



Evolution at two time frames: ancient structural variants involved in post-glacial divergence of the European plaice (*Pleuronectes platessa*)

Alan Le Moan ^{1,2} · Dorte Bekkevold¹ · Jakob Hemmer-Hansen¹

Received: 18 May 2020 / Revised: 5 November 2020 / Accepted: 5 November 2020 / Published online: 2 February 2021
© The Author(s), under exclusive licence to The Genetics Society 2021

Abstract

Changing environmental conditions can lead to population diversification through differential selection on standing genetic variation. Structural variant (SV) polymorphisms provide examples of ancient alleles that in time become associated with novel environmental gradients. The European plaice (*Pleuronectes platessa*) is a marine flatfish showing large allele-frequency differences at two putative SVs associated with environmental variation. In this study, we explored the contribution of these SVs to population structure across the North East Atlantic. We compared genome-wide population structure using sets of RAD-sequencing SNPs with the spatial structure of the SVs. We found that in contrast to the rest of the genome, the SVs were only weakly associated with an isolation-by-distance pattern. Indeed, both SVs showed important variation in haplogroup frequencies, with the same haplogroup increasing both along the salinity gradient of the Baltic Sea, and found in high frequency in the northern-range margin of the Atlantic. Phylogenetic analyses suggested that the SV alleles are much older than the age of the Baltic Sea itself. These results suggest that the SVs are older than the age of the environmental gradients with which they currently co-vary. Altogether, our results suggest that the plaice SVs were shaped by evolutionary processes occurring at two time frames, firstly following their origin, ancient spread and maintenance in the ancestral populations, and secondly related to their current association with more recently formed environmental gradients such as those found in the North Sea–Baltic Sea transition zone.

Introduction

Disentangling neutral from selective forces acting on the process of population divergence is a key part of population genetics, as the two forces provide different information about the mechanisms shaping population structure (Wright 1978). While neutral markers provide data about population connectivity and demographic processes

(Beaumont and Nichols 1996), markers under selection can reveal the existence of locally adapted populations or cryptic speciation processes in a meta-population (Gagnaire et al. 2015; Benestan et al. 2016). Access to high numbers of genetic markers obtained from high-throughput sequencing technologies has increased the power to detect genetic markers under selection. Recently, such genomic approaches have highlighted the importance of structural variants (SVs), which have often been associated with signals of selection in natural populations (see special issue: Wellenreuther et al. 2019). SVs can originate from changes in copy number (deletion, insertion and duplication), orientation (inversion) or position (translocation and fusion) of DNA sequences resulting in reduced levels of recombination in individuals that are heterozygous for different arrangements (Kirkpatrick 2010). Reduced recombination causes such chromosomal rearrangements to follow independent evolutionary pathways, often showing higher levels of population divergence than collinear regions of the genome (Navarro and Barton 2003a, 2003b; Farré et al. 2013). The SVs can harbor

Associate editor: Rui Faria

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-020-00389-3>.

✉ Alan Le Moan
alan.le.moan@gmail.com

¹ National Institute of Aquatic Resources, Technical University of Denmark, Vejlshøjvej 39, 8600 Silkeborg, Denmark

² Department of Marine Sciences at Tjärnö, University of Gothenburg, Laboratorievägen 10, Strömstad, Sweden

hundreds of genes and thus may have functional consequences for the organism (reviewed in Wellenreuther and Bernatchez 2018; Oomen et al. 2020).

SVs have been found to be important for promoting the evolution and maintenance of locally adapted populations in the face of gene flow. In a system of interconnected populations, gene flow results in the rapid homogenization of genetic variation and is the main evolutionary force acting against the process of divergence (Slatkin 1987). In the most extreme condition, when gene flow is too high, mutations that are favorable in a local environment are likely to recombine with maladapted genomic backgrounds, and can be lost through swamping effects (Lenormand 2002). However, if two (or more) locally advantageous mutations are located within the same SV, the absence of recombination with adjacent, maladapted genomic backgrounds increases their fitness locally. In comparison to independent genetic variants, mutations within SVs are less affected by the migration load and therefore more likely to maintain locally adapted alleles in a meta-population (Kirkpatrick and Barton 2006). The strength of population differentiation will depend on the balance between genetic drift, migration and selection, the effect of the latter being inflated if both locally adapted variation and genetic incompatibilities (decreased fitness in heterozygotes) are found at the same time in the population or even within the same SV (Bierne et al. 2011; Faria et al. 2019). Co-adaptation involving positive epistatic interactions between loci is also more likely to be maintained within a SV (Dobzhansky 1970; Feldman et al. 1996), resulting in a supergene that further increases the adaptive potential of SVs (Thompson and Jiggins 2014).

The age of the SVs is often older than the age of the contemporary populations in which they are studied, suggesting that their adaptive potential often relies on ancient polymorphisms (reviewed in Wellenreuther and Bernatchez 2018; Marques et al. 2019), representing a source of standing variation for population divergence and adaptation. For instance, ancient SVs have promoted the repeated evolution of ecotypes following the post-glacial recolonization of new environments, as described in systems undergoing parallel evolution (e.g., Jones et al. 2012; Nelson and Cresko 2018; Morales et al. 2019). It has been suggested that such adaptive genetic variation can be shaped during distinct time periods (i.e., “Evolution at two time-frames”, c.f. van Belleghem et al. 2018). The first period corresponding to the initial increase in the frequency of adaptive variants, can be facilitated by the presence of SVs (e.g., Barton and Kirkpatrick 2006) and/or a period of isolation (e.g., Feder et al. 2011; van Belleghem et al. 2018), while the second period, sometimes much later, corresponds to their association with environmental, physical and/or endogenous barriers to gene flow, observable in

contemporary meta-populations (Bierne et al. 2011; van Belleghem et al. 2018; Morales et al. 2019). While the origin of many SVs often remains unclear, they sometimes result from adaptive introgression that has positive fitness effects for the introgressed species or population. For example, adaptive introgression of a major inversion is responsible for mimicry of the wing color pattern between poisonous *Heliconius* butterflies (The *Heliconius* Genome Consortium et al. 2012; Jay et al. 2018).

The European plaice (*Pleuronectes platessa*) is a marine flatfish with biological traits such as long pelagic egg and larval phases that promote high levels of gene flow (Harding et al. 1978; Rijnsdorp 1991; Waples and Gaggiotti 2006). Plaice is distributed in the eastern Atlantic from the Iberian Peninsula in the South to the Barents Sea and Greenland in the North. Previous studies based on six microsatellite loci have found weak, often statistically non-significant, levels of population structure across Europe, except across the bathymetric barrier between the continental shelf and off-shelf regions (Europe vs. Iceland and Faroe Islands, respectively), where depth apparently acts as a strong physical barrier (Hoarau et al. 2002; Was et al. 2010). Although the demographic history of these populations is still poorly known, the off-shelf population shows a higher genetic diversity than most of the shelf populations (Hoarau et al. 2002), which could suggest that the divergence between the shelf and off-shelf is ancient, potentially linked to different glacial lineages. The plaice has also large populations established in the western parts of the Baltic Sea, a brackish environment formed 8000 years ago (8 kya). In a recent study of the genomic basis underlying the colonization of the Baltic Sea, we identified two putative SVs (Le Moan et al. 2019; Johannesson et al. 2020) showing a strong signal of linked selection over large portions of two chromosomes. These two putative SVs were associated with 66% of the top 0.1% of the North Sea–Baltic Sea F_{ST} outlier loci and showed allelic frequency clines associated with the Baltic Sea salinity gradient (Le Moan et al. 2019). However, whether these putative SVs evolved in connection with a founder event and adaptation to a Baltic Sea environment or are ancient polymorphisms remained unresolved. The European plaice is known to hybridize with its sister species, the European flounder (*Platichthys flesus*) (Kijewska et al. 2009). The flounder is a euryhaline species better adapted to low salinity than the plaice, and can be found in the innermost parts of the Baltic Sea where plaice does not occur (Hemmer-Hansen et al. 2007a). Therefore, it is also possible that the European flounder could be a source of the SVs found in plaice.

In the present study, our overall objective was to increase our understanding of the origin of the putative SVs and their effects in contemporary populations of the European plaice.

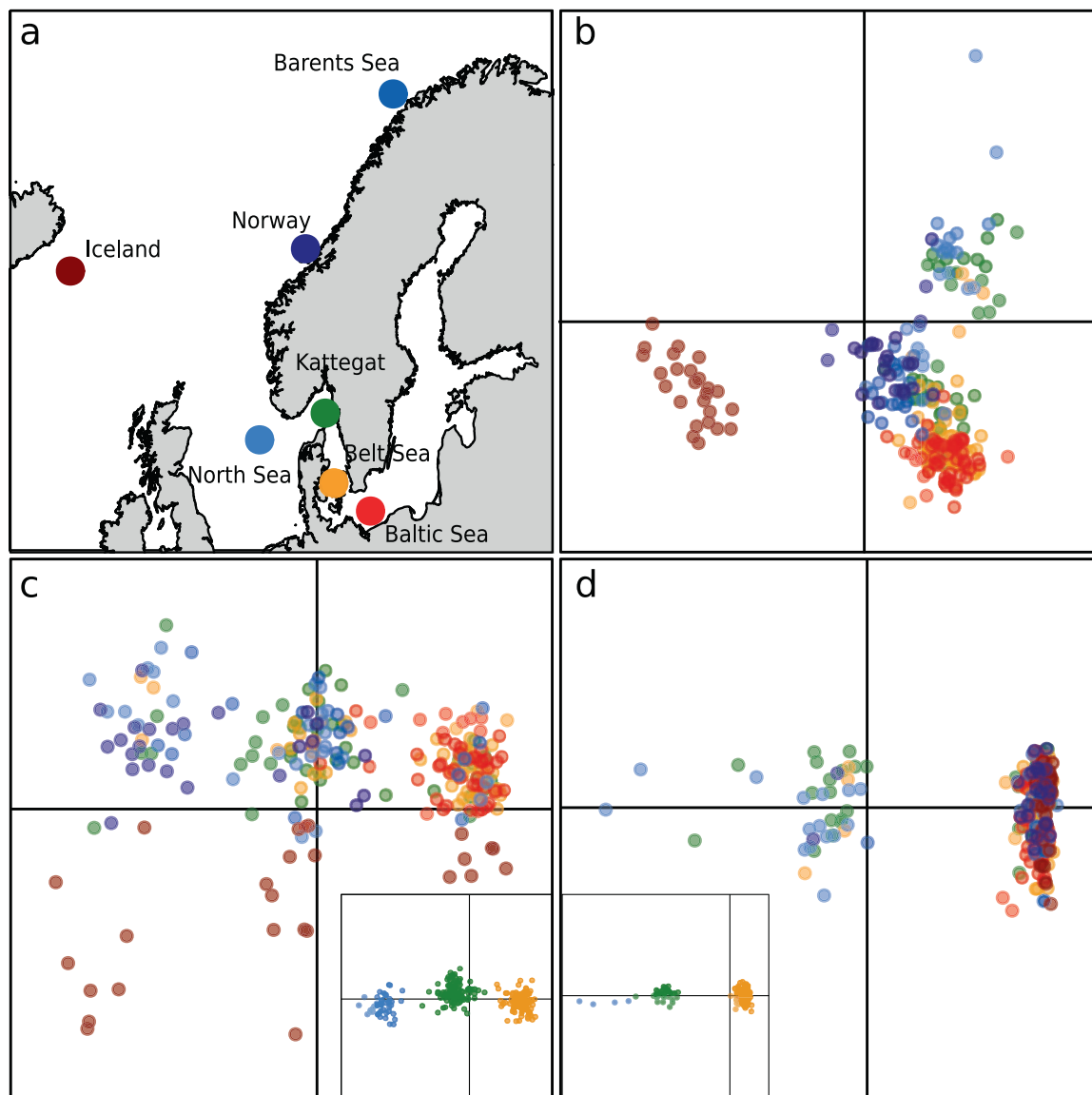


Fig. 1 Population structure of European plaice in northern Europe. Sampling design (a) and principal component analyses performed on individual diversity for the European plaice based on a genomic dataset of 3019 SNPs in the overall dataset (b), 222 SNPs on chromosome 19 (c), and 210 SNPs on chromosome 21 (d). Colors

correspond to sampling sites represented on the map in (a). Insets in (c) and (d) show DAPC results grouping individuals by haplogroup, where green dots are individuals from the heterozygote haplogroup and the yellow and blue dots are individuals from the homozygote haplogroups 1 and 3, respectively.

Specifically, the main goals of the study were to (i) test for a potential flounder origin of SVs, (ii) estimate the age of the SVs in plaice with a phylogenetic approach, (iii) re-assess the population structure in European plaice from northern Europe and from Iceland with the use of a population genomics approach, (iv) evaluate the contribution of SVs to population structure, and (v) provide relevant data to understand the extent to which selection is involved in maintaining the allelic clines observed at the SV. This work thus provides increased insight into the relative roles of demographic history, hybridization, environmental gradients and genomic structural changes in the evolution of marine species.

Materials and methods

Geographical sampling

European plaice samples were collected at seven sites distributed across Northern Europe and Iceland (Fig. 1a and Table 1) during the spawning season (Rijnsdorp 1991). Samples from the North Sea, Kattegat, the Belt Sea and the Baltic Sea were collected in 2016–2017. These samples were also analyzed by Le Moan et al. (2019) who studied the diversification process involved in the colonization of the Baltic Sea across multiple flatfish species, and where

Table 1 Details of the sampling locations, with the corresponding site ID, sample sizes, longitude, latitude and the date when the samples were collected.

Location	Site ID	Sample size	Sample analyzed	Longitude	Latitude	Dates of sampling
Iceland	Iceland	25	25	62.9	-15.8	Feb 2013
Barents Sea	Barents Sea	25	19	70.7	20.2	Feb 2013
Norway	Norway	25	24	64.8	8.8	Feb 2013
North Sea	North Sea	30	26	55.26	7.09	Feb 2016
Kattegat–Skagerrak	Kattegat	50	44	57.63	9.30	Feb 2017
Øresund–Belt Sea	Belt Sea	50	48	54.65	10.64	Mar 2016–2017
Southwest Baltic	Baltic Sea	50	48	54.98	14.54	Mar 2017

two large SVs were identified. To explore the spatial distribution of these SVs in greater detail, three additional northern sites were included in the current study. These samples were collected from the Barents Sea, Norway and Iceland in 2013. Most of the northern parts of the plaice distribution were covered with this sampling design. Analyses included 255 plaice in total, along with ten European flounders (the species hybridizing with plaice in the study area), and ten common dab (*Limanda limanda*), a closely related but reproductively isolated species from both European plaice and European flounder, to be used as an outgroup in a phylogenetic analysis.

ddRAD libraries and sequencing

Whole genomic DNA was extracted from gill tissue using the DNeasy Blood Tissue kit (Qiagen). The DNA concentration was measured with the Broad Range protocol of Qubit version 2.0[®] following the instruction manual. DNA extractions were diluted to 20 ng/μl. Four double-digest RAD (ddRAD) libraries were constructed following Poland and Rife (2012), using Pst1- and Msp1-restriction enzymes with rare and frequent cutting sites, respectively. Each library was made by randomly pooling between 60 and 75 barcoded individuals from various locations. The libraries were size-selected on agarose gels in order to retain insert sizes between 350 and 450 bp. After an amplification step (12 cycles), the libraries were purified with AMPure[®] beads and their quality was checked on a Bioanalyzer 2100 using the High Sensitivity DNA protocol (Agilent Technologies). The targeted size selection was successful in most libraries, except in one that was slightly shifted toward a shorter insert size (from 300 to 400 bp). The effects of different library size ranges are discussed further below. Each library was pair-end sequenced on one Illumina HiSeq4000[®] lane (2 × 101 bp).

Bioinformatics

Raw sequences were processed using the “ref-map” pipeline from Stacks version 2.1 (Catchen et al. 2013). Specifically, the samples were demultiplexed with “process radtag” by

removing reads with mean sequencing quality below 10 and reads with uncalled base pairs. On average, we obtained six million reads per sample (Fig. S1). The reads were trimmed to 85 bp using trimmomatic (Bolger et al. 2014), and aligned to the Japanese flounder (*Paralichthys olivaceus*) genome (Shao et al. 2017) using bwa-mem set with default parameters (Li and Durbin 2009). This reference genome is from a species of the same family of the European plaice, which has a relatively conserved genome structure (Robledo et al. 2017). On average, 65% of reads per sample mapped to the reference genome (Fig. S1). SNPs were called based on the mapping results using the “gstacks” function with the “marukilow” model, and alpha parameter of 0.05. Only biallelic SNPs genotyped in at least 80% of the individuals within each sampling site and with a maximum heterozygosity of 0.80 were called using the “population” function. All individuals with more than 10% missing data were removed (details in Table 1). Finally, SNPs with a significant departure from Hardy–Weinberg equilibrium (p value < 0.05) in more than 60% of the sampling sites, as well as singletons, were removed using vcftools (Danecek et al. 2011). The average coverage after filtering was 29× per sample (Fig. S1). Unfortunately, size selection was slightly shifted between the first three libraries (containing North Sea and Baltic Sea samples) and the last library (containing Barents Sea, Norway and Iceland samples). This shift resulted in a reduction of the genomic sampling when keeping only loci sequenced for all the sampling sites. Therefore, three datasets were constructed to take the differences in size selection into account: the “overall dataset” including all sampling sites (8587 RAD tags with 28,016 SNPs genotyped in 234 individuals), the “southern dataset” including the North Sea, Kattegat, Belt Sea and Baltic Sea (17,342 RAD tags with 56,740 SNPs genotyped in 166 individuals) and the “northern dataset” including Iceland, Norway and Barents (35,549 RAD tags with 92,706 SNPs genotyped in 68 individuals). These datasets were subsequently used for different analyses focusing on different aspects of population differentiation (see below), and further filtered to fulfill the requirements of the analyses performed (details Table S1).

Population structure and demographic history

The “overall dataset” was used to describe the overall population structure and was thinned by removing loci with minor-allele frequencies (MAF) below 5% and by keeping only one random SNP per bin of 1 kb to limit the effects from physical linkage but not the linkage disequilibrium (LD) within SVs (Fig. S2). In total, we kept 3019 SNPs from which individual genetic diversity was visualized using PCA analyses, conducted with the R package *ade-genet* (Jombart 2008). The same package was used to compute population-specific heterozygosity. Pairwise genetic differentiation (F_{ST}) between samples was estimated following the method of Weir and Cockerham (1984) using the R package *StAMPP* (Pembleton et al. 2013). We used 1000 bootstraps over loci to evaluate if pairwise F_{ST} values were significantly different from 0. The effect of isolation-by-distance was assessed with a Mantel test from the R base package (Ihaka and Gentleman 1996) between pairwise F_{ST} and geographical distances among sampling sites using 9999 permutations. All analyses were conducted by including and excluding the two chromosomes carrying the SVs, as well as using the information from these two chromosomes only. Additionally, the correlation between genome-wide population structure and the structure displayed by the SVs was assessed by Mantel tests on the genome-wide F_{ST} matrix without the SVs and both F_{ST} matrices of the SVs only.

We used an approximation-of-diffusion approach, as implemented in the software *moments* (Jouanous et al. 2017), to examine the demographic history associated with the major population breaks that separate populations from the continental shelf (Norway) and Iceland (Hoarau et al. 2002) using the “northern” dataset. The analysis was replicated with and without the two chromosomes carrying SVs. For this analysis, only polymorphic sites with a minimum allele count of 2 in at least one of the two populations were kept, which were then filtered for LD (1 SNP per 1 kb). Four basic models of demographic history were used, representing the scenario of a strict isolation (SI), an isolation-with-migration (IM), an ancestral migration (AM), and a secondary contact (SC). To include more realistic evolutionary histories, each of these four basic scenarios were further developed in eight variants, which lead to a comparison of 32 scenarios in total (Fig. 2). The 28 more elaborated models were tuned to include fluctuation of the effective population size (representing bottlenecks and/or expansions), which can happen in the ancestral population (“af” = ancient fluctuation) and/or in the Icelandic population (“rf” = recent fluctuation, Momigliano et al. 2020). Then, each of these demographic models included the possibility of heterogeneous genomic signatures involving variation in effective population size across loci (model 2N) for the model of SI

(SI, e.g., Cruickshank and Hahn 2014), or in migration rate (model 2m) for the models with gene flow (IM, SC and AM, e.g., Le Moan et al. 2016; Rougemont et al. 2017). These models were compared to the data using the folded version of the Joint Allelic Frequency Spectrum (JAFS) without considering singletons (-z option). We used the optimization routine developed for *δaδi* by Portik et al. (2017) and adapted for moments in Momigliano et al. (2020). This routine uses a four-step optimization procedure to infer the best parameter of each model. Each of the optimization steps were replicated 20, 20, 40, and 40 times, respectively, and the optimization routine was run 20 times for each model to control for convergence. We kept the 10 best runs for each model, and the best model was then selected based on the weighted Akaike Information Criterion (wAIC, e.g., Rougeux et al. 2017). The parameters inferred from the best model were subsequently transformed into meaningful biological numbers following Rougeux et al. (2017), using a mutation rate of 1×10^{-8} (Tine et al. 2014) and a generation time of 3.5 years (Erlandsson et al. 2017). Only the results of the best model and those with an AIC difference below 10 are shown (results for other models available in Supplementary File I).

Genomic analyses of the structural variants

The two chromosomes (C19 and C21) carrying the putative SVs (SV19 and SV21, respectively), were extracted from the overall dataset (LD and MAF pruned) to construct two independent sub-datasets to examine population structure in the SVs alone, using a PCA approach (on 222 and 210 SNPs, respectively). Initial PCA plots showed clustering into three distinct groups (Fig. 1) and hence suggested that each SV behaves like a Mendelian character with two major divergent alleles at multiple linked loci, leading to three distinct “haplogroups” (two homozygotes and one heterozygote, e.g., Mérot et al. 2020). Consequently, we performed DAPC analyses with *ade-genet* (Jombart and Ahmed 2011) set to three groups, to identify the haplogroup of each individual using the *find.cluster* function. The allele frequencies of the SVs for each population were calculated based on the DAPC clusters, using the formula

$$F = \frac{2 \times C_1 + C_2}{2N},$$

where C_1 is the number of individuals assigned to one of the homozygote haplogroups (haplogroup 1 or 3), C_2 the number of individuals assigned to the heterozygote haplogroup (haplogroup 2) in the DAPC, and N is the number of samples in the population. Then, we used the DAPC groups as genotype input to calculate the pairwise F_{ST} between populations at each SV using *hierfstat* (Goudet 2005).

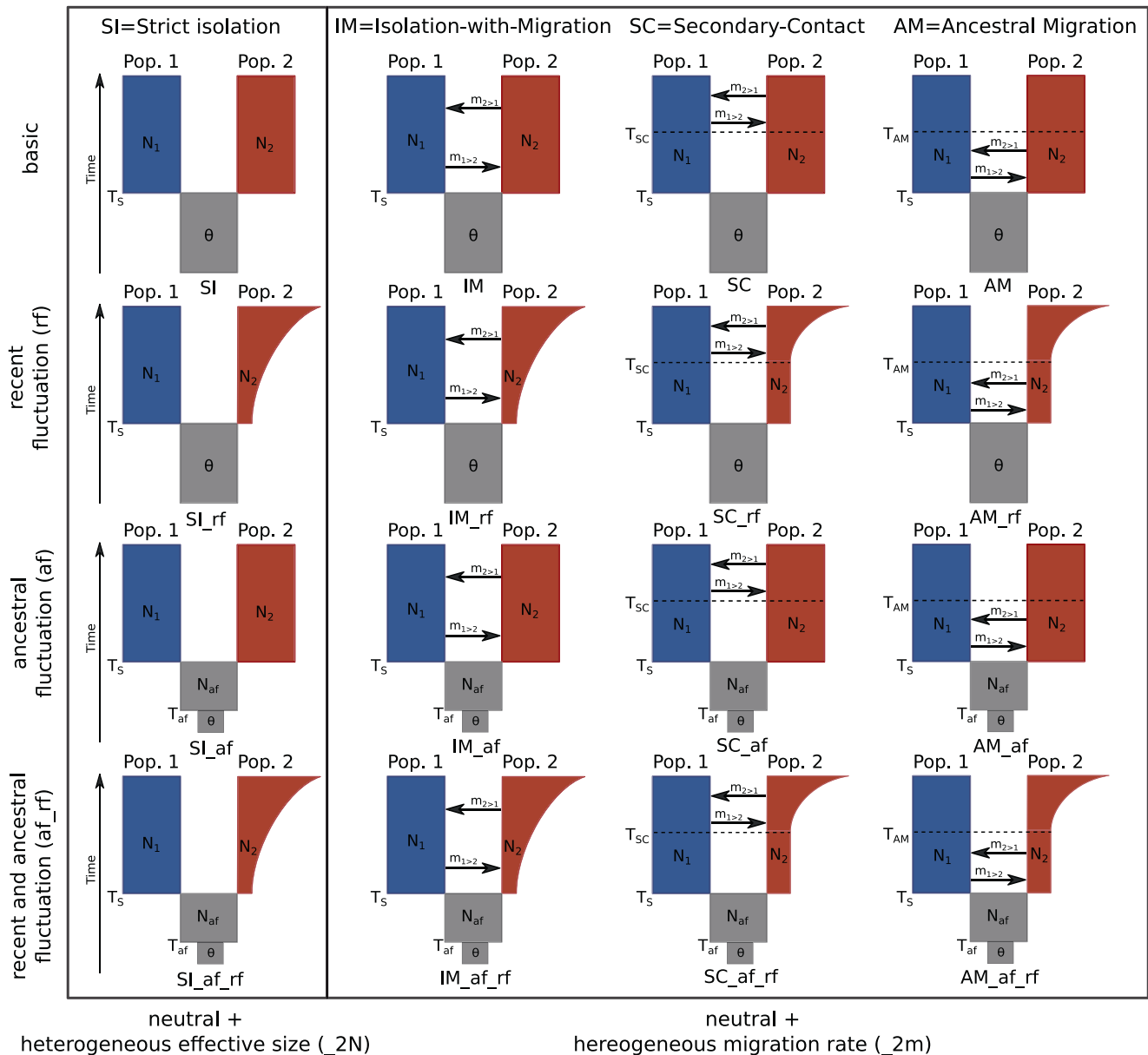


Fig. 2 Schematic representation of the tested models. An ancestral population splits into two derived populations of size N_1 and N_2 at time (T_S) in strict isolation (SI), or connected by continuous migration (IM), by ancient migration events (AM) or by a secondary contact phase (SC). In each scenario, the population evolved without any demographic changes (a, basic), with recent demographic fluctuation in

Iceland (b, rf), ancient demography fluctuation in the ancestral population (c, af) or both ancient and recent fluctuation at the time (d, af_rf). Each of these models were fitted with and without genomic heterogeneity (heterogeneous effective “ $2N$ ” size for SI models, and heterogeneous migration “ $2m$ ” for IM, AM, and SC models).

We analyzed the subsets of the overall data (“northern” and “southern” datasets combined) without pruning for LD to increase the genomic coverage and better describe the genomic heterogeneity along the chromosomes carrying SVs. Nucleotide diversity (π) and H_O were calculated per SNP for each haplogroup inferred in the DAPC. The genomic distribution of differentiation was examined using SNP-specific F_{ST} values. Specifically, this differentiation was calculated between proximal sites for Norway vs. Barents Sea, Norway vs. Iceland using the northern dataset

and for North Sea vs. Baltic Sea, and between homozygotes for haplogroups 1 and 3, using the southern dataset. We used a quantile regression from the quantreg R package to calculate the variation of the upper 1% F_{ST} quantile and average F_{ST} along bins of 100 kb across the chromosome in the different pairwise comparisons. We focused on the upper quantile to limit the effects from variable levels of diversity across the SVs that would tend to depress average F_{ST} in certain regions along the SVs (Fig. S3). Finally, we estimated the overall LD using the southern dataset by

calculating the pairwise correlation between loci localized on chromosomes 19 and 21. These statistics were computed using *vcftools* (Danecek et al. 2011) and the smoothed value of the statistics was plotted with *ggplot2* using local regression with *loess* (Wickham and Winston 2008) with a span parameter of 0.3. In addition, we used *LDheatmap* (Shin et al. 2006) that integrates the information about the physical position of the SNP on the chromosome to represent the heatmap of LD between every pair of SNPs.

Gene content of the structural variants

In order to understand the genetic composition of the two SVs, we extracted the filtered RAD tags from the two SVs into two individual fasta files. The genomic ranges of the SVs were defined based on visual observation of the breaking-point positions of the F_{ST} values between homozygote haplogroup 1 and 3 individuals, starting after the first and ending before the last SNP with $F_{ST} > 0.9$ (from 1.4 to 9.9 Mbp for SV19 and from 10.5 to 20 Mbp for SV21). The RAD tags within the SVs were then aligned back to a previous version of the Japanese flounder genome that is annotated (Shao et al. 2017) using *bwa-mem* set with default parameters (Li and Durbin 2009). All the genes localized between the two most distantly aligned RAD tags on the genome were then extracted from the annotation file using custom bash scripts. Finally, the gene lists were mapped to the Gene Ontology (GO) resource website (<http://geneontology.org/>) to test for functional enrichment using Fisher's exact tests to find GO terms that were over-represented in the gene list. The functional enrichment analyses were performed using both gene lists independently and with the gene lists combined using the zebrafish and the human databases.

Phylogenetic analyses

The ddRAD protocol can be used to identify orthologous sequences with restriction sites conserved across distantly related species. As such, this property was used in order to estimate the age of the SV polymorphisms by building a phylogeny of European plaice and two other species of the Pleuronectidae, the European flounder and the common dab. In order to obtain the sequences of the homozygote haplogroups and infer their divergence, we only retained haplogroups 1 and 3 plaice individuals (based on the DAPC) from the southern dataset. We focused on the southern dataset because it was the only dataset including haplogroup 3 for SV21 (Fig. 1d) and with the highest number of reads overlapping between the plaice and the two outgroup species.

Three independent phylogenies were constructed using concatenated ddRAD loci, one representing each SV from

chromosomes 19 and 21 and one representing loci localized outside the SVs on chromosome 19. The loci from the SVs were extracted based on the genomic ranges defined above for gene-content analyses, while the loci representative of the genome-wide divergence were selected from outside the SVs to get a sequence length similar to that of the SVs (from 15 to 25 Mbp on chromosome 19). Chromosome 19 was used to infer phylogenies both within and outside the SV in order to estimate any effects that the SVs may have on collinear regions of the same chromosome. Only ddRAD loci with sequence information for both homozygote haplogroups in plaice and in flounder and/or dab were used for the phylogeny. The full sequence of each RAD locus was extracted into individual fasta files with the population function of *Stacks* (Catchen et al. 2013) using a “whitelist” (-w) option comprising the RAD-tag ID from the filtered southern dataset. For each RAD locus, one random RAD allele per individual was retained. All alleles were concatenated into one pseudo-sequence using a custom script. The three phylogenies were inferred based on orthologous sequences of 10125, 6152, and 10341 bp for chromosomes 19 and 21 and “collinear” (i.e., outside the SV on chromosome 19), respectively. All phylogenies were estimated in *RaxML* (Stamatakis 2014), under the GTR+GAMMA model with a random number set as seed. Finally, we tested for potential gene flow between species using the “f4” statistic from *Treemix* (Pickrell and Pritchard 2012), evaluating the mismatch between the tree topology inferred with *RaxML* and individual SNP topologies.

The length of the inferred branch between each cluster represents the number of substitutions occurring after the split of the species/haplogroups, and it is directly proportional to the time of divergence under neutral processes (Kimura 1983). We applied a strict molecular clock to transform this nucleotide divergence into time since divergence in years. Specifically, we divided the substitution rate along each branch by the average SNP mutation rate per site, and multiplied this value by the average generation time described for the demographic inferences. Although this approach is commonly used to age SVs, the divergence estimates should be interpreted with caution as cross-recombination events in the center of the SV would decrease the estimate of divergence and selection acting on the haplogroup (background and disruptive) would tend to increase the estimate of divergence.

Results

Population structure and demographic history

The first axis of the PCA based on all markers explained 1.45% of the total inertia and distinguished Icelandic plaice

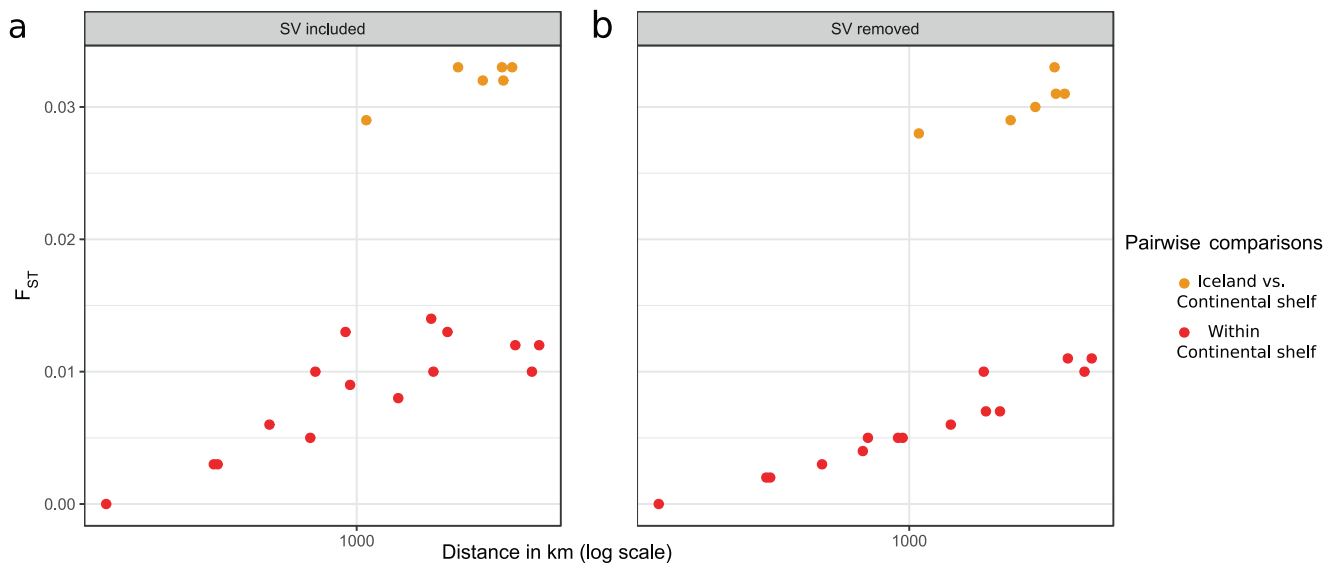


Fig. 3 Relationships between geographic distance and genetic differentiation. The relationships were estimated with (a; $r = 0.56$) and without (b; $r = 0.96$) structural variants. The correlation analyses

were performed using only the samples from the continental shelf (red dots), thus, comparison including the Icelandic sample (yellow dots) were excluded from the analyses.

from all continental shelf individuals, and to a lower extent also identified separation of (marine) Atlantic samples from (brackish) Baltic Sea samples (Fig. 1b). The second axis explained 1.2% of the inertia that roughly traced the gradient from the North Sea to the Baltic Sea. Observed heterozygosity was maximal in the North Sea ($H_O = 0.193/0.190$ with/without SVs) and decreased across the North Sea–Baltic Sea transition zone (Kattegat $H_O = 0.185/0.184$, Belt Sea $H_O = 0.182/0.184$, and Baltic Sea $H_O = 0.178/0.181$) as well as towards the north (Norway $H_O = 0.180/0.180$ and Barents Sea $H_O = 0.179/0.180$). The H_O of the Icelandic population was intermediate ($H_O = 0.184/0.185$). All pairwise F_{ST} estimates were significantly different from zero, except for the two sites in the North Sea–Baltic Sea transition zone (Kattegat vs. Belt Sea, Table S2). Pairwise comparisons including the Iceland population were homogeneously valued around $F_{ST} = 0.03$ (Fig. 3, yellow dots), while all other pairwise F_{ST} estimates were lower (Fig. 3, red dots). The demographic modeling revealed that the most likely scenario for the origin of the separation between Iceland and the continental shelf was a scenario of past isolation followed by a secondary phase of gene flow, including ancient fluctuation in N_e and heterogeneous migration rate during the secondary phase of gene flow (SC_af_2M model, wAIC > 0.5 with and without the SVs—Figs. S4, S5, Table 2 and Supplementary File I). The estimate of the divergence time, including the isolation phase, was two times higher than the estimated time under the SC phase (Table 2). By using a generation time of 3.5 years and a mutation rate of 10^{-8} per generation, the time of split was estimated to ~55–58 kya (SC estimated to 27–28 kya).

The results of the inferences were consistent with and without the chromosomes carrying the SVs (Table 2), suggesting that the demographic effects captured by our model were evident genome wide. The proportion of loci affected by reduced migration rate was nonetheless higher in the inferences performed with the SVs (1.47 vs. 1.05% with and without the SVs, Table 2).

In order to exclude the effects from ancient demography on inference about contemporary neutral structure, the Icelandic population was removed from the analyses of isolation-by-distance. The pairwise F_{ST} across the continental shelf was significantly correlated with geographical distance (Mantel test: $r = 0.59$, $p > 0.01$). Interestingly, removing the SVs from the analysis reduced the variation of the pairwise F_{ST} (Fig. 3, red dots) and resulted in a stronger correlation between genetic and geographical distances ($r = 0.96$, $p > 0.01$). However, this pattern of isolation-by-distance was not detected when only the chromosomes carrying SVs were analyzed ($r = 0.06$, $p = 0.89$ for C19 and $r = 0.11$, $p = 0.35$ for C21, Figs. S6, S7, Table S3). Consequently, the genome-wide pairwise F_{ST} (without the SVs) was not correlated with the pairwise F_{ST} calculated with the chromosomes carrying SVs ($r = 0.08$, $p = 0.77$ for C19, and $r = 0.11$, $p = 0.64$ for C21). Altogether, this result suggests that the population structure found genome-wide is different from the population structure displayed by the SVs. Interestingly, the pairwise F_{ST} values calculated for the two chromosomes with SVs were also not significantly correlated ($r = -0.2$, $p = 0.38$), which suggests unique geographical patterns of substructure for each SV.

Table 2 Summary table of the divergence history between the populations from Iceland and Norway inferred with (SV incl.) and without (SV excl.) the chromosomes carrying putative SVs.

Data	Model	AIC	Δ_{AIC}	W_{AIC}	T_S	$T_{\text{SC or AM}}$	P
SVs incl.	SC_af_2m	1678.00	0.00	0.67	57.91	28.19	1.47
	SC_af_rf_2m	1680.00	2.00	0.25	67.46	28.26	1.60
	IM_af_2m	1684.00	6.00	0.03	96.40	–	49.89
	IM_af_rf_2m	1684.00	6.00	0.03	96.60	–	0.55
	AM_af_2m	1687.00	9.00	0.01	96.35	0.27	49.62
	AM_af_rf_2m	1688.00	10.00	0.01	97.48	0.16	49.42
SVs excl.	SC_af_2m	1620.00	0.00	0.59	54.56	26.74	1.05
	SC_af_rf_2m	1622.00	2.00	0.22	60.71	31.36	0.87
	IM_af_rf_2m	1623.00	3.00	0.13	89.37	–	0.44
	IM_af_2m	1625.00	5.00	0.05	93.68	–	49.90
	AM_af_2m	1628.00	8.00	0.01	92.85	0.16	47.38
	AM_af_rf_2m	1630.00	10.00	0.00	96.32	0.19	48.65

In order of appearance, the table shows the dataset used for the inferences, the six best models sorted by AIC, their AIC, the difference of AIC with the best model (Δ_{AIC}), the weighted AIC (W_{AIC}), the time of split between populations (T_S), the time of secondary contact or ancestral migration ($T_{\text{AM or SC}}$), and the percentage of the genome affected by reduced migration rate (P).

The best model for each pairwise comparison is highlighted in bold.

Genomic variability and differentiation of the structural variants

A high dispersion of samples from the North Sea and Kattegat was observed on the second axis of the PCA (light blue and green samples in Fig. 1b). This dispersion mostly involved SNPs from chromosomes 19 and 21 carrying SVs (Fig. S8b). The PCA based on these chromosomes showed three clusters of samples, which were also inferred in the DAPC analyses (Fig. 1c, d). In both cases, the clusters were likely due to the presence of two divergent multilocus alleles segregating in the plaice populations for each SV, resulting in three haplogroups corresponding to two homozygotes (haplogroups 1 and 3 in blue and yellow, respectively) and one heterozygote cluster (haplogroup 2, in green).

Both SVs were polymorphic at most sampling sites (Fig. 4a, b). However, whereas SV19 allele frequencies were variable across most sampling sites (Fig. 4a), only the North Sea and Kattegat showed both SV21 alleles in high frequency (Figs. 4b, S9), which confirmed the different geographical patterns for the two SVs also identified by a lack of correlation of pairwise F_{ST} estimates. The important variation in allele frequencies leads to high F_{ST} between populations that are geographically close (Table S4, illustrated Fig. 4c). This large differentiation was evident across nearly one-half of the chromosomes, from 1.4 to 9.9 Mbp for SV19 (Fig. 4c), and from 10.5 to 20 Mbp for SV21 (Fig. 4b). The genome-wide differentiation outside the SVs was lower (e.g., mean F_{ST} North Sea vs. Baltic Sea = 0.004, sd = 0.031) than inside the SVs (mean F_{ST} SV19 = 0.203, sd = 0.177 and mean F_{ST} SV21 = 0.167, sd = 0.142, Figs. 4b, S10). The individuals from haplogroups 1 and 3 were differentially fixed for eight and 30 SNPs within

SV19 and SV21, respectively (also represented by the black dashed line in Fig. 4b). Strong LD occurred along the entire SVs, confirming that recombination between the SV alleles is rare (Fig. 4d). Only haplogroup 1 in each SV (in yellow in Fig. 4) showed reduced genetic diversity, whereas haplogroup 3 (in blue in Fig. 4) had a higher genetic diversity than the average observed across the genome. As expected, the individuals from haplogroup 2 showed the highest diversity of the three DAPC groups. The two SVs showed higher overall levels of LD than the average genome-wide (0.2 vs. 0.07), but lower levels than the average LD within the SVs (>0.35, Fig. S11). Chromosome 21 showed two peaks of LD separated by a distance of 5 Mbp of limited LD, and the LD within the two regions was equivalent to the LD between these regions (Fig. 4d).

Gene content and functional enrichment within the structural variants

We identified 200 and 256 genes located within SV19 and SV21, respectively (Supplementary File II). Of these genes, 48/66 in SV19/SV21 had a known function in zebrafish, and 145/191 in humans. For SV21, we detected functional enrichment linked to physiological or morphological membrane disruption (>100% enrichment, $p = 3.1 \times 10^{-10}$), immune and defense response to bacteria (16%, $p = 1.6 \times 10^{-05}$) and chromosome/chromatin organization (10%, $p = 6.5 \times 10^{-6}$) using the zebrafish database, and for purinergic receptor (40%, $p = 4.52 \times 10^{-7}$) using the human database. However, no functional enrichment was identified for genes in SV19. By pooling both gene lists, only the gene enrichment for SV21 was detected without additional function using the zebrafish database. However, using the

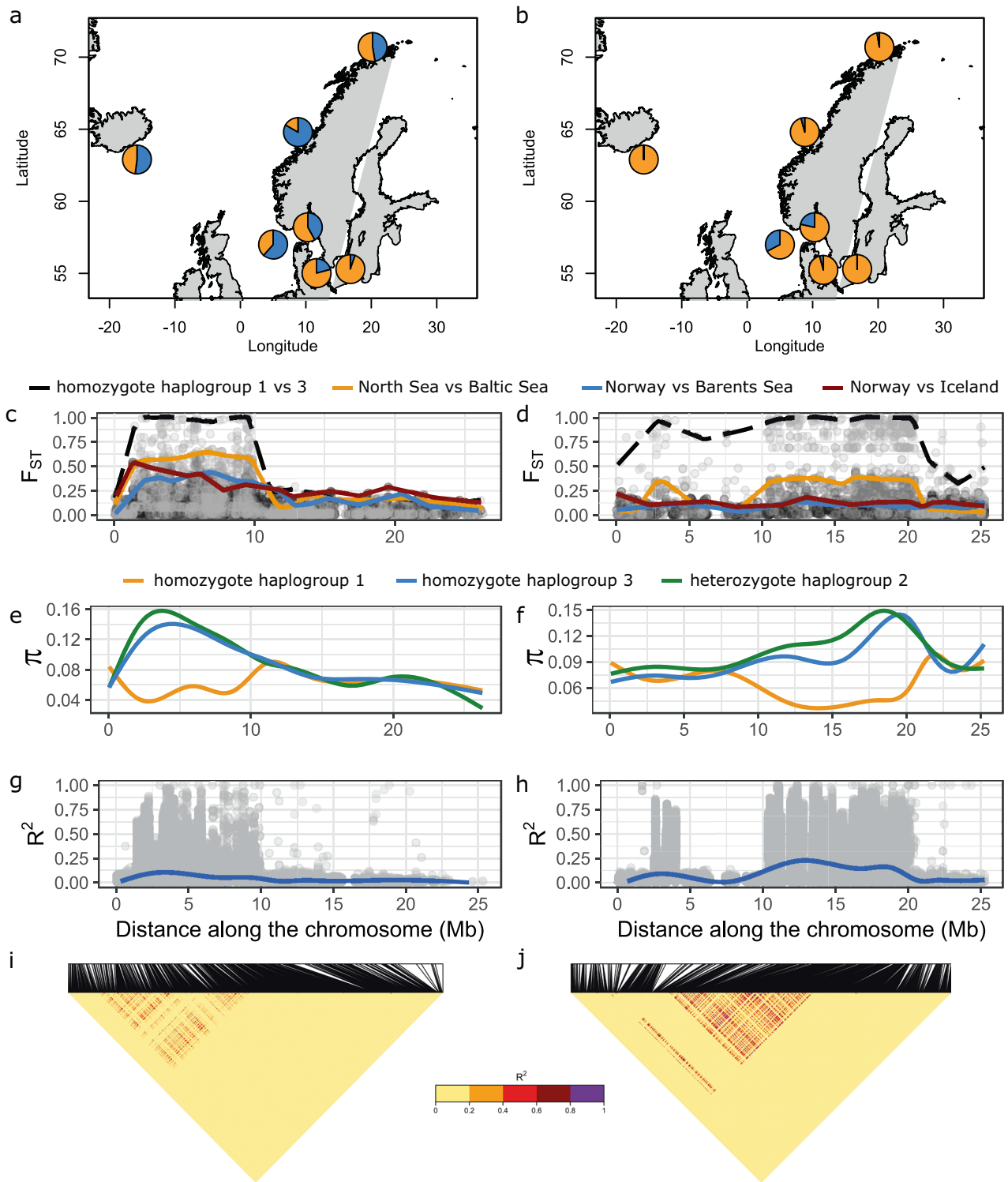
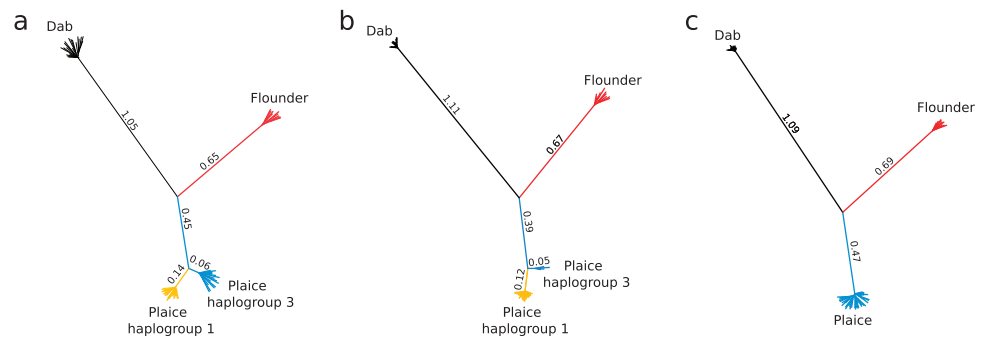


Fig. 4 Detailed analyses of structural variants. Sample-specific data for the European plume structural variants SV19 (a, c, e, g, i) and SV21 (b, d, f, h, j): a, b sample-specific allele frequencies (yellow=allele found twice in haplogroup 1 and once in haplogroup 2, and blue=allele found twice in haplogroup 3 and once in haplogroup 2); c, d differentiation between different pairs of populations and between

haplogroups 1 and 3 along the two chromosomes carrying the SVs, each dot is the F_{ST} value of an individual locus and the lines represent the smoothed upper 1% quantile (one color is one comparison, $\alpha=0.3$); e, f smoothed average π for the three haplogroups identified in the DAPC (span=0.3); g, h variation of LD along the chromosomes and i, j LD heatmaps for pairwise SNP comparisons.

Fig. 5 Phylogenies of European plaice, European flounder and common dab.

Phylogenetic trees are represented for the ddRAD loci within SV19 (a), SV21 (b), and outside the SVs (c). The lengths of the branches reflect the numbers of substitutions per site (multiplied by 100 in the figures).



human database, we identified functional enrichment for pre-B-cell allelic exclusion (62%, $p = 7.5 \times 10^{-5}$), which are genes involved in the immune system. Moreover, when looking specifically for known candidate genes involved in the local adaptation in other marine fishes (e.g., Hemmer-Hansen et al. 2014), we found the presence of two genes linked to heat-shock proteins (heat-shock factor-binding proteins 1 and 4) within SV19.

Phylogenetic analyses

The common dab was the most divergent species in all phylogenetic trees, with 0.024–0.025 substitutions per site compared with European flounder and European plaice (Fig. 5), which corresponds to time of split at 9 million years ago (Mya). All plaice individuals were equally distant from flounder individuals based on the concatenated ddRAD loci representative of the genome-wide divergence, with an average of 0.0110 substitutions per site (4 Mya, Fig. 5). However, the two haplogroups were clearly divergent in both SV phylogenies, with an average distance of 0.0020 and 0.0017 substitutions per site for SV19 and SV21 (Fig. 5), respectively. In each case, the deepest branches were observed for haplogroup 1 (yellow), which leads to a different estimate of divergence for each branch, around 460 (504–434) and 200 (224–179) kya for haplogroups 1 and 3, respectively. The longest branch of the SVs resulted in an average plaice–flounder divergence slightly higher than outside the SVs (0.0120 and 0.0115 substitutions per site, respectively).

We found a low, but statistically significant, effect of introgression between the plaice and the flounder ($f_4 = 0.004$, p value < 0.05). However, this signal was mostly carried by three loci from SV19 and one from SV21 that showed a departure with the tree phylogeny of the SVs. In all cases, the flounder and haplogroup 1 of the plaice were nearly fixed for the same allele, which was different from the major allele observed in the common dab and within haplogroup 3 of the plaice. Removing these SNPs leads to f_4 statistic values not significantly different from 0 (i.e., indicating no effect of introgression).

Discussion

The two SVs previously identified in Le Moan et al. (2019) along the North Sea–Baltic Sea transition zone were polymorphic across most of the north-eastern Atlantic distribution range of European plaice (Fig. 4a, b). Our analyses confirm the clear isolation of the Icelandic population and identify this population as likely originating from a different glacial refugium than the continental shelf populations. This knowledge has important implications for our understanding of the origin and evolution of SVs in the species. We have also confirmed weak but significant genetic structure among continental shelf samples. Removing the SVs from the analyses leads to a stronger pattern of isolation-by-distance among European plaice samples (from $r = 0.56$ to $r = 0.96$), which was not detected in previous European plaice studies based on six microsatellite markers (Hoarau et al. 2002; Was et al. 2010). Thus, the results highlight the power of increasing genomic resolution to detect subtle population structure for species with high gene flow. Below, we discuss the evolution and effects of SVs in European plaice in the context of population history and structure of the species in the northern Atlantic.

Origin of structural variants in European plaice

Type of the structural variants

The two SVs covered nearly half of chromosomes 19 and 21 of the Japanese flounder genome, where a strong LD was maintained over 9 Mbp. These large linkage blocks are expected with chromosomal rearrangements, such as inversions and translocations, which can formally be distinguished by use of a linkage map or genome sequencing (e.g., Faria et al. 2019), but not with the reduced representation approach used in this study. However, assuming a high degree of synteny between European plaice and Japanese flounder genomes, the size of the LD blocks and their central position are consistent with the presence of at least two major inversions in the genome (Kirkpatrick 2010). The observed second peak of LD on chromosome 21

could be due to the presence of a third and smaller inversion. However, the similar value of LD within and between the two LD blocks on this chromosome suggests that they may be part of the same inversion, and that a lack of synteny between the plaice and the Japanese flounder reference genome results in two distant peaks of LD on chromosome 21.

Age of the structural variants

The low genetic diversity of haplogroup 1 (yellow, Fig. 4c) and their long branches in the phylogenies (Fig. 5) suggest that they are the derived form of the SVs. The derived alleles of both SVs were found in both Iceland and continental shelf samples, presumably representing different glacial refugia, which were estimated to have diverged ~57 kya. Moreover, SV19 was polymorphic in both lineages, which could suggest that the SV polymorphism has been present as standing variation in European plaice since at least the time of divergence between the two glacial lineages. This hypothesis is also in agreement with the deep haplogroup divergence observed in both SV phylogenies. For the shortest branch, the split of the main haplotypes was estimated to 200 kya (470 kya for the longest branch). Although these split dates should be interpreted with caution since they do not take effects from selection or recombination between haplogroups into account, the age of the alleles and the fact that the derived alleles of both SVs are present in both glacial lineages, suggest that the SVs are older than the population in which they currently segregate.

Introgression of the structural variants

Adaptive introgression from sister taxa can sometimes explain the presence of ancient “supergenes” adapted to specific environmental conditions, such as those observed in the *Heliconius* butterflies (The *Heliconius* Genome Consortium et al. 2012; Jay et al. 2018). In order to test this hypothesis in the European plaice, we used the European flounder, a euryhaline species adapted to low salinity and known to hybridize with the European plaice, as a candidate for the source of the SVs. Under this hypothesis, the introgressed haplogroup in plaice should be less divergent from flounder than the rest of the genome on average. However, we found the opposite pattern, with each haplogroup being more than or as divergent from the European flounder as the plaice–flounder divergence inferred from outside the SVs. Thus, a potential flounder origin of SVs was rejected by the phylogeny. Introgression was also rejected by the f_4 statistics showing limited evidence of mismatches between the phylogenetic tree and individual SNP topologies. More data are necessary to assess if the few mismatches that were observed are due to a random process

of allele sorting or represent a case of more ancient introgression, pre-dating the formation of the SVs. Hence, our data suggest that the SVs originated after the split of European flounder and European plaice (estimated at 9 Mya from the phylogeny, Fig. 5). However, other potential introgression sources, involving species more closely related to plaice than flounder or ghost species, cannot be ruled out. Further analyses should thus be performed to fully understand the origin of the reported SVs in European plaice.

The population structure in European plaice and the contribution of the SVs

Population structure in European plaice

As mentioned above, we have confirmed the isolation of the Icelandic population, which has also been described in previous work (Hoarau et al. 2002). It has been hypothesized that this differentiation and the genetic diversity data of the Icelandic population (lower number of microsatellite alleles) could reflect the effects from the isolation on the edge of the distribution area and/or a recent bottleneck in the population (Hoarau et al. 2002). Our analyses provide additional information about the history of the Icelandic population, as the demographic analyses performed here suggest that the Icelandic population represents an old population, established from a different glacial refuge that diverged ~57 kya from the other populations of European plaice sampled in this study. This scenario would also explain why the observed heterozygosity is higher in the Icelandic population than in other populations located in other distributional edges of the species (Barents Sea or Baltic Sea), as observed with both the present genomic data and microsatellite data in previous work (Hoarau et al. 2002). In fact, Iceland itself may have been the glacial refuge where a relatively high diversity has been preserved (Maggs et al. 2008). Nevertheless, the physical barrier represented by deep oceanic regions may still be an important factor for maintaining the genetic differences that have evolved during the last glacial maximum, as suggested by Hoarau et al. (2002).

The reductions of diversity found from the North Sea to the Baltic Sea and from the North Sea to the Barents Sea were also reported by the previous studies with 6 microsatellite loci (Hoarau et al. 2002; Was et al. 2010). However, the numbers of markers used at the time were not sufficient to reliably detect any population differences within this geographical region. The populations studied here were sampled along the continental shelf coast lines, resembling a stepping-stone model of isolation, which represents an ideal condition to lead to patterns of isolation-by-distance (Kimura and Weiss 1964). However, other

processes may also be involved in maintaining this pattern (Jenkins et al. 2018), such as the effect of living on the edge of the species distribution range after a post-glacial recolonization (Hewitt 2000). Similar patterns have been reported for various species in association with the salinity gradient of the Baltic Sea (Johannesson and André 2006, Cuveliers et al. 2012), and along the South–North coast of Norway in taxa with lower dispersal capacities than European plaice (Hoarau et al. 2007; Morvezen et al. 2016).

Implication of structural variants for population structure of European plaice

In contrast to genome-wide patterns of weak population structure, the SVs were responsible for population differences much higher (mean F_{ST} on haplogroup frequencies overall pairwise comparisons for SV19 = 0.23 and SV21 = 0.12) than the average genome-wide F_{ST} (estimated at 0.01). Haplogroup 1, identified as the derived allele of the SVs (with the lowest diversity and the highest divergence), reached near-fixation along the environmental gradient of the Baltic Sea in both SVs ($f_{SV19 \text{ derived}} = 0.94$ and $f_{SV21 \text{ derived}} = 1.0$). However, the same allele was found in high frequency towards the northern margin of the plaice distribution, in the Barents Sea ($f_{SV19 \text{ derived}} = 0.52$ and $f_{SV21 \text{ derived}} = 0.97$), which was the most distant site from the Baltic Sea. This variation in allele frequency resulted in a geographical structure of the SVs, which was different from the genome-wide population structure. The distribution edge of the plaice represents a common feature associated with the increase of the derived SV allele frequencies (Fig. 4a, b), potentially resulting in allele-surfing effects in the marginal populations (Excoffier and Ray 2008). Selection may also explain the increased frequencies of haplogroup 1 towards the distribution edge, supported by a net divergence that was 2.4 times higher than the net divergence of haplogroup 3 in both SV phylogenies. This result could be explained by the accumulation of deleterious mutations and background selection in the derived haplogroup (Durrant et al. 2018; Perrier and Charmantier 2018). However, the accumulation of deleterious mutations alone would make the—near-fixation of the derived allele observed in several populations by drift unlikely, especially in species with a large effective size (Ohta 1973), like plaice.

Evidence for local adaptation

The derived allele of the SVs occurred under various habitat conditions, ranging from brackish to marine environments and along temperature and daylight/seasonal gradients within the Atlantic. It is possible that these associations are explained by selection along multiple gradients and on several genes within the SVs. The SNPs carried by the SVs

were previously detected as strong candidates for selection and with significant association with the salinity gradient of the North Sea–Baltic Sea transition zone (Le Moan et al. 2019; Johannesson et al. 2020). In the present study, the highest allele-frequency differences at the SVs were found along this transition zone (Fig. 4a). Interestingly, SV19 carried two heat-shock protein genes, which have important functions for cellular stress response (Sørensen et al. 2003), and have already been identified as candidate loci for local adaptation of Baltic Sea populations in several other marine fishes (Hemmer-Hansen et al. 2007b in European flounder; Nielsen et al. 2009 in Atlantic cod; Limborg et al. 2012 in Atlantic herring). Moreover, several biological functions seem to be enriched within the SVs, notably linked to the immune system, which may also be candidates for local adaptation associated with multiple environmental gradients. Further work focusing on the functional consequences of these SVs will help to understand their effects on the biology of European plaice.

Structural variants promoting evolution at two time frames

“Evolution at two time-frames” has been used to coin the process where several ancient alleles that are locally adapted can be quickly reassembled during more recent colonization of similar environmental conditions. This was initially described by van Belleghem et al. (2018) to explain the rapid parallel evolution of post-glacial ecotypes of salt-marsh beetle, which have repetitively evolved after the last glacial maximum, but for which the divergence of most alleles under selection can be traced back to a singular origin occurring about 190 kya. Our results suggest that similar processes may be involved in shaping the population structure of the European plaice. Indeed, the derived allele frequencies of the SVs were strongly associated with the North Sea–Baltic Sea transition zone that was connected to the Atlantic Ocean 8 kya (Le Moan et al. 2019), but the divergence of the alleles was estimated to be at least 25 times older than the age of the Baltic Sea itself. Moreover, the two SVs have similar estimated divergence times and show similar diversity patterns, suggesting that they share an evolutionary history and potentially have a common origin. Although our analyses did not completely clarify this origin, our study highlights an additional mechanism promoting the rapid reassembling of adaptive variation, i.e., the presence of SVs that maintain adapted alleles together.

Conclusions and perspectives

The SVs in European plaice are among a few examples of large SVs involved in the maintenance of population structure in marine fishes (e.g., Atlantic cod: Kirubakaran

et al. 2016; Capelin, Cayuela et al. 2020; Atlantic salmon: Lehnert et al. 2019; Atlantic herring, Pettersson et al. 2019; Atlantic silversides, Therkildsen et al. 2019; lingcod, Longo et al. 2020), but may be present in many other species (e.g., Australasian snapper, Catanach et al. 2019; lesser sandeel, Jimenez-Mena et al. 2019; sea horse, Riquet et al. 2019). Our data show the genomic heterogeneity of the European plaice population structure. While the differentiation outside the SVs was associated with demographic history and current connectivity patterns, the two SVs showed high variation in allele frequencies and were associated with more complex patterns of structure, which may be influenced by selection from standing genetic variation. Interestingly, two other fish species, the Atlantic cod and the Atlantic herring, show similar variation in SV frequencies, with one haplogroup increasing in frequency both in the Baltic Sea and the northern Atlantic (Barth et al. 2017, Kess et al. 2020 for cod and Pettersson et al. 2019 for herring, reviewed in Johannesson et al. 2020).

Additional experimental work focusing on the potential fitness effect of these SVs holds exciting perspectives for understanding their evolution and the role they may play in local adaptation and population structuring. Moreover, longer genomic sequencing reads, as provided by PacBio or nanopore technologies, could confirm if these SVs are chromosomal inversions. In addition, deeper sequencing coverage would allow an exploration of evolutionary signatures within the SV alleles that evolved in different geographical contexts. Such studies will provide an interesting framework to assess the evolutionary pathways involved in maintaining structure in this species where dispersal should normally limit population divergence.

Data availability

Individual raw fasta data, analyzed vcf file, and scripts (including the R scripts used for statistics and figures, and the two custom bash scripts used to extract the fasta file for the phylogeny and to extract the annotation information for the gene enrichment) are available as a Zenodo archive <https://zenodo.org/record/3724510>. The script used for the demographic inferences comprising the 4-step parameter optimization and the code for the tested models are available on https://github.com/alanlm-speciation/moments_optimization.

Acknowledgements We gratefully acknowledge Dorte Meldrup for her assistance with the library preparation, and Sigbjørn Mehl, Asgeir Aglen, Gróa Pétursdóttir, and late Jarle Mork for providing the samples from the coast of Norway, the Barents Sea and Iceland. We thank Paolo Momigliano and Juha Merilä for their support during the demographic analyses. Additionally, we thank Nicolas Bierne, Pierre-Alexandre Gagnaire, François Bonhomme and Thomas Broquet for their helpful advises in the early stage of this study, Romina Henriques

for her comments on the paper and Marjorlaine Rousselle for her advice on phylogenetic analyses, and Kerstin Johannesson and Carl André for support and discussions during later stages of the work. Version 5 of this paper has been peer-reviewed and recommended by Peer Community In Evolutionary Biology (<https://doi.org/10.24072/pci.evolbiol.100095>). We are grateful to associate editor Rui Faria, and the Peer Community in Evolutionary Biology, Maren Wellenreuther, and the three anonymous reviewers for the reviewing. This study received financial support from The European Regional Development Fund (Interreg V-A, project “MarGen”), the Ørsted Foundation and the Danish Ministry for Food, Agriculture and Fisheries and the European Union via the European Fisheries Fund (contract no. 33010-12-a-0253).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Barth JMI, Berg PR, Jonsson PR, Bonanomi S, Corell H, Hemmer-Hansen J et al. (2017) Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Mol Ecol* 26:4452–4466
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B Biol Sci* 263:1619–1626
- Belleghem SMV, Vangestel C, Wolf KD, Corte ZD, Möst M, Rastas P et al. (2018) Evolution at two time frames: polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLOS Genet* 14:e1007796
- Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJD (2016) Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol Ecol* 25:5073–5092
- Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol* 20:2044–2072
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Catanach A, Crowhurst R, Deng C, David C, Bernatchez L, Wellenreuther M (2019) The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost *Chrysophrys auratus*. *Mol Ecol* 28:1210–1223
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
- Cayuela H, Rougemont Q, Laporte M, Mérot C, Normandeau E, Dorant Y et al. (2020) Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Mol Ecol* 29:2379–2398
- Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157
- Cuveliers EL, Larmuseau MHD, Hellemans B, Verherstraeten SLNA, Volckaert FAM, Maes GE (2012) Multi-marker estimate of genetic connectivity of sole (*Solea solea*) in the North-East Atlantic Ocean. *Mar Biol* 159:1239–1253

- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Dobzhansky T (1970) *Genetics of the evolutionary process*. Columbia University Press, New York
- Duranton M, Allal F, Fraïsse C, Bierne N, Bonhomme F, Gagnaire PA (2018) The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nat Commun* 9:2518
- Erlandsson J, Östman Ö, Florin AB, Pekcan-Hekim Z (2017) Spatial structure of body size of European flounder (*Platichthys flesus* L.) in the Baltic Sea. *Fish Res* 189:1–9
- Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol* 23:347–351
- Faria R, Johannesson K, Butlin RK, Westram AM (2019) Evolving Inversions. *Trends Ecol Evol* 34:239–248
- Faria R, Chaube P, Morales HE, Larsson T, Lemmon AR, Lemmon EM et al. (2019) Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Mol Ecol* 28:1375–1393
- Farré M, Micheletti D, Ruiz-Herrera A (2013) Recombination rates and genomic shuffling in human and chimpanzee—a new twist in the chromosomal speciation theory. *Mol Biol Evol* 30:853–864
- Feder JL, Gejji R, Powell THQ, Nosil P (2011) Adaptive chromosomal divergence driven by mixed geographic mode of speciation. *Evolution* 65:2157–2170
- Feldman MW, Otto SP, Christiansen FB (1996) Population genetic perspectives on the evolution of recombination. *Annu Rev Genet* 30:261–295
- Gagnaire PA, Broquet T, Aurelle D, Viard F, Souissi A, Bonhomme F et al. (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evol Appl* 8:769–786
- Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. *Mol Ecol* 5:184–186
- Harding D, Nichols JH, Tungate DS (1978) The spawning of the plaice (*Pleuronectes platessa* L.) in the southern North Sea and English Channel. *Rapp et Procès-Verbaux Des Réunions Du Cons Perm Int Pour l'Exploration de la Mer* 172:102–113
- Hemmer-Hansen J, Nielsen EE, Grønkjær P, Loeschcke V (2007a) Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (*Platichthys flesus* L.). *Mol Ecol* 16:3104–3118
- Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007b) Adaptive divergence in a high gene flow environment: *Hsc70* variation in the European flounder (*Platichthys flesus* L.). *Hereditas* 99:592–600
- Hemmer-Hansen J, Therkildsen NO, Pujolar JM (2014) Population genomics of marine fishes: next-generation prospects and challenges. *Biol Bull* 227:117–132
- Hewitt G (2000) The genetic legacy of the quaternary ice ages. *Nature* 405:907
- Hoarau G, Coyer JA, Veldsink JH, Stam WT, Olsen JL (2007) Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Mol Ecol* 16:3606–3616
- Hoarau G, Rijnsdorp AD, Veer HWVD, Stam WT, Olsen JL (2002) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. *Mol Ecol* 11:1165–1176
- Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *J Comput Graph Stat* 5:299–314
- Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J et al. (2018) Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr Biol* 28:1839–1845
- Jenkins TL, Castilho R, Stevens JR (2018) Meta-analysis of northeast Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ* 6:e5684
- Jiménez-Mena B, Le Moan A, Christensen A, van Deurs M, Mosegaard H et al. (2019) Weak genetic structure despite strong genomic signal in lesser sandeel in the North Sea. *Evol Appl* 13:376–387
- Johannesson K, André C (2006) INVITED REVIEW: life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Mol Ecol* 15:2013–2029
- Johannesson K, Le Moan A, Perini S, André CA (2020) Darwinian laboratory of multiple contact zones. *Trends Ecol Evol* 35:1021–1036
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071
- Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J et al. (2012) The genomic basis of adaptive evolution in three-spine sticklebacks. *Nature* 484:55–61
- Jouganous J, Long W, Ragsdale AP, Gravel S (2017) Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics* 206:1549–1567
- Kess T, Bentzen P, Lehnert SJ, Sylvester EVA, Lien S, Kent MP et al. (2020) Modular chromosome rearrangements reveal parallel and nonparallel adaptation in a marine fish. *Ecol Evol* 10:638–653
- Kijewska A, Burzyński A, Wenne R (2009) Molecular identification of European flounder (*Platichthys flesus*) and its hybrids with European plaice (*Pleuronectes platessa*). *ICES J Mar Sci* 66:902–906
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge
- Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561–576
- Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLOS Biol* 8:e1000501
- Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* 173:419–434
- Kirubakaran TG, Grove H, Kent MP, Sandve SR, Baranski M, Nome T et al. (2016) Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol Ecol* 25:2130–2143
- Lehnert SJ, Bentzen P, Kess T, Lien S, Horne JB, Clément M et al. (2019) Chromosome polymorphisms track trans-Atlantic divergence and secondary contact in Atlantic salmon. *Mol Ecol* 28:2074–2087
- Le Moan A, Gagnaire PA, Bonhomme F (2016) Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol Ecol* 25:3187–3202
- Le Moan A, Gaggiotti O, Henriques R, Martínez P, Bekkevold D, Hemmer-Hansen J (2019) Beyond parallel evolution: when several species colonize the same environmental gradient. *BiorXiv*. <https://doi.org/10.1101/662569>
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends Ecol Evol* 17:183–189
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Limborg MT, Helyar SJ, De Bruyn M, Taylor MI, Nielsen EE, Ogden R et al. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Mol Ecol* 21:3686–3703
- Longo GC, Lam L, Basnett B, Samhoury J, Hamilton S, Andrews K et al. (2020) Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome. *Evol Appl* 00:1–19
- Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J et al. (2008) Evaluating signatures of glacial Refugia for North Atlantic Benthic Marine Taxa. *Ecology* 89:108–122

- Marques DA, Meier JI, Seehausen O (2019) A combinatorial view on speciation and adaptive radiation. *Trends Ecol Evol* 34:531–544
- Mérot C, Oomen RA, Tigano A, Wellenreuther M (2020) A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol* 35:561–572
- Momigliano P, Florin AB, Merilä J (2020) Biases in demographic modelling affect our understanding of the process of speciation. *bioRxiv*. <https://doi.org/10.1101/2020.06.03.128298>
- Morales HE, Faria R, Johannesson K, Larsson T, Panova M, Westram AM et al. (2019) Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Sci Adv* 5: eaav9963
- Morvezen R, Charrier G, Boudry P, Chauvaud L, Breton F, Strand Ø et al. (2016) Genetic structure of a commercially exploited bivalve, the great scallop *Pecten maximus*, along the European coasts. *Conserv Genet* 17:57–67
- Navarro A, Barton NH (2003a) Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science* 300:321–324
- Navarro A, Barton NH (2003b) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* 57:447–459
- Nelson TC, Cresko WA (2018) Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol Lett* 2:9–21
- Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T et al. (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evol Biol* 9:276
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246:96
- Oomen RA, Kuparinen A, Hutchings JA (2020) Consequences of single-locus and tightly linked genomic architectures for evolutionary responses to environmental change. *J Hered*. <https://doi.org/10.1093/jhered/esaa020>
- Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour* 13:946–952
- Perrier C, Charmantier A (2018) On the importance of time scales when studying adaptive evolution. *Evol Lett* 3:240–247
- Pettersson ME, Rochus CM, Han F, Chen J, Hill J et al. (2019) A chromosome-level assembly of the Atlantic herring genome—detection of a supergene and other signals of selection. *Genome Res* 29:1919–1928
- Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet* 8:e1002967
- Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5:92–102
- Portik DM, Leache AD, Rivera D, Blackburn DC, Rodel MO, Barej MF et al. (2017) Evaluating mechanisms of diversification in a Guineo-Congolian forest frog using demographic model selection. *Mol Ecol* 26:5245–5263
- Rijnsdorp AD (1991) Changes in fecundity of female North Sea plaice (*Pleuronectes platessa* L.) between three periods since 1900. *ICES J Mar Sci* 48:253–280
- Riquet F, Liautard-Haag C, Woodall L, Bouza C, Louisy P, Hamer B et al. (2019) Parallel pattern of differentiation at a genomic island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. *Evolution* 73:817–835
- Robledo D, Hermida M, Rubiolo JA, Fernández C, Blanco A et al. (2017) Integrating genomic resources of flatfish (Pleuronectiformes) to boost aquaculture production. *Comp Biochem Physiol Part D Genomics Proteomics* 21:41–55
- Rougemont Q, Gagnaire PA, Perrier C, Genthon C, Besnard AL, Launey S et al. (2017) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol* 26:142–162
- Rougeux C, Bernatchez L, Gagnaire PA (2017) Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of Lake Whitefish species pairs (*Coregonus clupeaformis*). *Genome Biol Evol* 9:2057–2074
- Shao C, Bao B, Xie Z, Chen X, Li B, Jia X et al. (2017) The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat Genet* 49:119–124
- Shin JH, Blay S, McNeney B, Graham J (2006) LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J Stat Softw* 16:1–9
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787–792
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Sørensen JG, Kristensen TN, Loeschcke V (2003) The evolutionary and ecological role of heat shock proteins. *Ecol Lett* 6:1025–1037
- The Heliconius Genome Consortium, Dasmahapatra KK, Walters JR, Briscoe AD, Davey JW, Whibley A et al. (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487:94–98
- Therkildsen NO, Wilder AP, Conover DO, Munch SB, Baumann H, Palumbi SR (2019) Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science* 365:487–490
- Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. *Heredity* 113:1–8
- Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST et al. (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* 5:5770
- Waples RS, Gaggiotti O (2006) INVITED REVIEW: what is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* 15:1419–1439
- 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. *Mar Biol* 157:447–462
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wellenreuther M, Bernatchez L (2018) Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol Evol* 33:427–440
- Wellenreuther M, Mérot C, Berdan E, Bernatchez L (2019) Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Mol Ecol* 28:1203–1209
- Wickham H, Winston C (2008) ggplot2: an implementation of the Grammar of Graphics. <http://CRAN.R-project.org/package=ggplot2>. Accessed 4 Nov 2019
- Wright S (1978) Evolution and the genetics of populations. University of Chicago Press, Chicago