**Neutral and adaptive loci reveal fine-scale population structure in *Eleginops maclovinus* from North Patagonia**

**Abstract**

Patagonia is an understudied area, especially when it comes to population genomic studies with relevance to fishery management. However, the dynamic and heterogeneous landscape in this area can harbor important but cryptic genetic population structure. Once such information is revealed, it can be integrated into the management of infrequently investigated species. *Eleginops maclovinus* is a protandrous hermaphrodite species with economic importance for local communities that is currently managed as a single genetic unit. In this study, we sampled five locations distributed across a salinity cline from Northern Patagonia to investigate the genetic population structure of *E*. *maclovinus*. We use Restriction-site Associated DNA (RAD) sequencing and outlier tests to obtain 11,712 neutral, 338 adaptive loci, and 78 loci associated to environmental variables. We identified a spatial pattern of structuration with gene flow and spatial selection by environmental association. Neutral and adaptive loci showed two and three genetic groups, respectively. The effective population sizes estimated ranged from 572 to 14,454 and were influenced more by locality than salinity cline. We found loci putatively associated with salinity suggesting that salinity may act as a selective driver in *E*. *maclovinus* populations. These results suggest a complex interaction between genetic drift, geneflow, and natural selection in this area. Our findings suggest that at least two to three genetic units exist in this area, and the information should be integrated into the management of this species. We discuss the significance of these results for fishery management and suggest future directions to improve our understanding of how *E*. *maclovinus* is adapted to the dynamic waters of Northern Patagonia.

**Keywords**: estuary; fjords; Notothenioidei; Patagonian blennie; protandrous hermaphrodite; salinity cline; SNPs

**Introduction**

Advances in genome sampling methods have reduced complexity (e.g., Restriction associated DNA sequencing, RADseq) and allowed the collection of an unusual amount of data to analyze the genome concerning conservation and management problems (Bernatchez *et al.* 2017; Xuereb *et al.* 2021). These data provide a great way to solve unanswered questions and have the advantage of allowing the quantification of adaptive variation comparing, for instance, microsatellites (Funk *et al.* 2012; Bernatchez *et al.* 2017). Currently researchers can differentiate neutral and adaptive variation across populations and incorporate this information into programs to obtain better management and conservation solutions (Xuereb *et al.* 2021).

Identification of evolutionary significant units (ESU) is important to guide management and conservation efforts (Funk *et al.* 2012) and maximize the evolutionary potential for environmental change (Bernatchez 2016). At fine-scale, a management unit, which is included in an ESU, refers to demographically independent populations and shows significant divergence results with low gene flow (Funk *et al.* 2012; von der Heyden 2017). Finding genetic differences at fine-scale is challenging because it depends on the biological characteristics of the organisms studied (e.g., vagility) and the geomorphological conformation and environmental heterogeneity of their geographical distribution (Jørgensen *et al.* 2005; Canales-Aguirre *et al.* 2016). For the latter, habitats such as fjords may greatly affect population genetic diversity in marine organisms due to unique environmental characteristics.

Patagonia in Chile includes a vast coastal area (240 000 km2; Pantoja *et al.* 2011). The northern region extends from latitude 41.5ºS (Reloncaví Fjord) to latitude 46.5ºS (San Rafael Lagune; Rodrigo 2008) with high ecosystem productivity and heterogeneous geomorphological and physical-chemical oceanographic conditions (Pérez-Santos *et al.* 2014; Ríos *et al.* 2016; Yevenes *et al.* 2017). For example, Patagonia has a saline cline pattern resulting from freshwater runoff from melting ice from the Andean Mountains (annual average caudal greater than 300 (m3/s)) and an annual average precipitation greater than 1000 mm (Garreaud *et al.* 2013). Salinity has been associated with population genetic structure in the marine realm (McCairns & Bernatchez 2008; Limborg *et al.* 2009; Berg *et al.* 2015) and can play an important role in the ontogeny of different organisms from early to adult stages. Salinity determines the vertical distribution of early stages (i.e., eggs and yolk sac larvae (Petereit *et al.* 2009), influences abundance as a result of freshwater discharges (Ojaveer & Kalejs 2010), can model the genetic diversity in space (McCairns & Bernatchez 2008), and can adapt to divergent salinity environments (Berg *et al.* 2015). Such landscape characteristics influence a high diversity of marine organisms and hierarchical levels, from populations to ecosystems (Olsen *et al.* 2002; Beuchel *et al.* 2006; Kristoffersen & Salvanes 2009; Canales-Aguirre *et al.* 2010, 2016), that support and sustain economically important fisheries. Unfortunately, this area has been understudied, and no population genomics studies with relevance to fisheries management have been conducted despite the fact that this unique landscape may promote large genetic differentiation and local adaptation. Genomic studies can reveal discrete genetic groups that can be integrated into future conservation and management programs (e.g., Larson *et al.* 2014b; a; McKinney *et al.* 2017a; Euclide *et al.* 2021).

*Eleginops maclovinus* (Cuvier and Valenciennes 1830) is a monotypic species and one of the few species of the Notothenioidei with a non-Antarctic distribution (Bargelloni *et al.* 2000; Near 2004; Matschiner *et al.* 2015). Endemic to South America, it is found close to estuaries from ~33°S on both sides of the Pacific and the Atlantic Ocean to the Beagle Channel (~54°S) (Pequeño 1989), including in the Malvinas/Falkland Islands (Gosztonyi 1974). This species is economically important for local communities and is caught by artisanal and recreational fishers (Gastaldi *et al.* 2009), with 182 tons landed in 2019 in Chile (Sernapesca 2019). In Chilean Patagonia, there is little scientific information about this fish that can be used to identify appropriate management units. *Eleginops maclovinus* is a partial spawner with a spawning peak during late autumn in estuaries of the Chilean coast (Ruiz 1993); it is also the most fecund of the Notothenioidei (~550,000 eggs per female, Gosztonyi 1979). This species is a protandrous hermaphrodite, and males are often smaller in length than females (10-52 cm males and >53cm females) (Brickle *et al.* 2005).

Connectivity and dispersal between populations of *E*. *maclovinus* seem to be biased by sex. Mechanical tags from females have been found up to 60 nautical miles away from their tagging location (Brickle *et al.* 2003), while parasites used as biological tags suggest that males are residents that do not move large distances (Brickle & MacKenzie 2007). This information supports the idea that *E*. *maclovinus* has a structured geographic distribution, however, this has not been supported by research based on molecular approaches. For instance, phylogeographic analysis using mtDNA, which only should reflect female dispersal, showed that *E*. *maclovinus* had weak genetic differentiation, shared haplotypes among locations, and recent population expansion that would have occurred as a result of Quaternary Glaciations (Ceballos *et al.* 2012). In addition, a subsequent study using microsatellites revealed a low but significant level of genetic differentiation between Pacific and Atlantic populations, with Atlantic populations showing a mixed membership from the two main genetic clusters. Some degree of genetic heterogeneity was suggested within the Pacific at a lower hierarchical level (Ceballos et al 2016). The microsatellites study also suggested that northern populations, at both Atlantic and Pacific Oceans, may harbor more genetic variability as revealed by the number of private alleles (Ceballos et al 2016). This suggests that Northern Patagonia fjords may harbor a cryptic population genetic structure at fine-scale, as a result of the geomorphological and heterogenous landscape, which was not observed due to lack of sampling locations from this region or low resolution of the molecular information used.

In this study, we investigate the population genomics of *E*. *maclovinus* from Northern Patagonia, using RADseq to assess both neutral and adaptive variation. We specifically aim to (i) evaluate the extent of neutral and adaptive genetic diversity, population differentiation between locations, and estimation of effective population size (ii) correlate putative adaptive loci to environmental variables, (iii) identify putative functions for candidate loci, and iv) discuss the implications of our results for conservation and management of the species.

**Materials and Methods**

*Sampling procedures*

We collected a total of 125 individuals using fishing nets (2"- 4”) or handlines from five sampling locations (25 each) between November 2018 and April 2019 (Figure 1). Sampling locations were Reloncaví Estuary (REL), Hornopirén (HOR), Manao (MAN), Chepu (CHE), Chaitén (CHA). REL, HOR and CHA correspond to continental locations, and MAN and CHE correspond to insular locations. 1For each specimen, we obtained a small piece of muscular tissue and stored it in 96% ethanol for further molecular procedures. Specimens used were collected in accordance with the national legislation of the country (Chile).

*RAD seq library and genotyping*

We obtained high-quality DNA using two steps. First, we used a traditional phenol-chloroform DNA extraction protocol to obtain a large amount of genomic DNA; second, we purified the DNA using DNeasy Blood & Tissue Kit (Qiagen®) following the manufacturer’s instructions but skipping the lysis step. We quantified the total double stranded DNA using a Qubit® 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and the dsDNA BR Assay Kit, following manufacturer’s instructions. Each DNA-sample was dried-down and then normalized with of sterile water to 20 ng/μl in a final volume of 10 μl (i.e., 200 ng per individual).

Three restriction site-associated DNA (RAD) libraries were prepared using the BestRAD protocol (Ali *et al.* 2016; Ackiss *et al.* 2020). Each DNA sample was digested using the SbfI-HF® restriction enzyme (New England Biolabs) and ligated with an eight bp unique barcode adaptor. Barcoded individual DNA samples were pooled into master libraries (i.e., 96 individuals each) and fragmented to ~300-500bp with 12-14 30s cycles in a Q500 sonicator (Qsonica, Newtown, CT). The fragmented DNA library was bound to Dynabeads™M-280 Streptavidin magnetic beads (Invitrogen). Subsequently, non-target fragments were removed by washing with a TLE buffer, and DNA was released from the beads using an incubation step. We conducted three purification steps for each library: i) non-ligated adaptors or small fragments of DNA and enzymes for each library were removed using AMPure XP beads (Beckman Coulter, Brea, CA); ii) all fragments of DNA that were not barcoded were removed using Dynabeads™M-280 Streptavidin magnetic beads leaving only the DNA fragments with the restriction site and barcode; iii) any residuals of small fragments of DNA and all enzymes and other impurities were also discarded using AMPure XP beads. NEBNext® Ultra™ DNA Library Prep Kit for Illumina® were used for ligation of master library barcodes, a 250-bp insert size-selection, and a 12-cycle PCR enrichment. We confirmed enrichment and size-selection by visualizing PCR products on a 2% agarose E-Gel (Invitrogen). A final AMPure XP purification clean-up step was conducted prior to quantifying the DNA with a Qubit® 2.0 Fluorometer. All prepared libraries for paired-end libraries were sent to Novogene (Sacramento, CA) for sequencing on the Illumina NovaseqS4 platform.

*SNP discovery*

To discover and genotype SNP from the raw RAD-seq data, we used a similar pipeline to the one suggested by Rochette and Catchen (2017) using the software STACKS v2.41 (Rochette *et al.* 2019). We selected a subset of 60 samples for testing and to select an optimal set of parameters as suggested by Paris et al. (2017). Raw sequences were demultiplexed by barcode using PROCESS\_RADTAGS program, where intact barcode and cut-site of SbfI restriction enzyme were checked, and whole reads with quality issues were discarded and trimmed to 140 bp (parameter flags: -e SbfI -c -q -r -t 140 --filter\_illumina --bestrad). Individuals that showed less than 950,000 reads retained were excluded from the downstream STACKS pipeline. We kept a total of 112 individuals. Reads were then assembled to build unique stacks to identify putative loci using a maximum likelihood framework (Hohenlohe *et al.* 2010) with USTACKS program (parameter flags: -m 3 -M 3 -d –disable\_gapped --model\_type bounded --bound\_high 0.25). A catalog of consensus loci was built with the CSTACKS program using five individuals from each sampling location that showed an amount of retained loci between average and median (parameter flags: -n 3 –disable\_gapped). Putative loci identified for all samples were matched against the catalog with the SSTACK program (parameter flag: --disable\_gapped). Data were transposed with the TSV2BAM program and built into a paired-end contig for calling variant sites and genotyping each individual with the GSTACKS program. Finally, we used the POPULATIONS program to call genotypes in one single VCF file for the posterior filtering process. POPULATIONS program was used without any filtering at this stage to have full control of the iterative filtering process described below.

*Bioinformatics and genotyping quality filters*

We used the iterative filtering process described by McKinney *et al*. (2020), which applies a soft and then a more stringent threshold to remove poor quality loci and samples. Filtering order was: i) minor allele frequencies, ii) genotype rate for loci, and iii) genotype rate for sample, after which we recalculated the proportion of missing data, and ran steps i), ii), and iii) with more stringent thresholds. For minor allele frequencies (MAF), we start by removing loci with MAF ≤ 0.05, and then MAF ≤ 0.1. For genotype rate for loci, the first threshold was set at 25% and the second at 90%. For the genotype rate for samples, the first threshold was set at 50% and the second at 85%. Loci in Hardy-Weinberg disequilibrium (p < 0.05) were removed if they deviated in three or more populations. For tags with more than one SNP, we kept the putative SNP with the highest *F*ST to reduce the influence from linked loci in the results. Additionally, paralog sequence variants were identified using HDPLOT (McKinney *et al.* 2017b) and then removed (parameter flags: H<0.6; |D|<5). Paralogs result from gene duplication events that have affected the evolution of the notothenioid genome (Chen *et al.* 2008, 2019). Paralogs are generally difficult to genotype reliably with RADseq data due to insufficient read-depth (McKinney *et al.* 2018). All variants that met with these criterion were retained for further analyses (Table 1).

*Data analyses: identifying neutral and adaptive loci*

We used two methods to detect loci under selection. The first method inferred outliers based on Principal Component Analysis (PCA), and was implemented using the PCADAPT package v4.3.3 (Privé *et al.* 2020). This method assumes that markers excessively related to population structure are candidates for local adaptation (Luu *et al.* 2017). The PCADAPT method uses the Mahalanobis distance (D) statistic, where a vector of the z-score is derived for regressing each SNP with K principal components (Luu et al. (2017). We applied Cattell’s rule to choose the K number of the principal components (Cattell 1966). The p-values were obtained from transforming Mahalanobis distance (D) based on the chi-square distribution (Cattell 1966). To avoid confounding effects of the population structure, we identified an optimal K-value testing from K=1 to K=10 and we checked the proportion of variance explained by each Principal Component using a scree plot using the *pcadapt* function and PCA (Figure S1). K=3 was retained and we calculated the False Discovery Rate of the p-values associated with Mahalanobis distances using the QVALUE package (Dabney *et al.* 2010). Finally, a list of putative adaptive loci was obtained under an expected False Discovery Rate of α = 0.1. The second method corresponds to FSTHET, which identifies candidate loci by calculating smoothed quantiles between loci with strong differentiation *F*ST relative to their expected heterozygosity (Flanagan & Jones 2017). This approach does not require any assumptions about the underlying population structure and is therefore more broadly applicable than other outlier detection methods. We calculated the empirical *F*ST based on Wright (1943) and expected heterozygosity. Loci were binned based on their expected heterozygosity values, sorted by *F*ST value, and quantiles were calculated. Loci that showed departures from a 95% confidence interval were considered under positive or balancing selection whether they surpassed superior or inferior confidence intervals, respectively (Figure S2). The VennDiagram package (Chen & Boutros 2011) was used to identify putative adaptive loci shared and unique between PCADAPT and FSTHET. Three data sub-sets were built after identifying putative loci under selection: i) neutral and ii) adaptive loci merged for both PCADAPT and FSTHET, and iii) adaptive loci shared between both PCADAPT and FSTHET, for further population genomic analyses.

*Summary statistics, population divergence, and effective population size*

The summary statistics of genetic diversity expected heterozygosity (HE) and observed heterozygosity (HO) for each sub-set were calculated by location and conducted using the HIERFSTAT v0.04-10 package (Goudet 2005). The number of polymorphic loci was obtained in NEESTIMATOR v2 (Do *et al.* 2014).

We estimated the individual ancestry coefficients based on sparse non-negative matrix factorization algorithms (sNMF) using the package ‘LEA’ in R (Frichot & François 2015). In this package we tested each data sub-set to reveal population genetic structure. We identified the best number of genetic clusters (K) based on cross-validation and on an information theoretic measure, the cross-entropy criterion (Alexander & Lange 2011; Frichot & François 2015). We iteratively tested from K=1 to K=10, with 10 replicates, and with 10,000 permutations per K using the function *obj.snmf* in LEA. We conducted the statistical procedure PCA to reduce the multivariate SNP multilocus data into two orthogonal axes using the ADEGENET v2.0 package (Jombart 2008; Jombart & Ahmed 2011). We used the PCA approach in all subsets obtained. We used the PCA to seek a summary of the genetic diversity among the sampled individuals ignoring the assumptions of the Hardy-Weinberg equilibrium and Linkage Disequilibrium which are often required in other individual-based models. Finally, we calculated pairwise *F*ST values for populations and performed significance tests for pairwise using 10,000 permutations in the STAMPP package (Pembleton *et al.* 2013). The *F*ST estimation was following the Wright (1949) method but corrected by the unequal population size as updated by Weir & Cockerham (1984) (see Pembleton *et al.* 2013). This analysis was completed for each sub-set. To understand the process that led to the population structure, we calculated directional migration rates among locations using the *divMigrate* function (Sundqvist *et al.* 2016) included in the diveRsity package (Keenan *et al.* 2013). Using the neutral data set we explored the migration rates using the D index, as statistic to calculate relative migration, 1000 bootstraps for statistical significance of directional migration, and a filter threshold between 0.5 to 0.85 to control the number of links.

The effective population size (Ne) of each location was estimated using the LD method (Waples 2006) updated for missing data and following Peel *et al*. (2013). The Ne was estimated only for a neutral data set. Values of Ne within corresponding 95% confidence intervals for each population were estimated using NEESTIMATOR v2 (Do *et al.* 2014) with the following parameters: a minimum allele frequency cutoff of 0.01 and a random mating model.

*Genotype-Environment Association*

Signatures indicative of local adaptation to environmental variables were investigated using Redundancy analysis (RDA). RDAis a multivariate ordination method that combines PCs from allele frequency and multivariate environmental distance matrices to produce canonical axes predicting relationships between environments and particular loci (Rellstab *et al.* 2015; Forester *et al.* 2018). We used the Bio-Oracle dataset (Tyberghein *et al.* 2012; Assis *et al.* 2018) to obtain environmental variables for salinity, temperature, pH, oxygen, silicate, current velocity, primary production, phosphate, phytoplankton, iron, nitrate, chlorophyll a, and calcite (https://www.bio-oracle.org/). For all these variables, we obtained values for the maximum, minimum, mean, and range of each variable when possible. RDA was performed using theVEGAN v. 2.3.4 R package (Oksanen *et al.* 2015). Variance inflation factor (*vif.cca* function of VEGAN) was used to ascertain lack of multi-collinearity among variables (Hair *et al.* 1995; Zuur *et al.* 2010) and excluded variables with a VIF ≥ 10 (Hair *et al.* 1995). Outliers were identified on each of the first three ordination axes as SNPs with a ‘locus score’ that was ±3 SD from the mean score for that axis RDA, as suggested by Forester et al. (2018) to minimize false-positive and false-negative results. We then determined the correlation between each candidate SNP and one or more environmental variables.

*Putative function from Blast2Go*

We conducted loci annotation to identify a putative function for candidate variants underlying positive selection obtained from PCADAPT, FSTHET and RDA. We used the software Blast2Go included in OmicsBox following the annotation pipeline described by Götz et al. (2008). Briefly, we compared our candidate loci against the NCBI genomic database translating the sequences from nucleotide to protein using BLASTX. Then we mapped homologous sequences to Gene Ontology (GO) terms. Finally, sequences were annotated applying the Blast2GO annotation rule (see Götz *et al.* 2008). We tabulated the Tag\_SNP from our read; method which identified the outlier, environmental variable associated, Gene name, Gene Ontology ID as well as their respective GO names.

**Results**

*Sequencing, genotyping quality filters and datasets*

After excluding individuals with low number of reads retained, we obtained RAD data from 112 individuals that ranged from 971,305 to 14,872,661 reads with a median of 5,821,883 reads. The STACKS pipeline without filter revealed a total of 1,334,812 (242,278 RADtag) putative SNPs. 98.41 % (1,313,546) of SNPs were removed by the iterative quality filters (MAF and missing data), increasing to 99.06% removed by the read depth filter (8); keeping only one SNP per tag (8,712), and finally the percentage increased to 99.07% after removing Hardy-Weinberg disequilibrium departure loci (41) and paralogs (123). The whole filtering process resulted in a final dataset of 101 individuals and 12,382 high quality SNPs. Outlier tests using PCADAPT identified 26 putative adaptive (0.2%) and 12,356 neutral loci (99.8%); while FSTHET (Figure S2) revealed 332 loci for positive selection (3%), 338 for balancing selection (3%), and 11,712 for neutral (94%) (Figure S2). A total of 20 adaptive loci were shared between both PCADAPT and FSTHET, while six loci were unique for PCADAPT and 312 for FSTHET (Figure S3).

*Summary statistics, population divergence, and effective population size*

Summary statistics of genetic diversity revealed similar values for each location within each dataset (Table 2). For the neutral dataset, HO ranged from 0.301 to 0.313, and HE from 0.317 to 0.324. For the adaptive loci merged dataset, HO ranged from 0.252 to 0.295 and HE from 0.284 to 0.321. For the adaptive loci shared dataset, HO ranged from 0.199 to 0.384 and HE from 0.204 to 0.442. The number of polymorphic loci was 11,680 for REL, 11,703 for MAN, 11,709 for HOR, 11,529 for CHE, and 11,632 for CHA (Table 2).

Based on PCA, we observed slight differences of structuration patterns between the three data-subset. For instance, the neutral loci dataset was only able to show clear differences between two groups while the adaptive loci dataset (338 SNPs) identified three distinct groups, i) REL (Reloncaví estuary population), ii) MAN, HOR, CHA (Inner Sea of Chiloé populations), and iii) CHE (oceanic population). The use of adaptive loci increased the variance explained by PCs (4.7% and 4.5% for PC1 and PC2, respectively) when compared to the neutral dataset (1.4% and 1.3% for PC1 and PC2, respectively) as well as decreased dispersion of individuals within groups (Figure 2B, Figure S4). Contrarily, the shared adaptive loci dataset (20 SNPs) did not show the same pattern for CHE when compared to the merged adaptive loci dataset. In this PCA, which showed the highest amount of variation explained by PCs (24.1 and10.8% for PC1 and PC2, respectively), CHE seems to be closer to CHA and HOR, but slightly differentiated from MAN (Figure 2C, Figure S4). The REL population was identified as a different genetic group than other populations in all analyses across the three datasets.

Pairwise *F*ST values, using Wright (1949) method, for neutral dataset ranged from 0.004 for the MAN-HOR, MAN-CHA, HOR-CHA comparisons to 0.007 for the CHE-REL comparison (Table 3). For the adaptive loci merged dataset, pairwise *F*ST values ranged from 0.043 for MAN-HOR to 0.118 for the CHE-CHA comparison; while the adaptive loci shared dataset values ranged from 0.005 for HOR-CHA to 0.191 for REL-MAN (Table 3). Overall, relative migration rates using neutral data showed high gene flow between locations (Figure 5A), with rates ranging from 0.56 to 1. Increasing the threshold, we notice that CHE has less connectivity (threshold > 0.56) followed by REL (threshold > 0.83). Putative adaptive loci show relative migration rates ranging from 0.22 to 1 (Figure 5B). There is a clearer low relative migration rate pattern in CHE. Varying the threshold, we notice that CHE is separated from others at > 0.38, REL and CHA at >0.81.

The lowest value of the cross-entropy suggested K=1 for neutral, K=4 for adaptive merged, and K=3 for adaptive shared dataset (Figure S6-S8). We surveyed the admixture result from K=2 to K=5 and decided to keep K=2 for neutral, K=3 for adaptive merged, and K=2 for the adaptive shared dataset (Figure 3) because of the shared similarity with PCA (Figure 2). For K= 2 in the neutral dataset the first group included REL with an admixture proportion of ~0.76 for Cluster 1, and the second group included MAN, CHE, HOR and CHA, with proportions ranging between ~0.65 and ~0.90 (Figure 4A). For K= 3 in the adaptive merged dataset, the first group included REL, the second included MAN, HOR and CHA, and the third group included CHE (Figure 4B). For K= 2 in the adaptive shared dataset, the first group included REL, the second included MAN, CHE, HOR and CHA (Figure 4C). Reloncaví Estuary location was clearly different in admixture analyses for all datasets (Figure 3, 4). Mean admixture proportions by location in putative neutral, adaptive merged, and adaptive shared loci are represented in Figures S9-S11.

Most estimates of effective population size for the neutral dataset were finite and varied by two or three orders of magnitude across locations; the Ne estimate ranged from 572.2 in Chepu to 14,454.8 for Chaitén (Table 1). Only the Chaitén confidence interval included an infinite value (Table 2).

*Genotype-Environment Association*

The global model of the multilocus genotype–environment RDA conducted using all loci was significant (ANOVA F= 1.21, p = 0.001) with the first three components explained 26.87%, 24.76% and 24.60 % of the variation for RDA1 to RDA3, respectively (Figure 6A, Figure S12 A-B). A total of 78 putative adaptive loci were found correlated to maximum dissolved molecular oxygen (25), mean of primary productivity (14), mean salinity (30), and minimum current velocity (9) (Figure 6B, Figure S12 C-D). A total of 75 of 78 outliers identified here were also identified in FSTHET and PCADAPT.

*Putative function from Blast2GO*

From 341 unique putative candidate loci (FSTHET, PCADAPT, and RDA), 124 loci were blasted in BlastX, 111 loci mapped were of homologue sequences to GO terms, and finally 98 loci were annotated to GO terms. Based on these results, we found a variety of candidate genes whose functions involved biological processes like signal transduction, metabolism, signaling pathways and gene expression among others (Table S1).

**Discussion**

Under the assumption that heterogeneous landscape, low vagility and biological conditions can result in population divergence in *Eleginops* *maclovinus*, we aimed to (i) disentangle the differences in neutral and adaptive genetic variation, (ii) correlate putative adaptive loci to environmental variables, and (iii) identify putative functions for candidate loci. PCA and membership analyses revealed two (neutral loci) and three (adaptive loci) clusters, none of them previously described for this species. Neutral loci suggest a spatial pattern of structuration with gene flow, while adaptive suggest spatial selection by environmental association. We identify candidate loci for divergent selection mainly associated with biological processes and metabolism. Contrasting Ne estimations among populations were found. Chepu, the population with the lowest Ne estimated, may be vulnerable to loss of genetic variation by genetic drift. Overall, our results uncover a hidden fine-scale population structure in *E*. *maclovinus* along with its North Patagonian distribution (i.e., Reloncaví Estuary). Also, a spatial selection by environmental variables arises in CHE, a location that is mainly influenced by oceanic conditions. Identification of these groups will facilitate the development of conservation and management measures for this species.

*Neutral genetic variation*

We found fine-scale genetic spatial pattern of structuration with gene flow based on neutral genomic data. We identified two genetic groups along the main PCA axis; the strongest genetic differentiation occurred between Reloncaví Estuary and all other populations. Studies in *E*. *maclovinus* using mtDNA and microsatellites showed contrasting results. For example, based on the Cyt-b fragment, Ceballos *et al*. (2012) showed low genetic differentiation between five populations of *E*. *maclovinus* located between Pacific and Atlantic distribution. They suggested current and historical connectivity between populations from their expansion from the middle Pleistocene. Later, using microsatellites, low but significant regional differentiation between the Pacific Ocean and Atlantic Ocean locations was found (Ceballos *et al.* 2016). Similar studies using microsatellites did not find population structure in the Patagonian area from the Pacific Ocean (Canales-Aguirre *et al.* 2010, 2018). Our findings contrast with previous results and demonstrate that genotyping tools such as RADseq increase the power to resolve shallow population structure in fish with geneflow (Luikart *et al.* 2003; Larson *et al.* 2014a; Hollenbeck *et al.* 2019) when other lower resolution genomic approaches cannot. Our findings can be explained through a combination of biological species’ attributes and geomorphological configurations.

Dispersal and reproductive attributes could contribute to the observed population pattern. Theoretically, hermaphroditic species tend to have more structured populations than gonochoristic species (Chopelet *et al.* 2009; Coscia *et al.* 2016), but information supporting this hypothesis is scarce. For example, Chopelet et al. (2009) conducted a metanalysis testing this hypothesis and found no supporting evidence. They suggested that dispersal capacities and environmental barriers can play an underlying role in the variance of the genetic structuring of marine fish populations. Our data may partially support this hypothesis, but only from the Reloncavi Estuary. Geographic distance between locations cannot be large enough to avoid connectivity and reinforce this theoretical issue. We suggest that this life history trait is not enough to promote differences. Passive dispersal by early stages (egg and larvae) is uncertain because there is no information whether *E*. *maclovinus* have pelagic or benthic eggs or whether spawning occurs in open sea as well as in estuaries. Hence, it is hard to suggest how drifting of early stages can connect populations through gene flow like has been suggested for several marine species (Benestan *et al.* 2021). Dispersal behavior in adults has been recorded using biological (parasites) and mechanical tags (Brickle *et al.* 2003; Brickle & MacKenzie 2007). Both studies indicate that juveniles (mainly males) tend to be residents and larger fish (mainly female) can migrate, comparatively, large distances (up to 60 nautical miles) (Brickle *et al.* 2003; Brickle & MacKenzie 2007). It is plausible to presume that *E*. *maclovinus* can have a sex-biased migration, however this has not been noticed using mitochondrial DNA (Ceballos *et al.* 2012) and also, there are hermaphroditic therefore sex is not a factor.

The geomorphology of the Northern Patagonia and the pattern of sea currents provide further evidence for the differences observed. The presence of channels, estuaries, close sounds and fjords can be an efficient barrier for dispersal at different life-history stages. For example, it has been suggested that genetic isolation is related to the shallow sill depth and life history behavior in *Benthosema glaciale* (Kristoffersen & Salvanes 2009). The Reloncaví Estuary is 55 km long, with geomorphological and environmental characteristics that make this a unique area (Castillo *et al.* 2016, 2017). For instance, superficial currents resulting from wind, flows-down to the Reloncaví Sound in winter and flows-up in spring and summer (Castillo *et al.* 2016, 2017). The latter could promote a safe environment for early stages of marine organisms. Our samples were collected in the head of the estuary, a sheltered area, less influenced by sea tidal change and more by freshwater runoff. The sea-current in Reloncaví Estuary mainly goes from the head to the mouth of the estuary (Castillo *et al.* 2016, 2017), however it seems that some individuals are more similar to Hornopirén and Chaitén suggesting dispersal based on PCA. This is supported by the admixture of some individuals in the membership analyses and the cluster distribution by location. This singular geomorphology could promote a similar pattern of genetic diversity for other species. Unfortunately, to date, there are no genetic studies in fish for this area specifically.

*Adaptive variation*

We identified three well-defined groups of samples in the adaptive merged dataset that include Reloncaví Estuary, Chepu, and all other populations. The adaptive shared dataset showed two groups of samples, however for further discussion we will focus on the adaptive merged dataset.

There are qualitative differences in the environment where these three genetic groups belong. The Reloncaví Estuary, as described above, shows idiosyncratic environmental characteristics that can result from being in an area sheltered from oceanic conditions (Castillo *et al.* 2017). Additionally, Reloncaví Estuary is influenced by runoff of freshwater from two main rivers, the Petrohué and Puelo. All these characteristics can act as a barrier for early stages and result in local adaptations. Chepu is mainly influenced by oceanic conditions, making it more exposed to the marine environment. In the exposed Pacific Ocean part of Chiloe Island, there are less estuary habitats where juveniles of *E*. *maclovinus* can use as a spawn and nursery area. The Manao, Hornopirén and Chaitén populations are distributed in inland waters of Chiloé. This area acts as a large estuary as it is described as a mix of freshwater runoff and marine (Pantoja *et al.* 2011).

All three genetic groups identified in the adaptive merged dataset are distributed in clinal marine environments and are more influenced by marine conditions than others. Divergence in adaptive loci reinforces the idea that such markers may experience selection by environmental pressures (Hollenbeck *et al.* 2019). For instance, in the estuarine fish *Sciaenops ocellatus* from the Atlantic Ocean in the United States and Mexico, neutral and adaptive loci show a similar pattern for population structure, but discordance in the *F*ST magnitude for outlier loci was greater than in neutral loci (Hollenbeck *et al.* 2019). Differences in habitat from dissolved inorganic phosphates, average wind speed and minimum ocean salinity can play a role in adaptive divergence. Although we did not test any specific environmental variables to explain the adaptive divergence in *E*. *maclovinus*, we could speculate that those differences can be associated with the different environments where the samples were collected. Environmental patterns in the Reloncaví Estuary indicate a higher influx of freshwater than Chepu, hence salinity may be a driver for this divergence. Additionally, winds in the Reloncaví Estuary result in a more protected area compared with the Chepu. Araneda et al. (2016) identified divergent adaptive loci for *Mytilus chilensis,* a native mollusk species*,* in the same area where samples were collected for this study. They found that factors such as salinity, water discharge by rivers, glacial melt and precipitation cause differences in the habitat between the Reloncaví Estuary and the exposed area in Chiloe (Araneda *et al.* 2016).

All putative candidate loci annotated as GO terms in Blast2GO revealed genes mainly associated with different biological processes, metabolic, and signaling pathways. Relevant examples include: sorting nexin-19 isoform X3 (SNX19), a disintegrin and metalloproteinase with thrombospondin motifs 1 (ADAMTS1), palladin isoform X2 (PALLD) Glutaminase kidney isoform (GLS), Rock1 kinase (ROCK1), Elongation factor-like GTPase 1 (EFL1), collagen alpha-1(IV) chain (COL4A1), and sorting nexin-33 (SNX33).

The Genotype-Environment Association analyses provide evidence for local adaptation for current velocity, primary productivity, oxygen, and salinity. These variables can have a major effect in early development stages, such as has been suggested for *Sardina pilchardus* (Antoniou *et al.* 2021). Sea current velocity provides a stable environment and the indirect proxy for food availability can be related to the old larval fish ecology hypotheses, which explains recruitment variability (Houde 2008). Oxygen can play a role in energetic metabolisms and neuromuscular processes, which appear during embryonic development and environmental stress (Epelboin *et al.* 2016; Moreira *et al.* 2018). In addition to the modification of egg and larvae distribution (Petereit *et al.* 2009), salinity can provide a challenging environment for adaptation in fish. Genes identified that are candidates for salinity adaptation are SNX19, which participates in regulating vesicle trafficking, which can confer salt tolerance (Deane-Coe *et al.* 2018), and DYNC2H1, which has been associated with “renal water homeostasis,” “vasopressin-regulated water reabsorption,” and “urea transport” (Zhou *et al.* 2018). Our findings reinforce the idea that salinity levels can act as a selective force in *E*. *maclovinus* populations.

*Management implications*

Our results provide the first report of fine-scale spatial pattern of structuration with gene flow and spatial selection by environmental association in *E*. *maclovinus.* This information can be used to improve the currently weak management measures that only cover regulations for the type of fishing gear used. The economic importance and conservation status of *E*. *maclovinus* populations makes the lack of regulatory measures for fish management an important issue. From an economic perspective, *E*. *maclovinus* is important for the activity of local artisanal fisherman and recreational anglers (Sernapesca 2019).

Our findings support the presence different conservation/management units at a fine-scale. Adaptive variation showed three genetic groups associated with different environments, represented by different locations: i) populations from the Reloncaví Estuary, which is sheltered from oceanic conditions influenced by freshwater from rivers such as the Petrohúe and Puelo, and less influenced by tides; ii) populations from Manao-Hornopirén-Chaitén located in the Chiloé inner sea, less protected and more influenced by tides regimens, and iii) Chepu, which is an exposed area influenced by marine seawater. Each genetic group identified can harbor variants that permit those populations to adapt to these environments. Additionally, the mixture observed for some individuals and the low *F*ST of neutral loci could suggest migration among locations and interdependence among populations. Also, it is unknown how these populations are connected with locations outside of the Patagonian area, i.e., northward Chile, where populations from Chepu may be more connected.

Stock assessment for hermaphroditic marine species is a challenge for fishery managers due to their sex ratio bias. In protandrous species, the sex ratio is skewed to males (Allsop & West 2004) and every year fisheries often remove larger/older individuals (females) from populations which can result in evolutionary changes in exploited populations. First, a decrease in population fitness or change of size for changing sex. Fecundity as a fitness trait in females increases with age and size; if older females are removed from the population the fitness of the population will decrease. This increases risk for populations that show small effective population size, such as Chepu (572) or the Reloncavi Estuary (1309). Second, Allsop and West (2003) found that hermaphrodite fish change their sex when they reach 80% of their maximum size and are 2.5 times their age of maturity. Nonetheless, the length for sex change can be strongly modified by fisheries. For instance, in *Semicossyphus pulcher*, found that males and females mature early in locations with intensified recreational or commercial fisheries (Hamilton et al. 2007). For *E*. *maclovinus* information about maturity is scarce. Brickle et al. (2005) indicate that males mature around 30.73 cm and females between 67–78 cm LT (maturation stage III). Currently, there is no information about maturity in both males or females in the areas sampled only for small-scale commercial fisheries. Therefore, the effect of removing larger individuals could bring evolutionary changes in a population of *E*. *maclovinus*, impacting their conservation and management.

*Future directions*

We showed that by using relatively dense genomic information it is possible to resolve population structure in *E*. *maclovinus*, whereas previous studies indicated weak genetic differentiation in a large geographical area (Ceballos *et al.* 2012, 2016). In this study we covered a segment of *E. maclovinus* distributional range. Future research should include the whole distribution of the species; making it possible to assess more populations along the heterogeneous landscape in Patagonia and identify whether environmental variables are associated with the current diversity and population structure.

Given the adaptive variation found, further studies should focus on identifying environmental selective pressure associated with phenotypical and ecological traits. Finally, this study opens a new path for research in *E*. *maclovinus*, that can be started in the near future, using long-term monitoring of genetic diversity.

**Literature cited**

Ackiss AS, Larson WA, Stott W (2020) Genotyping-by-sequencing illuminates high levels of divergence among sympatric forms of coregonines in the Laurentian Great Lakes. *Evolutionary Applications*, **13**, 1037–1054.

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, **12**, 246.

Ali OA, O’Rourke SM, Amish SJ *et al.* (2016) RAD Capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, **202**, 389–400.

Allsop DJ, West SA (2003) Constant relative age and size at sex change for sequentially hermaphroditic fish. *Journal of Evolutionary Biology*, **16**, 921–929.

Allsop DJ, West SA (2004) Sex-ratio evolution in sex changing animals. *Evolution*, **58**, 1019–1027.

Antoniou A, Manousaki T, Ramírez F *et al.* (2021) Sardines at a junction: seascape genomics reveals ecological and oceanographic drivers of variation in the NW Mediterranean Sea. *Authorea Preprints*.

Araneda C, Larraín MA, Hecht B, Narum S (2016) Adaptive genetic variation distinguishes Chilean blue mussels (Mytilus chilensis) from different marine environments. *Ecology and Evolution*, **6**, 3632–3644.

Assis J, Tyberghein L, Bosch S *et al.* (2018) Bio‐ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global ecology and biogeography: a journal of macroecology*, **27**, 277–284.

Bargelloni L, Marcato S, Zane L, Patarnello T (2000) Mitochondrial phylogeny of notothenioids: a molecular approach to Antarctic fish evolution and biogeography. *Systematic biology*, **49**, 114–129.

Benestan L, Fietz K, Loiseau N *et al.* (2021) Restricted dispersal in a sea of gene flow. *Proceedings Royal Society B*, **288**, 20210458.

Berg PR, Jentoft S, Star B *et al.* (2015) Adaptation to low salinity promotes genomic divergence in Atlantic Cod (Gadus morhua L.). *Genome biology and evolution*, **7**, 1644–1663.

Bernatchez L (2016) On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of fish biology*, **89**, 2519–2556.

Bernatchez L, Wellenreuther M, Araneda C *et al.* (2017) Harnessing the Power of Genomics to Secure the Future of Seafood. *Trends in ecology & evolution*, **32**, 665–680.

Beuchel F, Gulliksen B, Carroll ML (2006) Long-term patterns of rocky bottom macrobenthic community structure in an Arctic fjord (Kongsfjorden, Svalbard) in relation to climate variability (1980–2003). *Journal of Marine Systems*, **63**, 35–48.

Brickle P, Laptikhovsky V, Arkhipkin A (2005) Reproductive strategy of a primitive temperate notothenioid Eleginops maclovinus. *Journal of fish biology*, **66**, 1044–1059.

Brickle P, Laptikhovsky V, MacKenzie K, Arkhipkin A (2003) *The Falkland Islands mullet: biology and fisheries in Falkland Islands’ waters; Scientific Report*. Falkland Islands Government Fisheries Department.

Brickle P, MacKenzie K (2007) Parasites as biological tags for Eleginops maclovinus (Teleostei: Eleginopidae) around the Falkland Islands. *Journal of Helminthology*, **81**, 147.

Canales-Aguirre CB, Ferrada S, Hernandez CE, Galleguillos R (2010) Population structure and demographic history of Genypterus blacodes using microsatellite loci. *Fisheries research*, **106**, 102–106.

Canales-Aguirre CB, Ferrada-Fuentes S, Galleguillos R *et al.* (2018) High genetic diversity and low population differentiation in the Patagonian sprat (Sprattus fuegensis) based on mitochondrial DNA. *Mitochondrial DNA Part A*, **29**, 1148–1155.

Canales-Aguirre CB, Ferrada-Fuentes S, Galleguillos R, Hernández CE (2016) Genetic Structure in a Small Pelagic Fish Coincides with a Marine Protected Area: Seascape Genetics in Patagonian Fjords. *PloS one*, **11**, e0160670.

Castillo MI, Cifuentes U, Pizarro O, Djurfeldt L, Caceres M (2016) Seasonal hydrography and surface outflow in a fjord with a deep sill: The Reloncav?? fjord, Chile. *Ocean Science*, **12**, 533–534.

Castillo MI, Pizarro O, Ramírez N, Cáceres M (2017) Seiche excitation in a highly stratified fjord of southern Chile: the Reloncaví fjord. *Ocean Science*, **13**, 145–160.

Cattell RB (1966) The Scree Test For The Number Of Factors. *Multivariate behavioral research*, **1**, 245–276.

Ceballos SG, Lessa EP, Licandeo R, Fernández DA (2016) Genetic relationships between Atlantic and Pacific populations of the notothenioid fish Eleginops maclovinus: the footprints of Quaternary glaciations in Patagonia. *Heredity*, **116**, 372–377.

Ceballos SG, Lessa EP, Victorio MF, Fernández DA (2012) Phylogeography of the sub-Antarctic notothenioid fish Eleginops maclovinus: evidence of population expansion. *Marine biology*, **159**, 499–505.

Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, **12**, 35.

Chen Z, Cheng C-HC, Zhang J *et al.* (2008) Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 12944–12949.

Chen L, Lu Y, Li W *et al.* (2019) The genomic basis for colonizing the freezing Southern Ocean revealed by Antarctic toothfish and Patagonian robalo genomes. *GigaScience*, **8**.

Chopelet J, Waples RS, Mariani S (2009) Sex change and the genetic structure of marine fish populations. *Fish and fisheries* , **10**, 329–343.

Coscia I, Chopelet J, Waples RS, Mann BQ, Mariani S (2016) Sex change and effective population size: implications for population genetic studies in marine fish. *Heredity*, **117**, 251–258.

Dabney A, Storey JD, Warnes GR (2010) qvalue: Q-value estimation for false discovery rate control. *R package version*, **1**.

Deane-Coe P, Butcher BG, Greenberg R, Lovette IJ (2018) Whole genome scan reveals the multigenic basis of recent tidal marsh adaptation in a sparrow. *bioRxiv*, 360008.

Do C, Waples RS, Peel D *et al.* (2014) NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, **14**, 209–214.

Dong X, Wang J, Ji P *et al.* (2020) Seawater Culture Increases Omega-3 Long-Chain Polyunsaturated Fatty Acids (N-3 LC-PUFA) Levels in Japanese Sea Bass (Lateolabrax japonicus), Probably by Upregulating Elovl5. *Animals*, **10**.

Epelboin Y, Quintric L, Guévélou E *et al.* (2016) The Kinome of Pacific Oyster Crassostrea gigas, Its Expression during Development and in Response to Environmental Factors. *PloS one*, **11**, e0155435.

Euclide PT, MacDougall T, Robinson JM *et al.* (2021) Mixed-stock analysis using Rapture genotyping to evaluate stock-specific exploitation of a walleye population despite weak genetic structure. *Evolutionary Applications*, **14**, 1403–1420.

Flanagan SP, Jones AG (2017) Constraints on the FST-Heterozygosity Outlier Approach. *Journal of Heredity*, **108**, 561–573.

Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular ecology*, **27**, 2215–2233.

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

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

Garreaud R, Lopez P, Minvielle M, Rojas M (2013) Large-Scale Control on the Patagonian Climate. *Journal of climate*, **26**, 215–230.

Gastaldi M, Maggioni M, Reinaldo MO, González RAC (2009) Caracterización biológica y poblacional del róbalo Eleginops maclovinus (Pisces, Eleginopsidae) en la Bahía de San Antonio Oeste y zona de influencia. *Instituto de Biología Marina y Pesquera Serie Publicaciones*, **8**, 1–18.

Gosztonyi AE (1974) Edad y crecimiento del ‘róbalo’, Eleginops maclovinus (Osteichthyes, Nototheniidae) en aguas de la ría Deseado y sus adyacencias. *Physis*, **33**, 1–8.

Gosztonyi AE (1979) Biología del “Robalo” (Eleginops maclovinus (CUV. & VAL.,1830). Universidad Nacional de Buenos Aires.

Götz S, García-Gómez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, **36**, 3420–3435.

Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Hair JF, Anderson RE, Tatham RL, Black WC (1995) *Multivariate Data Analysis*. Macmillan Publishing Company, New York.

Hamilton SL, Caselle JE, Standish JD *et al.* (2007) Size-selective harvesting alters life histories of a temperate sex-changing fish. *Ecological Applications*, **17**, 2268–2280.

von der Heyden S (2017) Making evolutionary history count: biodiversity planning for coral reef fishes and the conservation of evolutionary processes. *Coral reefs* , **36**, 183–194.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**.

Hollenbeck CM, Portnoy DS, Gold JR (2019) Evolution of population structure in an estuarine-dependent marine fish. *Ecology and Evolution*, **9**, 3141–3152.

Houde ED (2008) Emerging from hjort’s shadow. *Journal of Northwest Atlantic fishery science*, **41**, 53–70.

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

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , **27**, 3070–3071.

Jørgensen HBH, Hansen MM, Bekkevold D, Ruzzante DE, Loeschcke V (2005) Marine landscapes and population genetic structure of herring (Clupea harengus L.) in the Baltic Sea. *Molecular ecology*, **14**, 3219–3234.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: AnRpackage for the estimation and exploration of population genetics parameters and their associated errors. *Methods in ecology and evolution / British Ecological Society*, **4**, 782–788.

Kristoffersen JB, Salvanes AGV (2009) Distribution, growth, and population genetics of the glacier lanternfish (Benthosema glaciale) in Norwegian waters: Contrasting patterns in fjords and the ocean. *Marine Biology Research* , **5**, 596–604.

Larson WA, Seeb LW, Everett MV *et al.* (2014a) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (Oncorhynchus tshawytscha). *Evolutionary Applications*, **7**, 355–369.

Larson WA, Seeb JE, Pascal CE, Templin WD, Seeb LW (2014b) Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (Oncorhynchus tshawytscha) from western Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 698–708.

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

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Luu K, Bazin E, Blum MGB (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular ecology resources*, **17**, 67–77.

Matschiner M, Colombo M, Damerau M *et al.* (2015) The Adaptive Radiation of Notothenioid Fishes in the Waters of Antarctica. In: *Extremophile Fishes: Ecology, Evolution, and Physiology of Teleosts in Extreme Environments* (eds Riesch R, Tobler M, Plath M), pp. 35–57. Springer International Publishing, Cham.

McCairns RJS, Bernatchez L (2008) Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large-scale estuarine environment. *Molecular ecology*, **17**, 3901–3916.

McKinney G, McPhee MV, Pascal C, Seeb JE, Seeb LW (2020) Network analysis of linkage disequilibrium reveals genome architecture in Chum salmon. *G3: Genes, Genomes, Genetics*, **10**, 1553–1561.

McKinney GJ, Seeb JE, Seeb LW (2017a) Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*.

McKinney GJ, Waples RK, Pascal CE, Seeb LW, Seeb JE (2018) Resolving allele dosage in duplicated loci using genotyping-by-sequencing data: A path forward for population genetic analysis. *Molecular Ecology Resources*, **18**, 570–579.

McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017b) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping‐by‐sequencing data from natural populations. *Molecular Ecology*, **17**, 656–669.

Moreira A, Figueira E, Mestre NC *et al.* (2018) Impacts of the combined exposure to seawater acidification and arsenic on the proteome of Crassostrea angulata and Crassostrea gigas. *Aquatic toxicology* , **203**, 117–129.

Near TJ (2004) Estimating divergence times of notothenioid fishes using a fossil-calibrated molecular clock. *Antarctic Science*, **16**, 37–44.

Ojaveer E, Kalejs M (2010) Ecology and long-term forecasting of sprat (Sprattus sprattus balticus) stock in the Baltic Sea: a review. *Reviews in fish biology and fisheries*, **20**, 203–217.

Oksanen J, Blanchet FG, Kindt R *et al.* (2015) vegan: Community Ecology Package. R package version 2.2-1.

Olsen RB, Richardson K, Simonsen V (2002) Population differentiation of eelpout Zoarces viviparus in a Danish fjord. *Marine Ecology Progress Series*, **227**, 97–107.

Pantoja S, Luis Iriarte J, Daneri G (2011) Oceanography of the Chilean Patagonia. *Continental Shelf Research*, **31**, 149–153.

Paris JR, Stevens JR, Catchen JM (2017) Lost in parameter space: a road map for stacks (S Johnston, Ed,). *Methods in Ecology and Evolution*, **8**, 1360–1373.

Peel D, Waples RS, Macbeth GM, Do C, Ovenden JR (2013) Accounting for missing data in the estimation of contemporary genetic effective population size (N(e) ). *Molecular Ecology Resources*, **13**, 243–253.

Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, **13**, 946–952.

Pequeño G (1989) Peces de Chile. Lista sistemática revisada y comentada. *Revista de Biología Marina, Valparaíso*, **24**, 1–132.

Pérez-Santos I, Garcés-Vargas J, Schneider W *et al.* (2014) Double-diffusive layering and mixing in Patagonian fjords. *Progress in Oceanography*, **129**, 35–49.

Petereit C, Hinrichsen H-H, Voss R *et al.* (2009) The influence of different salinity conditions on egg buoyancy and development and yolk sac larval survival and morphometric traits of Baltic Sea sprat (Sprattus sprattus balticus Schneider). *Scientia marina*, **73**, 59–72.

Privé F, Luu K, Vilhjálmsson BJ, Blum MGB (2020) Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. *Molecular Biology and Evolution*, **37**, 2153–2154.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular ecology*, **24**, 4348–4370.

Ríos F, Kilian R, Mutschke E (2016) Chlorophyll-a thin layers in the Magellan fjord system: The role of the water column stratification. *Continental Shelf Research*, **124**, 1–12.

Rochette NC, Catchen JM (2017) Deriving genotypes from RAD-seq short-read data using Stacks. *Nature protocols*, **12**, 2640–2659.

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, **28**, 4737–4754.

Rodrigo C (2008) Submarine topography in the Chilean North Patagonian channels. In: *Progress in the oceanographic knowledge of Chilean inner waters, from Puerto Montt to Cape Horn.* (eds Silva N, Palma S), pp. 19–23. Comité Oceanográfico Nacional-Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile.

Ruiz VH (1993) Ictiofauna del Rio Andalien (Concepción, Chile). *Gayana Zoologia*, **57**, 109–278.

Sernapesca (2019) *Anuario Estadístico de Pesca*. Ministerio de Economía Fomento y Reconstrucción, Santiago.

Sievers HA (2006) Temperature and salinity in the austral Chilean channels and fjords. In: *Progress in the oceanographic knowledge of Chilean interior waters, from Puerto Montt to Cape Horn.* (eds Silva N, Palma S), pp. 31–36. Comité Oceanográfico Nacional, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile.

Silva N, Calvete C, Sievers HA (1997) Características oceanográficas físicas y químicas de canales australes chilenos entre Puerto Montt y Laguna San Rafael (Crucero Cimar-Fiordo 1). *Ciencia y Tecnología del Mar*, **20**, 23–106.

Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D (2016) Directional genetic differentiation and relative migration. *Ecology and evolution*, **6**, 3461–3475.

Tyberghein L, Verbruggen H, Pauly K *et al.* (2012) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global ecology and biogeography: a journal of macroecology*, **21**, 272–281.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics* , **7**, 167–184.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications*, **3**, 244–262.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358–1370.

Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–138.

Wright S (1949) The genetical structure of populations. Annals of Human Genetics, 15, 323–354.

Xie D, Wang S, You C *et al.* (2015) Characteristics of LC-PUFA biosynthesis in marine herbivorous teleostSiganus canaliculatusunder different ambient salinities. *Aquaculture Nutrition*, **21**, 541–551.

Xuereb A, D’Aloia CC, Andrello M, Bernatchez L, Fortin M-J (2021) Incorporating putatively neutral and adaptive genomic data into marine conservation planning. *Conservation biology: the journal of the Society for Conservation Biology*, **35**, 909–920.

Yevenes MA, Bello E, Sanhueza-Guevara S, Farías L (2017) Spatial Distribution of Nitrous Oxide (N2O) in the Reloncaví Estuary–Sound and Adjacent Sea (41°–43° S), Chilean Patagonia. *Estuaries and Coasts*, **40**, 807–821.

Zhou X, Guang X, Sun D *et al.* (2018) Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nature communications*, **9**, 1276.

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution / British Ecological Society*, **1**, 3–14.

**Tables**

**Table 1.** Number of putative loci retained following each filtering step.

|  |  |  |
| --- | --- | --- |
| **Filtering steps** | **Sample size** | **Number of loci** |
| SNP after STACKS | 112 | 1334812 |
| 1MAF >= 0.05 | 112 | 829446 |
| 1Genotyped by locus (50%) | 112 | 168256 |
| 1Genotyped by sample (25%) | 109 | 168256 |
| 1Genotyped by locus (90%) | 109 | 38757 |
| 1Genotyped by sample (85%) | 101 | 38757 |
| 1MAF >= 0.1 | 101 | 21266 |
| min-meanDP 10 | 101 | 21264 |
| max-meanDP 100 | 101 | 21258 |
| one SNP by tag | 101 | 12546 |
| Hardy-Weinberg | 101 | 12505 |
| Singletons (HdPlot H<0.6; |D|<5) | 101 | 12382 |
| Putative Neutral PCADAPT | 101 | 12356 |
| Putative Adaptive PCADAPT | 101 | 26 |
| Putative Neutral FSTHET | 101 | 11712 |
| Putative Loci under Balancing Selection FSTHET | 101 | 338 |
| Putative Loci under Positive Selection FSTHET | 101 | 332 |

1 Steps included in the iterative filtering process

**Table 2.** Summary statistics of populations analyzed, observed heterozygosity (HO), expected heterozygosity (HE), and percentage of polymorphic loci (PL), effective population size \* (Ne).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neutral** | | | | | |  | **Adaptive1** | |  | **Adaptive2** | |
| **Location** | **N** | **HO** | **HE** | **% PL** | **Ne\*** | **CI** |  | **HO** | **HE** |  | **HO** | **HE** |
| REL | 20 | 0.312 | 0.324 | 0.997 | 1309.8 | 1199.2 - 1442.9 |  | 0.295 | 0.321 |  | 0.383 | 0.442 |
| MAN | 23 | 0.307 | 0.324 | 0.999 | 3516.8 | 2909.6 - 4443.6 |  | 0.274 | 0.303 |  | 0.213 | 0.215 |
| HOR | 27 | 0.301 | 0.317 | 1 | 3030.6 | 2624.1 - 3585.4 |  | 0.273 | 0.299 |  | 0.216 | 0.204 |
| CHE | 14 | 0.313 | 0.321 | 0.984 | 572.2 | 537.2 - 612.0 |  | 0.272 | 0.295 |  | 0.199 | 0.246 |
| CHA | 17 | 0.303 | 0.318 | 0.993 | 14454.8 | 6352.3 - Inf |  | 0.252 | 0.284 |  | 0.251 | 0.248 |

REL: Reloncaví Estuary, MAN: Manao, HOR: Hornopirén, CHE: Chepu, CHA: Chaitén

Index were calculated using 11 712 SNPs for neutral dataset, 338 SNPs for adaptive dataset merging loci identified by both PCADAPT and FSTHET analyses, and only using shared loci. \* Effective population size estimated based on Linkage Disequilibrium (LD; (Waples & Do 2010)). Inf: Infinite. 1 Putative adaptive merged loci for both PCADAPT and FSTHET analyses (338). 2 Only putative shared adaptive loci between both PCADAPT and FSTHET (20).

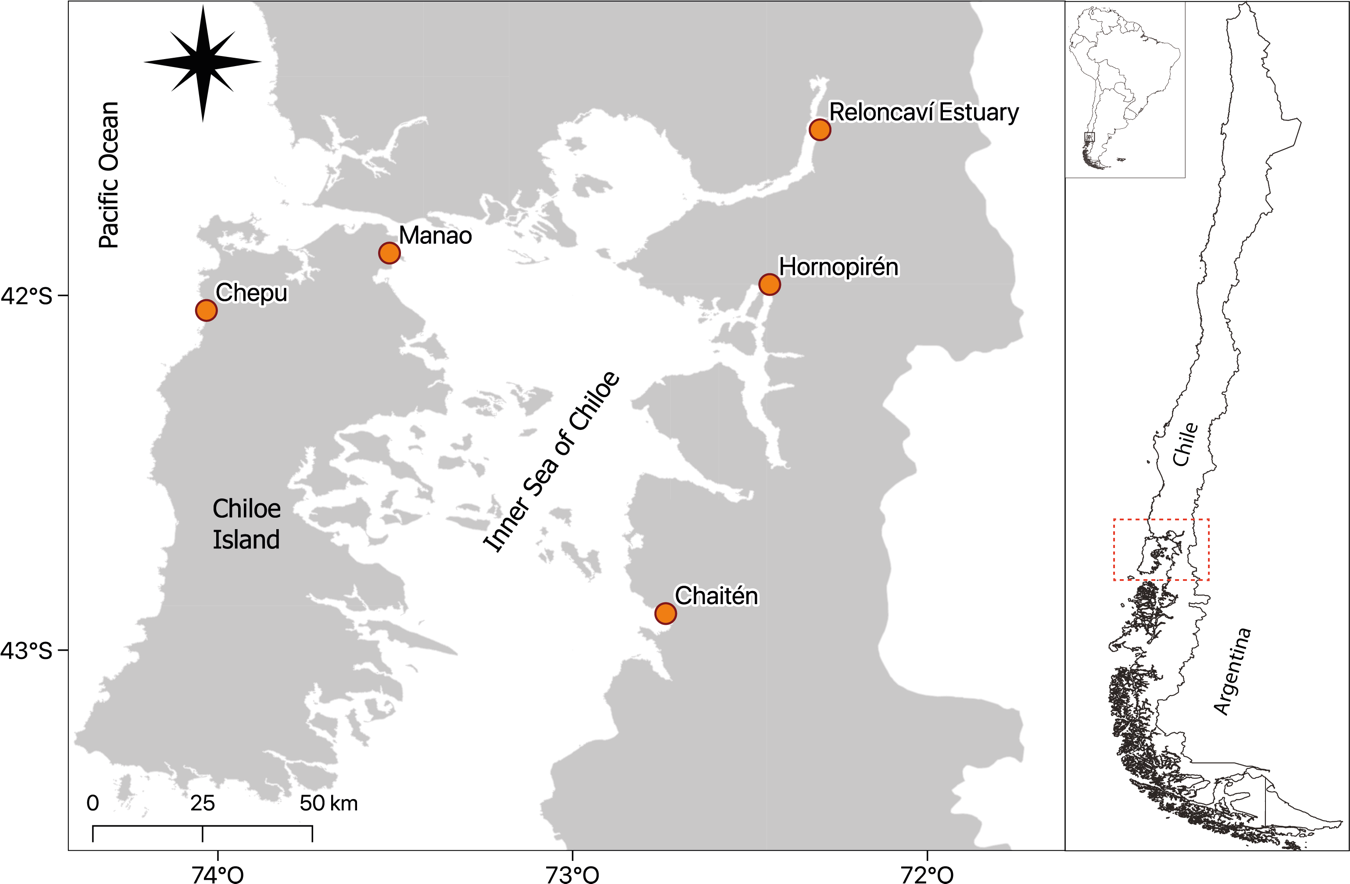
**Table 3.** Pairwise *F*ST values for neutral and adaptive datasets.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neutral** | | | | |  | **Adaptive1** | | | | |  | **Adaptive2** | | | | |
|  | **REL** | **MAN** | **HOR** | **CHE** | **CHA** |  | **REL** | **MAN** | **HOR** | **CHE** | **CHA** |  | **REL** | **MAN** | **HOR** | **CHE** | **CHA** |
| REL |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| MAN | 0.006 |  |  |  |  |  | 0.075 |  |  |  |  |  | 0.191 |  |  |  |  |
| HOR | 0.006 | 0.004 |  |  |  |  | 0.061 | 0.043 |  |  |  |  | 0.168 | 0.049 |  |  |  |
| CHE | 0.007 | 0.005 | 0.005 |  |  |  | 0.108 | 0.104 | 0.085 |  |  |  | 0.148 | 0.151 | 0.074 |  |  |
| CHA | 0.006 | 0.004 | 0.004 | 0.006 |  |  | 0.095 | 0.073 | 0.058 | 0.118 |  |  | 0.110 | 0.094 | 0.005 | 0.028 |  |

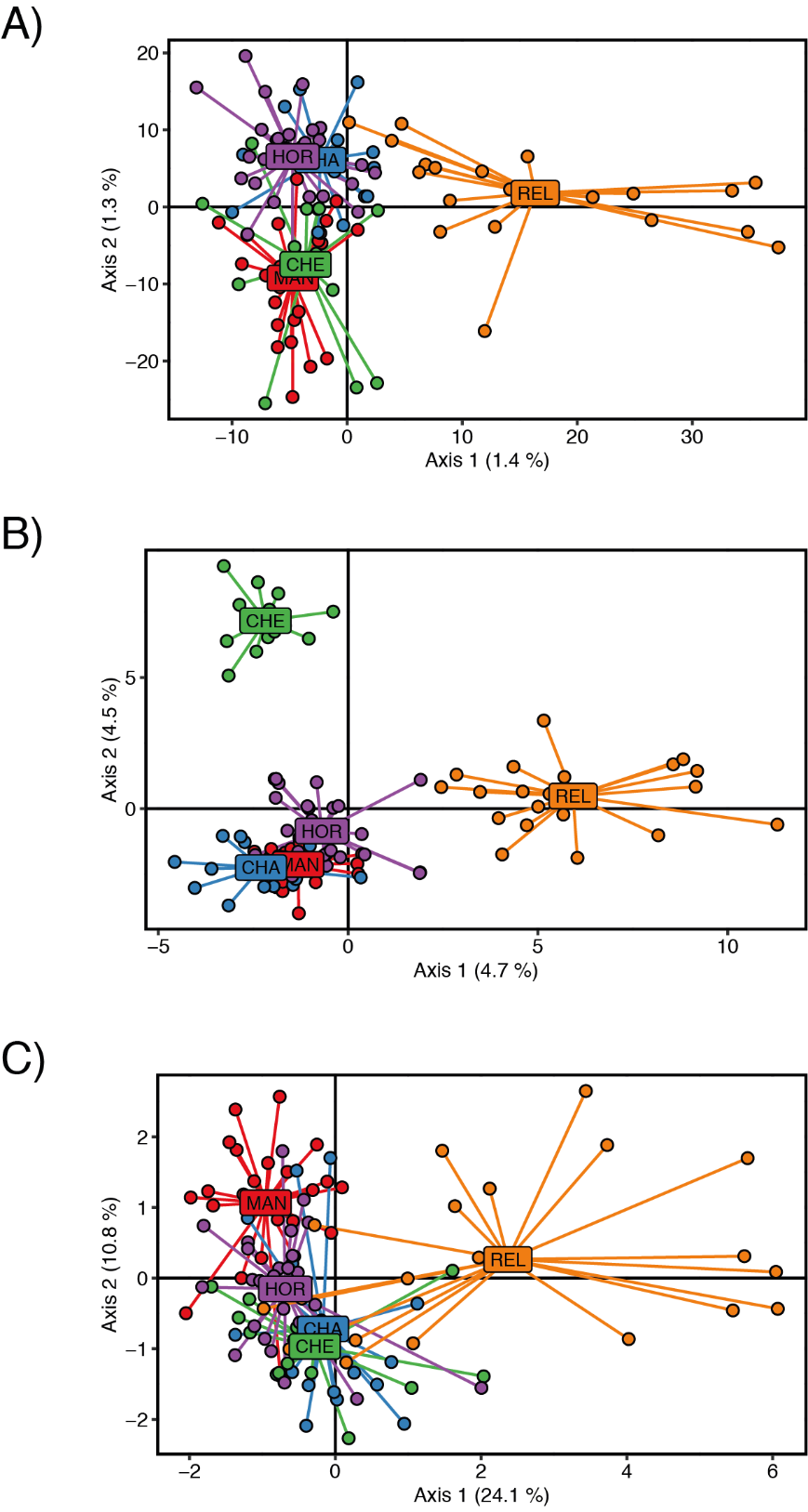
REL: Reloncaví Estuary, MAN: Manao, HOR: Hornopirén, CHE: Chepu, CHA: Chaitén

Index were calculated using 11,712 SNPs for neutral dataset, 338 SNPs for adaptive dataset merging loci identified by both PCADAPT and FSTHET analyses, and only using shared loci. 1 Putative adaptive merged loci for both PCADAPT and FSTHET analyses (338). 2 Only putative adaptive shared loci between both PCADAPT and FSTHET (20). *F*ST estimation was using Wright (1949) methods but corrected by Weir & Cockerham (1984) for uneven population size (see Pembleton et al. 2013).

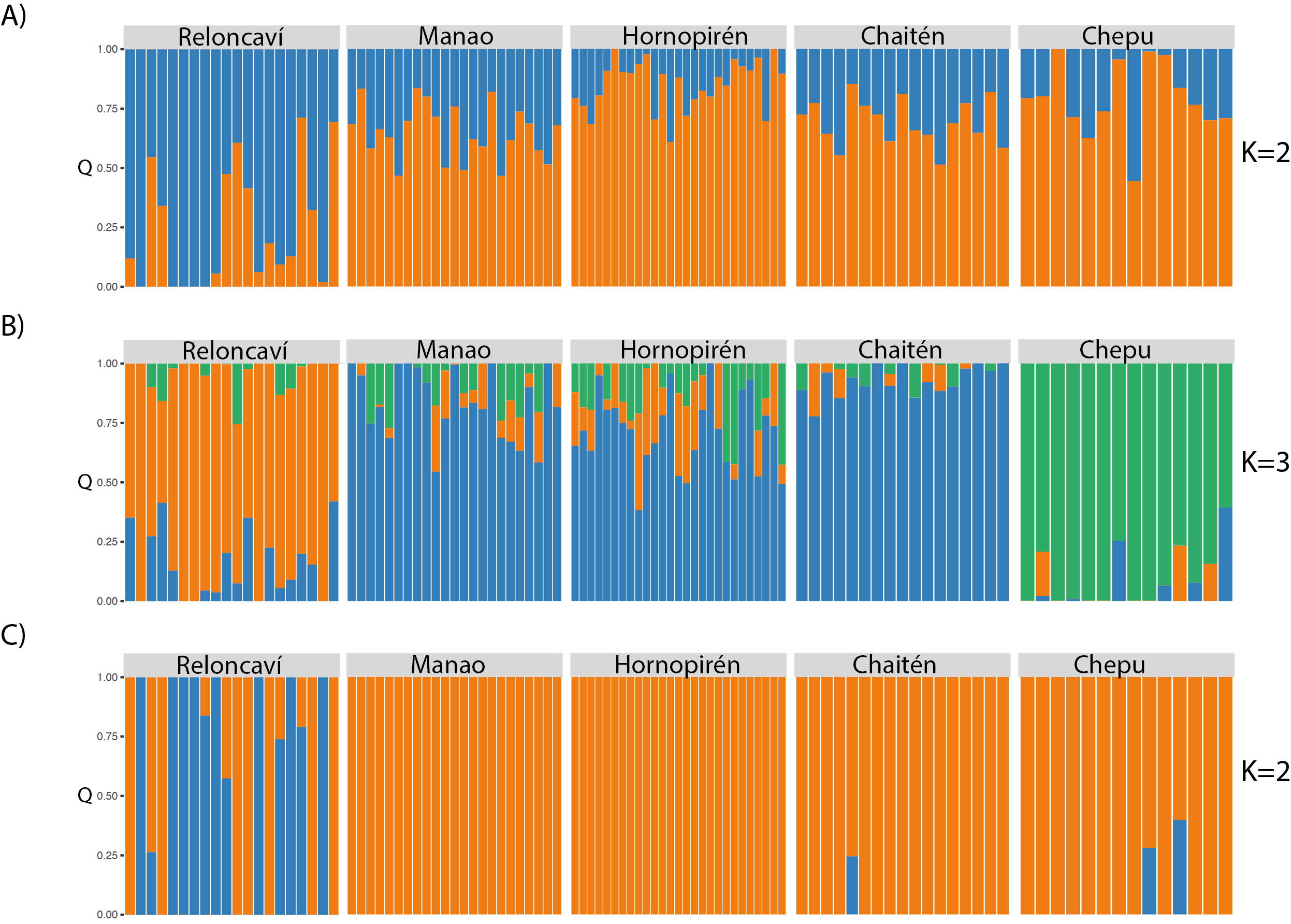
**Figures**

****

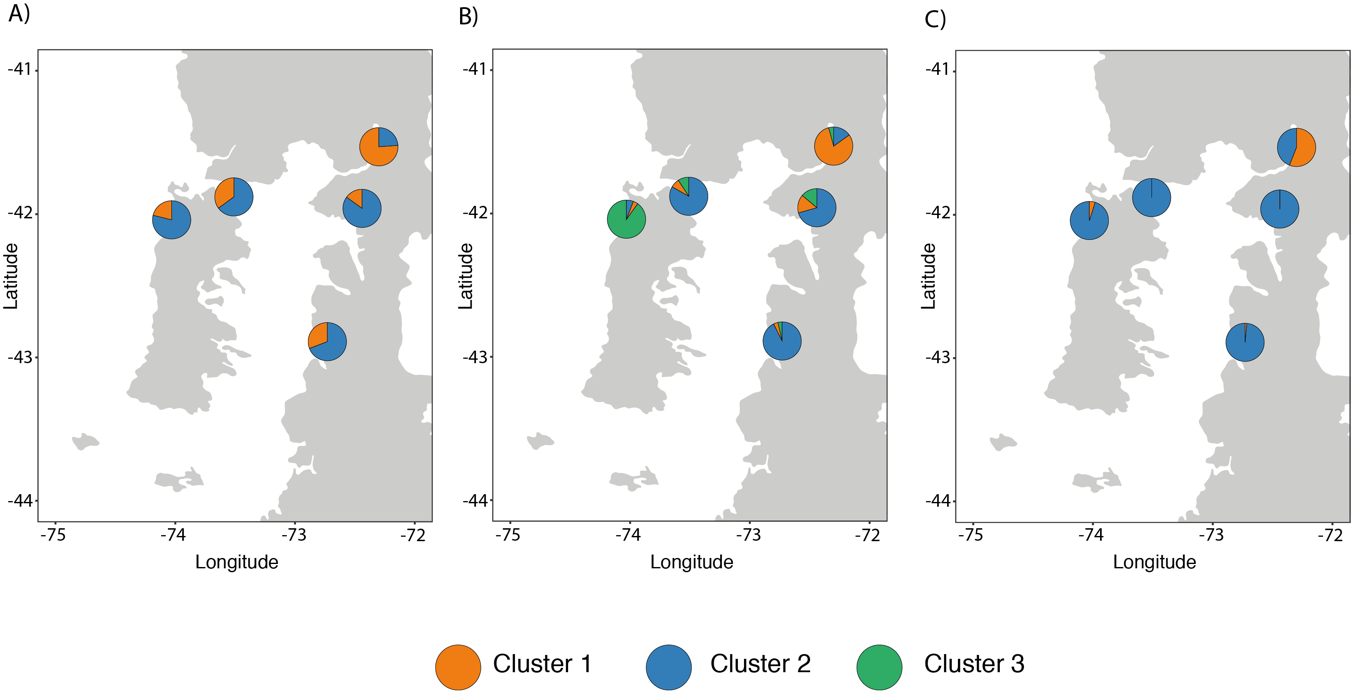
**Figure 1.** Map of sampling locations. REL: Reloncaví Estuary, MAN: Manao, HOR: Hornopirén, CHE: Chepu, CHA: Chaitén

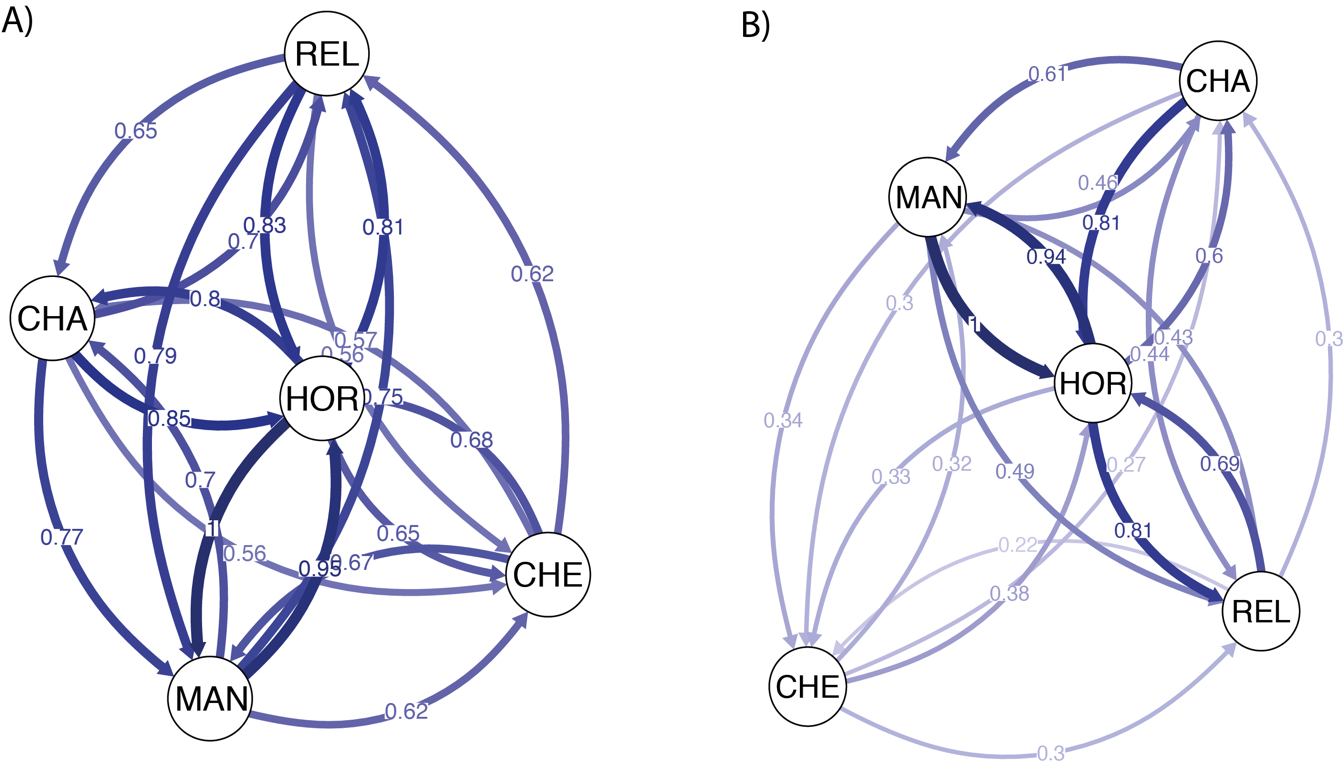


**Figure 2.** Individual-based principal component analysis for north Patagonian populations using (A) neutral dataset (11,712 SNPs), (B) adaptive merged loci (338 SNPs), and (C) adaptive shared loci for PCADAPT and FSTHET analyses (20 SNPs). REL: Reloncaví Estuary, MAN: Manao, HOR: Hornopirén, CHE: Chepu, CHA: Chaitén

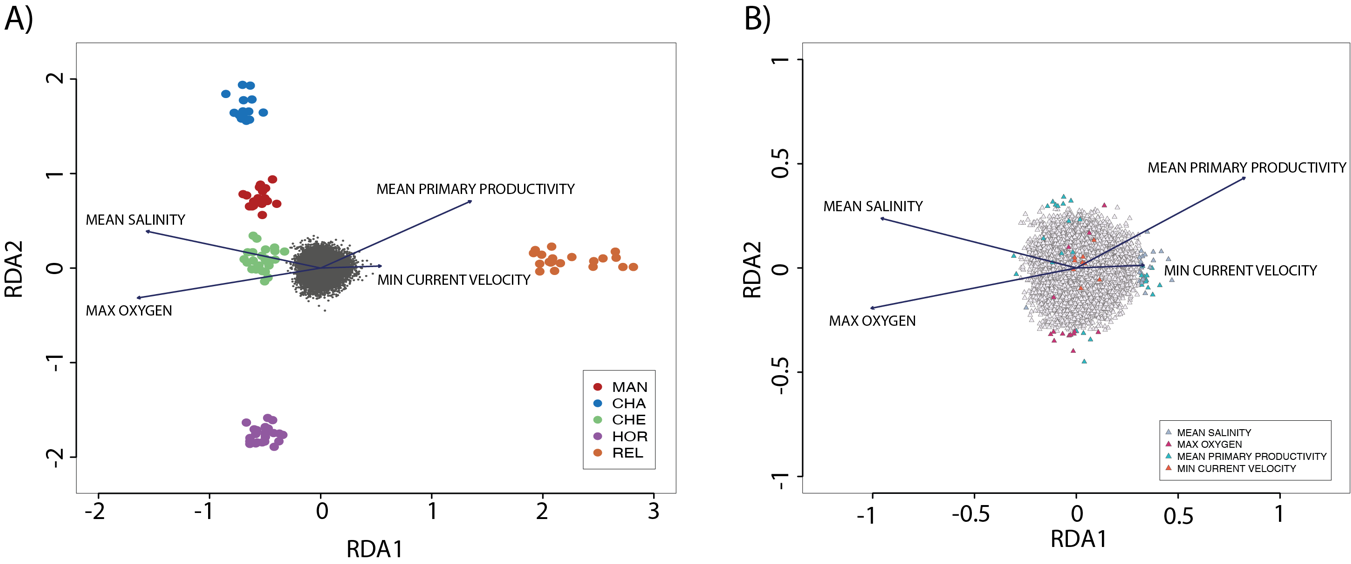


**Figure 3.** Admixture results showing the estimated population admixture coefficients (Q) for each individual, whose genome is broken into colored segments representing the proportion of that individual's genome derived from each of the K inferred clusters.Adaptive Individual-based principal component analysis for north Patagonian populations using (A) neutral dataset (11,712 SNPs), (B) adaptive merged loci (332 SNPs) and (C) adaptive shared loci (20 SNPs).

**Figure 4.** Mean admixture proportions by location in (A) neutral loci (11 712 SNPs), (B) adaptive merged loci for both PCADAPT and FSTHET analyses (338), and (C) adaptive shared loci between both PCADAPT and FSTHET (20 SNPs).



**Figure 5.** Directional relative migration network including all relative migration values among locations by (A) neutral and (B) adaptive merged dataset.

****

**Figure 6.** Triplots for RDA axes 1 and 2 for (A) individuals in sampling locations and (B) SNPs. In panel A, loci are represented by the dark gray cloud of points. Dots correspond to individuals which are colored by sampling location. Arrows (vectors) correspond to environmental predictors. In panel B, colored triangles represent loci associated with the environmental predictors.

**Data Archiving Statement**

Upon acceptance, demultiplexed sequence data used in this research will be archived in the NCBI sequence read archive. Genotypes for neutral and adaptive loci will be archived on DRYAD.