Evolution at two-time frames shape structural variants and population structure of European plaice (*Pleuronectes platessa*)

Alan Le Moan[1], Dorte Bekkevold[1] & Jakob Hemmer-Hansen[1]

Affiliations

[1]National Institute of Aquatic Resources, Technical University of Denmark, Denmark

Abstract

The presence of differentially adapted alleles within the standing genetic variation of a species can fuel population diversification when that organism faces new environmental conditions. Structural variant (SV) polymorphisms provide evidence for this type of evolution at two-time frames, where ancient alleles later become associated with new environmental gradients. The Baltic Sea basin was connected to the Atlantic ocean after the last glacial maximum (8 kya), and currently represents an environmental gradient responsible for a major transition zone for regional marine life. The European plaice (*Pleuronectes platessa*) is a marine flatfish that has populations established within the Baltic Sea and show strong genetic differentiation with North Sea populations at two SVs. In this study, using a set of RAD derived sequencing SNPs, we show that these SVs are old, having evolved around 220 kya. Interestingly, in contrast with the rest of the genome, these SVs are, at best, weakly associated with an isolation-by-distance pattern. In fact, one of the SVs is polymorphic across most of the northern range of the European plaice distribution, and populations at the edge of the distribution show increased frequency of the derived allele. These findings suggest that neutral demographic processes, such as allele surfing, might be involved in explaining the distribution of the SVs polymorphism across the global species distribution. Nevertheless, the highest extent of differentiation at the SVs was associated with the North Sea - Baltic Sea transition zone, at geographical scales where genome-wide differentiation was barely detectable, highlighting their likely role for the diversification of European plaice within the Baltic Sea.

Introduction

Structural variants (SVs) are heritable modifications in the chromosome structure that can be caused either by changes in copy number (deletion, insertion and duplication), orientation (inversion) and position (translocation, fusion) of DNA sequences. SVs represent a significant part of the genetic variation between individuals and species (see special issues: Wellenreuther *et al.*, 2019). When a new structural variant allele arises, it occurs mostly in heterokaryotype individuals and does not (or only rarely) recombine (Kirkpatrick, 2010). Therefore, chromosomal rearrangements follow independent evolutionary pathways, often showing higher levels of population divergence than collinear regions of the genome (Farré *et al.*, 2013; Navarro and Barton, 2003a, 2003b). The SVs can harbour several genes and, thus, may have functional consequences for the organism (reviewed in Wellenreuther and Bernatchez, 2018). For instance, SVs are likely to evolve incompatible alleles, i.e. Bateson, Dobzhansky Muller incompatibilities (BDMi), which are responsible for decreased fitness in hybrids (Coyne and Orr, 1997; Rieseberg, 2001). It has been predicted that incompatibilities can become trapped in environmental gradients and/or by physical barriers to gene flow, resulting in allelic clines between otherwise fully connected populations (Barton, 1979). This effect can initiate speciation or reinforce pre-existing barriers already resistant to gene flow between species (Kirkpatrick and Barton, 2006, Butlin and Smadja, 2017). The accumulation of genomic divergence may eventually lead to the formation of new species, in a so-called chromosomal speciation (Faria and Navarro, 2010; Navarro and Barton, 2003a).

Structural variants are also important for evolving and maintaining locally adapted populations in the face of gene flow. In a system of interconnected populations, gene flow results in the rapid homogenisation 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 favourable in a local environment can be swamped by an unadapted genomic background (Lenormand, 2002). This migration load may prevent locally adapted mutations from increasing in frequency. Therefore, adaptation of the population could more likely to involve several mutations with similar effect, due to the redundancy of the genetic variability remaining at low frequency in the population (Yeaman, 2015). However, if a locally advantageous mutation is located within a SV that does not have a strong negative effect under local environmental conditions, the absence of recombination with adjacent maladapted genomic background could increase its fitness locally. In

comparison to independent genetic variants, mutations within SVs are less affected by the migration load and therefore more likely to show allele frequency clines along environmental gradients (Kirkpatrick and Barton, 2006). The strength of population differentiation will depend on the balance between drift-migration-selection, the effect of the latter being inflated if both locally adapted variation and BDMis 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 together within a SV (Dobzhansky, 1970; Feldman *et al.*, 1996), leading to the development of a supergene that further increases the adaptive potential of SVs (Thompson and Jiggins, 2014). These co-adaptations are expected to arise continuously *de novo* after the SV associates with an environmental barrier.

Structural variants have often been found to be an order of magnitude older than the age of the populations in which they are found, suggesting that their adaptive potential often relies on ancient polymorphisms (reviewed in Marques *et al.*, 2019; Wellenreuther and Bernatchez, 2018), in effect representing a source of standing variation for population divergence and adaptation. The evolutionary process for a SV has therefore been divided into two distinct time periods (i.e. “Evolution at two-time frames”, c.f. Belleghem *et al.*, 2018). The first period is associated with the initial formation and maintenance of SVs following their first occurrence within a species, while the second period is associated with the sometimes much later developed association with an environmental/physical/endogenous barrier to gene flow, observable in contemporary populations. Established SVs can thus promote the evolution of ecotypes following colonization of new environments, as described in systems undergoing parallel evolution. For instance, in the threespine stickleback (*Gasterosteus aculeatus*) freshwater ecotypes have repetitively evolved after postglacial recolonization (less than 15 kya) through similar genomic pathways, originating in the marine ancestral population and often involved inversions that have evolved millions of years ago (Jones *et al.*, 2012; Nelson and Cresko, 2018). Similar processes have been reported to be involved in shaping ecotypic evolution in multiple species from distant taxa (Wellenreuther and Bernatchez, 2018). While the origin of SVs often remains unclear, they sometimes originate from adaptive introgression of loci that have positive fitness effects for the introgressed species. For example, adaptive introgression of a major inversion is responsible for mimicry of the wing colour pattern between poisonous and non-poisonous

Heliconius butterflies (The Heliconius Genome Consortium *et al.*, 2012). The *Heliconius* inversion contains several genes that have co-evolved to result in strong colour patterns providing a warning signal to predators. This “supergene“ has evolved in the poisonous species over millions of years and has recently introgressed into the non-poisonous species which now benefits from the protection of the colour signal (Jay *et al.*, 2018).

The European plaice (*Pleuronectes platessa*) is a marine flatfish found from the Iberian Peninsula to the Barents Sea and Greenland in the eastern Atlantic. This species is thought to have colonized the northern range of its distribution (from the North Sea to Greenland) after the Last Glacial Maximum (LGM) (Hoarau *et al.*, 2002), and has successfully established populations within the westernmost parts of the Baltic Sea, a brackish environment formed 8 kya, which represents a low salinity environment for a marine species (Johannesson and André, 2006). The biological traits of plaice are generally expected to be associated with low levels of population divergence (Waples and Gaggiotti, 2006; Ward *et al.*, 1994): large effective population size (N_E) (reducing genetic drift), external fertilization, non-determinant spawning season and pelagic egg and larval phases (promoting gene flow). Previous studies based on relatively few genetic markers have found weak, and often 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), where depth acts as a strong physical barrier (Hoarau *et al.*, 2002; Was *et al.*, 2010). However, in a recent study of the genomic basis underlying the colonization of the Baltic Sea by four flatfish species, we identified two large polymorphic SVs (Le Moan *et al*., 2019) responsible for most of the observed differentiation between the North Sea and the Baltic Sea populations. However, whether these SVs evolved in response to the local condition of the Baltic Sea, or are ancient polymorphisms, remained unresolved. In the present study, we explored the distribution of the SVs in European plaice, across a larger geographical scale to examine multiple hydrographic gradients. The main goals of the study were thus to: i) re-assess the population structure in European plaice from northern Europe with the use of a population genomics approach; ii) evaluate the contribution of SVs to population structure, and iii) provide relevant data to understand the extent to which selection is involved in maintaining the allelic clines observed along the North Sea - Baltic Sea transition zone. Furthermore, we wanted to evaluate the presence of introgression signatures between plaice and its sister-species the European flounder *(Platichthys flesus),*

which are known to hybridize in the area (Kijewska *et al.*, 2009). The flounder is a euryhaline species better adapted to low salinity than the plaice, and can be found in freshwater lakes and in the innermost parts of the Baltic Sea where plaice does not occur (Hemmer-Hansen *et al.*, 2007). Therefore, it is possible that the European flounder is the source of the SVs found in plaice. Hence, we also used our data to: i) test for a potential flounder origin of SVs, and ii) determine the age of the SVs in plaice with a phylogenetic approach. This work will thus provide a better insight into the relative roles of environmental gradients, hybridization and genomic structural changes in the evolution of a widespread and highly abundant species.

Material and method

Geographic sampling

European plaice samples were collected at seven sites distributed across Northern Europe and Iceland (Figure 1A & table S1). Samples from the North Sea (Nors), Kattegat (Katte), the Belt Sea (Bels) and the Baltic Sea (Bals) were collected during the spawning season in 2016-2017. These samples were also the subject of the study by Le Moan *et al.* (2019) that studied the diversification process involved in the colonization of the Baltic Sea in several flatfish species. In that study, two large SVs were identified with allele frequency differences associated with the salinity gradient between the North Sea and the Baltic Sea. 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 (Bars), Norway (Norw) and Iceland (Icel) in 2013. Most of northern limit of the plaice distribution was covered with this sampling design. Analyses included 250 plaice in total, along with ten European flounder (*Platichthys flesus*; 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 outgroup in the phylogenic analyses.

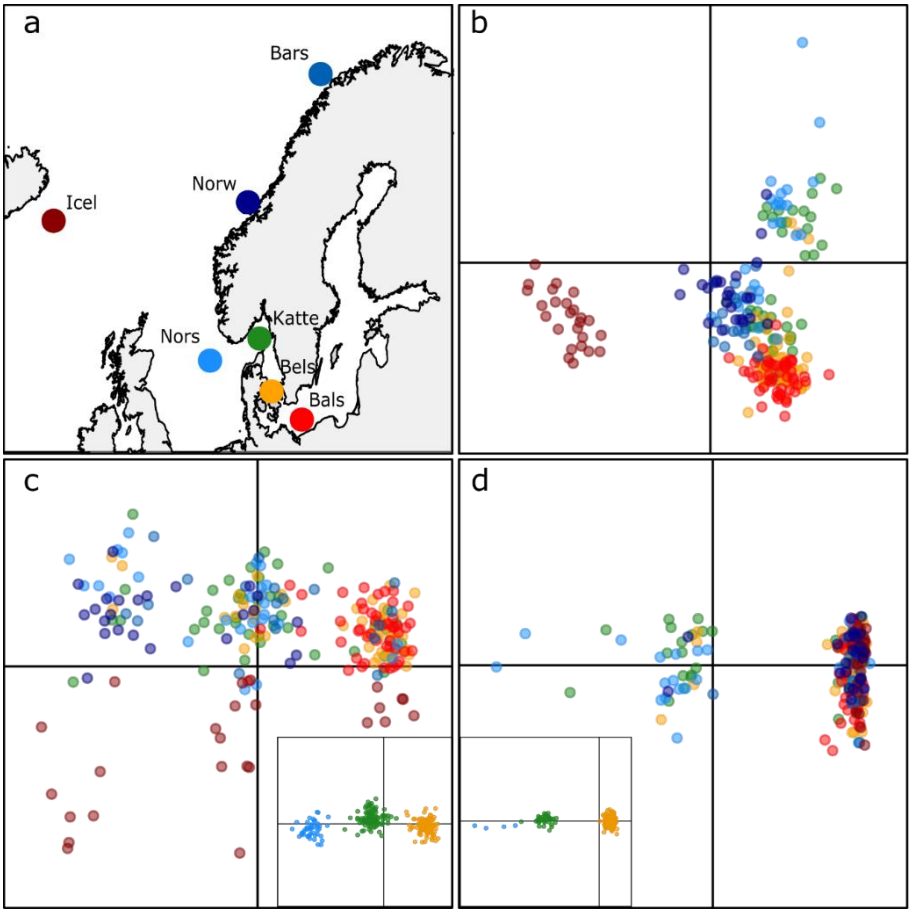

Figure 1: Sampling strategy (a) and multivariate analyses performed on individual diversity for the European plaice based on a genomic dataset of 3019 SNPs in the overall dataset (b), chromosome 19 (c) and chromosome 21 (d). Colours correspond to sampling sites represented on the map in a). Inserts in c) and d) show DAPC results grouping individuals by genotype, where green dots are the heterozygous individuals while the blue and yellow dots are individuals homozygous for the two respective haplotypes.

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 instructions manual. DNA extractions were standardized 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 450bp. 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). 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 (Figure 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 (Figure S1). SNPs were called based on the mapping results using the “gstacks” function with the “marukilow” model set with a minimum of coverage (m) of 5X to build a stacks of identical reads into a biological sequence, and alpha parameter of 0.05. Only bi-allelic SNPs present 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. 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 29X per samples (Figure 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 into account the differences in size selection: the “overall dataset” including all sampling sites, the “southern dataset” including Nors, Katte, Bels and Bals, and the “northern dataset” including Icel, Norw and Bars. These data sets were subsequently used for different analyses focusing on different aspects of population differentiation (see below).

Population structure

Population structure was assessed using the “overall dataset” which was thinned by removing loci with minor allele frequencies (MAF) below 5% and by keeping only one SNP per bin of 1kb to limit effects from physical linkage disequilibrium (LD). Individual genetic

diversity was visualized using PCA analyses, conducted with the R package adegenet (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 Spearman correlation test between pairwise F_{ST} and geographical distances among sampling sites and assessed with a Mantel test set after 9999 permutation with R base packages (Ihaka and Gentleman, 1996). All these analyses were conducted including and excluding the two chromosomes carrying the SVs, as well as only using the information from the two chromosomes.

We used an approximation-of-diffusion approach, as implemented in the software from δaδi (Gutenkunst *et al*., 2010), to examine the demographic histories associated with the major population breaks identified in the overall dataset. Specifically, we assessed the demographic history of Icel and its closest continental shelf population, Norw, because the Iceland population has been shown to be strongly differentiated from continental shelf populations (Hoarau *et al*., 2002). We used the “northern” dataset to compare four standard scenarios of demographic history: the Strict Isolation (SI), the Isolation-with-Migration (IM), the Ancestral Migration (AM), and the Secondary Contact (SC) models. These models were adjusted to the data using the folded version of the Joint Allelic Frequency Spectrum (JAFS) without considering singletons (-z option), using a modified version of δaδi from Tine *et al*. (2014). The best model was then selected based on its goodness of fit using the Akaike Information Criterion (AIC) selection.

Genotyping of the structural variants

The two chromosomes carrying the putative SVs (C19 and C21, and 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. Initial PCA plots showed clustering into three distinct groups (Figure 1) and hence suggested that each SV behaves like a Mendelian character with two haplotypes (Hap1 and Hap2) leading to three genotypes (homozygous for Hap1 or Hap2 and heterozygous). Consequently, we performed DAPC analyses with adegenet (Jombart and Ahmed, 2011)

set to three groups, to identify the genotype of each individual using the find.cluster function of adegenet. Then, haplotype allele frequencies for each population were calculated based on the DAPC clusters, using the formula

$$F = \frac{2 * C1 + C2}{2N}$$

Where C1 is the number of individuals assigned to one homozygous cluster and C2 the number of individual assigned to the heterozygous cluster in the DAPC, and N is the number of samples in the population. Several loci showed observed heterozygosity (H_O) of 1 in the heterozygous cluster (Figure S5) which also had F_{ST} = 1 between homozygous samples (Figure 3). These results are expected from SVs and therefore validated our genotyping procedure. Pairwise F_{ST} was then calculated between each population using the DAPC groups as genotype input independently for each SV in hierfstat (Goudet, 2005).

Genomic architecture of the structural variants

The genomic architecture of European plaice was characterized for the different pairs of populations with a particular focus on the two SVs. We analysed the subsets of the overall data (“northern” and “southern” dataset combined) to increase the genomic coverage along the SVs in these analyses. Nucleotide diversity (π) and H_O were calculated per SNP for each of the clusters inferred in the DAPC. The genomic architecture of differentiation was examined using SNP specific F_{ST} values. We used ggplot (Wickham and Winston, 2008) to represent and smooth the upper 5% F_{ST} quantile and average F_{ST} in different pairwise comparisons. Specifically, this differentiation was calculated for Norw vs. Bars, Norw vs. Icel using the northern dataset, and for Nors vs. Bals and the homozygous Hap1 vs. Hap2 of both SVs using the southern dataset. We focused on the upper quantile to limit effects from variable levels of variation across the SVs that would tend to depress average F_{ST} in certain regions along the SVs (Figure S6). Finally, we estimated LD by calculating the correlation between loci. These statistics were computed using vcftools (Danecek *et al*., 2011).

Gene content of the structural variants

In order to understand the genetic composition of the two SVs, we extracted the two SV sequences into two individual fasta files from the Japanese flounder genome using the same bins as defined for the phylogenic analyses below. These fasta files were aligned to the

Japanese flounder transcriptome (Shao *et al*., 2017) using blast (Johnson *et al*., 2008), and all genes with more than 80% mapping were recorded.

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 haplotypes and infer phylogenetic relationships, we only retained random subsets of the homozygous plaice individuals from the southern dataset. We focused on the southern dataset because it was the dataset 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 and one representing loci localized outside the SVs. Only ddRAD loci with sequence information for both haplotypes 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 filtrated southern dataset. The phylogeny of the first SV was based on loci located between 1.4 and 9.9 Mbp on chromosome 19, while loci located between 10.5 and 20.5 Mbp on the chromosome 21 were used for the second SV. The third data set consisted of loci located between 15 and 25 Mbp on chromosome 19 and represented the average genome-wide differentiation. The limits of the structural variants were defined based on the F_{ST} values between haplotypes, starting after the first and ending before the last SNP with $F_{ST} = 1$. The loci representative of the genome-wide divergence were selected to get a sequence size similar to that of the SVs. For each RAD locus, one random RAD allele per individual was kept. All alleles were concatenated into one pseudo-sequence using a custom script. The three phylogenies were inferred based on orthologous sequences of 10 125 bp, 6 152 bp and 10 341 bp for chromosomes 19, 21 and genome-wide, 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/haplotype, 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 (10^{-8}, The 1000 Genomes Project Consortium, 2010), and multiplied this value by 3.5, the average generation time of the Pleuronectidae (Erlandsson *et al*., 2017).

Results

Population structure

The first axis of the PCA explained 1.45% of the total inertia and distinguished the Icelandic plaice from all continental shelf individuals, and to a lower extent it also provided a separation of the marine Atlantic samples from the brackish samples of the Baltic Sea transition zone (Figure 1B). The second axis explained 1.2% of the inertia and was associated with differences among the remaining samples, roughly following 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 Baltic Sea transition zone (Katte H_O = 0.185/0.184, Bels H_O = 0.182/0.184 and Bals H_O = 0.178/0.181) as well as towards the northern populations (Norw H_O = 0.180/0.180 and Bars 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 (Katte vs Bels table S2). Pairwise comparisons including the Iceland population were homogeneously valued around F_{ST} = 0.03, (Figure 2, yellow dots), while all other pairwise F_{ST} were lower and correlated with geographical distance (r = 0.59, p = 0.01; excluding Iceland). Interestingly, pairwise F_{ST} values between the Barents Sea and the other samplings sites from the Baltics Sea and theTransition zone were lower than expected under an IBD scenario (Figure 2, blue dots), despite these being the most geographically distant sampling sites. This genetic similarity disappeared when the chromosomes carrying SVs were removed from analyses (Figure 2B). This removal resulted

in a higher correlation between genetic and geographic differences (r = 0.96, p = 0, excluding Iceland). However, this pattern was lost when only the chromosomes carrying SVs were analysed (r = 0.06, p = 0.89 for C19 & r = 0.11, p = 0.35 for C21, Figure S2 & Figure S3).

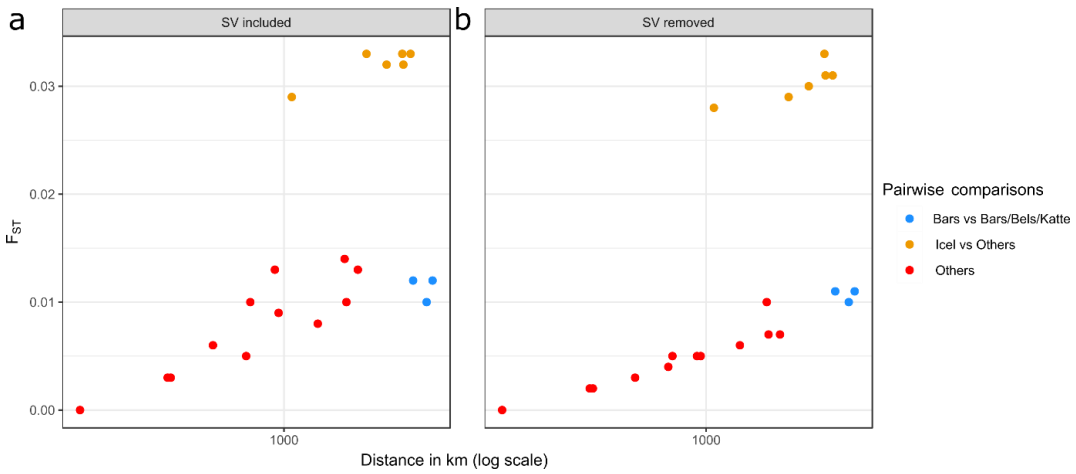

Figure 2: Relationships between geographic distance and genetic differentiation in European plaice sampling sites, including structural variants in a (r=0.56) and excluding structural variants in b (r=0.96).

The demographic modelling 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 (SC model – Figure S9 and table S5). The difference in AIC between this model and the second best (AM) was 2 912, which represents an extremely strong support for the SC among the tested models (Rougeux *et al.*, 2018). The estimate of the time of divergence, including the isolation phase, was ten times higher than the time under the secondary contact phase (table S5).

Genomic architecture of the structural variants

A high dispersion of samples from Nors and Katte (light blue and green samples) was observed on the second axis of the PCA (Figure 1B). This dispersion mostly involved SNPs from chromosome 21 carrying the structural variant SV21. The PCA based on this chromosome showed three clusters of samples which were also inferred in the DAPC analyses (Figure 1D). The other SV localized on chromosome 19 (SV19) showed similar patterns of clustering both in the PCA and the DAPC (Figure 1C). In both cases, the clusters on each side of the plots corresponded to homozygous individuals for the two haplotypes, the heterozygous being localized in the middle of the distribution.

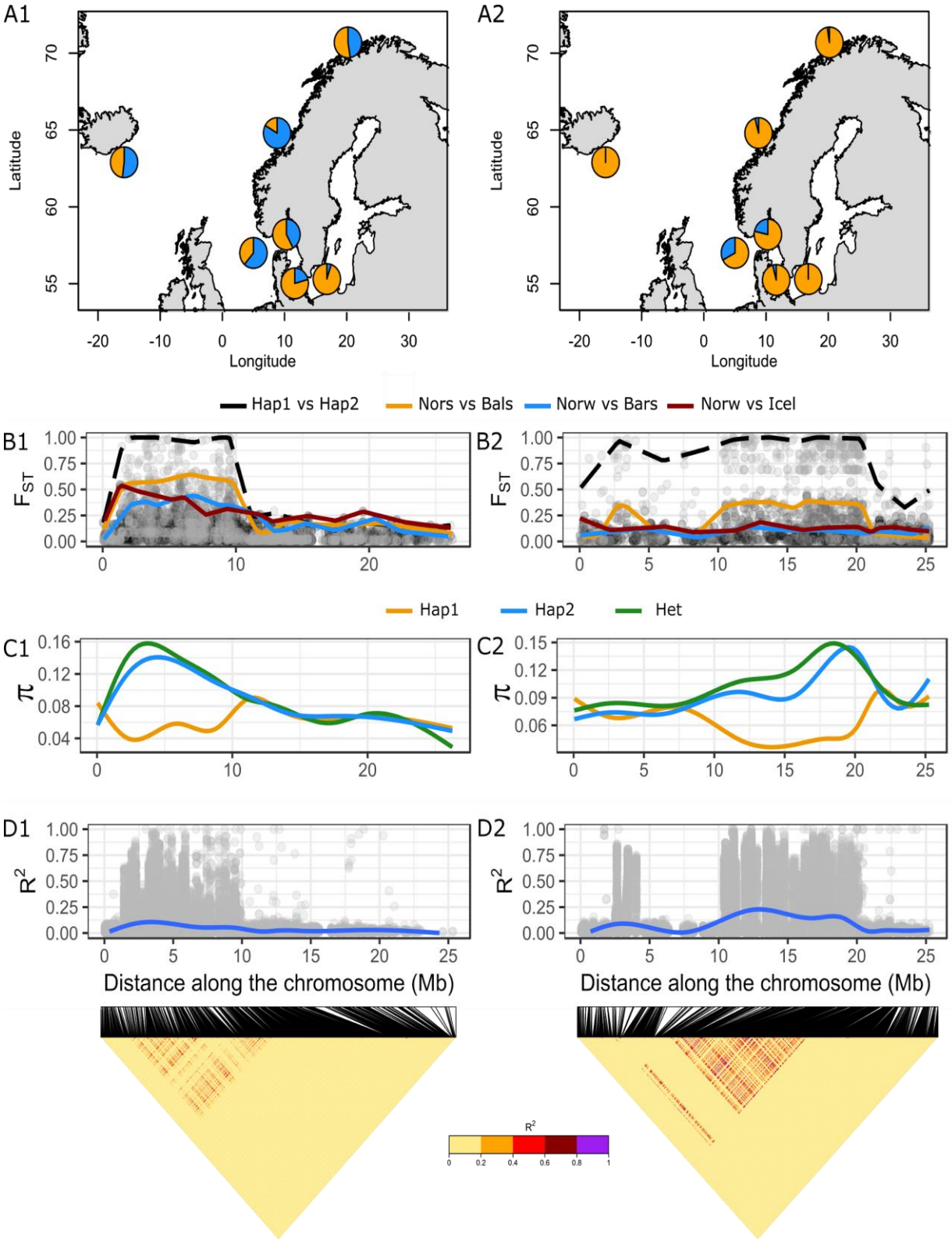

Figure 3: Sample specific data for the European plaice structural variants SV19 (left, 1) and SV21 (right, 2): A. sample specific allele frequencies (orange = Hap1 and blue = Hap2); B. differentiation for homozygous clusters (representing the two haplotypes) along the two chromosomes carrying the SVs, each dot is the F_{ST} value of an individual locus, and the solid line represents the smoothed upper 5% quantile (one colour = one comparison); C. smoothed average π for the three clusters identified in the DAPC; D. variation of LD along the chromosome and LD heatmaps for pairwise SNP comparisons.

The haplotype frequencies of SV19 and SV21 were inferred based on the DAPC clusters and were represented on the map (Figure 3A & Figure S4). Both SVs were polymorphic at most sampling sites. However, whereas SV19 haplotype frequencies were variable across most sampling sites, only Nors and Katte had both SV21 haplotypes in high frequency. Hence, the SV21 F_{ST} was elevated only in pairwise comparisons including Nors/Katte, while other comparisons also showed high F_{ST} for SV19 (Figure 3B). This high differentiation was evident across nearly one half of the chromosomes, from 1.5 Mbp to 10.5Mbp for SV19, and from 10.5 to 20.5 Mbp for SV21. The genome wide differentiation outside the SVs was generally lower than inside the SVs (Figure 3B and Figure S8). Several SNPs within the SVs were differentially fixed between homozygous individuals (as represented by the black dashed line in Figure 3B). Interestingly, a small decrease of F_{ST} appeared in the center of each SV. The small decrease in F_{ST} was aligned with a small increase of π (Figure 3C) which could be due to rare events of recombination between haplotypes. Nevertheless, strong LD occurred along the entire blocks, confirming that recombination between haplotypes is rare (Figure 3D). Only one of the haplotypes (in yellow in Figure 3) showed reduced genetic diversity whereas the other haplotype (in blue) had a higher genetic diversity than the average observed across the genome. As expected, the heterozygous individuals showed the highest diversity of the three DAPC groups. The two SVs were statistically more linked that the average genome-wide (0.2 vs. 0.07), but were less linked that the average LD within the SVs (Figure S7). The chromosome 21 showed nonetheless two peaks of LD separated by a distance of 5Mbp of limited LD (Figure 3D). The LD within the two regions were equivalent to the LD between these regions (Figure S7).

Gene content of the structural variant

We identified more than 900 genes located along both SVs (907 and 943 for SV19 and SV21, respectively). Several of the identified genes were involved in ion transport, while other gene functions were also identified, such as sexual recognition (Supplementary file I).

Phylogenetic analyses

The common dab was the most divergent species in all phylogenetic trees, with 0.024-0.025 substitution per site (9 million years ago - Mya) on average compared with the European flounder or the European plaice (Figure 4). All plaice individuals were equally distant from the flounder individuals based on the concatenated ddRAD loci representative of the

genome wide divergence, with an average of 0.011 substitutions per site (4 Mya, Figure 4). However, in both SV phylogenies the two haplotypes were clearly divergent, with an average distance of 0.0020 and 0.0017 substitutions per site for SV19 and SV21 (Figure 4), respectively. In each case, the deepest branches were observed for Hap1 which lead to a different estimate of divergence for each branch, around 550 and 220 thousand years ago (kya) for Hap1 and Hap2, 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, departure between the tree phylogeny and individual SNPs phylogeny of the structural variants (f4 = 0.004, p-value < 0.05). However, this signal was mostly carried by three loci from SV19 and one from SV21 that showed a strong mismatch from the phylogeny typology. In all cases, the flounder and the Hap1 of the plaice were nearly fixed for the same allele, which was different from the allele of the ancestral haplotype observed in the common dab and Hap2. Removing these SNPs lead to f4 statistic values not significantly different from 0. More data is necessary to assess if these mismatches between the SNPs trees and the phylogeny tree are due to a random process of allele sorting or represent a case of ancient introgression, pre-dating the formation of the SVs.

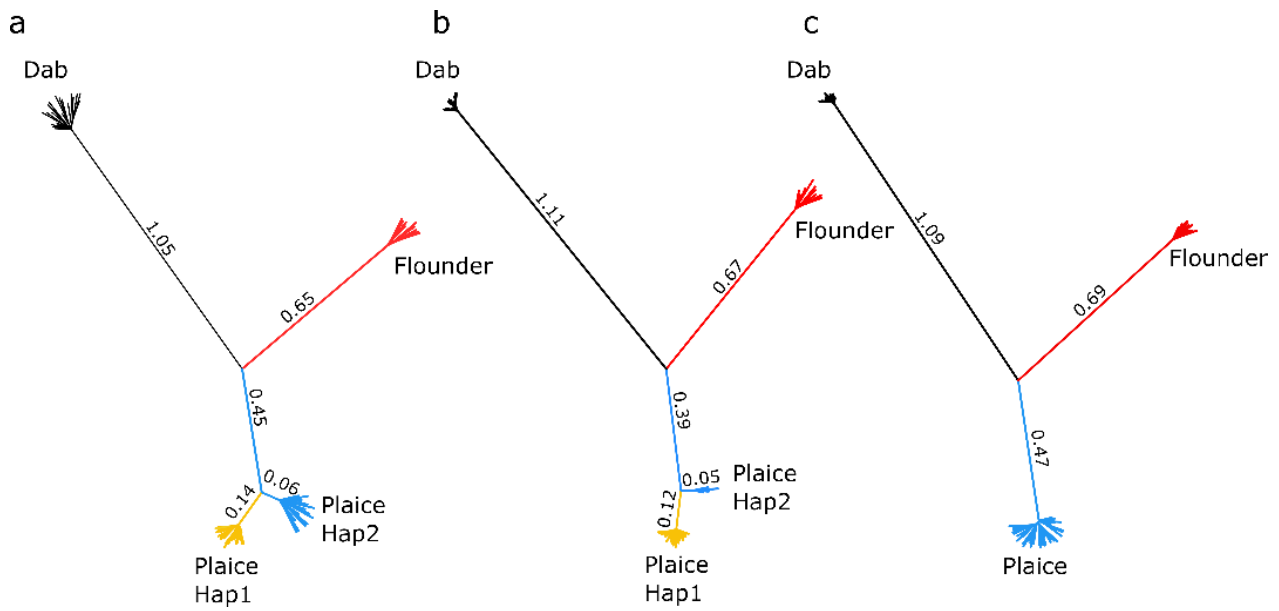

Figure 4: Plaice-flounder-dab phylogenetic trees for the ddRAD loci within SV19 (a), SV21 (b) and outside the SVs (c). The lengths of the branches reflect the frequency of numbers of substitutions per site (multiplied by 100 in the figures).

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 northeast Atlantic distribution range of the European plaice. Globally, the variation of haplotype frequencies among populations at both SVs result in important genomic differentiation, which is decoupled from the species' geographic distribution. The low genetic diversity of the Hap1 (orange, Figure 3C) in each SV and their long branches in the phylogenies suggest that they are the derived from of the SVs (Figure 4). The low diversity along the derived haplotypes, and its association with the edge the plaice distribution, suggest that these SVs are under some type of selective pressure, which is further supported by the deep branch inferred in the SVs phylogenies. Removing the SVs from the analyses leads to a strengthened pattern of IBD among European plaice samples, which to our knowledge is one of the clearest IBD patterns reported for a marine fish (r = 0.96). IBD was not detected in previous European plaice studies based on microsatellite markers (Hoarau *et al*., 2002; Was *et al*., 2010), and the results highlight once again the power of increasing genomic resolution to detect the subtleties of population structure for species with high gene flow.

Origin of the population structure in European plaice in the northern range of its distribution

The isolation of the Icelandic population previously described by Hoarau *et al*. (2002) was confirmed by our analyses. They hypothesized that differentiation was caused by a combination of effects from life on the edge of the distribution area and a founder effect following the postglacial recolonization of the coastal zone of the island, 10 kya. Hoarau *et al*. (2002) suggested that these differences were maintained by the deep oceanic regions separating the continental shelf and Iceland acting as a physical barrier to gene flow. However, this explanation does not fit well with the observed genetic diversity of the Icelandic population reported both with the microsatellite data (Hoarau *et al*. 2002) and in this study, which was somewhat higher than in most of the other populations postglacially established from the distributional edge (Barents Sea and Baltic Sea). Instead, the demographic analyses performed here suggest that the Icelandic population represents an old population, established from a different glacial refugium than the other populations of European plaice sampled in this study. 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 (LGM).

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 microsatellite studies (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. 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 observe an IBD (Kimura and Weiss, 1964). However, other processes might also be involved in maintaining a pattern of IBD (Jenkins *et al*., 2018). The most evident may be the effect of a population living on the edge of the species distribution range. Under this scenario, the reduction of genetic diversity from the North Sea to Barents Sea and from the North Sea to Baltic Sea are likely to be the result of two independent events of recent colonization, 15kya for the Barents Sea and 8 kya for the Baltic Sea. In this case, the general assumption is that colonization was initiated by only a subset of the North Sea population, which increased the effects of genetic drift, leading to the loss of genetic diversity in both the Barents Sea and Baltic Sea (Hewitt, 2000). In fact, both postglacial colonization and life on the edge effects were already suggested by Hoarau *et al*. (2002) to explain the continuous decrease of diversity in the area. Isolation-by-distance and edge effects are not mutually exclusive, and similar patterns have been reported for various species in association with the salinity gradient of the Baltic Sea (Johannesson and André, 2006), and along the South-North coast of Norway in taxa with lower dispersal capacities than the European plaice (Hoarau *et al*., 2007; Morvezen *et al*., 2016). However, to our knowledge, the observed pattern of IBD is the first example described in two independent postglacial recolonization routes within the same species.

The effect of structural variants on the population structure of European plaice

The two large SVs were polymorphic in most of the sites studied and were responsible for the main population differences observed here. The haplotype identified as the derived allele (with the lowest diversity and the highest divergence) reached fixation along the environmental gradient of the Baltic Sea in both SVs. However, the same haplotypes were

increasing in frequency towards the northern edge of the plaice distribution, in the Barents Sea, which was the most distant site from the Baltic Sea. These patterns resulted in a geographical decoupling of the SVs population structure from the population structure described genome-wide. The derived haplotypes occurred under various habitats 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. In principle, selection on any of the hundreds of genes within the SVs would result in haplotype frequency clines along the environmental gradients (Jay *et al*., 2018). In addition, the distributional edge of the plaice also represents a common feature associated with the increase of the derived haplotype frequency, potentially resulting in allele surfing effects in the marginal populations (Excoffier and Ray, 2008). In the North Sea-Baltic Sea transition zone, the ancestral haplotypes for SV19 disappeared in less than 300 km despite being highly frequent within the North Sea (>70/95 in Nors/Norw). Here, gene flow still affects the rest of the genome where differentiation is very low at this geographical scale (F_{ST} Katte vs. Bels not significantly different from 0). Since gene flow is acting homogeneously on the entire genome under pure migration-drift equilibrium, both haplotypes should be found in high frequency, as observed in the other range margin (Slatkin, 1987). Consequently, the steep haplotype clines identified along the Baltic Sea environmental gradient were established after the postglacial recolonization and may be maintained by (or coupled with) local adaptation (Kirkpatrick and Barton, 2006). Further work focusing on the functional consequences of these SVs that carry hundreds of genes (Supplementary file I) will help to understand their effect on the biology of the European plaice.

Origin and genomic architecture 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 9Mbp. These large linkage blocks are expected with chromosomal rearrangements such as inversions, duplications and translocations which can formally be distinguished by use of a linkage map or genome sequencing, but not with the reduced representation approach used in this study. However, our data filtration steps (filtering by heterozygosity) would have resulted in the loss of duplicated regions. Moreover, only large inversions are likely to recombine after being

rearranged through double crossover events between haplotypes (Andolfatto *et al*., 2001). Here, the increased diversity in the middle of the SVs associated with the reduction of F_{ST} between haplotypes strongly suggests that recombination still occurs, although rarely. Moreover, if most of the genome structure of the plaice is similar to that of the Japanese flounder genome, the central position of the LD block would also be unlikely by fission/fusion of chromosomes. Consequently, our data are consistent with the presence of major inversions in the genome. The observed second peak of LD on chromosome 21 could be due to the presence of a third small inversion. However, the similar value of LD within and between the two LD blocks suggest 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.

Interestingly, SV19 is polymorphic in both Iceland and continental shelf samples, presumably representing different glacial refugia, suggesting that the inversion polymorphism has been present as standing variation in European plaice for a long period of time. This hypothesis was confirmed by the deep haplotypes divergence observed in both SV phylogenies. For the shortest branch, the origin of the haplotype split was estimated to 220 kya (550 kya for the longest branch). These estimates are 20 to 50 times higher than the colonisation of northern Europe by the European plaice. Although these split dates are rough estimates and do not include effects from selection or recombination, the differences of an order of magnitude between the age of the populations and the age of the alleles suggest that the SVs are much older than the current populations in which they segregate. There is a growing evidence that ancient polymorphism may act as the fuel of adaptive divergence (reviewed in Marques *et al*., 2019; Wellenreuther and Bernatchez, 2018). In several cases, new alleles originate from adaptive introgression from sister taxa, such as observed in the *Heliconius* butterflies (Jay *et al*., 2018; The Heliconius Genome Consortium *et al*., 2012). In order to test this hypothesis of introgression as a possible source of the observed SVs, we used the European flounder, a euryhaline species adapted to low salinity and known to hybridize with European plaice, as a candidate for the source of the SVs. Under this hypothesis, the introgressed haplotype in the plaice should be less divergent from the flounder than the genome on average. However, we found the opposite pattern, with each haplotype being more than or as divergent from the flounder as the plaice-flounder divergence inferred from outside the SVs. Thus, a potential flounder origin of SVs was

rejected by the phylogenic data. It was also confirmed by the f4 statistics showing limited evidence of mismatches between the phylogenetic tree and individuals SNP topologies. Hence, our data suggest that the SVs originated after the split of the European flounder and the European plaice. However, other potential introgression sources, more closely related to the plaice than to the flounder, cannot be ruled out. For instance, introgressive hybridisation could have involved “ghost” species/populations which are now extinct. Interestingly, the two plaice SVs show similar divergence and diversity patterns, suggesting a common evolutionary history. Further analyses should thus be performed to fully understand the origin of the reported SVs in European plaice.

The evolution and maintenance of the structural variants

The substantial net divergence of the derived haplotype, which was approximately 2.5 times higher than in the ancestral haplotype in both SV phylogenies, may be linked to the accelerated rate of accumulation of deleterious mutations and background selection (Charlesworth, 1994; Cruickshank and Hahn, 2014; Duranton *et al*., 2018; Perrier and Charmantier, 2018; Faria *et al*. 2019). However, the accumulation of deleterious mutations would make it difficult for the derived allele to spread under drift equilibrium only, especially in species with large N_E (Ohta, 1973), like the European plaice. Therefore, some form of balancing selection is likely to have been involved in maintaining both haplotypes at the species level for a long period of time (reviewed in Faria *et al*., 2019; Wellenreuther and Bernatchez, 2018). The strength of balancing selection could be spatially varying in the ancestral marine pool, which is expected to maintain a stable polymorphism in a panmictic population living in a heterogeneous environment (Gagnaire *et al*., 2012; Nielsen et al., 2009). Alternatively, polymorphism could be maintained either through spatially varying selection across several populations, or through an inversion that carries genetic incompatibilities. Incompatibility-carrying inversions are expected to couple with other loci involved in local adaptation (Barton, 1979; Bierne *et al*., 2011; Kirkpatrick and Barton, 2006), hereby reinforcing population structure associated with environmental contrasts. Such incompatibilities may result in mosaic hybrid zones which could maintain the polymorphism in a heterogeneous environment (Fraïsse *et al*., 2016; Riquet *et al*., 2019; Simon *et al*., 2019). The two inversions in European plaice are large and may thus harbour both components of local adaptation and incompatibilities.

The structural variants in European plaice are among a few examples of large structural variants involved in the maintenance of population structure in marine fishes (Threespine stickleback: Jones *et al*., 2012; Atlantic cod: Kirubakaran *et al*., 2016; Atlantic salmon: Lehnert *et al*., 2019), but may be present in many other species (Australasian snapper: Catanach *et al*., 2019; sea horse: Riquet *et al*., 2019). In Atlantic cod, the inversion polymorphism is highly coupled with both the environment and geography and likely not completely independent from demographic history of the populations (Kirubakaran *et al.*, 2016). In other marine organisms with similar decoupling between inversion polymorphisms and geography as observed here, the inversions are often associated with genome-wide heterogeneous differentiation in collinear regions of the genome (Jones *et al.*, 2012; Morales *et al.*, 2018; Westram *et al.*, 2018), which can also reflect signatures of complex demographic histories (Belleghem *et al*., 2018; Le Moan *et al*., 2016; Rougemont *et al*., 2017; Rougeux *et al.*, 2017). In European plaice, the North Sea - Baltic Sea differentiation is associated with limited differentiation in the collinear regions of the genome (Figure S8) and several complex demographic models provided a poorer fit to the data than an isolation-with-migration model (Le Moan *et al.*, 2019). Hence, the plaice inversions could potentially be an example of the process examined in the theoretical study by Kirkpatrick and Barton (2006), who described the spread of differentially adapted alleles at inversions, but which has so far has only been demonstrated in few empirical studies. It thus makes the European plaice an interesting species to further studies of the effects of genomic inversions on colonization, adaptation and population structure of marine species, particularly for the populations living at the edge of the distribution.

Conclusions

Our data support the hypothesis that the structural variants described in European plaice have been influenced by evolution at two different time-frames (Belleghem *et al.*, 2018). The first followed their origins, estimated to be at least 220 000 year ago, while the second was associated with the contemporary distribution towards the edge of the plaice distributional area during a postglacial recolonization, which happened less than 15 000 years ago. The structural variants are likely two large chromosomal inversions with allele frequencies decoupled from geography, and it is likely that selection is involved in shaping the current geographical distribution of the alleles.

Additional experimental work focusing on the potential fitness effect of these structural variants holds exciting perspectives for understanding their evolution and the role they may be playing in local adaptation and population structuring. Moreover, deeper genomic sequencing coverage would allow an exploration of evolutionary signatures within the haplotypes that evolved in different geographical contexts. Finally, identifying the functional role of genes within the putative inversions would be a major step towards understanding the implication of these genomic regions for population divergence and local adaptation in the species. 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.

Acknowledgment

We gratefully acknowledge Dorte Meldrup for her assistance with the library preparation, and Jarle Mork, Sigbjørn Mehl, Asgeir Aglen, Gróa Pétursdóttir for providing the samples from the coast of Norway, the Barents Sea and Iceland. Additionally, we thanks Nicolas Bierne, Pierre-Alexandre Gagnaire and François Bonhomme for their helpful advises in the early stage of this study and Romina Henriques for her comment on the manuscript. This study received financial support from The European Regional Development Fund (Interreg V-A, project “MarGen”) and from Ørsted Foundation.

References

Andolfatto, P., Depaulis, F., Navarro, A., 2001. Inversion polymorphisms and nucleotide variability in Drosophila. Genet. Res. 77, 1–8. https://doi.org/10.1017/S0016672301004955

Barton, N.H., 1979. The dynamics of hybrid zones. Heredity 43, 341–359. https://doi.org/10.1038/hdy.1979.87

Belleghem, S.M.V., Vangestel, C., Wolf, K.D., Corte, Z.D., Möst, M., Rastas, P., Meester, L.D., Hendrickx, F., 2018. Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. PLOS Genet. 14, e1007796. https://doi.org/10.1371/journal.pgen.1007796

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. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Butlin, R.K., Smadja, C.M., 2017. Coupling, Reinforcement, and Speciation. Am. Nat. 191, 155–172. https://doi.org/10.1086/695136

Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., Wellenreuther, M., n.d. The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by three-fold in the marine teleost Chrysophrys auratus. Mol. Ecol. 0. https://doi.org/10.1111/mec.15051

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140. https://doi.org/10.1111/mec.12354

Charlesworth, B., 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. 63, 213–227. https://doi.org/10.1017/S0016672300032365

Coyne, J.A., Orr, H.A., 1997. “Patterns of Speciation in Drosophila” Revisited. Evolution 51, 295–303. https://doi.org/10.1111/j.1558-5646.1997.tb02412.x

Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23, 3133–3157. https://doi.org/10.1111/mec.12796

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., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dobzhansky, T., 1970. Genetics of the evolutionary process. Columbia University Press.

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., Gagnaire, P.-A., 2018. The origin and remolding of genomic islands of differentiation in the European sea bass. Nat. Commun. 9, 2518. https://doi.org/10.1038/s41467-018-04963-6

Erlandsson, J., Östman, Ö., Florin, A.-B., Pekcan-Hekim, Z., 2017. Spatial structure of body size of European flounder (Platichthys flesus L.) in the Baltic Sea. Fish. Res. 189, 1–9. https://doi.org/10.1016/j.fishres.2017.01.001

Excoffier, L., Ray, N., 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol. Evol. 23, 347–351. https://doi.org/10.1016/j.tree.2008.04.004

Faria, R., Johannesson, K., Butlin, R.K., Westram, A.M., 2019. Evolving Inversions. Trends Ecol. Evol. 34, 239–248. https://doi.org/10.1016/j.tree.2018.12.005

Faria, R., Navarro, A., 2010. Chromosomal speciation revisited: rearranging theory with pieces of evidence. Trends Ecol. Evol. 25, 660–669. https://doi.org/10.1016/j.tree.2010.07.008

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. https://doi.org/10.1093/molbev/mss272

Feldman, M.W., Otto, S.P., Christiansen, F.B., 1996. Population Genetic Perspectives on the Evolution of Recombination. Annu. Rev. Genet. 30, 261–295. https://doi.org/10.1146/annurev.genet.30.1.261

Fraïsse, C., Belkhir, K., Welch, J.J., Bierne, N., 2016. Local interspecies introgression is the main cause of extreme levels of intraspecific differentiation in mussels. Mol. Ecol. 25, 269–286. https://doi.org/10.1111/mec.13299

Gagnaire, P.-A., Normandeau, E., Côté, C., Hansen, M.M., Bernatchez, L., 2012. The Genetic Consequences of Spatially Varying Selection in the Panmictic American Eel (Anguilla rostrata). Genetics 190, 725–736. https://doi.org/10.1534/genetics.111.134825

Goudet, J., 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2010. Diffusion Approximations for Demographic Inference: DaDi. Nat. Preced. https://doi.org/10.1038/npre.2010.4594.1

Hemmer-Hansen, J., Nielsen, E.E., Grønkjær, P., Loeschcke, V., 2007. Evolutionary mechanisms shaping the genetic population structure of marine fishes; lessons from the European flounder (Platichthys flesus L.). Mol. Ecol. 16, 3104–3118. https://doi.org/10.1111/j.1365-294X.2007.03367.x

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907. https://doi.org/10.1038/35016000

Hoarau, G., Coyer, J.A., Veldsink, J.H., Stam, W.T., Olsen, J.L., 2007. Glacial refugia and recolonization pathways in the brown seaweed Fucus serratus. Mol. Ecol. 16, 3606–3616. https://doi.org/10.1111/j.1365-294X.2007.03408.x

Hoarau, G., Rijnsdorp, A.D., Veer, H.W.V.D., Stam, W.T., Olsen, J.L., 2002. Population structure of plaice (Pleuronectes platessa L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. Mol. Ecol. 11, 1165–1176. https://doi.org/10.1046/j.1365-294X.2002.01515.x

Ihaka, R., Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. J. Comput. Graph. Stat. 5, 299–314. https://doi.org/10.1080/10618600.1996.10474713

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M.Á., Nowell, R.W., Mallet, J., Dasmahapatra, K.K., Joron, M., 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. Curr. Biol. 28, 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072

Jenkins, T.L., Castilho, R., Stevens, J.R., 2018. Meta-analysis of northeast 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. INVITED REVIEW: Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Mol. Ecol. 15, 2013–2029. https://doi.org/10.1111/j.1365-294X.2006.02919.x

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9. https://doi.org/10.1093/nar/gkn201

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M.C., Myers, R.M., Miller, C.T., Summers, B.R., Knecht, A.K., Brady, S.D., Zhang, H., Pollen, A.A., Howes, T., Amemiya, C., Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, Baldwin, J., Bloom, T., Jaffe, D.B., Nicol, R., Wilkinson, J., Lander, E.S., Di Palma, F., Lindblad-Toh, K., Kingsley, D.M., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484, 55–61. https://doi.org/10.1038/nature10944

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. https://doi.org/10.1093/icesjms/fsp110

Kimura, M., 1983. The neutral theory of molecular evolution. Cambridge University Press.

Kimura, M., Weiss, G.H., 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. https://doi.org/10.1371/journal.pbio.1000501

Kirkpatrick, M., Barton, N., 2006. Chromosome Inversions, Local Adaptation and Speciation. Genetics 173, 419–434. https://doi.org/10.1534/genetics.105.047985

Kirubakaran, T.G., Grove, H., Kent, M.P., Sandve, S.R., Baranski, M., Nome, T., De Rosa, M.C., Righino, B., Johansen, T., Otterå, H., Sonesson, A., Lien, S., Andersen, Ø., 2016. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. Mol. Ecol. 25, 2130–2143. https://doi.org/10.1111/mec.13592

Lehnert, S.J., Bentzen, P., Kess, T., Lien, S., Horne, J.B., Clément, M., Bradbury, I.R., 2019. Chromosome polymorphisms track trans-Atlantic divergence and secondary contact in Atlantic salmon. Mol. Ecol. 28, 2074–2087. https://doi.org/10.1111/mec.15065

Le Moan, A., Gagnaire, P.-A., 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. https://doi.org/10.1111/mec.13627

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

Lenormand, T., 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 17, 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Maggs, C.A., Castilho, R., Foltz, D., Henzler, C., Jolly, M.T., Kelly, J., Olsen, J., Perez, K.E., Stam, W., Väinölä, R., Viard, F., Wares, J., 2008. Evaluating Signatures of Glacial Refugia for North Atlantic Benthic Marine Taxa. Ecology 89, S108–S122. https://doi.org/10.1890/08-0257.1

Marques, D.A., Meier, J.I., Seehausen, O., 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2019.02.008

Morales, H.E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A.M., Butlin, R., 2018. Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. bioRxiv 447854. https://doi.org/10.1101/447854

Morvezen, R., Charrier, G., Boudry, P., Chauvaud, L., Breton, F., Strand, Ø., Laroche, J., 2016. Genetic structure of a commercially exploited bivalve, the great scallop Pecten maximus, along the European coasts. Conserv. Genet. 17, 57–67. https://doi.org/10.1007/s10592-015-0760-y

Navarro, A., Barton, N.H., 2003a. Chromosomal Speciation and Molecular Divergence--Accelerated Evolution in Rearranged Chromosomes. Science 300, 321–324. https://doi.org/10.1126/science.1080600

Navarro, A., Barton, N.H., 2003b. Accumulating Postzygotic Isolation Genes in Parapatry: A New Twist on Chromosomal Speciation. Evolution 57, 447–459. https://doi.org/10.1111/j.0014-3820.2003.tb01537.x

Nelson, T.C., Cresko, W.A., 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evol. Lett. 2, 9–21. https://doi.org/10.1002/evl3.37

Nielsen, E.E., Hemmer-Hansen, J., Larsen, P.F., Bekkevold, D., 2009. Population genomics of marine fishes: identifying adaptive variation in space and time. Mol. Ecol. 18, 3128–3150. https://doi.org/10.1111/j.1365-294X.2009.04272.x

Ohta, T., 1973. Slightly Deleterious Mutant Substitutions in Evolution. Nature 246, 96. https://doi.org/10.1038/246096a0

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. Mol. Ecol. Resour. 13, 946–952. https://doi.org/10.1111/1755-0998.12129

Perrier, C., Charmantier, A., 2018. On the importance of time scales when studying adaptive evolution. Evol. Lett. 0. https://doi.org/10.1002/evl3.86

Pickrell, J.K., Pritchard, J.K., 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLOS Genet. 8, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Poland, J.A., Rife, T.W., 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5, 92–102. https://doi.org/10.3835/plantgenome2012.05.0005

Rieseberg, L.H., 2001. Chromosomal rearrangements and speciation. Trends Ecol. Evol. 16, 351–358. https://doi.org/10.1016/S0169-5347(01)02187-5

Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Béduneau, V., Briard, O., Ayari, T.E., Hochscheid, S., Belkhir, K., Arnaud-Haond, S., Gagnaire, P.-A., Bierne, N., 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. https://doi.org/10.1111/evo.13696

Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., Evanno, G., 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Mol. Ecol. 26, 142–162. https://doi.org/10.1111/mec.13664

Rougeux, C., Bernatchez, L., Gagnaire, P.-A., 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. https://doi.org/10.1093/gbe/evx150

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., Dong, Z., Gao, L., Gai, J., Sakamoto, T., Mo, S., Chen, W., Shi, Q., Li, H., Xiu, Y., Li, Y., Xu, W., Shi, Z., Zhang, G., Power, D.M., Wang, Q., Schartl, M., Chen, S., 2017. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. Nat. Genet. 49, 119–124. https://doi.org/10.1038/ng.3732

Simon, A., Arbiol, C., Nielsen, E.E., Couteau, J., Sussarellu, R., Burgeot, T., Bernard, I., Coolen, J.W.P., Lamy, J.-B., Robert, S., Skazina, M., Strelkov, P., Queiroga, H., Cancio, I., Welch, J.J., Viard, F., Bierne, N., 2019. Replicated anthropogenic hybridisations reveal parallel patterns of admixture in marine mussels. bioRxiv 590737. https://doi.org/10.1101/590737

Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. Science 236, 787–792. https://doi.org/10.1126/science.3576198

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

The 1000 Genomes Project Consortium, 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073. https://doi.org/10.1038/nature09534

The Heliconius Genome Consortium, Dasmahapatra, K.K., Walters, J.R., Briscoe, A.D., Davey, J.W., Whibley, A., Nadeau, N.J., Zimin, A.V., Hughes, D.S.T., Ferguson, L.C., Martin, S.H., Salazar, C., Lewis, J.J., Adler, S., Ahn, S.-J., Baker, D.A., Baxter, S.W., Chamberlain, N.L., Chauhan, R., Counterman, B.A., Dalmay, T., Gilbert, L.E., Gordon, K., Heckel, D.G., Hines, H.M., Hoff, K.J., Holland, P.W.H., Jacquin-Joly, E., Jiggins, F.M., Jones, R.T., Kapan, D.D., Kersey, P., Lamas, G., Lawson, D., Mapleson, D., Maroja, L.S., Martin, A., Moxon, S., Palmer, W.J., Papa, R., Papanicolaou, A., Pauchet, Y., Ray, D.A., Rosser, N., Salzberg, S.L., Supple, M.A., Surridge, A., Tenger-Trolander, A., Vogel, H., Wilkinson, P.A., Wilson, D., Yorke, J.A., Yuan, F., Balmuth, A.L., Eland, C., Gharbi, K., Thomson, M., Gibbs, R.A., Han, Y., Jayaseelan, J.C., Kovar, C., Mathew, T., Muzny, D.M., Ongeri, F., Pu, L.-L., Qu, J., Thornton, R.L., Worley, K.C., Wu, Y.-Q., Linares, M., Blaxter, M.L., ffrench-Constant, R.H., Joron, M., Kronforst, M.R., Mullen, S.P., Reed, R.D., Scherer, S.E., Richards, S., Mallet, J., Owen McMillan, W., Jiggins, C.D., 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487, 94–98. https://doi.org/10.1038/nature11041

Thompson, M.J., Jiggins, C.D., 2014. Supergenes and their role in evolution. Heredity 113, 1–8. https://doi.org/10.1038/hdy.2014.20

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir,

K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V.M., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5, 5770. https://doi.org/10.1038/ncomms6770

Waples, R.S., 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. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Ward, R.D., Woodwark, M., Skibinski, D.O.F., 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J. Fish Biol. 44, 213–232. https://doi.org/10.1111/j.1095-8649.1994.tb01200.x

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. https://doi.org/10.1007/s00227-009-1331-x

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wellenreuther, M., Bernatchez, L., 2018. Eco-Evolutionary Genomics of Chromosomal Inversions. Trends Ecol. Evol. 33, 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Wellenreuther, M., Mérot, C., Berdan, E., Bernatchez, L., n.d. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. Mol. Ecol. 0. https://doi.org/10.1111/mec.15066

Westram, A.M., Rafajlović, M., Chaube, P., Faria, R., Larsson, T., Panova, M., Ravinet, M., Blomberg, A., Mehlig, B., Johannesson, K., Butlin, R., 2018. Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. Evol. Lett. 2, 297–309. https://doi.org/10.1002/evl3.74

Wickham, H., Winston, C., 2008. ggplot2: An implementation of the Grammar of Graphics [WWW Document]. URL R package version 0.7, URL: http://CRAN. R-project. org/package= ggplot2 (accessed 4.11.19).

Yeaman, S., 2015. Local Adaptation by Alleles of Small Effect. Am. Nat. 186, S74–S89. https://doi.org/10.1086/682405

Supplementary material

Table S1 : Details of the sampling locations, with the corresponding site ID, sample sizes, longitude, latitude and the date where the samples were collected.

Location	Site ID	Sample size	Longiture	Lattitute	Dates of sampling
Iceland	Icel	25	62.9	-15.8	Feb. 2013
Barents Sea	Bars	25	70.7	20.2	Feb. 2013
Norway	Norw	25	64.8	8.8	Feb. 2013
North Sea	Nors	30	55.26	7.09	Feb. 2016
Skagerrak - Skagerak	Katte	50	57.63	9.30	Feb. 2017
Øresund-Belt Sea	Bels	50	54.65	10.64	Mar. 2016-2017
South-West of Baltic	Bals	50	54.98	14.54	Mar. 2017

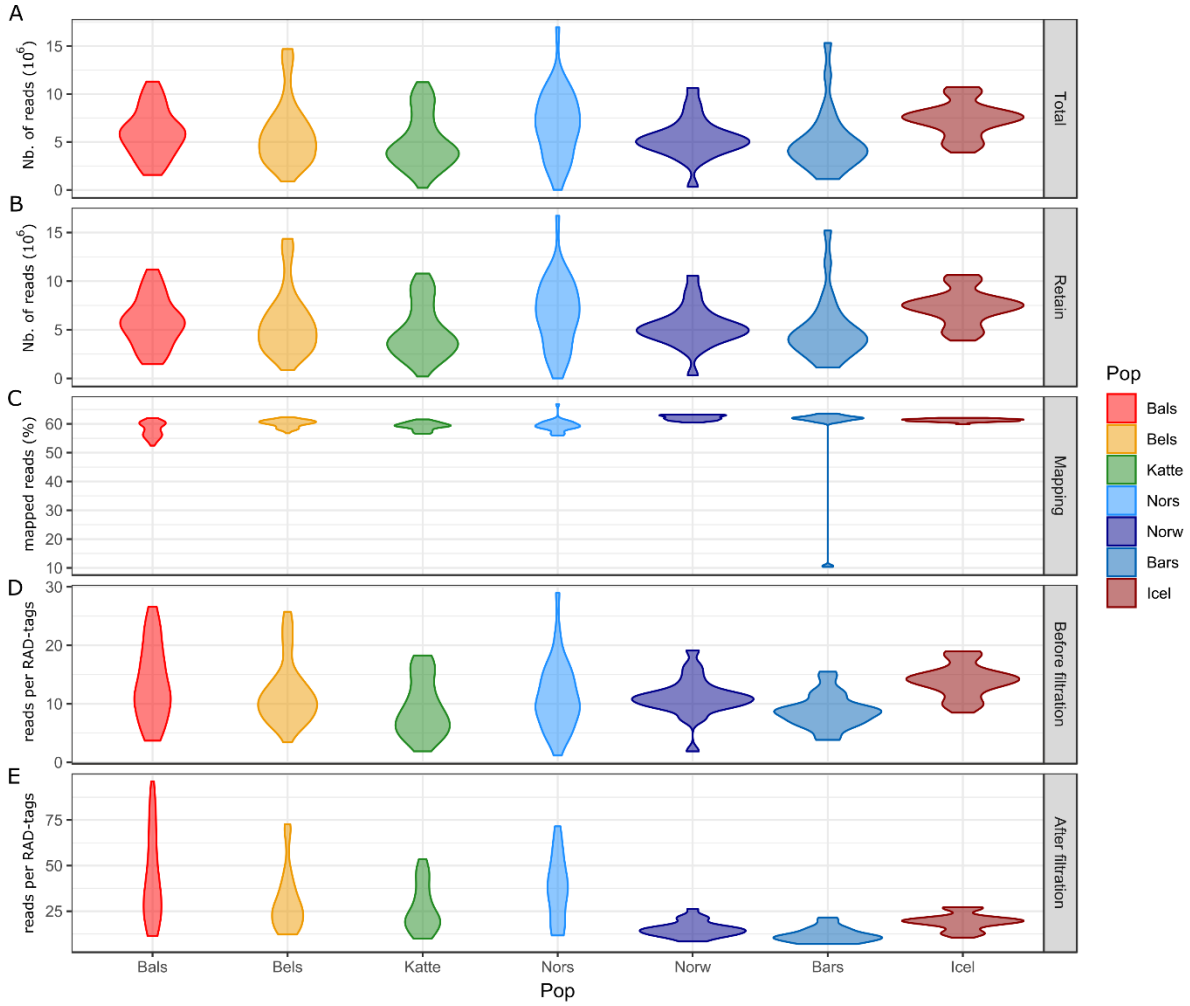

Figure S1: Violin plots of individual quality statistics of the data per population with A) the total number of reads, B) the number of reads with phred33 quality above 10, C) the percentage of reads mapped back to the Japanese flounder genome, D) the average coverage per RAD-tag and E) the average coverage per RAD-tag after data filtration

Table S1: Pairwise genomic differentiation (F_{ST}) among the European plaice samples with the SVs (above the diagonal) and with SVs removed (below the diagonal) calculated on the overall dataset. The significance levels are represented by * (p-value < 0.05)

	Katte	Bels	Bals	Nors	Norw	Bars	Icel
Katte	--	0.003*	0.006*	0.003*	0.008*	0.012*	0.032*
Bels	0.002*	--	0.000	0.010*	0.010*	0.010*	0.032*
Bals	0.003*	0.000	--	0.013*	0.013*	0.012*	0.033*
Nors	0.002*	0.005*	0.005*	--	0.009*	0.014*	0.033*
Norw	0.006*	0.007*	0.007*	0.005*	--	0.005*	0.029*
Bars	0.011*	0.010*	0.011*	0.010*	0.004*	--	0.033*
Icel	0.030*	0.031*	0.031*	0.029*	0.028*	0.033*	--

Table S2: Pairwise genomic differentiation (F_{ST}) among the European plaice samples based on the chromosomes carrying structural variants: chromosome 19 below the diagonal and chromosome 21 above the diagonal. The significance levels are represented by * (p-value < 0.05)

	Katte	Belts	Balts	Nors	Norw	Bars	Icel
Katte	--	0,023*	0,039*	0,014*	0,054*	0,036*	0,020*
Belts	0,009*	--	0,000	0,073*	0,018*	0,009*	0,002
Balts	0,034*	0,006*	--	0,098*	0,018*	0,010*	0,005*
Nors	0,011*	0,037*	0,074*	--	0,106*	0,075*	0,067*
Norw	0,036*	0,073*	0,120*	0,012*	--	0,017*	0,020*
Bars	0,005*	0,015*	0,038*	0,011*	0,050*	--	0,012*
Icel	0,047*	0,064*	0,092*	0,045*	0,000	0,050*	--

Table S3: Pairwise differentiation (F_{ST}) among the European plaice samples based on the DAPC genotypes for the SV19 (below diagonal) and SV21 (above diagonal)

	Katte	Belts	Balts	Nors	Norw	Bars	Icel
Katte	--	0.131	0.224	0.029	0.131	0.225	0.16
Betls	0.096	--	0.042	0.24	0.000	0.039	0.002
Balts	0.324	0.098	--	0.331	0.043	0.000	0.027
Nors	0.067	0.287	0.533	--	0.243	0.331	0.268
Norw	0.312	0.564	0.764	0.114	--	0.000	0.041
Bars	0.022	0.195	0.425	0.023	0.254	--	0.031
Icel	0.013	0.154	0.371	0.039	0.198	0.000	--

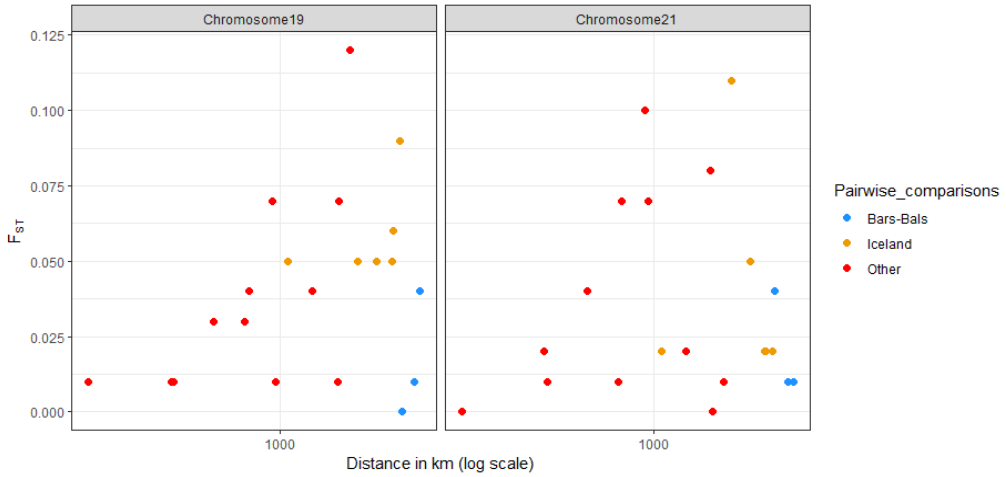

Figure S2: Relationships between F_{ST} and geographical distance in European plaice for the polymorphism data of the two chromosomes carrying SVs. Both relations are non-significant (R = 0.06, p = 0.89 for chromosome 19 and R = 0.11, p = 0.35 for chromosome 21)

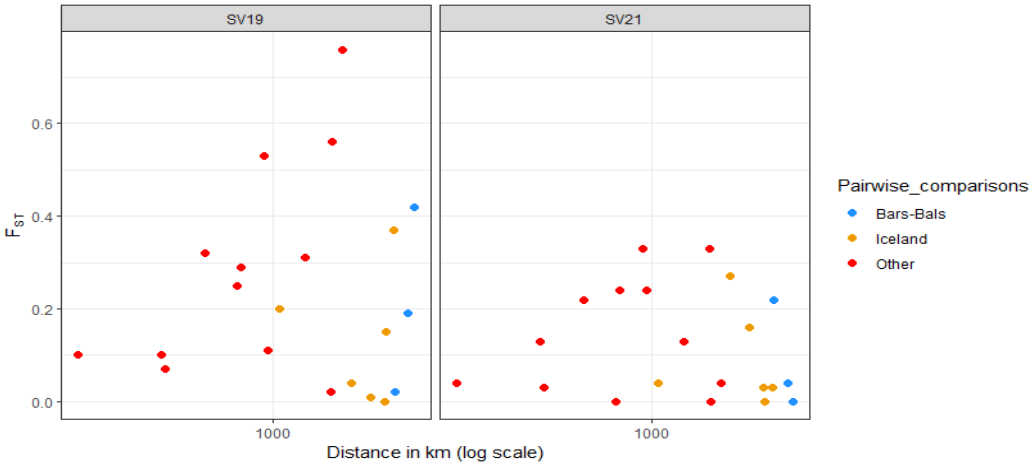

Figure S3: Relationships between F_{ST} for the genotypes of the structural variants inferred from DAPC and the log of geographical distance. Both relation are no significant (R = 0.00, p = 0.99 for the SV19 and R = -0.23, p = -0.29, for the SV21).

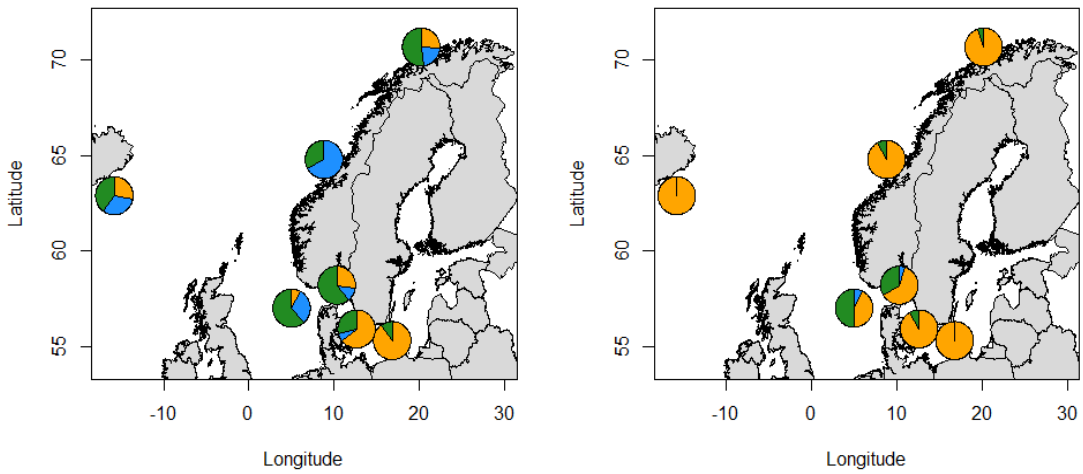

Figure S4: Map of the genotype frequencies across the different sampling sites for SV19 (left) and SV21 (right). The blue represent the homozygous for the minor allele frequency (= ancestral allele), the green represent the frequency of the heterozygous and the yellow represent the frequency of the major allele (= derived allele)

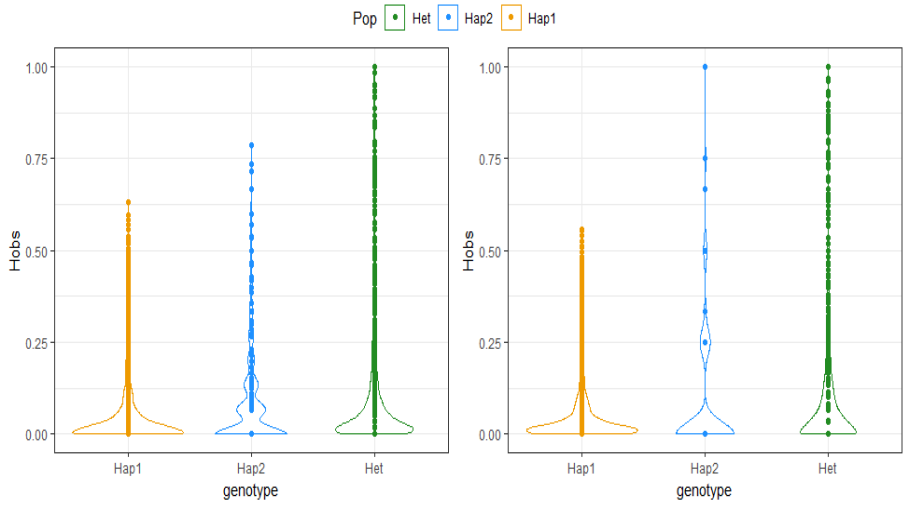

Figure S5: Boxplot of the observed heterozygosity for the DAPC clusters for the two chromosomes carrying SVs. The high variation of the homozygous individuals for Hap2 is due to the low number of individuals (n=4).

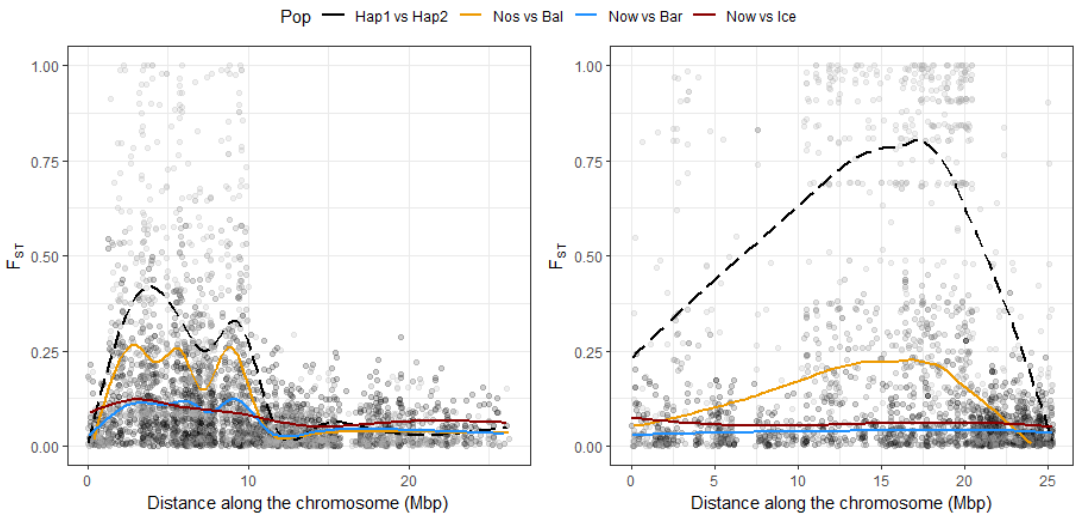

Figure S6: Estimated mean F_{ST} along the two chromosomes carrying SVs in the European plaice

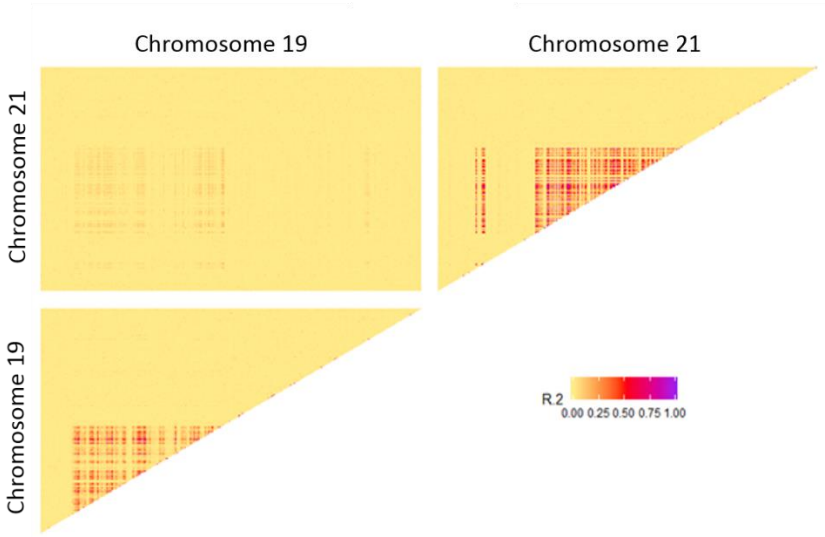

Figure.S7: Heatmap of Linkage Disequilibrium between SVs in the European plaice

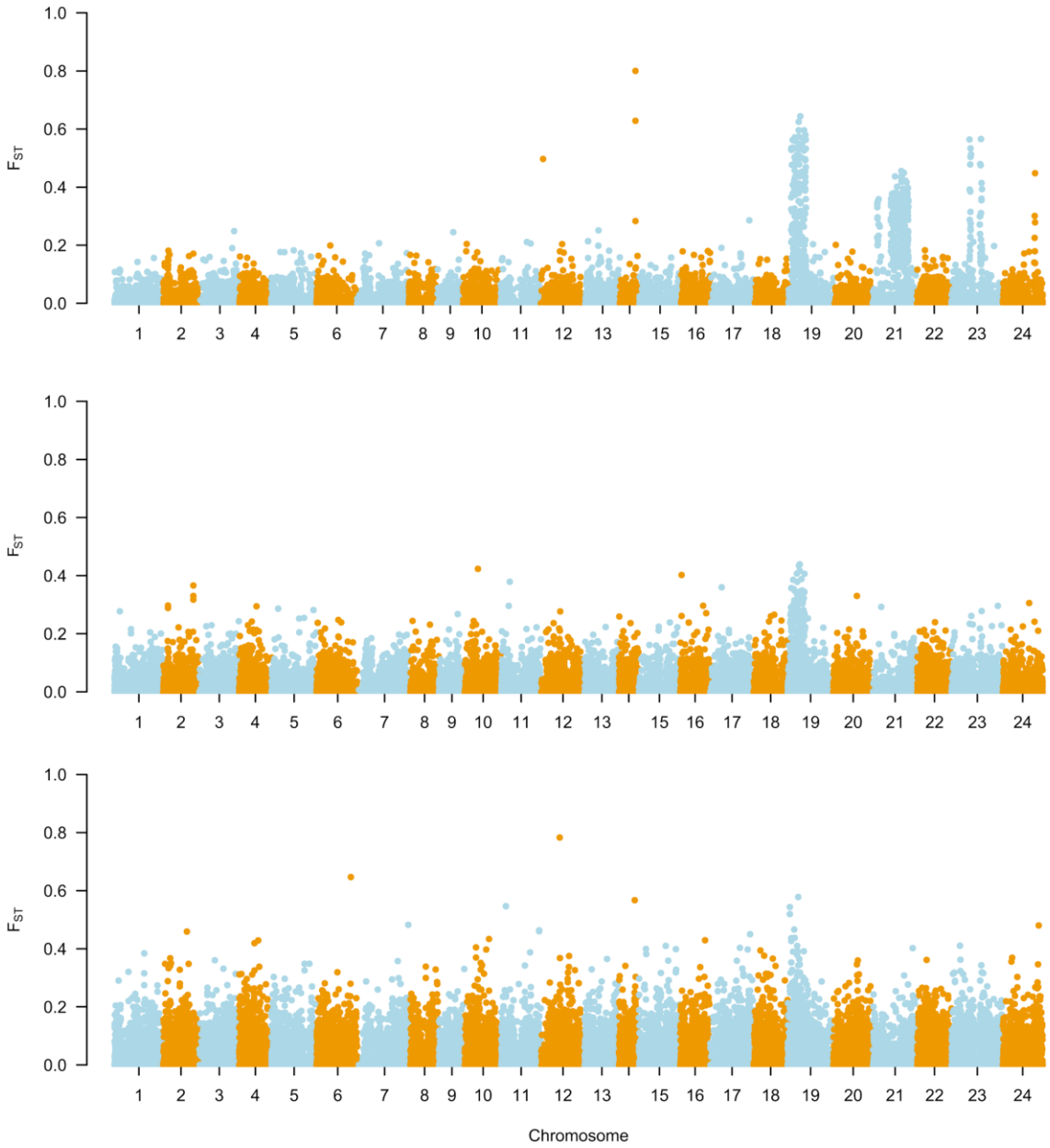

Figure S8: Manhattan plot representing the variation of F_{ST} for different pairwise comparisons in the European plaice. North Sea vs Baltic Sea (top), Norway vs Barents Sea (middle), and Norway vs Iceland (bottom). The two structural variants are located on chromosome 19 and 21.

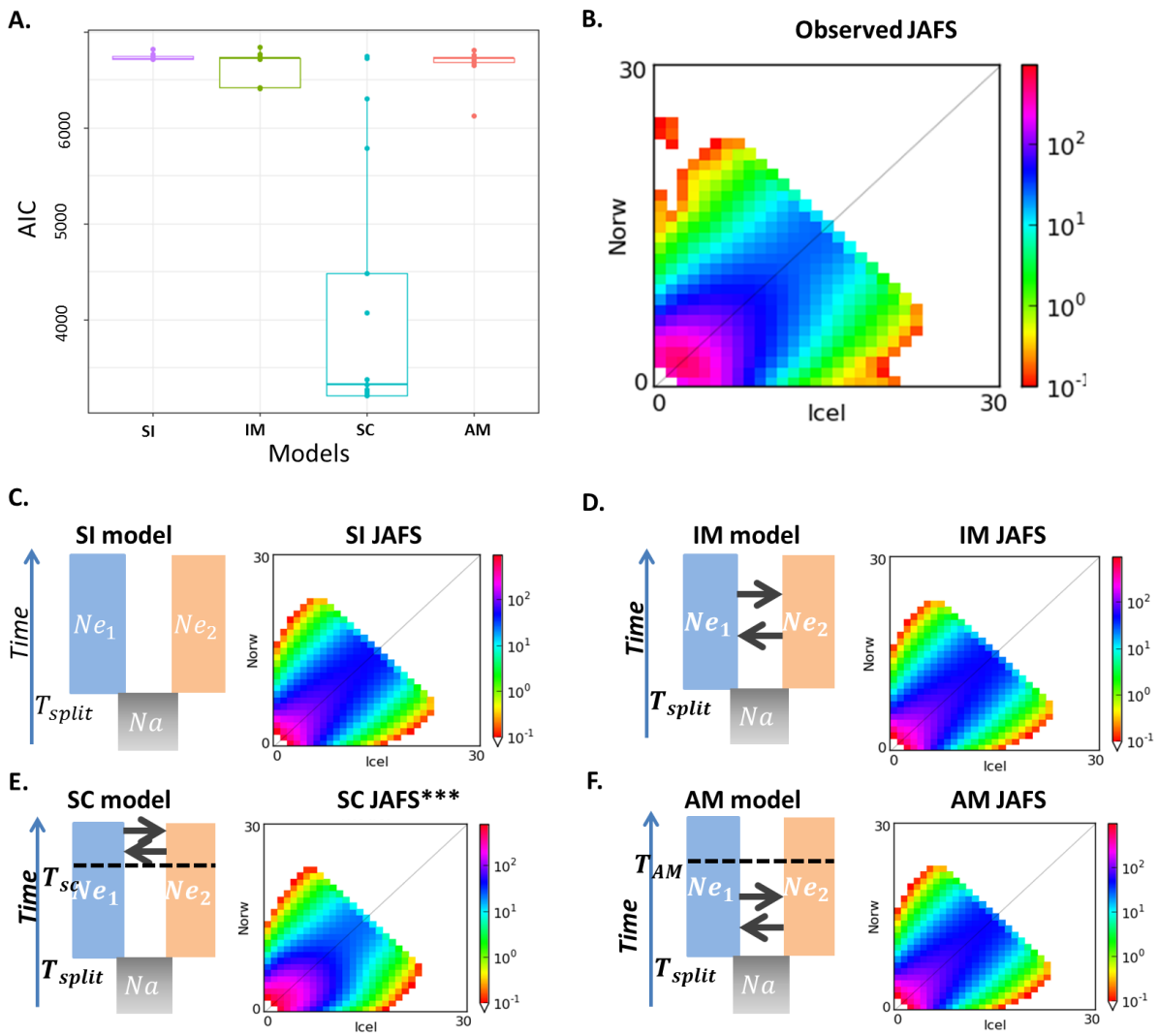

Figure S9: Demographic inferences between Norway and Iceland. A. boxplot of the AIC for the 20 replicates of inferences per model show a markedly better fit for the secondary contact model (SC). B. observed joint allelic frequency spectrum (JAFS) between Norway (Norw) and Iceland (Icel). C, D, E and F models SI, IM, SC, AM, respectively, with the theoretical JAFS of the best run for each model.

Table S5: Details of the pairwise inferences performed with δaδi for the 5 best fits for each model. The first columns shows the tested model, the following 3 columns details the statistics used for model selection, in order: the likelihood, the AIC and the delta AIC (relative to the best model). The remaining columns show the estimated parameters: Ɵ, the effective population size of Norway and Iceland (Ne_{Norw}, Ne_{Icel}), the migration rate ($m_{Norw>Icel}$, $m_{Icel>Norw}$) the time of split between the populations (T_S), the time of secondary contact and time of ancestral migration (T_{sc} / T_{am}). The best model is highlighted in bold.

Models	Likellhood	AIC	Δ_{AIC}	Ɵ	Ne_{Norw}	Ne_{Icel}	$m_{Norw>Icel}$	$m_{Icel>Norw}$	T_S	T_{sc} / T_{am}
SI	3357	6719	3509	5230	0.02	0.01	--	--	0.0004	--
SI	3357	6719	3509	5230	0.02	0.01	--	--	0.0004	--
SI	3357	6719	3509	5230	0.02	0.01	--	--	0.0004	--
SI	3357	6719	3509	5230	0.02	0.01	--	--	0.0004	--
SI	3357	6720	3510	5229	0.02	0.01	--	--	0.0004	--
IM	3199	6407	3197	5434	0.19	0.18	38.65	39.92	0.0185	--
IM	3199	6407	3197	5425	0.19	0.18	38.62	39.93	0.0181	--
IM	3199	6408	3198	5447	0.19	0.19	39.38	39.94	0.0195	--
IM	3199	6409	3199	5431	0.19	0.18	39.91	38.84	0.0182	--
IM	3200	6410	3200	5426	0.20	0.19	39.16	39.65	0.0191	--
AM	3055	6122	2912	752	0.59	10.01	13.68	1.35	0.990	0.000
AM	3320	6651	3441	5245	0.18	0.10	0.00	59.94	0.006	0.000
AM	3327	6666	3456	5272	0.17	0.16	59.97	0.00	0.008	0.000
AM	3333	6678	3468	5259	0.12	0.08	59.58	0.95	0.004	0.000
AM	3337	6686	3476	5239	0.09	0.06	0.01	59.55	0.003	0.000
SC	**1599**	**3210**	**0**	**1524**	**1.55**	**0.20**	**5.11**	**48.50**	**0.825**	**0.068**
SC	1599	3210	0	1322	1.79	0.23	4.47	41.69	0.968	0.078
SC	1599	3211	1	1314	1.82	0.22	4.41	41.97	0.969	0.078
SC	1600	3211	1	1416	1.67	0.21	4.65	43.78	0.901	0.075
SC	1600	3212	2	1313	1.81	0.24	4.22	40.15	0.974	0.080