











Post-glacial establishment of locally adapted fish populations over a steep salinity gradient

Erica H. Leder^{1,2,3}  | Carl André^{1,4}  | Alan Le Moan^{1,4} | Mats Töpel^{1,5}  | Anders Blomberg^{1,6}  | Jonathan N. Havenhand^{1,4}  | Kai Lindström⁷  | Filip A. M. Volckaert^{1,8}  | Charlotta Kvarnemo^{1,9}  | Kerstin Johannesson^{1,4}  | Ola Svensson^{1,9} 

¹Centre for Marine Evolutionary Biology, University of Gothenburg, Gothenburg, Sweden

²Department of Biology, University of Turku, Turku, Finland

³Natural History Museum, University of Oslo, Oslo, Norway

⁴Tjärnö Marine Laboratory, Department of Marine Sciences, University of Gothenburg, Strömstad, Sweden

⁵Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden

⁶Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

⁷Environmental and Marine Biology, Åbo Akademi University, Turku, Finland

⁸Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, Leuven, Belgium

⁹Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden

Correspondence

Erica H. Leder, Centre for Marine Evolutionary Biology, Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden. Email: eriled@utu.fi; erica.leder@gu.se

Present address

Ola Svensson, Department for Pre-School and School Teacher Education, University of Borås, Borås, Sweden

Funding information

Academy of Finland, Grant/Award Number: 136464; Vetenskapsrådet, Grant/Award Number: 2016-03343, 217-2008-1719 and 621-2011-4004; Fonds Wetenschappelijk Onderzoek, Grant/Award Number: W0.037.10 N

Abstract

Studies of colonization of new habitats that appear from rapidly changing environments are interesting and highly relevant to our understanding of divergence and speciation. Here, we analyse phenotypic and genetic variation involved in the successful establishment of a marine fish (sand goby, *Pomatoschistus minutus*) over a steep salinity drop from 35 PSU in the North Sea (NE Atlantic) to two PSU in the inner parts of the post-glacial Baltic Sea. We first show that populations are adapted to local salinity in a key reproductive trait, the proportion of motile sperm. Thereafter, we show that genome variation at 22,190 single nucleotide polymorphisms (SNPs) shows strong differentiation among populations along the gradient. Sequences containing outlier SNPs and transcriptome sequences, mapped to a draft genome, reveal associations with genes with relevant functions for adaptation in this environment but without overall evidence of functional enrichment. The many contigs involved suggest polygenic differentiation. We trace the origin of this differentiation using demographic modelling and find the most likely scenario is that at least part of the genetic differentiation is older than the Baltic Sea and is a result of isolation of two lineages prior to the current contact over the North Sea–Baltic Sea transition zone.

KEYWORDS

genetic differentiation, local adaptation, population genomics, salinity gradient, sperm traits

The peer review history for this article is available at <https://publons.com/publon/10.1111/JEB.13668>

[Correction added on 1 August 2020, after first online publication: URL for peer review history has been corrected.]

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 European Society For Evolutionary Biology. Journal of Evolutionary Biology © 2020 European Society For Evolutionary Biology

1 | INTRODUCTION

Genetic divergence of populations following establishment in new environments provides important insights into fundamental evolutionary processes, such as local adaptation, hybridization, speciation and post-glacial recolonization (Abbott et al., 2013; Hewitt, 2011; Schluter, 2009). In particular, recent colonization of a new environment gives several advantages for study; one being that the potential source populations are still present and available for analysis. A second advantage is that the accumulation of new genetic variation from new mutations and the continuous loss of genetic variation through genetic drift will have had less impact compared to populations established a long time ago. Finally, footprints of demographic events prior to and during colonization, such as prior isolation of populations and/or repeated bottlenecks at the edge of the expansion (Barton & Hewitt, 1985; Bierne, Welch, Loire, Bonhomme, & David, 2011; Excoffier, Hofer, & Foll, 2009; Ravinet et al., 2017) will remain intact and visible. For recent events, most of the observed genetic differentiation between the source and the new population is likely to be a consequence of sorting of genetic variation that is already present in the source population(s) by the altered selection pressure in the new environment (Barrett & Schluter, 2008). Today, genome-wide scans of genetic variation are used to analyse the genetic background of recent colonization events and, if complemented with experimental studies at the ecological and phenotypic level, will contribute understanding of when and how new environments are colonized.

Strong phenotypic and genetic changes over small-scale environmental shifts are common in a range of plant and invertebrate populations (Soria-Carrasco et al., 2014; Twyford & Friedman, 2015; Westram et al., 2018). Emblematic examples of such shifts in vertebrates include populations of three-spined stickleback *Gasterosteus aculeatus* invading recently formed habitats of lakes and streams over extremely short periods of time (50–150 years) (Lescak et al., 2015; Marques et al., 2016) and the repeated evolution of resistant populations of the Atlantic killifish *Fundulus heteroclitus* in polluted habitats (Whitehead, Clark, Reid, & Hahn, 2017). An opportunity to study strong divergence of populations at much larger spatial scales is to focus on post-glacial establishment in habitats that became available after the retreat of the last glacial ice cap in northern Europe and North America. Studies of colonization of post-glacial lakes have repeatedly suggested that large-scale phenotypic and genetic divergence is involved in local adaptation both among and within lakes (Bernatchez, Laporte, Perrier, Sirios, & Bernatchez, 2016; Gagnaire, Pavey, Normandeau, & Bernatchez, 2013; Wilson et al., 2004).

The transformation of the post-glacial Ancylus Lake into the brackish water Baltic Sea some 8–10,000 years ago (Björck, 1995; Zillén, Conley, Andrén, Andrén, & Björck, 2008) offers opportunities to study genetic differentiation in a number of marine species colonizing the new basin. After the post-glacial formation of the

Baltic Sea initially at a salinity $\sim 1/2$ of ocean salinity, but currently of less than $1/3$, some 10% of marine species present in the North Sea outside the Baltic Sea have colonized the new environment and established local populations and formed contact zones with the North Sea populations (Johannesson & André, 2006). The majority of the successful colonisers for which we have genetic data show rather dramatic genetic shifts coinciding with the steepest part of the salinity gradient southeast of the Danish straits (Barth et al., 2019; Johannesson & André, 2006; Le Moan, Gaggiotti, et al., 2019; Luttkhuizen, Drent, Peijnenburg, van der Veer, & Johannesson, 2012; Martinez-Barrio et al., 2016). However, there are also examples of shifts taking place further inside the Baltic Sea (e.g., Momigliano et al., 2017). Notably, some Baltic populations originate from recently derived Pacific lineages (Nikula, Strelkov, & Väinölä, 2007; Väinölä & Strelkov, 2011), whereas others have North Atlantic origins that had, prior to establishment in the Baltic Sea, been isolated in different refugia for some time (Le Moan, Gaggiotti, et al., 2019), or diverged very rapidly during an initial phase of isolation during establishment in the Baltic Sea (Momigliano et al., 2017). Finally, some appeared to have evolved gradually by local adaptation over the contact zone whereas under gene flow from their ancestral marine populations (Le Moan, Gaggiotti, et al., 2019; Le Moan, Jiménez-Mena, Jiménez-Mena, Bekkevold, & Hemmer-Hansen, 2019; Martinez-Barrio et al., 2016).

The sand goby (*Pomatoschistus minutus* Pallas, 1770), a small marine shallow-water annual fish, has successfully colonized the complete North Sea–Baltic Sea salinity gradient (Kullander, Nyman, Jilg, & Dellings, 2012; Wiederholm, 1987). As detailed below, salinity is likely to affect a multitude of traits in this species, including physiological, ecological and behavioural traits. The effects of salinity on reproductive traits might be particularly important, as one expects these traits to be under strong selection compared to other traits (e.g., Hendry, 2016). With external fertilization, gametes are also likely to be adapted to the local habitat (Browne et al., 2015; Griffin et al., 1998; Morisawa, 2008), and in the case of sand goby, it seems that salinity only affects the males' ability to reproduce since eggs will develop and hatch regardless of female origin as long as the male spawns in his native salinity (Svensson et al., 2017).

Here, we investigate whether the establishment of sand goby in the brackish water Baltic Sea involves adaptive genomic divergence and local adaptation by combining analyses of genetic variation with experimental tests of sperm motility, which is a key reproductive trait (Gage et al., 2004; Gasparini, Simmons, Beveridge, & Evans, 2010).

2 | METHODS

2.1 | Study species

The sand goby is a small annual fish distributed in shallow seas and estuaries in the NE Atlantic Ocean from northern Norway

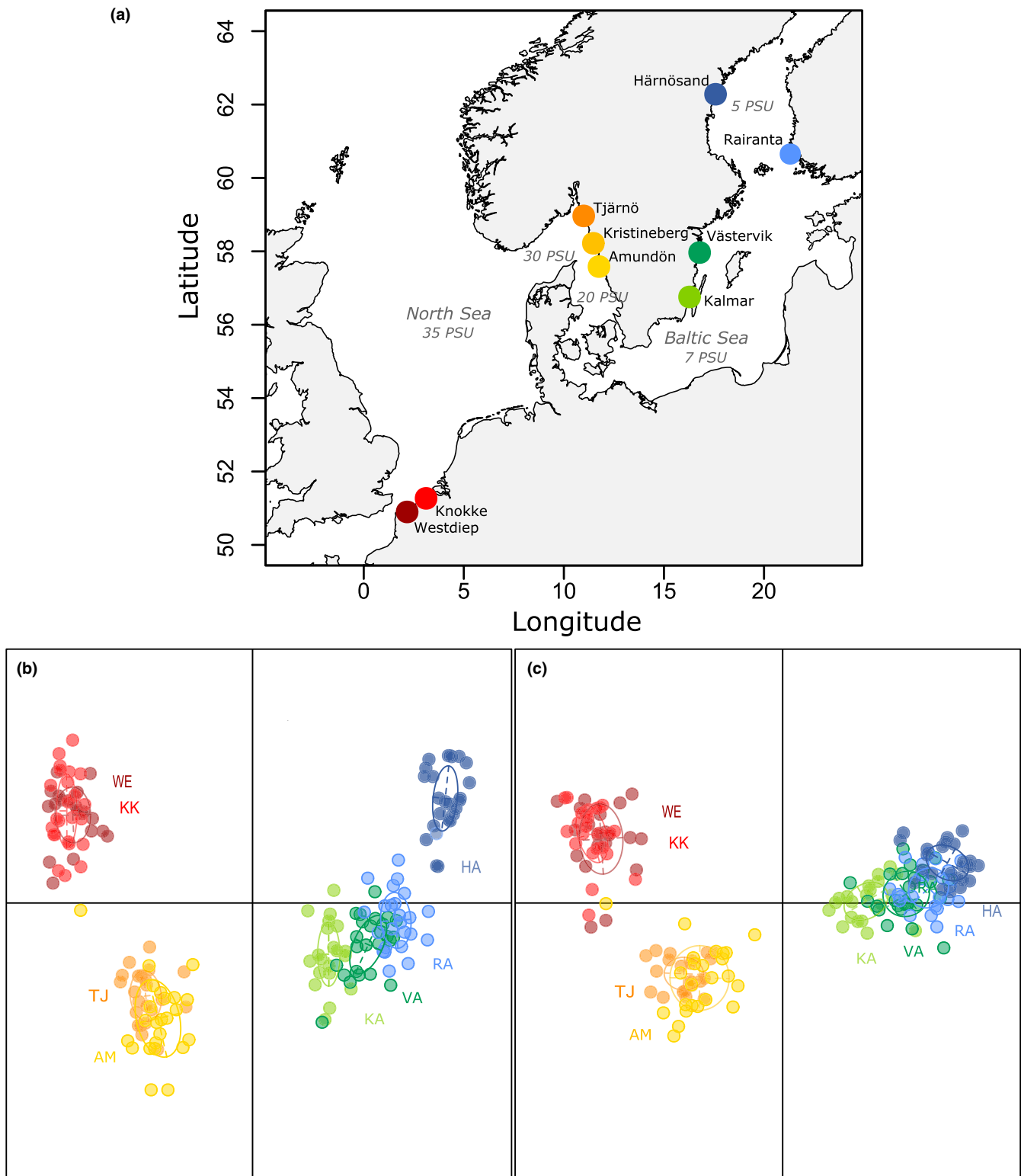


FIGURE 1 (a) Map of the North Sea–Baltic Sea transition zone with sampling sites. Kristineberg is the site for the sand goby reference genome individual. For details on sampling and site information, see Table S1. (b) Plot of principal component analysis (PCA) of all 22,190 sand goby SNP genotypes. Per cent of variation explained is 2.4982% on the x-axis and 0.9237% on the y-axis. (c) PCA plot of the significant outlier SNPs ($n = 314$) from BayeScEnv. Per cent of variation explained is 21.115% on the x-axis and 3.379% on the y-axis

(Tromsø) to southern Spain (Guadalquivir estuary), the northern Mediterranean Sea and the western Black Sea (Kullander et al., 2012; Miller, 1986). It is established in the Baltic Sea at least

up to Luleå and Oulu archipelagoes in the Bothnian Bay (Kullander et al., 2012), where salinity is as low as 1.8–3.8 PSU (Andersen et al., 2017). During reproduction, goby males provide paternal

care (Breder & Rosen, 1966). The sand goby male builds a nest under a substrate (e.g., bivalve shell, stone) and females attach their eggs to the ceiling of the nest (Forsgren, 1999). Both sexes breed repeatedly during their single breeding season. A male typically cares for eggs of multiple females and parasitic spawning is common (Järvi-Laturi, Lindström, Kvarnemo, & Svensson, 2011; Jones, Walker, Kvarnemo, Lindström, & Avise, 2001; Jones, Walker, Lindström, Kvarnemo, & Avise, 2001; Singer, Kvarnemo, Lindström, & Svensson, 2006; Svensson & Kvarnemo, 2007). Ecologically, gobies play an important role in the marine ecosystem both in the Baltic Sea and in the Atlantic. For example, seasonally some 50% of pelagic fish larvae in the northern Baltic proper are gobies (Parmanne & Lindström, 2003) with the sand goby being the most abundant (Ojaveer, Gross, Laur, Arula, & Klais, 2017). Sand gobies are benthic and as they are prey for many pelagic fish (Salvanes & Nordeide, 1993), they form an important trophic link between the benthic ecosystem and pelagic ecosystem.

Salinity has a direct impact on osmoregulation and many other traits of sand goby. Geographical differences in salinity also affect sand goby parasite composition (Zander & Kesting, 1998). Some parasites common to the North Sea (e.g., *Gyrodactylus gonda*) do not survive in the Baltic Sea (Huyse, Oeyen, Larmuseau, & Volckaert, 2017), whereas others such as *Aphalloides timmi* do (Zander & Reimer, 2002). The prevalence of some pathogens (especially *Saprolegnia*) that infect developing eggs increases with low salinities (Lehtonen & Kvarnemo, 2015a, 2015b). Consequently, the caring male of a brood increases its investment in pathogen prevention through behavioural adaptation, and females avoid nests that smell of *Saprolegnia* infected eggs (Lehtonen & Kvarnemo, 2015c). Salinity-driven differences in nest site availability shift sexual selection from female choice on the Swedish west coast to male–male competition for nest sites in the Baltic Sea (Forsgren, Kvarnemo, & Lindström, 1996; García-Berro, Yliportimo, Lindström, & Kvarnemo, 2019; Singer et al., 2006). Although sperm competition remains remarkably similar in the two environments, despite an expectation that nest site shortage in the Baltic Sea would have forced more males to resort to parasitic spawnings in that area (Jones, Walker, Lindström, et al., 2001; Singer et al., 2006). Additionally, two egg predator species, shore crab, *Carcinus maenas*, and netted dogwhelk, *Nassarius nitidus*, which are abundant at the Swedish west coast but absent in most of the Baltic Sea, affect both male care and nest-building behaviour (Järvi-Laturi et al., 2011; Lehtonen & Lindström, 2004; Lissåker & Kvarnemo, 2006).

2.2 | Sperm motility in relation to salinity

Reproductively mature males were collected from Härnösand and Kristineberg on 29 May 2012 (Figure 1a, Table S1). The fish were transported to the laboratory in temperature-controlled coolers filled with local water to within 24 hr of being caught. All fish were dissected 2 days after they had been captured. Only males

with breeding colouration were included in the study ($n = 8$ from Kristineberg; $n = 8$ from Härnösand).

Each male was euthanized (hit in the head followed by decapitation), immediately dissected, and had both testes removed. One testis was put into liquid nitrogen for later gene expression analysis. The other testis was used to provide a sperm sample for analysis of sperm motility (proportion motile sperm), assayed at high and low salinity. Before dissection, two tubes were prepared for each male, each with either 750 μ l of high salinity water (31 PSU, made from filtered natural seawater) or 750 μ l of low salinity water (six PSU, made from the same filtered natural water, but diluted with distilled water). Tubes were submerged in water of $15.5 \pm 0.5^\circ\text{C}$.

The assay of sperm motility followed Havenhand and Schlegel (2009). Briefly, 20 μ l Ca-free Ringer solution was added to each male testis after rupturing the tissue with a scalpel in an Eppendorf tube. From this solution, 8 μ l was suspended in the tube with high salinity water and 8 μ l in the tube with low salinity water. A 40 μ l drop of sperm suspension was placed between an albumin-coated microscope slide and coverslip, separated by a 0.75 mm thick O-ring. Sperm movement was recorded for 0.5 s at the midpoint of the drop, at 30 frames/s, using a digital video camera (PixelLink) mounted on a compound microscope (Leica DM-IL). Pilot experiments showed that illumination by the microscope lamp had no impact on the temperature inside the drop during video recording (time of slide on microscope: approx. 10 s). Nine sub-samples were taken as technical replicates: Three 40 μ l drops were placed on each of three slides, and each drop was recorded twice in immediate succession. These recordings of sperm movement were performed within 5 min after decapitation and within 30 s after placing a sample on the slide.

Videos were post-processed and analysed with ImageJ using the CASA plugin (Wilson-Leedy & Ingermann, 2007) to determine the proportion of motile sperm (i.e., sperm moving faster than 15 $\mu\text{m/s}$ on average). The effect of male origin on the proportion of motile sperm was tested with analysis of variance (ANOVA). We compared the males among themselves, with sperm of each male being tested at both high salinity and low salinity. Salinity was therefore used as repeated factor within each male. To achieve normality and homogeneous variances, the proportion motile sperm was arcsin square root transformed before analysis.

2.3 | Sample collection for DNA analyses

The fish used for the reference genome sequencing was a single individual male *P. minutus* caught in 2010 in Bökevik Bay ($58^\circ14'55''\text{N}$ $11^\circ26'51''\text{E}$) close to Kristineberg Marine Research Station. We also collected fish from eight localities along the salinity gradient, ranging from approximately 35 PSU to 4 PSU, from the southern North Sea (Belgium), Skagerrak, Kattegat, Baltic Sea proper and Gulf of Bothnia (Figure 1a, Table S1). We processed 25 individuals per population for genotyping. Sex was recorded for 89 of these 200 fish, and 42%

were female and 58% were male. Since the sex-determining mechanism is unknown, genotypic sexing was not possible. All fish were caught in sandy and shallow habitat from mid-spring to late-autumn. Each individual was euthanized in an overdose of MS222 or a hard hit on the head followed by decapitation. Samples were stored in alcohol prior to analyses. This study was performed under Swedish ethical permits 135-2010 and 143-2012.

2.4 | Laboratory methods

For the genome sequenced individual, genomic DNA was extracted using a simplified CTAB-protocol (Panova et al., 2016). In short, fragments of fins and muscles were digested in CTAB-buffer, mercaptoethanol and Proteinase K. RNA was digested by RNase A. Thereafter, DNA extracted with chloroform-isoamylate and precipitated with isopropanol. The DNA was washed with 70% alcohol and stored in 10% TE-buffer at -70°C before being sent to SciLife Stockholm for sequencing.

For the remaining samples, DNA was extracted from fin tissue or whole embryos using a salt extraction method. Tissue was digested in 200 μl of extraction buffer (0.4 M NaCl; 10 mM Tris-HCl pH 8.0; 2mM EDTA pH 8.0), 20 μl of 20% Sodium Dodecyl Sulphate (SDS) and 5 μl of Proteinase K (Promega) for 1.5–2 hr at 65°C . Proteins were precipitated with 150 μl 6M NaCl and then centrifuged at 11,000 g for 30 min. After centrifuging, the supernatant was removed to a new tube and isopropanol was added in 1:1 ratio by volume. The samples were inverted several times to mix, and DNA was allowed to precipitate for 15–30 min. Samples were then centrifuged at 13,000 g for 15 min at 4°C . The supernatant was removed, and 300 μl of ice cold 70% EtOH was added to wash the pellet. Samples were then centrifuged at 4°C for 5 min at 13,000 g. The ethanol was carefully removed, and the pellet was allowed to dry at RT, after which 30 μl of nuclease-free H_2O (SIGMA) was added to the tube. Quantity and quality of the DNA samples were estimated using the Nanodrop 1000x (Thermo Scientific, Vantaa, Finland) and by electrophoresis on a 1.5% agarose gel.

RNA was extracted from tissues with phenol-based phase separation (Tri-reagent, Ambion, Life Technologies) following the Trizol-protocol from SIGMA with minor modifications as follows: 1-bromo-3-chloropropane (SIGMA) was used instead of chloroform, and RNA samples were washed twice with 75% EtOH. After the first extraction with Tri-reagent DNase, treatment with RQ1 (RNase-free DNase, Promega) was performed to ensure no traces of DNA were left in the RNA. To remove any residue of the enzyme and DNA, the RNA samples were extracted a second time with amounts of reagents downscaled for 400 μl of Tri-reagent. The RNA samples were dissolved into 30 μl of nuclease-free H_2O . The RNA concentration was measured with Nanodrop prior to analysing quality on the Bioanalyzer (Agilent).

For genotyping by sequencing (GBS), we used the method of Elshire et al. (2011) with size selection using SPRI-Select beads (Beckman Coulter). DNA concentrations were measured using a

NanoDrop ND-1000 spectrophotometer. DNA (100 ng) was digested with the restriction enzyme Pst1 HF 1.5 U (New England Biolabs) and 1x Cut Smart Buffer in a total volume of 20 μl in a thermocycler (37°C , 1 hr; 75°C , 15 min; 4°C , 10 min). The digested DNA was ligated to adapters with T4-ligase 0.6x (New England Biolabs), 1x Ligase Buffer, 21 μl dH_2O and 50 nM of pooled forward and reverse adapters, which were prepared according to Elshire et al. (2011), (ligation program: 22°C , 1 hr; 65°C , 30 min; 4°C , 10 min). Up to 96 unique barcodes were used in each library to label individual samples. The ligation products were pooled into libraries and purified using SPRI-select beads. The purified libraries were PCR amplified with the following components: Purified ligated library (23.5 μl), 1.5 μl primer mix (12.5 μM) and 25 μl KAPA HIFI Hot start 2X Mastermix (amplification program: 72°C , 5 min; four cycles [95°C , 30 s; 95°C , 10 s; 65°C , 30 s; 70°C , 30 s]; 11 cycles [95°C , 10 s; 65°C , 30 s; 72°C , 20 s]; 72°C , 5 min; 4°C , 10 min). After amplification, libraries were purified and size selected using SPRI-select beads and a 0.9–0.7x selection range. The final size was then confirmed on a Bioanalyzer (Agilent). All GBS libraries were then used to generate 100 bp paired-end reads using the Illumina HiSeq 2500 platform at the Beijing Genome Institute (BGI).

2.5 | Genome assembly

Thirteen libraries of varied insert size (160–510 bases) and read length (101–301 bases) were generated and used for paired-end sequencing on an Illumina HiSeq 2000 or 2500 (Table S2). In addition, two long-insert size mate-pair libraries (2 and 5 kb) were sequenced on Illumina HiSeq 2500 (Table S2).

Sequences were processed with cutadapt v1.12 (Martin, 2011) to remove partial adapters and bad quality sequences. Paired reads with one read shorter than 40 bases were also removed. Flash v1.2.11 (Magoč & Salzberg, 2011) was used to merge paired-end reads that had 10 or more bases overlapping. Lighter v1.1.1 (Song, Florea, & Langmead, 2014) was then used to correct sequencing errors via kmer correction. Sequences were scanned using a kmer length of 23 and an estimated genome size of 0.9 GB. Reads were trimmed or discarded if they could not be corrected. A python script (https://github.com/enormandeu/Scripts/blob/master/fastq_CombinePairedEnd.py) was used to re-pair fastq files since Lighter may remove one sequence of a pair without removing the mate. Soapdenovo2 v2.04 (Luo et al., 2012) was then used to assemble a draft genome using kmer size 29, and GapCloser v1.12 was used to fill in the gaps in the scaffolds. Input sequence libraries, insert sizes, read numbers and the Soapdenovo2 config file are available in the Supplementary Information.

2.6 | Variant calling for genotyping by sequencing

The reads generated for the GBS analysis were demultiplexed and barcodes were removed using a custom Perl script. Low-quality bases

and leftover adapter sequences were then removed using cutadapt v1.8 (Martin, 2011). Paired-end reads for each individual were aligned to the sand goby reference genome using Bowtie2 v2.3.2 (Langmead & Salzberg, 2012). The SAM files were subsequently converted to sorted BAM files using samtools v0.1.19 (Li et al., 2009), and variants were called using the samtools mpileup function with extended BAQ computation, in combination with bcftools v0.1.19 (Li et al., 2009). Sites at which bcftools identified multiple variant types (SNPs, indels and multi-base polymorphisms) were removed, leaving 25,256 successfully genotyped variant sites. VCF files were converted to the input format for Plink v1.07 (Purcell et al., 2007). The following filters were applied in Plink: hwe <0.001 (run separately on populations), maximum missing genotypes = 0.20, minor allele frequency >0.02, >90% missing individual data removed (e.g. --mind 0.20 --geno 0.1). Removal of individuals with too much missing data left 24 individuals in Amundön, Knokke and Rairanta, 23 in Tjärnö, 21 in Westdiep and the original 25 in the remaining populations. Further filtration based on minor allele frequency was performed to remove SNPs with minor allele frequency <0.02 for the population genetic analyses (22,190 loci).

2.7 | Annotation of transcriptomes

RNA sequences from testicular tissue from one individual collected at Kristineberg in 2012 and brain, liver, spleen and ovary from two individuals (brain from one and liver, spleen and ovary from the other) collected at Tjärnö in June 2013 were used to assemble transcriptomes using Trinity v2.6.6 (Grabherr et al., 2011) and parameters for a stranded library (RF), no read normalization and a minimum kmer coverage of 2. GMAP v2019-06-10 (Wu, Reeder, Lawrence, Becker, & Brauer, 2016) was used to map the assembled transcripts to the reference genome in order to annotate coding regions. We used Diamond (Buchfink, Xie, & Huson, 2015) to functionally annotate each transcript, and similarity searches were performed for each transcript against the zebrafish, *Danio rerio* (Ensembl 96 and Uniprot Reference Proteome, release 2019_04) and mudskipper, *Periophthalmus magnuspinnatus* (Ensembl 96) protein databases using the parameters --sensitive --frameshift 15 --top 5 --min-score 100 --id 60 --unal 1 --outfmt 6 and keeping the blast trace-back operations string (btop) and the query sequence matching the target (qseq). The logic for this procedure is that frameshifts are often introduced into the transcript sequence during the Trinity assembly. By using the strand information provided by the btop string to adjust the frameshift, a single open reading frame can be obtained from the transcript. We parsed the query sequence according to the btop string, thus removing any frameshifts and then translated them using the tool fastatranslate (in reading frame one) from Exonerate v2.2. CD-HIT v4.6.8 (Li & Godzik, 2006) was used to cluster the transcript sequences in order to remove some of the redundant amino acid sequences (one representative sequence per cluster where retained). The resulting sequences were used in Orthofinder to determine the most likely orthologues (Emms & Kelly, 2015). In order to identify orthogroups in Orthofinder, we used the Ensembl 96 protein databases for tigertail seahorse

(*H. comes*), three-spined stickleback, zebrafish, and mudskipper as well as the Uniprot reference proteome (release 2019_4) for zebrafish.

2.8 | Differential expression in testes

Testes from the fish in the sperm motility experiment were compared for differential expression. Sequences obtained from the RNA of the testes were processed using Cutadapt v1.8 (Martin, 2011) to remove partial adapter sequences and bad quality bases, as well as read pairs when one of the pairs was less than 40 bases. Reads were mapped to the sand goby draft genome using STAR v2.7.2 (Dobin et al., 2013). A reference genome index was built in STAR using the testes transcriptome from the multiple tissue samples (above). This sample was used for annotation because the sequences were generated using a stranded library protocol whereas the testis samples from the salinity experiment were not. The reads from this one sample were mapped to the genome using STAR with a pair-end overlap threshold of 10 and other default parameters. Cufflinks v2.2.1 (Trapnell et al., 2010) was used to assemble the transcripts and create the gene transfer format (GTF) file. A new genome index was then created using this GTF file.

All testis samples from the salinity experiment were mapped using this genome index, and the splice junction files were retrieved. Splice junction files were concatenated and filtered by choosing only junctions that were supported by more than five uniquely mapped reads in one individual. All samples were then mapped again using both the gtf file and the splice junction file, and gene-level read counts were determined using the `quantMode GeneCounts` option. EdgeR v3.5.2 (Robinson, McCarthy, & Smyth, 2010) was used to normalize expression using the TMM method (Robinson & Oshlack, 2010) and remove genes with low expression (minimum count = 25, minimum total count = 100). Then, a linear model, limma trend analysis (Law, Chen, Shi, & Smyth, 2014), was implemented using the R package limma v3.5.1 (Ritchie et al., 2015) to assess differential expression. Transcripts were considered differentially expressed if an adjusted p-value less than 0.05 was observed (Benjamini & Hochberg, 1995). For transcripts for which no orthogroup was identified, a search was performed using BLASTX, with default parameters and the nonredundant protein database (nr) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

2.9 | Population differentiation

Gene diversity (expected heterozygosity) was calculated and averaged over loci for each of the eight samples. Pairwise F_{ST} values (Weir & Cockerham, 1984) between all populations were calculated using ARLEQUIN v3.5 (Excoffier & Lischer, 2010), and their significance assessed using 10,000 permutation steps. Significance was adjusted for multiple testing using FDR at $q = 0.05$ (Benjamini & Hochberg, 1995). Genetic isolation by geographic distance was estimated by correlating the most direct waterway distances calculated with Google earth and pairwise genetic differentiation, $F_{ST}/(1 - F_{ST})$ (Rousset, 1997).

Individual ancestry and the number of genetic clusters (K) were assessed using STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000), allowing admixture, with uncorrelated allele frequencies but no locprior. Three replicates of 50,000 Monte Carlo Markov chain (MCMC) iterations (discarding the first 10,000 iterations as burn-in) were performed, testing for $K = 1-5$. Convergence was assumed by observing consistent results in all three replicates. In addition, principal component analysis (PCA) was performed to visualize genetic differentiation among individuals using pcadapt (Luu, Bazin, & Blum, 2017).

2.10 | Inference of the demographic history

We used the software $\delta a \delta i$ v2.0.1 (Diffusion Approximations for Demographic Inference; Gutenkunst, Hernandez, Williamson, & Bustamante, 2010) using the four step likely optimization procedure implemented by Portik et al. (2017) to understand the demographic history at the origin of the main population break, localized between the North Sea and the Baltic Sea. The software infers the demographic history by diffusion approximation, which in the absence of an outgroup uses the folded version of the joint Site Frequency Spectrum (jSFS) derived from the minor allele frequency. Here, we compared 13 biologically relevant scenarios of divergence (Figure 2) using only the singletons and the SNPs polymorphic in the focal population. These were then further pruned for physical linkage disequilibrium by keeping 1 random SNPs per 1kb. Analysis included one scenario of strict allopatry ('Strict Isolation, SI' Figure 2a), and three scenarios with asymmetric migration, which can be continuous along the process of divergence ('Isolation with migration, IM', Figure 2b), or discontinuous either after a primary phase of allopatry ('Secondary Contact, SC', Figure 2c) or limited to the beginning of the divergence ('Ancestral Migration, AM', Figure 2d). Each scenario including migration was further tuned to include the possibility of exponential growth rate (G) of the populations (Neutral + Growth in Figure 2e-g). This growth occurred after the split for the IM model (IMG), after the secondary contact for the SC model (SCG) and after the isolation in the AM model (AMG). An inferred growth rate above one results in a positive growth as represented with blue population (Figure 2e-g), and a growth rate between zero and one results in a negative growth of the population as represented for the red population (Figure 2e-g). The remaining model (IM2m, SC2m, AM2m, IM2mG, SC2mG and AM2mG, Figure 2h-m) included the possibility of heterogeneous gene flow (2m) to the model with migration in order to represent case of ongoing speciation (Wu, 2001, and see Rougemont et al., 2016; Rougeux, Bernatchez, & Gagnaire, 2017 for further example using $\delta a \delta i$). Precisely, these models with heterogeneous gene flow incorporate two categories of loci, one with a neutral migration (m) and the other one with reduced migration (mr), which affect a proportion P and $P-1$ of the genome, respectively (Figure 2d,e). Each model was run with 40 independent replicates, each of which performed 40 pseudo-replicate during the four step optimization after the four optimization procedure (Portik

et al., 2017). The 10 best replicates for each of the models were then kept and used to compare the ability of each scenario to reproduce the observed jSFS using the Akaike information criterion (AIC). As, in particular, the samples from the Swedish west coast (Tjörn, Amundön) are found at the Baltic Sea transition zone, they are likely to be affected by recent gene flow that might swamp the effect of ancient demographic history. Therefore, we only used the sampling site Westdiep from Belgium as representative of the Atlantic lineage and inferred the demographic history with Härnösand (the most northern site in the Baltic Sea) and with Kalmar (the most southern site in the Baltic Sea) to evaluate whether the inference converges to a similar demographic scenario. The inferences were based on subset of 5,727 and 5,754 SNPs from the 20,727 and 21,464 polymorphic SNPs in the two pairs of focal populations. As the parameters estimated by $\delta a \delta i$ are given relative to θ (the population mutation rate), we used the approach described in Rougeux et al. (2017) to convert the parameter estimates into meaningful biological values using a generation time of 1 year and a mutation rate of 10^{-8} .

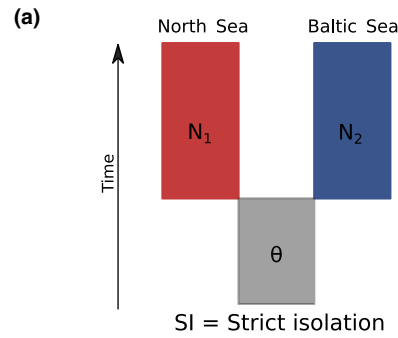
2.11 | Genetic variation associated with salinity

To detect high F_{ST} outliers potentially underlying traits involved in local adaptation, we used two methods. First, the Bayesian regression approach implemented in Bayescan v2.1. (Foll & Gaggiotti, 2008), using default settings with the following exceptions: 75,000 Markov chain Monte Carlo (MCMC) iterations with thinning 15 and prior odds 1,000. False discovery rate (FDR) cut-off was set at $q = 0.05$ (Benjamini & Hochberg, 1995). Convergence of the MCMC chain was assessed using the software CODA (Plummer, Best, Cowles, & Vines, 2006). Second, we used the frequentist approach FDIST2 by Beaumont and Nichols (1996) implemented in the software ARLEQUIN v3.5.2.2 (Excoffier & Lischer, 2010). Here, we used a hierarchical island model to reduce the rate of false positives which may be inflated if samples belong to a hierarchically subdivided population or if some samples have a recent shared history, for example at range expansion (Excoffier et al., 2009). We used four groups representing the four main water basins: the southern North Sea (Westdiep and Knokke), the eastern North Sea (Tjörn and Amundön), the southern Baltic Sea (Västervik and Kalmar), and northern Baltic Sea (Rairanta and Härnösand). The number of coalescent simulations performed was 50,000 and we used an FDR cut-off at $q = 0.05$.

In addition to the high F_{ST} outliers, we identified outlier loci associated with salinity, using BayeScEnv v1.1 (de Villmereuil & Gaggiotti, 2015). This software is an extension of Bayescan and can test hypotheses for specific drivers of local adaptation. We used default settings and prior jump 10, thin 15 and FDR $q = 0.05$. Convergence of the MCMC chain was assessed using CODA (Plummer et al., 2006).

Outlier SNP positions from the BayeScEnv analysis were converted to bed format and the bedtools window command from the BedTools Suite v2.28.0 (Quinlan & Hall, 2010) was used to identify transcripts within a window ranging 4,000 bp upstream and

Without migration:



With migration:

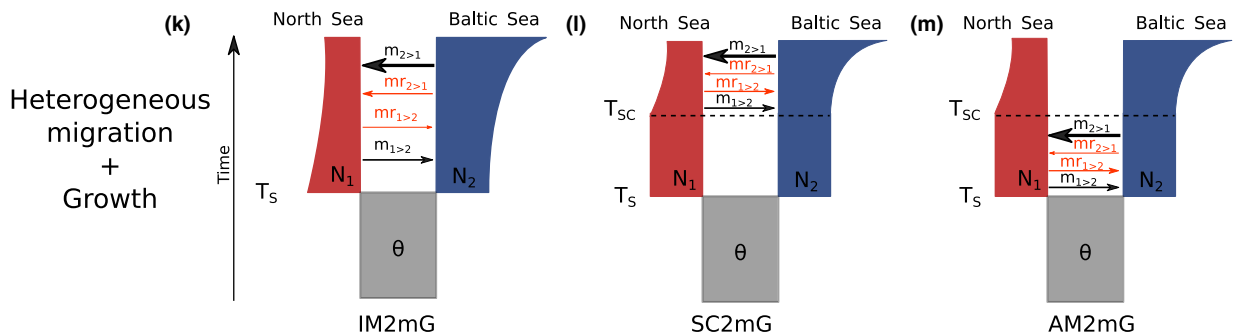
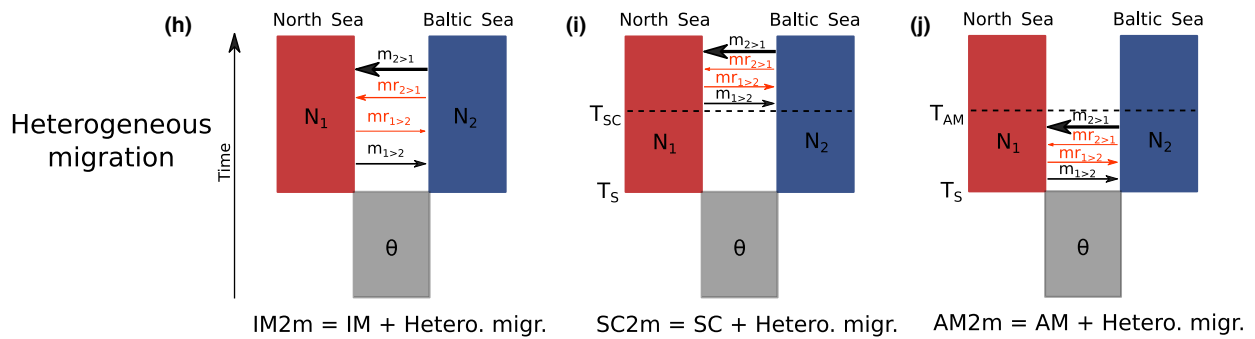
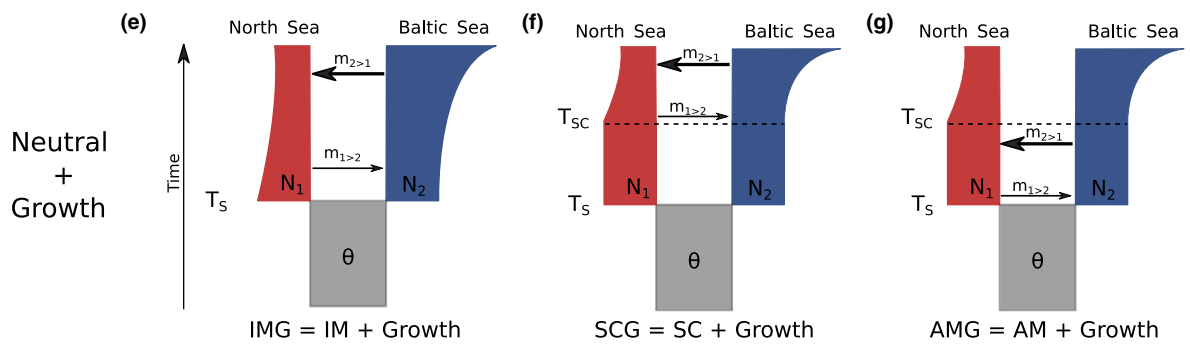
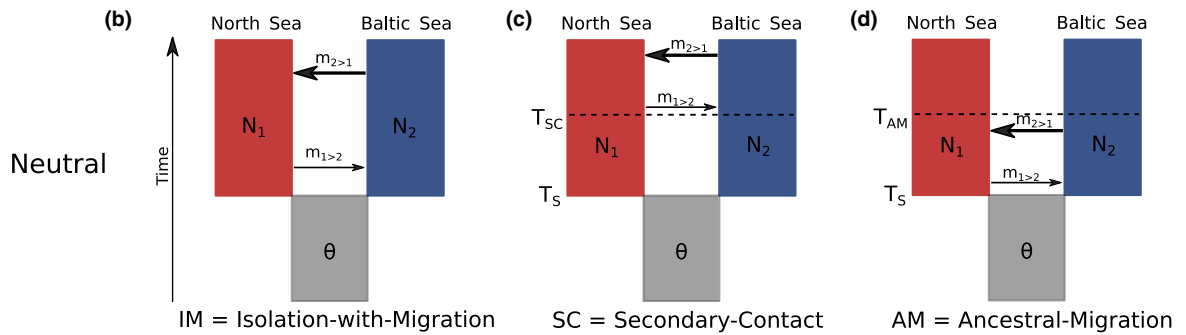


FIGURE 2 Schematic representation of the 13 demographic models tested where an ancestral population of size N_a splits into two populations of size N_{e1} and N_{e2} at a time (T). Populations are split and evolved in (a) absence of gene flow (SI: strict isolation), (b) in the face of continuous gene flow (IM: isolation with migration), (c) with gene flow after a primary phase of strict isolation (SC: secondary contact), (d) gene flow at the beginning of the divergence (AM: Ancestral migration), (e–g) model with gene flow IM, SC and AM including a population growth (IMG, SCG and AMG), (h–j) model with gene flow IM, SC and AM including heterogeneous migration rate (IM2m, SC2m, AM2m), and (k–m) model with gene flow IM, SC and AM including heterogeneous migration rate and population growth (IM2mG, SC2mG and AM2mG)

downstream of outlier SNPs. The transcripts were annotated using the three-spined stickleback and zebrafish orthogroup members (from above annotation) when possible. For the transcripts falling within the overlap window, the zebrafish identifiers were searched for protein–protein interactions and functional enrichment using STRING v11.0 (Szklarczyk et al., 2019).

2.12 | Availability of data and material

The head of the sequenced sand goby male has been deposited as a voucher at the Ichthyology Collection of the Swedish Museum of Natural History (NRM) with catalogue number NRM 69,326. In addition, 20 males and 20 females caught at the same location

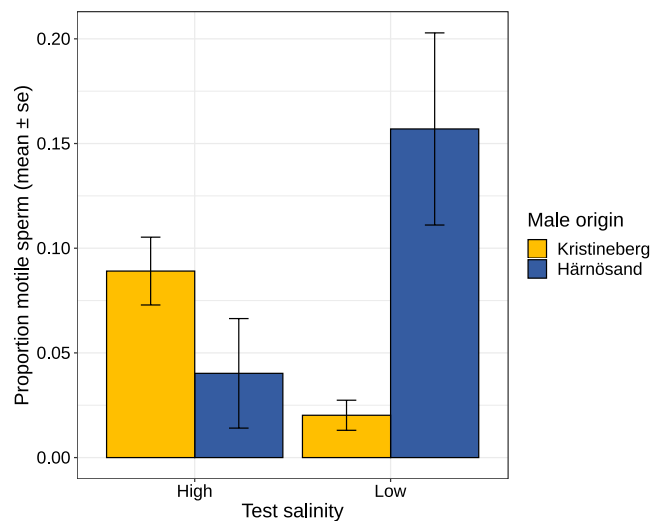


FIGURE 3 Sperm motility (proportion motile sperm) in relation to male origin (Kristineberg = high salinity and Härnösand = low salinity) and test salinity (high = 31 PSU, low = 6 PSU). Nontransformed values are shown, although the analysis was done on arcsin square root transformed data

(58°14'55"N 11°26'51"E) in May 2016 are deposited at NRM (catalog numbers NRM 69,327, NRM 70,138, NRM 70,139 and NRM 70,140). The draft scaffolded genome assembly is available in Dryad.

3 | RESULTS

3.1 | Sperm motility: effects of male origin and salinity

Sperm motility (proportion of sperm that were motile) was affected by test salinity and male origin, as shown by a significant interaction between the two factors. Specifically, a larger proportion of sperm were motile in the salinity that matched with the male origin than in the reciprocal salinity (repeated measures ANOVA: male origin * test salinity: $F_{1,14} = 114.45$, $p < .001$; Figure 3).

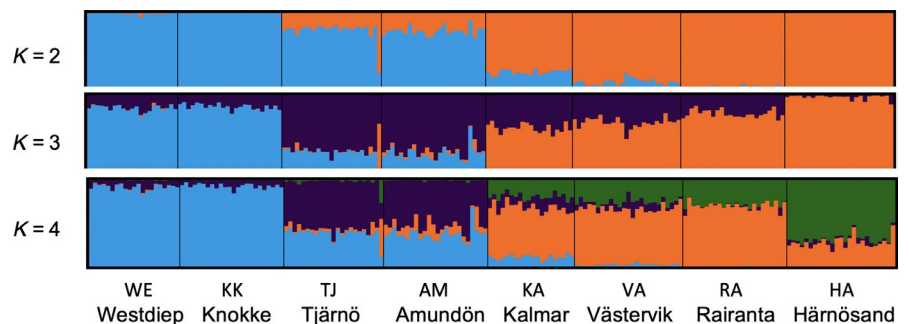
3.2 | Genome sequencing and assembly

The assembled genome contains 204,891 scaffolds/contigs ranging in length from 100 to 2,618,418 bases with 21,093 contigs longer than 500 bases. Assuming a genome size of 1 Gb, the NG50 is 127,739 bases. Using the actinopterygii_odb9 database with BUSCO v 3.1.0 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015), we obtained a score of 90.2% complete genes (87.9% complete and single-copy; 4.5% fragmented genes) and 5.3% missing genes.

3.3 | Population structure

Heterozygosity levels were slightly lower in the North Sea, but overall similar for the eight populations (Table S3).

FIGURE 4 STRUCTURE analysis showing the three most likely scenarios of clustering, with $2 \leq K \leq 4$ genetic clusters. Locations are ordered from high to low salinity. The steepest decline in salinity occurs between Amundön and Kalmar



Population-based analysis of pairwise F_{ST} (Table S4a), as well as individual-based analyses using PCA and STRUCTURE, identified a major genetic division between Baltic Sea and North Sea populations of sand goby (Figures 1b and 4, and Figure S2). The main split overlaps with the steepest salinity shift (Figure 1a). The two main populations were further subdivided into an eastern and a southern cluster of subpopulations in the North Sea, and a northern (with only one subpopulation, Hännösand) and a south-eastern cluster in the Baltic Sea (Figure 1b). Notably, the differentiation between the Hännösand subpopulation and the other Baltic Sea subpopulations largely disappeared when only outlier loci were analysed, suggesting that this difference was more related to genetic drift effects than to divergent selection (Figure 1c). If plotted on waterway distance, the population structure corroborates a pattern expected from isolation by distance ($r^2 = .87$; $p < .001$; Figure S1).

3.4 | Demographic history

Both model comparisons (Westdiep vs. Hännösand, and Westdiep vs. Kalmar) showed that the most likely scenario describing the origin of the North Sea–Baltic Sea divergence involved past isolation followed by a secondary period of gene flow with heterogeneous migration rate along the genome (i.e. secondary contact with heterogeneous migration, SC2m) (Table 1 and Figure S3). Under such scenario, the isolation period was estimated to be 10 to 15 times longer than the period of secondary contact (ratio T_{split}/T_{sc}). By using a generation time of 1 year and a mutation rate of 10^{-8} , the divergence started 200kya (Table S5), whereas the secondary contact period was estimated to have happened 17 kya (15 kya for WE vs. HA and 20 kya for We vs. KA, Table 1 and Table S5). Interestingly, the population of Hännösand (inner Baltic Sea) was inferred to be half of the size of Westdiep (400k vs. 200k, Table S5), whereas Kalmar, the

TABLE 1 Results from the best run for each of the four demographic scenarios (secondary contact, SC; ancestral migration, AM; isolation with migration, IM; and strict isolation, SI) for the two comparisons (Westdiep in the North Sea vs. a northern Baltic Sea population, Hännösand, or vs. a southern Baltic Sea population, Kalmar)

Comp.	Model	Lik.	AIC	Δ_{AIC}	wAIC	θ	N_{eNS}
Westdiep vs. Hännösand	SI	-1,278	2,562	1,196	0	1,200	2.71
	IM	-751	1,512	146	0	562	5.43
	IMG	-756	1,523	157	0	474	0.94
	IM2m	-686	1,388	22	0	499	5.58
	IM2mG	-674	1,368	2	0.25	497	1.89
	SC	-702	1,415	49	0	613	2.96
	SCG	-690	1,397	30	0	706	3.10
	SC2m	-674	1,366	0	0.75	614	3.43
	SC2mG	-719	1,461	95	0	469	3.21
	AM	-781	1,573	207	0	558	5.83
	AMG	-777	1,571	205	0	614	4.47
	AM2m	-718	1,455	89	0	134	8.90
	Westdiep vs. Kalmar	AM2mG	-701	1,424	58	0	155
SI		-1,247	2,500	1,250	0	1,178	8.31
IM		-705	1,420	169	0	470	6.30
IMG		-701	1,417	166	0	285	0.88
IM2m		-631	1,278	27	0	514	4.03
IM2mG		-623	1,267	17	0	262	0.87
SC		-651	1,314	64	0	516	3.42
SCG		-634	1,284	33	0	518	4.67
SC2m		-616	1,250	0	1	530	2.82
SC2mG		-639	1,299	49	0	554	5.56
AM		-707	1,425	175	0	287	16.16
AMG		-721	1,458	208	0	452	1.36
AM2m		-682	1,381	130	0	181	3.92
AM2mG	-655	1,331	80	0	136	4.01	

Note: The table shows the difference in AIC with the best model (Δ_{AIC}) and the weighted AIC (Rougeux et al., 2017), the likelihood of the model (lik.), the population mutation rate (θ), the effective population size of the North Sea and the Baltic Sea populations (N_{eNS} and N_{eBS}), the growth factor (b_{NS} and b_{BS}), the per generation migrants ($m_{NS>BS}$ and $m_{BS>NS}$), the reduced per generation migrants ($mr_{NS>BS}$ and $mr_{BS>NS}$), the time of split (T_s) and the time of secondary contact or time of ancestral migration (T_{SCorAM}). All parameter estimates are relative to θ (see $\delta a \delta i$ manual for more explanation), and see Suppl. Table S5 for transformations to values in biological meaningful units. Models providing the best fits are highlighted in bold.

other Baltic Sea population, shows a similar effective population size to Westdiep. In both scenarios, the proportion of the genome resisting gene flow from the North Sea was higher in Hännösand (13%) than in Kalmar (7%), suggesting a higher introgression rate in the population closer to the main break between lineages. The model with continuous divergence including heterogeneous gene flow and rapid population growth (IM2mG) was the second-best model in both replicates (Table 1) and provides nearly as good fit as the SC2m scenarios for the Westdiep versus Hännösand comparison ($\Delta_{AIC} < 2$, fit Figure 5). This model suggests that the Baltic Sea was initiated by a founder population of 10% of the ancestral population that experienced an exponential growth at a factor of 16 (Table 1). In this model, the split between the North Sea lineages and the Baltic Sea lineages was estimated to happen about 300 kya (Table S5), which predates the last glacial maximum and the opening of the Baltic Sea about 8,000 years ago.

3.5 | Outliers

A significant number of outliers were identified, 252 ($q < 0.05$) from the Arlequin hierarchical analysis, 231 from the Bayescan analysis and 314 from the BayeScEnv environmental association analysis. Of these, 148 were shared between all three methods (Figure 6). The pairwise differentiation for the outliers was about ten times higher as compared to the differentiation using all SNPs (Table S4b). Of the BayescEnv outliers, 191 SNPs from 147 different scaffolds were within a gene or within 4,000 bases of the start or end of a gene region identified by mapped transcripts. However, no significant Gene Ontology terms or Kegg Pathways were enriched when all annotated loci overlapping the outliers were examined; however, there was an enrichment for the Uniprot key words 'Magnesium' and 'Mitochondrion'. When examining the gene lists, several likely candidate genes for

N_{eBS}	b_{NS}	b_{BS}	$m_{NS>BS}$	$m_{BS>NS}$	$mr_{NS>BS}$	$mr_{BS>NS}$	T_{split}	T_{SCorAM}	P
24.39							0.13		
0.75			2.19	9.50			1.03		
1.87	9.92	0.54	1.39	7.17			1.94		
0.98			0.35	19.58	0.24	1.55	1.34		0.80
0.10	4.16	16.69	2.41	9.66	0.15	1.44	1.66		0.83
1.50			2.56	9.31			0.83	0.13	
1.80	0.21	12.49	10.31	18.22			0.59	0.03	
1.94			6.42	13.91	4.40	1.41	0.77	0.07	0.87
1.61	3.85	0.28	0.27	19.78	5.33	0.09	0.95	8.54	0.98
0.53			0.05	27.52			1.05	0.01	
4.59	0.85	0.17	5.37	18.22			0.79	0.02	
9.84			2.28	1.40	0.02	0.20	8.17	0.02	0.88
5.75	0.48	1.30	1.52	7.45	0.08	0.20	6.72	0.04	0.88
8.32	0.13								
1.15			3.59	9.61			1.33		
8.73	0.94	2.83	18.18	7.26			2.62		
2.57			8.94	8.13	0.38	0.25	1.09		0.91
1.84	5.92	6.10	7.33	2.97	0.44	0.12	4.03		0.85
2.89			3.08	9.92			1.04	0.12	
1.39	9.81	14.61	1.02	6.56			0.99	0.19	
2.91			9.46	8.66	1.28	0.56	1.00	0.10	0.93
1.23	0.54	1.99	4.62	10.19	1.72	9.61	0.50	0.10	0.93
2.89			5.23	4.52			2.94	0.03	
6.71	8.96	0.24	0.66	14.00			1.76	0.01	
10.90			4.19	1.97	0.09	0.18	5.47	0.03	0.98
15.81	10.33	5.03	8.14	1.10	0.23	0.15	7.37	0.04	0.88

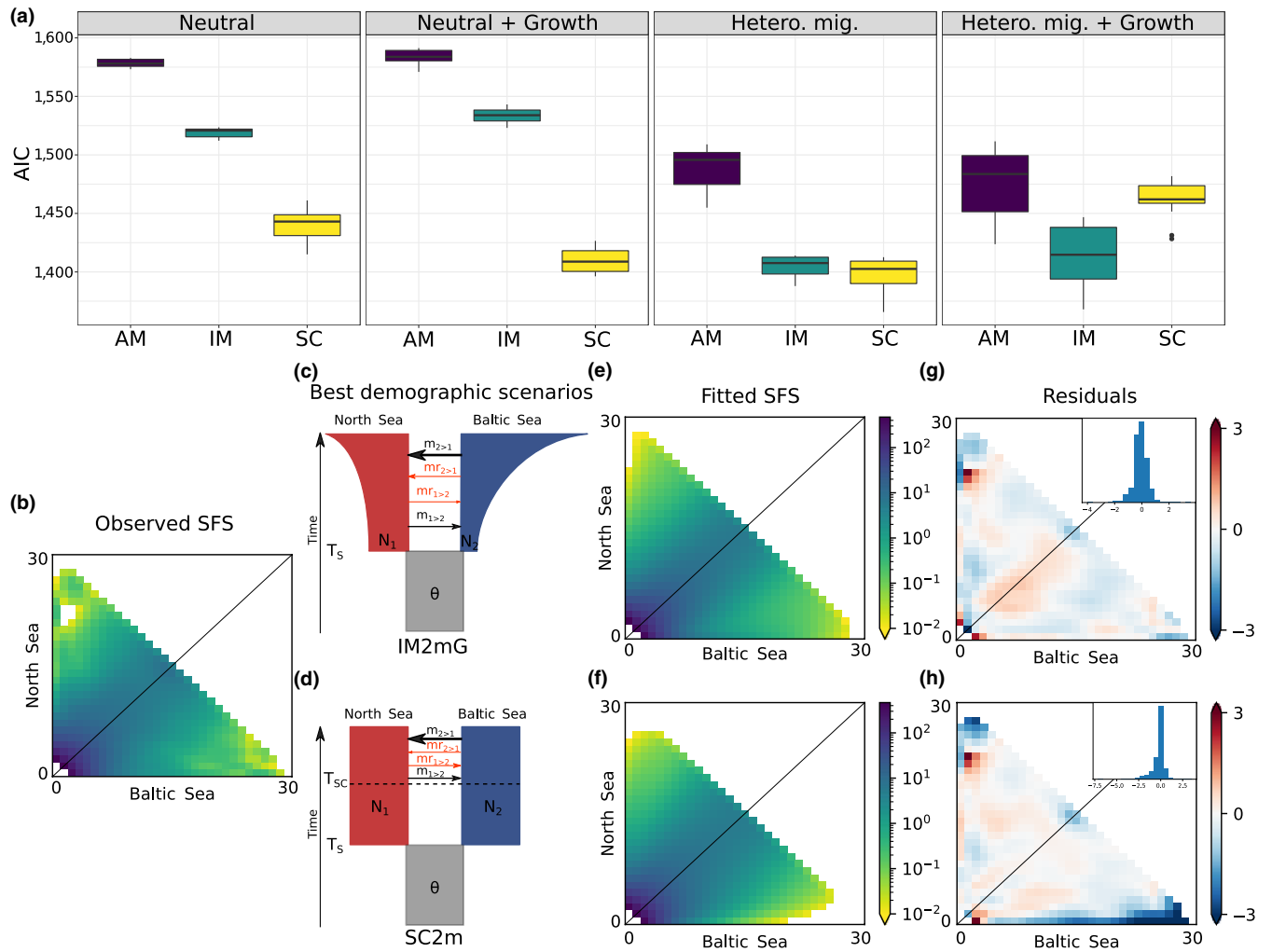


FIGURE 5 Demographic inferences for Westdiep in the North Sea versus a northern Baltic Sea population, Hårnösand: (a) shows the boxplot of the 10 best runs for each of the models (SI not shown as it performed poorly), (b) shows the observed jSFS, (c and d) show the schematic representation of the second most likely (IM2mG) and the most likely (SC2m) scenarios at the origin of population differences, (e and f) show the fitted SFS resulting from the scenarios c and d, and (g and h) show the residuals of the models c and d together with the distribution of the residuals in the subplot

environmental differentiation were observed. Genes involved in ion transport (eight genes) immune function (19 genes), visual perception (one gene) and mitochondrial interactions (10 genes) were found to be close to outlier SNPs (Table S6). Twenty-one BayescEnv outliers with high statistical significance ($q < 0.0001$) were, however, not adjacent to genes identified in our transcriptome analyses.

3.6 | Testes gene expression: effect of male origin

Differential transcript expression in testicular tissue compared between the Hårnösand and Kristineberg populations revealed only seven differentially expressed genes, and only three of these seem to be protein-coding genes (serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit beta isoform, translocator protein, and one uncharacterized protein). The other four genes did not appear to be protein coding since they lack unambiguous open reading frames and a significant similarity to known protein-coding genes found in NCBI.

4 | DISCUSSION

4.1 | Population structure, demographic history and outlier loci

Our study identified two distinct populations of sand goby in the NE Atlantic Ocean, a fully marine population living in the North Sea and a brackish water population living in the Baltic Sea. In addition, we found differentiation among subpopulations within each basin and an overall pattern of isolation by distance, both supporting substructure within each population. Nevertheless, most differentiation is present over the saltwater-brackish water transition at the entrance of the Baltic Sea (Figures 1 and 4), and a large number of outlier loci correlate with this salinity transition (Table S6). Distinguishing secondary contact and isolation with migration models is difficult, and with multiple models that fit almost similarly well, interpretations should be done with caution, also taking into account that other alternative models may fit better than the models currently being compared. This said the demographic inference of both models

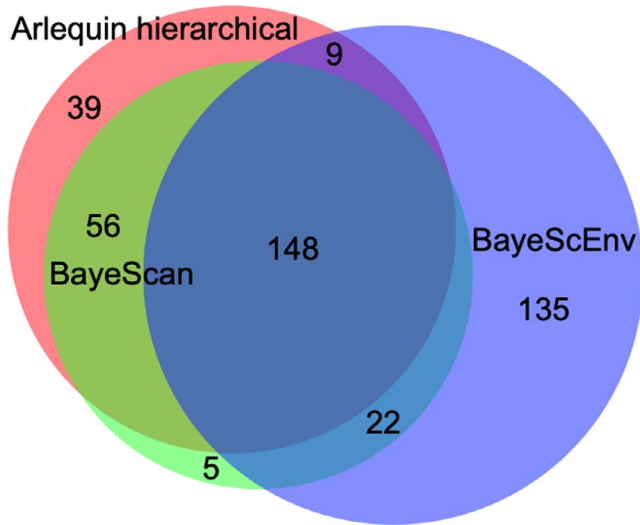


FIGURE 6 Venn diagram illustrating the overlap of outlier loci derived from three different methods. Arlequin and Bayescan identify loci with high F_{ST} values whereas BayescEnv identifies loci where allele frequencies are correlated with local salinity

(divergence with and without gene flow) suggests that the separation of the two populations started long before the opening of the Baltic Sea. The interpretation of this is that two lineages have colonized the North Sea–Baltic Sea from different refugia and thereafter formed a contact zone with genetic exchange. Currently, the contact zone between the populations overlaps with the steepest part of the salinity gradient in the Baltic Sea. This demographic history has led to a semi-permeable barrier to gene flow, inferred by strong support for a model with heterogeneous gene flow, which are classically associated with recent events of speciation (Wu, 2001). We found several markers within genes or closely linked to genes that were much more differentiated than most other loci over the contact zone, suggesting that genes underlying functional adaptation involved in, for example osmoregulation, vision and reproduction, are segregating over the salinity gradient. Differences in sperm motility that we found suggest local adaptation to the different salinities in this key reproductive trait, although only slight evidence of differential adaptation was evident in the male gonad transcriptome data.

A history of colonization of the Baltic Sea by gradual transition from the North Sea sand goby population was previously suggested based on the finding that many haplotypes were shared between the North Sea and Baltic Sea groups (Gysels, Hellemans, Patarnello, & Volckaert, 2004; Larmuseau, Raeymaekers, Ruddick, Van Houdt, & Volckaert, 2009; Larmuseau, Vancampenhout, Raeymaekers, Van Houdt, & Volckaert, 2010). Here, models with gradual transition from the North Sea provide good fit of the data only when they account for a strong bottleneck followed by rapid population growth. However, limited evidence of such strong demographic effect appeared in the data, as the diversity of the Baltic Sea populations is similar to those of the North Sea populations (Table S3). In addition, the divergence obtained under the IM (isolation with migration) scenario (>300 kya) is difficult to fit to the history of the Baltic Sea, whereas the divergence time provided by the secondary contact

fits much closer to the glacial history of Europe. Indeed, this model inferred that the populations have evolved at least a large part of the current genetic differences in allopatry and only recently experienced a second period of gene flow. This secondary contact phase roughly corresponds to the age of the glacier retreat after the Last Glacial Maximum in the northern hemisphere (~15 kya, Kettle, Morales-Muñiz, Roselló-Izquierdo, Heinrich, & Vøllestad, 2011). Interestingly, sand gobies also occur in other salinity gradients elsewhere in Europe, and so divergence observed here between marine and brackish environments might well predate the age of the opening of the Baltic Sea. Under this scenario, the sharing of haplotypes (Larmuseau et al., 2010) might be either ancestrally shared variation, or a consequence of recent gene flow over the contact zone (Gagnaire et al., 2015). The importance of secondary gene flow in the structure of the Baltic Sea is corroborated in our inferences by lower proportion of the genome resisting to gene flow in the population close to the transition zone. However, the presence of ancestral divergence does not reject the possibility of in situ genetic differentiation due to divergent selection over the salinity transition from standing variation, as divergence can be very efficient, in particular, if high numbers of genes are contained in large haplotype blocks or inversions (Faria, Johannesson, Butlin, & Westram, 2019). In addition, following a mixing of gene pools due to introgression between two ancestral lineages might lead to evolution of new and unique lineages (Marques, Meier, & Seehausen, 2019), as observed in Baltic blue mussels and Baltic clams (Nikula et al., 2007; Väinölä & Strelkov, 2011).

4.2 | Genetic differentiation involves multiple traits

The large number of outlier loci and the lack of functional enrichment suggest the involvement of several physiological processes in adaptation to the Baltic Sea. For example, among outliers, we found several genes related to ion transport located in close proximity (within 4 kb) to outlier SNPs. For instance, we identified: solute carrier family 5 member 3 (*slc5a3*), a well-known cotransporter that affects osmolyte levels in mammals (Kwon et al., 1992); acid-sensing ion channel 2 (*asic2*), which is involved in sodium ion transport (Price, Snyder, & Welsh, 1996); adenylate cyclase type 8 (*adcyl8*), which is involved in the regulation of cytosolic calcium ion concentration and water balance (Steel & Anderson, 2002). Additionally, we found SNPs near genes related to cellular water balance, such as aquaporin-4 (*aqp4*) and guanine nucleotide-binding protein subunit beta-4 (*GNB4*). The latter is a molecule that provides energy for vasopressin and thus affects membrane permeability allowing for water transport through aquaporin channels (Moeller, Fenton, Zeuthen, & Macaulay, 2009).

The myo-inositol synthesis pathway is linked to osmoregulation (Gardell et al., 2013; Kalujnaia, Hazon, & Cramb, 2016; Sacchi, Li, Villarreal, Gardell, & Kültz, 2013) and upregulated in response to increased salinity. Myo-inositol is a small molecule used to adjust cellular osmolality, and by increasing the concentration,

organisms can cope with hyperosmotic stress (Burg, Ferraris, & Dmitrieva, 2007). We found the inositol-3-phosphate synthase 1 (*isynd1*) gene located on one end of scaffold 315. The gene has three outlier SNPs, and one of the few differentially expressed transcripts in sand goby testis tissue is located immediately downstream of this gene on the antisense strand. Increasing evidence of antisense noncoding transcripts involved in transcriptional regulation (Pelechano & Steinmetz, 2013) suggests this transcript is involved in regulation of the *isynd1* gene. Additionally, in humans, the highest expression of this gene is found in testicular tissue (Guan, Dai, & Shechter, 2003), suggesting an important role in reproduction. Another gene involved in inositol metabolism is inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase (*ppip5k1b*) which is also adjacent to an outlier SNP in another scaffold. A third gene related to inositol is the previously mentioned *slc5a3b* gene, which is a sodium/myo-inositol cotransporter. Whether ancient or relatively recent, this broad spectrum of genetic variation in osmoregulatory genes seems to make the sand goby well prepared for establishment in a truly marginal marine habitat.

Genes linked to outliers were enriched for the Uniprot key word 'mitochondrion' and more than 10 genes linked to outlier SNPs are involved in mitochondrial functions (Table S5). There are hundreds of nuclear genes that interact with mitochondria (Burton & Barreto, 2012), but these interactions have been overlooked as a mechanism for population and species divergence until recently (e.g., Hill, 2017; Hill & Johnson, 2013; Wernick et al., 2019; Wolff, Ladoukakis, Enríquez, & Dowling, 2014). Oxidative phosphorylation (OXPHOS), transcription, translation and replication, all involve mitonuclear interactions (Burton & Barreto, 2012). To function properly, proteins from the mitochondria and the nucleus need to bind optimally requiring co-evolution of the binding regions. This can create a situation where populations have locally co-adapted mitonuclear genes, and divergence between locally co-adapted mitonuclear lineages can occur rapidly, especially if gene flow is disrupted (Hill, 2017).

We also found that many outlier loci were related to immune function (Table S5). These genes function in both adaptive and innate immunity as well as cytokine signalling in the immune system. Interestingly, most outliers in a genome scan of the plaice population inhabiting the Baltic Sea transition zone are linked to the immune system and are significantly enriched (Le Moan, Bekkevold, & Hemmer-Hansen, 2019). Since habitats vary dramatically in pathogen communities, it is not surprising that immune genes are important for local adaptation to the Baltic Sea (Eizaguirre, Lenz, Kalbe, & Milinski, 2012).

An additional outlier was the visual perception gene, phosphodiesterase 6H, cGMP-specific, cone, gamma, paralog a (*PDE6HA*) that likely has a role in visual adaptation in sand goby (Shimizu-Matsumoto et al., 1996). Sand gobies are sight predators active at low light levels (Larmuseau et al., 2009). Both day length and sea ice cover, which both correlate with the salinity gradient, affect the visual environment during the sand goby life cycle. Sea ice restricts the penetration of light and reduces turbidity from wind-induced mixing (Thomas et al., 2017).

4.3 | Reproductive incompatibilities

Limited gene flow due to selection against immigrants through inviability, infecundity or reproductive dysfunction is likely to facilitate genetic differentiation between populations inhabiting different habitats (Lemel, Belichon, Clobert, & Hochberg, 1997; Nosil, 2012; Nosil, Vines, & Funk, 2005; Svensson et al., 2017). We found clear signs of local adaptation to the native salinity in sperm motility, with a much larger proportion of sperm being motile in the salinity that matched the male's origin. These results are further supported by sperm velocity showing similar response (Svensson et al., 2017). We cannot completely rule out that acclimation may have affected our results, since sperm were sampled from fish that were kept in original salinity. However, other results support immigrant reproductive dysfunction in other traits between individuals of the same sand goby populations and salinities as tested here, after 4.5–8 months of acclimation to treatment salinities, (Svensson et al., 2017). Also, the round goby, *Neogobius melanostomus*, shows similarly strong effects of salinity on sperm motility and velocity, immediately and after 2 months of acclimation (Green, 2020). Since sperm are free-swimming single cells, all the osmoregulation genes discussed above would be relevant for sperm adaptation to salinity. Additionally, several genes involved in microtubule formation are present among the genes adjacent to outlier SNPs. Microtubules are an important component of sperm cell structure as well as motility. Rho-kinase 2 (*rock2*) was identified adjacent to an outlier SNP and has a role in regulating intracellular pH during the acrosome reaction in sea urchin (de la Sancha et al., 2007).

5 | CONCLUSION

Isolation prior to establishment over the North Sea–Baltic Sea salinity gradient, followed by secondary contact with gene flow and divergent selection can explain the observed genetic structure of North Sea and Baltic Sea populations of the sand goby. Whether all genetic divergence is ancient or some divergence is more recent, the current Baltic Sea lineage is well adapted to the low salinity in traits related to osmoregulation, immunity, vision and reproduction. Many loci also appear to be involved in these traits, but the specific functional mechanism (e.g., coding sequence, regulatory loci) remains to be clarified. We conclude that the first steps on the speciation continuum trajectory have been taken and show interesting parallels with other taxa inhabiting the North Sea–Baltic Sea transition.

ACKNOWLEDGMENTS

We thank SciLife Lab, Stockholm (Mattias Ormestad), Magnus Alm-Rosenblad and Tomas Larsson (for bioinformatic support), Sanna Pausio, Heidi Viitaniemi, Marina Panova, Ulrika Lind and Ricardo Pereyra (for help with DNA extraction, RNA extraction, and GBS library preparation), Eva-Lotta Blom, Johanna Gräns, Ulf Lindahl, Sami Merilaita, Andreas Svensson, Ingvar Svensson (for help with catching

gobies), Gry Sagebakken (for fish maintenance), Sofie Schöld (for sperm analysis and salinity data), Ellika Faust (for help with statistical analyses) and Malin Celander and Joanna Wilson for discussion and support. We acknowledge CSC – IT CENTER FOR SCIENCE LTD for providing computing resources to EL. We thank the Swedish Research Councils (VR and Formas) for funding to the Centre for Marine Evolutionary Biology (to EL, OS, TL) and the Swedish Research Council VR to LK (2016-03343& 621-2011-4004), and the Finnish Academy (to EL, grant nr 136464) and the ASSEMBLE program to KL and EL. FV acknowledges the Petra och Karl Erik Hedborgs Stiftelse (Sweden) and the Scientific Research Network 'Eco-evolutionary dynamics in natural and anthropogenic communities' (grant W0.037.10 N) of the Research Foundation Flanders (Belgium). This study was done within the IMAGO initiative within the Centre for Marine Evolutionary Biology at the University of Gothenburg (www.cemeb.science.gu.se), supported by a Linnaeus-grant (to KJ, AB, CA, CK, JH) from the Swedish Research Councils VR and Formas (grant nr 217-2008-1719).

DATA AVAILABILITY STATEMENT

The draft scaffolded genome assembly is available in Dryad: <https://doi.org/10.5061/dryad.rfj6q5780>.

ORCID

Erica H. Leder  <https://orcid.org/0000-0002-7160-2290>
 Carl André  <https://orcid.org/0000-0003-4404-7292>
 Mats Töpel  <https://orcid.org/0000-0001-7989-696X>
 Anders Blomberg  <https://orcid.org/0000-0002-1260-3920>
 Jonathan N. Havenhand  <https://orcid.org/0000-0001-6721-7763>
 Kai Lindström  <https://orcid.org/0000-0002-8356-5538>
 Filip A. M. Volckaert  <https://orcid.org/0000-0003-4342-4295>
 Charlotta Kvarnemo  <https://orcid.org/0000-0001-8983-2900>
 Kerstin Johannesson  <https://orcid.org/0000-0003-0176-7986>
 Ola Svensson  <https://orcid.org/0000-0003-3752-3131>

REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26, 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Andersen, J. H., Carstensen, J., Conley, D. J., Dromph, K., Fleming-Lehtinen, V., Gustafsson, B. G., ... Murray, C. (2017). Long-term temporal and spatial trends in eutrophication status of the Baltic Sea. *Biological Reviews*, 92, 135–149. <https://doi.org/10.1111/brv.12221>
- Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, 23, 38–44. <https://doi.org/10.1016/j.tree.2007.09.008>
- Barth J. M. I., Villegas-Ríos D., Freitas C., Moland E., Star B., André C., Knutsen H., Bradbury I., Dierking J., Petereit C., Righton D., Metcalfe J., Jakobsen K. S., Olsen E. M., Jentoft S. (2019). Disentangling structural genomic and behavioural barriers in a sea of connectivity. *Molecular Ecology*, 28(6), 1394–1411. <http://dx.doi.org/10.1111/mec.15010>.
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16, 113–148. <https://doi.org/10.1146/annurev.es.16.110185.000553>
- Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B*, 263, 1619–1626.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B - Statistical Methodology*, 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bernatchez, S., Laporte, M., Perrier, C., Sirios, P., & Bernatchez, L. (2016). Investigating genom and phenotypic parallelism between piscivorous and planktivorous lake trout (*Salvelinus namaycush*) ecotypes by means of RADseq and morphometrics analyses. *Molecular Ecology*, 25, 4773–4792.
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20, 2044–2072. <https://doi.org/10.1111/j.1365-294X.2011.05080.x>
- Björck, S. (1995). A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quaternary International*, 27, 19–40. [https://doi.org/10.1016/1040-6182\(94\)00057-C](https://doi.org/10.1016/1040-6182(94)00057-C)
- Breder, C. M., & Rosen, D. E. (1966). *Modes of reproduction in fishes*. Garden City, NY: The Natural History Press.
- Browne, R. K., Kaurova, S. A., Uteshev, V. K., Shishova, N. V., McGinnity, D., Figiel, C. R., ... Cosson, J. (2015). Sperm motility of externally fertilizing fish and amphibians. *Theriogenology*, 83, 1–13. <https://doi.org/10.1016/j.theriogenology.2014.09.018>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59. <https://doi.org/10.1038/nmeth.3176>
- Burg, M. B., Ferraris, J. D., & Dmitrieva, N. I. (2007). Cellular response to hyperosmotic stresses. *Physiological Reviews*, 87, 1441–1474. <https://doi.org/10.1152/physrev.00056.2006>
- Burton, R. S., & Barreto, F. S. (2012). A disproportionate role for mt DNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology*, 21, 4942–4957. <https://doi.org/10.1111/mec.12006>
- de la Sancha, C. U., Martínez-Cadena, G., López-Godínez, J., Castellano, L. E., Nishisaki, T., Darszon, A., & García-Soto, J. (2007). Rho-kinase (ROCK) in sea urchin sperm: Its role in regulating the intracellular pH during the acrosome reaction. *Biochemical and Biophysical Research Communications*, 364, 470–475. <https://doi.org/10.1016/j.bbrc.2007.10.050>
- de Villemereuil, P., & Gaggiotti, O. E. (2015). A new FST - based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, 6, 1248–1258.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Eizaguirre, C., Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters*, 15, 723–731. <https://doi.org/10.1111/j.1461-0248.2012.01791.x>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16, 157. <https://doi.org/10.1186/s13059-015-0721-2>
- Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103, 285–298. <https://doi.org/10.1038/hdy.2009.74>
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
- Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology and Evolution*, 34, 239–248. <https://doi.org/10.1016/j.tree.2018.12.005>

- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, *180*, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Forsgren, E. (1999). Sexual selection and sex roles in the sand goby. In V. C. Almada, R. F. Oliveira, & E. J. Gonçalves (Eds.), *Behaviour and conservation of littoral fishes* (pp. 249–274). Lisboa: ISPA.
- Forsgren, E., Kvarnemo, C., & Lindström, K. (1996). Modes of sexual selection determined by resource abundance in two sand goby populations. *Evolution*, *50*, 646–654.
- Gage, M. J. G., Macfarlane, C. P., Yeates, S., Ward, R. G., Searle, J. B., & Parker, G. A. (2004). Spermatozoal traits and sperm competition in Atlantic salmon: Relative sperm velocity is the primary determinant of fertilization success. *Current Biology*, *14*, 44–47. [https://doi.org/10.1016/S0960-9822\(03\)00939-4](https://doi.org/10.1016/S0960-9822(03)00939-4)
- Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., ... Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, *8*, 769–786. <https://doi.org/10.1111/eva.12288>
- Gagnaire, P. A., Pavey, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, *67*, 2483–2497. <https://doi.org/10.1111/evo.12075>
- García-Berro, A., Yliportimo, J., Lindström, K., & Kvarnemo, C. (2019). Understanding resource driven female-female competition: Ovary and liver size in sand gobies. *Royal Society Open Science*, *6*, 190886. <https://doi.org/10.1098/rsos.190886>
- Gardell, A. M., Yang, J., Sacchi, R., Fanguie, N. A., Hammock, B. D., & Kültz, D. (2013). Tilapia (*Oreochromis mossambicus*) brain cells respond to hyperosmotic challenge by inducing myo-inositol biosynthesis. *Journal of Experimental Biology*, *216*, 4615–4625. <https://doi.org/10.1242/jeb.088906>
- Gasparini, C., Simmons, L. W., Beveridge, M., & Evans, J. P. (2010). Sperm swimming velocity predicts competitive fertilization success in the green swordtail *Xiphophorus helleri*. *PLoS One*, *5*, e12146. <https://doi.org/10.1371/journal.pone.0012146>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, *29*, 644–652. <https://doi.org/10.1038/nbt.1883>
- Green, L. (2020). *Reproductive traits in euryhaline gobies – Insights into physiology, adaptations and biological invasions*. PhD thesis. University of Gothenburg, Gothenburg.
- Griffin, F. J., Pillai, M. C., Vines, C. A., Kääriä, J., Hibbard-Robbins, T., Yanagimachi, R., & Cherr, G. N. (1998). Effects of salinity on sperm motility, fertilization, and development in the Pacific herring, *Clupea pallasii*. *Biological Bulletin*, *194*, 25–35.
- Guan, G., Dai, P., & Shechter, I. (2003). cDNA cloning and gene expression analysis of human myo-inositol 1-phosphate synthase. *Archives of Biochemistry and Biophysics*, *417*(2), 251–259. [https://doi.org/10.1016/S0003-9861\(03\)00388-6](https://doi.org/10.1016/S0003-9861(03)00388-6)
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2010). Diffusion approximations for demographic inference: DaDi. *Nature Precedings*, 1–1. <https://doi.org/10.1038/npre.2010.4594.1>
- Gysels, E. S., Hellemans, B., Patarnello, T., & Volckaert, F. A. M. (2004). Current and historic gene flow of the sand goby *Pomatoschistus minutus* on the European Continental Shelf and in the Mediterranean Sea. *Biological Journal of the Linnean Society*, *83*, 561–576. <https://doi.org/10.1111/j.1095-8312.2004.00411.x>
- Havenhand, J. N., & Schlegel, P. (2009). Near-future levels of ocean acidification do not affect sperm motility and fertilization kinetics in the oyster *Crassostrea gigas*. *Biogeosciences*, *6*, 3009–3015.
- Hendry, A. P. (2016). *Eco-evolutionary dynamics*. Princeton, NJ: Princeton University Press.
- Hewitt, G. M. (2011). Quaternary phylogeography: The roots of hybrid zones. *Genetica*, *139*, 617–638. <https://doi.org/10.1007/s10709-011-9547-3>
- Hill, G. E. (2017). The mitonuclear compatibility species concept. *The Auk: Ornithological Advances*, *134*, 393–409. <https://doi.org/10.1642/AUK-16-201.1>
- Hill, G. E., & Johnson, J. D. (2013). The mitonuclear compatibility hypothesis of sexual selection. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1768), 20131314. <https://doi.org/10.1098/rspb.2013.1314>
- Huyse, T., Oeyen, M., Larmuseau, M. H. D., & Volckaert, F. A. M. (2017). Co-phylogeographic study of the flatworm *Gyrodactylus gondae* and its goby host *Pomatoschistus minutus*. *Parasitology International*, *66*, 119–125. <https://doi.org/10.1016/j.parint.2016.12.008>
- Järvi-Laturi, M., Lindström, K., Kvarnemo, C., & Svensson, O. (2011). Sand goby males trade off between defence against egg predators and sneak intrusions. *Journal of Zoology*, *283*(4), 269–275. <https://doi.org/10.1111/j.1469-7998.2011.00788.x>
- Johannesson, K., & André, C. (2006). Life on the margin – Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, *15*, 2013–2029.
- Jones, A. G., Walker, D., Kvarnemo, C., Lindström, K., & Avise, J. C. (2001). How cuckoldry can decrease the opportunity for sexual selection: Data and theory from a genetic parentage analysis of the sand goby, *Pomatoschistus minutus*. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 9151–9156. <https://doi.org/10.1073/pnas.171310198>
- Jones, A. G., Walker, D., Lindström, K., Kvarnemo, C., & Avise, J. C. (2001). Surprising similarity of sneaking rates and genetic mating patterns in two populations of sand goby experiencing disparate sexual selection regimes. *Molecular Ecology*, *10*, 461–469. <https://doi.org/10.1046/j.1365-294x.2001.01193.x>
- Kalujnaia, S., Hazon, N., & Cramb, G. (2016). Myo-inositol phosphate synthase expression in the European eel (*Anguilla anguilla*) and Nile tilapia (*Oreochromis niloticus*): Effect of seawater acclimation. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *311*, R287–R298.
- Kettle, A. J., Morales-Muñiz, A., Roselló-Izquierdo, E., Heinrich, D., & Vøllestad, L. A. (2011). Refugia of marine fish in the northeast Atlantic during the last glacial maximum: Concordant assessment from archaeozoology and palaeotemperature reconstructions. *Climate of the Past Discussions*, *6*, 1351–1389. <https://doi.org/10.5194/cpd-6-1351-2010>
- Kullander, S. O., Nyman, L., Jilg, K., & Delling, B. (2012). *Nationalnyckeln till Sveriges flora och fauna. Strålfeninga fiskar. Actinopterygii*. Uppsala, Sweden: Artdatabanken, SLU.
- Kwon, H. M., Yamauchi, A., Uchida, S., Preston, A. S., Garcia-Perez, A., Burg, M. B., & Handler, J. S. (1992). Cloning of the cDNA for a Na⁺/myo-inositol cotransporter, a hypertonicity stress protein. *Journal of Biological Chemistry*, *267*, 6297–6301.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. <https://doi.org/10.1038/nmeth.1923>
- Larmuseau, M. H. D., Raeymaekers, J. A. M., Ruddick, K. G., Van Houdt, J. K. J., & Volckaert, F. A. M. (2009). To see in different seas: Spatial variation in the rhodopsin gene of the sand goby (*Pomatoschistus minutus*). *Molecular Ecology*, *18*, 4227–4239. <https://doi.org/10.1111/j.1365-294X.2009.04331.x>
- Larmuseau, M. H. D., Vancampenhout, K., Raeymaekers, J. A. M., Van Houdt, J. K. J., & Volckaert, F. A. M. (2010). Differential modes of selection on the rhodopsin gene in coastal Baltic and North Sea populations of the sand goby, *Pomatoschistus minutus*. *Molecular Ecology*, *19*, 2256–2268. <https://doi.org/10.1111/j.1365-294X.2010.04643.x>
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*, R29. <https://doi.org/10.1186/gb-2014-15-2-r29>

- Le Moan, A., Bekkevold, D., & Hemmer-Hansen, J. (2019). Evolution at two time frames shape structural variants and population structure in European plaice (*Pleuronectes platessa*). *bioRxiv* 662577. <https://doi.org/10.1101/662577v5>
- Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D., & Hemmer-Hansen, J. (2019). Beyond parallel evolution: when several species colonize the same environmental gradient. *bioRxiv*, 662569. <https://doi.org/10.1101/662569>
- Le Moan, A., Jiménez-Mena, B., Bekkevold, D., & Hemmer-Hansen, J. (2019). Fine scale population structure linked to neutral divergence in the common sole (*Solea solea*), a marine fish with high dispersal capacity. *bioRxiv*, 662619. <https://doi.org/10.1101/662619>
- Lehtonen, T. K., & Kvarnemo, C. (2015a). Density effects on fish egg survival and infections depend on salinity. *Marine Ecology Progress Series*, 540, 183–191. <https://doi.org/10.3354/meps11517>
- Lehtonen, T. K., & Kvarnemo, C. (2015b). Infections may select for filial cannibalism by impacting egg survival in interactions with water salinity and egg density. *Oecologia*, 178, 673–683. <https://doi.org/10.1007/s00442-015-3246-1>
- Lehtonen, T. K., & Kvarnemo, C. (2015c). Odour cues from suitors' nests determine mating success in a fish. *Biology Letters*, 11, 20150021. <https://doi.org/10.1098/rsbl.2015.0021>
- Lehtonen, T., & Lindström, K. (2004). Changes in sexual selection resulting from novel habitat use in the sand goby. *Oikos*, 104, 327–335. <https://doi.org/10.1111/j.0030-1299.2004.12489.x>
- Lemel, J. Y., Belichon, S., Clobert, J., & Hochberg, M. E. (1997). The evolution of dispersal in a two-patch system: Some consequences of differences between migrants and residents. *Evolutionary Ecology*, 11, 613–629. <https://doi.org/10.1007/s10682-997-1516-z>
- Lescak, E. A., Bassham, S. L., Catchen, J., Gelmond, O., Sherbick, M. L., von Hippel, F. A., & Cresko, W. A. (2015). Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proceedings of the National Academy of Sciences of the United States of America*, 112, E7204–E7212. <https://doi.org/10.1073/pnas.1512020112>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Lissåker, M., & Kvarnemo, C. (2006). Ventilation or nest defence – Parental trade-offs in a fish with male care. *Behavioral Ecology and Sociobiology*, 60, 864–873.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., ... Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1, 18. <https://doi.org/10.1186/2047-217X-1-18>
- Luttikhuisen, P. C., Drent, J., Peijnenburg, K. T. C. A. J., van der Veer, H. W., & Johannesson, K. (2012). Genetic architecture in a marine hybrid zone: Comparing outlier detection and genomic clines analysis in the bivalve *Macoma balthica*. *Molecular Ecology*, 21, 3048–3061.
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). *pcadapt*: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.
- Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27, 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics*, 12, e1005887. <https://doi.org/10.1371/journal.pgen.1005887>
- Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view of speciation and adaptive radiation. *Trends in Ecology and Evolution*, 34, 531–544.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet.journal*, 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martinez-Barrio, A. M., Lamichhaney, S., Fan, G. Y., Rafati, N., Pettersson, M., Zhang, H., ... Andersson, L. (2016). The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *Elife*, 5, e12081.
- Miller, P. J. (1986). Gobiidae. In P. J. P. Whitehead (Ed.), *Fishes of the north-eastern Atlantic and the Mediterranean* (pp. 1019–1085). Paris, France: Unesco.
- Moeller, H. B., Fenton, R. A., Zeuthen, T., & Macaulay, N. (2009). Vasopressin-dependent short-term regulation of aquaporin 4 expressed in *Xenopus oocytes*. *Neuroscience*, 164, 1674–1684. <https://doi.org/10.1016/j.neuroscience.2009.09.072>
- Momigliano, P., Jokinen, H., Fraimout, A., Florin, A.-B., Norkko, A., & Merilä, J. (2017). Extraordinarily rapid speciation in a marine fish. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 6074–6079. <https://doi.org/10.1073/pnas.1615109114>
- Morisawa, M. (2008). Adaptation and strategy for fertilization in the sperm of teleost fish. *Journal of Applied Ichthyology*, 24, 362–370. <https://doi.org/10.1111/j.1439-0426.2008.01126.x>
- Nikula, R., Strelkov, P., & Väinölä, R. (2007). Diversity and trans-arctic invasion history of mitochondrial lineages in the North Atlantic *Macoma balthica* complex (Bivalvia: Tellinidae). *Evolution*, 61, 928–941. <https://doi.org/10.1111/j.1558-5646.2007.00066.x>
- Nosil, P. (2012). *Ecological speciation*. Oxford, UK: Oxford University Press.
- Nosil, P., Vines, T. H., & Funk, D. J. (2005). Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, 59, 705–719. <https://doi.org/10.1554/04-428>
- Ojaveer, H., Gross, R., Laur, K., Arula, T., & Klais, R. (2017). Genetic analysis reveals the diversity of larval Gobiidae in a temperate estuary. *Journal of Fish Biology*, 91, 1048–1061. <https://doi.org/10.1111/jfb.13407>
- Panova, M., Aronsson, H. R., Cameron, A., Dahl, P., Godhe, A., Lind, U., ... Johannesson, K. (2016). DNA extraction protocols for whole genome sequencing in marine organisms. In S. Bourlat (Ed.), *Marine genomics: Methods and protocols* (pp. 13–44). New York, NY: Springer.
- Parmanne, R., & Lindström, K. (2003). Annual variation in gobiid larval density in the northern Baltic Sea. *Journal of Fish Biology*, 62, 413–426. <https://doi.org/10.1046/j.1095-8649.2003.00037.x>
- Pelechano, V., & Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nature Reviews Genetics*, 14, 880–893. <https://doi.org/10.1038/nrg3594>
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- Portik, D. M., Leache, A. D., Rivera, D., Blackburn, D. C., Rodel, M.-O., Barej, M. F., ... Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian forest frog using demographic model selection. *Molecular Ecology*, 26, 5245–5263.
- Price, M. P., Snyder, P. M., & Welsh, M. J. (1996). Cloning and expression of a novel human brain Na channel. *Journal of Biological Chemistry*, 271, 7879–7882. <https://doi.org/10.1074/jbc.271.14.7879>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559–575. <https://doi.org/10.1086/519795>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape

- of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30, 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Research*, 43, e47. <https://doi.org/10.1093/nar/gkv007>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- Rougemont, Q., Gagnaire, P.-A., Perrier, C., Genthon, C., Besnard, A.-L., Launey, S., & Evanno, G. (2016). Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, 26, 142–162.
- 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 Biology and Evolution*, 9, 2057–2074. <https://doi.org/10.1093/gbe/evx150>
- Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145(4), 1219–1228.
- Sacchi, R., Li, J., Villarreal, F., Gardell, A. M., & Kültz, D. (2013). Salinity-induced regulation of the myo-inositol biosynthesis pathway in tilapia gill epithelium. *Journal of Experimental Biology*, 216, 4626–4638. <https://doi.org/10.1242/jeb.093823>
- Salvanes, A. G. V., & Nordeide, J. T. (1993). Dominating sublittoral fish species in a west Norwegian fjord and their trophic links to cod (*Gadus morhua* L.). *Sarsia*, 78, 221–234.
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323, 737–740. <https://doi.org/10.1126/science.1160006>
- Shimizu-Matsumoto, A., Itoh, K., Inazawa, J., Nishida, K., Matsumoto, Y., Kinoshita, S., ... Okubo, K. (1996). Isolation and chromosomal localization of the human cone cGMP phosphodiesterase γ cDNA (PDE6H). *Genomics*, 32, 121–124. <https://doi.org/10.1006/geno.1996.0085>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Singer, A., Kvarnemo, C., Lindström, K., & Svensson, O. (2006). Genetic mating patterns studied in pools with manipulated nest site availability in two populations of *Pomatoschistus minutus*. *Journal of Evolutionary Biology*, 19, 1641–1650. <https://doi.org/10.1111/j.1420-9101.2006.01114.x>
- Song, L., Florea, L., & Langmead, B. (2014). Lighter: Fast and memory-efficient sequencing error correction without counting. *Genome Biology*, 15, 509. <https://doi.org/10.1186/s13059-014-0509-9>
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344, 738–742. <https://doi.org/10.1126/science.1252136>
- Steel, H. C., & Anderson, R. (2002). Dissociation of the PAF-receptor from NADPH oxidase and adenylate cyclase in human neutrophils results in accelerated influx and delayed clearance of cytosolic calcium. *British Journal of Pharmacology*, 136, 81–89. <https://doi.org/10.1038/sj.bjp.0704685>
- Svensson, O., Gräns, J., Celander, M. C., Havenhand, J., Leder, E. H., Lindström, K., ... Kvarnemo, C. (2017). Immigrant reproductive dysfunction facilitates ecological speciation. *Evolution*, 71, 2510–2521. <https://doi.org/10.1111/evo.13323>
- Svensson, O., & Kvarnemo, C. (2007). Parasitic spawning in sand gobies: An experimental assessment of nest-opening size, sneaker male cues, paternity, and filial cannibalism. *Behavioural Ecology*, 18(2), 410–419.
- Szklarczyk D., Gable A.L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N.T., Morris J.H., Bork P., Jensen L.J., von Mering C. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47, (D1), D607–D613. <http://dx.doi.org/10.1093/nar/gky1131>
- Thomas, D. N., Kaartokallio, H., Tedesco, L., Majaneva, M., Piiparinen, J., Eronen-Rasimus, E., ... Granskog, M. A. (2017). Life associated with Baltic Sea ice. In P. Snoeijs-Leijonmalm, H. Schubert, & T. Radziejewska (Eds.), *Biological oceanographic of the Baltic Sea* (pp. 333–357). Dordrecht, The Netherlands: Springer.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., ... Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28, 511. <https://doi.org/10.1038/nbt.1621>
- Twyford, A. D., & Friedman, J. (2015). Adaptive divergence in the monkey flower *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution*, 69, 1476–1486.
- Väinölä, R., & Strelkov, P. (2011). *Mytilus trossulus* in Northern Europe. *Marine Biology*, 158, 817–833. <https://doi.org/10.1007/s00227-010-1609-z>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Wernick, R. I., Christy, S. F., Howe, D. K., Sullins, J. A., Ramirez, J. F., Sare, M., ... Estes, S. (2019). Sex and mitonuclear adaptation in experimental *Caenorhabditis elegans* populations. *Genetics*, 211, 1045–1058.
- Westram, A. M., Rafajlović, M., Chaube, P., Faria, R., Larsson, T., Panova, M., ... Butlin, R. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, 2, 297–309. <https://doi.org/10.1002/evl3.74>
- Whitehead, A., Clark, B. W., Reid, N. M., & Hahn, M. (2017). When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evolutionary Applications*, 10, 762–783.
- Wiederholm, A.-M. (1987). Distribution of *Pomatoschistus minutus* and *P. microps* (Gobiidae, Pisces) in the Bothnian Sea: Importance of salinity and temperature. *Memoranda Societatis pro Fauna et Flora Fennica*, 63, 56–62.
- Wilson, A. J., Gislason, D., Skúlason, S., Snorrason, S. S., Adams, C. E., Alexander, G., ... Ferguson, M. M. (2004). Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. *Molecular Ecology*, 13, 1129–1142. <https://doi.org/10.1111/j.1365-294X.2004.02149.x>
- Wilson-Leedy, J. G., & Ingermann, R. L. (2007). Development of a novel CASA system based on open source software for characterization of zebrafish sperm motility parameters. *Theriogenology*, 67, 661–672. <https://doi.org/10.1016/j.theriogenology.2006.10.003>
- Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A., & Dowling, D. K. (2014). Mitonuclear interactions: Evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1646), 20130443. <https://doi.org/10.1098/rstb.2013.0443>
- Wu, C.I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14, 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Wu, T. D., Reeder, J., Lawrence, M., Becker, G., & Brauer, M. J. (2016). GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. In E. Mathé, & S. Davis (Eds.),

- Statistical genomics. Methods in molecular biology* (Vol. 1418, pp. 283–334). New York, NY: Humana Press.
- Zander, C. D., & Kesting, V. (1998). Colonization and seasonality of goby (Gobiidae, Teleostei) parasites from the outhwestern Baltic Sea. *Parasitology Research*, 84, 459–466.
- Zander, C. D., & Reimer, L. W. (2002). Parasitism at the ecosystem level in the Baltic Sea. *Parasitology*, 124, S119–S135. <https://doi.org/10.1017/S0031182002001567>
- Zillén, L., Conley, D., Andrén, T., Andrén, E., & Björck, S. (2008). Past occurrences of hypoxia in the Baltic Sea and the role of climate variability, environmental change and human impact. *Earth-Science Reviews*, 91, 77–92. <https://doi.org/10.1016/j.earscirev.2008.10.001>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Leder EH, André C, Le Moan A, et al. Post-glacial establishment of locally adapted fish populations over a steep salinity gradient. *J Evol Biol.* 2020;00:1–19. <https://doi.org/10.1111/jeb.13668>