic processes of aquatic ectotherms (Pörtner & Knust, 2007). Temperature has also been shown multiple times to be a selective pressure in the marine environment (Conover et al., 2006; Pörtner et al., 2001). Nevertheless, we are unable to provide evidence to support the functionality of the associations reported in our study. In a broader context, we argue that these patterns are likely to relate to spatially varying selection, a scenario previously reported in the American eel (*Anguilla rostrata*), a catadromous species with an equally diverse distribution range and complex life cycle (Gagnaire et al., 2012).

4.3 | Concluding remarks

We show that sea lampreys in the Eastern Atlantic coast constitute a nearly-panmictic population, with North Sea imposing some limits to the species' dispersal capacity. Within the limitations of our sampling strategy, we support the idea that the North Sea-Baltic Sea transition can be a reservoir of unique sea lamprey diversity (whether it be genetic, life-history-based, etc), and thus monitoring and conservation of this anadromous species should be integrated into activities that aim to preserve biodiversity in this region. Regarding the factors mediating the dispersal ability of a species without homing behavior, we showed one possibility would be spatially varying selection shaping the adaptive seascape across sea lamprey distribution.

While we find particularly interesting the association between candidate loci mapping to jawless-specific immune genes and host abundances, we acknowledge more work is necessary to further validate these results. Nevertheless, these observations could provide the background to investigate how host–parasite co-evolution could have mediated the evolution of earlier versions of vertebrate adaptive immune repertoire under natural contemporary conditions.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

Raw RADseq reads will be deposited in NCBI's Short Read Archive (SRA). All scripts utilized are available at: https://github.com/Miguel-BSoares/SeaLamprey_POPGEN.

ORCID

Miguel Baltazar-Soares https://orcid.org/0000-0002-5919-6197
J. Robert Britton https://orcid.org/0000-0003-1853-3086
Victoria Dominguez Almela https://orcid.org/0000-0003-4877-5967

REFERENCES

Agranoff, B. W. (2009). Turtles all the way: Reflections on myo-inositol. *Journal of Biological Chemistry*, *284*(32), 21121–21126.

Almada, V. C., Pereira, A. M., Robalo, J. I., Fonseca, J. P., Levy, A., Maia, C., & Valente, A. (2008). Mitochondrial DNA fails to reveal genetic structure in sea-lampreys along European shores. *Molecular Phylogenetics and Evolution*, *46*, 391–396.

Antognazza, C. M., Britton, R. J., Read, D. S., Goodall, T., Mantzouratou, A., De Santis, V., Davies, P., Aprahamian, M., Franklin, E., Hardouin, E. A., & Andreou, D. (2021). Application of eDNA metabarcoding in a fragmented lowland river: Spatial and methodological comparison of fish species composition. *Environmental DNA*, *3*(2), 458–471.

Asaduzzaman, M., Igarashi, Y., Wahab, M. A., Nahiduzzaman, M., Rahman, M. J., Phillips, M. J., Huang, S., Asakawa, S., Rahman, M. M., & Wong, L. L. (2020). Population genomics of an anadromous Hilsa Shad Tenualosa ilisha species across its diverse migratory habitats: Discrimination by fine-scale local adaptation. *Genes*, *11*(1), 46.

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2018). Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3), 277–284.

Beamish, F. (1980). Biology of the north American anadromous sea lamprey, *Petromyzon marinus*. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*(11), 1924–1943.

Benestan, L. M., Ferchaud, A. L., Hohenlohe, P. A., Garner, B. A., Naylor, G. J., Baums, I. B., Schwartz, M. K., Kelley, J. L., & Luikart, G. (2016). Conservation genomics of natural and managed populations: Building a conceptual and practical framework. *Molecular Ecology*, *25*(13), 2967–2977.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., Jakobsen, K. S., & André, C. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution*, *7*(6), 1644–1663.

Bergstedt, R. A., & Seelye, J. G. (1995). Evidence for lack of homing by sea lampreys. *Transactions of the American Fisheries Society*, *124*(2), 235–239.

Bevan, S., Vakharia, V., & Parker, D. (2008). Changes in gene expression and integrin-mediated structural changes are associated with long-term plasticity of a spinal cord locomotor network. *Neuroscience*, *152*(1), 160–168.

Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., & Cooper, M. D. (2012). VLR-based adaptive immunity. *Annual Review of Immunology*, *30*, 203–220.

Boulêtreau, S., Carry, L., Meyer, E., Filloux, D., Menchi, O., Mataix, V., & Santoul, F. (2020). High predation of native sea lamprey during spawning migration. *Scientific Reports*, *10*(1), 1–9.

Branco, P., Segurado, P., Santos, J., Pinheiro, P., & Ferreira, M. (2012). Does longitudinal connectivity loss affect the distribution of freshwater fish? *Ecological Engineering*, *48*, 70–78.

Bullingham, O. M., Firkus, T. J., Goetz, F. W., Murphy, C. A., & Alderman, S. L. (2022). Lake charr (Salvelinus namaycush) clotting response may act as a plasma biomarker of sea lamprey (*Petromyzon marinus*) parasitism: Implications for management and wound assessment. *Journal of Great Lakes Research*, *48*(1), 207–218.

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications*, *12*(10), 1971–1987.

Conover, D., Clarke, L., Munch, S., & Wagner, G. (2006). Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, *69*, 21–47.

Cui, W., Ma, A., Farhadi, A., Saqib, H. S. A., Liu, S., Chen, H., & Ma, H. (2022). How myo-inositol improves the physiological functions of aquatic animals: A review. *Aquaculture*, *553*, 738118. https://doi.org/10.1016/j.aquaculture.2022.738118

Daan, N., Gislason, H., Pope, J. G., & Rice, J. C. (2011). Apocalypse in world fisheries? The reports of their death are greatly exaggerated. *ICES Journal of Marine Science*, *68*(7), 1375–1378.

Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S., & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC Genomics*, *19*, 1–13.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(2), 163–182. https://doi.org/10.1017/S1464793105006950

Ebert, D., & Fields, P. D. (2020). Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, *21*(12), 754–768.

Eizaguirre, C., & Baltazar-Soares, M. (2014). Evolutionary conservation— Evaluating the adaptive potential of species. *Evolutionary Applications*, *7*(9), 963–967. https://doi.org/10.1111/eva.12227

Eizaguirre, C., Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters*, *15*(7), 723–731. https://doi.org/10.1111/j.1461-0248.2012.01791.x

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, *10*(3), 564–567.

FAO. (2016). Fisheries and aquaculture software. In *FishStatJ – software for fishery and aquaculture statistical time series*. FAO.

Farmer, A. (2018). Phosphate pollution: A global overview of the problem. In C. Schaum (Ed.), *Phosphorus: Polluter and resource of the future—Removal and recovery from wastewater* (pp. 35–55). Christian Schaum, IWA Publishing.

Ferreira-Martins, D., Coimbra, J., Antunes, C., & Wilson, J. (2016). Effects of salinity on upstream-migrating, spawning sea lamprey, *Petromyzon marinus*. *Conservation Physiology*, *4*(1), cov064.

Finstad, J., & Good, R. A. (1964). The evolution of the immune response: III. Immunologic responses in the lamprey. *The Journal of Experimental Medicine*, *120*(6), 1151–1168.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699.

Froese, R., Zeller, D., Kleisner, K., & Pauly, D. (2012). What catch data can tell us about the status of global fisheries. *Marine Biology*, *159*(6), 1283–1292.

Gagnaire, P.-A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology*, *62*(6), 603–616. https://doi.org/10.1093/cz/zow088

Gagnaire, P. A., Normandeau, E., Cote, C., Hansen, M. M., & Bernatchez, L. (2012). The genetic consequences of spatially varying selection in the panmictic American eel (*Anguilla rostrata*). *Genetics*, *190*(2), 725–U703. https://doi.org/10.1534/genetics.111.134825

Garibaldi, L. (2012). The FAO global capture production database: A six-decade effort to catch the trend. *Marine Policy*, *36*(3), 760–768.

Genner, M. J., Hillman, R., McHugh, M., Hawkins, S. J., & Lucas, M. C. (2012). Contrasting demographic histories of European and north American sea lamprey (*Petromyzon marinus*) populations inferred from mitochondrial DNA sequence variation. *Marine and Freshwater Research*, *63*(9), 827–833.

Goetz, F., Smith, S. E., Goetz, G., & Murphy, C. A. (2016). Sea lampreys elicit strong transcriptomic responses in the lake trout liver during parasitism. *BMC Genomics*, *17*(1), 1–16.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M., Meng, J., Mulligan, M., … Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221.

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

Guo, B., Li, Z., & Merilä, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular Ecology*, *25*(12), 2833–2852.

Guo, D., & Britton, J. (2017). Sea lamprey *Petromyzon marinus* biology and management across their native and invasive ranges: Promoting conservation by knowledge transfer. *Reviews in Fisheries Science & Aquaculture*, *25*(1), 84–99.

Hansen, M. J., Madenjian, C. P., Slade, J. W., Steeves, T. B., Almeida, P. R., & Quintella, B. R. (2016). Population ecology of the sea lamprey (*Petromyzon marinus*) as an invasive species in the Laurentian Great Lakes and an imperiled species in Europe. *Reviews in Fish Biology and Fisheries*, *26*(3), 509–535.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Johannesson, K., Le Moan, A., Perini, S., & André, C. (2020). A Darwinian laboratory of multiple contact zones. *Trends in Ecology & Evolution.*, *35*, 1021–1036.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Kapralova, K., Morrissey, M., Kristjánsson, B., Ólafsdóttir, G. Á., Snorrason, S., & Ferguson, M. (2011). Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity*, *106*(3), 472–487.

Lança, M. J., Machado, M., Mateus, C. S., Lourenço, M., Ferreira, A. F., Quintella, B. R., & Almeida, P. R. (2014). Investigating population structure of sea lamprey (*Petromyzon marinus* L.) in western Iberian Peninsula using morphological characters and heart fatty acid signature analyses. *PLoS One*, *9*(9), e108110.

Leder, E. H., André, C., le Moan, A., Töpel, M., Blomberg, A., Havenhand, J. N., Lindström, K., Volckaert, F. A. M., Kvarnemo, C., Johannesson, K., & Svensson, O. (2021). Post-glacial establishment of locally adapted fish populations over a steep salinity gradient. *Journal of Evolutionary Biology*, *34*(1), 138–156.

Leitwein, M., Garza, J. C., & Pearse, D. E. (2017). Ancestry and adaptive evolution of anadromous, resident, and adfluvial rainbow trout (*Oncorhynchus mykiss*) in the San Francisco bay area: Application of adaptive genomic variation to conservation in a highly impacted landscape. *Evolutionary Applications*, *10*(1), 56–67.

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv preprint arXiv:1303.3997.

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, J., Das, S., Herrin, B. R., Hirano, M., & Cooper, M. D. (2013). Definition of a third VLR gene in hagfish. *Proceedings of the National Academy of Sciences*, *110*(37), 15013–15018.

Limburg, K. E., & Waldman, J. R. (2009). Dramatic declines in North Atlantic diadromous fishes. *Bioscience*, *59*(11), 955–965. https://doi.org/10.1525/bio.2009.59.11.7

Lin, H., Li, S., & Shu, H.-B. (2019). The membrane-associated MARCH E3 ligase family: Emerging roles in immune regulation. *Frontiers in Immunology*, *10*. https://doi.org/10.3389/fimmu.2019.01751

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*(5), 1031–1046.

Majumder, A. L., Chatterjee, A., Dastidar, K. G., & Majee, M. (2003). Diversification and evolution of L-myo-inositol 1-phosphate synthase. *FEBS Letters*, 553(1–2), 3–10.

Mateus, C. S., Almeida, P. R., Quintella, B. R., & Alves, M. J. (2011). MtDNA markers reveal the existence of allopatric evolutionary lineages in the threatened lampreys *Lampetra fluviatilis* (L.) and *Lampetra planeri* (Bloch) in the Iberian glacial refugium. *Conservation Genetics*, 12(4), 1061–1074.

Mateus, C. S., Rodríguez-Muñoz, R., Quintella, B. R., Alves, M. J., & Almeida, P. R. (2012). Lampreys of the Iberian Peninsula: Distribution, population status and conservation. *Endangered Species Research*, 16(2), 183–198.

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, 7, 999–1007. https://doi.org/10.1111/eva.12193

Moser, M. L., Almeida, P. R., King, J. J., & Pereira, E. (2020). Passage and freshwater habitat requirements of anadromous lampreys: Considerations for conservation and control. *Journal of Great Lakes Research*, 47, S147–S158. https://doi.org/10.1016/j.jglr.2020.07.011

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11, 184–194.

Naylor, R. L., Hardy, R. W., Buschmann, A. H., Bush, S. R., Cao, L., Klinger, D. H., Little, D. C., Lubchenco, J., Shumway, S. E., & Troell, M. (2021). A 20-year retrospective review of global aquaculture. *Nature*, 591(7851), 551–563.

Nielsen, E. S., Henriques, R., Beger, M., Toonen, R. J., & von der Heyden, S. (2020). Multi-model seascape genomics identifies distinct environmental drivers of selection among sympatric marine species. *BMC Evolutionary Biology*, 20(1), 121. https://doi.org/10.1186/s12862-020-01679-4

Ormerod, S. J., Dobson, M., Hildrew, A. G., & Townsend, C. R. (2010). Multiple stressors in freshwater ecosystems. *Freshwater Biology*, 55, 1–4. https://doi.org/10.1111/j.1365-2427.2009.02395.x

Pauly, D., & Zeller, D. (2016). Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining. *Nature Communications*, 7(1), 1–9.

Perrier, C., Le Gentil, J., Ravigne, V., Gaudin, P., & Salvado, J.-C. (2014). Origins and genetic diversity among Atlantic salmon recolonizing upstream areas of a large south European river following restoration of connectivity and stocking. *Conservation Genetics*, 15(5), 1095–1109.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135.

Pörtner, H. O., Berdal, B., Blust, R., Brix, O., Colosimo, A., De Wachter, B., Giuliani, A., Johansen, T., Fischer, T., Knust, R., Lannig, G., Naevdal, G., Nedenes, A., Nyhammer, G., Sartoris, F. J., Serendero, I., Sirabella, P., Thorkildsen, S., & Zakhartsev, M. (2001). Climate induced temperature effects on growth performance, fecundity and recruitment in marine fish: Developing a hypothesis for cause and effect relationships in Atlantic cod (Gadus morhua) and common eelpout (Zoarces viviparus). *Continental Shelf Research*, 21(18–19), 1975–1997.

Pörtner, H. O., & Knust, R. (2007). Climate change affects marine fishes through the oxygen limitation of thermal tolerance. *Science*, 315(5808), 95–97.

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., Hotaling, T. J., Spidle, A. P., & Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, 3(8), e1603198.

Privé, F., Luu, K., Vilhjálmsson, B. J., & Blum, M. G. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution*, 37(7), 2153–2154.

Quintella, B., Clemens, B., Sutton, T., Lança, M., Madenjian, C., Happel, A., & Harvey, C. (2021). At-sea feeding ecology of parasitic lampreys. *Journal of Great Lakes Research.*, 47, S72–S89.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589.

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M. N., & Replogle, J. M. (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*, 344, 519–523. https://doi.org/10.1126/science.1249547

Reid, S. B., & Goodman, D. H. (2020). Natural recolonization by Pacific lampreys in a Southern California coastal drainage: Implications for their biology and conservation. *North American Journal of Fisheries Management*, 40(2), 335–341.

Renaud, C. B., & Cochran, P. A. (2019). Post-metamorphic feeding in lampreys. In *Lampreys: Biology, conservation and control* (pp. 247–285). Springer.

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12), 2640–2659.

Rodríguez-Muñoz, R., Waldman, J., Grunwald, C., Roy, N., & Wirgin, I. (2004). Absence of shared mitochondrial DNA haplotypes between sea lamprey from north American and Spanish rivers. *Journal of Fish Biology*, 64(3), 783–787.

Roesti, M., Salzburger, W., & Berner, D. (2012). Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, 12, 1–7.

Shavalier, M. A., Faisal, M., Moser, M. L., & Loch, T. P. (2021). Parasites and microbial infections of lamprey (order Petromyzontiformes Berg 1940): A review of existing knowledge and recent studies. *Journal of Great Lakes Research*, 47, S90–S111.

Silva, S., Araújo, M. J., Bao, M., Mucientes, G., & Cobo, F. (2014). The haematophagous feeding stage of anadromous populations of sea lamprey *Petromyzon marinus*: Low host selectivity and wide range of habitats. *Hydrobiologia*, 734(1), 187–199.

Silva, S., Servia, M. J., Vieira-Lanero, R., Barca, S., & Cobo, F. (2013). Life cycle of the sea lamprey *Petromyzon marinus*: Duration of and growth in the marine life stage. *Aquatic Biology*, 18(1), 59–62.

Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., Bloom, O. E., Morgan, J. R., Buxbaum, J. D., Sachidanandam, R., Sims, C., Garruss, A. S., Cook, M., Krumlauf, R., Wiedemann, L. M., Sower, S. A., … Weiming, L. (2013). Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nature Genetics*, 45(4), 415–421.

Sorensen, P. W., Fine, J. M., Dvornikovs, V., Jeffrey, C. S., Shao, F., Wang, J., Vrieze, L. A., Anderson, K. R., & Hoye, T. R. (2005). Mixture of new sulfated steroids functions as a migratory pheromone in the sea lamprey. *Nature Chemical Biology*, 1(6), 324–328.

Sower, S. A. (2018). Landmark discoveries in elucidating the origins of the hypothalamic-pituitary system from the perspective of a basal vertebrate, sea lamprey. *General and Comparative Endocrinology*, 264, 3–15. https://doi.org/10.1016/j.ygcen.2017.10.016

Spice, E. K., Goodman, D. H., Reid, S. B., & Docker, M. F. (2012). Neither philopatric nor panmictic: Microsatellite and mtDNA evidence suggests lack of natal homing but limits to dispersal in Pacific lamprey. *Molecular Ecology*, 21(12), 2916–2930.

Sunga, J., Wilson, J. M., & Wilkie, M. P. (2020). Functional re-organization of the gills of metamorphosing sea lamprey (*Petromyzon marinus*): Preparation for a blood diet and the freshwater to seawater transition. *Journal of Comparative Physiology B*, 190(6), 701–715.

Therkildsen, N. O., Hemmer-Hansen, J., Hedeholm, R. B., Wisz, M. S., Pampoulie, C., Meldrup, D., Bonanomi, S., Retzel, A., Olsen, S. M., & Nielsen, E. E. (2013). Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evolutionary Applications*, 6(4), 690–705. https://doi.org/10.1111/eva.12055

Thiel, R., Winkler, H. M., Riel, P., Neumann, R., Gröhsler, T., Böttcher, U., Spratte, S., & Hartmann, U. (2009). Endangered anadromous lampreys in the southern Baltic Sea: Spatial distribution, long-term trend, population status. *Endangered Species Research*, 8(3), 233–247.

Thompson, B. J. (2022). Par-3 family proteins in cell polarity & adhesion. *The FEBS Journal*, 289(3), 596–613.

Turbek, S. P., Funk, W. C., & Ruegg, K. C. (2023). Where to draw the line? Expanding the delineation of conservation units to highly mobile taxa. *Journal of Heredity*, esad011. https://doi.org/10.1093/jhered/esad011

Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., & De Clerck, O. (2012). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, 21(2), 272–281.

Van Drecht, G., Bouwman, A. F., Harrison, J., & Knoop, J. M. (2009). Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050. *Global Biogeochemical Cycles*, 23, GB0A03. https://doi.org/10.1029/2009GB003458

Vrieze, L. A., Bergstedt, R. A., & Sorensen, P. W. (2011). Olfactory-mediated stream-finding behavior of migratory adult sea lamprey (*Petromyzon marinus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 68(3), 523–533.

Waldman, J., Grunwald, C., & Wirgin, I. (2008). Sea lamprey *Petromyzon marinus*: An exception to the rule of homing in anadromous fishes. *Biology Letters*, 4(6), 659–662.

Waples, R. S., & Lindley, S. T. (2018). Genomics and conservation units: The genetic basis of adult migration timing in Pacific salmonids. *Evolutionary Applications*, 11(9), 1518–1526.

Ward, D. L., Clemens, B. J., Clugston, D., Jackson, A. D., Moser, M. L., Peery, C., & Statler, D. P. (2012). Translocating adult Pacific lamprey within the Columbia River basin: State of the science. *Fisheries*, 37(8), 351–361.

Wennerström, L., Laikre, L., Ryman, N., Utter, F. M., Ghani, N. I. A., André, C., DeFaveri, J., Johansson, D., Kautsky, L., Merilä, J., Mikhailova, N., Pereyra, R., Sandström, A., Teacher, A. G. F., Wenne, R., Vasemägi, A., Zbawicka, M., Johannesson, K., & Primmer, C. R. (2013). Genetic biodiversity in the Baltic Sea: Species-specific patterns challenge management. *Biodiversity and Conservation*, 22(13), 3045–3065.

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, 30(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Youson, J. H., & Potter, I. C. (1979). A description of the stages in the metamorphosis of the anadromous sea lamprey, *Petromyzon marinus* L. *Canadian Journal of Zoology*, 57(9), 1808–1817. https://doi.org/10.1139/z79-235

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.