Article

Genomic Signatures of Local Adaptation under High Gene Flow in Lumpfish—Implications for Broodstock Provenance Sourcing and Larval Production

Simo Njabulo Maduna [1,*], Ólöf Dóra Bartels Jónsdóttir [2], Albert Kjartan Dagbjartarson Imsland [2,3], Davíð Gíslason [4], Patrick Reynolds [5], Lauri Kapari [6], Thor Arne Hangstad [7], Kristian Meier [8] and Snorre B. Hagen [1]

[1] Department of Ecosystems in the Barents Region, Svanhovd Research Station, Norwegian Institute of Bioeconomy Research, 9925 Svanvik, Norway; snorre.hagen@nibio.no
[2] Akvaplan-Niva Iceland Office, Akralind 6, 201 Kópavogur, Iceland; odj@akvaplan.niva.no (Ó.D.B.J.); aki@akvaplan.niva.no (A.K.D.I.)
[3] Department of Biological Sciences, High Technology Centre, University of Bergen, 5020 Bergen, Norway
[4] Matís, Vínlandsleið 12, 113 Reykjavík, Iceland; davidg@matis.is
[5] GIFAS AS, Gildeskål, 8140 Inndyr, Norway; pat.reynolds@gifas.no
[6] Akvaplan-Niva, Framsenteret, 9296 Tromsø, Norway; lauri.kapari@akvaplan.niva.no
[7] Lumarine AS, Stadionveien 21, 4632 Kristiansand, Norway; tah@lumarine.no
[8] Lerøy Aurora, Stortorget 1D, 9008 Tromsø, Norway; kristian.meier@leroyaurora.no
*   Correspondence: simo.maduna@nibio.no; Tel.: +47-902-89-810

Abstract: Aquaculture of the lumpfish (*Cyclopterus lumpus* L.) has become a large, lucrative industry owing to the escalating demand for "cleaner fish" to minimise sea lice infestations in Atlantic salmon mariculture farms. We used over 10K genome-wide single nucleotide polymorphisms (SNPs) to investigate the spatial patterns of genomic variation in the lumpfish along the coast of Norway and across the North Atlantic. Moreover, we applied three genome scans for outliers and two genotype–environment association tests to assess the signatures and patterns of local adaptation under extensive gene flow. With our 'global' sampling regime, we found two major genetic groups of lumpfish, i.e., the western and eastern Atlantic. Regionally in Norway, we found marginal evidence of population structure, where the population genomic analysis revealed a small portion of individuals with a different genetic ancestry. Nevertheless, we found strong support for local adaption under high gene flow in the Norwegian lumpfish and identified over 380 high-confidence environment-associated loci linked to gene sets with a key role in biological processes associated with environmental pressures and embryonic development. Our results bridge population genetic/genomics studies with seascape genomics studies and will facilitate genome-enabled monitoring of the genetic impacts of escapees and allow for genetic-informed broodstock selection and management in Norway.

Keywords: 3RAD; adaptive variation; aquaculture; cleaner fish; *C. lumpus*; genotype–environment association; seascape genomics; selective breeding

1. Introduction

Land- and marine-based farming of aquatic animals (aquaculture) is a lucrative industry that produces high amounts of seafood estimated at a record high 87.5 million tonnes, valued at USD 264.8 billion in 2022, to meet the increasing consumer demand [1]. Aquaculture production alleviates commercial fishing pressure on natural populations, and it is an increasingly important component of global food security [1,2]. One of the most industrialised fish farming enterprises in the world is the Atlantic salmon (*Salmo salar* L.) aquaculture industry in Norway [3,4]. Seven other countries also produce farmed Atlantic salmon and include (in descending order of production after Norway) Chile, UK, North America, Faroe Islands, Ireland, New Zealand, and Australia [5,6]. Currently, one of the

major constraints of further expansion of Atlantic salmon production in sea-based farming (mariculture) is disease outbreaks, specifically an ectoparasitic infestation of Atlantic salmon by copepod sea lice, which causes economic losses in Norway worth USD 436 million per year [4,7,8].

Sea lice cause direct damage to farmed stocks by feeding on their hosts' skin, mucous, and blood [9], with severe infestations causing skin erosion, physical damage, osmoregulatory failure, increased disease incidence, stress, and immunosuppression [10–12]. Disease outbreaks are primarily caused by the salmon louse (*Lepeophtheirus salmonis* K.; specialist; specific for salmonids) and the sea louse (*Caligus elongates* N.; generalist; less host-specific) [7,9,13], collectively referred to as "sea lice" [8]. Consequently, one of the key objectives in the Atlantic salmon aquaculture industry is the designing and integrating of pest management plans to minimise sea lice infestations in sea cages [4,7,14]. Numerous treatment methods have been developed and used over the years to combat infestations of salmon lice [7,15]. Sea lice control methods initially involved the use of chemical treatments [16,17] and then started using "environmentally friendly" alternative or supplement biocontrol methods, involving the use of "cleaner fish", because of the concerns surrounding the development of chemical-resistant strains of sea lice [8,18,19]. Consequently, the use of "cleaner fishes" that eat attached pre-adult and adult lice stages directly off the Atlantic salmon as an alternative to chemotherapeutics is not harmful to either the salmon, the environment, or the consumer, making this approach a leading contender [8,14,20].

Aquaculture of the lumpfish has become a large, lucrative industry owing to the escalating demand for cleaner fish to minimise sea lice infestations in Atlantic salmon mariculture farms [21–24]. Consequently, after Atlantic salmon and rainbow trout (*Oncorhynchus mykiss* W.), lumpfish is currently the third most profitable marine aquaculture species in Norway [25,26]. However, the stocking of cleaner fish in salmon sea cages has raised increasing concerns regarding the genetic integrity of natural populations following (accidental) escape events of cultured cleaners [27,28]. Accumulating genetics and genomics studies on natural populations of the lumpfish reveal varying patterns of population genetic structure across spatial gradients in the pan-Atlantic [28–33]. At large spatial scales, short tandem repeat (STR)-based studies revealed the existence of three major genetic groups of lumpfish: (1) western Atlantic, including samples from USA, Canada, and Greenland, (2) eastern Atlantic, including samples from Iceland and Norway, and (3) the Baltic [29–31], a pattern later confirmed with a panel of genome-wide SNPs [33]. At finer scales, a genetic break among northwestern and southwestern populations in Greenland was found [30] while no such break was apparent between northeastern and southeastern lumpfish populations in Norway [28,32]. Although, a divergent subpopulation in mid-Norway was found to be highly differentiated from other Norwegian sampling populations [31]. Moreover, a subset panel of 139 'diagnostic' SNPs revealed that there were at least two genetic clusters in Norway, although not entirely concordant with geography [33]. Further, evidence for the adaptation to local oceanographic features across the pan-Atlantic was also evident in lumpfish [33]. Indeed, the abiotic environment acts as a selective force on natural populations, shaping and maintaining variation in heritable traits (morphology, physiology, and behaviour) that provide significantly greater fitness in their local environment (i.e., local adaptation) when compared to populations from other environments [34–37]. Understanding the extent and scale of local adaptation is critical in determining how quickly and to what extent specific populations will respond to habitat changes, climate change, fisheries—or farming-induced evolution and interactions with a hatchery—or captive-reared counterparts [36,38–40].

Gene flow plays a complex and multifaceted evolutionary role in local adaptation as it has the capacity to either enhance or disrupt local adaptation, specifically by spreading advantageous ("adaptive") alleles or by homogenizing the gene pool [41–45]. Population genetics theory predicts that if a diversifying selection is not strong enough to prevent the loss in adaptive alleles, homogenizing gene flow between adaptively differentiated populations will "swamp out" whatever differences evolved in response to local environmental

conditions (i.e., gene swamping) [46,47] and can, at worst, transfer maladaptive alleles into differentially adapted recipient populations (i.e., migration meltdown) [47–49]. Mounting empirical evidence challenges the classic concepts of the "swamping" and "meltdown" effects of gene flow by demonstrating that not only can gene flow promote local adaptation, but adaptive polymorphisms can also be conserved within populations subject to high gene flow [43–45,50–53]. These contrasting effects of gene flow are a key part of the eco-evolutionary and conservation puzzle [44], where in the context of rising anthropogenic pressures on the natural world, understanding the role of gene flow in local adaptation has become critical for biodiversity conservation [44,51,54].

In order to further investigate the genetic basis of local adaptation in lumpfish along the Norwegian coastline, this study builds upon our previous research in population genetics. Our earlier work focused on the lumpfish population genetics using non-functional short tandem repeats (STRs) as genetic markers [28], followed by a subsequent study utilizing functional gene-associated STR markers [32]. In this current study, we aim to expand our analysis to the genomic level by examining genetic variation potentially associated with the local adaptation in lumpfish. To achieve this, we employ the triple-enzyme restriction-site-associated DNA sequencing (3RADseq) technique, as described by Bayona-Vásquez et al. [55], which provides genome-wide molecular data in the form of single nucleotide polymorphisms (SNPs) and copy number variants (CNVs). Specifically, the present study aimed to investigate whether local adaptation occurs in the face of high gene flow in the species and to identify potential environmental selective pressures that drive adaptation at a finer scale. We also hypothesised that selection coupled with stochastic oceanic dispersal and hierarchical spatial structure results in chaotic genetic patchiness, where lumpfish exhibits genetic differentiation between subpopulations at large spatial scales, with limited fine-scale geographic patterns of population differentiation.

2. Materials and Methods

2.1. Study Species

The lumpfish is a benthopelagic, cold-adapted marine species found throughout the North Atlantic that is characterised by a globose body, skin that is typically covered with tubercles, the pelvic fins modified to form a sucking disc, and the usual presence of two short dorsal fins [56–58]. Lumpfish are often found in low densities and dispersed over a vast geographical area, and adults are largely solitary with limited social interaction compared to schooling fish [56]. Studies on movement ecology show that lumpfish are capable of facultative iteroparity and display philopatry to specific spawning sites, i.e., reproductive philopatry [59,60]. Spawning occurs annually between late spring and early summer (April to May) in shallow coastal waters, where males establish territory and nesting sites prior to the arrival of females, and subsequent to their arrival and effective egg laying in the nests, males fertilise and guard the eggs until they hatch [56]. During brooding, the male is the sole provider of the care that is necessary for offspring survival (i.e., paternal care), and after hatching, the larvae attach to substrates (e.g., seaweed and floating seaweed clumps) via the sucking disc [56,58]. Juveniles remain in shallow water areas for approximately 6–12 months before gradually making their way to the feeding grounds offshore. Juveniles eat copepods, gammarids, and polychaetes [61], while adults eat crustaceans, ctenophores, jellyfish, and polychaetes [62].

2.2. Sampling Regime

We concentrated our lumpfish sampling in the northeastern Atlantic coastal region, with a particular emphasis on the Norwegian coast. Here, the coast of Norway stretches for nearly 2000 km from south to north, where coastal areas are characterised by complex bathymetry and topography [63–65]. The shelf along the coast has an average depth of 300 m and many fjords stretch in from the coast, some with deep basins reaching 600–1300 m, and further offshore is the much deeper Norwegian Sea. The Norwegian Coastal Current (NCC) is the principal oceanic current along the Norwegian coast. It is a

surface water mass that originates in the south (in the Skagerrak Strait) and combines with freshwater runoff from Norwegian fjords as it progresses northwards along the coast [64,65]. The North Atlantic Current (NAC) is another important oceanic current that flows beneath the NCC and periodically intrudes into the bank, bringing warm, saline, and nutrient-rich waters into some offshore locations, such as the Norwegian Froan archipelago. We collected a total of 107 tissue samples from 10 sites spanning coastal Norway (Figure 1), and sampling details are largely described in Jónsdóttir et al. [28] and Whittaker et al. [31]. In addition, we included two outgroup sample areas: Western Greenland (*n* = 15) and Canada (*n* = 15) in the northwestern Atlantic (Figure 1). Samples consisted of a small fin clip or muscle tissue from 5 to 15 lumpfish per site (Table 1).

2.3. Genetic Data

2.3.1. DNA Extraction, 3RAD Library Preparation, and Sequencing

We primarily utilised the same samples from previous studies to avoid sampling and temporal variation issues between studies [28,31], which was appropriate for our research objectives. For tissue samples from Canada, Greenland, and Norway (samples not part of Jónsdóttir et al. [28] and Whittaker et al. [31]), we used the DNeasy Tissue Kit following the manufacturer's instructions (Qiagen, Venlo, The Netherlands) to extract total genomic DNA from each sample. Then, we quantified all DNA stocks with the Quantus™ Fluorometer using the QuantiFluor® ONE dsDNA System (Promega, Promega, WI, USA). We normalised each DNA sample to 10 ng/μL in nuclease and ion-free water and stored the working stocks at −21 °C prior to 3RAD library preparation.

We prepared RADseq libraries using the Adapterama III library preparation protocol of Bayona-Vásquez et al. [55] (their Supplemental File SI), which is a modified version of double-digest (dd) RAD [66] that uses three enzymes for digesting genomic DNA (3RAD). This procedure reduces the amount of DNA required since the third enzyme inhibits the formation of adapter dimers and DNA chimeras during the simultaneous digestion and ligation reactions [55,67,68]. We surveyed the literature and identified the frequently used enzymes in RADseq experiments of fish, and we chose a pair of enzymes that were compatible with one of the 3RADseq designs (Table 1 in [55]). We selected *MspI* (C|CGG) as the frequent cutter and *BamHI-HF* (G|GATCC) as the rare cutter, which fit the 3RAD Design 2 with *ClaI* (AT|CGAT) as the third enzyme for suppressing phosphorylated ends in *MspI* recognition sites. A detailed account of the laboratory procedure of the 3RAD library preparation can be found within the Supplementary Materials.

We included 16 technical replicates to estimate genotyping error rates within and between lanes. Cleaned and indexed library pools per design were sent to the Norwegian Sequencing Centre (NSC) for quality control and subsequent final size selection using a one-sided bead clean-up (0.7:1 ratio) to capture 550 bp ± 10% fragments, and the final paired-end (PE) 150 bp sequencing on one lane of the Illumina HiSeq 4000 platform.
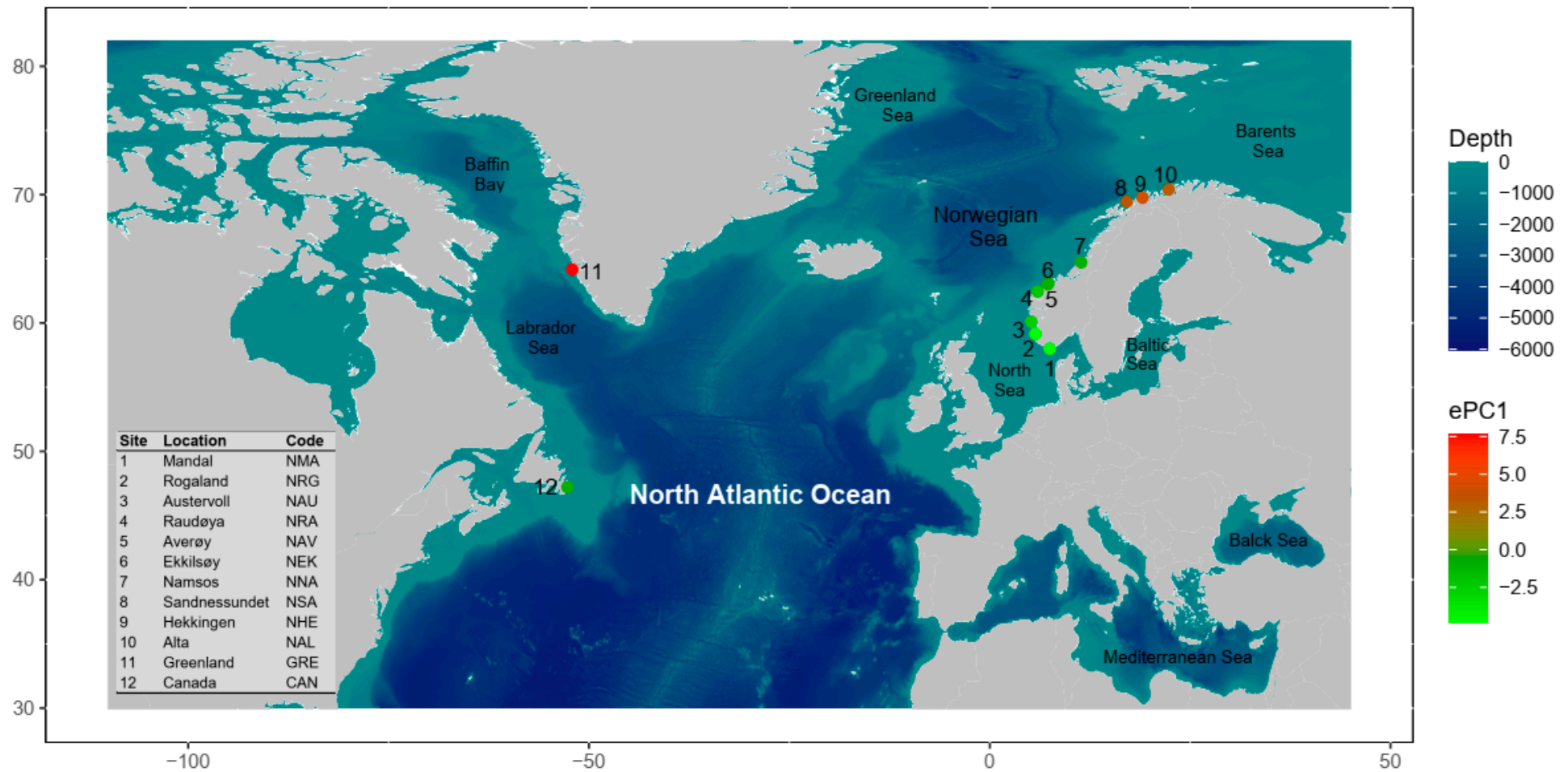
Figure 1. Sampling localities of lumpfish *C. lumpus* along the Norwegian coast in the northeastern Atlantic (Sites 1–10) and western Greenland (Site 11) and Canada (Site 12) in the northwestern Atlantic Ocean. Sampling sites are colour coded based principal component (PC) analysis (PCA) of environmental variables to describe patterns of the environmental variation in our sampling regime. The ePC1 represents the first PC loadings that explain the largest part of the total variance of 41%.

Table 1. Information about sampling localities for lumpfish along the Norwegian coast and genetic variation descriptors: number of samples before filtering, followed by the number of individuals, which passed all filtering steps within parentheses (N), number of private alleles (A_P), allelic richness (A_R), observed heterozygosity (H_O), gene diversity (H_S), inbreeding coefficient (F_{IS}) with 95% confidence interval, and sampling population-specific F_{ST} (β_{WT}) with 95% confidence interval.

Location	Code	Region	Latitude	Longitude	N	A_P	A_R	H_O	H_S	F_{IS} (95% CI)	β_{WT} (95% CI)
Mandal	NMA	Southern Norway	57.99	7.48	10 (9)	1	1.11	0.159	0.161	0.012 (0.005–0.020)	0.023 (0.014–0.031)
Rogaland	NRG	Southern Norway	59.15	5.70	11 (8)	1	1.15	0.155	0.151	−0.023 (−0.032−−0.015)	0.102 (0.092–0.113)
Hardangerfjord	NHA	Southern Norway	59.75	5.55	11 (0)	-	-	-	-	-	-
Austervoll	NAU	Southern Norway	60.10	5.19	7 (7)	0	1.17	0.158	0.165	0.041 (−0.011–0.013)	0.022 (0.010–0.033)
Raudøya	NRA	Southern Norway	62.45	5.97	5 (4)	0	1.17	0.163	0.165	0.014 (0.003–0.025)	0.020 (0.007–0.032)
Averøy	NAV	Southern Norway	63.01	7.23	10 (10)	4	1.16	0.162	0.163	0.006 (−0.001–0.014)	0.028 (0.020–0.037)
Ekkilsøy	NEK	Southern Norway	63.07	7.33	5 (4)	0	1.16	0.160	0.156	−0.022 (−0.034−−0.010)	0.066 (0.054–0.079)
Namsos	NNA	Central Norway	64.72	11.41	10 (10)	1	1.17	0.162	0.165	0.017 (0.010–0.024)	0.021 (0.013–0.030)
Sandnessundet	NSA	Northern Norway	69.76	19.05	12 (11)	3	1.16	0.159	0.156	−0.020 (−0.027−−0.013)	0.063 (0.055–0.071)
Hekkingen	NHE	Northern Norway	69.45	17.10	11 (11)	8	1.16	0.161	0.160	−0.008 (−0.015−−0.001)	0.047 (0.040–0.055)
Alta	NAL	Northern Norway	70.40	22.31	15 (14)	25	1.16	0.155	0.164	0.056 (0.050–0.063)	0.026 (0.018–0.033)
Sermersooq	GRE	Western Greenland	64.17	−52.09	15 (15)	243	1.15	0.151	0.154	0.020 (0.014–0.027)	0.087 (0.074–0.099)
Canada	CAN	Eastern (Atlantic) Canada	47.21	−52.69	15 (14)	180	1.13	0.135	0.135	−0.001 (−0.008–0.006)	0.188 (0.173–0.201)
		Overall			137 (117)	38.8	1.15	0.157	0.159	0.008 (−0.004–0.013)	0.058 (0.054–0.061)

2.3.2. SNP Discovery and Genotyping

After assessing the quality of the sequencing run post demultiplexing to plate-level using the plate-specific i7 indexes using FASTQC [69] plus MULTIQC v.2.31 [70], we processed the data as follows: First, we demultiplexed to the sample level using internal combinatorial barcodes and cleaned and trimmed the raw sequences with the perl script *process_radtags.pl* included as part of STACKS v.2.51 [71]. We ran the script using the "inline_inline" mode because the internal barcodes are part of the sequence in both Illumina reads. We removed reads with uncalled bases (-*c*) and discarded reads with low quality scores (-*q*) with a default sliding window of 15% of the length of the read and raw Phred score of 20 to retain high-quality reads. We specified Read 1 (*MspI*) and Read 2 (*BamHI-HF*) restriction enzymes, and we rescued sequence tags (internal tags) and RAD tags (enzyme overhang) within 2 mismatches of their expected sequence (-*r*); otherwise, reads were discarded. We truncated (-*t*) PE150 reads to 140 nt to have equal length among all reads with different barcodes. Second, we downloaded the reference genome of the lumpfish (accession number GCF_009769545.1; assembly version: fCycLum1.pri; scaffolds: 48; contigs: 395; N50: 4,950,682; L50: 35) from NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/; accessed on 10 March 2020) under BioProject PRJNA562003, the Vertebrate Genomes Project (G10K Consortium) [72]. Third, we then constructed the FM-index for the reference genome and individually aligned the cleaned, demultiplexed PE reads from each fish to the indexed reference genome using the BWA-MEM algorithm of BWA v.0.7.17 aligner [73], excluding reads with a minimum quality score of <30. Alignments were sorted and indexed, and read pairs were fixed using tools from the SAMTOOLS v.1.9 suite [74]; we obtained alignment quality control (QC) statistics for the sorted and indexed alignments using BAMQC as implemented in QUALIMAP v.2.2.20 [75]. Finally, we used the alignments that passed QC to assemble RAD loci, call SNPs, and construct genotypes for individual fish using the perl script *ref_map.pl* also implemented in STACKS, which consisted of two components (*gstacks* and *populations*). We ran population analysis in *ref_map.pl*, with (i) a minimum of 75% of individuals in a population required to process a RAD locus for that population (-*r* 0.75), (ii) a minimum of 75% of populations a RAD locus must be present to process a locus (-*p* 9), (iii) by discarding unpaired reads (--*rm-unpaired-reads*), (iv) a minimum mapping quality of 20 to consider a read (--*min-mapq* 20), and (v) a minimum minor allele count (MAC) of 3 required to process a SNP at a RAD locus (--*min-mac* 3).

2.3.3. Data Filtering

Recent research on marine species indicates that copy number variants (CNVs) are a significant source of genetic variability in genomes [76,77]. Therefore, we based our data filtering on the prior work of McKinney et al. [76] and Dorant et al. [77] to divide our 3RAD-seq dataset for the lumpfish containing all called SNPs from STACKS into two datasets, one containing single-copy loci (hereafter singleton SNPs) and the other containing multiple-copy loci (hereafter duplicated SNPs, CNV). We filtered genotype data and characterised singleton and duplicated SNP loci using filtering procedures and custom scripts available in STACKS WORKFLOW v.2 (https://github.com/enormandeau/stacks_workflow; accessed on 25 March 2023). First, we filtered the 'raw' VCF file keeping only SNPs that (i) showed a minimum depth of four (-*m* 4) and (ii) were called in at least 80% of the samples in each site (-*p* 80), (iii) for which at least two samples had the rare allele, i.e., Minor Allele Sample (MAS; -*S* 2), using the python script 05_*filter_vcf_fast.py*. Second, we excluded those samples with more than 20% missing genotypes from the dataset. Third, we calculated pairwise relatedness between samples with the Yang et al. [78] algorithm and individual-level heterozygosity in VCFTOOLS v.0.1.17 [79]. Additionally, we calculated pairwise kinship coefficients among individuals using the KING-robust method [80] with the R package SNPRELATE v.1.28.0 [81]. Then, we estimated genotyping error rates between technical replicates using the software TIGER [82]. Finally, we removed one of the pairs of closely related individuals exhibiting the higher level of missing data along with samples that showed extremely low heterozygosity (< -0.2) from graphical observation

of individual-level heterozygosity per sampling population. Fourth, we conducted a secondary dataset filtering step using 05_*filter_vcf_fast.py*, keeping the above-mentioned data filtering cut-off parameters (i.e., -*m* = 4; -*p* = 80; -*S* = 3). Fifth, we calculated a suit of four summary statistics to discriminate high-confidence SNPs (singleton SNPs) from SNPs exhibiting a duplication pattern (duplicated SNPs): (i) median of allele ratio in heterozygotes (*MedRatio*), (ii) proportion of heterozygotes (*PropHet*), (iii) proportion of rare homozygotes (*PropHomRare*), and (iv) inbreeding coefficient (F_{IS}). We calculated each parameter from the filtered VCF file using the python script 08_*extract_snp_duplication_info.py*. The four parameters calculated for each locus were plotted against each other to visualise their distribution across all loci using the R script 09_*classify_snps.R*. Based on the methodology of McKinney et al. [76] and Dorant et al. [77], and by plotting different combinations of each parameter, we graphically fixed cut-offs for each parameter. Sixth, we then used the python script 10_*split_vcf_in_categories.py* for classify SNPs to generate two separate datasets: the "SNP dataset", based on SNP singletons only, and the "CNV dataset", based on duplicated SNPs only. Seventh, we postfiltered both datasets by keeping all unlinked SNPs within each 3RAD locus using the 11_*extract_unlinked_snps.py* script with a minimum difference of 0.5 (-*diff_threshold* 0.5) and a maximum distance 600 bp (-*max_distance* 500) to consider linked SNPs. Then, for the "SNP dataset", we filtered out SNPs that were located in unplaced scaffolds, i.e., contigs that were not part of the 25 chromosomes of the lumpfish genome. We used PGDSPIDER v.2.1.1.5 [83] to transform some VCF files into input files for multiple programs. Lastly, to construct the CNV dataset, we extracted the locus read depth of SNPs identified as duplicated using VCFTOOLS and performed normalisation using the trimmed mean of *M*-values method originally described for RNAseq count normalisation and implemented in the R package *edgeR* [84]. The correction adjusts for the fact that an individual with a higher copy number at a particular locus will contribute proportionally more to the sequencing library than an individual with a lower copy number at that locus. The resulting CNV dataset comprised a matrix of normalised read counts for each individual at each CNV locus [77].

2.3.4. Genome Scans and Signatures of Selection

We investigated if SNPs were putatively under selection using a population-level method (BAYESCAN) and two individual-level methods (PCADAPT and OUTFLANK). For BAYESCAN v.1 [85,86], we set the prior odds (*pr_odds*) at 10,000 after explorative runs (*pr_odds* = 100; 1000), which is appropriate for the number of markers in our datasets [86,87], and ran the model using 10,000 iterations with a thinning interval of 10, a burn-in of 200,000 steps, and 20 pilot runs of 5000 iterations. Then, we assessed convergence by visual inspection of trace plots, and additional diagnostic tests were carried out to confirm convergence, such as obtaining estimates of autocorrelation and effective sample size, Geweke's convergence diagnostic, and Geweke–Brooks plots [88] with the R-package CODA v.0.19.4 [89]. For further post-processing of BAYESCAN output in R, we defined a false discovery rate (FDR) *q*-value threshold of 0.05 and defined the respective α and *q* combinations to classify SNPs: (i) positive α and $q \leq 0.05$ as suggestive of diversifying selection, (ii) negative α and $q \geq 0.05$ as suggestive of balancing selection, and (iii) positive α and $q \geq 0.05$ as suggestive of neutrality. For the respective individual-level methods, we used PCADAPT v.4.3.3 [90] and OUTFLANK v.0.2 [91] to carry out the analyses, and for each, we examined screen plots to determine the number of populations (*K)* and used the first 10 components that captured the majority of the population structure in the data. To control the false positive, we set the F_{ST} threshold value to 0.05 in the OUTFLANK approach and the FDR threshold value to 0.05 in the PCADAPT approach. While BAYESCAN is suitable for our population-level sampling design, PCADAPT and OUTFLANK are more reliable for species with complex, hierarchical population structure and are less sensitive to admixed individuals and outliers in the data [85,90,92]. Thus, to derive a putatively neutral SNP dataset for estimating population genetics parameters for GEA analysis, we removed outliers identified in any of the three methods to be putatively under selection.

2.3.5. Genetic Diversity and Differentiation

To characterise genetic diversity, we calculated the average observed heterozygosity (H_O), and within population gene diversity (H_s), Wright's inbreeding coefficient (F_{IS}), and a population-specific index of differentiation relative to all other sampling populations (β_{WT}) [93] using the R package *hierfstat* v.0.5-11 [94,95]. We computed the 95% confidence intervals of F_{IS} and β_{WT} values using the bootstrap method (1000 bootstraps) as implemented in *hierfstat*. We also calculated the number of private alleles at each locus and sampling population with the function '*privateAlleles*' of the R package *StrataG* v.2.5.01 [96]. We calculated genome-wide pairwise F_{ST} estimates using the Weir and Cockerham [97] method with the function '*pairwise.WCfst*' in *hierfstat* and computed the 95% confidence intervals of F_{ST} values using the bootstrap function '*boot.ppfst*' (*nboot* = 1000) as implemented in *hierfstat*. Indices were considered significant if the 95% credible set did not include zero. We estimated the population genetic differentiation of loci identified as CNVs by calculating the variant fixation index V_{ST} [98], an analog of the F_{ST} estimator of population differentiation [97], commonly used to identify differentiated CNV profiles between populations and species [77,98–100]. For each pairwise population comparison, we calculated V_{ST} as follows:

$$V_{ST} = \frac{\left(V_T - \left(V_i \times N_i + V_j \times N_j\right)\right)/N_T}{V_T} = \frac{V_T - V_S}{V_T}$$

where V_T is the total variance of normalised read depths among all individuals from populations *i* and *j*; V_i and V_j are the normalised read depth variance for populations *i* and *j*, respectively; N_i and N_j are the sample size for populations *i* and *j*, respectively; N_T is the total sample size; and -V_S is the average of the variance within each population, weighed for population size [77,98,100]. We performed permutation tests on the normalised read depths, where we randomly permuted all population *i* and *j* individuals and calculated a new V_{ST} for every CNV. This process was repeated 1000 times creating a distribution of V_{ST} values for each CNV using the R script available at https://github.com/DaRinker/PolarBearCNV (accessed on 10 February 2023). We calculated 95% confidence intervals, and indices were considered significant if the 95% credible set did not include zero. We also established genome-wide standard cut-offs of $V_{ST} > 0.1 < 0.3$ and $V_{ST} > 0.3$ based on the results from population comparisons of a marine species [77] and available species comparisons [100], respectively, to identify outlier CNV likely under selection.

2.3.6. Genetic Clustering and Connectivity

To investigate population clustering, we first computed a matrix of Nei's genetic distance D_A [101,102] for each pair of populations with the function *stamppNeisD()* of the R package STAMPP v.1.6.3 [103]. The distance matrix was then used to perform a principal coordinate analysis (PCoA) using the function *pcoa()* of the R package APE v.5.6. [104]. Second, we examined genetic clustering with Discriminant Analysis of Principal Components (DAPC) using the R package ADEGENETv.2.1.5 [105,106], both with and without prior population assignment. We first used the adegenet function *find.clusters()* testing up to 30 clusters and using Bayesian information criteria (BIC) to identify the best-fit number of clusters for our data. Using the number of clusters with the lowest BIC value, we performed a DAPC with the ADEGENET function *dapc()* and plotted the results in R. The number of PCs was set based on the *a*-score. Third, we inferred the genomic admixture of lumpfish using a sparse non-negative matrix factorisation (snmf) as implemented in the *snmf()* function in the R-package LEA v. 3.8.0 [107,108]. We calculated estimates of individual admixture coefficients over a range of *K* values (1–10). We determined the number of ancestral populations (*K*) using the entropy criterion [107,109]. Ancestry coefficients (*Q*) were visualised by plotting the *Q*-values of each individual in a bar plot using the R package POPHELPER v.2.3.1 [110]. Last, we assessed the genomic co-ancestry among individuals with FINERADSTRUCTURE [111]. We first inferred a co-ancestry matrix with the script *RADpainter*. Subsequently, clustering was performed with the Markov chain Monte

Carlo method of FINERADSTRUCTURE, running for 500,000 generations and sampling every 1000 generations; the first 200,000 generations were discarded as burn-in (non-default parameters: -*x* 200,000 -*y* 300,000 -*z* 1000). We also inferred a tree for visualisation with FINERADSTRUCTURE using the tree-building algorithm of Lawson et al. [112] with 10,000 attempts (non-default parameters: -*m* T -*x* 10,000). FINERADSTRUCTURE results were plotted with R scripts included in the FINERADSTRUCTURE package. We only used PCAs to visualise the pattern of genetic differentiation based on all CNVs in R.

2.4. Environmental Data

Seascape genomics analyses require a comprehensive characterisation of the marine environment in order to avoid the confounding influence of collinear gradients [113–115]. As a result, the seascape characterisation we employed included 16 environmental variables, which include climate and biogeochemical variables: sea surface temperature (SST), sea surface salinity (SSS), eastern sea current velocity (ESCV), northern sea current velocity (NSCV), chlorophyll concentration (CHLa), chlorophyll mass concentration (CHLm), phytoplankton concentration (PHYC), net primary production (PP), suspended matter (SPM), dissolved oxygen (O_2) concentration (hereafter abbreviated as DO), nitrate (NO_3) concentration (hereafter abbreviated as NO3), phosphate (PO_4) concentration (hereafter abbreviated as PO4), silicate (SiO_4^{4-}) concentration (SI), dissolved iron (Fe) concentration (hereafter abbreviated as FE), surface partial pressure of carbon dioxide in sea water (pCO_2^{sea}; hereafter abbreviated as SPCO2), and pH (hereafter abbreviated as PH). We obtained these variables from five respective georeferenced datasets describing atmospheric and seawater conditions from the Copernicus Marine Environment Monitoring Service (CMEMS: https://marine.copernicus.eu/, accessed on 10 November 2022; Table S1). We obtained the average monthly values down to 156 m in depth between 1993 and 2014, with a spatial resolution ranging from ~9 to 4 km, from CMEMS, before the genetic data were sampled (2015–2017) [28,31]. We chose to acquire average daily values for SST, SSS, ESCV, NSCV, and CHLa because these variables are typically the main environmental drivers of local adaptation in the marine environment for species with a planktonic larval phase [116–119].

We processed these variables in the R environment using the RASTER package v.2.8 [120] to compute the overall mean (OM) and standard deviation (OSD) associated with the mean for each variable. CMEMS provides information on sea current velocity in two separate datasets, ESCV and NSCV, which represent the water movement towards east and towards north, respectively. Therefore, prior to computing the overall mean and standard deviation of the overall sea current velocity (SCV), we considered the ESCV and NSCV as perpendicular vectors of water velocity, where the Euclidean norm of these vectors is the overall SCV. We computed the overall SCV as the square root of the sum of squares of the ESCV and NSCV.

Moreover, given the importance of temperature in local adaptation, we also computed four seasonal statistics for SST, which include the mean and standard deviation temperature of the coldest and hottest month, respectively. In total, we computed 34 environmental variables, which we used to summarise the patterns of environmental variation across sampling sites using principal component and hierarchical clustering analyses in base R with the *prcomp()* and *hclust()* functions.

In order to avoid multicollinearity and select the most seemingly ecologically relevant variable, we used the *corSelect()* function of the R program *fuzzySim* v.4.3 [121] with a threshold value of the correlation coefficient $-0.5 \leq |r| \geq 0.5$ and excluded the variable with the highest Variance Inflation Factor (VIF) [122,123] among correlated variables. Nevertheless, we identified groups of highly collinear environmental variables using a hierarchical clustering implemented in the R package KLAR v.1.7-2 [124] for the downstream interpretation of genotype–environment associations to facilitated comparison with other studies.

2.5. Spatial Data

We implemented the spatial eigenfunction analysis (SEA) and isolation-by-distance (IBD) approaches to evaluate the relative contribution of the geographic distance on multiscale spatial patterns of genetic variation in our system. For both analyses, we used the F_{ST}-based genetic distance matrix calculated from the neutral SNP datasets consisting of the (i) global (Norway, including outgroups) and (ii) local (Norway) sampling sites. For SEA, we used distance-based Moran Eigenvector Maps (dbMEM) [125,126] to model spatial structure across our sampling sites. Briefly, dbMEMs represent structure in a dataset with orthogonal axes to describe complex patterns of spatial structure in rectangular form suitable for use as a cofactor to control for spatial correlation in downstream statistical tests of y~x relationships, e.g., partial redundancy analysis (RDA) in GEA studies [127,128]. Importantly, these independent vectors can model both broad- and fine-scale patterns, which may affect genetic variation partitioning across large and/or highly heterogeneous landscapes. To calculate dbMEMs, we first converted degrees latitude and longitude to Cartesian coordinates with the *geoXY()* function available in the R package SODA v.1.0-6.1 [129]. Then, we computed a Euclidian distance matrix on the Cartesian coordinates using the *dist()* function and also transformed geographical distances as in-water distances using the *lc.dist()* function of the R package MARMAP v.1.0.6 [130]. We generated two sets of dbMEM variables by decomposing (i) Euclidean (geographic) and (ii) in-water (least-cost) distances between sites, respectively, with the *dbmem()* function implemented in the R package ADESPATIAL v. 0.3-20 [131]. We used the *cor.test()* function to determine the correlation between 'Euclidean' and 'in-water' dbMEMs. Last, we performed two site-level RDAs on sites' allelic frequencies and geographical distances using (i) in-water distances and (ii) significant dbMEM with the *rda()* function. To test for IBD, we used the *mantel.rtest()* function of the R package ADE4 v.1.7-20 [132] with 10,000 permutations using Euclidean and in-water distances.

2.6. Seascape Genomics

We carried out a seascape genomics analysis to investigate the possible correlation between environmental variables and the frequency of particular genotypes of lumpfish globally and regionally. Associations of this kind might reveal an environmental constraint requiring adaptation in lumpfish at different scales, as well as the genetic features conferring the selective advantage. For genotype–environment association analyses, we combined univariate and multivariate methods, namely latent factor mixed models (LFMM) and redundancy analysis (RDA). While LFMM identifies associations between single loci and single predictors, RDA can detect multilocus signatures of selection as a function of a multivariate set of predictors [133,134]. The LFMM and RDA approaches have been used extensively in the field [135,136] and provide a good compromise between detection power and error rates and are robust to a variety of sampling designs and underlying demographic models [113,133]. Both methods assume a linear relationship between allele frequency and environmental variables and require complete datasets (without missing values). For this reason, we imputed missing genotypes, however, using the most common genotype at each SNP across all individuals instead of ancestry-based missing data imputation given that our study species is characterised by genetic homogeneity between many sampling populations and the low levels of missing data (<5% in our datasets, see Results). To apply LFMM, we first used the R function *rda()* function in VEGAN v. 2.6-4 [137] to perform a principal component analysis (PCA) on environmental predictor variables to generate synthetic predictors, instead of using raw predictor variables, in order to minimise the number of tests. We determined the optimal number of principal components (PCs) to retain using the broken-stick criterion [138] with the *screeplot()* function in VEGAN, which suggests retaining only those components whose eigenvalue is larger than the value given by the broken-stick distribution [138,139]. We used the *scores()* function as implemented in VEGAN to determine correlations between the PC axis and predictors. Similarly, we applied the same approach to estimate population structure in the genotypic data and determined

the optimal number of principal components (PCs) to retain. After determining how many PCs explained most of the genotypic variance (*K*, for setting latent factors) and predictor variance, we conducted the LFMM analysis with the newer *lfmm2* implementation that uses a least-squares estimation method to compute LFMMs using the R package LFFM v.1.1 [134]. We implemented the *lfmm_ridge()* function for estimating the parameters of LFMMs, where the ridge estimates are based on minimising a regularised least-squares problem with an L_2 penalty [134]. We then post-processed the model outputs for each PC with the *lfmm_test()* to perform association testing using the fitted model and calibrated *p*-values using the *gif* option that calculated the genomic inflation factor (GIF; λ). The *lfmm_test()* function returns *p*-values describing the statistical significance of every genotype–environment association under different values of *K*. For each association model related to the same environmental variable, we converted GIF-adjusted *p*-values to *q*-values and applied the FDR control method to correct for multiple testing with the *qvalue()* function of the R package QVALUE v.2.28.0 [140], and the values were deemed significant if *q* < 0.05 under at least one level of *K*.

We then performed individual-level RDA using the *rda()* function for modelling genotypes as a function of predictor variables and produced as many constrained axes as predictors [133]. We assessed the significance of RDA constrained axes using the '*ANOVA.cca*' function, and we then used the significant axes to identify candidate loci. Candidate loci were identified using a Mahalanobis distance-based approach [141], which made the RDA result comparable with those obtained with LFMM since it allowed adjusting *p*-values using the genomic inflation factor (λ) and setting FDRs to *q* = 0.05, as described above.

2.7. Functional Annotation Analysis

Given that landscape/seascape genomics analyses can suffer from the issue of false positives, we annotated the genomic neighbourhood of each SNP associated with a particular environmental variable(s), i.e., environment-associated SNP, and determined whether the molecular functions of the genes surrounding an environment-associated locus were consistent with an assumed adaptive role. For every environment-associated SNP, we performed the annotation procedure as follows. First, we annotated environment-associated SNP with the NCBI General Feature Format (GFF) annotation file of the reference genome of the lumpfish using SNPEFF v.5.1 [142], and we retrieved all the predicted genes falling within the ±50 kilobases (kb) window. This maximum window size was chosen because genes associated with a mutation can be found hundreds of kbs away [143–145]. Second, we retrieved the predicted protein sequences related to these genes and ran a similarity search with BLASTP [146] against metazoan protein sequences in the UNIPROTKB/SWISS-PROT database (Release 2023_01) [147]. For every predicted gene, we kept only the best significant match with an E-score threshold < 10^{-7}. Lastly, we performed pathway and process enrichment analysis for Gene Ontology terms, KEGG Pathway, Reactome Gene Sets, and WikiPathways among the predicted genes using the METASCAPE web server v.3.5.20230101 [148].

3. Results

3.1. Genomic Datasets and Signatures of Selection

Although our HiSeq4000 run yielded approximately 330 million PE reads, the total number of raw reads reported in this study is 162,423,062, and the rest of the reads correspond to those from other studies on the species in our group. Processing of raw Illumina data by the program *process_radtags* recovered a high percentage of retained reads (92.7%) with a mean of 1.2 million reads per sample. After removing samples that did not sequence well, with <500,000 retained reads, the *gstacks* program generated a catalog of 457,751 putative RAD loci with a mean insert length of 442.8 bp (±155.7 bp) and a mean effective per-sample coverage of 18× (±14.3×). After the initial SNP filtering with the *populations* program, we retained 57,840 RAD loci composed of 40,972 SNPs for the global dataset and 60,540 RAD loci composed of 40,877 SNPs for the local dataset. After

secondary filtering, thinning the step in STACKS WORKFLOW, purging closely related individuals, and removing SNPs within non-chromosomal scaffolds, we retained a final panel of 21,202 and 14,613 polymorphic SNPs for the global (genotyping rate = 0.95; n = 117) and regional (genotyping rate = 0.98; n = 88) datasets, respectively (Figure S1). We detect 5053 putative SNPs under balancing selection and 9 SNPs putatively under divergent selection with BAYESCAN, 213 putative SNPs under divergent selection with PCADAPT and 0 SNP putatively under divergent selection with OUTFLANK. We derived a putatively adaptive SNP dataset comprised of outlier loci under divergent selection. We removed SNPs putatively under selection and retained datasets of 15,953 and 10,153 putatively neutral SNPs for the global (genotyping rate = 0.95; n = 117) and regional (genotyping rate = 0.98; n = 88) sampling populations. From these datasets each, we derived a SNP subset dataset of 150 SNPs with the highest locus-specific F_{ST}. We identified 803 CNV, and after normalizing the data and removing all CNV with a variance of zero, we retained a final panel of 27 CNV for each dataset.

3.2. Genetic Diversity and Differentiation

The overall patterns of the genome-wide diversity dataset for global and regional samples were similar across sampling populations, with H_O and H_S ranging from 0.135 to 0.163 and 0.135 to 0.165, and F_{IS} ranging from −0.023 to 0.056 (Table 1). Most notably, A_P patterns among global and regional sampling population varied, ranging from 0–243, where lumpfish from the northwestern Atlantic (CRE and CAN) harboured the highest A_P of 243 and 180 alleles. Within Norway, the northernmost sampling populations NAL and NHE harboured the highest A_P of 25 and 8. The overall β_{WT} index for the global dataset was moderate (0.058; 95% CI: 0.054–0.061), where the highest observed β_{WT} within sampling populations was among lumpfish from NRG (0.10, 95% CI: 0.092–0.113) and CAN (0.188, 95% CI: 0.173–0.201). Whereas the overall β_{WT} index for the regional dataset was shallow (0.012; 95% CI: 0.009–0.015), where the highest observed β_{WT} was among lumpfish from NRG (0.065, 95% CI: 0.053–0.076) and NSA (0.015, 95% CI: 0.007–0.024).

We uncovered shallow-to-great population differentiation among sampling populations with the global dataset, with pairwise F_{ST} ranging from −0.002 to 0.182 (global F_{ST} = 0.072), where outgroup sampling populations were the most divergent followed by the southern Norway sampling population NRG (Table 2). We detected shallow population differentiation among Norwegian sampling populations with the regional dataset similar to the global dataset, with pairwise F_{ST} ranging from −0.001 to 0.036 (global F_{ST} = 0.011), where NRG and NAU were the highly differentiated sampling populations. We found a similar trend of a limited discernible geographic pattern with the CNV dataset (Table 2). We uncovered similar patterns of genetic diversity and population differentiation with the subset dataset of 150 SNPs with the highest locus-specific F_{ST} and 'adaptive' SNP dataset (Figures S2 and S3).

Table 2. Pairwise F_{ST} (SNPs) and V_{ST} (CNV) and their confidence interval in brackets among sampling populations of lumpfish estimated using the full neutral SNP and CNV datasets. F_{ST}-values are below the diagonal and V_{ST}-values are above the diagonal.

	NMA	NRG	NAU	NRA	NAV	NEK	NNA	NSA	NHE	NAL	GRE	CAN
NMA		0.111 (0.067–0.156)	0.345 (0.232–0.458)	0.134 (0.057–0.212)	0.089 (0.037–0.141)	0.182 (0.11–0.254)	0.159 (0.065–0.252)	0.219 (0.123–0.314)	0.095 (0.056–0.134)	0.046 (0.019–0.072)	0.147 (0.082–0.213)	0.317 (0.213–0.421)
NRG	0.019 (0.017–0.021)		0.127 (0.065–0.188)	0.19 (0.098–0.281)	0.139 (0.074–0.203)	0.243 (0.148–0.338)	0.301 (0.193–0.408)	0.116 (0.065–0.168)	0.078 (0.042–0.115)	0.057 (0.033–0.08)	0.219 (0.13–0.309)	0.163 (0.099–0.227)
NAU	0.009 (0.006–0.012)	0.032 (0.028–0.035)		0 (0–0)	0.2 (0.105–0.295)	0.271 (0.18–0.363)	0.169 (0.062–0.277)	0.272 (0.168–0.377)	0.127 (0.065–0.189)	0.054 (0.02–0.088)	0.103 (0.039–0.166)	0.276 (0.179–0.374)
NRA	0.004 (0.002–0.007)	0.032 (0.028–0.035)	0.01 (0.006–0.015)		0.174 (0.107–0.242)	0.248 (0.148–0.349)	0.325 (0.212–0.437)	0.181 (0.087–0.275)	0.238 (0.151–0.325)	0.077 (0.035–0.119)	0.121 (0.048–0.195)	0.287 (0.189–0.385)
NAV	0.007 (0.005–0.009)	0.029 (0.026–0.031)	0.015 (0.012–0.018)	0.006 (0.003–0.009)		0.185 (0.084–0.286)	0.198 (0.125–0.271)	0.154 (0.068–0.24)	0.207 (0.132–0.283)	0.115 (0.058–0.172)	0.138 (0.061–0.216)	0.226 (0.137–0.315)
NEK	0.003 (0–0.005)	0.03 (0.027–0.034)	0.008 (0.004–0.013)	0.005 (0.001–0.009)	0.001 (−0.002–0.005)		0.191 (0.083–0.3)	0.258 (0.144–0.372)	0.123 (0.07–0.176)	0.1 (0.039–0.161)	0.126 (0.054–0.198)	0.313 (0.198–0.428)
NNA	0.004 (0.003–0.006)	0.027 (0.025–0.03)	0.011 (0.009–0.014)	0.002 (−0.001–0.005)	0.006 (0.004–0.008)	−0.002 (−0.004–0.001)		0.076 (0.037–0.114)	0.087 (0.051–0.123)	0.088 (0.05–0.126)	0.087 (0.046–0.128)	0.308 (0.208–0.409)
NSA	0.006 (0.004–0.008)	0.03 (0.027–0.033)	0.016 (0.013–0.019)	0.008 (0.005–0.011)	0.004 (0.002–0.005)	0.005 (0.002–0.008)	0.005 (0.003–0.006)		0.27 (0.162–0.378)	0.094 (0.049–0.138)	0.057 (0.013–0.101)	0.331 (0.22–0.442)
NHE	0.005 (0.004–0.007)	0.026 (0.024–0.029)	0.014 (0.011–0.017)	0.003 (0.001–0.006)	0.005 (0.003–0.007)	0.001 (−0.002–0.003)	0.002 (0–0.003)	0.003 (0.001–0.004)		0.098 (0.064–0.131)	0.113 (0.062–0.163)	0.158 (0.094–0.222)
NAL	0.008 (0.007–0.01)	0.031 (0.028–0.033)	0.015 (0.012–0.018)	0.003 (0.001–0.006)	0.01 (0.008–0.012)	0.002 (−0.001–0.005)	0.007 (0.005–0.008)	0.008 (0.007–0.01)	0.006 (0.005–0.008)		0.061 (0.036–0.086)	0.179 (0.12–0.238)
GRE	0.118 (0.114–0.122)	0.141 (0.137–0.146)	0.134 (0.129–0.138)	0.132 (0.126–0.137)	0.127 (0.122–0.131)	0.124 (0.118–0.129)	0.122 (0.118–0.126)	0.126 (0.122–0.131)	0.123 (0.119–0.127)	0.118 (0.114–0.122)		0.048 (0.023–0.074)
CAN	0.153 (0.148–0.159)	0.182 (0.176–0.187)	0.17 (0.164–0.176)	0.176 (0.169–0.182)	0.165 (0.159–0.171)	0.165 (0.158–0.171)	0.156 (0.151–0.162)	0.163 (0.158–0.168)	0.157 (0.152–0.163)	0.152 (0.147–0.157)	0.068 (0.065–0.071)	

3.3. Population Genetic Structure and Connectivity

We explored population structure using several independent methods, and we did detect genetic structuring with the global, regional, and subset SNP datasets. First, globally, the PCoA revealed geographic clustering of lumpfish from the three different regions (Canada, Greenland, and Norway), where samples from respective regions formed individual clusters with a hierarchical pattern of genetic structuring within Norway. Overall PCoA clustering patterns revealed at least three main genetic clusters based on the first three principal components (PCs), which together accounted for 26% of the genetic variation in the global datasets (Figure 2A). A similar clustering pattern was recovered with the regional 150-SNP dataset (Figure 2B). Whereas at least two main genetic clusters were evident in Norway with the regional dataset, the first three PCs together accounted for 9% of the genetic variation (Figure 3A). Although, the regional 150-SNP dataset recovered finer scale patterns of genetic clustering with the first three PCs accounting for 52.9% of the genetic variation (Figure 3B). Second, the DAPC analysis also uncovered similar geographical clustering patterns of lumpfish as a priori, and when excluding location prior to where three and two genetic clusters were identified based on the BIC score. The genetic clusters corresponded with geography at the global scale but not at the regional scale, where Norway constituted two sub-clusters. Third, the post-processing of the *snmf* results based on the lowest average CV score across replicates identified $K = 3$ and $K = 1$ as the most likely number of genetic clusters for the global and regional levels, respectively, which was consistent with the number of clusters based on the first two principal components and DAPC, respectively (Figure 4A,B). Notably, the genetic composition of NRG was largely consistent among the genetic differentiation estimates, where it was the most divergent sampling population of lumpfish in Norway (Figure 4A,B). Finally, the FINERADSTRUCTURE co-ancestry between genotype pairs revealed geographic clustering of genotypes by region (Canada, Greenland, and Norway) and the northwestern Atlantic sampling populations shared more co-ancestry with each other than between Norwegian sampling populations in the northeastern Atlantic (Figure 5). Genotypes in Norway were weakly clustered, and the separation into two subgroups was marginally supported, where southern Norway individuals from NMA, NRA, and NAU largely constituted one group (Figure 5). Moreover, we uncovered that there was a finer scale of genetic structuring observed from lumpfish collected from the northernmost Altafjord.
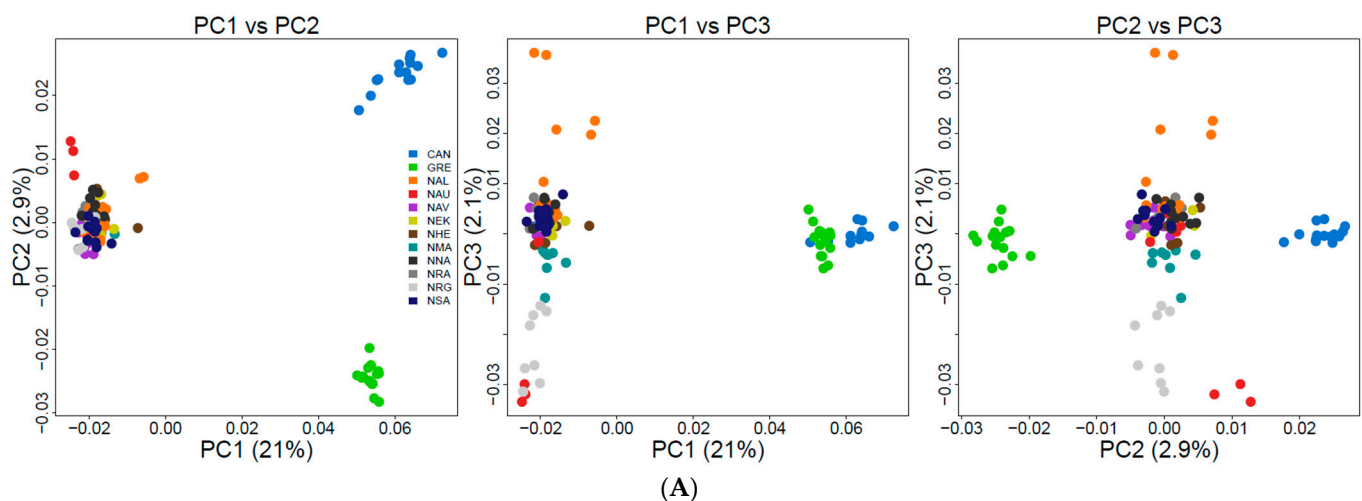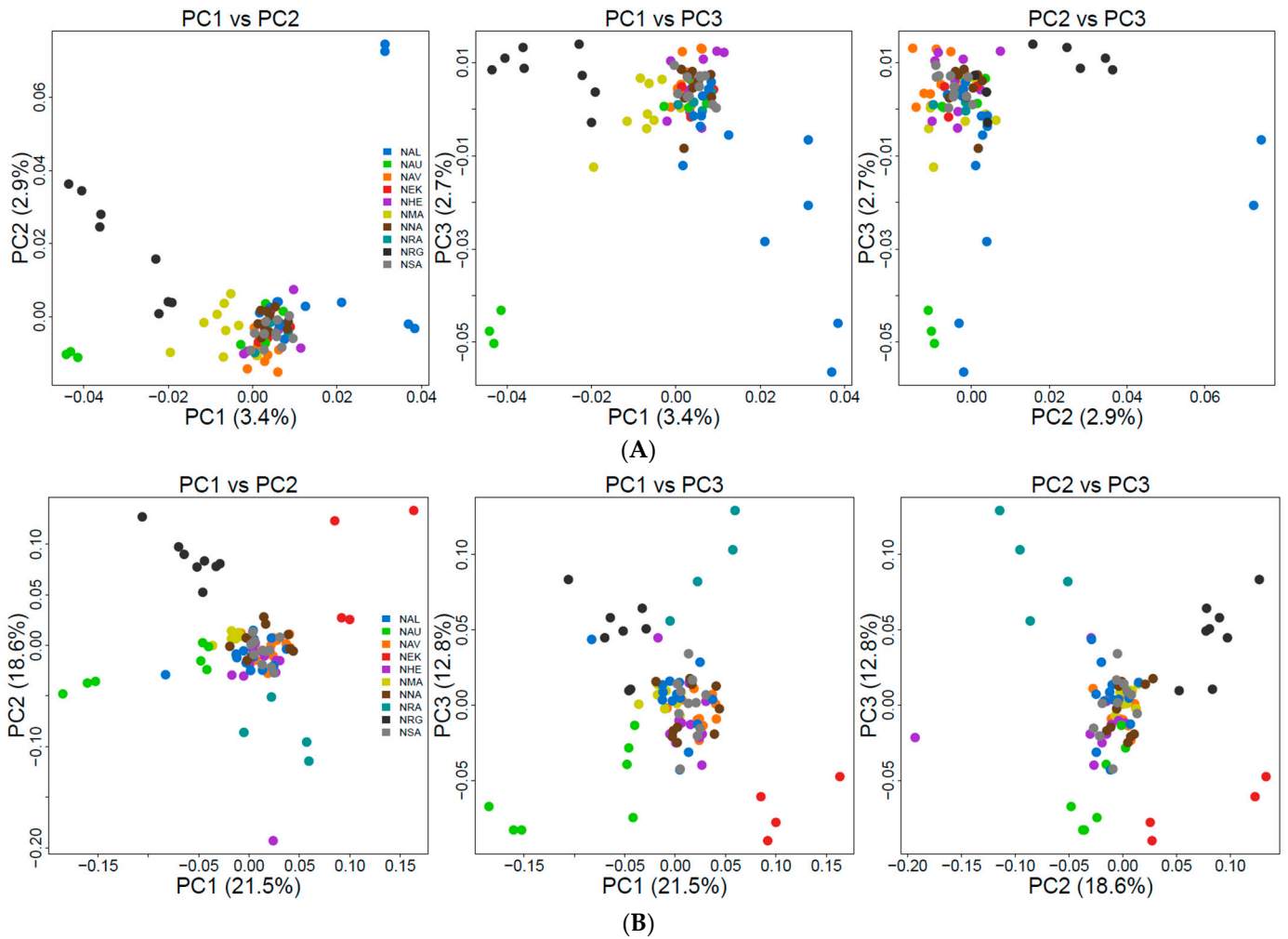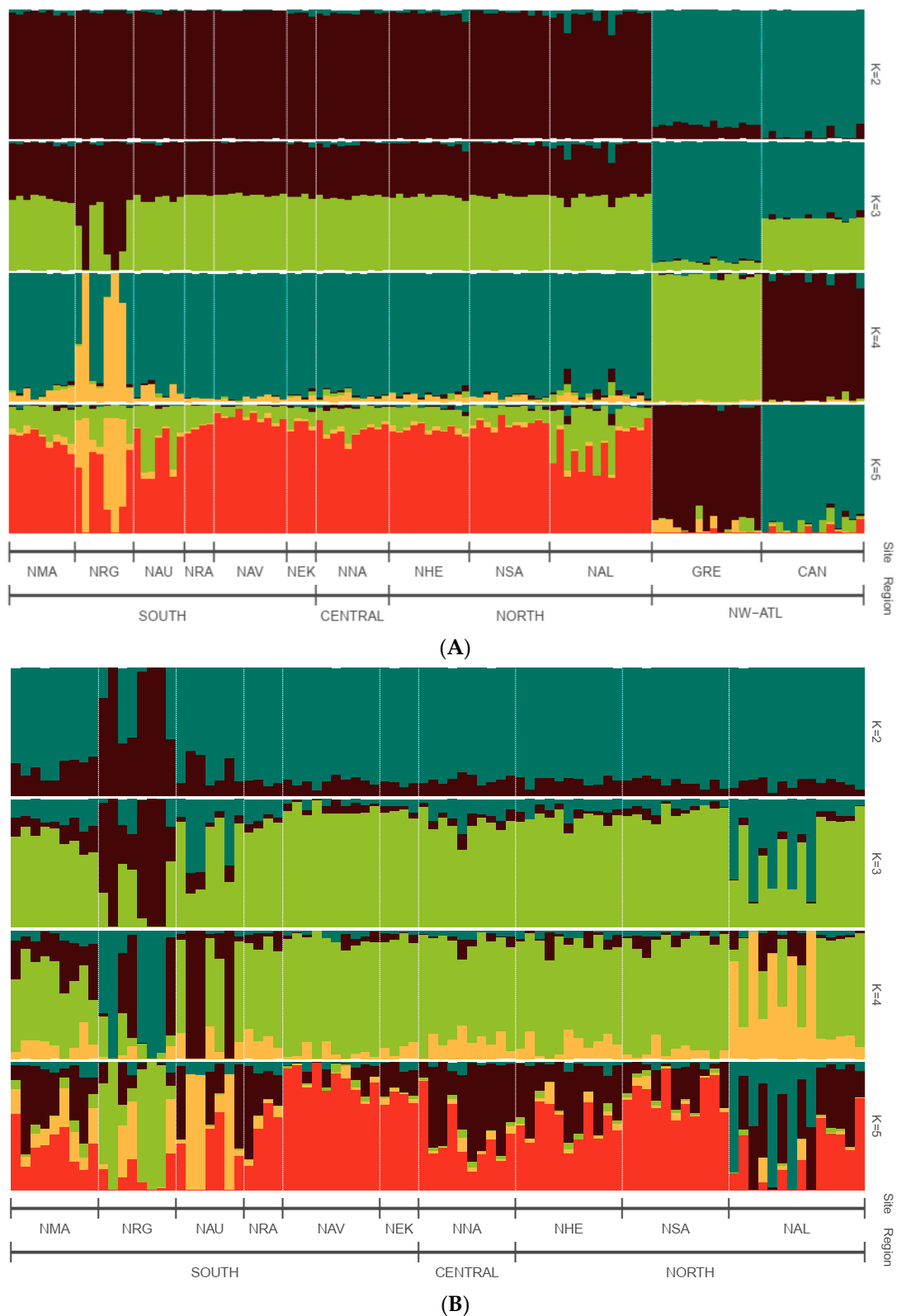

Figure 2. *Cont.*
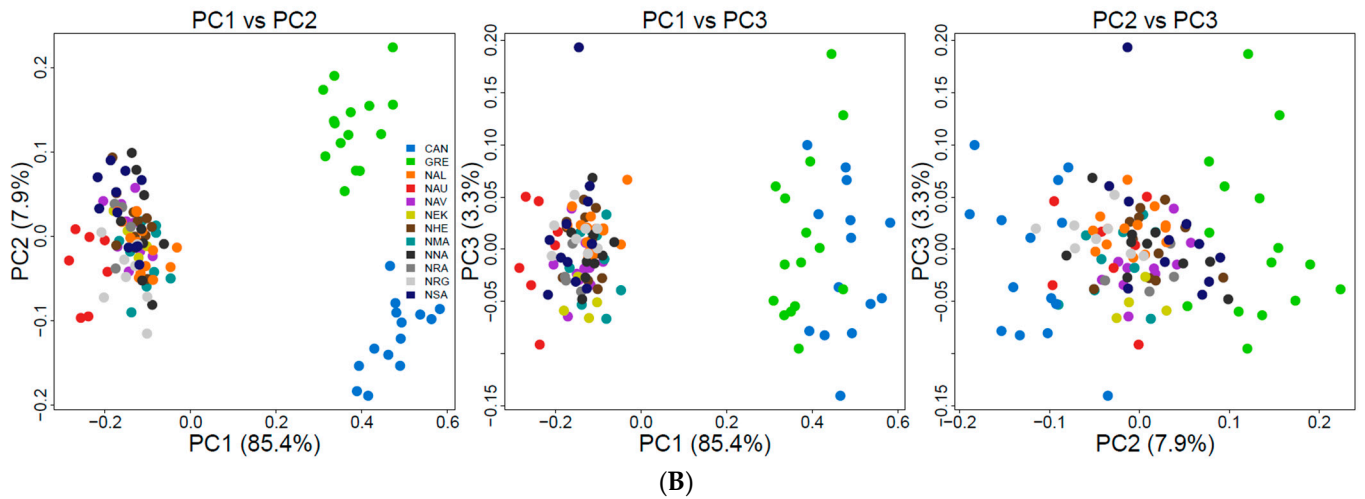
Figure 2. Principal coordinate analysis (PCoA) based on the full (**A**) and 150 genome-wide SNPs (**B**) for the global sampling scheme across the Pan Atlantic.


Figure 3. Principal coordinate analysis (PCoA) based on the full neutral (**A**) and 150 genome-wide SNPs (**B**) for the regional sampling scheme in Norway.

Figure 4. Sparse non-negative matrix factorisation (snmf) clustering analyses for a global (**A**) and regional (**B**) picture of genetic cluster patterns of lumpfish across the north Atlantic for 10 independent runs of K = 2–5 based on the full neutral SNP dataset. NW-ATL, North-western Atlantic Ocean. A single vertical line represents each individual; a black line separates sample sites; the whole sample is divided into *K* colours representing the number of clusters inferred. The colours show the estimated individual proportions of the cluster membership, where each colour represents a genetic cluster.
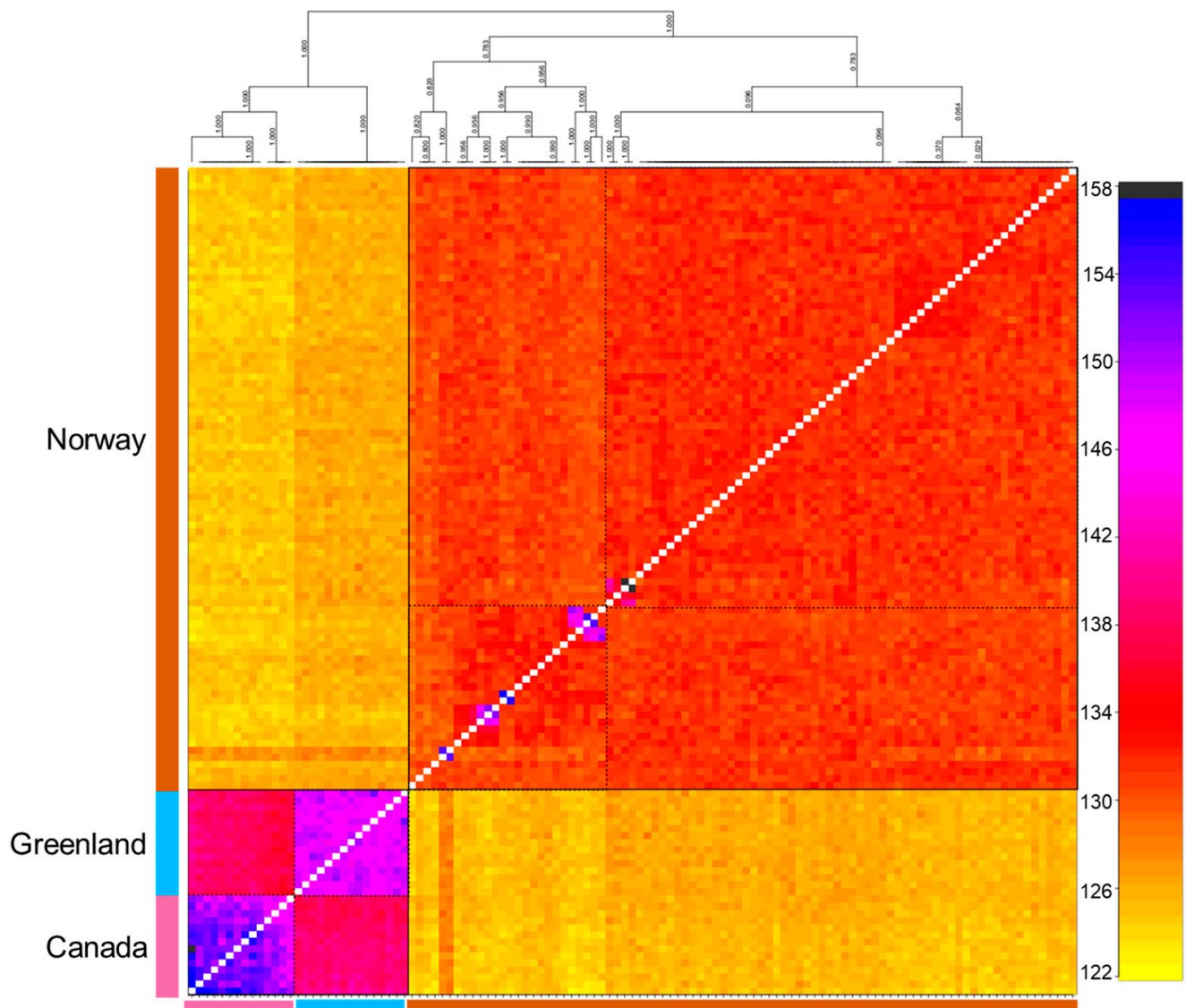
Figure 5. FINERADSTRUCTURE pairwise co-ancestry matrix and tree based on the global full neutral SNP datasets. The central matric panel shows co-ancestry between genotypes, with light yellow indicating low co-ancestry and darker yellows, oranges, and reds to black indicating progressively higher co-ancestry. Solid black squares indicate the major clades corresponding to ocean basin. Dashed squares indicate the sub-clades.

In contrast to the SNP datasets, the global and regional CNV data revealed three and two clusters, respectively, which do not correspond to the geographic proximity among the sampling population (Figures 6 and 7). The global dataset revealed that GRE, NNA, and NRA were differentiated from the rest of the study sampling populations (Figure 6A–C). The regional dataset corroborated the global dataset with regard to the differentiation of NNA and NRG (Figure 7A–C).
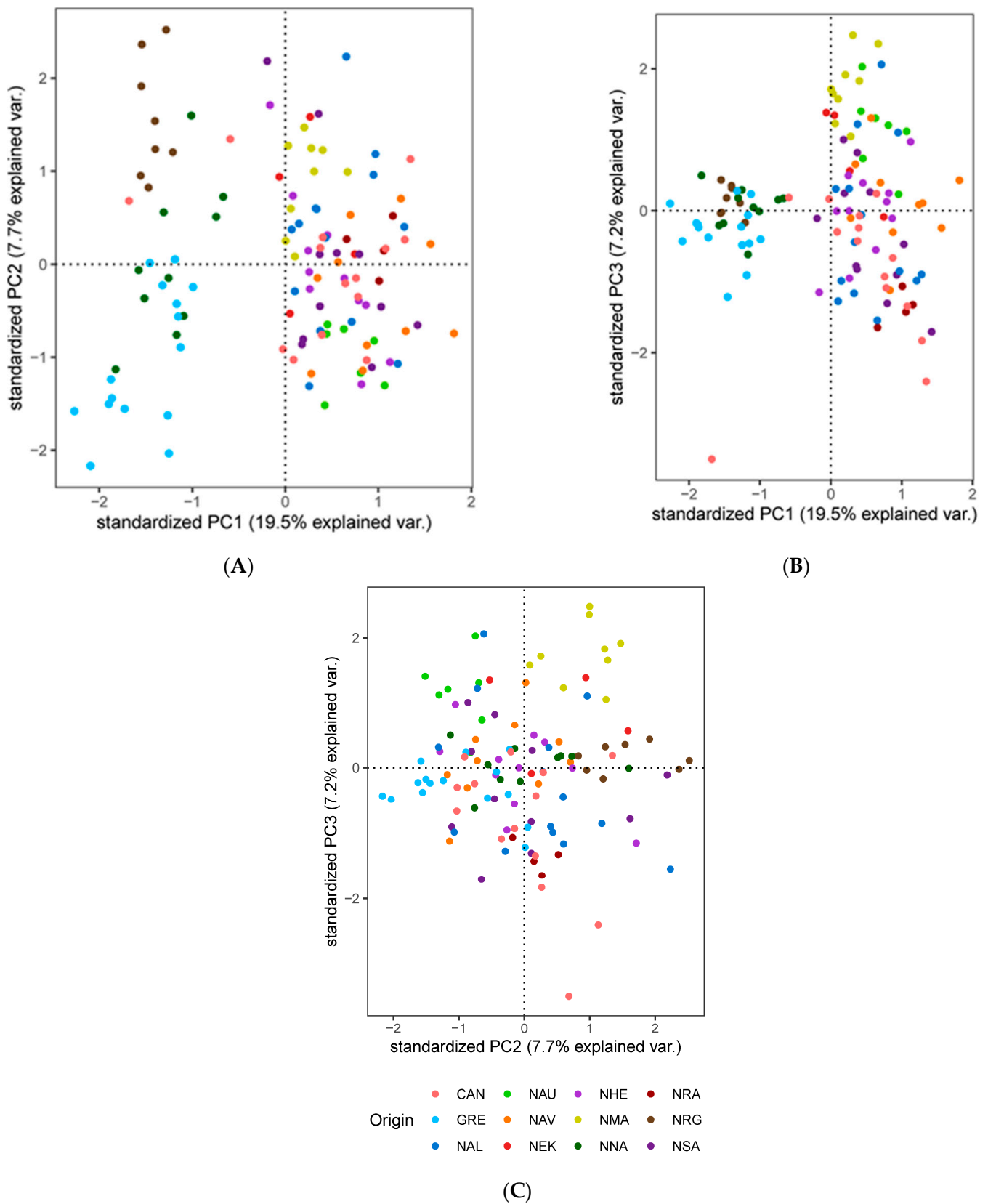
Figure 6. Principal component analysis (PCA) based on the 27 genome-wide CNVs based on PC axes 1 and 2 (**A**), PC axes 1 and 3 (**B**), and PC axes 1 and 3 (**C**) for the global sampling scheme in the North Atlantic.
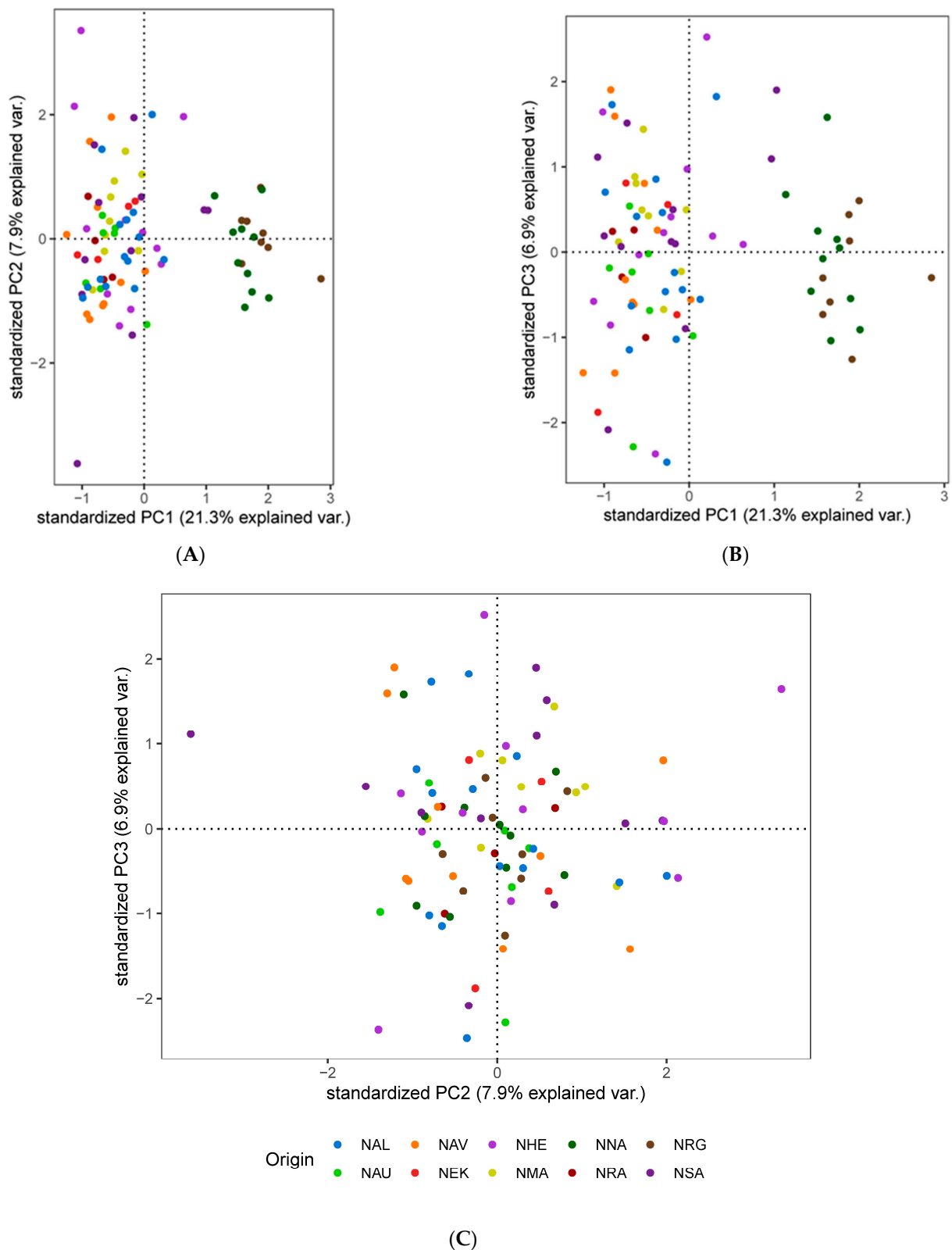
Figure 7. Principal component analysis (PCA) based on the 27 genome-wide CNVs based on PC axes 1 and 2 (**A**), PC axes 1 and 3 (**B**), and PC axes 1 and 3 (**C**) for the global sampling scheme in the North Atlantic for the regional sampling scheme in Norway.

3.4. Spatial Structure and Environmental Associations

We found a strong, significant correlation between pairwise F_{ST} and Euclidean geographic distance (Mantel $r = 0.947$, $p = 0.0041$) and in-water geographic distance (Mantel $r = 0.952$, $p = 0.0029$) with the Mantel test, indicating a significant effect of IBD in the global dataset. However, within Norway, we detected no significant effect of IBD with the regional dataset (Euclidean: Mantel $r = 0.110$, $p = 0.2767$; in-water: Mantel $r = 0.166$, $p = 0.217$). We retained the first four axes from the PCA of Hellinger-transformed allele frequencies, which cumulatively explained 56.8% of the total neutral genetic variation for the Norwegian sampling populations. Forward selection of the Euclidean dbMEM variables identified two significant predictors (dbMEM6 and dbMEM8) representing regional-scale spatial structure. RDAs with only the selected dbMEM variables were significant ($p = 0.013$) with a coefficient of determination (adjusted R^2) of 22.3%. The first two RDA axes were marginally significant and explained 20.8% and 18.8% of the genetic variation summarised in the PC axes, respectively (Figure 8).
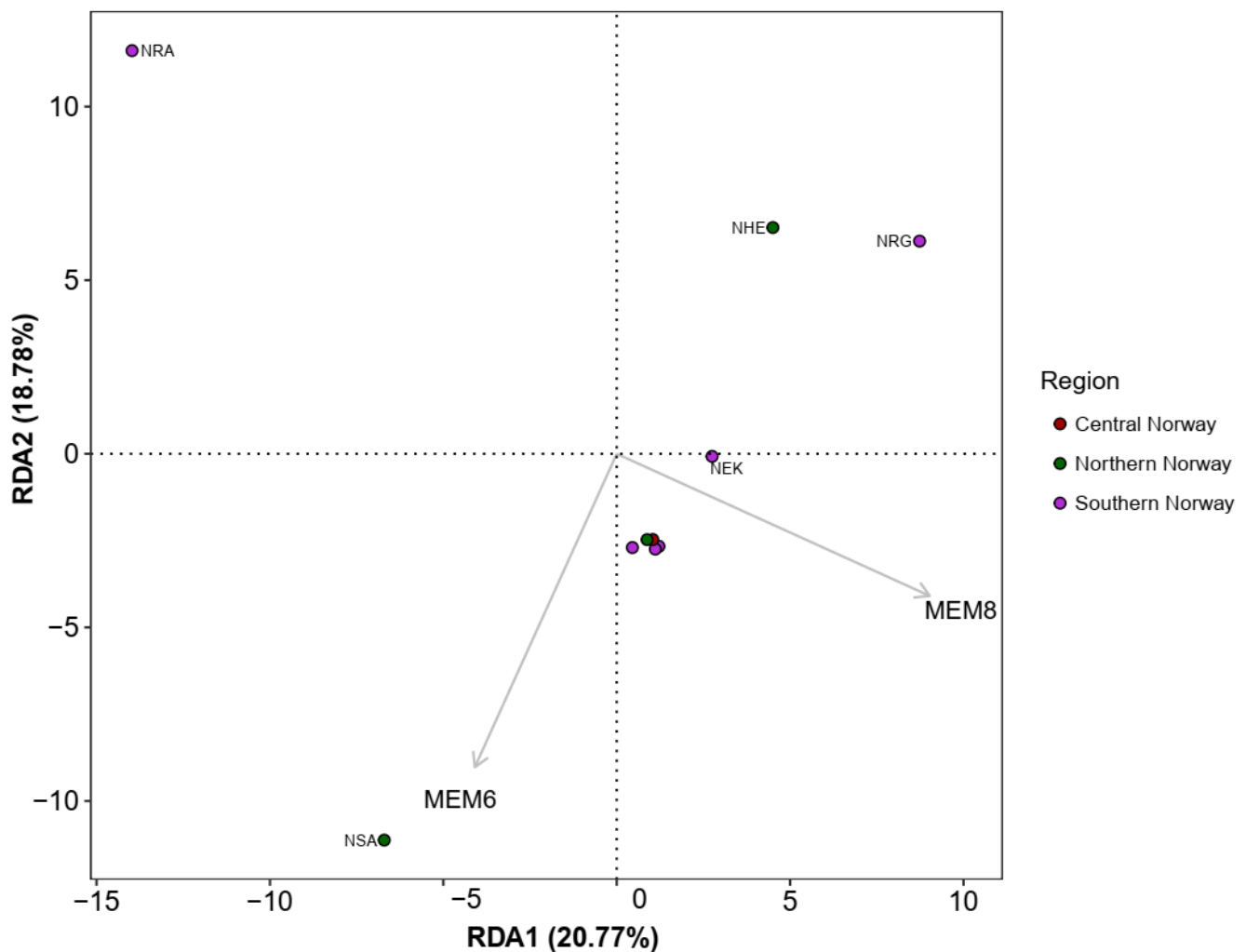
Figure 8. RDA biplot showing the significant Moran Eigenvector Maps (MEMs) predictor variables (arrows) on RDA axes 1 and 2.

For seascape genomics analysis, we found that our sampling regime constituted of at least four environmental clusters based on principal component and hierarchical clustering analyses (Figure 1). Our sampling scheme harboured adequate environmental gradients for both global and regional seascape genomic analysis. We identified seven independent environmental variables based on an analysis of the correlation among environmental variables

for the global sampling populations, whereas four independent variables were identified for the regional sampling populations. For the global dataset, we detected 542 significant genotype–environment associations (with 5 duplicate detections; 537 unique GEAs) based on RDA analysis without accounting for spatial structure using partially constrained ordination (Figure 9A,B). Environment-associated SNPs were related to CHLm_OM (151 SNPs), CHLm_OSD (53 SNPs), FE_OSD (48 SNPs), CHLa_OM (82 SNPs), MEM4 (107 SNPs), O2_OSD (45 SNPs), and SCV_OM (51 SNPs). While 382 unique environment-associated SNPs were detected when accounting for spatial structure with pRDA (Figure 10A,B): CHLm_OM (131 SNPs), CHLm_OSD (69 SNPs), FE_OSD (15 SNPs), CHLa_OM (64 SNPs), MEM4 (12 SNPs), O2_OSD (50 SNPs), and SCV_OM (41 SNPs). Using LFMM, we identified 33 environment-associated SNPs with the global dataset. Regionally along the Norwegian coastline, we detected 70 significant genotype–environment associations with RDA (CHLm_OSD: 1 SNP; FE_OM: 42 SNPs; MEM6: 22 SNPs; MEM8: 5 SNPs), 48 with pRDA (CHLm_OSD: 13 SNPs; FE_OM: 33 SNPs; MEM6: 2 SNPs; MEM8: 0 SNPs), and 5 with LFMM. We did not detect any environment-associated CNVs with either the global or the regional dataset, but uncovered a similar pattern of genetic structure revealed by PCA analysis (Figure 11A,B).
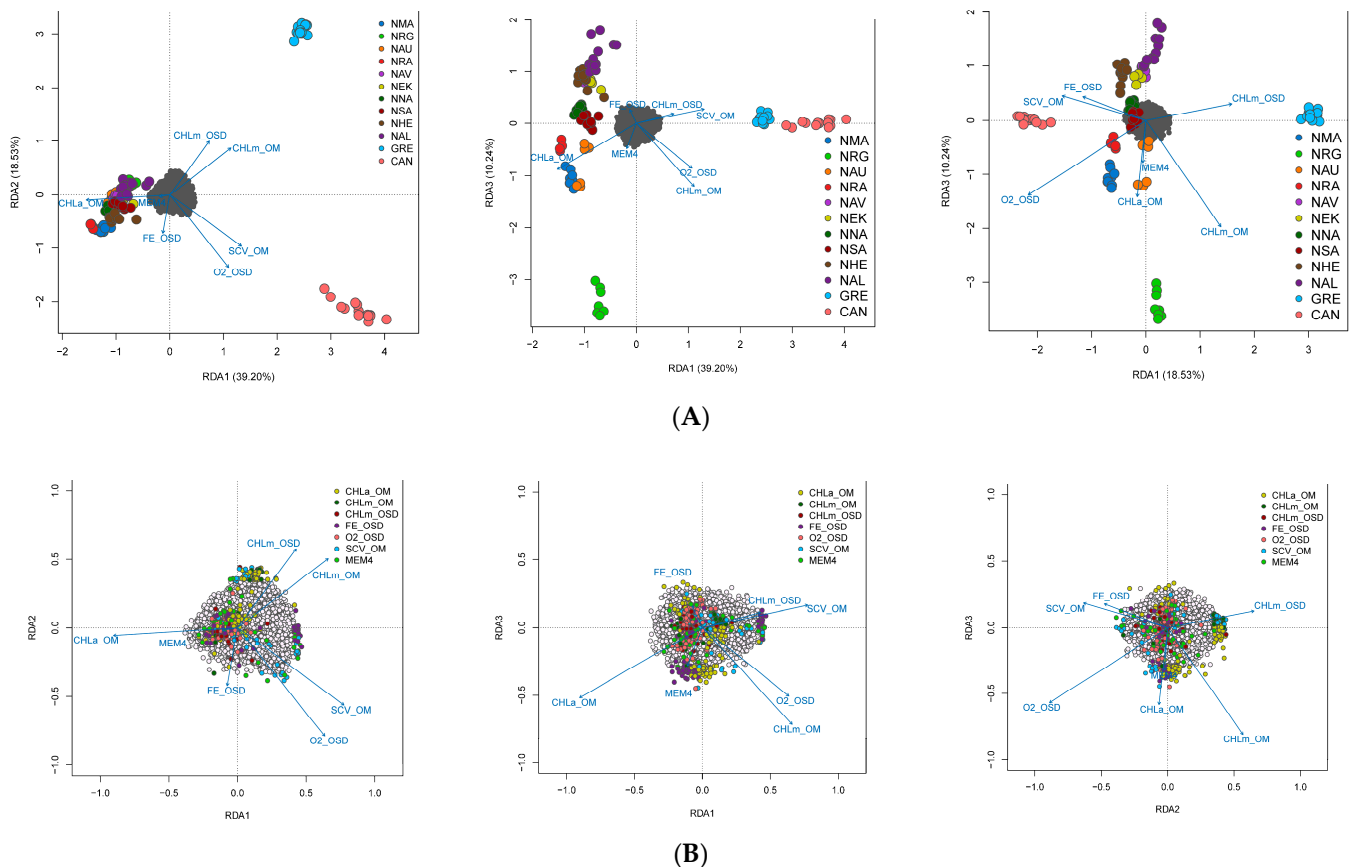
Figure 9. RDA biplot of sampling sites (**A**) and SNPs (**B**) based on RDA axes 1 and 2, RDA axes 1 and 3, and RDA axes 1 and 3 for the global sampling scheme in the North Atlantic.
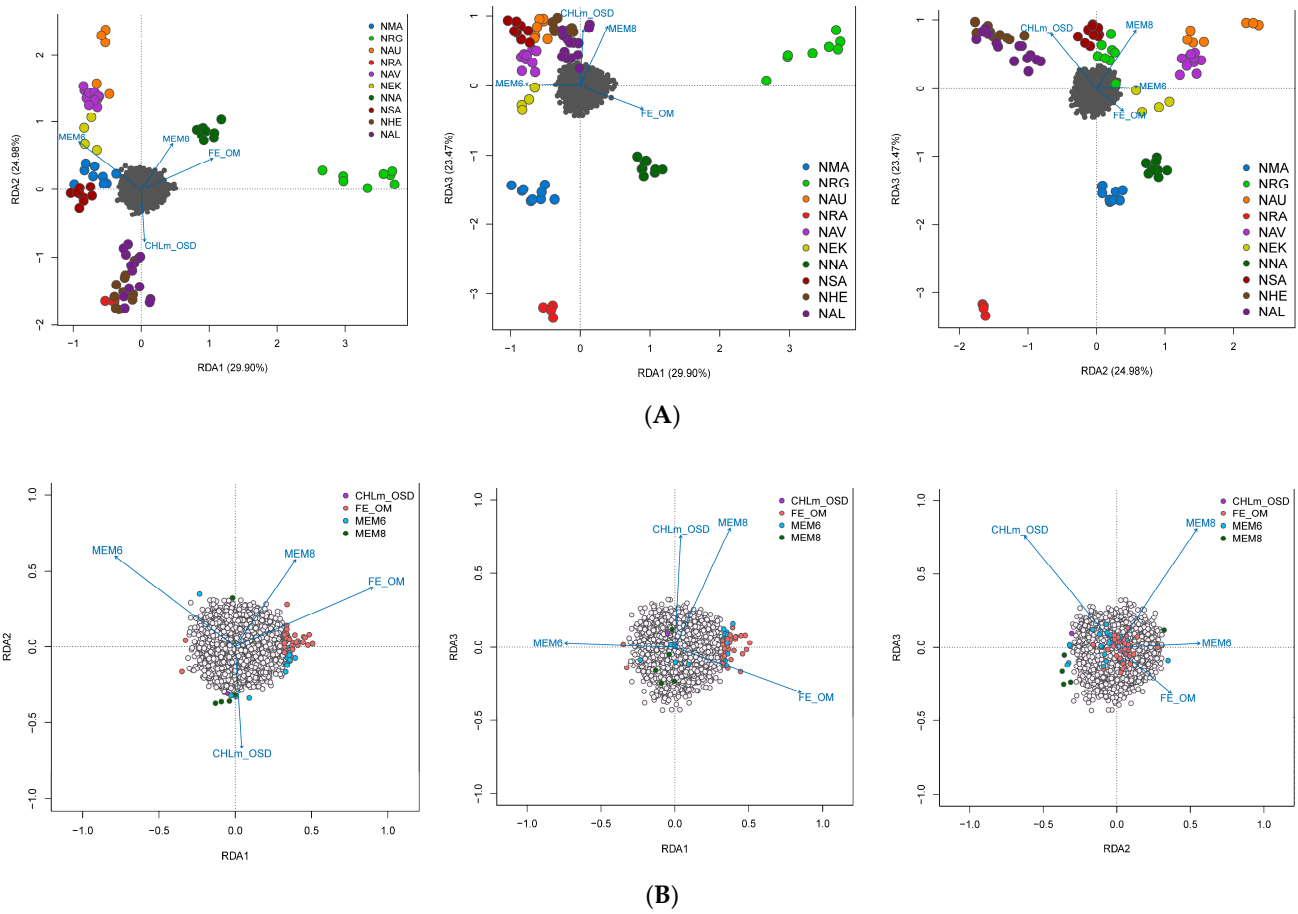
Figure 10. RDA biplot of sampling sites (**A**) and SNPs (**B**) based on RDA axes 1 and 2, RDA axes 1 and 3, and RDA axes 1 and 3 for the regional sampling scheme in Norway.

Figure 11. *Cont.*

(**B**)
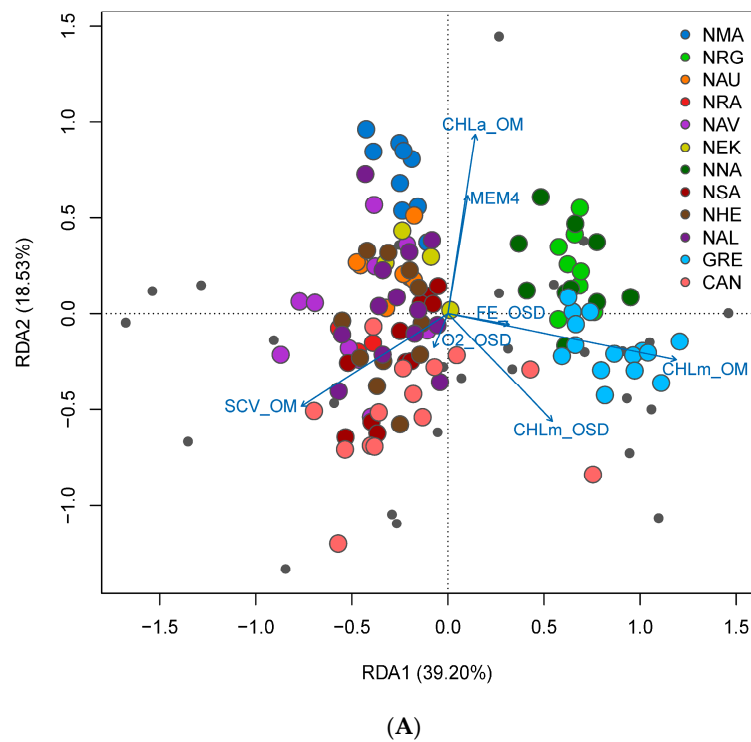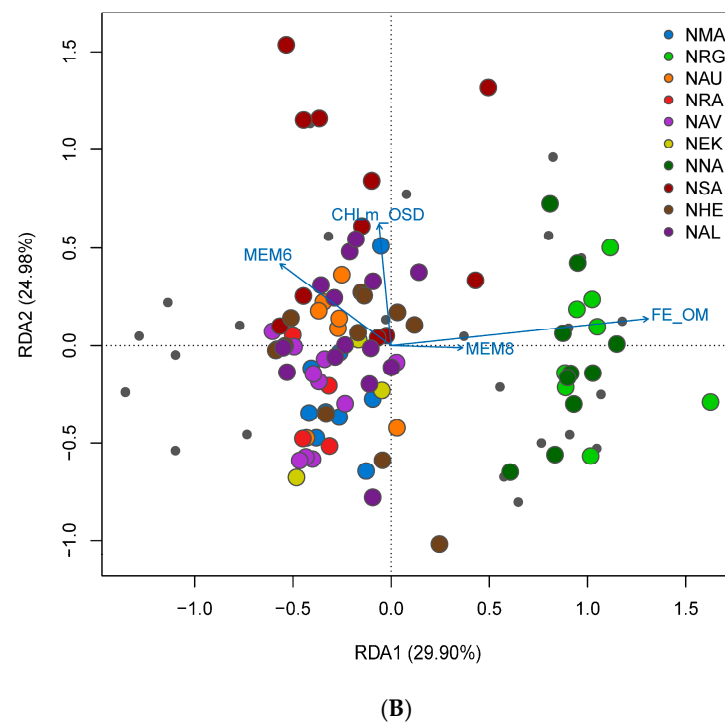
Figure 11. RDA biplot of sampling sites based on RDA axes 1 and 2 for the global (**A**) and regional (**B**) CNV datasets.

3.5. Functional Annotation

Across GEA methodologies, we detected 654 and 106 unique environment-associated RAD loci with the global and regional datasets, respectively, with an overall of 35 loci (Table S2). The functional annotations surrounding the global panel of environment-associated SNPs revealed that the top-level Gene Ontology (GO) categories were developmental process, metabolic process, localisation, cellular process, homeostatic process, response to stimulus, biological regulation, positive regulation of biological process, multicellular organismal process, and negative regulation of biological process (Figures S4A and S5A,B). The genes were involved in four Reactome Gene Sets (phase 0—rapid depolarisation, EPHA-mediated growth cone collapse, gap junction trafficking and regulation, and SUMO transferred from E1 to E2 (UBE2I, UBC9)), whereas the regional panel of environment-associated SNPs were involved in the cellular process, developmental process, positive regulation of the biological process, response to stimulus, growth, biological regulation, metabolic process, and locomotion (Figures S4B and S5C,D). The genes were involved in two KEGG pathways (MAPK signalling pathway and RIG-I-like receptor signalling pathway) and four Reactome Gene Sets (nucleotide-like (purinergic) receptors, lysosome vesicle biogenesis, postmitotic nuclear pore complex reformation, and signalling by PDGF)). The 35 overlapping loci were involved in three GO biological processes (cellular process, developmental process, and response to stimulus) and three KEGG pathways (Glycerophospholipid metabolism, WNT signalling, and Hedgehog signalling pathways) (Figures S4C and S6A–C).

4. Discussion

It is widely accepted that local adaptation to the abiotic environment plays a major role in both intraspecific and interspecific diversification; however, its contribution relative to other evolutionary forces, such as gene flow is rarely quantified [44,52,54]. In this study, we investigated whether local adaptation occurs despite high gene flow in lumpfish, and if it does, we identified the principal environmental drivers of local adaptation at both larger and finer scales. We found associations between genetic and climatic structure in this

system and quantified the contributions of individual environmental variables to patterns of genomic variation and dissected the genetic targets of these putative selective gradients in a multivariate statistical framework at a finer scale. Furthermore, population genomic analyses using genome-wide SNPs revealed that genetic homogeneity was prevalent in Norwegian sampling populations, correlating with previous findings using microsatellite markers [28,32]. This study is one of the largest population and seascape genomics efforts performed to date on the Norwegian lumpfish and is predicted to facilitates genome-enabled monitoring of the genetic impacts of escapees and allow for genetic-informed broodstock selection and management.

4.1. Genome Scans for Outliers

Genome scans for the selection indicated both signatures of divergent and balancing selection acting on the lumpfish genome. The putative genomic regions under divergent selection that we uncovered revealed that sampled populations from Norway were not too adaptively divergent. We found that the balancing selection was the predominant force operating on the lumpfish genome, indicating that multiple alleles were actively maintained at frequencies higher than expected from genetic drift alone [42,149]. The balancing selection, such as the negative frequency-dependent selection, plays a crucial role in maintaining allelic and phenotypic polymorphisms within populations, including behavioural traits, such as foraging [150], personality [151,152], behaviourally mediated life history strategies [153], and migration [154,155]. Our results represent the first genetic evidence for the balancing selection in lumpfish that could be linked to various commercially important traits, including sea lice foraging habits. Although, there have been issues surrounding the sustainability and delousing efficacy of wild-caught cleaner fish [20,156], mounting evidence suggests that delousing-related traits, such as small body size (40–140 g), personality, and grazing efficacy, are likely parentally controlled and can be optimised under selective and targeted breeding programs [26,157–159]. Accordingly, our results are foundational towards understanding whether selection operates differentially within populations, between individuals, or not for traits of interest, such as sea lice foraging behaviour.

4.2. Understanding the Population Structure of Lumpfish

Delineating the patterns and levels of population structure at various scales in lumpfish has major implications for conservation, fisheries, and aquaculture. With our 'global' sampling regime, we found two major genetic groups of lumpfish, i.e., the western (Canada and Greenland) and eastern Atlantic (Norway), congruent with previous findings [29–33]. Our results revealed that lumpfish from the western Atlantic (outgroup) group shared higher co-ancestry with each other than between populations from the eastern Atlantic group. Within the eastern Atlantic group, conflicting findings have been presented based on genomic STRs (g-STRs) [28,31], expressed sequence tag (EST) STRs (EST-STRs) [32], and genome-wide SNPs [33] on whether there is genetic structuring among Norwegian lumpfish. Our results revealed that Norwegian lumpfish were weakly clustered into two marginally supported subgroups, where individuals from southern Norway largely constituted one group. With both SNP and CNV, we found no evidence of population structure with sampling populations from Jónsdóttir et al. [28,32], suggesting that the lack of genetic structure in the previous studies was not preordained by the marker of choice, i.e., g-STRs [28] or EST-STRs [32]. In fact, a similar trend of genetic homogeneity was observed in Jansson et al. [33] between representative sampling populations from Norway (Flatanger and Flekkefjord) with genome-wide and 'diagnostic' SNPs datasets, where weak population differentiation and structure between populations was uncovered regardless of the SNP panel used and the large geographical distance between the sampling locations. Overall, our results of population genetic structure of the Norwegian lumpfish agree with the findings of the existence of one [28,32] and two [31,33] genetic groups depending on the finer scale coverage of the Norwegian coastline. Nonetheless, it is

clear from these studies that population genetic structure in lumpfish is influenced by the interactions between evolutionary forces (gene flow and genetic drift), reproductive behavioural traits (philopatry), and seascape features (hydrodynamic, thermal, or eco-physiological boundaries).

4.3. Local Adaptation under High Gene Flow in Lumpfish

The intensity of divergent natural (or Darwinian) selection frequently fluctuates in heterogeneous environments across ecological clines (i.e., selective gradients), resulting in genotype–environment interactions for Darwinian fitness [160,161]. Marine species are largely characterised by the lack of apparent physical barriers to gene flow across vast geographic distances [118,162,163]. However, dispersal potential may differ over a fragmented seascape depending on patterns and gradients of environmental variables, such as ocean currents, temperature, pH, and salinity, resulting in varying levels of population connectivity [116,118,163]. Consequently, marine species provide an ideal system for understanding how diversifying selection operates on maintaining divergence between populations at adaptive loci, whilst allowing homogenisation in other parts of the genome, i.e., local adaptation with gene flow. Our results provided novel insights into the existence of local adaptation under high gene flow in the Norwegian lumpfish governed by selective gradients along the coastline, with temperature and salinity correlated variables and sea current velocity as the main environmental drivers of local adaption in lumpfish. Within Norway, we found that the northern and southern populations are differentially adapted to local conditions. These findings add to the growing body of knowledge that oceanographic currents aiding larval transport are among the major drivers of both genetic homogeneity and population structure in species with a passive dispersal phase depending on the oceanographic regime and spatial scale [28,33,117–119,164]. Moreover, our results were consistent with those of previous research demonstrating that gene flow can promote local adaptation and that adaptive polymorphisms can be conserved within populations subject to high gene flow, contrary to the traditional notion of the 'swamping' and 'meltdown' effects of gene flow [43–45,50–53]. Given the predominant genome-wide homogeneity in Norwegian lumpfish, our results uncovered that some parts of the lumpfish genome are targets of divergent selection while other parts are under the influence of high gene flow.

4.4. The Genomic Profile of Environmentally-Associated RAD Loci

Functional annotation analyses of the environment-associated RAD loci revealed that the gene sets linked to these loci play a key role in biological processes associated with environmental pressures. The retrieved top GO categories (e.g., cellular process, developmental process, metabolism, biological regulation, and response to stimulus) have been reported in other studies that measure transcriptomic responses under experimentally induced heat stress [165,166], ocean acidification [167,168], biomineralisation, energy metabolism, and heat, disease, or hypoxia tolerance [40,169]. Moreover, the gene sets reported in the present study were enrichment in four functional pathways, which included the WNT-, MAPK-, RIG-I-like receptor and Hedgehog signalling pathways and Glycerophospholipid metabolism. The WNT and Hedgehog signalling pathways play a crucial role in embryonic development [170–172], suggesting that divergent natural selection largely operates during larval development. In addition, the MAPK pathway is essential for the response to signals or stresses from changing environmental conditions and various other stimuli [173]. Whereas the RIG-I-like receptor pathway is involved in the regulation of the innate immune response against viruses [174]. Overall, these findings broaden our understanding of the genetic basis of local adaptation in natural populations and add to the growing database of functional annotated environment-associated adaptive loci.

4.5. Conclusions, Conservation, and Aquaculture Implications

In the upheaval of environmental change and commercial exploitation of the lumpfish, applied genomics is paramount for efficient conservation and sustainable species

management. This study conducted analyses of population structure on all SNPs but also independently on putatively neutral and outlier SNPs to aid in the inference of neutral versus adaptive processes underlying the observed genetic patterns. Our study provides important clues on the definition and delineation of stocks in Norway and their potential to respond to environmental variability. Additionally, our analyses of population structure support the subdivision of lumpfish into at least two stocks in Norway, where one of the southernmost sampling populations, NRG, was the most divergent sampling population. Our regional sampling regime allowed us to resolve the contradictory evidence relating to genetic homogeneity versus structuring among Norwegian lumpfish and identify the likely causes for the conflicting results of the spatial scale of sampling design given the complex bathymetry, topography, and oceanography of the coastline. Moreover, the patterns of local adaptation in lumpfish uncovered in our study provide further valuable information that can help identify the best potential source populations for broodstock development and provenance sourcing and to identify escapees from farms.

Future sampling design should focus on embayment and/or fjord locations shielded from ocean currents to account for habitat patchiness with increased likelihood to capture genetic disjunction at smaller spatial scales. This is crucial in order to accurately investigate the number of genetic clusters and to avoid misattributing them to interbreeding between farmed escapees and wild individuals. Furthermore, it is imperative to conduct a comprehensive investigation into the mechanisms of local adaptation under high gene flow by employing a population transcriptomics framework. This approach will allow for a conclusive examination of the genetic and expression patterns, as well as an exploration of adaptive divergence at various levels, including sequences, genes, and biological metabolic pathways, among natural populations. This will further establish fundamental data regarding the genetic makeup of broodstock and assess the impact of artificial selection on adaptive polymorphisms, as well as their potential linkage with causal variants underpinning production-related traits in breeding programs.

References

1. FAO. *The State of World Fisheries and Aquaculture 2022—Towards Blue Transformation*; FAO: Rome, Italy, 2022. [CrossRef]
2. Pradeepkiran, J.A. Aquaculture role in global food security with nutritional value: A review. *Transl. Anim. Sci.* **2019**, *3*, 903–910. [CrossRef] [PubMed]
3. Misund, B.; Nygård, R. Big fish: Valuation of the world's largest salmon farming companies. *Mar. Resour. Econ.* **2018**, *33*, 245–261. [CrossRef]
4. Bailey, J.L.; Eggereide, S.S. Indicating sustainable salmon farming: The case of the new Norwegian aquaculture management scheme. *Mar. Policy* **2020**, *117*, 103925. [CrossRef]
5. Iversen, A.; Asche, F.; Hermansen, Ø.; Nystøyl, R. Production cost and competitiveness in major salmon farming countries 2003–2018. *Aquaculture* **2020**, *522*, 735089. [CrossRef]
6. Salmon Farming Industry Handbook. 2023. Available online: https://mowi.com/wp-content/uploads/2023/06/2023-Salmon-Farming-Industry-Handbook-2023.pdf (accessed on 5 August 2023).
7. Cerbule, K.; Godfroid, J. Salmon louse (*Lepeophtheirus salmonis* (Krøyer)) control methods and efficacy in Atlantic salmon (*Salmo salar* (Linnaeus)) aquaculture: A literature review. *Fishes* **2020**, *5*, 11. [CrossRef]
8. Imsland, A.K.D.; Reynolds, P. In lumpfish We Trust? The Efficacy of Lumpfish *Cyclopterus lumpus* to Control *Lepeophtheirus salmonis* Infestations on Farmed Atlantic Salmon: A Review. *Fishes* **2022**, *7*, 220. [CrossRef]
9. Torrissen, O.; Jones, S.; Asche, F.; Guttormsen, A.; Skilbrei, O.T.; Nilsen, F.; Horsberg, T.E.; Jackson, D. Salmon lice—Impact on wild salmonids and salmon aquaculture. *J. Fish Dis.* **2013**, *36*, 171–194. [CrossRef]
10. Bowers, J.M.; Mustafa, A.; Speare, D.J.; Conboy, G.A.; Brimacombe, M.; Sims, D.E.; Burka, J.F. The physiological response of Atlantic salmon, *Salmo salar* L., to a single experimental challenge with sea lice, *Lepeophtheirus salmonis*. *J. Fish Dis.* **2000**, *23*, 165–172. [CrossRef]
11. Grave, K.; Horsberg, T.E.; Lunestad, B.T.; Litleskare, I. Consumption of drugs for sea lice infestations in Norwegian fish farms: Methods for assessment of treatment patterns and treatment rate. *Dis. Aquat. Org.* **2004**, *60*, 123–131. [CrossRef]
12. Hamre, L.A.; Eichner, C.; Caipang, C.M.A.; Dalvin, S.T.; Bron, J.E.; Nilsen, F.; Boxshall, G.; Skern-Mauritzen, R. The Salmon Louse *Lepeophtheirus salmonis* (Copepoda: Caligidae) life cycle has only two Chalimus stages. *PLoS ONE* **2013**, *8*, e73539. [CrossRef]

13. Contreras, M.; Karlsen, M.; Villar, M.; Olsen, R.H.; Leknes, L.M.; Furevik, A.; Yttredal, K.L.; Tartor, H.; Grove, S.; Alberdi, P.; et al. Vaccination with ectoparasite proteins involved in midgut function and blood digestion reduces salmon louse infestations. *Vaccines* **2020**, *8*, 32. [CrossRef] [PubMed]

14. Bui, S.; Stien, L.H.; Nilsson, J.; Trengereid, H.; Oppedal, F. Efficiency and welfare impact of long-term simultaneous in situ management strategies for salmon louse reduction in commercial sea cages. *Aquaculture* **2020**, *520*, 734934. [CrossRef]

15. Barrett, L.T.; Oppedal, F.; Robinson, N.; Dempster, T. Prevention not cure: A review of methods to avoid sea lice infestations in salmon aquaculture. *Rev. Aquac.* **2020**, *12*, 2527–2543. [CrossRef]

16. Johnson, S.C.; Constible, J.M.; Richard, J. Laboratory investigations on the efficacy of hydrogen peroxide against the salmon louse *Lepeophtheirus salmonis* and its toxicological and histopathological effects on Atlantic salmon *Salmo salar* and Chinook salmon *Oncorhynchus tshawytscha*. *Dis. Aquat. Org.* **1993**, *17*, 197–204. [CrossRef]

17. Overton, K.; Dempster, T.; Oppedal, F.; Kristiansen, T.S.; Gismervik, K.; Stien, L.H. Salmon lice treatments and salmon mortality in Norwegian aquaculture: A review. *Rev. Aquac.* **2019**, *11*, 1398–1417. [CrossRef]

18. Skiftesvik, A.B.; Bjelland, R.M.; Durif, C.M.; Johansen, I.S.; Browman, H.I. Delousing of Atlantic salmon (*Salmo salar*) by cultured vs. wild ballan wrasse (*Labrus bergylta*). *Aquaculture* **2013**, *402*, 113–118. [CrossRef]

19. Imsland, A.K.; Reynolds, P.; Eliassen, G.; Hangstad, T.A.; Foss, A.; Vikingstad, E.; Elvegård, T.A. The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeophtheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar* L.). *Aquaculture* **2014**, *424*, 18–23. [CrossRef]

20. Overton, K.; Barrett, L.T.; Oppedal, F.; Kristiansen, T.S.; Dempster, T. Sea lice removal by cleaner fish in salmon aquaculture: A review of the evidence base. *Aquac. Environ. Interact.* **2020**, *12*, 31–44. [CrossRef]

21. Bolton-Warberg, M. An overview of cleaner fish use in Ireland. *J. Fish Dis.* **2018**, *41*, 935–939. [CrossRef]

22. Treasurer, J.; Prickett, R.; Zietz, M.; Hempleman, C.; Garcia de Leaniz, C. Cleaner fish rearing and deployment in the UK. In *Cleaner Fish Biology and Aquaculture Applications*; Treasurer, J.W., Ed.; 5M Publishing Ltd.: Sheffield, UK, 2018; pp. 376–391.

23. Imsland, A.K.D.; Reynolds, P.; Hangstad, T.A.; Kapari, L.; Maduna, S.N.; Hagen, S.B.; Jónsdóttir, Ó.D.B.; Spetland, F.; Lindberg, K.S. Quantification of grazing efficacy, growth and health score of different lumpfish (*Cyclopterus lumpus* L.) families: Possible size and gender effects. *Aquaculture* **2021**, *530*, 735925. [CrossRef]

24. Norway Directorate of Fisheries. Akvakulturstatistikk: Rensefisk. 2022. Available online: https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Rensefisk (accessed on 15 December 2022).

25. Holmyard, N. Lumpfish Production Becoming Big Business in Norway. SeafoodSource. 2018. Available online: https://www.seafoodsource.com/news/aquaculture/lumpfish-production-becoming-big-business-in-norway (accessed on 5 August 2022).

26. Sae-Lim, P.; Khaw, H.L.; Nielsen, H.M.; Puvanendran, V.; Hansen, Ø.; Mortensen, A. Genetic variance for uniformity of body weight in lumpfish (*Cyclopterus lumpus*) used a double hierarchical generalized linear model. *Aquaculture* **2020**, *514*, 734515. [CrossRef]

27. Jónsdóttir, Ó.D.B.; Imsland, A.K.; Kennedy, J. Lumpfish biology and population genetics. In *Cleaner Fish Biology and Aquaculture Applications*; Treasurer, J., Ed.; 5M Publishing Ltd.: Sheffield, UK, 2018.

28. Jónsdóttir, Ó.D.B.; Schregel, J.; Hagen, S.B.; Tobiassen, C.; Aarnes, S.G.; Imsland, A.K. Population genetic structure of lumpfish along the Norwegian coast: Aquaculture implications. *Aquac. Int.* **2018**, *26*, 49–60. [CrossRef]

29. Pampoulie, C.; Skirnisdottir, S.; Olafsdottir, G.; Helyar, S.J.; Thorsteinsson, V.; Jónsson, S.; Fréchet, A.; Durif, C.M.F.; Sherman, S.; Lampart-Kałużniacka, M.; et al. Genetic structure of the lumpfish *Cyclopterus lumpus* across the North Atlantic. *ICES J. Mar. Sci.* **2014**, *71*, 2390–2397. [CrossRef]

30. Garcia-Mayoral, E.; Olsen, M.; Hedeholm, R.; Post, S.; Nielsen, E.E.; Bekkevold, D. Genetic structure of West Greenland populations of lumpfish *Cyclopterus lumpus*. *J. Fish Biol.* **2016**, *89*, 2625–2642. [CrossRef]

31. Whittaker, B.A.; Consuegra, S.; Garcia de Leaniz, C. Genetic and phenotypic differentiation of lumpfish (*Cyclopterus lumpus*) across the North Atlantic: Implications for conservation and aquaculture. *PeerJ* **2018**, *6*, e5974. [CrossRef]

32. Jónsdóttir, Ó.D.B.; Gíslason, D.; Ólafsdóttir, G.; Maduna, S.N.; Hagen, S.B.; Reynolds, P.; Sveinsson, S.; Imsland, A.K. Lack of population genetic structure of lumpfish along the Norwegian coast: A reappraisal based on EST-STRs analyses. *Aquaculture* **2022**, *555*, 738230. [CrossRef]

33. Jansson, E.; Faust, E.; Bekkevold, D.; Quintela, M.; Durif, C.; Halvorsen, K.T.; Dahle, G.; Pampoulie, C.; Kennedy, J.; Whittaker, B.; et al. Global, regional, and cryptic population structure in a high gene-flow transatlantic fish. *PLoS ONE* **2023**, *18*, e0283351. [CrossRef]

34. Williams, G.C. *Adaptation and Natural Selection*; Princeton University Press: Princeton, NJ, USA, 1966.

35. Kawecki, T.J.; Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **2004**, *7*, 1225–1241. [CrossRef]

36. Blanquart, F.; Kaltz, O.; Nuismer, S.L.; Gandon, S. A practical guide to measuring local adaptation. *Ecol. Lett.* **2013**, *16*, 1195–1205. [CrossRef]

37. Savolainen, O.; Lascoux, M.; Merilä, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* **2013**, *14*, 807–820. [CrossRef]

38. Fraser, D.J.; Cook, A.M.; Eddington, J.D.; Bentzen, P.; Hutchings, J.A. Mixed evidence for reduced local adaptation in wild salmon resulting from interbreeding with escaped farmed salmon: Complexities in hybrid fitness. *Evol. Appl.* **2008**, *1*, 501–512. [CrossRef] [PubMed]

39. Bourret, V.; O'Reilly, P.T.; Carr, J.W.; Berg, P.R.; Bernatchez, L. Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity* **2011**, *106*, 500–510. [CrossRef] [PubMed]

40. Antoniou, A.; Manousaki, T.; Ramírez, F.; Cariani, A.; Cannas, R.; Kasapidis, P.; Magoulas, A.; Albo-Puigserver, M.; Lloret-Lloret, E.; Bellido, J.M.; et al. Sardines at a junction: Seascape genomics reveals ecological and oceanographic drivers of variation in the NW Mediterranean Sea. *Mol. Ecol.* **2023**, *32*, 1608–1628. [CrossRef] [PubMed]

41. Levins, R. The theory of fitness in a heterogeneous environment. IV. The adaptive significance of gene flow. *Evol. Int. J. Org. Evol.* **1964**, *18*, 635–638. [CrossRef]

42. Endler, J.A. Gene flow and population differentiation. *Science* **1973**, *179*, 243–250. [CrossRef]

43. Garant, D.A.N.Y.; Forde, S.E.; Hendry, A.P. The multifarious effects of dispersal and gene flow on contemporary adaptation. *Funct. Ecol.* **2007**, *21*, 434–443. [CrossRef]

44. Fitzpatrick, S.W.; Gerberich, J.C.; Kronenberger, J.A.; Angeloni, L.M.; Funk, W.C. Locally adapted traits maintained in the face of high gene flow. *Ecol. Lett.* **2015**, *18*, 37–47. [CrossRef]

45. Tigano, A.; Friesen, V.L. Genomics of local adaptation with gene flow. *Mol. Ecol.* **2016**, *25*, 2144–2164. [CrossRef]

46. Haldane, J.B.S. A mathematical theory of natural and artificial selection (Part VI, Isolation.). In *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: Cambridge, UK, 1930; Volume 26, pp. 220–230.

47. Lenormand, T. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **2002**, *17*, 183–189. [CrossRef]

48. Allendorf, F.W.; Luikart, G. *Conservation and the Genetics of Populations*; Blackwell Publishing: Oxford, UK, 2007.

49. Bridle, J.R.; Vines, T.H. Limits to evolution at range margins: When and why does adaptation fail? *Trends Ecol. Evol.* **2007**, *22*, 140–147. [CrossRef]

50. Tallmon, D.A.; Luikart, G.; Waples, R.S. The alluring simplicity and complex reality of genetic rescue. *Trends Ecol. Evol.* **2004**, *19*, 489–496. [CrossRef]

51. Bontrager, M.; Angert, A.L. Gene flow improves fitness at a range edge under climate change. *Evol. Lett.* **2018**, *3*, 55–68. [CrossRef]

52. Clark, J.D.; Benham, P.M.; Maldonado, J.E.; Luther, D.A.; Lim, H.C. Maintenance of local adaptation despite gene flow in a coastal songbird. *Evol. Int. J. Org. Evol.* **2022**, *76*, 1481–1494. [CrossRef]

53. Salloum, P.M.; Santure, A.W.; Lavery, S.D.; de Villemereuil, P. Finding the adaptive needles in a population-structured haystack: A case study in a New Zealand mollusc. *J. Anim. Ecol.* **2022**, *91*, 1209–1221. [CrossRef] [PubMed]

54. Hoffmann, A.; Griffin, P.; Dillon, S.; Catullo, R.; Rane, R.; Byrne, M.; Jordan, R.; Oakeshott, J.; Weeks, A.; Joseph, L.; et al. A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Clim. Chang. Responses* **2015**, *2*, 1. [CrossRef]

55. Bayona-Vásquez, N.J.; Glenn, T.C.; Kieran, T.J.; Pierson, T.W.; Hoffberg, S.L.; Scott, P.A.; Bentley, K.E.; Finger, J.W.; Louha, S.; Troendle, N.; et al. Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ* **2019**, *7*, e7724. [CrossRef] [PubMed]

56. Davenport, J. *Synopsis of Biological Data on the Lumpsucker, Cyclopterus lumpus (Linnaeus, 1758)*; Food & Agriculture Organization: Rome, Italy, 1985; Volume 147.

57. Bañón, R.; Garazo, A.; Fernández, A. Short communication Note about the presence of the lumpsucker *Cyclopterus lumpus* (Teleostei, Cyclopteridae) in Galician waters (NW Spain). *J. Appl. Ichthyol.* **2007**, *24*, 108–109. [CrossRef]

58. Voskoboinikova, O.S.; Kudryavtseva, O.Y.; Orlov, A.M.; Orlova, S.Y.; Nazarkin, M.V.; Chernova, N.V.; Maznikova, O.A. Relationships and evolution of lumpsuckers of the family Cyclopteridae (Cottoidei). *J. Ichthyol.* **2020**, *60*, 154–181. [CrossRef]

59. Kennedy, J.; Durif, C.M.F.; Florin, A.-B.; Fréchet, A.; Gauthier, J.; Hüssy, K.; Jónsson, S.; Ólafsson, H.G.; Post, S.; Hedeholm, R.B. A brief history of lumpfishing, assessment, and management across the North Atlantic. *ICES J. Mar. Sci.* **2019**, *76*, 181–191. [CrossRef]

60. Kennedy, J.; Jónsson, S.Þ.; Ólafsson, H.G.; Kasper, J.M. Observations of vertical movements and depth distribution of migrating female lumpfish (*Cyclopterus lumpus*) in Iceland from data storage tags and trawl surveys. *ICES J. Mar. Sci.* **2016**, *73*, 1160–1169. [CrossRef]

61. Zvetkov, V.I.; Kalyakina, N.M. Relationships between White Sea fish and macrophytes. *Ecology* **1987**, *6*, 40–43.

62. Kudryavzeva, O.J. The Lumpsucker (*Cyclopterus lumpus*) in the Barents Sea and Adjusted Waters. Ph.D. Thesis, Nauka Press, Moskow, Russia, 2008.

63. Gade, H.G. (Ed.) Features of fjord and ocean interaction. In *The Nordic Seas*; Springer: New York, NY, USA, 1986; pp. 183–190.

64. Sætre, R. (Ed.) *The Norwegian Coastal Current: Oceanography and Climate*; Fagbokforlaget: Oslo, Norway, 2007.

65. Skagseth, Ø.; Drinkwater, K.F.; Terrile, E. Wind-and buoyancy-induced transport of the Norwegian Coastal Current in the Barents Sea. *J. Geophys. Res. Ocean.* **2011**, *116*, C08007. [CrossRef]

66. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* **2012**, *7*, e37135. [CrossRef] [PubMed]

67. Graham, C.F.; Glenn, T.C.; McArthur, A.G.; Boreham, D.R.; Kieran, T.; Lance, S.; Manzon, R.G.; Martino, J.A.; Pierson, T.; Rogers, S.M.; et al. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Mol. Ecol. Resour.* **2015**, *15*, 1304–1315. [CrossRef]

68. Hoffberg, S.L.; Kieran, T.J.; Catchen, J.M.; Devault, A.; Faircloth, B.C.; Mauricio, R.; Glenn, T.C. RADcap: Sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Mol. Ecol. Resour.* **2016**, *16*, 1264–1278. [CrossRef]

69. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 29 June 2023).

70. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef]

71. Rochette, N.C.; Rivera-Colón, A.G.; Catchen, J.M. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **2019**, *28*, 4737–4754. [CrossRef]

72. Holborn, M.K.; Einfeldt, A.L.; Kess, T.; Duffy, S.J.; Messmer, A.M.; Langille, B.L.; Brachmann, M.K.; Gauthier, J.; Bentzen, P.; Knutsen, T.M.; et al. Reference genome of lumpfish *Cyclopterus lumpus* Linnaeus provides evidence of male heterogametic sex determination through the AMH pathway. *Mol. Ecol. Resour.* **2022**, *22*, 1427–1439. [CrossRef]

73. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef]

74. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]

75. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **2016**, *32*, 292–294. [CrossRef]

76. McKinney, G.J.; Waples, R.K.; Seeb, L.W.; Seeb, J.E. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol. Ecol. Resour.* **2017**, *17*, 656–669. [CrossRef] [PubMed]

77. Dorant, Y.; Cayuela, H.; Wellband, K.; Laporte, M.; Rougemont, Q.; Mérot, C.; Normandeau, E.; Rochette, R.; Bernatchez, L. Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species. *Mol. Ecol.* **2020**, *29*, 4765–4782. [CrossRef] [PubMed]

78. Yang, J.; Benyamin, B.; McEvoy, B.P.; Gordon, S.; Henders, A.K.; Nyholt, D.R.; Madden, P.A.; Heath, A.C.; Martin, N.G.; Montgomery, G.W. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **2010**, *42*, 565–569. [CrossRef] [PubMed]

79. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. 1000 Genomes Project Analysis Group. (Aug 1, 2011) The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]

80. Manichaikul, A.; Mychaleckyj, J.C.; Rich, S.S.; Daly, K.; Sale, M.; Chen, W.M. Robust relationship inference in genome-wide association studies. *Bioinformatics* **2010**, *26*, 2867–2873. [CrossRef] [PubMed]

81. Zheng, X.; Levine, D.; Shen, J.; Gogarten, S.M.; Laurie, C.; Weir, B.S. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **2012**, *28*, 3326–3328. [CrossRef]

82. Bresadola, L.; Link, V.; Buerkle, C.A.; Lexer, C.; Wegmann, D. Estimating and accounting for genotyping errors in RAD-seq experiments. *Mol. Ecol. Resour.* **2020**, *20*, 856–870. [CrossRef]

83. Lischer, H.E.; Excoffier, L. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **2012**, *28*, 298–299. [CrossRef]

84. Robinson, M.D.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11*, R25. [CrossRef]

85. Foll, M. BayeScan v2.1 User Manual. 2012. Available online: http://cmpg.unibe.ch/software/BayeScan/ (accessed on 30 June 2022).

86. Foll, M.; Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **2008**, *180*, 977–993. [CrossRef] [PubMed]

87. Lotterhos, K.E.; Whitlock, M.C. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* **2015**, *24*, 1031–1046. [CrossRef] [PubMed]

88. Geweke, J.F. *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments (No. 148)*; Federal Reserve Bank of Minneapolis: St. Louis, MO, USA, 1991.

89. Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence diagnosis and output analysis for MCMC. *R News* **2006**, *6*, 7–11.

90. Luu, K.; Bazin, E.; Blum, M.G. pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **2017**, *17*, 67–77. [CrossRef]

91. Whitlock, M.C.; Lotterhos, K.E. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *Am. Nat.* **2015**, *186* (Suppl. S1), S24–S36. [CrossRef] [PubMed]

92. Privé, F.; Luu, K.; Vilhjálmsson, B.J.; Blum, M.G.B. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Mol. Biol. Evol.* **2020**, *37*, 2153–2154. [CrossRef]

93. Weir, B.S.; Goudet, J. A unified characterization of population structure and relatedness. *Genetics* **2017**, *206*, 2085–2103. [CrossRef]

94. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **2005**, *5*, 184–186. [CrossRef]

95. Goudet, J.; Jombart, T. Hierfstat: Estimation and Tests of Hierarchical F-Statistics. 2022. Available online: https://www.r-project.org (accessed on 15 November 2022).

96. Archer, F.I.; Adams, P.E.; Schneiders, B.B. stratag: An r package for manipulating, summarizing and analysing population genetic data. *Mol. Ecol. Resour.* **2017**, *17*, 5–11. [CrossRef]

97. Weir, B.S.; Cockerham, C.C. Estimating F-statistics for the analysis of population structure. *Evolution* **1984**, *1*, 1358–1370.

98. Redon, R.; Ishikawa, S.; Fitch, K.R.; Feuk, L.; Perry, G.H.; Andrews, T.D.; Fiegler, H.; Shapero, M.H.; Carson, A.R.; Chen, W.; et al. Global variation in copy number in the human genome. *Nature* **2006**, *444*, 444–454. [CrossRef] [PubMed]

99. Dennis, M.Y.; Harshman, L.; Nelson, B.J.; Penn, O.; Cantsilieris, S.; Huddleston, J.; Antonacci, F.; Penewit, K.; Denman, L.; Raja, A.; et al. The evolution and population diversity of human-specific segmental duplications. *Nat. Ecol. Evol.* **2017**, *1*, 0069. [CrossRef] [PubMed]

100. Rinker, D.C.; Specian, N.K.; Zhao, S.; Gibbons, J.G. Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 13446–13451. [CrossRef] [PubMed]

101. Nei, M. Genetic distance between populations. *Am. Nat.* **1972**, *106*, 283–292. [CrossRef]

102. Nei, M.; Takezaki, N. Estimation of genetic distances and phylogenetic trees from DNA analysis. In Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph, ON, Canada, 7–12 August 1994; Volume 21, pp. 405–412.

103. Pembleton, L.W.; Cogan, N.O.; Forster, J.W. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* **2013**, *13*, 946–952. [CrossRef] [PubMed]

104. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [CrossRef]

105. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 94. [CrossRef]

106. Jombart, T.; Ahmed, I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **2011**, *27*, 3070–3071. [CrossRef]

107. Frichot, E.; François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **2015**, *6*, 925–929. [CrossRef]

108. Gain, C.; François, O. LEA 3: Factor models in population genetics and ecological genomics with R. *Mol. Ecol. Resour.* **2021**, *21*, 2738–2748. [CrossRef]

109. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [CrossRef] [PubMed]

110. Francis, R.M. pophelper: An R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **2017**, *17*, 27–32. [CrossRef] [PubMed]

111. Malinsky, M.; Trucchi, E.; Lawson, D.J.; Falush, D. RADpainter and fineRADstructure: Population inference from RADseq data. *Mol. Biol. Evol.* **2018**, *35*, 1284–1290. [CrossRef] [PubMed]

112. Lawson, D.J.; Hellenthal, G.; Myers, S.; Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **2012**, *8*, e1002453. [CrossRef] [PubMed]

113. Rellstab, C.; Gugerli, F.; Eckert, A.J.; Hancock, A.M.; Holderegger, R. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* **2015**, *24*, 4348–4370. [CrossRef]

114. Riginos, C.; Crandall, E.D.; Liggins, L.; Bongaerts, P.; Treml, E.A. Navigating the currents of seascape genomics: How spatial analyses can augment population genomic studies. *Curr. Zool.* **2016**, *62*, 581–601. [CrossRef]

115. Leempoel, K.; Duruz, S.; Rochat, E.; Widmer, I.; Orozco-terWengel, P.; Joost, S. Simple rules for an efficient use of geographic information systems in molecular ecology. *Front. Ecol. Evol.* **2017**, *5*, 33. [CrossRef]

116. Riginos, C.; Liggins, L. Seascape genetics: Populations, individuals, and genes marooned and adrift. *Geogr. Compass* **2013**, *7*, 197–216. [CrossRef]

117. Benestan, L.; Quinn, B.K.; Maaroufi, H.; Laporte, M.; Clark, F.K.; Greenwood, S.J.; Rochette, R.; Bernatchez, L. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol. Ecol.* **2016**, *25*, 5073–5092. [CrossRef]

118. Selkoe, K.; D'aloia, C.; Crandall, E.; Iacchei, M.; Liggins, L.; Puritz, J.; von der Heyden, S.; Toonen, R. A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Mar. Ecol. Prog. Ser.* **2016**, *554*, 1–19. [CrossRef]

119. Cayuela, H.; Rougemont, Q.; Laporte, M.; Mérot, C.; Normandeau, E.; Dorant, Y.; Tørresen, O.K.; Hoff, S.N.K.; Jentoft, S.; Sirois, P.; et al. Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Mol. Ecol.* **2020**, *29*, 2379–2398. [CrossRef] [PubMed]

120. Hijmans, R.J. Raster: Geographic data analysis and modeling. *R Package* **2016**, *734*, 473.

121. Barbosa, A.M. fuzzySim: Applying fuzzy logic to binary similarity indices in ecology. *Methods Ecol. Evol.* **2015**, *6*, 853–858. [CrossRef]

122. Marquardt, D.W. Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics* **1970**, *12*, 591–612. [CrossRef]

123. Mansfield, E.R.; Helms, B.P. Detecting multicollinearity. *Am. Stat.* **1982**, *36*, 158–160.

124. Roever, C.; Raabe Luebke, K.; Ligges, U.; Szepannek, G.; Zentgraf, M.; Meyer, D. klaR: Classification and Visualization, R Package Version 0.6-15. 2023. Available online: https://cran.r-project.org/web/packages/klaR/index.html (accessed on 15 June 2023).

125. Dray, S.; Legendre, P.; Peres-Neto, P.R. Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Model.* **2006**, *196*, 483–493. [CrossRef]

126. Dray, S.; Pélissier, R.; Couteron, P.; Fortin, M.-J.; Legendre, P.; Peres-Neto, P.R.; Bellier, E.; Bivand, R.; Blanchet, F.G.; De Cáceres, M.; et al. Community ecology in the age of multivariate spatial analysis. *Ecol. Monogr.* **2012**, *82*, 257–275. [CrossRef]

127. Peres-Neto, P.R.; Legendre, P. Estimating and controlling for spatial structure in the study of ecological communities. *Glob. Ecol. Biogeogr.* **2010**, *19*, 174–184. [CrossRef]

128. Borcard, D.; Gillet, F.; Legendre, P.; Borcard, D.; Gillet, F.; Legendre, P. Spatial analysis of ecological data. In *Numerical Ecology with R*; Springer: New York, NY, USA, 2018; pp. 299–367. [CrossRef]

129. Chambers, J.M. *No. 1., Software for Data Analysis: Programming with R*; Springer: New York, NY, USA, 2008; Volume 2.

130. Pante, E.; Simon-Bouhet, B. marmap: A package for importing, plotting and analyzing bathymetric and topographic data in R. *PLoS ONE* **2013**, *8*, e73051. [CrossRef]

131. Dray, S.; Bauman, D.; Blanchet, G.; Borcard, D.; Clappe, S.; Guénard, G.; Madi, N.H. Adespatial: Multivariate Multiscale Spatial Analysis. *R Package* **2022**, 1–138. Available online: https://cran.r-project.org/src/contrib/Archive/adespatial/ (accessed on 20 August 2023).

132. Dray, S.; Dufour, A.B. The ade4 Package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **2007**, *22*, 1–20. [CrossRef]

133. Forester, B.R.; Lasky, J.R.; Wagner, H.H.; Urban, D.L. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* **2018**, *27*, 2215–2233. [CrossRef] [PubMed]

134. Caye, K.; Jumentier, B.; Lepeule, J.; François, O. LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Mol. Biol. Evol.* **2019**, *36*, 852–860. [CrossRef] [PubMed]

135. Ahrens, C.W.; Rymer, P.D.; Stow, A.; Bragg, J.; Dillon, S.; Umbers, K.D.; Dudaniec, R.Y. The search for loci under selection: Trends, biases and progress. *Mol. Ecol.* **2018**, *27*, 1342–1356. [CrossRef]

136. Waldvogel, A.M.; Schreiber, D.; Pfenninger, M.; Feldmeyer, B. Climate change genomics calls for standardized data reporting. *Front. Ecol. Evol.* **2020**, *8*, 242. [CrossRef]

137. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'hara, R.B.; Simpson, G.L.; Solymos, P.; et al. Package 'Vegan'. Community Ecology Package, Version. 2019. Available online: https://CRAN.R-project.org/package= vegan (accessed on 25 November 2022).

138. Macarthur, R.H. On the relative abundance of bird species. *Proc. Natl. Acad. Sci. USA* **1957**, *43*, 293–295. [CrossRef]

139. King, J.R.; Jackson, D.A. Variable selection in large environmental data sets using principal components analysis. *Environmetrics Off. J. Int. Environmetrics Soc.* **1999**, *10*, 67–77. [CrossRef]

140. Storey, J.D. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Stat.* **2003**, *31*, 2013–2035. [CrossRef]

141. Capblancq, T.; Luu, K.; Blum, M.G.B.; Bazin, E. Evaluation of redundancy analysis to identify signatures of local adaptation. *Mol. Ecol. Resour.* **2018**, *18*, 1223–1233. [CrossRef]

142. Cingolani, P. Variant annotation and functional prediction: SnpEff. In *Variant Calling: Methods and Protocols*; Springer: New York, NY, USA, 2018; pp. 289–314.

143. Brodie, A.; Azaria, J.R.; Ofran, Y. How far from the SNP may the causative genes be? *Nucleic Acids Res.* **2016**, *44*, 6046–6054. [CrossRef]

144. Visel, A.; Rubin, E.M.; Pennacchio, L.A. Genomic views of distant-acting enhancers. *Nature* **2009**, *461*, 199–205. [CrossRef] [PubMed]

145. Selmoni, O.; Rochat, E.; Lecellier, G.; Berteaux-Lecellier, V.; Joost, S. Seascape genomics as a new tool to empower coral reef conservation strategies: An example on north-western Pacific *Acropora digitifera*. *Evol. Appl.* **2020**, *13*, 1923–1938. [CrossRef] [PubMed]

146. Madden, T.; Coulouris, G. *BLAST Command Line Applications User Manual BLAST Command Line Applications User Manual-BLAST*; National Center for Biotechnology Information: Bethesda, MD, USA, 2008. Available online: http://www.ncbi.nlm.nih.gov/ books/NBK279690/ (accessed on 15 March 2023).

147. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370. [CrossRef]

148. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **2019**, *10*, 1523. [CrossRef] [PubMed]

149. Charlesworth, B. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **2009**, *10*, 195–205. [CrossRef]

150. Greene, J.S.; Brown, M.; Dobosiewicz, M.; Ishida, I.G.; Macosko, E.Z.; Zhang, X.; Butcher, R.A.; Cline, D.J.; McGrath, P.T.; Bargmann, C.I. Balancing selection shapes density-dependent foraging behaviour. *Nature* **2016**, *539*, 254–258. [CrossRef]

151. Nettle, D. The evolution of personality variation in humans and other animals. *Am. Psychol.* **2006**, *61*, 622–631. [CrossRef]

152. Penke, L.; Denissen, J.J.; Miller, G.F. The evolutionary genetics of personality. *Eur. J. Personal.* **2007**, *21*, 549–587. [CrossRef]

153. Christie, M.R.; McNickle, G.G.; French, R.A.; Blouin, M.S. Life history variation is maintained by fitness trade-offs and negative frequency-dependent selection. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4441–4446. [CrossRef]

154. Cavedon, M.; Goubili, C.; Heppenheimer, E.; Vonholdt, B.; Mariani, S.; Hebblewhite, M.; Hegel, T.; Hervieux, D.; Serrouya, R.; Steenweg, R.; et al. Genomics, environment and balancing selection in behaviourally bimodal populations: The caribou case. *Mol. Ecol.* **2019**, *28*, 1946–1963. [CrossRef]

155. Carrier, A.; Prunier, J.; Poisson, W.; Trottier-Lavoie, M.; Gilbert, I.; Cavedon, M.; Pokharel, K.; Kantanen, J.; Musiani, M.; Côté, S.D.; et al. Design and validation of a 63K genome-wide SNP-genotyping platform for caribou/reindeer (*Rangifer tarandus*). *BMC Genom.* **2022**, *23*, 687. [CrossRef] [PubMed]

156. Halvorsen, K.T.; Larsen, T.; Sørdalen, T.K.; Vøllestad, L.A.; Knutsen, H.; Olsen, E.M. Impact of harvesting cleaner fish for salmonid aquaculture assessed from replicated coastal marine protected areas. *Mar. Biol. Res.* **2017**, *13*, 359–369. [CrossRef]

157. Imsland, A.K.; Reynolds, P.; Nytrø, A.V.; Eliassen, G.; Hangstad, T.A.; Jónsdóttir, Ó.D.; Emaus, P.-A.; Elvegård, T.A.; Lemmens, S.C.; Rydland, R.; et al. Effects of lumpfish size on foraging behaviour and co-existence with sea lice infected Atlantic salmon in sea cages. *Aquaculture* **2016**, *465*, 19–27. [CrossRef]

158. Imsland, A.K.; Reynolds, P.; Eliassen, G.; Mortensen, A.; Hansen, Ø.J.; Puvanendran, V.; Hangstad, T.A.; Jónsdóttir, Ó.D.; Emaus, P.A.; Elvegård, T.A.; et al. Is cleaning behaviour in lumpfish (*Cyclopterus lumpus*) parentally controlled? *Aquaculture* **2016**, *459*, 156–165. [CrossRef]

159. Whittaker, B.A.; Consuegra, S.; de Leaniz, C.G. Personality profiling may help select better cleaner fish for sea-lice control in salmon farming. *Appl. Anim. Behav. Sci.* **2021**, *243*, 105459. [CrossRef]

160. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*; John Murray: London, UK, 1859. [CrossRef]

161. Hedrick, P.W.; Ginevan, M.E.; Ewing, E.P. Genetic polymorphism in heterogeneous environments. *Annu. Rev. Ecol. Syst.* **1976**, *7*, 1–32. [CrossRef]

162. Palumbi, S.R. Genetic divergence, reproductive isolation, and marine speciation. *Annu. Rev. Ecol. Syst.* **1994**, *25*, 547–572. [CrossRef]

163. Selkoe, K.A.; Henzler, C.M.; Gaines, S.D. Seascape genetics and the spatial ecology of marine populations. *Fish Fish.* **2008**, *9*, 363–377. [CrossRef]

164. Coscia, I.; Wilmes, S.B.; Ironside, J.E.; Goward-Brown, A.; O'Dea, E.; Malham, S.K.; McDevitt, A.D.; Robins, P.E. Fine-scale seascape genomics of an exploited marine species, the common cockle *Cerastoderma edule*, using a multimodelling approach. *Evol. Appl.* **2020**, *13*, 1854–1867. [CrossRef]

165. Huang, J.; Li, Y.; Liu, Z.; Kang, Y.; Wang, J. Transcriptomic responses to heat stress in rainbow trout *Oncorhynchus mykiss* head kidney. *Fish Shellfish Immunol.* **2018**, *82*, 32–40. [CrossRef]

166. Huang, Z.; Ma, A.; Yang, S.; Liu, X.; Zhao, T.; Zhang, J.; Wang, X.A.; Sun, Z.; Liu, Z.; Xu, R. Transcriptome analysis and weighted gene co-expression network reveals potential genes responses to heat stress in turbot Scophthalmus maximus. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2020**, *33*, 100632. [CrossRef] [PubMed]

167. Uthicke, S.; Deshpande, N.P.; Liddy, M.; Patel, F.; Lamare, M.; Wilkins, M.R. Little evidence of adaptation potential to ocean acidification in sea urchins living in "Future Ocean" conditions at a CO_2 vent. *Ecol. Evol.* **2019**, *9*, 10004–10016. [CrossRef] [PubMed]

168. Petit-Marty, N.; Nagelkerken, I.; Connell, S.D.; Schunter, C. Natural CO_2 seeps reveal adaptive potential to ocean acidification in fish. *Evol. Appl.* **2021**, *14*, 1794–1806. [CrossRef] [PubMed]

169. De Wit, P.; Palumbi, S.R. Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. *Mol. Ecol.* **2013**, *22*, 2884–2897. [CrossRef] [PubMed]

170. Hammerschmidt, M.; Brook, A.; McMahon, A.P. The world according to hedgehog. *Trends Genet.* **1997**, *13*, 14–21. [CrossRef] [PubMed]

171. Chuang, P.T.; McMahon, A.P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **1999**, *397*, 617–621. [CrossRef]

172. Katoh, M.; Katoh, M. WNT signaling pathway and stem cell signaling network. *Clin. Cancer Res.* **2007**, *13*, 4042–4045. [CrossRef]

173. Chen, R.E.; Thorner, J. Function and regulation in MAPK signaling pathways: Lessons learned from the yeast Saccharomyces cerevisiae. *Biochim. Biophys. Acta* **2007**, *1773*, 1311–1340. [CrossRef]

174. Rehwinkel, J.; Gack, M.U. RIG-I-like receptors: Their regulation and roles in RNA sensing. *Nat. Rev. Immunol.* **2020**, *20*, 537–551. [CrossRef]