Limited Parallelism in Genetic Adaptation to Brackish Water Bodies in European Sprat and Atlantic Herring

Mats E. Pettersson [1,†], María Quintela[2,†], François Besnier[2], Qiaoling Deng[1], Florian Berg [2], Cecilie Kvamme[2], Dorte Bekkevold [3], Mai-Britt Mosbech[4], Ignas Bunikis[4], Roger Lille-Langøy[2], Iole Leonori [5], Andreas Wallberg [1], Kevin A. Glover[2], Leif Andersson [1,6,*]

[1]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

[2]Institute of Marine Research, 5817 Bergen, Norway

[3]DTU-Aqua National Institute of Aquatic Resources, Technical University of Denmark, Silkeborg, Denmark

[4]Uppsala Genome Centre, SciLifeLab, Uppsala, Sweden

[5]CNR IRBIM, Italian National Research Council, Institute for Marine Biological Resources and Biotechnology, 60125 Ancona, Italy

[6]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA

[†]Equal contribution.

*Corresponding author: E-mail: leif.andersson@imbim.uu.se.

Accepted: June 14, 2024

Abstract

The European sprat is a small plankton-feeding clupeid present in the northeastern Atlantic Ocean, in the Mediterranean Sea, and in the brackish Baltic Sea and Black Sea. This species is the target of a major fishery and, therefore, an accurate characterization of its genetic population structure is crucial to delineate proper stock assessments that aid ensuring the fishery's sustainability. Here, we present (i) a draft genome assembly, (ii) pooled whole genome sequencing of 19 population samples covering most of the species' distribution range, and (iii) the design and test of a single nucleotide polymorphism (SNP)-chip resource and use this to validate the population structure inferred from pooled sequencing. These approaches revealed, using the populations sampled here, three major groups of European sprat: Oceanic, Coastal, and Brackish with limited differentiation within groups even over wide geographical stretches. Genetic structure is largely driven by six large putative inversions that differentiate Oceanic and Brackish sprats, while Coastal populations display intermediate frequencies of haplotypes at each locus. Interestingly, populations from the Baltic and the Black Seas share similar frequencies of haplotypes at these putative inversions despite their distant geographic location. The closely related clupeids European sprat and Atlantic herring both show genetic adaptation to the brackish Baltic Sea, providing an opportunity to explore the extent of genetic parallelism. This analysis revealed limited parallelism because out of 125 independent loci detected in the Atlantic herring, three showed sharp signals of selection that overlapped between the two species and contained single genes such as *PRLRA*, which encodes the receptor for prolactin, a freshwater-adapting hormone in euryhaline species, and *THRB*, a receptor for thyroid hormones, important both for metabolic regulation and the development of red cone photoreceptors.

Key words: European sprat, genome assembly, whole genome sequencing, population structure, genetic adaptation, genetic parallelism.

Significance

An interesting question in genome biology is to which extent the same genes contribute to the same or similar genetic adaptation in different species. Here, we explore this question by a population genetic analysis of the European sprat and by comparing its genetic adaptation to brackish water with the corresponding adaptation in the closely related clupeid, the Atlantic herring. The results reveal limited genetic parallelism because we see overlapping signals of selection at three distinct loci harboring single genes, but the majority of loci detected in the European sprat do not overlap with those noted in the Atlantic herring.

Introduction

A proper description of population structure is essential for sustainable fisheries, as it allows for appropriate distribution of fishing pressure across stocks. The possibility to perform whole genome sequencing (WGS) now provides outstanding resolution to characterize population structure and genetic differentiation (Andersson et al. 2024). The European sprat (*Sprattus sprattus*), hereafter denoted sprat, is an economically important clupeid that constitutes the basis for a substantial fishery in Northern European waters. It inhabits an extensive range and is found all along the European Atlantic coastline, from Norway to Portugal, as well as in the Mediterranean Sea. Additionally, it has colonized several brackish environments, including the Baltic Sea, the Black Sea, and Landvikvannet in southern Norway. The latter is a former freshwater lake that became brackish after the opening of a small canal in the 19th century (Whitehead 1985; Eggers et al. 2014; Berg et al. 2018). This colonization of multiple brackish environments provides an ideal situation to study the genetics of salinity-related adaptation, as it increases the power to infer specific salinity-related genetic differentiation.

Currently, genetic differentiation among sprat populations has been assessed using reduced-representation sequencing approaches, resulting in a set of relatively internally homogenous genetic groups: (i) coastal Norwegian; (ii) northeast Atlantic including the North Sea, Celtic Sea, and Bay of Biscay; and (iii) the Baltic Sea (McKeown et al. 2020; Quintela et al. 2020). Likewise, distinct groups were found in the Adriatic and Black Seas (Quintela et al. 2020) as well as in Landvikvannet, Norway (Quintela et al. 2021). However, the sparseness of markers assessed in these studies, coupled with the lack of a reference genome, has precluded further characterization of the putative adaptive signals.

A major aim of this study has been to compare genetic adaptation of this marine species to a brackish environment with the corresponding process in the closely related Atlantic herring (*Clupea harengus*; estimated split ~10 million years ago; Jamsandekar et al. 2023) to explore to what extent genetic parallelism occurs in evolutionary processes. Both sprat and Atlantic herring have colonized the brackish Baltic Sea subsequent to the last glaciation, and recent studies in the latter species have revealed hundreds of loci showing strong genetic differentiation between herring from the Atlantic Ocean and the Baltic Sea (Martinez Barrio et al. 2016; Pettersson et al. 2019; Han et al. 2020). Although the species are phylogenetically relatively close, their split far exceeds the formation of the shared brackish waters, the Baltic Sea (approximately 12,000 years old; Lass and Matthäus 2008), and Landvikvannet, Norway (150 years old), ensuring that adaptation took place independently in each species. Furthermore, there is no reported evidence for any gene flow between sprat and herring.

Here, we provide a draft assembly of the sprat genome, whole genome pool-seq data from 19 populations spanning from the Baltic to the Black Sea and use the recently released MultiFishChip (Andersson et al. 2024) to genotype 369 specimens. We identify strong signals of genetic adaptation to brackish waters, including six putative inversions. Additionally, we identify three sharp peaks of divergence, involving single genes that each overlap with a similarly narrow signal in Atlantic herring, providing evidence for some genetic parallelism between the two species in regard to genetic adaptation to low salinity.

Results

Genome Assembly and Pool Resequencing

The primary assembly, based on 14 M PacBio CSS reads, is composed of 2,115 contigs (contig N50 = 1.0 Mb) with a total length of 971 Mb. This contig assembly was scaffolded based on 287 M HiC read pairs, using pin-HiC (v3.0.0; https://github.com/dfguan/pin_hic). It was thereafter manually curated using Juicebox (v1.11.08; Dudchenko et al. 2018) and a custom deduplication procedure based on mapped read depth and location of duplicated BUSCOs. This resulted in a total length of 785 Mb, with scaffold N50 = 23.4 Mb and scaffold L50 = 12 Mb. The total length is almost identical to the 786 Mb genome assembly of the Atlantic herring, including unplaced scaffolds (Pettersson et al. 2019).

The signal-to-noise ratio of the HiC mappings was unusually poor by modern standards, possibly due to the comparatively high diversity and apparent high rate of structural

variation between the chromosome copies, and the underlying PacBio assembly proved resistant to commonly used deduplication methods. This leads to some ambiguity regarding scaffold breaks in the final assembly. In particular, the largest scaffolds are substantially longer than expected, given the size distribution of the herring chromosomes. Thus, in a subset of the following analyses, we relied on liftover to the current herring reference assembly to organize the results. Overall, this liftover revealed a good correspondence between the two assemblies, with 19 of 26 herring chromosomes mapping near exclusively to a single sprat scaffold (supplementary figs. S1 and S2, Supplementary Material online). However, it also revealed that the two very large sprat scaffolds (s1111 and s1118) correspond to two (s1111: Chrs 21 and 23) and three (s1118: Chrs 8, 24, and 26) herring chromosomes, respectively, with the physical order along the sprat scaffolds strongly correlating with the order on herring chromosomes, i.e. herring Chr 21 maps to the beginning of s1111, while Chr 23 maps to the end. The Darwin Tree of Life project has recently published a chromosome-level assembly of the sprat genome (fSprSpr1.1; GCA_963457725.1) that confirms the presence of these fusions.

This scaffolded assembly was used to call single nucleotide polymorphisms (SNPs) in the 19 sample pools. The number of individuals per pool was in the range 16 to 24 (supplementary table S1, Supplementary Material online). The pools were sequenced to an average depth of 36× coverage of paired Illumina NovaSeq reads, resulting in 2.8 M SNPs retained after stringent filtration. Samples were collected from a broad geographic area (Fig. 1), from Nordfjord on the Norwegian coast in the North to the Celtic Sea in the West and the Black Sea in the Southeast.

Nucleotide Diversity

Since the PacBio assembly provides complementary primary and alternative contigs, corresponding to the two chromosome copies in the reference individual, for most of the genome, we can use these to estimate a genomic average for nucleotide diversity (π). Based on the median distribution of observed diversities in 362 alignment blocks (supplementary fig. S3, Supplementary Material online), we estimated average π to be 1.2%—four times the value in Atlantic herring (4)—but also note that the variance among blocks is substantial (SD = 1.1%), indicating that the average number is not representative of all genomic regions.

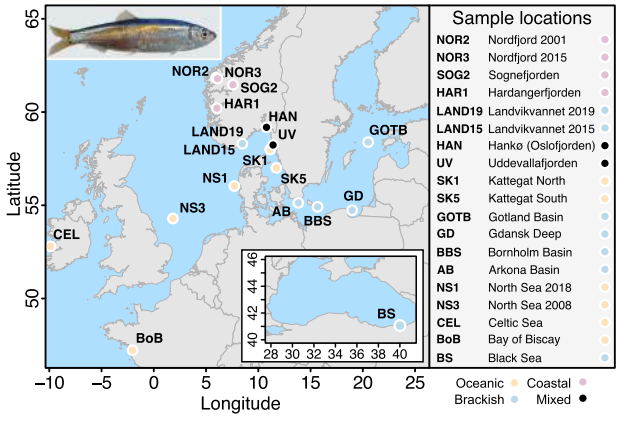

Fig. 1. Sample locations. Geographic location of the 19 population samples of European sprat (*S. sprattus*) used for pool-seq analysis. Photo: Merete Kvalsund, Institute of Marine Research, used with permission.

Overall Population Structure

Based on the approximately 2.8 million SNPs that passed quality filtering, we obtained the neighbor-joining tree shown in Fig. 2a, which shows that, based on the entire genome, samples group by habitat rather than geographic location. The three groups indicated in the tree correspond to distinct ecotypes: Oceanic, Coastal, and Brackish. These will form the basis for the genetic contrast reported below. The "Brackish" group comprises three separate localities, the Baltic Sea (samples Gotland Basin, Gdańsk Deep, Bornholm Basin, and Arkona Basin), the Black Sea, and Landvikvannet in Norway, and, in spite of their geographic dispersion, they form a well-defined clade. The "Oceanic" group consists of oceanic samples from Kattegat, North Sea, Celtic Sea, and Bay of Biscay whereas the "Coastal" group represents Norwegian coastal waters and fjords. The samples from Oslofjorden and Uddevallafjorden were placed between the Oceanic and Coastal groups, suggesting they may represent admixed populations, and, to reduce complexity, these two samples were therefore not included in the contrasts described below.

Contrasting the average allele frequencies in the three groups defined above reveals that the great majority of SNPs occur at very similar allele frequencies across populations, within the expected fluctuation due to sample size. However, the distribution contains a long tail of SNPs (supplementary fig. S4, Supplementary Material online). This mimics the situation in the Atlantic herring, in which the F_{ST} distribution among populations deviates significantly from the one expected for selectively neutral alleles under a genetic drift model (Lamichhaney et al. 2017).

To validate the patterns of genetic differentiation detected using pooled WGS, we employed the recently released MultiFishChip (Andersson et al. 2024), a resource developed to support cost-efficient typing of informative markers in several teleost species, to generate genotype information for 369 sprat individuals from 19 sampling sites largely overlapping the WGS sample set (supplementary table S1, Supplementary Material online). The sprat component of the MultiFishChip (Andersson et al. 2024) was designed based on the WGS data presented herein (see Materials and Methods), and the SNP designs are included as supplementary data S1, Supplementary Material online.

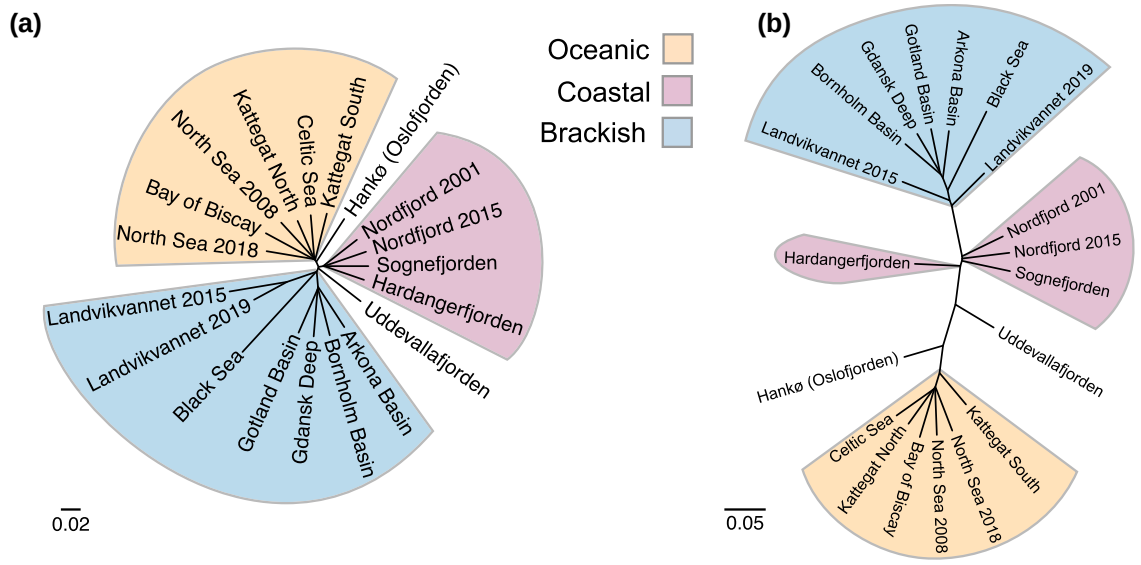

Fig. 2. Neighbor-joining tree based on allele frequency distances among pools. a) Based on neutral markers' genome. b) Using only highly differentiated markers (*DAF* > 0.5). The major groupings that have been used throughout the study are indicated

A principal component analysis (PCA) biplot built using 2,063 linkage disequilibrium (LD)-pruned SNPs differentiated the Oceanic, Coastal, Landvikvannet, Baltic, and Adriatic–Black Sea populations (Fig. 3a). Most of the individuals from the Uddevalla fjord (UV) clustered with samples from the Coastal population while a few joined the Oceanic samples, thus demonstrating that this is an admixed sample as suggested from pooled sequencing (see above). In contrast with the homogeneity of the Oceanic sprat, some individuals from Landvikvannet (LAND19) clustered with Coastal samples whereas other from the same location clustered with the brackish group, indicating that this is also a mixed sample. Discriminant analysis of principal components (DAPCs) revealed that the distribution of the individuals along the first axis (59.7% of the variation) seemed to follow a longitudinal gradient with the sample from the Black Sea occupying the farthest extreme (supplementary fig. S5a, Supplementary Material online), whereas the second axis (16.5%) discriminated the Baltic sprat and placed UV in an intermediate position between Oceanic and Baltic fish. The third axis (11.9%) further separated Landvikvannet sprat samples leaving the recent one (LAND19) closer to the Coastal sprat (supplementary fig. S5b, Supplementary Material online). However, the dendrogram coupled with the pairwise F_{ST} matrix revealed a first dichotomic division between brackish (plus the Adriatic Sea) and nonbrackish environments (Fig. 3b). Nonbrackish sprat was divided into Coastal and Oceanic with UV stemming from the Coastal branch. The genetic differentiation was essentially null among population samples within Coastal, Oceanic, and Baltic sprat, respectively (F_{ST} in the range 0 to 0.006). In Landvikvannet, the average differentiation toward the Coastal sprat was 0.078 in 2015 and circa half this amount (0.037) 4 years later. If we, instead of pruning by LD, use all high-quality markers selected from the high differentiation regions found in the WGS analysis ($n = 2{,}354$), the pattern somewhat changes; the dichotomy between Oceanic sprat and the remaining ones is emphasized, while the Black Sea comes closer to the Baltic Sea samples (Fig. 3c and d), mirroring the WGS results. The reason for this is that SNPs from the six inversions, shared between Baltic Sea and Black Sea populations, have a prominent impact on the pattern.

The outcome of STRUCTURE (Pritchard et al. 2000) analyzed through Evanno's test revealed $K = 2$ as the most likely number of genetic groups (supplementary fig. S6a, Supplementary Material online), which discriminated Landvikvannet, Baltic, Adriatic, and Black Sea from the Oceanic samples and leaving admixed Coastal sprat and hybrid zone sprat. In contrast, Puechmaille's statistics reported $K = 5$ (supplementary fig. S6b, Supplementary Material online) and revealed four rather homogenous clusters (Landvikvannet, Oceanic, Baltic, and Adriatic–Black Sea, respectively).

Genomic Regions of Differentiation among Subpopulations

Using the scaffolded version of the sprat genome assembly, we performed a genome-wide compilation of independent regions of differentiation, defined as having a gap of at least 500 kb to the next highly significant SNPs (Fig. 4). The threshold was chosen to minimize the risk of artificially dividing a region that represents a single signal at the

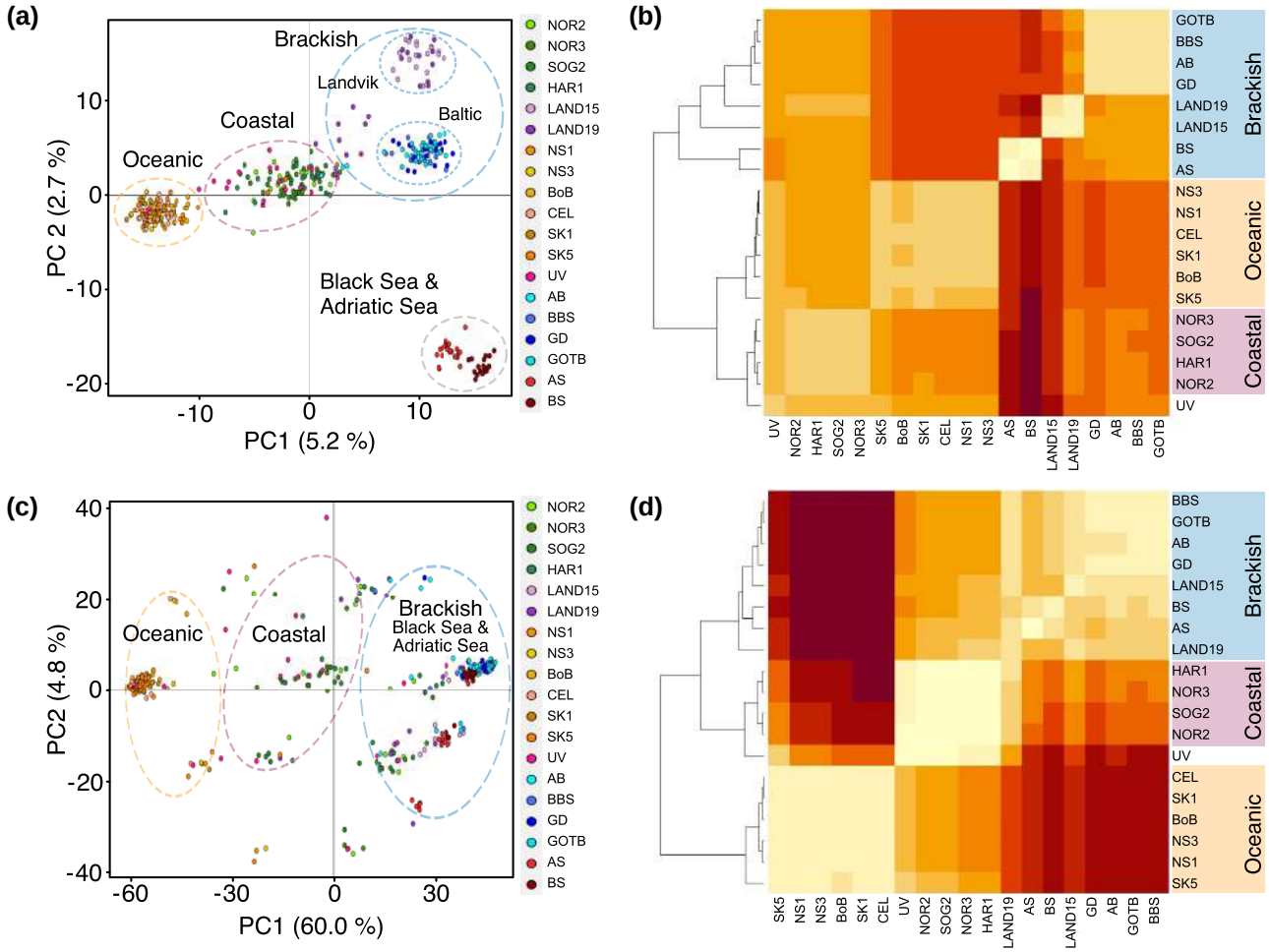

Fig. 3. SNP-chip–based PCA and F_{ST} heatmaps. a) PCA analysis based on 2,063 LD-pruned markers. b) Heatmap showing pairwise F_{ST} values, based on the markers used in a). c) PCA analysis based on 2,354 highly differentiated markers. d) Heatmap showing pairwise F_{ST} values, based on the markers used in c). Sample codes are given in Fig. 1 and supplementary table S1, Supplementary Material online.

expense of possibly combining nearby, but independent, signals. We estimated differentiation for two contrasts: Oceanic versus Brackish (OvsB) and Oceanic versus Coastal (OvsC). This yielded a total of 103 regions in the OvsB contrast. If we exclude regions consisting of a single SNP, this number drops to 68. The regions cover 14.0 Mb of the genome, with six putative inversions comprising the majority of that total. All identified regions are listed in supplementary table S2, Supplementary Material online. These six genomic regions were identified as putative inversions because they showed the characteristic of equally strong genetic differentiation over a large genomic region, on the megabase scale, and sharp borders to the flanking region not showing differentiation (Fig. 5).

The OvsC contrast contains more regions than the OvsB contrast: 286 in total, 138 if excluding single SNP regions, and also covering a substantially larger part of the genome (48.9 Mb). However, this should not be taken as evidence for a stronger genetic differentiation in this contrast. The lower delta allele frequencies (*DAFs*) of the inversion regions, in particular, lower the SD of the overall distribution in the OvsC contrast and thus the significance threshold (0.65 in OvsB vs. 0.50 in OvsC). In essence, the peaks of divergence in OvsC, while more numerous, are noticeably less differentiated than in OvsB (Fig. 4a and b). Furthermore, there is a strong overlap with the OvsB contrast, with a total of 11.2 Mb being called as significant in both contrasts. In general, the Coastal populations are more similar to the Brackish populations than to the Oceanic ones, in particular at strongly differentiated loci, causing less differentiation overall in the Brackish versus Coastal (BvsC) contrast (Fig. 4c), as well as a relatively strong correlation ($r^2 = 0.30$) between *DAF* values in the OvsC and OvsB contrasts (Fig. 4d). Nevertheless, there are signals specific to the Coastal populations, as can be seen by a cluster of SNPs with DAF in the range 0.4 to 0.8 in the OvsC contrast but $DAF < 0.1$ in the OvsB contrast (Fig. 4d). The majority of those SNPs are located within a

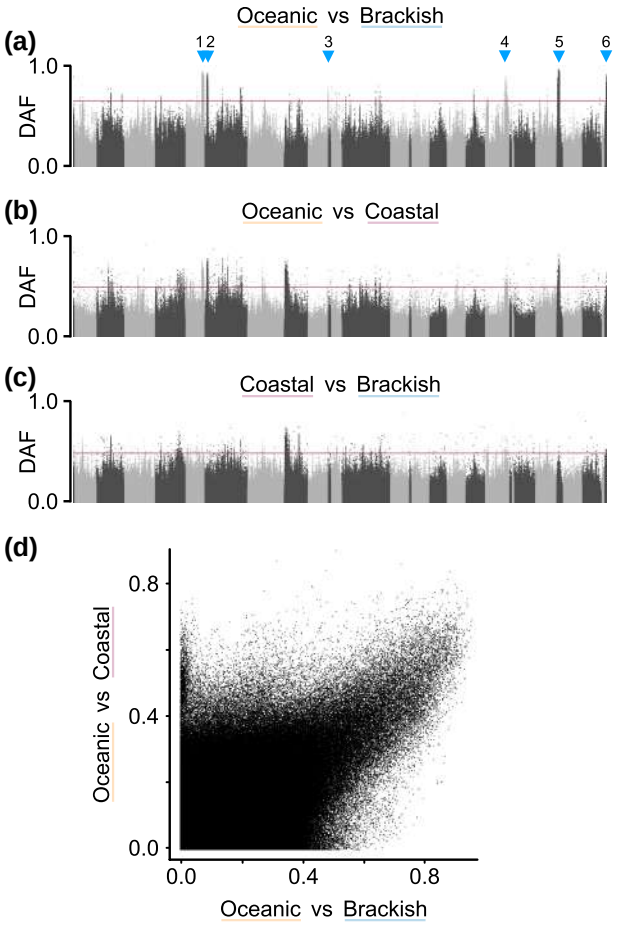

Fig. 4. *DAFs* of SNPs in three genome-wide contrasts. a) Oceanic versus Brackish populations. Arrows mark the location of six putative inversions. b) Oceanic versus Coastal. c) Coastal versus Brackish. d) Scatter plot of the SNP data from a) and b).

region on scaffold s1113, corresponding to 0.1 to 1.0 Mb on herring Chr 12. This region harbors another putative inversion, for which all Coastal populations have intermediate to high frequencies of a haplotype that is essentially missing from all other samples (supplementary fig. S7, Supplementary Material online).

We used the data from the SNP-chip analysis for a redundancy analysis (RDA) as an alternative method to identify loci potentially under selection, resulting in 84 detected outlier SNPs. Most of the outlier SNPs flagged by the RDA were associated with the environmental variable salinity ($n = 60$), followed by dissolved oxygen ($n = 11$), and temperature of sea surface ($n = 9$), whereas only two outliers were associated with pH and current velocity, respectively. Comparing the 60 salinity-related outliers with the WGS-based regions of differentiation, we note that 25 out of 103 regions contain at least one outlier, while 38 new loci are tagged by the remaining outliers. The outlier SNPs for dissolved oxygen and salinity are found in supplementary tables S3 and S4, Supplementary Material online, respectively, while supplementary tables S5 and S6, Supplementary Material online, contain the per-population frequencies of the same SNPs.

Six Putative Inversions Associated with Ecological Adaptation to Brackish Waters from the Black Sea to the Baltic Sea

We identified six putative megabase-scale inversions strongly associated with adaptation to brackish waters (Fig. 5). Inspection of allele frequencies for SNPs in the putative inversions revealed a common pattern, i.e. that the Coastal population samples were intermediate between the Oceanic and Brackish ones, and that in many cases, the Black Sea and the Baltic samples appeared to carry closely related haplotypes at high frequencies (Fig. 6; supplementary fig. S8, Supplementary Material online). This provides an explanation for the topology of the phylogenetic trees shown in Fig. 2, because many of the most differentiated SNPs are located within the putative inversion regions. Using only differentiated markers, defined as having DAF higher than 0.5 in any contrast, reinforces the "Brackish" and "Oceanic" groupings while positioning the "Coastal" samples along the main branch (Fig. 2b).

In addition to the signals of divergence consistent with the main groupings, we also detected other genome regions with strong differentiation between samples but with patterns that do not exactly conform to the three main groups. Instead, they are restricted to a subset of the brackish samples and include a region around 5.8 Mb on HiC scaffold 1144, where the signal is restricted to the Baltic Sea (supplementary fig. S9, Supplementary Material online), and a 1 Mb region from 18.5 to 19.5 Mb on HiC scaffold s4, where the Landvikvannet and Black Sea samples display similar allele frequencies that are distinct from that in other populations (supplementary fig. S10, Supplementary Material online).

Limited Genetic Parallelism between European Sprat and Atlantic Herring

The sprat and herring are closely related clupeids that show genetic adaptation to brackish waters. The liftover procedure allowed us to explore to which extent this adaptation shows genetic parallelism, i.e. that genetic variation in the same genes have contributed to adaptation. The main outcome of these analyses is that the observed parallelism is limited to a few loci. Firstly, none of the sprat's six putative inversions appears to have direct counterparts in the Atlantic herring, judged by the lack of block-like genetic differentiation between Atlantic and Baltic herring in the corresponding regions. Secondly, there are about 125 independent loci that show strong genetic differentiation between Atlantic and Baltic herring (Pettersson et al.

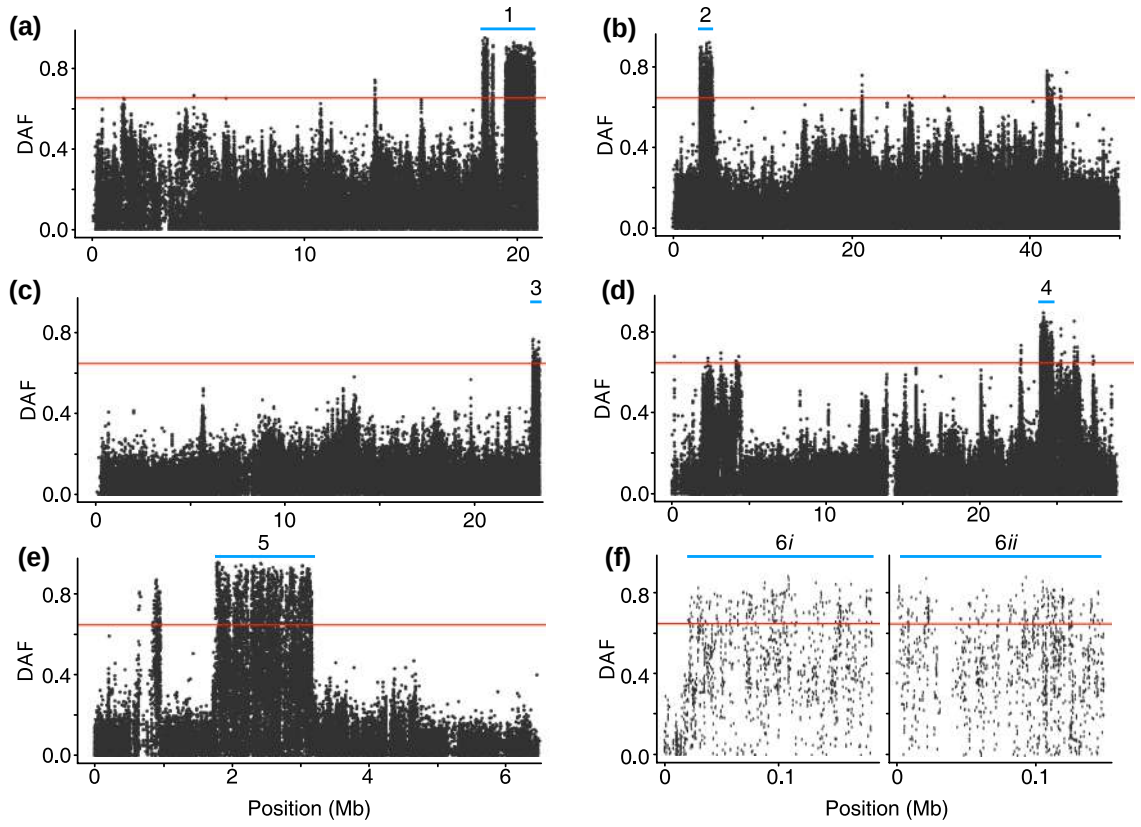

Fig. 5. Six putative inversions detected in the "Oceanic" versus "Brackish" contrast. Per-SNP *DAFs* highlight putative inversions, and the regions are found on the following scaffolds: s1110 a), s40 b), s1114 c), s13 d), s1111 e), and s173* and s374* f). *This signal is located on two minor scaffolds in the sprat assembly that both map to nearby regions on *C. harengus* chromosome 13.

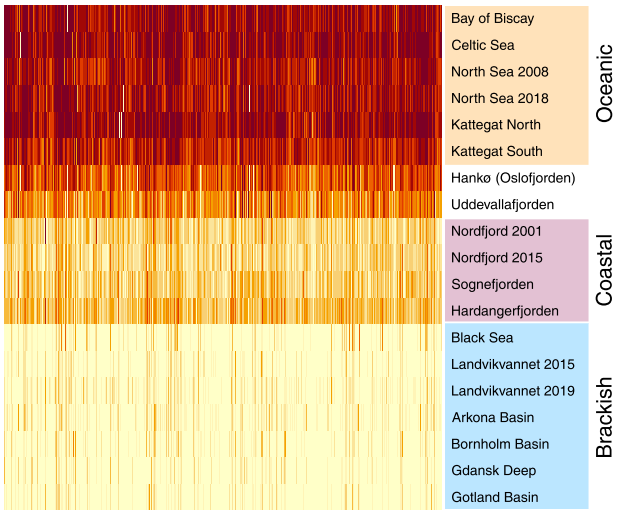

Fig. 6. Heatmap showing allele frequencies at the putative inversion on sprat scaffold s40. Population groupings are indicated, in accordance with Fig. 2. This 1.5 Mb region in sprat (shown in Fig. 5e) corresponds to two regions on *C. harengus* Chr 18 (13.9 to 16.5 Mb and 17.9 to 19.3 Mb), suggesting a rearrangement between the two species. Heatmaps for the other five putative inversions are shown in supplementary fig. S8, Supplementary Material online.

2019), and the great majority of these loci do not show an obvious overlap with signals of selection detected in the marine brackish contrast in sprat. However, for three genes, the narrow signals of selection overlap. Firstly, one of these regions is located around 2.3 Mb on chromosome 12 in Atlantic herring and harbors the prolactin receptor (*PRLRA*) gene only (Fig. 7). A similar narrow signal involves herring chromosome 19 containing the thyroid receptor beta (*THRB*) gene (supplementary fig. S11a, Supplementary Material online). Finally, there is a signal covering a cluster of troponin I2 (*TNNI2*) genes on herring chromosome 3 (supplementary fig. S11b, Supplementary Material online). We did not find any missense mutations in any of these three genes, neither in sprat nor in herring.

Discussion

Here, we have shown that the European sprat is a species with high, albeit not extreme, genetic variability, with overall nucleotide diversity estimated to be in excess of 1%. This is considerably higher than the 0.3% found in the Atlantic herring (Martinez Barrio et al. 2016), which we tentatively attribute to the fact that the sprat has a more southerly

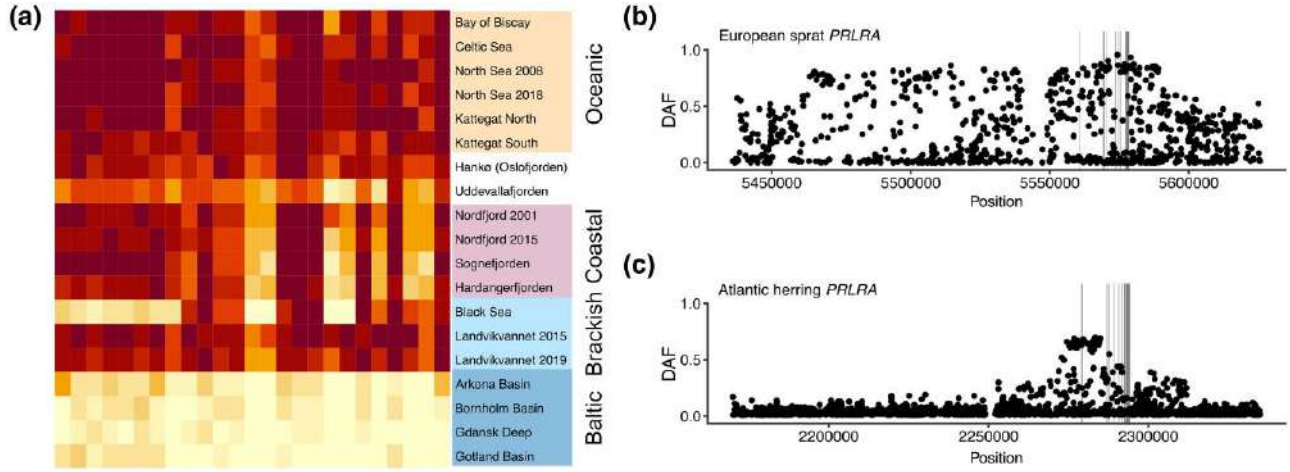

Fig. 7. Genetic parallelism between European sprat and Atlantic herring—*PRLRA*. a) Heatmap of the *PRLRA* region in European sprat. Samples are indicated as in Fig. 2, with the exception of Baltic samples being highlighted within the Brackish group. b) Zoom-in showing *DAF* in the Oceanic and Brackish contrasts, in European sprat, across the *PRLRA* locus (found on scaffold s1113 in the European sprat assembly). c) Zoom-in showing *DAF* between Atlantic and Baltic herring populations across the *PRLRA* locus on Chr 12 (Han et al. 2020). In b) and c), gray boxes indicate location of exons.

habitat range, reducing the risk of severe bottlenecks during periods of extensive polar ice cap expansion. The colonization of brackish waters has led to marked genetic differentiation, approaching fixation of different alleles, at a discrete set of loci across the genome. This adaptive response is, on the other hand, qualitatively similar to that previously shown in Atlantic herring (Martinez Barrio et al. 2016; Pettersson et al. 2019; Han et al. 2020). However, while this study has shed light on the differentiation between major groupings of sprat populations, the differentiation within these groups remains very limited with F_{ST} values in the range 0 to 0.006 between population samples. However, this result is based on a limited number of sampled populations, not quite reaching a quarter of the number used by Han et al. (2020) for Atlantic herring. A cost-efficient way of remedying this situation is to use a SNP-chip, and, to this end, the signals identified here have been used to select markers for the sprat component of the recently released MultiFishChip (Andersson et al. 2024) containing thousands of SNPs from each of seven species including the European sprat and Atlantic herring.

Like the herring, sprat has adapted to spawning in brackish conditions. This is important, since heterogeneity in environmental conditions during spawning has recently been brought up as a critical aspect resulting in genetic differentiation between subpopulations due to natural selection (Fuentes-Pardo et al. 2023; Andersson et al. 2024). This puts species that spawn in diverse environmental conditions, but may intermingle as adults, in contrast with species that spawn communally and then later on disperse. The archetypical example of the latter is the European eel (*Anguilla anguilla*), in which all spawning takes place in the Sargasso Sea. Consequently, that species has no discernible signs of genetic adaptation in response to the environments encountered by the adults, in spite of the habitat range for the adult stage extending, in geographic terms, from the Baltic to the Mediterranean and, in terms of salinity, from fully marine to freshwater (Enbody et al. 2021). The sprat and, according to a recent study, the Atlantic horse mackerel (*Trachurus trachurus*; Fuentes-Pardo et al. 2023) appear intermediate between two extremes, the European eel being one panmictic population and Atlantic herring showing extensive genetic differentiation (Martinez Barrio et al. 2016). An important difference between sprat and Atlantic horse mackerel, on one hand, and herring, on the other, is that the former are pelagic spawners (de Silva 1973; Abaunza et al. 2003), whereas herring deposits fertilized eggs at specific bottom types or on vegetation (Runnström 1941; Aneer et al. 1983). The latter are therefore exposed to a much more heterogenous environment than eggs from a pelagic spawner, possibly resulting in a difference in local selection pressure. Furthermore, it is expected that a homing behavior is less precise for a pelagic spawner than for a fish where the fertilized eggs are developing at spawning grounds.

We have strong indications that at least six putative inversions play important roles in the adaptation to brackish conditions in sprat. In Atlantic herring, there are at least four major inversions with prominent frequency gradients along a North–South axis, but no inversion appears to be underlying adaptation to low salinity (Han et al. 2020). The observation that the same inversion haplotype groups are present at high frequencies in geographically distant brackish populations separated by marine habitats (Fig. 1) suggests that the brackish haplotypes may occur at low frequencies in marine populations and, in particular, in coastal

populations that display temporary brackish conditions. These haplotypes can then be subject to strong positive selection when sprat colonize brackish environment. This must have happened when Landvikvannet in Norway was colonized about 150 years ago after this former freshwater lake became a brackish environment when a canal to the ocean was established for transporting timber. This mimics the situation in sticklebacks where marine populations constitute a reservoir of alleles critical for adaptation to freshwater conditions (Jones et al. 2012).

The putative inversions here reported for sprat and the four inversions in herring confirmed by PacBio long-read sequencing (Han et al. 2020; Jamsandekar et al. 2023) all associated with local adaptation, adding to a growing list of so-called supergenes contributing to ecological adaptation in marine organisms. For instance, in Atlantic cod, four large chromosome inversions (on chromosomes 1, 2, 7, and 12) are allegedly linked to a migratory lifestyle and environmental adaptations such as salinity tolerance (Berg et al. 2016, 2017; Barth et al. 2019; Matschiner et al. 2022). Similarly, three large putative chromosomal inversions have been found to be associated with sea surface temperature in the king scallop (*Pecten maximus*; Hollenbeck et al. 2022).

An important question in evolutionary biology is how often the same gene contributes to a similar genetic adaptation in different species (Conte et al. 2012). One example is that selection at the human *EPAS1* gene has contributed to adaptation to high altitude in both Tibetan and Andean highlanders (Lawrence et al. 2024). Another striking example of convergence at the molecular level concerns genes encoding visual opsins that often respond to selection related to differences in light conditions among species habitats (Musilova et al. 2021). For instance, one-third of all fish species adapted to freshwater or brackish waters, including Baltic herring, express a rhodopsin protein with tyrosine at residue 261, whereas essentially all marine fish have phenylalanine at this position (Hill et al. 2019). However, most biological traits have a highly polygenic background (Pritchard and Di Rienzo 2010), which implies that such striking genetic parallelism is expected to be uncommon. The genome-wide screens of the two closely related clupeids European sprat and Atlantic herring as regards their adaptation to the brackish Baltic Sea provide an excellent opportunity to explore how common genetic parallelism is. Our conclusion is that it is limited because the adaptation in sprat is dominated by six putative inversions not being present in Atlantic herring, and out of the 125 loci identified in Atlantic herring (Han et al. 2020), only three show a striking overlap with signals in sprat. One of these three genes, *TNNI2*, does not have an obvious link to adaptation to a brackish environment; *TNNI2* encodes troponin I2, an important component of fast twitch muscles. *THRB* encodes thyroid hormone receptor beta, a nuclear receptor that affects gene expression in many cell types. Of potential relevance to the selection signals in sprat and herring is that *THRB* in zebrafish has a critical role for development of retinal red cones and long-wave vision (Volkov et al. 2020). This is of interest because of the strong selection acting on visual opsins in fish in general (Musilova et al. 2021) and the documented selection for improved vision in the red spectra in Baltic herring (Hill et al. 2019). The third case of genetic parallelism concerns the prolactin receptor (*PRLRA*) gene (Fig. 7), which has an obvious link to adaptation to a brackish environment due to its important role in osmoregulation (Manzon 2002). The hormone prolactin has many functions in vertebrates, including stimulation of milk secretions in mammals as indicated by its name. However, in fish, it constitutes the freshwater-adapting hormone in euryhaline species (Manzon 2002). Prolactin is released from the pituitary but its effect on target tissues (gills, kidney, intestine, urinary bladder, and skin) is mediated by prolactin receptor signaling. The fact that we did not find any *PRLRA* missense mutations neither in sprat nor in herring implies that genetic adaptation to the brackish environment is mediated by an altered expression pattern. The results imply that *PRLRA* may be a gene that often shows genetic parallelism in fish species adapted to differences in salinity.

The results of this study have important implications for fishery management of the sprat. The study confirms the basic population structure reported in previous work showing clear genetic differentiation between Oceanic, fjord, and populations from different brackish environments (McKeown et al. 2020; Quintela et al. 2020; Quintela et al. 2021). It also provided data for the construction of a SNP-chip tool that can be used for stock assessment in areas where genetically distinct populations are mixed. Our data show that the results obtained with these carefully selected SNPs match the resolution obtained with WGS. However, further WGS studies are justified because the present study is based on a limited set of sample locations given the broad geographic distribution from the northeast Atlantic Ocean and the Baltic Sea to the Mediterranean Sea and the Black Sea. Such studies will reveal, for instance, whether the sprat in the Baltic Sea constitutes a single panmictic population or not and if Oceanic and Coastal ecotypes occur throughout the species distribution.

Materials and Methods

Sample Collection

All the individuals typed in the different steps of this study (supplementary table S1, Supplementary Material online), with the exception of samples HAN and AS (the latter of which was not used for resequencing due to low DNA quality), were formerly used to outline sprat management

boundaries (ICES 2018; Quintela et al. 2020) and, in the case of the samples from Landvikvannet (LAND15 and LAND19), to describe the genetic response to human-induced habitat changes (Quintela et al. 2021). Sample HAN was collected in the Oslofjorden area in 2018 (59.2°N, 10.9°E), whereas sample AS was collected in 2021 in the Adriatic Sea (43.1°N, 13.9°E) during the MEDIAS survey (Leonori et al. 2021), following a common protocol. DNA was extracted from fin clips stored in ethanol using the Qiagen DNeasy 96 Blood & Tissue Kit in 96-well plates; each of which contained two or more negative controls.

Preparation of High-Molecular-Weight DNA

Flash-frozen tissue from one individual sprat was collected from Langenuen in the outer part of Hardangerfjorden, Norway (59.975°N, 5.376°E). High-molecular-weight DNA for PacBio long-read sequencing was extracted using the Nanobind Tissue Big DNA Kit v1.0, standard TissueRuptor protocol. The genomic DNA was subsequently cleaned using the Pacific Bioscience "Guidelines for Using a Salt: Chloroform Wash to Clean Up gDNA." DNA concentration was measured using the Qubit Broad Range Kit (Thermo Fisher Scientific, #Q32850), purity was checked by UV absorbance, and fragment lengths were determined using the Genomic DNA 165 kb Kit (Agilent #FP-1002-0275) on the Femto Pulse System.

Long-Read Sequencing and Draft Assembly Construction

A total of 1.47×10^6 M PacBio CCS reads (≥Q20), covering a total of 24.6×10^9 bases, were generated from a single flow cell (SMRT link v8.0.0.79519). This read-set was used to create the contig version of the draft assembly, using IPA v1.3.2 (https://github.com/PacificBiosciences/pbipa). The scaffolded version was done based on the abovementioned contig assembly and 287 M HiC read pairs, using pin-HiC (v3.0.0; https://github.com/dfguan/pin_hic). The pin-HiC output was manually curated using Juicebox (v1.11.08; Dudchenko et al. 2018) and a custom deduplication procedure based on read depth and the location of duplicated BUSCO hits (https://github.com/LeifAnderssonLab/Sprat_pool_reseq).

Pooled WGS

Genomic DNA was extracted individually and then pooled in equimolar quantities. Illumina short-read sequencing of pooled genomic DNA from the 19 populations was carried out using the standard configuration of paired 150 bp reads, and Illumina short-read sequencing of pooled population samples was performed at NGI Uppsala—SNP&SEQ Technology Platform. Data were generated using one lane of Illumina NovaSeq S4, comprising 3.2×10^9 read pairs ($1.69 \pm 0.38 \times 10^8$ per pool), resulting in 9.7×10^{11} bases sequenced. The reads are available at NCBI's Short Read Archive (BioProject: PRJNA1023385).

Reads were mapped to the draft sprat assembly (see above) using bwa mem (v 0.7.17-r1188; Li 2013), and polymerase chain reaction duplicates were marked using "MarkDuplicates" from Picard Tools (http://broadinstitute.github.io/picard/). The resulting bam files were then processed using GATK (v4.1.1.0; McKenna et al. 2010), with the following workflow: First, we used "HaplotypeCaller" to generate per-sample gvcf files. These were merged using "CombineGVCFs" and genotyped using "GenotypeGVCFs." The raw SNP genotypes were then passed through the "VariantFiltration" module, with the following arguments:

--filter-name "stringent_combined_filter" --filter-expression "QD < 8.0 || FS > 50.0 || MQ < 30.0 || MQRankSum < -10.0 || ReadPosRankSum < -6.0" --filter-name "depth_filter" --filter-expression "DP < 200.0 || DP > 2500" --missing-values-evaluate-as-failing.

The called genotypes were, however, not used; instead, we calculated frequency estimates based on the ratio of mapped reference and alternative reads at each called SNP position. The bioinformatic analysis of pooled WGS data was performed using custom R (R_Core_Team 2019) scripts; these are deposited in the "Sprat_pool_reseq" repository at the LeifAnderssonLab GitHub page (https://github.com/LeifAnderssonLab/Sprat_pool_reseq).

Genome-Wide Nucleotide Diversity

We estimated genome-wide nucleotide diversity (π) by comparing the sequences in the primary and alternative assembly from the reference individual. First, we performed a "satsuma chromosemble" (Grabherr et al. 2010) run to align the primary and alternative assemblies. Then, we randomly selected a set of positions ($n = 600$) from the primary assembly to serve as starts for alignment blocks. Starting at these positions, we extracted 25 kb regions and queried the Satsuma output for a matching contig in the alternative assembly. These were then aligned using Clustal Omega (Sievers and Higgins 2018), and then diversity in the block was calculated using the "dist.dna" method from the R package "ape" (Paradis and Schliep 2019). Blocks yielding either no, or very short, alignments—typically indicative of the corresponding contig being missing for the alternative assembly—were eliminated, as were those with observed divergence > 5%. The latter outcome is not consistent with regular sequence divergence, but rather likely to be the result of alignment artifacts induced by structural differences between the two haplotypes. Similar issues are likely to contribute to the highest of the retained values as well, which is why we are using the median, rather than mean, observed divergence as our estimate of π. Retaining all

alignments would lead to a small increase in estimated π, from 1.2% to 1.3%.

Frequency-Based Neighbor-Joining Tree

The distance matrix used was the average frequency difference per SNP, calculated as the sum of absolute allele frequency differences divided by the number of observed SNPs. The tree was constructed using the "bionj" function from "ape" (Paradis and Schliep 2019).

Population Contrasts

All contrasts were based on the absolute *DAFs* of the groups involved. Significance was estimated based on a *Z*-score conversion, with mean and SD calculated from the entire marker set, Bonferroni corrected for the number of SNPs that passed filtering. The resulting threshold value was then converted to an absolute *DAF* value. This approach is somewhat conservative, as all markers, even those potentially under selection, are included. However, it avoids any potential bias by attempting to predefine a neutral set of markers to estimate background divergence.

Liftover to the Atlantic Herring Assembly

Liftover was achieved using the "chromosemble" tool from "satsuma" (Grabherr et al. 2010) to generate a set of matching genome locations. The matches were then used, by way of a custom R script, to assign an approximate location on the Atlantic herring genome to each sprat SNP that fell into a mapped region.

MultiFishChip Design

The sprat component of the MultiFishChip (Andersson et al. 2024) was based on the data presented herein and comprises three subsets: first, a neutral set containing SNPs with low variation between groups but comparatively high minor allele frequencies ($MAF > 0.3$), estimated across the entire set of samples; second, a set selected from the genome-wide contrasts presented here, with representative SNPs selected from identified signals of divergence; and lastly, a set of SNPs with high variance across populations, not tied to any particular predefined contrast. Together, the design resulted in 7,742 candidate markers (supplementary data S1, Supplementary Material online), out of which 5,916 were eventually included on the MultiFishChip.

MultiFishChip Analysis

A subset of 381 individuals from 19 samples was sent to IdentiGEN (Ireland) for genotyping with the MultiFishChip SNP array (Andersson et al. 2024); 18 of the samples were overlapping with the WGS set; the sample from Hankø (HAN) had to be discarded due to amplification issues whereas the sample from the Adriatic Sea was not available for WGS but could be acquired and genotyped with the SNP-chip at a later stage of the project. Twelve of the individuals failed to amplify, thus leaving 369 fish ranging between 16 (BoB) and 24 (AS) per sample. Data were obtained for 4,602 out of the 5,916 SNPs of the chip, 75 of which were discarded due to >10% missing data. PLINK (Purcell et al. 2007) was used to LD prune the data set ($r^2 = 0.25$, $MAF > 0.05$), thus leaving 2,063 SNPs for statistical analyses. PCA was conducted using the function "dudi.pca" in ade4 (Dray and Dufour 2007). The relationship among geographically explicit samples was assessed using pairwise F_{ST} (Weir and Cockerham 1984) and through the DAPCs (Jombart et al. 2010) implemented in adegenet (Jombart 2008) using the cross-validation function (Jombart and Collins 2015; Miller et al. 2020) to avoid overfitting. In addition, STRUCTURE (v.2.3.4; Pritchard et al. 2000) analysis was conducted using the software ParallelStructure (Besnier and Glover 2013) to identify genetic groups under a model assuming admixture and correlated allele frequencies without using population information. Ten runs with 100,000 burn-in and 1,000,000 MCMC iterations were performed for $K = 1$ to $K = 5$ clusters. STRUCTURE output was analyzed using (i) the ad hoc summary statistic ΔK of Evanno et al. (2005) and (ii) the four statistics of Puechmaille (2016), both implemented in StructureSelector (Li and Liu 2018). Finally, runs for the selected *K*s were averaged with CLUMPP v.1.1.1 (Jakobsson and Rosenberg 2007) using the FullSearch algorithm and the G′ pairwise matrix similarity statistic and graphically displayed using bar plots.

RDA is a genotype–environment association method to detect loci under selection (Forester et al. 2018). Environmental data (temperature, salinity, pH, dissolved oxygen, current velocity, and chlorophyll) were obtained for the different sampling points using the Bio-ORACLE database (https://www.bio-oracle.org; Tyberghein et al. 2012; Assis et al. 2018). Collinearity between variable pairs was investigated, and only noncorrelated ones were retained for analyses. Analysis was conducted with the R package vegan v.2.5–7 (Oksanen et al. 2019) using the 2,063 LD-pruned loci with samples classified into Brackish, Oceanic, and Coastal. The samples from the hybrid zone and the Adriatic Sea were discarded for not falling in any of these categories.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

The National Genomics Infrastructure (NGI)/Uppsala Genome Center provided service in PacBio long-read sequencing. Short-read Illumina sequencing was performed by the

SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. The computational infrastructure was provided by the National Academic Infrastructure for Supercomputing in Sweden, partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

Author Contributions

L.A. and M.Q. conceived the study. M.E.P. was responsible for the genome assembly and bioinformatic analysis of WGS data. M.Q. and Fr.B. were responsible for the SNP-chip analysis. Q.D. and A.W. contributed to the bioinformatic analysis of WGS data. I.B. produced the PacBio assembly. M.-B.M. isolated HMW-DNA needed for long-read sequencing. Fl.B., C.K., D.B., R.L.-L., I.L., and K.A.G. contributed to the collection of sprat samples. M.E.P., L.A., and, M.Q. wrote the paper with input from all other authors. All authors approved the manuscript before submission.

Funding

The project was financially supported by the Research Council of Norway (CoastRisk project—299554), Vetenskapsrådet (2017-02907; to L.A.), and Knut and Alice Wallenberg Foundation (KAW 2023.0160; to L.A.).

Conflict of Interest

None declared.

Data Availability

The sequence data generated in this study and genome assemblies have been submitted to NCBI (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1023385). The analyses of data have been carried out with publicly available software, and all are cited in the Materials and Methods section. Codes associated with bioinformatic analyses are available at https://github.com/LeifAnderssonLab/Sprat_pool_reseq. Correspondence and requests for materials should be addressed to L.A. (leif.andersson@imbim.uu.se).

Literature Cited

Abaunza P, Gordo L, Karlou-Riga C, Murta A, Eltink ATGW, García Santamaría MT, Zimmermann C, Hammer C, Lucio P, Iversen SA, et al. Growth and reproduction of horse mackerel, *Trachurus trachurus* (*carangidae*). Rev. Fish Biol. Fish. 2003:13(1):27–61. https://doi.org/10.1023/A:1026334532390.

Andersson L, Bekkevold D, Berg F, Farrell ED, Felkel S, Ferreira MS, Fuentes-Pardo AP, Goodall J, Pettersson M. How fish population genomics can promote sustainable fisheries: a road map. Annu Rev Anim Biosci. 2024:12(1):1–20. https://doi.org/10.1146/annurev-animal-021122-102933.

Aneer G, Florell G, Kautsky U, Nellbring S, Sjöstedt L. In-situ observations of Baltic herring (*Clupea harengus membras*) spawning behaviour in the Askö-Landsort area, northern Baltic proper. Mar Biol. 1983:74(2):105–110. https://doi.org/10.1007/BF00413912.

Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O. Bio-ORACLE v2.0: extending marine data layers for bioclimatic modelling. Glob Ecol Biogeogr. 2018:27(3):277–284. https://doi.org/10.1111/geb.12693.

Barth JMI, Villegas-Ríos D, Freitas C, Moland E, Star B, André C, Knutsen H, Bradbury I, Dierking J, Petereit C, et al. Disentangling structural genomic and behavioural barriers in a sea of connectivity. Mol Ecol. 2019:28(6):1394–1411. https://doi.org/10.1111/mec.15010.

Berg F, Slotte A, Andersson L, Folkvord A. Genetic origin and salinity history influence the reproductive success of Atlantic herring. Mar Ecol Prog Ser. 2018:2018:81–94. https://doi.org/10.3354/meps12680.

Berg PR, Star B, Pampoulie C, Bradbury IR, Bentzen P, Hutchings JA, Jentoft S, Jakobsen KS. Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. Heredity (Edinb). 2017:119(6):418–428. https://doi.org/10.1038/hdy.2017.54.

Berg PR, Star B, Pampoulie C, Sodeland M, Barth JMI, Knutsen H, Jakobsen KS, Jentoft S. Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. Sci Rep. 2016:6(1):23246. https://doi.org/10.1038/srep23246.

Besnier F, Glover KA. ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multicore computers. PLoS One. 2013:8(7):e70651. https://doi.org/10.1371/journal.pone.0070651.

Conte GL, Arnegard ME, Peichel CL, Schluter D. The probability of genetic parallelism and convergence in natural populations. Proc. R. Soc. B Biol. Sci. 2012:279(1749):5039–5047. https://doi.org/10.1098/rspb.2012.2146.

de Silva SS. Aspects of the reproductive biology of the sprat, *Sprattus sprattus* (L.) in inshore waters of the west coast of Scotland. J. Fish Biol. 1973:5(6):689–705. https://doi.org/10.1111/j.1095-8649.1973.tb04505.x.

Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. J Stat Softw. 2007:22(4):1–20. https://doi.org/10.18637/jss.v022.i04.

Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, Hilaire BS, Yao W, Stamenova E, et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv. 254797. 2018 https://doi.org/10.1101/254797, preprint: not peer reviewed.

Eggers F, Slotte A, Libungan LA, Johannessen A, Kvamme C, Moland E, Olsen EM, Nash RDM. Seasonal dynamics of Atlantic herring (*Clupea harengus* L.) populations spawning in the vicinity of marginal habitats. PLoS One. 2014:9(11):e111985. https://doi.org/10.1371/journal.pone.0111985.

Enbody ED, Pettersson ME, Sprehn CG, Palm S, Wickström H, Andersson L. Ecological adaptation in European eels is based on phenotypic plasticity. Proc Natl Acad Sci U S A. 2021:118(4):e2022620118. https://doi.org/10.1073/pnas.2022620118.

Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol. 2005:14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Forester BR, Lasky JR, Wagner HH, Urban DL. Comparing methods for detecting multilocus adaptation with multivariate genotype–

environment associations. Mol Ecol. 2018:27(9):2215–2233. https://doi.org/10.1111/mec.14584.
Fuentes-Pardo AP, Farrell ED, Pettersson ME, Sprehn CG, Andersson L. The genomic basis and environmental correlates of local adaptation in the Atlantic horse mackerel (*Trachurus trachurus*). Evol Appl. 2023:16(6):1201–1219. https://doi.org/10.1111/eva.13559.
Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Di Palma F, Lindblad-Toh K. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics. 2010:26(9): 1145–1151. https://doi.org/10.1093/bioinformatics/btq102.
Han F, Jamsandekar M, Pettersson ME, Su L, Fuentes-Pardo AP, Davis BW, Bekkevold D, Berg F, Casini M, Dahle G, et al. Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. Elife. 2020:9:e61076. https://doi.org/10.7554/eLife.61076.
Hill J, Enbody ED, Pettersson ME, Sprehn CG, Bekkevold D, Folkvord A, Laikre L, Kleinau G, Scheerer P, Andersson L. Recurrent convergent evolution at amino acid residue 261 in fish rhodopsin. Proc Natl Acad Sci U S A. 2019:116(37):18473–18478. https://doi.org/10.1073/pnas.1908332116.
Hollenbeck CM, Portnoy DS, Garcia de la Serrana D, Magnesen T, Matejusova I, Johnston IA. Temperature-associated selection linked to putative chromosomal inversions in king scallop (*Pecten maximus*). Proc. R. Soc. B Biol. Sci. 2022:289:20221573. https://doi.org/10.1098/rspb.2022.1573.
ICES. Benchmark Workshop on Sprat (WKSPRAT 2018). ICES Expert Group reports (until 2018). Report. 2018. https://doi.org/10.17895/ices.pub.19291145.v1
Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007:23(14): 1801–1806. https://doi.org/10.1093/bioinformatics/btm233.
Jamsandekar M, Ferreira MS, Pettersson ME, Farell E, Davis BW, Andersson L. 2023. The origin and maintenance of supergenes contributing to ecological adaptation in Atlantic herring. bioRxiv. 2023.10.23.562618. https://doi.org/10.1101/2023.10.23.562618., preprint: not peer reviewed.
Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008:24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129.
Jombart T, Collins C. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 1. Vienna: R-Forge; 2015. p. 3–4. https://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf.
Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010:11(1):94. https://doi.org/10.1186/1471-2156-11-94.
Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012:484(7392):55–61. https://doi.org/10.1038/nature10944.
Lamichhaney S, Fuentes-Pardo AP, Rafati N, Ryman N, McCracken GR, Bourne C, Singh R, Ruzzante DE, Andersson L. Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. Proc Natl Acad Sci U S A. 2017:114(17): E3452–E3461. https://doi.org/10.1073/pnas.1617728114.
Lass HU, Matthäus W. General oceanography of the Baltic Sea. State and evolution of the Baltic Sea, 1952-2005: a detailed 50-year survey of meteorology and climate, physics, chemistry, biology, and marine environment. New York: John Wiley & Sons; 2008. p. 5–43. https://doi.org/10.1002/9780470283134.ch2.
Lawrence ES, Gu W, Bohlender RJ, Anza-Ramirez C, Cole AM, Yu JJ, Hu H, Heinrich EC, O'Brien KA, Vasquez CA, et al. Functional EPAS1/HIF2A missense variant is associated with hematocrit in Andean highlanders. Sci Adv. 2024:10(6):eadj5661. https://doi.org/10.1126/sciadv.adj5661.
Leonori I, Tičina V, Giannoulaki M, Hattab T, Iglesias M, Bonanno A, Costantini I, Canduci G, Machias A, Ventero A, et al. History of hydroacoustic surveys of small pelagic fish species in the European Mediterranean Sea. Mediterr. Mar. Sci. 2021:22(4):751–768. https://doi.org/10.12681/mms.26001.
Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. bioRxiv. 1303.3997v1 [q-bio.GN]. 2013. https://doi.org/10.48550/arXiv.1303.3997, preprint: not peer reviewed.
Li YL, Liu JX. StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol Ecol Resour. 2018:18(1):176–177. https://doi.org/10.1111/1755-0998.12719.
Manzon LA. The role of prolactin in fish osmoregulation: a review. Gen Comp Endocrinol. 2002:125(2):291–310. https://doi.org/10.1006/gcen.2001.7746.
Martinez Barrio A, Lamichhaney S, Fan G, Rafati N, Pettersson M, Zhang HE, Dainat J, Ekman D, Höppner M, Jern P, et al. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. Elife. 2016:5:e12081. https://doi.org/10.7554/eLife.12081.
Matschiner M, Barth JM, Tørresen OK, Star B, Baalsrud HT, Brieuc MS, Pampoulie C, Bradbury I, Jakobsen KS, Jentoft S, et al. Supergene origin and maintenance in Atlantic cod. Nat Ecol Evol. 2022:6: 469–481. https://doi.org/10.1038/s41559-022-01661-x.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010:20(9): 1297–1303. https://doi.org/10.1101/gr.107524.110.
McKeown NJ, Carpi P, Silva JF, Healey AJE, Shaw PW, van der Kooij J. Genetic population structure and tools for the management of European sprat (*Sprattus sprattus*). ICES J Mar Sci. 2020:77(6): 2134–2143. https://doi.org/10.1093/ICESJMS/FSAA113.
Miller JM, Cullingham CI, Peery RM. The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. Heredity (Edinb). 2020:125(5): 269–280. https://doi.org/10.1038/s41437-020-0348-2.
Musilova Z, Salzburger W, Cortesi F. The visual opsin gene repertoires of teleost fishes: evolution, ecology, and function. Annu Rev Cell Dev Biol. 2021:37(1):441–468. https://doi.org/10.1146/annurev-cellbio-120219-024915.
Oksanen J, Simpson GL, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Henry H M, Stevens ES, Wagner H, et al. Vegan: community ecology package (version 2.5-6). Vienna: The Comprehensive R Archive Network; 2019.
Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019:35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633.
Pettersson ME, Rochus CM, Han F, Chen J, Hill J, Wallerman O, Fan G, Hong X, Xu Q, Zhang H, et al. A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. Genome Res. 2019:29(11):1919–1928. https://doi.org/10.1101/gr.253435.119.
Pritchard JK, Di Rienzo A. Adaptation—not by sweeps alone. Nat Rev Genet. 2010:11(10):665–667. https://doi.org/10.1038/nrg2880.
Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000:155(2):945–959. https://doi.org/10.1093/genetics/155.2.945.
Puechmaille SJ. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol Ecol Resour. 2016:16(3):608–627. https://doi.org/10.1111/1755-0998.12512.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007:81(3):559–575. https://doi.org/10.1086/519795.

Quintela M, Kvamme C, Bekkevold D, Nash RDM, Jansson E, Sørvik AG, Taggart JB, Skaala Ø, Dahle G, Glover KA. Genetic analysis redraws the management boundaries for the European sprat. Evol Appl. 2020:13(8):1906–1922. https://doi.org/10.1111/eva.12942.

Quintela M, Richter-Boix À, Bekkevold D, Kvamme C, Berg F, Jansson E, Dahle G, Besnier F, Nash RDM, Glover KA, et al. Genetic response to human-induced habitat changes in the marine environment: a century of evolution of European sprat in Landvikvannet, Norway. Ecol Evol. 2021:11(4):1691–1718. https://doi.org/10.1002/ece3.7160.

R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.

Runnström S. Quantitative investigations on herring spawning and its yearly fluctuations at the west coast of Norway. FiskDir Skr Ser HavUnders. 1941:6:1–71. https://imr.brage.unit.no/imr-xmlui/handle/11250/114834.

Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2018:27(1):135–145. https://doi.org/10.1002/pro.3290.

Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F, De Clerck O. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. Glob Ecol Biogeogr. 2012:21(2):272–281. https://doi.org/10.1111/j.1466-8238.2011.00656.x.

Volkov LI, Kim-Han JS, Saunders LM, Poria D, Hughes AEO, Kefalov VJ, Parichy DM, Corbo JC Thyroid hormone receptors mediate two distinct mechanisms of long-wavelength vision. Proc Natl Acad Sci U S A. 2020:117(26):15262–15269. https://doi.org/10.1073/pnas.1920086117.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984:38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.

Whitehead PJP. Clupeoid fishes of the world (suborder Clupeioidei). In: An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. FAO Species Catalogue; 1985. Vol 7. p. 42–50.

Associate editor: Bonnie Fraser