



## RESEARCH ARTICLE

# The role of genomic signatures of directional selection and demographic history in the population structure of a marine teleost with high gene flow

Peggy Weist<sup>1</sup>  | Sissel Jentoft<sup>2</sup>  | Ole K. Tørresen<sup>2</sup> | Franziska M. Schade<sup>3</sup> | Christophe Pampoulie<sup>4</sup> | Uwe Krumme<sup>3</sup> | Reinhold Hanel<sup>1</sup>

<sup>1</sup>Thünen Institute of Fisheries Ecology, Bremerhaven, Germany

<sup>2</sup>Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

<sup>3</sup>Thünen Institute of Baltic Sea Fisheries, Rostock, Germany

<sup>4</sup>Marine and Freshwater Research Institute, Hafnafjörður, Iceland

## Correspondence

Peggy Weist, Thünen Institute of Fisheries Ecology, Herwigstraße 31, 27572 Bremerhaven, Germany.  
Email: [weist.peggy@gmail.com](mailto:weist.peggy@gmail.com)

## Funding information

European Maritime and Fisheries Fund (EMFF) of the European Union (EU)

## Abstract

Recent studies have uncovered patterns of genomic divergence in marine teleosts where panmixia due to high gene flow has been the general paradigm. These signatures of divergent selection are often impacted by structural variants, acting as “supergenes” facilitating local adaptation. The highly dispersing European plaice (*Pleuronectes platessa*)—in which putative structural variants (i.e., inversions) have been identified—has successfully colonized the brackish water ecosystem of the Baltic Sea. Thus, the species represents an ideal opportunity to investigate how the interplay of gene flow, structural variants, natural selection, past demographic history, and gene flow impacts on population (sub)structuring in marine systems. Here, we report on the generation of an annotated draft plaice genome assembly in combination with population sequencing data—following the salinity gradient from the Baltic Sea into the North Sea together with samples from Icelandic waters—to illuminate genome-wide patterns of divergence. Neutral markers pointed at large-scale panmixia across the European continental shelf associated with high gene flow and a common postglacial colonization history of shelf populations. However, based on genome-wide outlier loci, we uncovered signatures of population substructuring among the European continental shelf populations, i.e., suggesting signs of ongoing selection. Genome-wide selection analyses (xp-EHH) and the identification of genes within genomic regions of recent selective sweeps—overlapping with the outlier loci—suggest that these represent the signs of divergent selection. Our findings provide support for genomic divergence driven by local adaptation in the face of high gene flow and elucidate the relative importance of demographic history versus adaptive divergence in shaping the contemporary population genetic structure of a marine teleost. The role of the putative inversion(s) in the substructuring—and potentially ongoing adaptation—was seemingly not substantial.

## KEYWORDS

adaptive variation, candidate gene, divergence hitchhiking, genome assembly, ice ages, single nucleotide polymorphism

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

TAXONOMY CLASSIFICATION  
Evolutionary ecology, Genomics

## 1 | INTRODUCTION

Recent progress in state-of-the-art genomic tools has been most helpful in our understanding of the evolutionary trajectory of non-model organisms in the last decade (Ekblom & Galindo, 2011; Helyar et al., 2011). Such comprehensive characterizations have included genome-wide patterns of divergence, e.g., loci that are under adaptive selection or genetic drift, as well as neutral genomic variation, which are indicative of demographic processes such as gene flow, migration, and dispersal (Ellegren & Galtier, 2016; Gagnaire et al., 2015). In marine teleosts, the absence of physical barriers to migration and larval dispersal promotes genetic connectivity even between geographically distant populations (Cowen et al., 2000). Genomic signatures of local adaptation are expected to be rare in such species since strong directional selection pressure is needed to override the effect of random genetic drift and of migration between populations (Tigano & Friesen, 2016). However, there is an increasing number of studies demonstrating patterns of genetic and genomic divergence and thus, evidence for signatures of local adaptation in marine teleosts despite high gene flow (Barth et al., 2017; Hemmer-Hansen et al., 2007). These investigations have largely contributed to our understanding of the molecular basis of adaptation and identified key sources of adaptive evolution (Feder et al., 2012; Wellenreuther et al., 2019). Structural variants—i.e., smaller or larger genomic reorganizations including chromosomal inversions—have been demonstrated as potent forces in driving local adaptation and diversification allowing for the inheritance of co-adapted gene complexes (Wellenreuther et al., 2019) and play an important role in maintaining genomic regions of divergence even in a milieu of high gene flow (Pettersson et al., 2019). Such chromosomal inversions can act as “supergenes” facilitating local adaptation, by the fact that they encompass multiple neighboring genes with potential beneficial mutations that are inherited together because of close genetic linkage due to lower recombination within these larger haplotype blocks (Tigano & Friesen, 2016). For instance, in Atlantic cod (*Gadus morhua*) four larger chromosomal inversions have been demonstrated to discriminate between populations throughout its geographical distribution, as well as the nonmigratory versus migratory ecotypes (Barth et al., 2017; Berg et al., 2015). Such patterns can promote parts of the genome underlying reproductive isolation or adaptation and prevent them from the homogenizing effects of gene flow leading to so-called genomic islands of divergence. Such genomic islands could also be facilitated by divergence hitchhiking, i.e., when gene exchange between diverging populations is reduced over large genomic regions of several megabases as an accompanying effect of strong divergent selection on genes involved in local adaptation (Via, 2012), and thus increase the degree of differentiation and eventually lead to reproductive isolation among locally adapted populations (Feder et al., 2012). Linked selection has also recently

been shown to form genomic islands of divergence between populations of European sea bass (*Dicentrarchus labrax*) during their past geographic isolation (Duranton et al., 2018).

The North Sea and the Baltic Sea represent a marine and a brackish water ecosystem connected by a relatively steep salinity gradient along the shallow straits of the Danish archipelago. During the late Pleistocene, closed freshwater pools were formed in the southern Baltic region from meltwater as the ice retreated northward. The history of the Baltic Sea started when the preceding freshwater lake opened to the North Sea about 8000 years BP (Björck, 1995). Most of the marine species present today in the Baltic Sea descend from a diverse postglacial flora and fauna established circa 8000–4000 years BP at the time when the Baltic Sea was more saline than today (Winterhalter et al., 1981). Postglacial rebound lowered the connection to the North Sea allowing less saltwater to enter the Littorina Sea, which gradually evolved into the current brackish Baltic Sea. The successive colonization of the Baltic Sea and its characteristics—deeper basins that are connected by shallower sills—has stimulated unique adaptations to life in a stratified brackish environment.

Given its evolutionarily young age, a restricted water exchange with the North Atlantic Ocean, and its predominantly brackish conditions, the Baltic Sea constitutes an ideal study system to investigate patterns of genetic variability and to identify footprints of selection. Over the past thousands of years, marine species colonizing the Baltic Sea were exposed to strong selection pressures imposed by the estuarine conditions such as increasing salinity and decreasing oxygen levels with depth. Many species within the Baltic Sea have evolved locally adapted populations showing increased resilience towards lower salinity, oxygen depletion, and higher water temperatures (Johannesson & André, 2006). Several studies revealed the presence of genetic signatures of directional selection in local Baltic Sea fish populations, which were associated with environmental conditions such as temperature, salinity, or oxygen supply (Guo et al., 2016; Nielsen et al., 2009). For example, in Atlantic cod, genome scans detected outlier polymorphisms strongly associated with adaptation to the local environmental conditions, both within the identified chromosomal inversions and at the genome-wide level, indicating reproductive isolation between the low-salinity adapted Baltic cod and the adjacent cod populations (Berg et al., 2015). Among others, candidate genes under selection were functionally related to hydration and the development of oocytes. Indeed, eggs from the eastern Baltic cod stock display remarkable phenotypic differences in neutral egg buoyancy compared with populations inhabiting a more saline environment, likely to compensate for water loss by being less permeable (Petereit et al., 2014).

The European plaice (*Pleuronectes platessa*) is a commercially exploited flatfish species distributed along the Atlantic coasts of Europe including the Baltic Sea (Nielsen, 1986). Both fecundity

and dispersal capacities (mostly during the larval phase) are considered to be high (Rijnsdorp, 1991). In the North Sea, spawning occurs off the Scottish coast (De Veen, 1962) and offshore in the deeper, more saline parts of the southern North Sea (Harding et al., 1978). The main spawning season lasts from January to March (Rijnsdorp, 1989). After spawning, the pelagic eggs and larval stages are advected north-eastwards from the spawning to the nursery grounds in the shallow coastal areas of the Wadden Sea (Cushing, 1990). Juveniles remain here for up to 2 years before they move deeper and recruit to adult feeding grounds. Tagging experiments have further uncovered highly directional seasonal migrations from the winter spawning area to the summer feeding ground and indications of strong spawning site fidelity, indicative of potential subpopulation structuring of plaice in the North Sea area (Hunter et al., 2004). Spawning activities were also reported for the Skagerrak and Kattegat areas (Nielsen et al., 2004), as well as in the Baltic Sea (Nissling et al., 2002). In the Baltic Sea, spawning occurs in the deeper, more saline basins of the western Baltic Sea, the Arkona Sea, and the Bornholm basin from November to March, peaking in the first quarter. In this area, a higher frequency of plaice migrations between the basins, and thus, lower site fidelity is reported (Aro, 1989).

Past genetic studies of plaice showed high levels of gene flow across large geographic distances with a sharp genetic break between the European continental shelf and the off-shelf population around Iceland (Hoarau et al., 2004; Was et al., 2010). The clear genetic differentiation was attributed to tectonic barriers involving bathymetry and ocean currents (Was et al., 2010). However, while microsatellite analyses revealed genetic homogeneity over much of the European continental shelf, samples from the North Sea, Norway, and the Baltic Sea were weakly distinguishable using mitochondrial DNA (Hoarau et al., 2004; Was et al., 2010). The lack of population structure over substantial parts of the species' distributional range has mainly been attributed to a combination of past colonization events, imperfect spawning site fidelity, and the high dispersal ability of the pelagic early developmental stages (Hoarau et al., 2004; Was et al., 2010). Recent attempts using modern sequencing techniques (ddRAD-sequencing) uncovered differentiation between plaice from the North Sea and the Baltic Sea (Le Moan et al., 2019, 2021; Ulrich et al., 2013), when mapping their sequencing data towards the chromosome-anchored genome of the closely related Japanese flounder (*Paralichthys olivaceus*; Shao et al., 2017). This divergence was mainly driven by two putative structural variants (Le Moan et al., 2019, 2021)—on chromosome 19 and chromosome 21—that showed to some degree an allele frequency differentiation following the salinity gradient from the North Sea into the Baltic Sea. Higher overall levels of linkage disequilibrium (LD) and increased genomic divergence ( $F_{ST}$  and nucleotide diversity  $\pi$ ) were detected in these regions compared with average genome wide (Le Moan et al., 2021). Additionally, they detected a population genetic differentiation among continental shelf samples driven by isolation by distance (Le Moan et al., 2021).

Given the recent improvements in genomic sequencing techniques and data acquisition as well as in the theoretical frameworks

of adaptive evolution in the face of high gene flow, divergent selection and its role in shaping locally adapted populations are and will be continually better understood (Tigano & Friesen, 2016). To improve our understanding of plaice population structure and the impact of gene flow and local adaptation in the Baltic Sea, we here, report on a study that has fully exploited the possibilities of high-throughput sequencing. An annotated draft genome assembly of European plaice was generated using a combination of short- and long-read sequencing technologies, Illumina and PacBio, respectively. Access to a draft genome assembly for the species of interest is crucial to improve mappability, obtain a reliable baseline for variant calling to discover and survey genome-wide polymorphisms, describe the population genetic structure of a high gene flow species, and illuminate how local adaptation affects genomic differentiation between populations (Eklom & Galindo, 2011; Wellenreuther et al., 2019). Additionally, we have whole-genome sequenced (to approx. 8x mean coverage) specimens from locations following the salinity gradient from the hypo-haline Gdańsk deep (GDA) and Bornholm basin (BOR) in the Baltic Sea towards the Arkona basin (ARK) and the transition zone (Belt Sea [BEL] and Kattegat [KAT]) and further into the North Sea (i.e., the Eastern North Sea [ENO] and the central North Sea [CNO]) together with samples from the Western North Sea (WNO) and Icelandic waters (ICE). With the new resources in place, we first re-evaluate the hypothesis of genetic panmixia of plaice from the European continental shelf. We then elucidate genome-wide signatures of divergent selection and recent selective sweeps and identify genes that might be functionally associated with adaptation to different environments. We consider the different environmental conditions experienced by plaice across both, small and large spatial scales. Finally, we confirm the previously detected past demographic history of plaice using Markovian coalescent analyses of past effective population size and reconstruct the colonization of the European continental shelf and the Baltic Sea based on a high-resolution genome-wide SNP set.

## 2 | MATERIAL AND METHODS

### 2.1 | Sequencing and de novo assembly of a reference genome

To assemble a reference genome of the European plaice, we combined data obtained from long- and short-read sequencing from a single female specimen of *Pleuronectes platessa* from the Belt Sea (Table S1) since females are expected to be the heterogametic sex in plaice (Purdom, 1972). High molecular weight DNA was extracted from liver tissue using a high salt precipitation method (<https://www.mn.uio.no/cees/english/people/researcher-postdoc/jentoft/sop-038-high-salt-dna-extraction.pdf>). DNA was purified using HighPrep polymerase chain reaction (PCR) magnetic beads (MAGBIO). DNA integrity was assessed on a 1%-agarose gel and DNA content was quantified using a Qubit 3.0 Fluorometer (Invitrogen). Long-read libraries were prepared using the Pacific Biosciences 20kb library preparation protocol and sequenced

Sample location	Sample code	Region	No individuals	Month	Year
Iceland	ICE	ICE	10	11	2018
Western North Sea	WNO	SHELF, NS	4	2	2016
Central North Sea	CNO	SHELF, NS	3	2	2016
Eastern North Sea	ENO	SHELF, NS	3	2	2016
Kattegat	KAT	SHELF <sup>a</sup>	5	10	2015
Belt Sea	BEL	SHELF, BS	10	2, 3	2016
Arkona basin	ARK	SHELF, BS	10	11, 2, 3	2015, 2016
Bornholm basin	BOR	SHELF, BS	10	2, 3	2016
Gdańsk deep	GDA	SHELF, BS	8	10	2015

Note: Sample location and number of individuals are given per location with the respective population affiliation, as well as the sampling month and year (see Table S1 for more details). Sampling locations were grouped according to geography for the demographic inference and the detection of regions under divergence (BS, Baltic Sea; ICE, Iceland; NS, North Sea; SHELF, European continental shelf).

<sup>a</sup>All individuals originating from the transition area were excluded from comparisons NS versus BS.

aiming at a coverage of 30x on a Pacific Biosciences RS II instrument. All library preparation and sequencing steps mentioned in this study were performed at the Norwegian Sequencing Centre (NSC), CEES, UiO, Norway. Short-insert (550 bp) libraries were constructed using the TruSeq DNA PCR-Free LT Library Preparation Kit (Illumina), and libraries were paired-end sequenced with read lengths of 250 bp (Rapid Run Mode) aiming at a coverage of 120x on an Illumina HiSeq 2500 system.

The PacBio raw reads were then assembled based on an overlap size set at 2500bp and initially scaffolded using the software Flye v2.3 (Kolmogorov et al., 2019). The draft genome was further scaffolded using SGA v0.10.15 (Simpson & Durbin, 2012) and polished by applying the Arrow algorithm provided within the GenomicConsensus package v2.1.0 (Chin et al., 2013) by mapping back raw PacBio reads onto the preliminary genome. A second polishing step used the short reads generated by Illumina HiSeq 2500 to further improve and error correct the genome assembly as implemented in the software Pilon v1.8 (Walker et al., 2014).

The genome size for the plaice was calculated based on the 17-mer distribution with Jellyfish v2.1.4 (Marçais & Kingsford, 2011). We assessed the genome completeness by detecting a set of universal single-copy orthologs ( $N = 4584$ ) specific to Actinopterygii using BUSCO v3.0.1 (Seppey et al., 2019). The repetitive component of the final plaice assembly was estimated using RepeatMasker v4.0.6 (Smit et al., 2015) and the vertebrate database (Bao et al., 2015).

## 2.2 | RNA-sequencing, transcriptome assembly and genome annotation

We isolated RNA from gill, spleen, and liver tissue from a total of 12 adult female plaice individuals caught in the Baltic Sea (Table S1). RNA was extracted using the Total RNA Kit (peqGold), and 150bp paired-end libraries were prepared using the TruSeq

TABLE 1 Sample characteristics of plaice individuals ( $N = 63$ ) considered for sequencing and analysis of population structure.

DNA PCR-Free Library prep kit for sequencing on an Illumina HiSeq 4000 instrument. Raw reads were adapter trimmed and quality filtered using Trimmomatic v0.36 (Bolger et al., 2014) and assembled using Trinity v3.2.3 (Grabherr et al., 2011) with default parameters.

For gene prediction and annotation, we followed the pipeline implemented in Funannotate v1.5.2 (<http://www.github.com/nextgenusfs/funannotate>), which combines multiple ab initio gene model predictions with evidence (transcripts or proteins) aligned to the genome and outputs consensus gene models. We used the Trinity assembly as transcript evidence. We further generated a second transcript evidence by mapping the RNA reads to the plaice reference genome with Hisat v2.1.0 (Kim et al., 2015) and assembled the resulting alignments into potential transcripts with Stringtie v1.3.1 (Pertea et al., 2015). Uncertain splice junctions were identified and filtered from the transcripts using Portcullis v1.2.2 (Mapleson et al., 2018). The software Mikado v1.2.4 (Venturini et al., 2018) was used to define expressed loci and a set of "best" gene models. We passed the transcripts to BRAKER v1.9.3 (Hoff et al., 2016), a tool for ab initio gene prediction. Transcript evidence and gene models were then submitted to Funannotate and gene prediction was run with the zebrafish (*Danio rerio*) set as seed species and a maximum intron length of 100kb. We assessed the completeness of our transcriptome annotation with BUSCO v3.0.1 (Seppey et al., 2019).

## 2.3 | Genome sequencing and data analysis

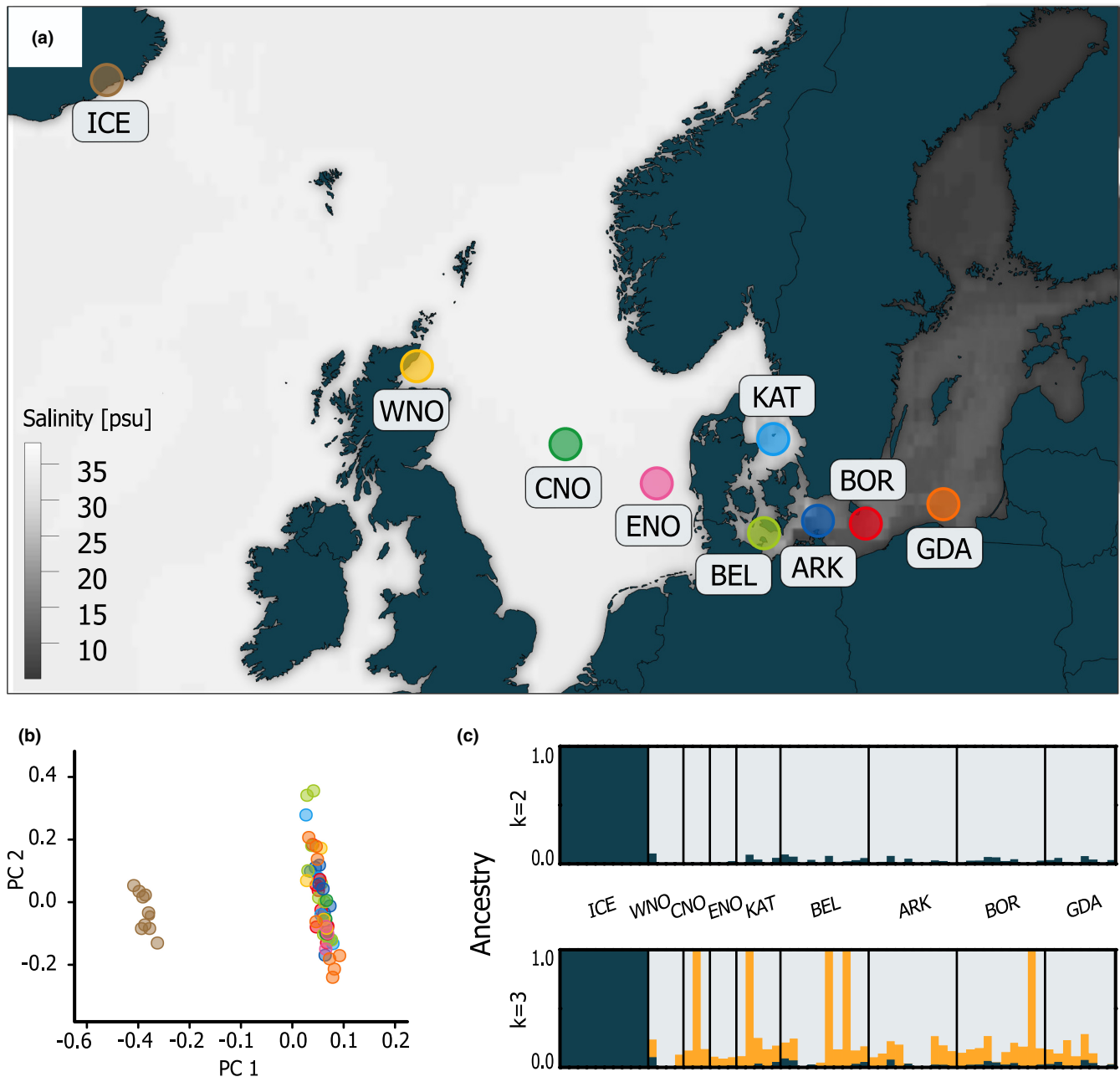
### 2.3.1 | Sampling and sequencing

Tissue from 53 individuals was collected at eight sampling sites covering the species' distribution from Moray Firth in the western North Sea to Gdańsk Bay in the eastern Baltic Sea, representing the edge of the species distribution area (Table 1, Table S1, Figure 1a). Adult

mature individuals, either in ripening or spawning conditions were obtained during scientific surveys and from commercial catches. In addition, we included 10 samples from Iceland as outgroup samples, since they are known to be divergent from plaice from the continental shelf (Was et al., 2010). DNA was extracted from each individual using the high salt extraction protocol. Each of these 63 individuals was 150bp paired-end sequenced (insert size 350bp) with the Illumina HiSeq 4000 sequencing technology.

### 2.3.2 | Mapping to the reference genome and variant calling

Raw reads were adapter and quality trimmed using Trim Galore v0.3.3 (Krueger, 2015) and mapped against the final plaice genome assembly using the BWA-MEM algorithm in BWA v0.7.12 (Li, 2013). Duplicates were marked with Picard tools and indels were locally realigned with GATK v3.7 (McKenna et al., 2010). Variants were



**FIGURE 1** Population structure analysis from a total of 63 European plaice samples based on 83,011 genome-wide distributed single nucleotide polymorphisms (SNPs). (a) Map of sampling locations of plaice individuals included in this study (see also Table S1). (b) Principal component analysis (PCA) of population differentiation. Individuals are color-coded corresponding to the sampling locations in (a). (c) Genetic clustering according to an analysis with admixture for  $k = 2$  (above; most likely number of putative ancestral populations) and  $k = 3$  (below). Genetic clusters are represented by different colors with vertical bars showing the ancestry proportion for each individual genotype.



called using GATK and were initially filtered to include only biallelic SNPs. Snpeff v4.3 (Cingolani et al., 2012) was used to annotate and predict the effects of genetic variants. We applied adjusted hard filters following GATK's recommendations (Van der Auwera et al., 2013):  $FS > 20$ ,  $MQRankSum < 12.5$ ,  $ReadPosRankSum < 2$ ,  $QD < 5$ ,  $MQ < 40$ .

Additionally, we investigated the utility of a high-quality reference genome for SNP discovery. Here, using the reference genome of the Japanese flounder (*Paralichthys olivaceus*, Shao et al., 2017), the closest relative of *P. platessa* for which a chromosome-anchored genome was available, to complement the linkage decay analysis at the chromosome level and investigate the heterogenetic differentiation at the putative inversions. We applied exactly the same filtering steps as described below, see Appendix S1 for more detailed information on the methods and an interpretation of the results.

### 2.3.3 | Identifying levels of neutral gene flow

To investigate the population structure of European plaice, variants within spanning deletions were removed using BCFtools v1.1 (Li, 2011). SNPs with a genotype quality score  $< 20$ , and read depth (DP)  $< 5$ , or  $DP > 20$  were excluded with VCFtools v0.1.13 (Danecek et al., 2011). We pruned SNPs so that no SNP was within 10 bp of another SNP. We excluded SNPs displaying a minor-allele count  $< 2$  across all populations and SNPs deviating from Hardy-Weinberg equilibrium with a  $p$ -value  $< .0001$  using Plink v1.9 (Purcell et al., 2007). Furthermore, variants with a minor-allele frequency  $< 0.03$  and more than 20% missing data per site were removed (hereafter "linked dataset"). Linkage disequilibrium (LD) decay was calculated across each of the 20 largest scaffolds. Additionally, LD decay was calculated across each of the 24 Japanese flounder chromosomes by mapping our raw reads directly towards the reference genome of the Japanese flounder (see Appendix S1). Pairwise  $r^2$  values were calculated between all SNPs per scaffold (and chromosomes) using VCFtools. Decay plots were created per population (ICE, BEL, BOR) by binning the distance between SNPs in increments of 1 kb and averaging the  $r^2$  values within each bin. SNPs with pairwise  $r^2$  values  $> .1$  within sliding windows of 50 kb SNPs were omitted with Plink (hereafter "pruned datasets").

Genetic differentiation between sampling locations was then inspected via PCA as implemented in the Eigensoft software v6.0.1 (Patterson et al., 2006). Ancestry fractions of individual plaice from all sampling locations were inferred based on a maximum likelihood estimation with the software Admixture v1.23 (Alexander et al., 2009) with  $k$  ranging from 2 to 10, where the most likely number of genetic ancestries was defined based on the lowest cross-validation error. Pairwise fixation indices ( $F_{ST}$ ) were calculated using the package Stampp v1.5.1 (Pembleton et al., 2013) in R v3.6.0 (R Development Core Team, 2017). Bootstrapping was performed to obtain 95% confidence intervals and  $p$ -values, which were adjusted

for multiple testing by applying sequential Bonferroni correction (Rice, 1989).

### 2.3.4 | Identifying local adaptation and genomic signatures of selection

Locus-specific  $F_{ST}$  estimates between sampling locations with a sufficient number of individuals (i.e., ICE, KAT, BEL, ARK, BOR, GDA) were calculated pairwise on the pruned dataset using VCFtools and loci were screened for highly differentiated SNPs (i.e., above the 99th percentile;  $F_{ST} \geq 0.4$ ). Based on this set of high  $F_{ST}$  SNPs we explored the genetic structure by conducting a PCA analysis using Eigensoft.

To detect genomic regions under recent selection and potential candidate genes involved in driving the heterogeneity or local adaptation between the genetic clusters, we conducted genome scans between the Icelandic (ICE) versus all the continental shelf (SHELF) samples and between the North Sea (NS) versus the Baltic Sea (BS) samples (see Table 1). Nucleotide diversity was calculated separately per scaffold in 20 kb sliding windows with 5 kb overlap using VCFtools for ICE, SHELF, NS (i.e., WNO, CNO, and ENO), and BS (BEL, ARK, BOR, and GDA). Allele frequencies per site were calculated with VCFtools. Allele frequency differences (AFD) between ICE/SHELF and NS/BS were computed and for sites with an AFD above the upper 5% percentile, all surrounding genes within 50 kb up- and downstream were extracted.

To discover regions with signs of recent positive selection among the populations, we calculated cross-population extended haplotype homozygosity scores (xp-EHH) using the R package rehh v3.0.0 (Gautier & Vitalis, 2012). Comparisons were performed between the pooled ICE/SHELF and NS/BS. Therefore, haplotypes were estimated for scaffolds  $> 500$  kb using the unpruned dataset with Beagle v4.1 and default settings (Browning & Browning, 2007), and xp-EHH scores were calculated for window sizes of 50 kb with 10 kb overlap. Since the ancestral allele was unknown, we set "polarized = F." Significant SNPs for each of the two comparisons were defined by selecting the top 0.05% of  $\log_{10}$   $p$ -values. Again, we extracted all surrounding genes within 50 kb upstream and downstream. GO terms were annotated by mapping gene names to the UniProtKB database (The UniProt Consortium, 2018). Furthermore, gene functions were categorized into three different categories (cellular components, molecular functions, and biological processes) with PANTHER (Mi et al., 2013) against the human GO database.

### 2.3.5 | Linking signals of local adaptation to potential structural variants

To investigate the existence of the two putative structural variants on chromosome 19 and chromosome 21 and other potential SVs and further infer their impact on genetic divergence

and substructuring of plaice in this geographical region, we here mapped our plaice scaffolds towards the reference genome with Minimap2 (Li, 2018) and calculated the linkage decay for these scaffolds. This was done to take full advantage of the higher mappability to our plaice reference genome, resulting in a higher number of SNPs, and as an add-on to the LD decay analyses and mapping our raw reads directly towards the Japanese flounder reference genome (see description above and Appendix S1). Additionally, we took a closer look at the scaffolds mapping inside or nearby the putative structural variants using VCFtools to calculate the heterozygosity per site over all populations and to perform pairwise calculations of the genetic differentiation along these two chromosomes, comparing ICE versus BOR, ICE versus BEL and BEL versus BOR in sliding windows of 200kb with steps of 50kb. The genetic variation between samples was inspected via PCA. These analyses were also conducted for other scaffolds showing the tendency of higher LD—found on other chromosomes.

For two representative scaffolds mapping to the putative structural variant on chromosome 19, individuals were genotyped and assigned to the respective haplogroup based on the PCA clusters (see Mérot, 2020 for concept). Pairwise differentiation ( $F_{ST}$ ) and nucleotide diversity ( $\pi$ ) between individuals representing the noninverted (HOM\_R) and the inverted (HOM\_I) haplotypes and linkage decay for individuals from the two haplogroups and from heterozygote individuals (HET) along scaffold\_302 and scaffold\_159 were calculated as described in the previous sections.

### 2.3.6 | Inferring rates of migration and past population history

To understand the direction of neutral gene flow among the plaice sampling locations we estimated relative migration rates between each of the locations with the R package *diveRsity* v1.9.90 (Keenan et al., 2013) using  $G_{ST}$  as a measure of genetic distance. We used the pruned set of SNPs and further removed the high  $F_{ST}$  SNPs potentially under selection leading to a set of 78,508 SNPs. The relative migration rates were calculated on a dataset down-sampled to five random individuals per location to ensure a balanced sampling scheme.

We inferred past effective population sizes ( $N_e$ ) separately for ICE and BS using the multiple sequentially Markovian coalescent (MSMC) method as implemented in the software MSMC v2.0.0 (Schiffels & Durbin, 2014). We prepared the input data by following the author's recommendations (<https://github.com/stschiff/msmc-tools>) with variant sites filtered for a mean depth of 7x. Scaffolds shorter than 500kb in length were excluded from the analysis. Variant sites falling within repeat regions were masked to include only sites at which the mapping is of high quality.

We performed five MSMC runs with default settings for individuals from ICE or BS, respectively, where each run included three randomly selected individuals. Since the mutation rate is unknown for *P. platessa*, we applied three different mutation rates (low:

$\mu = 2 \times 10^{-9}$ , medium:  $\mu = 1.5 \times 10^{-8}$ , high:  $\mu = 3.7 \times 10^{-8}$ ) to cover the known spectrum of mutation rates in marine teleosts (Feng et al., 2017; Liu et al., 2016). The generation time for plaice was set to 3 years (Rijnsdorp, 1989).

## 3 | RESULTS

### 3.1 | De novo assembly and annotation of a plaice reference genome

Short- and long-read sequencing yielded 73 and 20Gb of raw data for the reference individual (see Table S1 for detailed sample characteristics), respectively. Reads were assembled into 4512 scaffolds with a total length of 583,260,484 bases. The longest scaffold was 1.4 Mb in size and scaffold N50 was 0.28 Mb (Table S2). Based on 17-mer analysis, the estimated genome size was 587 Mb (Figure S1). The quality of the genome assembly was confirmed by successfully identifying 94.7% of complete BUSCOs (Table S3). Only 2.5% of the BUSCO genes were not found in the final assembly. Repetitive sequences make up to 8.8% of the plaice genome (Table S4). These are grouped into interspersed repeats (5.5%), small RNA (0.07%), satellites (0.11%), simple repeats (2.71%), and low-complexity motifs (0.35%).

RNA-sequencing initially produced 247Gb of data, which were used to construct a de novo plaice transcriptome. The final transcriptome was 617Mb in size, contained 800,892 Trinity genes, and had an N50 contig length of 785bp based on all transcript contigs. Both genome and transcriptome information were integrated for annotation. Finally, a total of 49,037 gene models were annotated. Running BUSCO on the final plaice annotation revealed 88.7% of the searched orthologs being present and complete, of which 84.7% and 4.0% were single-copy or duplicated orthologs, respectively. About 7.2% of the BUSCO genes were fragmented and 4.1% of the genes were missing in the annotation.

### 3.2 | Genome sequencing and variant identification

The whole genomes of 63 plaice individuals were generated to a mean coverage of 8x ( $\pm$  SD 2x), trimmed, and aligned against the plaice reference genome with an average alignment rate of 98% ( $\pm$  SD 2%) with 92% ( $\pm$  SD 3%) of the mapped reads properly paired (Table S5). Variant calling using the GATK pipeline resulted in the discovery of 23,278,376 biallelic Single Nucleotide Polymorphisms (SNPs). These variants contributed to 42,693,312 predicted effects. For intergenic regions, 10,022,005 predicted effects (23.47%) were identified (Tables S6 and S7). Many SNPs were located upstream (22.81%), or downstream (22.66%) of genes, or were introns (27.18%) (Table S6). Among SNPs located in protein-coding regions, 808,682 (1.89%) were synonymous variants while 669,227 (1.56%) were nonsynonymous variants (Table S7). Additionally, we used the reference genome of the Japanese flounder to investigate the utility

of a cross-species reference genome for mapping and SNP discovery (Appendix S1, including Table S8 for full details).

### 3.3 | Population differentiation of European plaice

In general, we detected a rapidly decreasing linkage decay when analyzing the 20 longest plaice scaffolds for each population (ICE, BEL, BOR) (Figure S2) with LD being higher for BOR than for ICE and BEL.

The population differentiation among all samples of European plaice was investigated (see Table 1, Figure 1a) based on 83,011 genome-wide distributed pruned SNPs (236,458 SNPs prior to pruning). The level of gene flow was first inferred without prior knowledge of the population structure using a principal component analysis (PCA). Two distinct clusters were observed with the Icelandic samples clearly separated from the continental shelf samples (the first axis explains 2.73% of the observed variation, Figure 1b). The second axis, explaining 2.02% of the variation, revealed high individual variation within clusters and low population structuring among the continental shelf plaice samples, i.e., with no apparent differentiation between the North Sea (i.e., WNO, CNO, ENO) and the Baltic Sea samples (BEL, ARK, BOR, GDA).

We further examined the population genetic structure through admixture analysis with different numbers of potential ancestries ranging from  $k = 2$  to 10. We observed a minimum cross-validation (CV) error at  $k = 2$ , at which all individuals from Iceland were distinguished from the remaining individuals (Figure 1c). The individuals from the North Sea, Kattegat, and the Baltic Sea were mostly composed of the same ancestry, although some shared a minor ancestral portion with the Icelandic samples. At  $k = 3$ , we observed a third ancestry component for single individuals from the North Sea, the Kattegat, and the western and the eastern Baltic, although a general pattern was not apparent.

The degree of genetic differentiation measured as  $F_{ST}$  value was hence highest between the Icelandic samples and all the other samples ( $F_{ST} = 0.02$ ,  $p < .001$ , Table 2). Comparisons among European continental shelf populations revealed lower  $F_{ST}$  values and most

comparisons were nonsignificant. However, we detected subtle levels of genetic substructure among North Sea samples and between samples from the North Sea and the Baltic Sea (Table 2). Within the Baltic Sea, we observed low levels of genetic differentiation among all four sampling locations.

Genome-wide estimates of the overall nucleotide diversity  $\pi$  were extremely low (Table S9). Nucleotide diversity was highest in the Icelandic population and lowest for plaice in the Baltic Sea.

### 3.4 | Genomic regions of divergence

Even if the LD measurements are on the lower side, it should be mentioned that we did observe some variation between the scaffolds tested. The complementary LD analysis by mapping the plaice reads directly towards the chromosome-anchored genome of the Japanese flounder confirmed a similar pattern of linkage decay (Figure S3). However, it should be noted, that even if values are very low, we see a tendency towards a slight increase in LD on chromosome 19 where one of the previous putative SVs for plaice has been detected (Le Moan et al., 2021). Of the two putative SVs reported by Le Moan et al. (2021), we only found evidence for increased LD combined with genomic divergence within a larger genomic region on chromosome 19 (Figure S4), while there was no evidence for higher LD and/or genomic differentiation on chromosome 21 or within the putative structural variant on chromosome 21 (Figure S5). However, we did detect other scaffolds with high genomic divergence coupled to high LD that mapped to chromosomes 4, 5, and 6 on the Japanese flounder genome (see Figures S6–S9a–l).

The pattern of population separation seen for this region is showing subtle allele frequency variation from the Baltic Sea into the North Sea (Figure 2a). It should, however, be noted, that within the putative SV on chromosome 19 (SV19) where we localized scaffolds/regions with a higher LD and higher degree of differentiation (Figures 2b–i), we also detected interspersed scaffolds/regions displaying low LD and low differentiation (Figures S4, S10 and S11a–h). Additionally, plaice scaffolds mapping downstream

TABLE 2 Estimates of genome-wide differentiation between samples of European plaice based on 83,011 SNPs

	ICE	WNO	CNO	ENO	KAT	BEL	ARK	BOR	GDA
ICE		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
WNO	<b>0.019</b>		1	1	1	0.684	1	<0.001	<0.001
CNO	<b>0.021</b>	0.002		1	<0.001	1	<0.001	<0.001	<0.001
ENO	<b>0.015</b>	−0.003	0.000		1	1	1	1	1
KAT	<b>0.019</b>	0.001	<b>0.003</b>	0.000		1	1	1	0.072
BEL	<b>0.019</b>	0.001	0.001	−0.004	0.000		0.072	0.648	1
ARK	<b>0.020</b>	0.001	<b>0.002</b>	−0.003	0.001	0.001		0.072	1
BOR	<b>0.019</b>	<b>0.001</b>	<b>0.004</b>	−0.001	0.001	0.001	0.001		1
GDA	<b>0.018</b>	<b>0.002</b>	<b>0.005</b>	−0.001	0.002	0.000	0.000	0.000	

Note: Pairwise  $F_{ST}$  estimates are presented below the diagonal and corresponding  $p$ -values above the diagonal. Significant estimates are indicated in bold.



of the structural variant, all demonstrated low LD and no clear genomic separation between locations and/or populations (see Figure S11i–l).

In addition to the allele frequency differentiation along the environmental gradient seen for the potential inversion on chromosome 19, we observed distinct genomic loci of greater differentiation. Notably, 4503 singular loci out of 83,011 pruned SNPs were spotted across all scaffolds with pronounced levels of pairwise differentiation ( $F_{ST}$  values  $\geq 0.4$ ; Table S10, Figure S12). These outlier loci were also distributed all over the genome when scaffolds were mapped against the Japanese flounder genome (figure not shown). Only considering these high  $F_{ST}$  outlier SNPs, PCA analysis revealed a population substructure that was mostly in congruence with the geographic origin of the individuals (Figure 3a). Considering PC1 versus PC2, the Icelandic and the Kattegat plaice formed isolated clusters, whereas the Baltic Sea samples were not distinguishable. Considering, PC2 versus PC3, samples from the eastern Baltic Sea (GDA) separated from the Baltic Sea cluster (Figure 3b).

Moreover, analyses uncovering genome-wide signatures of divergent selection and recent selective sweeps and thus, identification of genes that could be functionally associated with adaptation to the different environments of the North Sea and the Baltic Sea were conducted. Here, different environmental conditions experienced by plaice across both, small and large spatial scales were considered. We detected 486 and 358 distinct loci with increased shifts in allele frequency between samples from Icelandic (ICE) and the European continental shelf (SHELF) and from the North Sea (NS) and the Baltic Sea (BS), respectively. Within a distance of 50 kb surrounding the focal SNP in both directions, 167 and 116 annotated genes were identified, respectively (Table S11).

Genomic regions of divergence were further investigated by identifying signatures of recent selective sweeps. We calculated the cross-population extended haplotype homozygosity (xp-EHH) statistic, which compares the integrated extended haplotype homozygosity profiles between two populations at the same locus and points to alleles that have swept to near-fixation within one of the two populations (Sabeti et al., 2007). Haplotypes were inferred only for the 184 largest scaffolds (i.e., >500 kb), representing 22% of the genome. A total of 232 genomic loci on 44 scaffolds were noted as outliers between ICE versus SHELF and 235 loci on 67 scaffolds for NS versus BS (Table S11, Figure S13). Of these, 54 and 55 SNPs were located in annotated genes, respectively. For ICE versus SHELF, 112 outlier SNPs were located in exons and 20 SNPs were nonsynonymous coding substitutions (Table S11, Figure 4a). For NS versus BS, 110 SNPs were located in exon regions and 36 SNPs were nonsynonymous coding substitutions (Figure 4b). Gene ontology (GO) analysis revealed that many of the genes identified were functionally involved in cellular, metabolic or regulatory processes (Table S11, Figure S14). In terms of biological processes, genes were related to adaptation to low salinity (e.g., PKN1, PKD2), reproduction (e.g., ANTXR1, JAG2), and immune system response (e.g., GPER2, POLR3E).

### 3.5 | Direction of gene flow and demographic history

Elucidating directional patterns of genetic connectivity between all pairs of samples from each location revealed moderate to high relative migration rates across the European continental shelf (Figure S15a) supporting the findings regarding the similarity of populations from the PCA analysis (Figure 1b). Highest exchange rates were observed among the Baltic Sea populations. Removing the Icelandic samples revealed a similar pattern among the North Sea and the Baltic Sea populations (Figure S15b). Again, the highest rates of gene flow could be observed within the Baltic Sea.

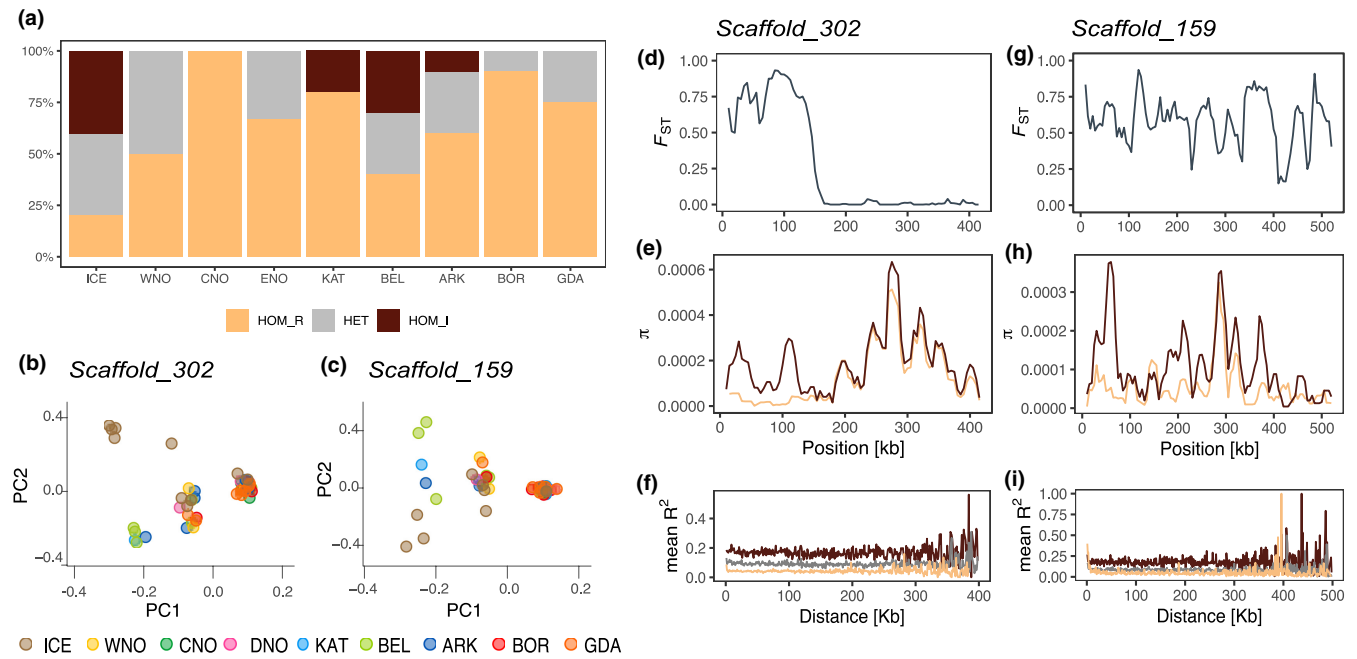
Estimates of historical population size ( $N_e$ ) and demographic trends were inferred using MSMC2 analyses for three different mutation rates and a fixed generation time of 3 years (Figure 5, Figures S16 and S17). After experiencing sharp reductions in  $N_e$  during the Late Pleistocene, both populations gradually expanded in size towards recent times, with the Icelandic populations showing lower levels of  $N_e$  compared with the ancestral Baltic population since the Last Glacial Maximum (LGM). Replicate runs with three randomly chosen individuals yielded consistent  $N_e$  estimates for the Icelandic and the inner Baltic Sea past history. With an assumed mutation rate of  $1.5 \times 10^{-8}$  mutations/site/generation,  $N_e$  fluctuated between ~300,000 and ~100,000 for the ancestral Baltic Sea and between ~400,000 and ~50,000 for the Icelandic plaice, respectively.

## 4 | DISCUSSION

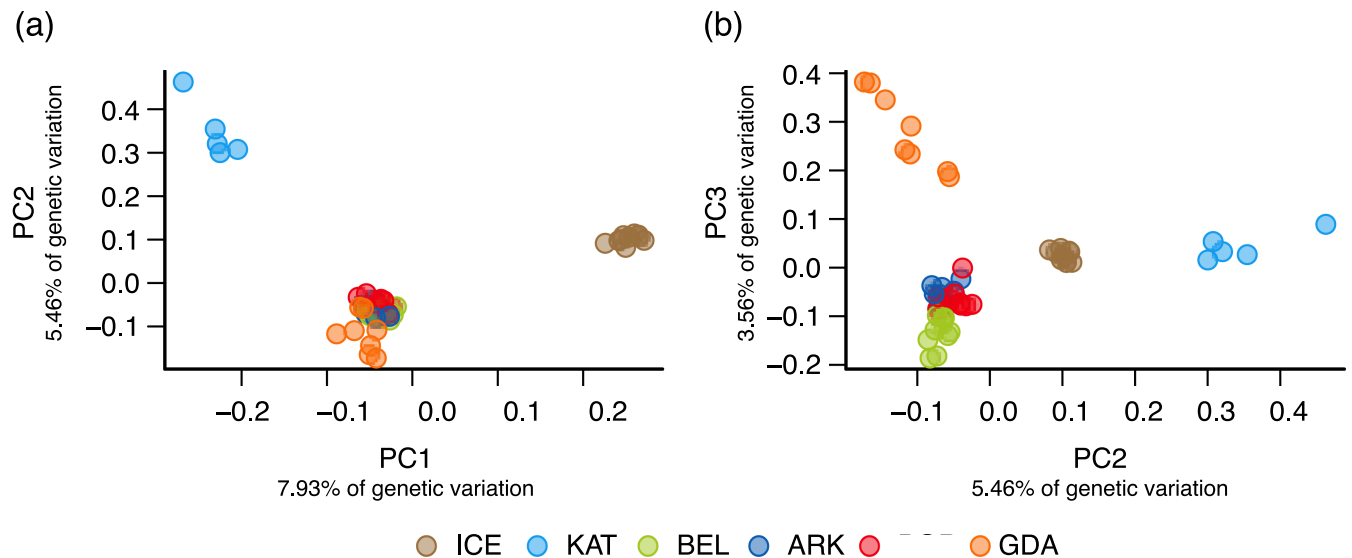
By generating the first annotated reference genome assembly for the European plaice combined with whole-genome sequencing of specimens (to approx. 8x coverage) across large parts of its geographical distribution, we uncovered a population substructure that was linked to genomic signatures of local adaptation despite high gene flow within the European continental shelf. This population substructure at smaller geographical scales was only revealed with SNPs potentially under strong selection, i.e., defined as high  $F_{ST}$  outliers. Neutral markers, however, pointed at the clear separation of the Icelandic population from samples collected along the gradient of the European continental shelf as earlier studies have already revealed. Moreover, by using the whole-genome information we were able to unfold the demographic history of plaice and detected that the plaice populations investigated experienced a major population bottleneck during the LGM following similar past demographic trajectories.

### 4.1 | Genome assembly of the European plaice

We obtained a plaice draft genome assembly of high completeness, covering 99% of the k-mer-based estimated genome size. The size of plaice genome assembly (583 Mb) falls within the expected



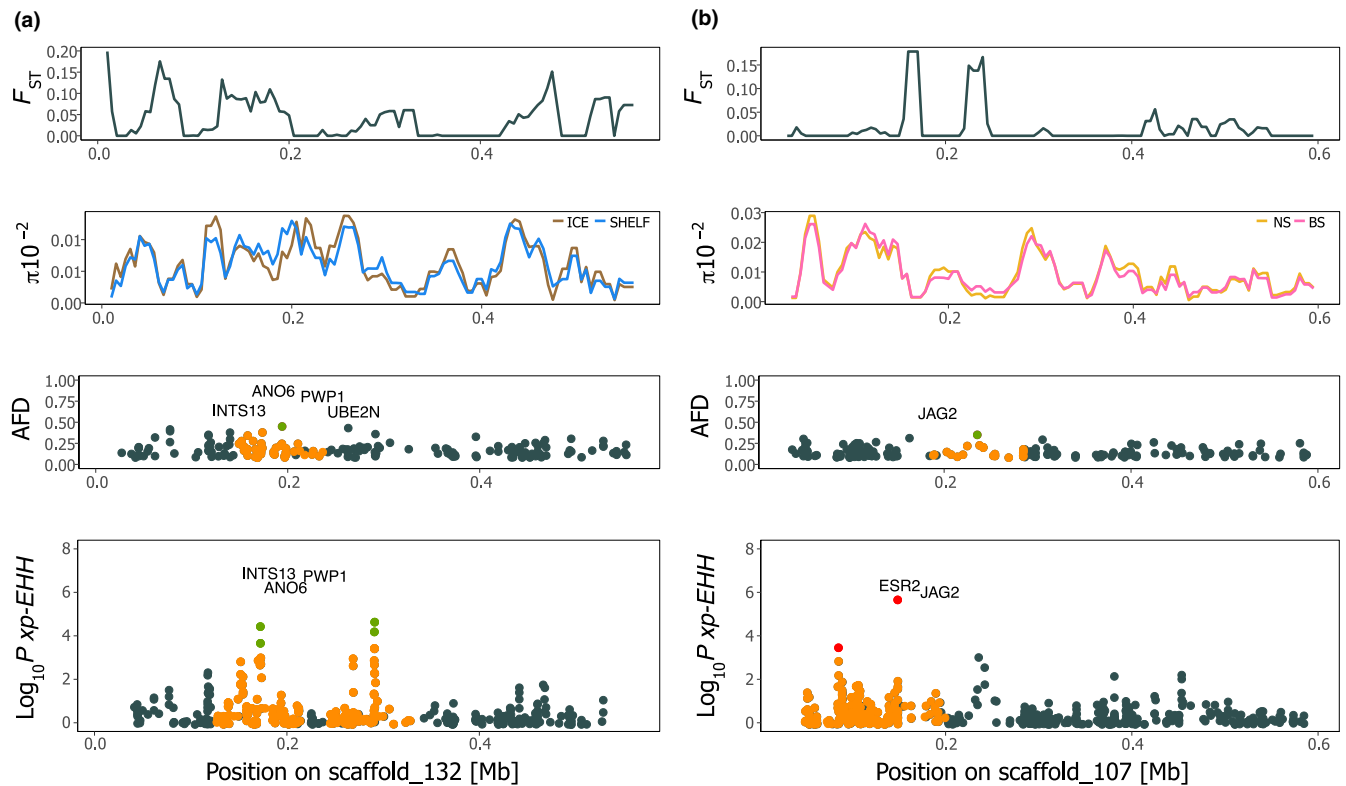
**FIGURE 2** Analysis of linked regions in European plaice for scaffolds mapping to a putative structural variant located in the corresponding chromosome 19 of the Japanese flounder reference genome. (a) Allele frequencies of the putative structural variant based on genotype clustering in (b) and (c) and Figure S11 at different locations following the salinity gradient from the North Sea into the Baltic Sea. The proportion of individuals being homozygous for the presumed collinear allele is shown in yellow (HOM\_R), and the proportions of individuals heterozygous or homozygous for the rearranged allele are shown in gray (HET) and dark red (HOM\_I), respectively. (b, d–f) Principal component analysis of population differentiation, pairwise differentiation ( $F_{ST}$ ), and nucleotide diversity ( $\pi$ ) between individuals representing the noninverted (HOM\_R) and the inverted (HOM\_I) haplotypes and linkage decay for individuals from the two haplogroups and from heterozygote individuals (HET) along scaffold\_302 and scaffold\_159 (c, g–i).



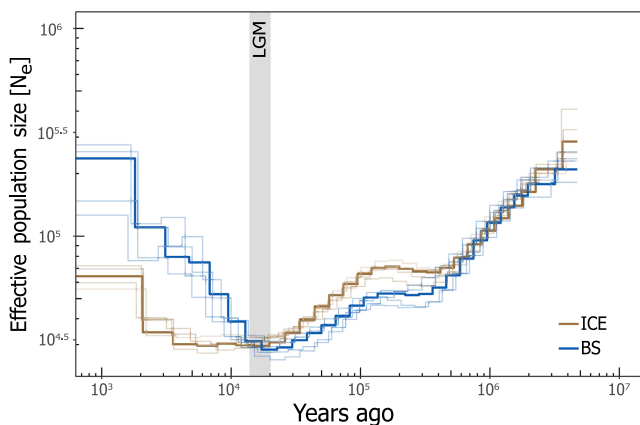
**FIGURE 3** Analysis of 4503 highly differentiated SNPs ( $F_{ST} \geq 0.4$ ) based on principal component analysis (PCA) of 53 plaice samples for (a) PC1, PC2 and (b) PC2, PC3.

range of flatfish genomes (Chen et al., 2014; Figueras et al., 2016; Shao et al., 2017). However, the estimated genome size is smaller than the genome size calculated with fluorometric methods for Pleuronectidae (Hinegardner & Rosen, 1972). Genomic completeness was further estimated by gene content and gave a high

percentage (94.7%) of core gene presence. Compared with other flatfish assemblies the repetitive content in plaice is slightly higher (Chen et al., 2014; Figueras et al., 2016). Along with the repeat content, the level of heterozygosity and low-complexity regions are known factors impeding the contiguity of a genome assembly (Alkan



**FIGURE 4** Closer investigation of the genomic differentiation. (a) Genomic region of divergence between individuals from ICE and SHELF (see Table 1) for scaffold\_132. Pairwise fixation index ( $F_{ST}$ ) between samples and nucleotide diversity ( $\pi$ ) calculated in sliding windows of 20 kb with 5 kb overlap size. Allele frequency difference (AFD) and cross-population extended haplotype homozygosity (xp-EHH) were calculated between samples. Outlier loci above the 99.95th percentile are colored in green, if substitution is synonymous, or red, if substitution is nonsynonymous. Loci within 50 kb distance to the outlier SNP are denoted in orange. (b) Genomic differentiation between NS and BS for scaffold\_107.



**FIGURE 5** Inference of effective population size ( $N_e$ ) over time. Past population history was inferred based on five independent runs of MSMC2 for three randomly chosen individuals from the Icelandic (ICE) population and the Baltic (BS) population (Table 1). Results are shown for a medium mutation rate of  $\mu = 1.5 \times 10^{-8}$  mutations/site/generation and a generation time of 3 years. The gray bar visualizes the duration of the last glacial maximum (LGM).

et al., 2011). Future studies could thus aim for anchoring the 4512 scaffolds that we have detected into the expected number of 48 chromosomes ( $2n$ ) revealed by the karyotypic analysis in this species

(Fan & Fox, 1991) to gain further insight into the genomic architecture of plaice.

Annotation with the Funannotate pipeline initially yielded 49,037 gene models. This is comparatively more than for other flatfish genomes (*Cynoglossus semilaevis*: 21,516 [Chen et al., 2014]; *Paralichthys olivaceus*: 21,787 [Shao et al., 2017]; *Scophthalmus maximus*: 22,751 [Figueras et al., 2016]) for whose annotation different pipelines have been applied. Since the draft assembly is fragmented, it is likely that some of the genes are broken up among several scaffolds leading to an inflation of the overall gene count if some genes are located on different scaffolds (Salzberg, 2019).

## 4.2 | Extensive gene flow across the continental shelf

Using whole-genome sequencing data based on the neutral loci we confirmed the already documented genetic/genomic divergence of the Icelandic and the European continental shelf plaice population and the large-scale genetic homogeneity across the shelf area (Le Moan et al., 2021; Was et al., 2010). Bathymetry and hydrography have been suggested to act as the main drivers in the formation of the two genetic clusters promoting the reproductive separation of

shelf and off-shelf populations (Hoarau et al., 2004; Was et al., 2010). Moreover, our neutral SNP set uncovered high levels of gene flow between specimens sampled at the North Sea versus the Baltic Sea spawning grounds. However, based on pairwise  $F_{ST}$  analyses on the pruned/neutral marker set subtle differentiation between plaice from CNO, WNO versus the eastern Baltic Sea samples (ARK, BOR, GDA) was detected (see Table 2). Though, it should be noted that the low number of samples (3–10 individuals per sampling location) used in this study could potentially bias the allele frequency distributions. However, since our results are in line with the previous findings (Hoarau et al., 2004; Was et al., 2010), we propose that several mechanisms including (i) historical events and (ii) selection are important drivers of the contemporary genetic structure of European plaice at different geographical scales.

Congruently, Le Moan et al. (2021) and our study observed the greatest genetic differentiation between plaice from Iceland and the European continental shelf using modern sequencing techniques. Additionally, Le Moan et al. (2021), discovered two putative SVs displaying allele frequency differentiation following the salinity gradient from the North Sea into the Baltic Sea. With the generation of a draft reference genome assembly combined with whole-genome sequencing data—using a smaller sample set compared with Le Moan et al. (2021)—our results support the existence of a larger genomic region with higher LD on chromosome 19, which shows a pattern of genetic divergence resembling an inversion (see Figure 2b–i, Figure S11a–d), displaying three distinct clusters for the homokaryotypes, the heterokaryotypes, and the homokaryotypes of the alternative arrangement (Mérot, 2020). Moreover, the fact that high LD scaffolds within the putative SV19 were interspersed with regions of low LD and low genetic differentiation (see Figures S4 and S11) could be caused by (i) discrepancy at the micro-synteny level between the Japanese flounder and the European plaice on chromosome 19, leading to this mosaic pattern of high and low LD regions when mapping towards the Japanese flounder genome, or it could be indicative of (ii) misassembled genomic regions in plaice or (iii) the existence of multiple smaller regions of repressed recombination rather than a single huge block (Pettersson et al., 2019). Alternatively, the alternating pattern of regions with high and low LD is potentially caused by double crossovers (Kennington & Hoffmann, 2013), which have been observed for other species (Kennington & Hoffmann, 2013; Pettersson et al., 2019). Nevertheless, the potential inversion on chromosome 19 is seemingly under balancing selection since all karyotypes were evident at all locations (and subpopulations)—with subtle allele frequency differences—indicative of maintaining all variants within the populations investigated in this study. We further identified additional larger haplotype blocks displaying genomic divergence on chromosomes 4, 5, and 6 pointing at further linked genomic regions that are under selection and thus important for local adaptation. However, a chromosome-anchored genome of the European plaice could help to improve our understanding of the genomic architecture and patterns of synteny in the teleost family Pleuronectidae.

For many marine fish species, genetic population differentiation is related to their egg and larval dispersal potential, population

history, and adult migration (Waples, 1998). Hydrodynamic processes and ocean currents play a major role in linking geographically distant populations by facilitating dispersal during early life stages (White et al., 2010). High fecundity, a long pelagic larval duration, and larval behavior offer great dispersal capacities over hundreds of kilometers by passive transport in plaice (Volckaert, 2013) and by such leading to a high degree of gene flow—as demonstrated by our neutral SNP set—between locations within the European continental shelf at various spatial scales. For instance, in the macrotidal North Sea, currents are important for the dispersion of plaice eggs and larvae from the commonly used spawning grounds in the southern North Sea towards the nursery areas following the direction of the general circulation pattern (Cushing, 1990; Erftemeijer et al., 2009; Hunter et al., 2004) and, thus, enable mixing and genetic homogeneity. In the Skagerrak/Kattegat area larval stages from the North Sea populations passively drifted by currents are retained and thought to return to the North Sea to spawn (Ulrich et al., 2013, 2017), hinting at possible homing behavior in this species. However, the Kattegat samples were genetically similar to the Eastern North Sea and the Baltic Sea samples pointing at imperfect homing and/or the existence of migrants between these sites, respectively. In the Baltic Sea, the exchange of eggs between adjacent basins is suggested to be minor given the species' egg density and that floating eggs are restricted to the more saline bottom layers in each basin (Petereit et al., 2014), potentially resulting in reduced gene flow between spawning grounds in neighboring basins. However, regular major Baltic inflows and periods of strong easterly winds (Mohrholz, 2018) may cause pulsed mixing of eggs and larvae across the sill-connected adjacent basins of the western and the eastern Baltic Sea. On the other hand, subtle differentiation was noticed between the westernmost North Sea (WNO) and the easternmost Baltic Sea sampling sites (GDA). The subtle pattern might indicate reduced gene flow impacted by multiple evolutionary mechanisms such as isolation-by-distance, different selection pressures, and colonization history.

### 4.3 | Signatures of local adaptation and putative population substructuring

Although no significant population structure was identified among plaice specimens caught throughout the steep salinity gradient from the Baltic Sea following the shallow straits of the archipelago into the North Sea area based on neutral SNP markers, different results were obtained with highly differentiated SNPs. The indication of a potential population substructuring, i.e., genetic divergence based on the high  $F_{ST}$  outliers might be explained by strong divergent selection pressures at the various locations and thus point to signatures of local adaptation.

For the North Sea samples, pairwise comparisons of genomic differentiation were omitted due to the limited resolution of allele frequency estimates caused by low sample sizes. Given the subtle signs of genomic differentiation detected among these populations, future studies including more samples from these locations could

aim at elucidating the genomic divergence of these populations. Further, the remaining samples from the continental shelf form three genetic clusters (KAT, BEL/ARK/BOR, GDA) based on the high  $F_{ST}$  outliers roughly corresponding to their geographic origin along the environmental gradient between the North Sea and Baltic Sea. The observed population substructuring within the Baltic Sea area highlights that even within smaller spatial scales, the plaice encounters distinct selection pressures and thus, likely mirrors local adaptations to life in the brackish water environment with deep basins. However, the differentiated GDA might also reflect the effects of living on the edge of the species' distributional range and genetic drift (Johannesson & André, 2006).

Throughout its distribution range, plaice experiences diverse physical environments and variable abiotic conditions, such as salinity, temperature, oxygen content, and tidal currents (in the North Sea), which can ultimately lead to adaptive genomic divergence between locations. Apart from single highly differentiated loci, several genomic regions were strongly differentiated among specimens from the North Sea (WNO, CNO, ENO) and the Baltic Sea (BEL, ARK, BOR, GDA). We detected loci that stood out above the 99.95th percentile in the global comparisons considering either AFD or xp-EHH (Figure S13) indicating recent selective sweeps. Many of them contained at least one candidate gene with a functional connection to either reproduction, immune system, or adaptation to environmental conditions. For example, on scaffold\_132 we detected a region possibly under divergent selection between the continental shelf and off-shelf populations containing ANO6 (Figure 4a). The gene encodes a transmembrane protein that has been shown to be involved in angiogenesis in zebrafish, a process that is triggered by low-oxygen conditions (Delcourt et al., 2015). Tolerance against variation in oxygen levels and hypoxia might play an important role in ecological adaptation in plaice and in other demersal species, as well (Barrio et al., 2016), particularly for those living in the Baltic Sea.

On scaffold\_107 a nonsynonymous substitution was located in the estrogen receptor ESR2 (Figure 4b), which is suggested to play a vital role during maturation (Vidal et al., 2000). Estrogen receptors have been proposed as candidate genes of selection in Atlantic cod (Clucas et al., 2019) and in three-spined stickleback (*Gasterosteus aculeatus*; Shimada et al., 2010). In close proximity to ESR2, we identified JAG2, which is involved in spermatogenesis (Hayashi et al., 2001), indicating that the entire genomic region is under strong directional selection. Additionally, we detected multiple genes throughout the genome related to sperm motility (e.g., GAPDH2, INTS13, DDX6), and some were already suggested to be of importance for efficient reproduction in brackish water environments (Berg et al., 2015). Furthermore, some of the genomic regions of divergence were coupled to specific genes important for immunosystem functioning and signaling pathways (e.g., OPRK1 on scaffold\_13, DENND1B on scaffold\_140).

Our approach to identify outlier loci was rather conservative, i.e., based on the empirical distribution of allele frequencies and haplotype statistics, but a commonly used approach to detect loci that are divergent among populations (Cuevas et al., 2021; Kusakabe

et al., 2017; Ravinet et al., 2016). Selection scans such as those applied in this study have the highest power when selection is strong and the genetic architecture underlying a trait under selection is simple (Stephan, 2016). We further note the limited number of samples used for our divergence analysis and caution is still warranted when interpreting the data, but our study takes advantage of genome-wide variation providing great power in resolving signatures of adaptation.

#### 4.4 | Reconstruction of plaice population history

Glaciations during the Pleistocene have shaped the contemporary geographical distribution of populations and the degree of genetic connectivity among them (Jenkins et al., 2018). To further illuminate the origin of the population substructure indicated by the outlier SNPs, we reconstructed the demographic history of plaice from individuals from ICE and BS using Markovian coalescent analyses since high gene flow could have masked past history. It should be noted that the algorithm used here, has some limitations when it comes to low-coverage data (e.g., as in our dataset). The method has been shown to recover essential features of a demographic history even with lower coverage datasets, however, the estimates of the effective population size might not be accurate (Mather et al., 2020). Assuming a generation time of 3 years and a medium mutation rate of  $1.5 \times 10^{-8}$  mutations/site/generation, the results indicate that both populations (ICE vs. BS) have already split more than 200,000 years ago during the mid-Pleistocene, hinting at two glacial refugia, which has also been suggested by Le Moan et al. (2021). The ancestral population of the Icelandic plaice was probably located close to the Rockall Plateau as hypothesized for Atlantic cod (Pampoulie et al., 2008). The ancestral population of the Baltic plaice likely survived in a glacial refugium located in the northern parts of the North Sea (Kettle et al., 2010) and gradually expanded southwards following the retreat of the glacial coverage during the late Weichselian, a process that was reflected by a decreasing genetic diversity towards the Baltic Sea (this study). Additionally, the lack of population differentiation between the North Sea and the Baltic Sea and the low pairwise differences between sampling sites suggests a substantial genetic exchange between the source population and the Baltic Sea colonizers during past colonization events. After successful colonization, the ancestral Baltic Sea population has rapidly expanded and reached high numbers of contemporary effective breeders. A large effective population size is often correlated with a high evolutionary potential to respond to changing environments (Ellegren & Galtier, 2016). However, a genetic separation may stay unnoticed for a long time since differences in allele frequencies will only slowly accumulate and the signal of population genetic differentiation is likely to be obscured at the level of neutral genomic variation (Hauser & Carvalho, 2008; Waples, 1998). Hence, the combination of high gene flow among European continental shelf plaice, high effective population size, and the relatively young age of the Baltic Sea could have prevented plaice populations living in the North Sea and the



Baltic Sea from establishing a distinct population structure at the neutral markers.

## 5 | CONCLUSION

Here, genomic markers have unfolded the relative importance of demographic history versus adaptive divergence in shaping the population genetic structure of a marine teleost. We report on the emergence of a putative population substructuring in European plaice whose detection is confounded by high levels of gene flow, a common colonization history of European continental shelf populations, and a high effective population size. While neutral markers revealed the reproductively separated Icelandic and European continental shelf population, outlier loci pointed at a local population structure at smaller geographical scales indicating spatially heterogeneous selection and not related to the potential inversion on chromosome 19. Our findings suggest the importance of having access to a reference genome and high-resolution markers with full-genome coverage to elucidate how the genomic landscape in plaice is shaped by neutral and adaptive genetic variation and to better understand the evolutionary history of populations. Interestingly, this study provides evidence that adaptive divergence between subpopulations uncoupled from structural variants can be maintained in high gene flow species. Throughout the genome, we have identified loci that are putatively under directional selection and that are therefore of potential importance for adaptation. Hence, our findings have implications for the fisheries management of the European plaice in order to sustain its high adaptive capacity in the North Sea and the Baltic Sea.

### AUTHOR CONTRIBUTIONS

**Peggy Weist:** Formal analysis (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Sissel Jentoft:** Conceptualization (supporting); methodology (supporting); resources (supporting); supervision (supporting); writing – original draft (supporting). **Ole K. Tørresen:** Formal analysis (supporting); methodology (supporting); resources (supporting); writing – original draft (supporting). **Franziska M. Schade:** Methodology (supporting); resources (supporting); writing – original draft (supporting). **Christophe Pampoulie:** Resources (supporting); writing – original draft (supporting). **Uwe Krumme:** Conceptualization (equal); funding acquisition (equal); methodology (supporting); project administration (equal); resources (supporting); writing – original draft (supporting). **Reinhold Hanel:** Conceptualization (equal); funding acquisition (equal); project administration (equal); supervision (lead); writing – original draft (supporting).

### ACKNOWLEDGMENTS

The authors thank three anonymous reviewers for their helpful comments on an earlier version of the manuscript. The authors thank the technical staff at the Thünen-Institute of Baltic Sea Fisheries for assistance in sampling and sample processing as well as Vivian Kuhlenkamp for technical assistance during the DNA quality assessment. PacBio and Illumina library creation and high-throughput

sequencing were carried out at the Norwegian Sequencing Centre (NSC; <https://www.sequencing.uio.no>), University of Oslo, Norway. Computational analyses were performed on the Abel Cluster owned by the UiO and the Norwegian metacenter for High Performance Computing (NOTUR) and operated by the UiO Department for Research Computing (<https://www.hpc.uio.no>). The project, P.W., and F.M.S. were partly funded by the European Maritime and Fisheries Fund (EMFF) of the European Union (EU) under the Data Collection Framework (DCF, Regulation 2017/1004 of the European Parliament and the Council).

### DATA AVAILABILITY STATEMENT

All raw sequences have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under the umbrella project with the accession number PRJEB36682 (<https://www.ebi.ac.uk/ena/data/view/PRJEB36682>). Individual accession numbers for samples are provided in Table S1. The ENA accession numbers to the annotated reference genome assembly and to the transcriptome assembly are ERZ1305001 and ERZ1300624, respectively.

### ORCID

Peggy Weist  <https://orcid.org/0000-0002-3715-6936>

Sissel Jentoft  <https://orcid.org/0000-0001-8707-531X>

### REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Alkan, C., Sajjadian, S., & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1), 61–65. <https://doi.org/10.1038/nmeth.1527>
- Aro, E. (1989). A review of fish migration patterns in the Baltic. *Rapports et Procès-Verbaux des Réunions*, 190, 72–96. <http://en.scientificcommons.org/26047672>
- Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(1), 11. <https://doi.org/10.1186/s13100-015-0041-9>
- Barrio, A. M., Lamichaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., Dainat, J., Ekman, D., Höppner, M., Jern, P., Martin, M., Nystedt, B., Liu, X., Chen, W., Liang, X., Shi, C., Fu, Y., Ma, K., Zhan, X., ... Andersson, L. (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife*, 5, e12081. <https://doi.org/10.7554/eLife.12081>
- Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., Jakobsen, K. S., Johannesson, K., Jorde, P. E., Knutsen, H., Moksnes, P. O., Star, B., Stenseth, N. C., Svedäng, H., Jentoft, S., & André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, 26(17), 4452–4466. <https://doi.org/10.1111/mec.14207>
- Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., Jakobsen, K. S., & André, C. (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). *Genome Biology and Evolution*, 7(6), 1644–1663. <https://doi.org/10.1093/gbe/evv093>
- Björck, S. (1995). A review of the history of the Baltic Sea, 13.0–8.0 ka BP. *Quaternary International*, 27, 19–40. [https://doi.org/10.1016/1040-6182\(94\)00057-C](https://doi.org/10.1016/1040-6182(94)00057-C)
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, 81(5), 1084–1097. <https://doi.org/10.1086/521987>
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J. N., Hong, Y., Li, Q., Sha, Z., Zhou, H., Xie, M., Yu, Q., Liu, Y., Xiang, H., Wang, N., ... Wang, J. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*, 46(3), 253–260. <https://doi.org/10.1038/ng.2890>
- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, 10(6), 563–569. <https://doi.org/10.1038/nmeth.2474>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications*, 12(10), 1971–1987. <https://doi.org/10.1111/eva.12861>
- Cowen, R. K., Lwiza, K. M. M., Sponaugle, S., Paris, C. B., & Olson, D. B. (2000). Connectivity of marine populations: Open or closed? *Science*, 287(5454), 857–859. <https://doi.org/10.1126/science.287.5454.857>
- Cuevas, A., Ravinet, M., Saetre, G. P., & Eroukhanoff, F. (2021). Intraspecific genomic variation and local adaptation in a young hybrid species. *Molecular Ecology*, 30(3), 791–809. <https://doi.org/10.1111/mec.15760>
- Cushing, D. (1990). Hydrographic containment of a spawning group of plaice in the Southern Bight of the North Sea. *Marine Ecology Progress Series*, 58(3), 287–297. <https://doi.org/10.3354/meps058287>
- 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., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- De Veen, J. F. (1962). On the subpopulations of plaice in the southern North Sea. *International Council for the Exploration of the Sea*, 94, 1–6.
- Delcourt, N., Quevedo, C., Nonne, C., Fons, P., O'Brien, D., Loyaux, D., Diez, M., Autelitano, F., Guillemot, J. C., Ferrara, P., Muriana, A., Callot, C., Hérault, J. P., Herbert, J. M., Favre, G., & Bono, F. (2015). Targeted identification of sialoglycoproteins in hypoxic endothelial cells and validation in zebrafish reveal roles for proteins in angiogenesis. *The Journal of Biological Chemistry*, 290(6), 3405–3417. <https://doi.org/10.1074/jbc.M114.618611>
- Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P. A. (2018). The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nature Communications*, 9(1), 2518. <https://doi.org/10.1038/s41467-018-04963-6>
- Eklom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1–15. <https://doi.org/10.1038/hdy.2010.152>
- Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), 422–433. <https://doi.org/10.1038/nrg.2016.58>
- Erfteimeijer, P. L. A., van Beek, J., Bolle, L. J., Dickey-Collas, M., & Los, H. F. J. (2009). Variability in transport of fish eggs and larvae. I. Modelling the effects of coastal reclamation. *Marine Ecology Progress Series*, 390, 167–181. <https://doi.org/10.3354/meps08173>
- Fan, Z., & Fox, D. P. (1991). Robertsonian polymorphism in plaice, *Pleuronectes platessa* L., and cod, *Gadus morhua* L., (Pisces Pleuronectiformes and Gadiformes). *Journal of Fish Biology*, 38(5), 635–640. <https://doi.org/10.1111/j.1095-8649.1991.tb03152.x>
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Feng, C., Pettersson, M., Lamichhane, S., Rubin, C. J., Rafati, N., Casini, M., Folkvord, A., & Andersson, L. (2017). Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *eLife*, 6, e23907. <https://doi.org/10.7554/eLife.23907>
- Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J. A., Gómez-Garrido, J., Carreté, L., Bello, X., Gut, M., Gut, I. G., Marcet-Houben, M., Forn-Cuní, G., Galán, B., García, J. L., Abal-Fabeiro, J. L., Pardo, B. G., Taboada, X., Fernández, C., ... Martínez, P. (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Research*, 23(3), 181–192. <https://doi.org/10.1093/dnares/dsw007>
- Gagnaire, P. A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8(8), 769–786. <https://doi.org/10.1111/eva.12288>
- Gautier, M., & Vitalis, R. (2012). rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, 28(8), 1176–1177. <https://doi.org/10.1093/bioinformatics/bts115>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Guo, B., Li, Z., & Merilä, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular Ecology*, 25(12), 2833–2852. <https://doi.org/10.1111/mec.13657>
- Harding, D., Nichols, J. H., & Tungate, D. S. (1978). The spawning of plaice (*Pleuronectes platessa* L.) in the Southern North Sea and English Channel. *Rapports et Procès-Verbaux des Réunions*, 172, 102–113.
- Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4), 333–362. <https://doi.org/10.1111/j.1467-2979.2008.00299.x>
- Hayashi, T., Kageyama, Y., Ishizaka, K., Xia, G., Kihara, K., & Oshima, H. (2001). Requirement of Notch 1 and its ligand jagged 2 expressions for spermatogenesis in rat and human testes. *Journal of Andrology*, 22(6), 999–1011. <https://doi.org/10.1002/j.1939-4640.2001.tb03441.x>
- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R., & Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, 11, 123–136. <https://doi.org/10.1111/j.1755-0998.2010.02943.x>
- Hemmer-Hansen, J., Nielsen, E. E., Frydenberg, J., & Loeschcke, V. (2007). Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity*, 99(6), 592–600. <https://doi.org/10.1038/sj.hdy.6801055>
- Hinegardner, R., & Rosen, D. E. (1972). Cellular DNA content and the evolution of Teleostean fishes. *The American Naturalist*, 106(951), 621–644. <https://doi.org/10.1086/282801>
- Hoarau, G., Piquet, A. M. T., van der Veer, H. W., Rijnsdorp, A. D., Stam, W. T., & Olsen, J. L. (2004). Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: A comparison of resolving power between microsatellites and mitochondrial DNA data. *Journal*

- of *Sea Research*, 51(3–4), 183–190. <https://doi.org/10.1016/j.seares.2003.12.002>
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, 32(5), 767–769. <https://doi.org/10.1093/bioinformatics/btv661>
- Hunter, E., Metcalfe, J. D., Arnold, G. P., & Reynolds, J. D. (2004). Impacts of migratory behaviour on population structure in North Sea plaice. *Journal of Animal Ecology*, 73(2), 377–385. <https://doi.org/10.1111/j.0021-8790.2004.00801.x>
- Jenkins, T. L., Castilho, R., & Stevens, J. R. (2018). Meta-analysis of north-east Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ*, 6, e5684. <https://doi.org/10.7717/peerj.5684>
- Johannesson, K., & André, C. (2006). Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, 15(8), 2013–2029. <https://doi.org/10.1111/j.1365-294X.2006.02919.x>
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). DiveRsim: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4(8), 782–788. <https://doi.org/10.1111/2041-210X.12067>
- Kennington, W. J., & Hoffmann, A. A. (2013). Patterns of genetic variation across inversions: Geographic variation in the In(2L)t inversion in populations of *Drosophila melanogaster* from eastern Australia. *BMC Evolutionary Biology*, 13(1), 100. <https://doi.org/10.1186/1471-2148-13-100>
- Kettle, A. J., Morales-Muñiz, A., Roselló-Izquierdo, E., Heinrich, D., & Vøllestad, L. A. (2010). Refugia of marine fish in the northeast Atlantic during the last glacial maximum: Concordant assessment from archaeozoology and palaeotemperature reconstructions. *Climate of the Past Discussions*, 6(4), 1351–1389. <https://doi.org/10.5194/cpd-6-1351-2010>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Krueger, F. (2015). *Trim Galore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.* [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
- Kusakabe, M., Ishikawa, A., Ravinet, M., Yoshida, K., Makino, T., Toyoda, A., Fujiyama, A., & Kitano, J. (2017). Genetic basis for variation in salinity tolerance between stickleback ecotypes. *Molecular Ecology*, 26(1), 304–319. <https://doi.org/10.1111/mec.13875>
- Le Moan, A., Bekkevold, D., & Hemmer-Hansen, J. (2021). Evolution at two time frames: Ancient structural variants involved in post-glacial divergence of the European plaice (*Pleuronectes platessa*). *Heredity*, 126(4), 668–683. <https://doi.org/10.1038/s41437-020-00389-3>
- Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D., & Hemmer-Hansen, J. (2019). Beyond parallel evolution: When several species colonize the same environmental gradient. *bioRxiv*, 662569. <https://doi.org/10.1101/662569>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.* <http://arxiv.org/abs/1303.3997>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Liu, S., Hansen, M. M., & Jacobsen, M. W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, 25(20), 5187–5202. <https://doi.org/10.1111/mec.13827>
- Mapleson, D., Venturini, L., Kaithakottil, G., & Swarbreck, D. (2018). Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience*, 7(12), giy131. <https://doi.org/10.1093/gigascience/giy131>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Mather, N., Traves, S. M., & Ho, S. Y. W. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecology and Evolution*, 10(1), 579–589. <https://doi.org/10.1002/ece3.5888>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Molecular Ecology*, 29(14), 2513–2516. <https://doi.org/10.1111/mec.15500>
- Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1), D377–D386. <https://doi.org/10.1093/nar/gks1118>
- Mohrholz, V. (2018). Major Baltic inflow statistics - revised. *Frontiers in Marine Science*, 5, 384. <https://doi.org/10.3389/fmars.2018.00384>
- Nielsen, E., Støttrup, J. G., Heilmann, J., & MacKenzie, B. R. (2004). The spawning of plaice *Pleuronectes platessa* in the Kattegat. *Journal of Sea Research*, 51(3–4), 219–228. <https://doi.org/10.1016/j.seares.2003.08.005>
- Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., Mittelholzer, C., Taranger, G. L., Ogden, R., & Carvalho, G. R. (2009). Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, 9(1), 276. <https://doi.org/10.1186/1471-2148-9-276>
- Nielsen, J. G. (1986). Pleuronectidae. In P. J. P. Whitehead, M.-L. Bauchot, J.-C. Hureau, J. Nielsen, & E. Tortonese (Eds.), *Fishes of the North-eastern Atlantic and the Mediterranean* (pp. 1305–1306). UNESCO.
- Nissling, A., Westin, L., & Hjerne, O. (2002). Reproductive success in relation to salinity for three flatfish species, dab (*Limanda limanda*), plaice (*Pleuronectes platessa*), and flounder (*Pleuronectes flesus*), in the brackish water Baltic Sea. *ICES Journal of Marine Science*, 59(1), 93–108. <https://doi.org/10.1006/jmsc.2001.1134>
- Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., & Daniëlsdóttir, A. K. (2008). Recolonization history and large-scale dispersal in the open sea: The case study of the North Atlantic cod, *Gadus morhua* L. *Biological Journal of the Linnean Society*, 94(2), 315–329. <https://doi.org/10.1111/j.1095-8312.2008.00995.x>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), 2074–2093. <https://doi.org/10.1371/journal.pgen.0020190>
- Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, 13(5), 946–952. <https://doi.org/10.1111/1755-0998.12129>
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, 33(3), 290–295. <https://doi.org/10.1038/nbt.3122>



- Petereit, C., Hinrichsen, H. H., Franke, A., & Köster, F. W. (2014). Floating along buoyancy levels: Dispersal and survival of western Baltic fish eggs. *Progress in Oceanography*, 122, 131–152. <https://doi.org/10.1016/j.pocean.2014.01.001>
- Pettersson, M. E., Rochus, C. M., Han, F., Chen, J., Hill, J., Wallerman, O., Fan, G., Hong, X., Xu, Q., Zhang, H., Liu, S., Liu, X., Haggerty, L., Hunt, T., Martin, F. J., Flicek, P., Bunikis, I., Folkvord, A., & Andersson, L. (2019). A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Research*, 29(11), 1919–1928. <https://doi.org/10.1101/gr.253435.119>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Purdom, C. E. (1972). Induced polyploidy in plaice (*Pleuronectes platessa*) and its hybrid with the flounder (*Platichthys flesus*). *Heredity*, 29(1), 11–24. <https://doi.org/10.1038/hdy.1972.61>
- R Development Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://doi.org/10.1038/sj.hdy.6800737>
- Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Molecular Ecology*, 25(1), 287–305. <https://doi.org/10.1111/mec.13332>
- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 43(1), 223–225.
- Rijnsdorp, A. D. (1989). Maturation of male and female North Sea plaice (*Pleuronectes platessa* L.). *ICES Journal of Marine Science*, 46(1), 35–51. <https://doi.org/10.1093/icesjms/46.1.35>
- Rijnsdorp, A. D. (1991). Changes in fecundity of female North Sea plaice (*Pleuronectes platessa* L.) between three periods since 1900. *ICES Journal of Marine Science*, 48(3), 253–280.
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., ... Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. <https://doi.org/10.1038/nature06250>
- Salzberg, S. L. (2019). Next-generation genome annotation: We still struggle to get it right. *Genome Biology*, 20(1), 92. <https://doi.org/10.1186/s13059-019-1715-2>
- Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46(8), 919–925. <https://doi.org/10.1038/ng.3015>
- Seppy, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, 1962(19), 227–245. [https://doi.org/10.1007/978-1-4939-9173-0\\_14](https://doi.org/10.1007/978-1-4939-9173-0_14)
- Shao, C., Bao, B., Xie, Z., Chen, X., Li, B., Jia, X., Yao, Q., Ortí, G., Li, W., Li, X., Hamre, K., Xu, J., Wang, L., Chen, F., Tian, Y., Schreiber, A. M., Wang, N., Wei, F., Zhang, J., ... Chen, S. (2017). The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nature Genetics*, 49(1), 119–124. <https://doi.org/10.1038/ng.3732>
- Shimada, Y., Shikano, T., & Merilä, J. (2010). A high incidence of selection on physiologically important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Molecular Biology and Evolution*, 28(1), 181–193. <https://doi.org/10.1093/molbev/msq181>
- Simpson, J. T., & Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. *Genome Research*, 22(3), 549–556. <https://doi.org/10.1101/gr.126953.111>
- Smit, A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0*. <http://www.repeatmasker.org>
- Stephan, W. (2016). Signatures of positive selection: From selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology*, 25(1), 79–88. <https://doi.org/10.1111/mec.13288>
- Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10), 2144–2164. <https://doi.org/10.1111/mec.13606>
- Ulrich, C., Boje, J., Cardinale, M., Gatti, P., LeBras, Q., Andersen, M., Hemmer-Hansen, J., Hintzen, N. T., Jacobsen, J. B., Jonsson, P., Miller, D. C. M., Nielsen, E. E., Rijnsdorp, A. D., Sköld, M., Svedäng, H., & Wennhage, H. (2013). Variability and connectivity of plaice populations from the Eastern North Sea to the Western Baltic Sea, and implications for assessment and management. *Journal of Sea Research*, 84, 40–48. <https://doi.org/10.1016/j.seares.2013.04.007>
- Ulrich, C., Hemmer-Hansen, J., Boje, J., Christensen, A., Hüsey, K., Sun, H., & Clausen, L. W. (2017). Variability and connectivity of plaice populations from the Eastern North Sea to the Baltic Sea, part II. Biological evidence of population mixing. *Journal of Sea Research*, 120, 13–23. <https://doi.org/10.1016/j.seares.2016.11.002>
- The UniProt Consortium. (2018). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 46(5), 2699. <https://doi.org/10.1093/nar/gky092>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L., & Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*, 7(8), giy093. <https://doi.org/10.1093/gigascience/giy093>
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1587), 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Vidal, O., Lindberg, M. K., Hollberg, K., Baylink, D. J., Andersson, G., Lubahn, D. B., Mohan, S., Gustafsson, J. Å., & Ohlsson, C. (2000). Estrogen receptor specificity in the regulation of skeletal growth and maturation in male mice. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), 5474–5479. <https://doi.org/10.1073/pnas.97.10.5474>
- Volckaert, F. A. M. (2013). (Flat)fish stocks in an ecosystem and evolutionary perspective. *Journal of Sea Research*, 75, 19–32. <https://doi.org/10.1016/j.seares.2012.06.011>
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11), e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89, 438–450. <https://doi.org/10.1093/jhered/89.5.438>
- Was, A., Gosling, E., & Hoarau, G. (2010). Microsatellite analysis of plaice (*Pleuronectes platessa* L.) in the NE Atlantic: Weak genetic structuring in a milieu of high gene flow. *Marine Biology*, 157, 447–462. <https://doi.org/10.1007/s00227-009-1331-x>
- Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology*, 28(6), 1203–1209. <https://doi.org/10.1111/mec.15066>

- White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences*, 277(1688), 1685–1694. <https://doi.org/10.1098/rspb.2009.2214>
- Winterhalter, B., Floden, T., Ignatius, H., Axberg, S., & Niemisto, L. (1981). Geology of the Baltic Sea. In A. Voipio (Ed.), *Elsevier oceanography series* (pp. 1–121). Elsevier. [https://doi.org/10.1016/S0422-9894\(08\)70138-7](https://doi.org/10.1016/S0422-9894(08)70138-7)

## SUPPORTING INFORMATION

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

**How to cite this article:** Weist, P., Jentoft, S., Tørresen, O. K., Schade, F. M., Pampoulie, C., Krumme, U., & Hanel, R. (2022). The role of genomic signatures of directional selection and demographic history in the population structure of a marine teleost with high gene flow. *Ecology and Evolution*, 12, e9602. <https://doi.org/10.1002/ece3.9602>