

# Biases in Demographic Modeling Affect Our Understanding of Recent Divergence

Paolo Momigliano <sup>\*,1</sup> Ann-Britt Florin <sup>2</sup> and Juha Merilä <sup>1,3</sup>

<sup>1</sup>Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland

<sup>2</sup>Department of Aquatic Resources, Institute of Coastal Research, Swedish University of Agricultural Sciences, Öregrund, Sweden

<sup>3</sup>Division of Ecology and Biodiversity, Faculty of Science, The University of Hong Kong, Hong Kong SAR

\*Corresponding author: E-mail: paolo.momigliano@gmail.com.

Associate editor: Yuseob Kim

## Abstract

Testing among competing demographic models of divergence has become an important component of evolutionary research in model and non-model organisms. However, the effect of unaccounted demographic events on model choice and parameter estimation remains largely unexplored. Using extensive simulations, we demonstrate that under realistic divergence scenarios, failure to account for population size ( $N_e$ ) changes in daughter and ancestral populations leads to strong biases in divergence time estimates as well as model choice. We illustrate these issues reconstructing the recent demographic history of North Sea and Baltic Sea turbot (*Scophthalmus maximus*) by testing 16 isolation with migration (IM) and 16 secondary contact (SC) scenarios, modeling changes in  $N_e$  as well as the effects of linked selection and barrier loci. Failure to account for changes in  $N_e$  resulted in selecting SC models with long periods of strict isolation and divergence times preceding the formation of the Baltic Sea. In contrast, models accounting for  $N_e$  changes suggest recent (<6 kya) divergence with constant gene flow. We further show how interpreting genomic landscapes of differentiation can help discerning among competing models. For example, in the turbot data, islands of differentiation show signatures of recent selective sweeps, rather than old divergence resisting secondary introgression. The results have broad implications for the study of population divergence by highlighting the potential effects of unmodeled changes in  $N_e$  on demographic inference. Tested models should aim at representing realistic divergence scenarios for the target taxa, and extreme caution should always be exercised when interpreting results of demographic modeling.

**Key words:** demographic modeling, allele frequency spectrum, secondary contact, isolation with migration.

## Introduction

Since Alfred Wallace noted that “Every species has come into existence coincident both in space and time with a pre-existing closely allied species” (Wallace 1855), understanding the processes by which new species arise (speciation) has been one of the major quests of evolutionary biology. In the case of sexual organisms, speciation can be defined as the evolution of reproductive isolation among populations, leading to distinct gene pools (Bolnick and Fitzpatrick 2007). The study of population pairs where reproductive isolation is incomplete can therefore provide insight into the processes leading to the evolution of different species. According to the genic view of speciation (Wu 2001), in the early stages of divergence reproductive isolation is a byproduct of differential adaptations and/or genetic incompatibilities, and therefore restricted to regions of the genome under exogenous and endogenous selection (often referred to as “barrier loci”). Such process results in heterogeneous differentiation across the genome, with barrier loci appearing as areas (“islands”) of higher differentiation due to selection, whereas unimpeded gene flow homogenizes the rest of the genome (Nosil 2012; Nosil and

Feder 2012; Roux et al. 2016; Ravinet et al. 2017). Partial reproductive isolation could arise via the gradual erosion of gene flow (primary divergence), for example because of multifarious selection across environments (Nosil 2008; Nosil et al. 2009; Nosil and Feder 2012). Alternatively, successive stages of strict isolation and secondary contact (secondary divergence) may facilitate the evolution of reproductive barriers among populations, whether they are due to ecological selection or to the evolution of genetic incompatibilities (Roux et al. 2013, 2014; Rougeux et al. 2017; Rougemont and Bernatchez 2018). Both processes can lead to similar genomic landscapes of differentiation, as following secondary contact gene flow can erode genetic differentiation across the genome (with the exception of barrier loci and regions around them) to the point at which any signature of the initial stage of strict isolation is lost (Ravinet et al. 2017). Therefore, distinguishing whether divergence initiated in the presence or absence of gene flow, while important to understand how reproductive isolation arises, is not a trivial task.

Recent events of primary divergence and secondary introgression among ancient lineages are however expected to

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

generate distinctive genomic landscapes surrounding barrier loci. Recent selection on rare or novel mutations is likely to temporally reduce genetic diversity ( $\pi$ ) surrounding barrier loci in the population experiencing selection (Smith and Haigh 1974), revealing signatures typical of selective sweeps (Tavares et al. 2018). Absolute divergence among populations ( $d_{xy}$ ) will initially remain low, as in the early stages of divergence  $d_{xy}$  in regions surrounding a barrier locus is expected to reflect ancestral genetic diversity (see discussion in Ravinet et al. 2017). Instead, if barrier loci are of ancient origin, increased genetic differentiation ( $F_{ST}$ ) around barrier loci is likely to be driven by an increase in  $d_{xy}$  rather than a decrease in  $\pi$  (Cruickshank and Hahn 2014). Indeed, there are several examples where islands of divergence that originated during long allopatric phases show both elevated  $F_{ST}$  and elevated  $d_{xy}$  with respect to the rest of genome, where the original signatures of divergence have been eroded by unimpeded gene flow (Duranton et al. 2018, 2020; Gagnaire et al. 2018; Nelson and Cresko 2018). Unfortunately, the interpretation of genomic landscapes of differentiation is not always strait-forward (reviewed in Ravinet et al. [2017]), and it requires data (highly contiguous genome assembly) that are still lacking for most non-model organisms.

Demographic modeling provides a framework to reconstruct how gene flow has changed through the evolutionary history of diverging populations. Within the past two decades, several computational methods have been developed to reconstruct demographic history from genomic data. Such approaches usually rely on comparing summary statistics obtained from empirical data to simulations performed under competing divergence scenarios, of which the most commonly tested ones include isolation with continuous migration (IM), secondary contact (SC), strict isolation (SI), and ancient migration (AM) (Roux et al. 2013). Approximate Bayesian computation (ABC) approaches (Beaumont et al. 2002; Excoffier et al. 2013), as well as methods based on the diffusion approximation of the joint site frequency spectrum (jAFS) (*dadi*, Gutenkunst et al. 2009) or its direct computation using a model of ordinary differential equations (*moments*, Jouganous et al. 2017), have been broadly applied to test among competing demographic models of divergence. The models usually assume that an ancestral population of size  $N_{ANC}$  gives rise to two populations of size  $N_1$  and  $N_2$  respectively at a time of split  $T_S$ , after which several migration scenarios are contrasted. ABC, *dadi* and *moments* allow users to define complex demographic scenarios, explicitly modeling the effect of barrier loci—modeled as heterogeneous migration rates across the genome (Roux et al. 2013; Tine et al. 2014)—as well as the effect of linked selection—modeled as heterogeneous effective population size ( $N_e$ ) across the genome (Roux et al. 2016; Rougemont et al. 2017). Failing to account for heterogeneity in linked selection and migration rates may lead to biases in model choice and parameter estimation (Roux et al. 2014, 2016; Ewing and Jensen 2016; Pouyet et al. 2018).

Such models have been used to test among competing gene flow scenarios across a broad range of divergence times, from a few thousand to millions of generations. For example,

demographic modeling has been extensively used to test whether sympatric and parapatric lineages within environments that were shaped during the last glacial cycle arose via rapid ecologically driven divergence or are the result of postglacial secondary contact between more ancient lineages (Tine et al. 2014; Le Moan et al. 2016, 2019; Rougeux et al. 2017, 2019; Van Belleghem et al. 2018; Jacobs et al. 2020). The same approach has been extensively used to infer demographic models of divergence among incipient species that diverged hundreds of thousands to millions of generations ago (Roux et al. 2013, 2014, 2016; Stuglik and Babik 2016; Bourgeois et al. 2019). Most of these studies concluded that contemporary heterogeneous gene flow is a result of recent secondary contact (Roux et al. 2013, 2014, 2016; Tine et al. 2014; Le Moan et al. 2016; Rougemont et al. 2017; Gagnaire et al. 2018; Rougemont and Bernatchez 2018; Rougeux et al. 2019) providing support to the hypothesis that initial allopatric phases of differentiation play a central role in the evolution of reproductive isolation (Roux et al. 2014), and hence primary divergence with gene flow due to ecological selection is rarer than suggested by some authors (Nosil 2008). However, although these models can provide important insight into demographic history, they also show significant limitations. If the models tested are not close enough to the real divergence scenario, both model choice (e.g., the choice between an SC and IM model) and parameter estimation may be affected. Although great effort has been placed recently to overcome potential biases due to barrier loci and linked selection (Roux et al. 2013, 2016; Bhaskar and Song 2014; Tine et al. 2014; Le Moan et al. 2016; Rougemont et al. 2017; Gagnaire et al. 2018; Rougemont and Bernatchez 2018), it remains unclear how unmodeled demographic events, such as size changes in both ancestral and daughter populations, may affect model choice and parameter estimation.

Most recent studies of non-model organisms, where prior knowledge of past demographic events is limited, assume that divergence starts from an ancestral population at mutation–drift equilibrium, with an instantaneous split into two populations of constant size (as in Roux et al. [2013]). If, for example, there was an unmodeled size change in the ancestral population (such as a population expansion or contraction), we can expect an overestimation of divergence time, since the models allows changes in  $N_e$  only at time  $T_S$ , pushing estimates of  $T_S$  toward the time of ancestral population size change. Similarly, unmodeled bottlenecks followed by exponential growth in one of the daughter populations may bias estimates of  $T_S$ , as small populations experience faster genetic drift (potentially leading to overestimate recent divergence). Both bottlenecks (Luikart et al. 1998) and SC (Alcala et al. 2015) are expected to generate an excess of middle-frequency variants, and hence an unmodeled bottleneck could bias model choice toward SC. It is less clear how changes in  $N_e$  in ancestral populations affect model choice, as to the best of our knowledge no one has addressed this question. This is a matter of concern, as few studies explicitly model growth in daughter populations, and very few studies of non-model organisms have tested for changes in ancestral population size. A brief search in Web of Science for published studies

using demographic modeling in non-model organisms (using combinations of keywords “Isolation with Migration,” “Secondary Contact,” “dadi,” “abc,” “fastsimcoal”) suggests that less than one fifth of papers published between 2016 and 2020 accounted for changes in  $N_e$  in the ancestral population or in at least one of the daughter populations.

Here, we used both simulations and empirical data to demonstrate that unmodeled demographic events in both ancestral and daughter populations can strongly affect both model choice and parameter estimation. Using coalescent simulations of IM and SC scenarios under the Wright–Fisher neutral model, we demonstrate that failure to account for changes in  $N_e$  in ancestral and daughter populations leads to extreme biases in estimates of  $T_S$  and to a strong bias toward the choice of SC models. We then reconstruct the demographic history of the Atlantic and Baltic Sea populations of the turbot *Scophthalmus maximus*, which a recent study (Le Moan et al. 2019) suggested have diverged before the last glacial maximum (>50 kya) and experienced secondary contact following the end of the last glaciation. We argue that these inferences were likely biased because of the failure to account for a past demographic expansion (which led to overestimate  $T_S$ ) and to model a bottleneck during the invasion of the Baltic Sea (leading to the erroneous choice of a SC model). Furthermore, we show that genomic patterns of differentiation are also consistent with a scenario of very recent divergence with gene flow. We discuss the potential implications of our findings for inferring the demographic history of non-model organisms in general.

## Results

### Analyses of Simulated Data

For all simulated data (fig. 1A and B), we optimized parameters for the simple IM and SC models and tested, using Akaike weights ( $W_{AIC}$  see Materials and Methods) the support for the correct gene flow scenario (IM). For the recent divergence scenarios, we tested (and optimized parameters for) all basic demographic models (fig. 1C), and used  $W_{AIC}$  to test the support for the correct gene flow scenario (SC or IM depending on simulation). Details of all models used for demographic inference are given in supplementary table 1, Supplementary Material online.

### Biases in IM and SC Models in Older Divergence Scenarios

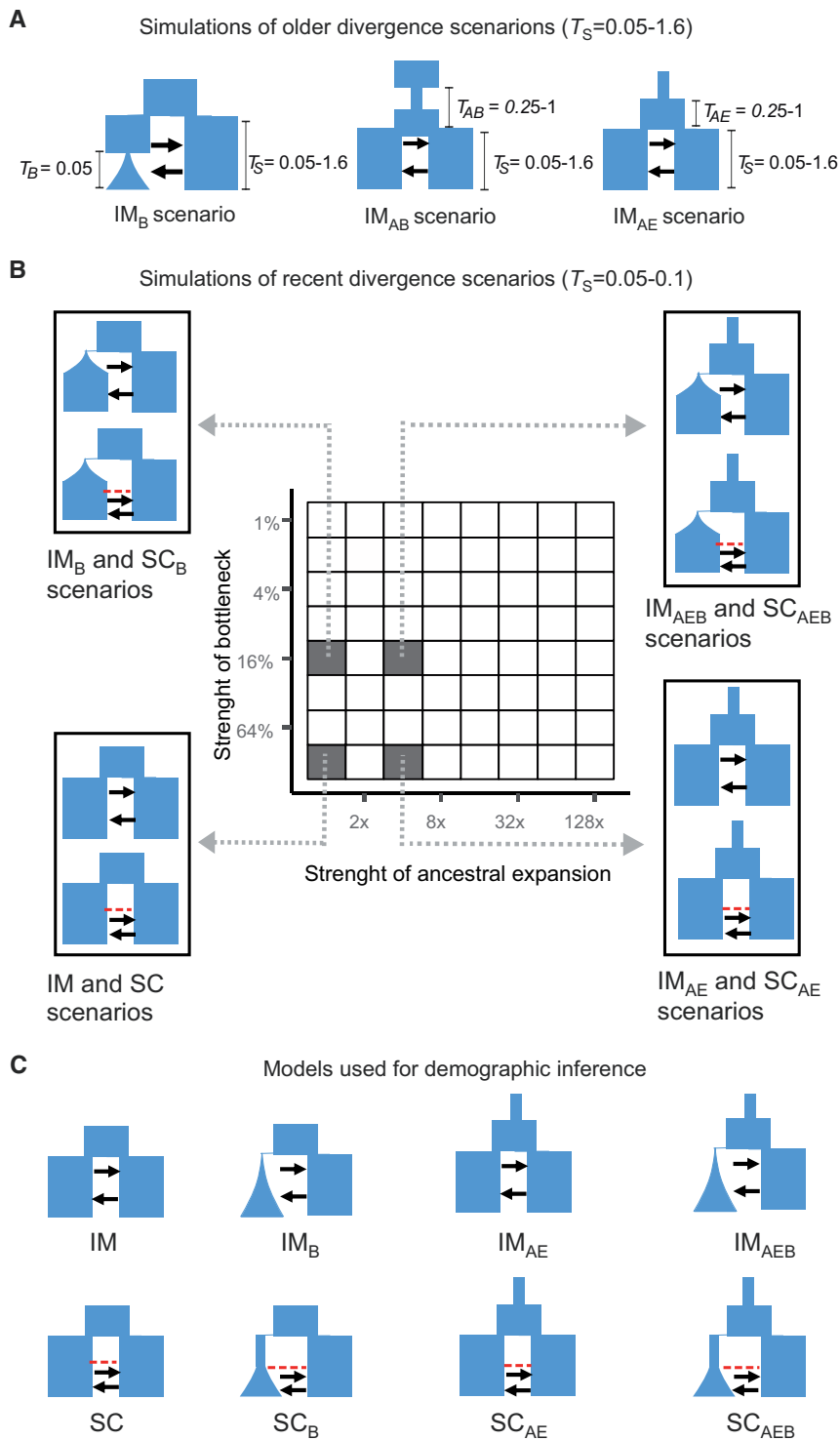
An unmodeled recent bottleneck always led to severe biases in model choice (favoring the SC model, as shown by low  $W_{AIC}$ ) and estimates of  $T_S$ , even when the simulated bottleneck was mild (fig. 2C). Estimates of  $T_S$  tended to reflect the strength of the bottleneck rather than divergence time (fig. 2B), suggesting the effects of a recent bottleneck on the joint allele frequency spectrum (jAFS) obscure signatures of divergence. This led to a severe overestimation of  $T_S$  in recent divergence scenarios (simulation  $T_S = 0.05$ ) and a severe underestimation of  $T_S$  in older divergence scenarios (simulation  $T_S > 0.4$ ). The estimated time of strict isolation as a proportion of divergence time ( $T_{SI}/T_S$ ) is a positive function of the strength of the unmodeled bottleneck (fig. 2D).

Non-equilibrium states in the ancestral population also led to biases in both model choice and parameter estimation (fig. 3), but the severity of these biases was affected by both  $T_S$  and the time before  $T_S$  at which the last ancestral change in  $N_e$  occurred ( $T_{AE}$  for AE models and  $T_{AB}/2$  for AB models). When  $T_S$  was relatively shallow ( $T_S < 0.4$ ) unaccounted changes in  $N_e$  in the ancestral population led to overestimate  $T_S$ . This bias was most severe when the last ancestral change in  $N_e$  ( $T_{AE}$  in AE model, and  $T_{AB}/2$  in AB models) occurred  $0.5 \times 4N_{REF}$  generations before  $T_S$ . Biases toward the choice of SC models were most severe at intermediate  $T_S$ , whereas the effects of not accounting for non-equilibrium in the ancestral population become irrelevant when  $T_S > 0.8$  (fig. 3).

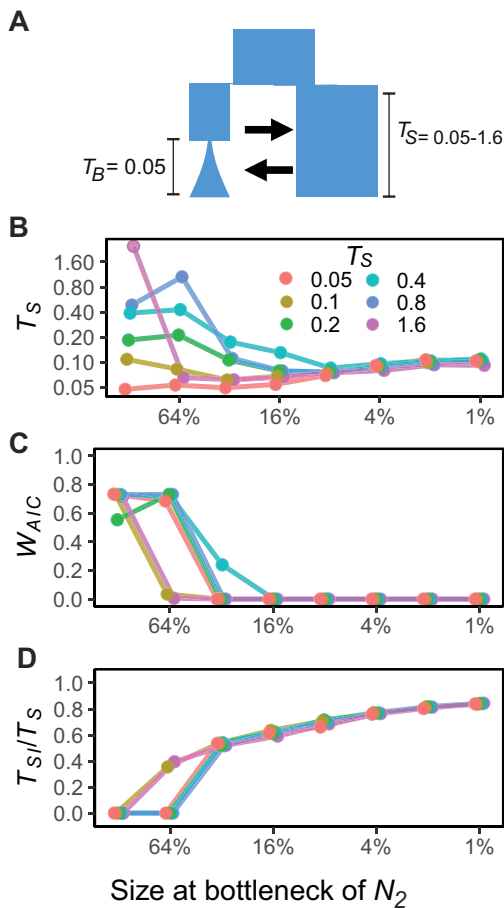
### Biases in IM and SC Models in Recent Divergence Scenarios

For the simulations of recent divergence scenarios, we conducted analyses based on simulations of 1 million or 100,000 loci. These simulations gave similar results (compare fig. 4 and supplementary fig. 2, Supplementary Material online with supplementary figs. 16 and 17, Supplementary Material online, respectively), with the exception that there was more stochastic variation when analyzing the smaller data sets. Hence, we report results based on the larger data sets in the main body of the manuscript, whereas the results for the smaller numbers of loci are presented in the Supplementary Material online. Models generally converged very well with the optimization scheme utilized (see examples in supplementary figs. 3–10, Supplementary Material online).

Parameter estimation and model choice from basic IM and SC models were severely biased when the demographic scenario of the simulations deviated from the models (fig. 4). Although for SC simulations the underlying gene flow scenario was almost always identified (fig. 4B), when the true divergence scenario of the simulations had continuous gene flow even moderate unmodeled bottlenecks resulted in strong support for SC (fig. 4A). Stronger bottlenecks were associated with an overestimation of divergence time, and the bias was much more severe for simulations with constant gene flow (IM; fig. 4C and D; supplementary fig. 11A and B, Supplementary Material online). There was also a clear relationship between the length of the estimated period of strict isolation ( $T_{SI}$ ) in IM simulations and the strength of the unmodeled bottleneck (fig. 4E). Hence, in recent demographic scenarios unmodeled bottlenecks generally led to choice of SC models with long periods of strict isolation even when the simulation scenario had constant gene flow. Unmodeled size changes in the ancestral population led to severe overestimates of divergence time in IM scenarios (fig. 4C) and to a lesser degree in SC scenarios (fig. 4D). The stronger the unmodeled ancestral expansion, the more estimates of  $T_S$  approached  $T_{AE}$  (which is  $10 \times T_S$  under IM scenarios and  $5 \times T_S$  under SC scenarios, see Materials and Methods; fig. 4C and D; supplementary fig. 11A and B, Supplementary Material online). Unmodeled ancestral expansions also biased model choice toward SC when the true scenario had constant migration, but only when the unfolded jAFS was used for demographic inference.



**FIG. 1.** Simulation scenarios and models used for demographic inference. Times are given in units of  $4N_{REF}$  generations. (A) We simulated data under an IM scenario with divergence times from 0.05 to 1.6. In IM<sub>B</sub> scenarios one population experiences a contraction to 1–64% of its previous size at time  $T_B$  followed by exponential growth. The IM<sub>AB</sub> scenario reflects an ancestral population size contraction followed by expansion. In the IM<sub>AE</sub> scenario, the ancestral population undergoes an expansion. For both AB and AE scenarios, we tested a range of times (0.25, 0.5, 1) and strengths of the ancestral contraction ( $AB = 1/4, 1/16, 1/64$  of previous size) and expansion ( $AE = 4\times, 16\times, 64\times$  the ancestral  $N_e$ ). (B) Simulations of recent divergence scenarios, with a  $T_S$  of 0.05 for IM scenarios and 0.1 for SC scenarios. Data were simulated under the IM and SC scenarios with 64 possible combinations of ancestral expansions and bottlenecks. Simulations were ran both with symmetric and asymmetric migration, for a total of 256 demographic scenarios. (C) demographic models used for inference with *moments* and *dadi*, that is, the basic IM and SC models as well as modifications that include an ancestral population size change as well as a bottleneck followed by growth in one of the daughter populations. Migration rate is asymmetric (i.e., different  $m_{12}$  and  $m_{21}$  parameters) in all inference models. For the analyses of the empirical data, we used the models graphically represented in panel C, and modifications of these models accounting for heterogeneity in migration rates (2M) and  $N_e$  (2N models) across the genome, as well as modifications accounting for both (2N2M models).



**FIG. 2.** Effects of unmodeled bottlenecks in a daughter population on model choice and parameter estimates. Model choice and parameter estimates for the simple IM and SC models for simulations with  $T_S = 0.05 - 1.6$  and recent bottlenecks of different strengths at time  $T_B = 0.05$ . Panel (A) shows the simulation model, panel (B) shows estimated time of divergence  $T_S$ , panel (C) shows Akaike weights ( $W_{AIC}$ ) for the correct model (IM) and panel (D) shows the inferred time of strict isolation ( $T_{SI}$ ) as a proportion of total divergence time.

However, when an unmodeled ancestral expansion led to incorrectly choose the SC model, the proportion of time of strict isolation inferred was low ( $< 50\%$  of the total divergence time) unless the simulation included also a strong bottleneck (supplementary fig. 12A and B, Supplementary Material online). Interestingly, in SC simulations without strong bottlenecks, an unmodeled ancestral expansion led in some cases to a bias toward IM models (fig. 4B), and in general to an underestimate of the proportion of divergence time for which the model inferred strict isolation (supplementary fig. 12A and B, Supplementary Material online). Not surprisingly, unmodeled ancestral expansions led also to strong overestimates of the ancestral population size  $N_{ANC}$  (supplementary fig. 13, Supplementary Material online).

All the above-reported biases were more severe when the unfolded jAFS was used for demographic inference. Particularly,  $T_S$  misestimation was roughly twice as high when inference was carried out using the unfolded jAFS compared with the folded jAFS (fig. 4C and D; supplementary fig. 11A and B, Supplementary Material online). Similarly, bias in

model choice toward SC models was stronger when using the unfolded jAFS (fig. 4A; supplementary fig. 11, Supplementary Material online). It should be noted, however, that estimates of contemporary  $N_e$  and migration rates were always fairly accurate (see supplementary figs. 14 and 15, Supplementary Material online). Furthermore, in general, these biases were less severe when the smaller data set was used for inference (supplementary fig. 16, Supplementary Material online). These biases were not unique to the simulation engine we used, as the analyses of a subset of the data using *dadi* gave nearly identical results (supplementary fig. 18, Supplementary Material online). Gross model mis-specifications were, however, identifiable by inspecting the jAFS residuals produced by *dadi* and *moments* (supplementary fig. 19, Supplementary Material online).

### Biases in 8-Model Comparisons in Recent Divergence Scenarios

Testing demographic scenarios that more closely approximate the real demographic history of simulated populations resulted in much less severe biases in model choice. Weight of evidence for the correct gene flow scenario was much stronger, with all simulations of SC scenarios being correctly identified and IM simulations being sometimes misidentified as SC only when population two experienced very severe bottlenecks followed by exponential growth (supplementary fig. 2A, Supplementary Material online). However, when SC scenarios were incorrectly chosen, the length of the inferred periods of strict isolation was usually very short ( $< \frac{1}{3}$  of total divergence), suggesting that IM and SC models were converging toward the same demographic history (supplementary figs. 2C and 12C and D, Supplementary Material online). Although ancestral expansions and bottlenecks still caused a systematic bias toward overestimating divergence times, this bias was negligible (supplementary figs. 2B and 11C and D, Supplementary Material online).

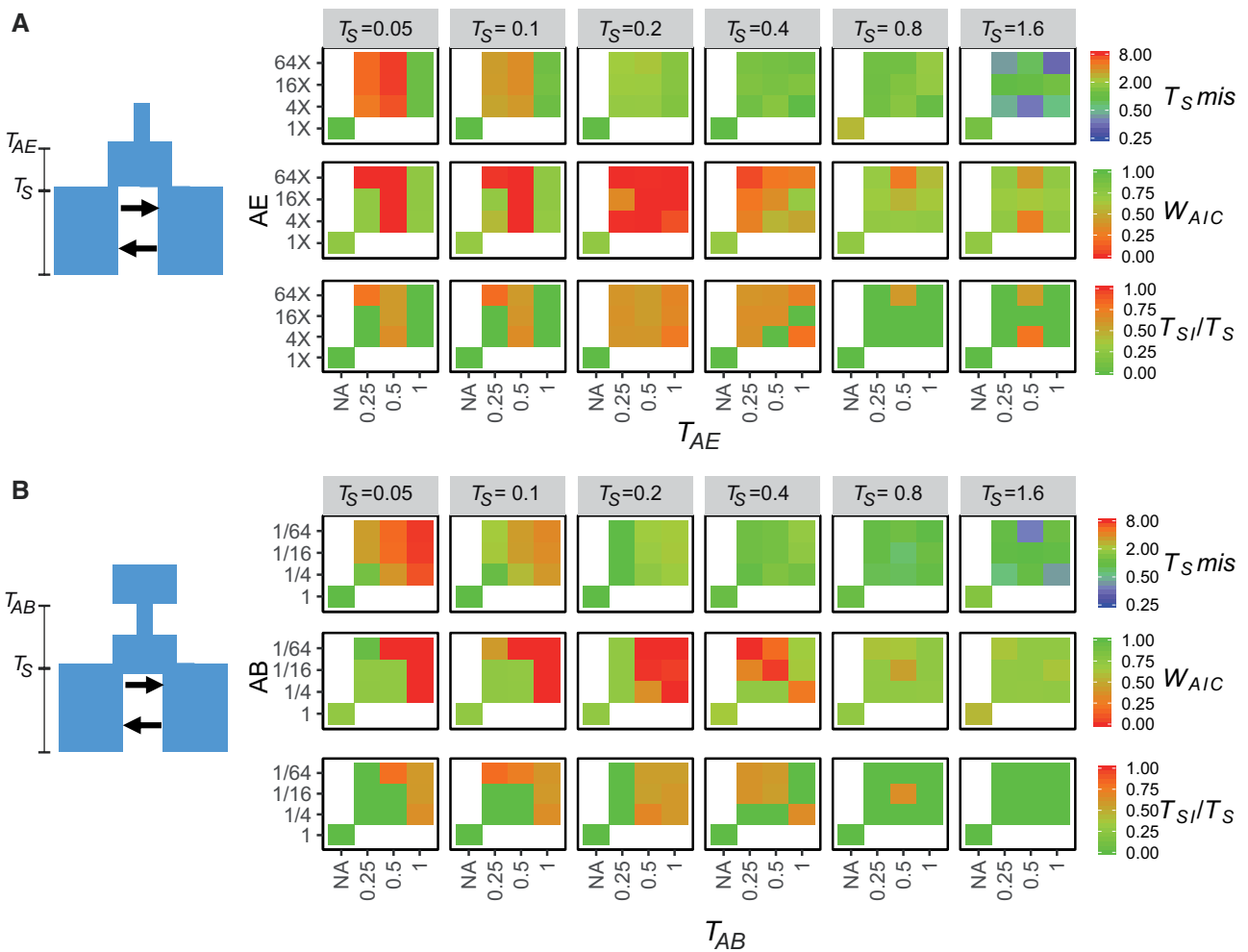
### Analyses of Empirical Data

#### Population Genetics Analyses

A PCA based on genotype likelihoods from 9,063 biallelic SNPs with MAF  $> 0.02$  shows a clear partitioning along the first PC between the North Sea and Baltic Sea individuals (fig. 5A and C), with several individuals from the transition zone (locations VD and ÖS in fig. 5C) showing intermediate genotypes. FastSTRUCTURE analyses (fig. 5B) show very similar results, with two clear genetic clusters (North Sea and Baltic Sea) and individuals from VD and ÖS (the transition Zone) exhibiting high proportion of admixture. Based on these results, individuals from admixed populations (VD and ÖS) were excluded from further analyses, with the exception of genome scans for selection using PCAngsd (since this analysis does not assume discrete populations).

#### Genomic Landscape of Differentiation

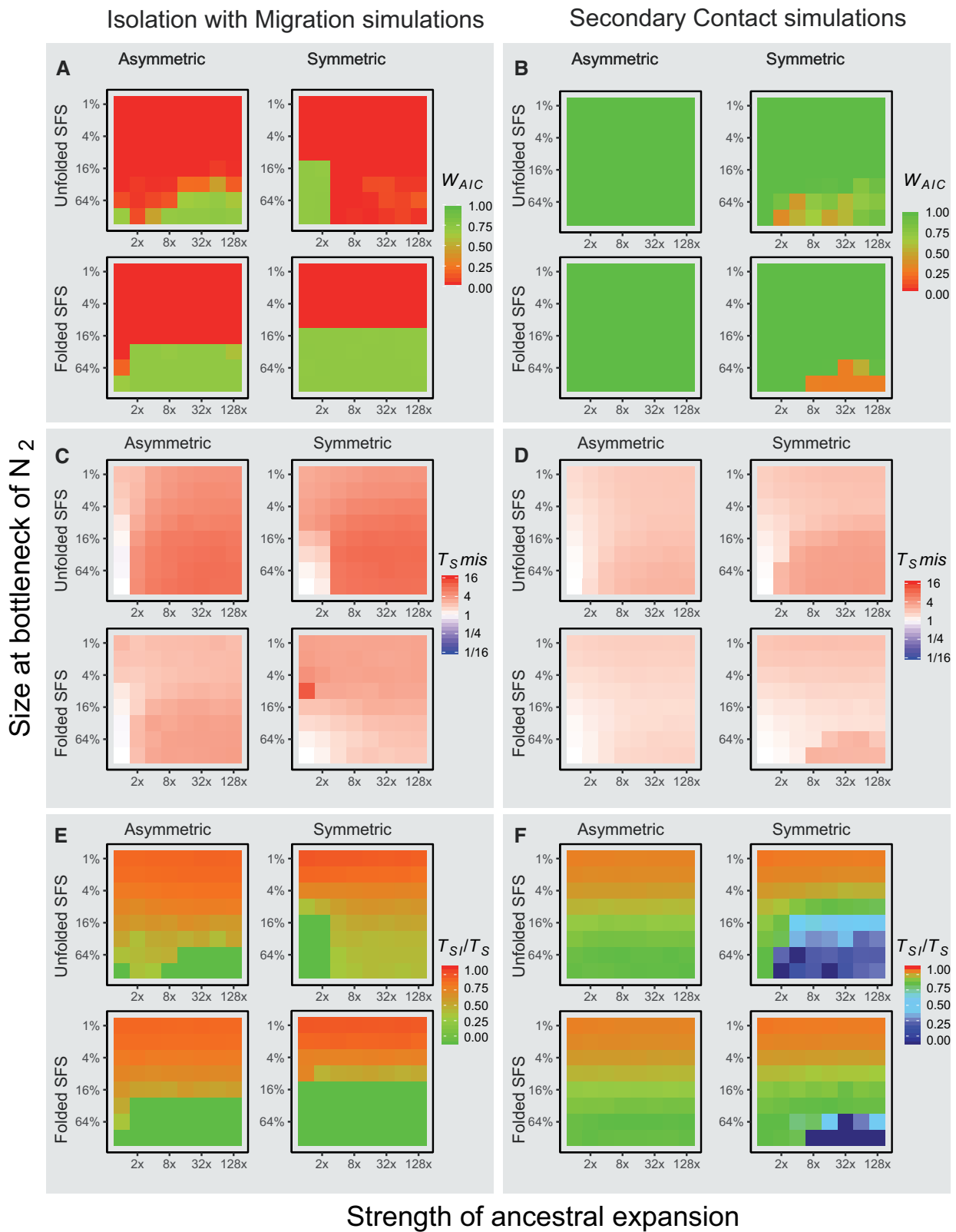
Average differentiation between the North Sea and Baltic Sea populations was weak (mean  $F_{ST} = 0.017$ ), however several SNPs across the genome showed marked  $F_{ST}$ , notably



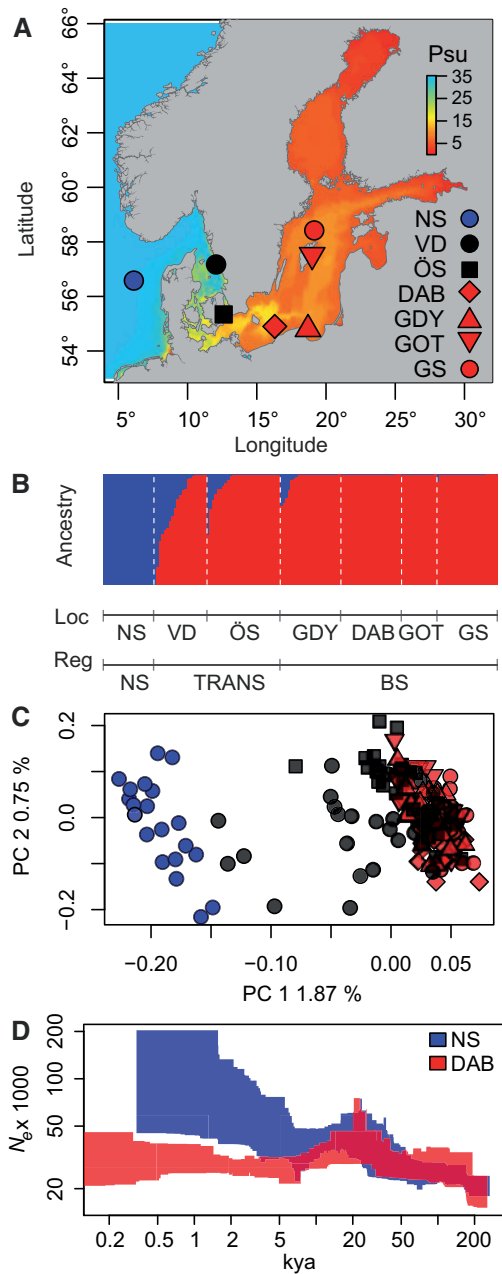
**FIG. 3.** Effect of unmodeled demographic changes in the ancestral population on model choice and parameter estimates. In this figure, we report biases in model choice and parameter estimation of simple IM and SC models for IM simulations with  $T_S = 0.05 - 1.6$  including an ancestral demographic expansion (A) and an ancestral bottleneck followed by expansion (B). On the left of panels (A) and (B) is a graphical representation of the simulation model. In panel (A), the y axis represents the ancestral population size after the expansion as a multiplier of its size preceding the expansion, and the x axis represents the time before  $T_S$  at which the ancestral expansion happened ( $T_{AE}$ ). In panel (B), the y axis represents the size of the ancestral population after contraction as fraction of the population size preceding contraction and the x axis represents the time before  $T_S$  at which the ancestral bottleneck happened ( $T_{AB}$ ). Times are given in units of  $4N_{REF}$  generations. For (A) and (B) and for each simulated  $T_S$ , we present the misestimation of  $T_S$  ( $T_{Smis} = T_{Sestimated}/T_{Ssimulated}$ ), the weight of evidence ( $W_{AIC}$ ) for the correct (IM) gene flow scenario as well as the estimated time of strict isolation as a proportion of total divergence time ( $T_{Si}/T_S$ ). In each graph, the bottom-left square represents estimates for simulations with no ancestral changes in  $N_e$ .

SNPs located around the center of chromosome 1 and in chromosome 13 (fig. 6A). Genome scans, performed using the extended model of fastPCA (Galinsky et al. 2016) implemented by PCAngsd and a Hidden Markov model (HMM) approach to detect genomic islands (Hofer et al. 2012; Marques et al. 2016), identified 32 outlier loci with a false discovery rate  $< 0.1$ , located in 15 distinct genomic regions (fig. 6B). Most of these 15 regions included one or two outlier loci, and were located at the very end of chromosome arms (supplementary fig. 20, Supplementary Material online), where the chance of false positives is highest. However, two genomic islands, located in chromosomes 1 and 13, had several outlier SNPs (10 and 13, respectively) spanning a distance of over 1 Mb and showing extreme levels of differentiation (fig. 6A–D).

Levels of genetic diversity ( $\pi$ ) and absolute divergence ( $d_{xy}$ ) were highly correlated across the genome (fig. 6C and D; supplementary fig. 20, Supplementary Material online).  $d_{xy}$  was highly correlated with  $\pi$  in the North Sea ( $R^2 = 0.85$ ,  $P < 1e^{-16}$ ), which is the expectation since in early stages of divergence  $d_{xy}$  is expected to approximate  $\pi$  in the ancestral population, which is likely still represented by the North Sea. The correlation between  $\pi$  in the North Sea and Baltic Sea populations breaks down within the two genomic islands of differentiation in chromosomes 1 and 13, where increased allelic differentiation ( $F_{ST}$ ) is driven by a dramatic decrease in genetic diversity in the Baltic Sea, rather than an increase in absolute divergence (fig. 6C and D). Such genomic landscapes are classic signatures of recent selective sweeps acting on novel or rare variants, resulting in a transient loss of genetic diversity surrounding the site of selection.



**FIG. 4.** Model choice and parameter misestimation for the simple IM and SC models for all simulations of the larger data sets (1 million loci). Left panels (**A**, **C**, **E**) show results from simulations with constant migration (IM), right panels (**B**, **D**, **F**) show results from simulation with a period of strict isolation (SC). Within each panel, results are shown for simulations with symmetric and asymmetric migration, and for inference using the folded or unfolded jAFS. Within each panel, each graph represents the values for all 64 simulations as per Fig. 1B. Panel **A** and **B** show weight of evidence for the correct gene flow scenario (0-1). Panel **C** and **D** show misestimation of the parameter  $T_S$  ( $T_S$  estimate /  $T_S$  of simulation). Panels **E** and **F** show the estimated proportion of the divergence time for which the model inferred strict isolation (green represents the correct time, i.e. 0 for IM model and 0.75 for SC models).



**Fig. 5.** Sampling sites, genetic structure, and historical change in  $N_e$  of *Scophthalmus maximus* populations in the North Sea (blue) the transition zone (black) and Baltic Sea (red). (A) Sampling locations showing modeled bottom salinity of the Baltic Sea. (B) Individual ancestry reconstructed from fastSTRUCTURE. (C) PCA performed from genotype likelihoods of 9,063 biallelic SNPs, color, and population codes as per (A). (D) Changes in  $N_e$  across time inferred for a representative population from the North Sea (NS) and the Baltic Sea (DAB). Polygons represent 95% confidence intervals.

### Demographic Modeling

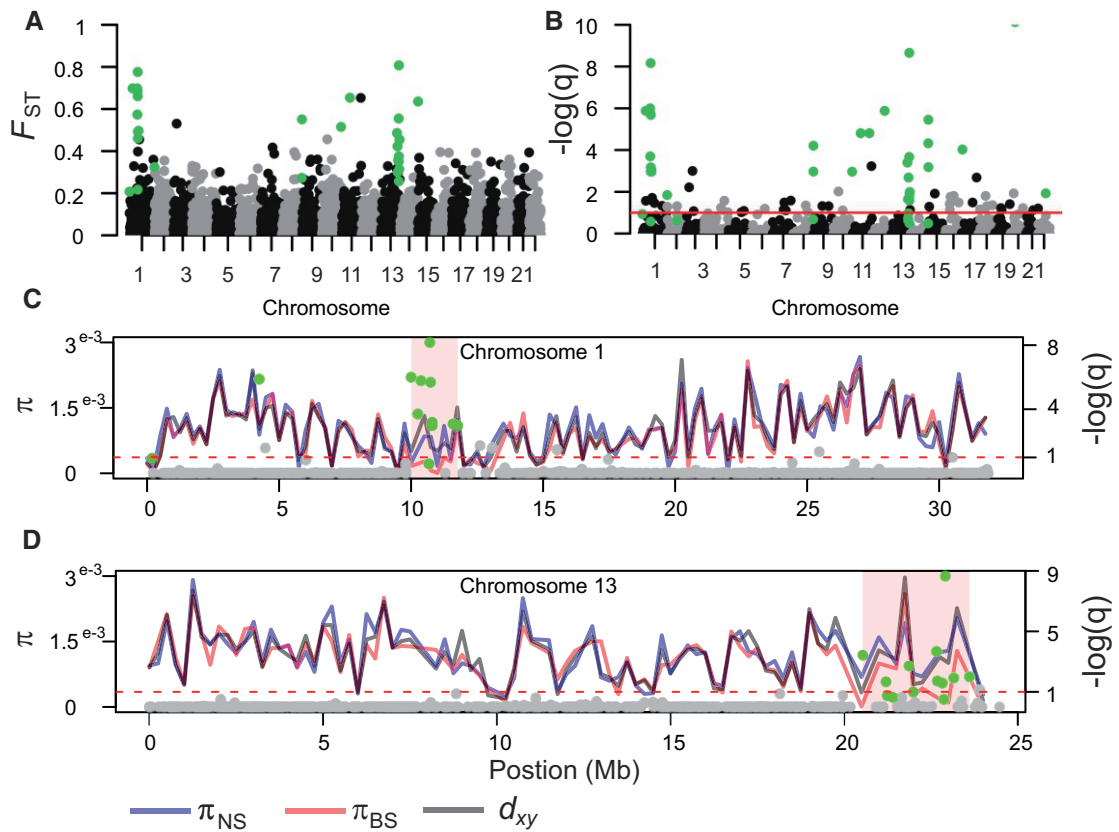
Stairway plots revealed that Baltic Sea and North Sea turbot shared a similar demographic history until 10–20 kya, with evidence of a population expansion between 20 and 100 kya (fig. 5D). The representative population for the Baltic Sea (DAB) shows signs of population contraction followed by growth following the end of the last glaciation, and lower contemporary  $N_e$  compared with the North Sea. We also

tested three simple one-population models using *moments* on all individuals from the North Sea and Baltic Sea separately: a standard neutral model (SNM, assuming constant population size at equilibrium), a two epochs model (2EP, assuming a sudden population change at time  $T_1$ ), and a three epochs model including a sudden demographic change at time  $T_1$  followed by a bottleneck at time  $T_2$  and exponential growth (3EP<sub>B</sub>). The results give strong support for population size changes, rejecting the SNM neutral mode for both populations. In the North Sea, the 2EP model had stronger support than the 3EP<sub>B</sub> model ( $W_{AIC}$  of 0.88 and 0.12, respectively), whereas for the Baltic Sea population, both models had similar support (0.57 and 0.43, respectively). One-population models provide guidance in selecting realistic demographic scenarios to test in more complex models but should otherwise be interpreted with extreme caution as they ignore the effects of gene flow. Here, results from both stairway plots and one-population models performed in *moments* suggest that both an ancestral population expansion and a bottleneck followed by growth in the Baltic Sea population should be formally tested in two-population models.

The folded jAFS used for testing two-population models included 16,270 biallelic SNPs. We tested the eight basic scenarios presented in figure 1C (IM, SC, IM<sub>B</sub>, SC<sub>B</sub>, IM<sub>AE</sub>, SC<sub>AE</sub>, IM<sub>AEB</sub>, and SC<sub>AEB</sub>) as well as modifications of these eight basic models accounting for heterogeneous migration rates (2M models) and  $N_e$  (2N models) across the genome as well as a combination of the two (2M2N models). Model choice and parameter estimates in two-population models were strongly affected by the inclusion or exclusion of an ancestral expansion and a bottleneck in the Baltic Sea population (fig. 7). Scenarios including an ancestral population expansion (AE) and a bottleneck (B) had lower AIC than simple IM and SC models, with AEB models performing best (fig. 7A). Within basic IM and SC models, SC models fitted the data significantly better showing much lower AIC than IM models. However, as more complex demographic scenario (AE and B models) were compared, the difference in AIC between IM and SC models became smaller, with IM models showing the lowest AIC in AEB scenarios. Interestingly a very similar pattern was observed in our simulations. In figure 7C, we show the AIC of IM and SC models (and their AE, B, and AEB variations) for a simulation with an ancestral expansion ( $N_{AE} = 2 \times N_{ANC}$ ) and bottleneck (where  $N_2$  at time of split is 4% of current  $N_2$ ). Models accounting for heterogeneous migration rates (2M) and  $N_e$  (2N) across the genome fitted the data better (fig. 7A), but the inclusion or exclusion of these parameters in the models did not change the relative support to IM and SC models.

The two best-fitting models were IM<sub>AEB2N</sub> and SC<sub>AEB2N</sub>, with a  $W_{AIC}$  of 0.66 and 0.31, respectively (fig. 7B). However, these two models converged to approximately the same scenario (fig. 8). Both models suggest an ancestral population expansion approximately 30–100 kyr before divergence to about 1.7 times the ancestral size, a colonization of the Baltic Sea <5.5 kya, and a strong reduction in  $N_e$  at the time of colonization followed by growth in the Baltic Sea population. The estimate of the period of strict isolation





**FIG. 6.** Genomic landscape of differentiation between North Sea and Baltic Sea turbot. (A) Per site  $F_{ST}$  from hard-called SNPs across the genome, SNPs significant in the HHM test are shown in green. (B)  $q$ -values (FDR) on a negative log scale from fastPCA genome scan for selection. The red line marks an FDR of 0.1, and green circles above this line denote SNPs which are significant based on an FDR threshold of 0.1 as well as based on results from the HHM test. (C) Patterns of genetic diversity across chromosome 1 calculated in non-overlapping windows of 250 kb. The red line represents  $\pi$  in the Baltic Sea, the blue line  $\pi$  in the North Sea, and the black line represents  $d_{xy}$ . Dots along the chromosomes represent  $q$ -values (FDR) on a negative log scale from fastPCA genome scan for selection for each individual SNP, the dotted line demarks the 0.1 FDR and green dots represent significant outliers according to the HHM test. (D) same as (C) but for chromosome 13.

( $T_{SI}$ ) from the SC model is very small (0.6 kyr, [fig. 8D](#)). Formal testing among these two competing, nested models using a likelihood ratio test (LRT), and adjusting the D statistics to account for possible effects of linkage ([Coffman et al. 2016](#)), gives no statistical support for the  $SC_{AEB2N}$  model ( $D_{adj} = 0.1293$ ,  $P = 0.3596$ ). LRT, on the other hand, provide support for the inclusion of a bottleneck in the model (LRT between  $IM_{2N}$  and  $IM_{B2N}$ :  $D_{adj} = 20.8501$ ,  $P < 10^{-4}$ ) and of an ancestral expansion (LRT between  $IM_{2N}$  and  $IM_{AE2N}$ :  $D_{adj} = 7.4156$ ,  $P = 0.0155$ ). Unscaled parameters for the two best models and their standard deviation estimated using the Fisher Information Matrix and the Godambe Information Matrix ([Coffman et al. 2016](#)) are given in [supplementary table 3](#), [Supplementary Material](#) online.

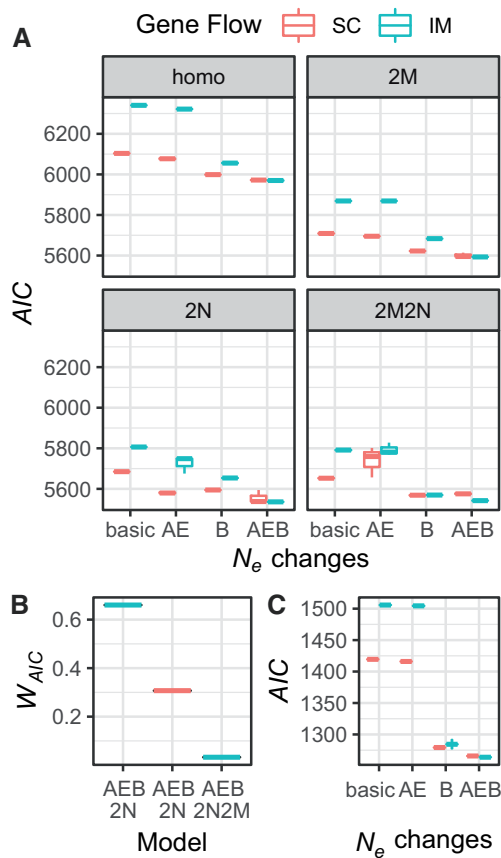
## Discussion

Reconstructing the demographic history of diverging populations is of central importance to understand the role of gene flow and periods of strict isolation in shaping the process of speciation. Here, we demonstrate that when the demographic history of the simulated taxa deviates from the tested scenarios, model choice and parameter estimation can be severely biased. Specifically, unmodeled changes in both

ancestral and daughter populations led to biases in estimates of divergence times and to favor scenarios that include periods of strict isolation. These biases can be minimized by comparing more realistic models, but a small systematic bias toward the choice of secondary contact models always remained. Using data from turbot populations from the Baltic Sea and the North Sea, we further demonstrate using an empirical case study that not accounting for potential demographic changes in both ancestral and daughter populations can lead to overestimate divergence times and conclude that these populations diverged during a long allopatric phase ([Le Moan et al. 2019](#)) whereas our results indicate a very recent divergence with constant gene flow.

## Lessons from Simulations

Testing basic models of divergence, such as AM, SC, IM, and SI, relies on the assumption that model choice is largely robust to unmodeled demographic events. Surprisingly, until now this assumption was never formally tested, though it has been shown that in the original IM program ([Hey and Nielsen 2004](#)) departures from assumptions can lead to strong biases in parameter estimation ([Becquet and Przeworski 2009](#)). Using extensive simulations, we demonstrate that under a



**Fig. 7.** Results from the 32 models optimized for the *Scophthalmus maximus* data (A and B) and data simulated with similar parameters (C). (A) shows AIC from the best three replicates of SC (red boxplots) and IM (blue boxplots) models for any combination of ancestral expansion/bottleneck (AE, B, and AEB models). Top left panel (homo): homogenous gene flow and  $N_e$ . Panel 2M: heterogeneous migration rates across the genome. Panel 2N: heterogenous  $N_e$  across the genome. Panel 2M2N: heterogenous migration rates and  $N_e$  across the genome (B)  $W_{AIC}$  of the three best models. (C) Results from simulations representing a scenario similar to the best inferred model (IM scenario with  $N_{AE} = 2 \times N_{ANC}$  and a bottleneck in population two at time  $T_S$  to 4% of the current  $N_e$ ).

very broad range of divergence scenarios this assumption does not hold. Recent bottlenecks followed by growth in a daughter population always lead to strong support for SC models, regardless of  $T_S$ . However, not accounting for a recent bottleneck can lead to overestimate or underestimate  $T_S$  depending on whether divergence is recent or older. This is because estimates of  $T_S$  tend to reflect the strength of the recent bottleneck, rather than divergence time. Another pattern emerging from our simulations is that failure to model a change in  $N_e$  in the ancestral population can lead to biases in model choice and to overestimate divergence time, but this effect depends both on how recent divergence is and on how much time has passed between  $T_{AE}$  or  $T_{AB}$  and  $T_S$ . Biases caused by unmodeled demographic changes in the ancestral population are most severe for recent divergence scenarios (i.e., when  $T_S \ll T_{AE}$  or  $T_{AB}$ ) and their effect on demographic inference fades when  $T_S \approx 1$ . Similarly, the effect of a change in  $N_e$  in the ancestral population on parameter estimation

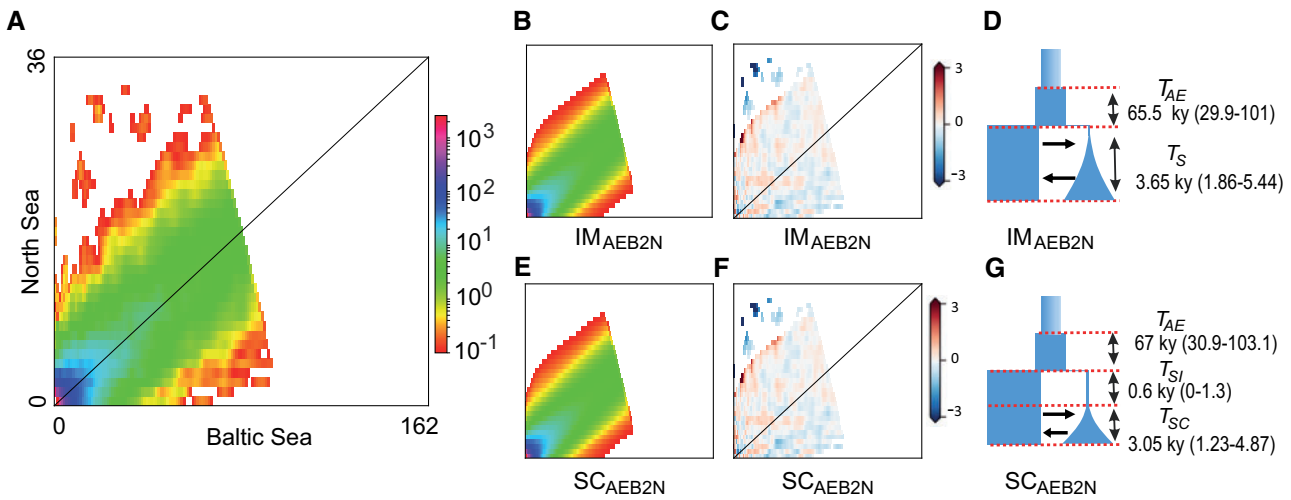
and model choice becomes irrelevant when the last change in  $N_e$  occurred  $\approx 4N_e$  generations before  $T_S$ . This is theoretically expected, as it takes approximately  $4N_e$  generations for a population to reach mutation–drift equilibrium (Kimura and Ohta 1969; Lande 1980).

The most commonly tested demographic models assume an ancestral population at mutation–drift equilibrium and a change in  $N_e$  is permitted only at the time  $T_S$ . Therefore, an unmodeled ancestral expansion or contraction could push estimates of  $T_S$  back to the time of the ancestral change in  $N_e$ . When  $T_S$  is small, demographic models that did not account for ancestral expansion tended to overestimate  $T_S$  by up to a factor of ten (as  $T_{AE} + T_S = 0.5$ , i.e.,  $10 \times T_S$ ). In our SC coalescent simulations of recent divergence scenarios  $T_S = 0.1$  and  $T_{AE} + T_S = 0.5$ , and as expected, the overestimation of  $T_S$  was up to a factor of five. In IM simulations, unmodeled ancestral expansions also led to a bias toward SC models.

In our simulations of recent divergence scenarios under the SC model, an ancestral population expansion in some cases led to a slight bias toward choosing IM models. This is most likely because the effect of not modeling an ancestral expansion is to push  $T_S$  back in time, which under the scenario of long strict isolation will result in much longer divergence times. Roux et al. (2016) also demonstrated that when the period of strict isolation preceding SC is a small proportion (<60%) of the total divergence time, distinguishing between IM and SC can be very difficult. When the true divergence scenario is a basic SC model, we observed that IM models that included abrupt population size changes performed better than IM models that did not (but always worse than SC models). For example,  $IM_B$  models fitted better than IM models, and  $IM_{AEB}$  models had even stronger support (supplementary figs. 2–9, Supplementary Material online).

All these observations taken together suggest that extreme caution should be exercised when choosing among competing divergence scenarios using methods based on the jAFS or its summary statistics. It is known that, if one allows competing models to be arbitrarily complex, there is an infinite number of demographic histories that can produce the same AFS (Myers et al. 2008). In reality, when comparing more biologically realistic models, a unique function producing the expected AFS is identifiable (Bhaskar and Song 2014; Rosen et al. 2018). However, we demonstrate that when the models compared do not match closely the demographic history of the simulated populations, several demographic parameters in the model can yield better fits but lead to the wrong biological conclusion. Herein lies the major issue in interpreting results from demographic modeling: models are always extreme simplification of complex biological processes, and we often do not know what complexities can be safely excluded. Several studies (Roux et al. 2014, 2016; Ewing and Jensen 2016; Pouyet et al. 2018) clearly showed that the effects of barrier loci and linked selection should be accounted for and here we demonstrate that historical changes in  $N_e$  cannot be ignored.

The systematic bias toward the choice of SC models when the real scenario generating the data had continuous gene flow, and the general overestimation of the proportion of



**FIG. 8.** Inferred demographic history and model fit from the two models with highest support. Observed jAFS (A), modeled jAFS (B and E), model residuals (C and F), and parameters (D and G) for the two best models. (B) modeled jAFS, (C) residuals, and (D) graphical representation of the isolation with migration model including a bottleneck in the Baltic Sea population, an ancestral change in  $N_e$  and heterogenous  $N_e$  across the genome (IM<sub>AEB2N</sub> model) as well as estimated time parameters with 95% confidence intervals. (E) modeled jAFS, (F) residuals, and (G) graphical representation of the secondary contact model including a bottleneck in the Baltic Sea population, an ancestral change in  $N_e$  and heterogenous  $N_e$  across the genome (SC<sub>AEB2N</sub> model) as well as estimated time parameters with 95% confidence intervals estimated from the Godambe Information Matrix.

strict isolation for SC models, suggest that the use of simple IM and SC models to differentiate between primary and secondary divergence may often lead to the wrong conclusion. These findings have important repercussions on how we interpret demographic analyses of recent divergence scenarios; changes in  $N_e$  in ancestral populations have almost inevitably happened during past glacial cycles, and bottlenecks followed by population expansions are a classic signature of colonization of novel environments (Hewitt 1999; Liu et al. 2016; Feng et al. 2020). Furthermore, as testing on a smaller number of simulations demonstrated, this bias is not unique to the main method employed in this manuscript (*moments*), but also applies to another very widely utilized approach to estimate demographic parameters based on the jAFS of multiple populations, *dadi* (Gutenkunst et al. 2009, supplementary fig. 18, Supplementary Material online). It is unclear at this stage what the effect of unmodeled changes in  $N_e$  in ancestral and daughter populations would be on model choice and parameter estimation under an ABC framework. Most likely this will depend upon the choice of summary statistics. Most summary statistics commonly used (Watterson's  $\Theta$ ,  $\pi$ , Tajima's  $D$ ,  $F_{ST}$ , and  $d_{xy}$ ) are summaries of the AFS and therefore are also expected to be affected. However, statistics such as the decay of linkage disequilibrium are not, and could perhaps be less sensitive to these biases (Jay et al. 2019). It is interesting to notice that when reconstructing the demographic history of model organisms, it is common to test more realistic demographic scenarios modeling past demographic changes in ancestral and daughter populations as well as bottlenecks followed by population growth (Gutenkunst et al. 2009; Gravel et al. 2011; Garud et al. 2015; Jouganous et al. 2017). However, these realistic demographic scenarios are more seldom considered when testing IM and SC models in non-model species. This possibly is in part due to the

assumption that the data are inadequate to deal with such model complexity, and in part due to the assumption that model choice and parameter estimation is robust to such unmodeled demographic events. Here, we showed that neither of these assumptions is correct. It should also be noted that a novel unsupervised approach for inferring demographic histories that performs jointly model structure and parameter optimization (GADMA; Noskova et al. 2020) has the potential to alleviate the biases we describe here.

It is not our intention to suggest that most studies that compared IM and SC models and found strong support for SC have likely chosen the wrong scenario. Several very recent studies have indeed modeled the potential effects of bottlenecks (Montano et al. 2015; Christe et al. 2017; Rougeux et al. 2017, 2019; Hartmann et al. 2020; Rougemont et al. 2020), and formal model testing is often only one of several lines of evidence suggesting secondary contact (Roux et al. 2014; Tine et al. 2014; Le Moan et al. 2016; Rougemont et al. 2017). For example, a correlation between  $F_{ST}$  and  $d_{xy}$  (i.e., elevated divergence in genomic islands of differentiation) provides further evidence for SC (Cruickshank and Hahn 2014; Duranton et al. 2018, 2020; Gagnaire et al. 2018; Nelson and Cresko 2018). Nevertheless, the results clearly show that realistic demographic scenarios can generate strong false support for SC models with long periods of strict isolation. This has clear implications for studying incipient speciation and recent population divergence, for example, as a result of range expansions and colonization of novel habitats following the end of the last glaciation.

It must also be considered that we explored a limited number of unmodeled demographic events. Spatial genetic structure, recent range expansions and admixture from ghost populations are all common demographic events, and these scenarios can also lead to biases in demographic inference

(Delser et al. 2019). Similarly, although it is relatively simple to account (albeit in a coarse way) for linked background selection by modeling heterogeneity of  $N_e$  across the genome, it is more difficult to model the possible effects of a reduction in  $N_e$  in genomic regions of a single population (i.e., the expectation for linked positive selection). Since selective sweeps have the same local effect of a bottleneck (and can indeed lead to false inferences of changes in  $N_e$ , e.g. Schrider et al. 2016), it is reasonable to assume that the presence of large recent selective sweeps may also lead to a bias toward SC models.

### Lessons from the Turbot's Demographic History

A clear example of when the biases we describe in this study are especially problematic is given by the study of the origin of the Baltic Sea biodiversity. The Baltic Sea is a large body of brackish water which became connected to the North Sea about 8 kya. Its marine fauna has probably more than one origin, with evidence of populations and species in the Baltic Sea being the result of both primary and secondary divergence (reviewed in Johannesson et al. 2020). The evolutionary origin of some specific taxa, such as the Baltic Sea populations of Pleuronectiformes, remains controversial (Momigliano et al. 2017, 2018; Jokinen et al. 2019; Le Moan et al. 2019). In our empirical study, our model fit suggests that *S. maximus* from the Baltic Sea originated from a very recent invasion (<6 kya) from the North Sea and diverged with continuous gene flow. One-population models suggest that North Sea and Baltic Sea *S. maximus* share the same demographic history, with an ancestral population expansion that occurred 35–102 kya, until approximately 5 kya, roughly the divergence time estimated by our two-population models. After this, both stairway plots and two-population models show support for a founder event coincident with the time at which the Baltic Sea had the highest salinity in its history (Gustafsson and Westman 2002). At this time, as noted by Momigliano et al. (2017), there would have been broad opportunity for marine fish to colonize the Baltic Sea. Failure to account for these realistic demographic events would have resulted in a very strong ( $W_{AIC} > 0.999$ ) support for SC models and estimates of divergence times that predate the origin of the Baltic Sea.

Le Moan et al. (2019) reconstructed the demographic history of North Sea–Baltic Sea population pairs for five flatfish species (including *S. maximus*) using similar data and approaches as in this study, but assuming an ancestral population at equilibrium and no bottleneck followed by population growth associated with the invasion of the Baltic Sea. The authors found strong support for SC in four out of the five populations studied, with estimates of divergence time for each of these population pairs predating the origin of the Baltic Sea by at least a factor of five (Le Moan et al. 2019). Since their estimates of timing of secondary contact was different for the four species, the authors concluded that the population-pairs diverged in strict isolation in several unidentified marine refugia, and colonized the Baltic Sea at different times following the end of the last glaciation (Le Moan et al. 2019). As there is no other evidence for such scenario apart

from testing of IM and SC demographic models, it is possible that Le Moan et al. (2019) results are a product of the biases we described here. We also note that when we do not assume an ancestral population at equilibrium and model potential bottlenecks, IM and SC models converge toward the same demographic scenario (fig. 8) giving strong support for postglacial divergence between flatfish populations in the North Sea and the Baltic Sea. Furthermore, the likelihood ratio test provided no support for the secondary contact model. Interestingly, Momigliano et al. (2017) used ABC to model the divergence of the flounder species pair (*Platichthys flesus* and *P. solemdali*) in the Baltic Sea taking into account both potential changes in  $N_e$  in ancestral and daughter populations (but with no formal testing of IM and SC scenarios), finding support for postglacial colonization of these two flatfish species. Improved modeling in this study has led to support for a more biologically plausible history of *S. maximus*'s invasion of the Baltic Sea but we wish to caution that it is entirely possible that new data and/or better modeling approaches will provide in the future support for an alternative evolutionary scenario. We also wish to caution against overinterpreting scaled parameters, since their value is dependent on the mutation rate, the exact value of which is not known.

The genomic landscape of differentiation between North Sea and Baltic Sea turbot populations also points toward shallow divergence and recent selection in the Baltic Sea. Differentiation across the genome was generally low, and genomic islands of high allelic differentiation ( $F_{ST}$ ) were driven by strongly reduced  $\pi$  in the Baltic Sea rather than by a local increase in  $d_{xy}$ , that is, the classic signature of a recent selective sweep. Indeed within these genomic islands,  $d_{xy}$  approximates  $\pi$  in the North Sea; assuming the North Sea still represents ancestral population diversity, this suggests that net divergence among the selected haplotypes is close to 0. As noted by Cruickshank and Hahn (2014), genomic islands of differentiation that evolved in allopatry and resist introgression (the expected pattern under a model of SC and heterogeneous gene flow) are expected to instead show increased levels of absolute divergence.

### Conclusion

In conclusion, using very extensive simulations as well as empirical data on turbot, we demonstrate that testing IM and SC models can be difficult when the demographic history of the studied taxa deviates from the tested scenarios. We conclude that statistical support for SC or IM in model testing can often be an artifact of unmodeled demographic events, and that estimates of  $T_S$  can often reflect recent changes in  $N_e$  rather than divergence time. Given the centrality of formal testing between competing divergence scenarios in current research on local adaptation and speciation, these biases should not be ignored. Testing one-population models can provide guidance in identifying what demographic scenarios need to be incorporated in formal model testing, and testing more realistic demographic scenarios is paramount for avoiding at least the most severe biases described in this manuscript. However, even when testing more realistic divergence

models, extreme caution should be exercised when interpreting results.

## Materials and Methods

### Analyses of Simulated Data

#### *Simulations of Older Divergence Scenarios*

We tested the effect of recent bottlenecks as well as ancestral expansions and contractions on model choice and parameter estimation within divergence scenarios with  $T_S$  ranging from 4,000 to 128,000 generations. To do this, we first generated several simulations under an Isolation with Migration model using the software *ms* (Hudson 2002). All simulated scenarios within this study share a few common parameters. The simplest models represent a scenario where an ancestral population of size  $N_{ANC}$  splits at time  $T_S$  in two populations of size  $N_1$  (which is always fixed at 20,000 individuals, and is used as the reference population size  $N_{REF}$ ) and  $N_2$  ( $N_2 = 0.25 \times N_1$ , i.e., 5,000 individuals). The migration rate  $m$  is, unless explicitly stated, symmetrical and set to four. The migration rate is given in units of  $M = 4N_{REF}m$ , where  $M$  is the fraction of each population which is made up of migrants at a given generation. We explored the effects of unmodeled demographic events across six divergence times, with  $T_S$  values ranging from 0.05 to 1.6 in units of  $4N_{REF}$  generations (i.e., 4,000 to 128,000 generations) in log2 steps (i.e.,  $T_S = 0.05, 0.1, 0.2, 0.4, 0.8, 1.6$ ).

For each  $T_S$ , we generated simulations including a bottleneck in population two (fig. 1A). The time of the bottleneck remains constant at 0.05 ( $T_B$ ), as we aim to represent the effect of a recent bottleneck associated with fluctuations in  $N_e$  within the last glacial cycle. We simulated bottlenecks of different strengths, so that  $N_2$  at times  $T_B$  ranges from 1% to 64% in log2 steps of current  $N_2$ , and following the bottleneck population two experiences an exponential growth that starts at time  $T_B$  and continues until present. This resulted in six simulations without bottlenecks (one for each  $T_S$ ) as well as 48 simulations including a bottleneck in population two (six  $T_S$  and eight strengths of bottleneck).

We further investigated the effects of changes in effective population size in the ancestral population, that is, an ancestral expansion (AE) and an ancestral bottleneck (AB) (fig. 1A). We included nine scenarios of population expansion at time  $T_{AE} + T_S$ , and ancestral expansions were modeled as different values of ancestral population size  $N_{ANC}$  (to  $\frac{1}{4}$ ,  $\frac{1}{16}$  and  $\frac{1}{64}$  of  $N_{REF}$ ). We modeled every possible combination of six values of  $T_S$  (as above), three values for  $T_{AE}$  ( $T_{AE} = 0.025, 0.5, 1$ ) and three strengths of ancestral expansion for a total of 54 independent simulations.

We simulated, in the same way, scenarios where the ancestral population underwent first a demographic contraction followed by expansion (fig. 1A). The time of the demographic contraction is set to  $T_{AB} + T_S$ , where  $T_{AB}$  represents the number of generations (in units of  $4N_{REF}$ ) before  $T_S$  at which the demographic contraction takes place. At time  $T_S + T_{AB}/2$ , the ancestral population returns to its original size, which is equal to  $N_{REF}$ . As per the AE scenario, we modeled every possible combination of six values of  $T_S$  (as above),

three values for  $T_{AB}$  ( $T_{AB} = 0.025, 0.5, 1$ ) and three strengths of ancestral contraction for a total of 54 independent simulations.

For each of the simulations above, we simulated sampling of 20 individuals per populations and of 1 million unlinked loci of a length of 36 bp, which, when using a standard germ-line mutation rate ( $\mu$ ) of  $1 \times 10^{-8}$  and a  $N_{REF}$  of 20,000 individuals, results in roughly 80,000–120,000 unlinked SNPs (depending on the specific model). We then used the unfolded jAFS from the simulated data for demographic inference.

#### *Simulations of Recent Divergence Scenarios*

We furthermore tested more extensively the effects of unmodeled demographic events on model choice and parameter estimation on recent divergence scenarios (when  $T_S = 0.05$  and  $0.1$ , i.e., 4,000–8,000 generations). These simulated scenarios are particularly relevant to the empirical case study we present later (the divergence of turbot populations in the North Sea and Baltic Sea). Here, we simulated data under four gene flow scenarios: SC and IM, each with symmetric ( $M = 4$ ) and asymmetric gene flow ( $m_{12} = 4$  and  $m_{21} = 16$  where  $M_{ij} = 4N_{REF}m_{ij}$  and  $M_{ij}$  is the proportion of individuals in population  $i$  which is made up of migrants from population  $j$ ). Modeling gene flow as the proportion of migrants in a given population (rather than the proportion of individuals migrating from a population), while not always realistic, has the advantage to keep gene flow constant even while  $N_e$  fluctuates among populations exchanging genes. For each of these gene flow scenarios, we modeled 64 combinations of demographic events in a fully orthogonal design (fig. 1B). The simplest models represent IM and SC scenarios where an ancestral population of size  $N_{ANC} = N_{REF}$  splits at time  $T_S$  in two populations of size  $N_1$  and  $N_2$  (which have the same values as given above). We then included seven scenarios of population expansion at time  $T_{AE}$  ( $T_{SC} + T_{AE} = 0.5$ , in units of  $4 \times N_{REF}$  generations, i.e., 40,000 generations), ranging from  $2 \times N_{ANC}$  to  $128 \times N_{ANC}$  on log2 steps. Since our reference population size ( $N_{REF}$ ) for parameter scaling is always  $N_1$ , ancestral expansions were modeled as different values of  $N_{ANC}$  (to  $\frac{1}{2}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$ ,  $\frac{1}{16}$ ,  $\frac{1}{32}$ ,  $\frac{1}{64}$  and  $\frac{1}{128}$  of  $N_1$ ). Furthermore, we modeled seven scenarios of bottlenecks followed by exponential growth in population two. In these scenarios  $N_2$  at time  $T_S$  can be 1%, 2%, 4%, 8%, 16%, 32%, and 64% of contemporary  $N_2$ , reflecting a range of strong to very mild reduction in  $N_e$  at time of divergence. Following the bottleneck, exponential growth took place during a period lasting 2,000 generations following  $T_S$ , after which the  $N_e$  of population two reached  $N_2$ . This kind of scenario is meant to reflect the invasion of a novel habitat, for example, a new environment that became available after the end of the last glaciation. In IM models, the time of divergence is set at 4,000 generations ago ( $T_S = 0.05 \times 4N_{REF}$ ). In SC models, the time of divergence is set at 8,000 generations ago ( $T_S = 0.1 \times 4N_{REF}$ ), whereas secondary contact is established at time  $T_{SC}$  ( $T_{SC} = 0.25 \times T_S$ ), that is, 2,000 generations ago.

Fully orthogonal combinations of all the demographic scenarios outlined above (SC and IM, symmetric and asymmetric gene flow, ancestral expansions, and bottlenecks) resulted in a total of 256 simulated scenarios. Firstly, for each of the 256 scenarios, we simulated sampling of 20 individuals per population and of 1 million unlinked loci of a length of 36 bp (as for the simulations above). We then estimated the folded and unfolded jAFS from the simulated data. These data sets represent standard data sets when working with high-quality whole-genome data, assuming both scenarios whereby the genome of a closely related species is and is not available to polarize the jAFS. Secondly, for each of the 256 scenarios, we simulated sampling of ten individuals per population and of 100,000 unlinked loci of a length of 36 bp, resulting in roughly 8–12,000 unlinked SNPs (depending on the specific model) and estimated again both the unfolded and the folded jAFS. These data sets represent a standard, small-scale 2b-RAD data set for a non-model species (i.e., the kind of data that most people working of non-model organisms can easily access). This led to a total of 512 coalescent simulations of recent divergence scenarios.

#### Demographic Modeling of Simulated Data

Demographic modeling of simulated data was carried out using the software package *moments* (Jouganous et al. 2017), which is a development of the *dadi* (Gutenkunst et al. 2009) method for demographic inference from genetic data based on diffusion approximation of the allele frequency spectrum. *Moments* introduces a new simulation engine based on the direct computation of the jAFS using a model of ordinary differential equations for the evolution of allele frequencies that is closely related to the diffusion approximation used in *dadi* but avoids some of its limitations (Jouganous et al. 2017). Firstly, for all simulated scenarios, we tested whether a simple isolation with migration (IM) or a secondary contact (SC) model fitted the data best (IM and SC models in fig. 1C). The models consist of an ancestral population of size  $N_{\text{ANC}}$  that splits into two populations of sizes  $N_1$  and  $N_2$  at time  $T_S$ . In the IM model, there is continuous asymmetric migration. In the SC model, there is a period of strict isolation starting at time  $T_S$  followed by a period of secondary contact with asymmetric migration starting at time  $T_{\text{SC}}$ .

For the 256 simulations of recent divergence scenarios, we also tested models that accounted for population size changes in the ancestral population (AE models) and bottlenecks followed by growth in population two (B models), as well as both demographic changes (AEB) models (fig. 1C). Therefore, for both IM and SC scenarios, we had four alternative models: a basic scenario assuming an ancestral population at equilibrium and instantaneous size changes at time  $T_S$  and the three above-mentioned combinations of demographic changes in the ancestral population and in population two (AE, B, and AEB models). It is to be noted that models including bottlenecks intentionally do not match exactly the coalescent simulations (fig. 1B), but rather mirror how growth is modeled in previously utilized models in speciation research (Rougeux et al. 2017). In  $\text{IM}_B$  models,

exponential growth starts at time  $T_S$  and continues until the present, whereas in  $\text{SC}_B$  models, exponential growth starts at time  $T_{\text{SC}}$  and continues until the present (fig. 1B).

Models were optimized for three rounds following an approach similar to Portik et al. (2017). In the first round, a Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm optimization (function “optimize.log”; max-iter = 10) was run for ten sets of three-fold randomly perturbed parameters. In the second round, the parameters from the replicate with the highest likelihood from each model were used as a starting point and the same optimization algorithm was used on ten sets of two-fold randomly perturbed parameters, increasing max-iter to 20. We repeated the same procedure for round three, but using one-fold perturbed parameters and setting max-iter to 30. For models including ancestral sizes changes and bottlenecks, we optimized parameters using four rounds of optimization. We ran the entire optimization procedure ten times to check for convergence among independent optimizations. We selected the replicate with the highest likelihood for each of the two-population models and calculated the Akaike Information Criterion as  $-2(\log -\text{likelihood}) + 2K$  where  $K$  is the number of model parameters, and the  $\Delta\text{AIC}_i = \text{AIC}_i - \text{AIC}_{\text{min}}$ . We then calculated Akaike weight of evidence ( $W_{\text{AIC}}$ ) as outlined in Rougeux et al. (2017). The equation is outlined below, and  $R$  represents the total number of models compared.

$$W_{\text{AIC}} = \frac{e^{\frac{-\Delta\text{AIC}_i}{2}}}{\sum_{i=1}^R e^{\frac{-\Delta\text{AIC}_i}{2}}}$$

Firstly, we looked at the effect of unaccounted demographic events when our model choice was restricted to the basic IM and SC models (i.e., excluding models that account for demographic size changes in the ancestral population and a bottleneck and growth in population two). We evaluated not only if the IM and SC models were correctly identified but also whether not accounting for ancestral expansion and bottlenecks affected parameter estimation.

Second, for all scenarios of recent divergence (i.e., the ones more relevant to our empirical study), we looked at whether including models with ancestral population expansions and bottlenecks improved model choice and parameter estimation. When comparing the eight full models, we calculated the  $W_{\text{AIC}}$  for each gene flow scenario (IM vs. SC) as the sum of  $W_{\text{AIC}}$  for all IM and SC models. This comparison among family of models was carried out since at times the correct gene flow scenario was identified, but one of the demographic changes modeled was not (i.e., this process maximized our chance to recover the simulated gene flow scenario). For the simulations of ancestral expansions (AE) and ancestral bottlenecks (AB), we used only the unfolded jAFS. Since in our empirical study we use the folded jAFS, for all 512 simulations of recent divergence scenarios, model optimization and parameter estimation were carried out using both the folded and unfolded jAFS, to test whether demographic inference using the unpolarized jAFS is less or more susceptible to biases in model choice and parameter estimation.

### Comparison between Methods

To determine whether the biases reported in this study were specific to *moments* or reflected a general issue in inferring demographic histories from the jAFS, we repeated a portion of the analyses using the diffusion approximation approach implemented in *dadi*. We performed model choice and parameter estimation for the 126 simulations of recent divergence scenarios under the IM model (fig. 1B), with symmetric and asymmetric gene flow and using only the simulations of the larger data sets (1 million loci). We used exactly the same optimization strategy, but only ran three independent optimization routines, which were sufficient to get convergence for independent runs of the simple IM and SC models.

### Analysis of Empirical Data

#### Sampling

We obtained a total of 172 samples of *S. maximus* from seven locations and three biogeographic regions (supplementary table 2, Supplementary Material online and fig. 5A): the North Sea (one location,  $N = 20$ ), the transition zone separating the North Sea from the Baltic Sea (two sampling locations, Vendelsö:  $N = 27$  and Öresund:  $N = 35$ ), and the Baltic Sea (four populations, Dabki:  $N = 24$ , Gdynia:  $N = 23$ , Gotland:  $N = 20$  and Gotska Sandön:  $N = 24$ ). Individuals from the transition zone and the Baltic Sea are a subsample of the individuals analyzed by Florin and Höglund (2007), whereas samples from the North Sea were originally collected by Nielsen et al. (2004).

#### Library Preparation

We built 2b-RAD libraries following the approach described by Wang et al. (2012), but with degenerate adaptors to allow identification of PCR duplicates. The protocol is described in detail by Momigliano et al. (2018). In short, DNA was extracted using a modified salting out protocol, and about 200 ng of DNA was digested with the type II b enzyme Bcgl (New England Biolabs). This enzyme cuts both upstream and downstream of the 6-bp recognition site, creating fragments of a length of exactly 36 bp with 2-bp overhangs. Adaptors, one of which included degenerate bases, were ligated and the fragments amplified via ten cycles of PCR as described in Momigliano et al. (2018). Fragments of the expected size were isolated using a BluePippin machine (Sage Science). Libraries were sequenced on Illumina machines (NextSeq500 and HiSeq 4,000) to achieve a mean coverage of approximately  $20\times$ .

#### Bioinformatics and Basic Population Genetic Analyses

Raw reads were demultiplexed and PCR duplicates were removed as per Momigliano et al. (2018). Reads were then mapped to the latest version of *S. maximus* reference genome (Figueras et al. 2016, Assembly ASM318616v1, GenBank accession number: GCA003186165) using Bowtie2 (Langmead and Salzberg 2012). SAM files were converted to BAM files and indexed using SAMTOOLS (Li et al. 2009). A genotype likelihood file in beagle format was produced in the software ANGSD (Korneliusson et al. 2014), using the following filters:

no more than 20% missing data, retaining only biallelic loci, removing bases with mapping quality below 30 and Phred quality scores below 20. See supplementary figure 1, Supplementary Material online, for summary statistics. A principal component analysis (PCA) based on genotype likelihoods was performed using the software PCAngsd (Meisner and Albrechtsen 2018) using only variants with a minor allele frequency above 0.02. The folded AFS for each population as well as the jAFS were produced in ANGSD. ANGSD calculates folded jAFS where the minor allele is computed separately for each AFS, whereas *moments* expects minor alleles to be estimate for the jAFS. Thus, for the jAFS for North Sea and Baltic Sea, we produced the unfolded jAFS in ANGSD in the form of a *dadi* data dictionary. We then selected a random SNP within each locus and folded the jAFS in *moments*.

We produced also a variant call file (VCF) using the UnifiedGenotyper function from GATK v.3.8. Following UnifiedGenotyper, we removed individuals with an average read depth below seven. Then, we used four technical replicate pairs (i.e., four pairs of individuals for which we constructed and sequenced two independent libraries) to generate a list of SNPs for which we have high confidence (i.e., which show 100% matches between all replicate pairs). We used this list of SNPs to carry out variant quality score recalibration (VQSR), following GATK best practice (Dixon et al. 2015). Finally, we removed genotype calls with low sequencing depth ( $<7$ ), removed indels, triallelic SNPs, SNPs with minor allele frequencies below 0.01 and with more than 10% missing data. This resulted in a final VCF containing 12,678 sites genotyped for 154 individuals. This data set was used to calculate Weir and Cockerham  $F_{ST}$  for each SNP. Furthermore, we thinned the data retaining, for each tag, the SNP with the highest minor allele frequency and used this data for inferring population structure using fastSTRUCTURE (Raj et al. 2014). Bioinformatic steps, scripts for analyses, and the jAFS used are publicly available (see Data Availability).

#### Inferring the Genomic Landscape of Differentiation

We used several approaches to identify potential islands of differentiation across the genome between North Sea and Baltic Sea turbot. Firstly, we calculated from the called genotypes  $F_{ST}$  between the North Sea and Baltic (excluding individuals from the transition zone) using VCFTOOLS (Danecek et al. 2011). This is the only measure of differentiation we calculated from called genotypes. Secondly, we used the software package PCAngsd to run a selection scan using an extended model of fastPCA (Galinsky et al. 2016) working directly on genotype likelihoods, based on the input beagle file we used for the PCA. This approach identifies variants whose differentiation along a specific principal component (in our case the first PC) is greater than the null distribution under genetic drift. To account for multiple comparisons, we converted  $P$  values to  $q$ -values (false discovery rate, FDR) following Benjamini and Hochberg (1995). As a second approach to classify SNPs as outliers, we used a Hidden Markov model (HMM) approach to detect genomic islands,

based on the uncorrected  $P$  values from PCAnsd selection scan (Hofer et al. 2012; Marques et al. 2016). We counted as candidate outliers SNPs that show an FDR <0.1 and that simultaneously were identified as outliers by the HMM test. If adjacent SNPs identified by both approaches lied within a distance of <500 kb, we identified them as part of the same genomic island of differentiation.

We then obtained estimates of within population genetic diversity ( $\pi$ ) and absolute divergence ( $d_{xy}$ ) across the genome. In order to maximize the usable data and account for differences in coverage among samples, we performed all analyses in ANGSD directly from genotype likelihoods. Firstly, we generated windows of 250 kb across the genome using BEDTools (Quinlan and Hall 2010). Then, we calculated the unfolded AFS (for the North Sea and Baltic Sea individuals) as well as the jAFS for each 250-kb window across the genome with ANGSD, using only filters that do not distort the AFS (-uniqueOnly 1 -remove\_bads 1 -minMapQ 20 -minQ 20 -C 50). We finally used custom R scripts to calculate  $\pi$  and  $d_{xy}$  for each window, retaining only windows for which the AFS was derived from at least 1,000 sequenced bases. All scripts to calculate the AFS in windows and derive summary statistics are available from GitHub and Zenodo (see Data Availability). Note that, for diversity analyses, we used the unfolded AFS even if we do not have an appropriate outgroup, assuming the reference allele as the ancestral state. However, this is not an issue since the summary statistics calculated are based on allele frequencies, which are symmetrical with respect to folding.

### Demographic Modeling of Empirical Data

The demographic history of the North Sea and Baltic Sea populations of *S. maximus* were reconstructed using several approaches based on the 1d-AFS (for one-population models) and the jAFS (for two-population models). Since taking into account possible changes in effective population size may have important effects on the estimation of parameters such as migration rate and divergence times (Gravel et al. 2011), we first determined the demographic history of each population independently using the 1d-AFS, using both *moments* and stairway plots (Liu and Fu 2015). We then proceeded to compare 32 two-population models to determine the demographic history from the jAFS. We used a  $\mu$  of  $1 \times 10^{-8}$  and a generation time of 3.5 years for scaling demographic events to make results directly comparable to Le Moan et al. (2019).

### One-Population Models

Firstly, we estimated past demographic changes in the North Sea populations and from the Baltic Sea populations (i.e., excluding samples from the transition zone) using the multi-epoch model implemented in the software Stairway Plot v2 (Liu and Fu 2015). Stairway plots use composite likelihood estimations of  $\Theta$  at different epochs, which is then scaled using the mutation rate. For Baltic Sea populations, we estimated past demographic changes from the 1d-SFS from each sampling location independently. Stairway plots were

generated including singletons, using 2/3 of the sites for training and four numbers of random break points for each try ( $\frac{1}{4}$ ,  $\frac{1}{2}$ ,  $\frac{2}{3}$  and 1 time the number of samples in each population). Since demographic histories were similar in all locations, and there was no evidence of population structure from other analyses, all subsequent analyses were performed using the jAFS estimated from pooling all samples from populations in the Baltic Sea.

Secondly, we compared three simple one-population models using *moments*: a standard neutral model (SNM, assuming constant population size at equilibrium), a two-epoch model (2EP, assuming a sudden population change at time  $T_1$ ), and a three-epoch model including a sudden demographic change at time  $T_{AE}$  followed by a bottleneck at time  $T_B$  followed by exponential growth (3EP<sub>B</sub>). The 2EP model represents a single demographic change and a scenario where a demographic expansion/contraction occurred either in the ancestral population from which the North Sea and Baltic Sea populations are derived or in the North Sea and Baltic Sea populations themselves. The 3EP<sub>B</sub> model represents a scenario where, in addition to an ancestral expansion/contraction, there was a recent bottleneck followed by growth; this could be a realistic scenario for the Baltic Sea populations, which must have invaded the Baltic Sea following its connection to the North Sea within the past 8 kyr.

### Two-Population Models

Given the existence of a well-known hybrid-zone between the Baltic Sea and the North Sea (Nielsen et al. 2004), we tested the two main divergence scenarios that include contemporary migration: the isolation with migration model (IM), and the secondary contact (SC) models. All models consist of an ancestral population of size  $N_{ANC}$  that splits into two populations of size  $N_1$  and  $N_2$  at time  $T_S$ . Migration is continuous and asymmetric in the IM model. The SC model includes a period of isolation starting at time  $T_S$  and a period of secondary contact when asymmetric migration starts at time  $T_{SC}$ . For each of these basic models, we tested models that included heterogeneous migration rates across the genome (2M, i.e., islands resisting migration), and heterogeneous  $N_e$  across the genome (2N, a way to model linked selection) as described in Rougeux et al. (2017). We therefore had four possible variations of each of the basic model (e.g., SC, SC<sub>2M</sub>, SC<sub>2N</sub>, SC<sub>2M2N</sub>), for a total of eight basic divergence models.

Both stairway plots and *moments* analyses of the empirical data suggest (see Results) a demographic expansion 20–100 kya, and that the Baltic Sea population may also have undergone a recent bottleneck. If such demographic changes had happened at the time of split between the two populations, this would be well captured by the eight models described above which take into account a single change in  $N_e$  from  $N_{ANC}$  to  $N_1$  and  $N_2$  at time  $T_S$ . However, given that the timing and magnitude of the population expansion are very similar in all populations, another possibility is that the ancestral population underwent a demographic expansion prior to the split between the North Sea and the Baltic Sea. We



incorporated this hypothesis by extending the eight models described above to include an ancestral population expansion (AE, ancestral expansion models), a recent bottleneck followed by population growth in the Baltic Sea (B, bottleneck models) or both (AEB models). In the AE models, the ancestral population undergoes a demographic change at time  $T_{AE}$ , after which population size remains constant until time of split ( $T_S$ ). In the B models, the Baltic Sea population undergoes a bottleneck followed by population growth at time  $T_S$ . This scenario mimics a possible invasion of the Baltic Sea from a small founder population. For both IM and SC models, all possible combinations of heterogeneous  $N_e$  (2N models), heterogeneous migration rates (2M models), ancestral expansion (AE models), and bottlenecks (B models) were tested, yielding 16 variations of the IM and SC models and a total of 32 models tested. It should be noted that in 2N2M models, it is assumed that regions experiencing lower migration rates and regions experiencing lower effective population size do not overlap. This made convergence of the complex models easier, and it should not be problematic assuming the proportion of the genome experiencing lower migration rates and with lower  $N_e$  are relatively small (as is our case, see [supplementary table 3, Supplementary Material](#) online).

#### Model Optimization and Model Selection

With the exception of the SNM models, which has no free parameters, all one- and two-population models were optimized in five independent optimization routines, each consisting of five rounds of optimization using an approach similar to [Portik et al. \(2017\)](#). In the first round, we generated 30 sets of three-fold randomly perturbed parameters and for each ran a BFGS optimization setting max-iter to 30. In the second round, we chose for each model the replicate with the highest likelihood from the first round and generated 20 sets of three-fold randomly perturbed parameters, followed by the same optimization strategy. We repeated the same procedure for rounds three, four, and five, but using two-fold (round three) and one-fold (round four and five) perturbed parameters, respectively. In the final round, we also estimated 95% confidence intervals (CI) of the estimated parameters using the Fisher Information Matrix, as described by [Coffman et al. \(2016\)](#), and 95% CI of  $T_S$  and  $T_{SC}$  were estimated according to the rules of propagation of uncertainty. We selected the best replicate for each of the final models (three one-population models and 32 two-population models). We then calculated  $W_{AIC}$  as a relative measure of support for each model. Parameters in coalescent units were scaled based on estimates of  $N_{ANC}$  as outlined in *dadi's* manual.  $N_{ANC}$  was calculated as  $\Theta / (4L\mu)$ , where  $L$  is the total sequence length from which SNPs used in demographic analyses originated and  $\mu = 1 \times 10^{-8}$ . The total sequence length was calculated as  $S \times (V_U / V_T)$ , where  $S$  is the number of sites (variant and invariant) retained in ANGSD to calculate the jAFS,  $V_U$  is the number of unlinked SNPs retained for demographic modeling and  $V_T$  is total number of variants in the jAFS before linked SNPs were removed. Time parameters were scaled assuming a

generation time of 3.5 years to be directly comparable with a previous study on the same species ([Le Moan et al. 2019](#)).

The use of AIC to rank models and of the Fisher Information Matrix to estimate parameter uncertainty relies on the assumption that genetic data are independent. For RAD data, it is generally assumed that keeping one SNP per RAD locus is sufficient to satisfy this assumption. Nevertheless, we carried out further analyses to ensure that unaccounted linkage did not lead to 1) biased estimates of parameter uncertainty and 2) favoring more complex models. Firstly, we estimated parameter uncertainty for the two best models (that according to  $W_{AIC}$  had similar support) using block-bootstraps and the Godambe Information Matrix (GIM). Secondly, after the two best-fitting models had been identified (in this case,  $IM_{AEB2N}$  and  $SC_{AEB2N}$ ), we used LRT to test for the support of specific parameters among nested models, controlling for Type I error by normalizing the difference in log-likelihoods as outlined in [Coffman et al. \(2016\)](#). We used LRT to test support for an SC scenario by comparing  $IM_{AEB2N}$  and  $SC_{AEB2N}$  models. Since the test failed to reject the simpler IM scenario, we further used LRT to test for support for a past bottleneck in the Baltic Sea population (comparing  $IM_{2N}$  and  $IM_{B2N}$  models) and for an ancestral expansion (comparing  $IM_{2N}$  and  $IM_{AE2N}$  models). The correct weighting of  $\chi^2$  distributions for the LRT were calculated according to [Ota et al. \(2000\)](#).

#### Acknowledgments

We thank Roger Butlin, Ryan Gutenkunst, Paul Blischak, Xin Huang, Petri Kempainen, Bohao Fang, Pierre Nouhaud, and Sean Stankowski for providing feedback on the manuscript. We thank Einar Nielsen for donating turbot samples from the North Sea. This study was funded by the Academy of Finland (Grant Nos. 316294 to P.M. and 218343 to J.M.) and the Helsinki Institute of Life Science (HiLIFE to J.M.).

#### Author Contributions

P.M. conceived the study, performed laboratory work, analyzed the data, and wrote the manuscript. J.M. contributed to study design, provided constructive feedback on the manuscript, and funded sequencing. A.B.F. contributed most of the samples and provided constructive feedback on the manuscript.

#### Data Availability

All raw sequence data are publicly accessible via GenBank Sequence Read Archive (Bioproject ID PRJNA699405). Scripts and processed data are available at Zenodo data repository (doi: 10.5281/zenodo.4518372), and models for demographic inference are also available via GitHub (<https://github.com/Nopaoli/Demographic-Modelling>).

#### References

Alcala N, Jensen JD, Telenti A, Vuilleumier S. 2015. The genomic signature of population reconnection following isolation: from theory to HIV. *G3 (Bethesda)* 6(1):107–120.

- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025–2035.
- Becquet C, Przeworski M. 2009. Learning about modes of speciation by computational approaches. *Evolution* 63(10):2547–2562.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol.* 57(1):289–300.
- Bhaskar A, Song YS. 2014. Descartes' rule of signs and the identifiability of population demographic models from genomic variation data. *Ann Stat.* 42(6):2469–2493.
- Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Syst.* 38(1):459–487.
- Bourgeois Y, Ruggiero RP, Manthey JD, Boissinot S. 2019. Recent secondary contacts, linked selection, and variable recombination rates shape genomic diversity in the model species *Anolis carolinensis*. *Genome Biol Evol.* 11(7):2009–2022.
- Christe C, Stöltzing KN, Paris M, Frañsse C, Bierne N, Lexer C. 2017. Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Mol Ecol.* 26(1):59–76.
- Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN. 2016. Computationally efficient composite likelihood statistics for demographic inference. *Mol Biol Evol.* 33(2):591–593.
- Cruikshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* 23(13):3133–3157.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
- Delsler PM, Corrigan S, Duckett D, Suwalski A, Veuille M, Planes S, Naylor GJ, Mona S. 2019. Demographic inferences after a range expansion can be biased: the test case of the blacktip reef shark (*Carcharhinus melanopterus*). *Heredity* 122(6):759–769.
- Dixon GB, Davies SW, Aglyamova GV, Meyer E, Bay LK, Matz MV. 2015. Genomic determinants of coral heat tolerance across latitudes. *Science* 348(6242):1460–1462.
- Duranton M, Allal F, Fraïsse C, Bierne N, Bonhomme F, Gagnaire P-A. 2018. The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nat Commun.* 9(1):1–11.
- Duranton M, Allal F, Valière S, Bouchez O, Bonhomme F, Gagnaire P-A. 2020. The contribution of ancient admixture to reproductive isolation between European sea bass lineages. *Evol Lett.* 4(3):226–242.
- Ewing GB, Jensen JD. 2016. The consequences of not accounting for background selection in demographic inference. *Mol Ecol.* 25(1):135–141.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9(10):e1003905.
- Feng X, Merilä J, Löytynoja A. 2020. We shall meet again-genomics of historical admixture in the sea. *bioRxiv*. Available from: <https://doi.org/10.1101/2020.06.03.128298>.
- Figueras A, Robledo D, Corvelo A, Hermida M, Pereiro P, Rubiolo JA, Gómez-Garrido J, Carreté L, Bello X, Gut M, et al. 2016. Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA Res.* 23(3):181–192.
- Florin A-B, Höglund J. 2007. Absence of population structure of turbot (*Psetta maxima*) in the Baltic Sea. *Mol Ecol.* 16(1):115–126.
- Gagnaire P-A, Lamy J-B, Cornette F, Heurtebise S, Dégremont L, Flahauw E, Boudry P, Bierne N, Lapegue S. 2018. Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biol Evol.* 10(9):2518–2534.
- Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, Price AL. 2016. Fast principal-component analysis reveals convergent evolution of *adh1b* in Europe and East Asia. *Am J Hum Genet.* 98(3):456–472.
- Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet.* 11(2):e1005004.
- Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD, 1000 Genomes Project. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A.* 108(29):11983–11988.
- Gustafsson BG, Westman P. 2002. On the causes for salinity variations in the Baltic Sea during the last 8500 years. *Paleoceanography* 17(3):12–1–12-14.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5(10):e1000695.
- Hartmann FE, Rodríguez de la Vega RC, Gladieux P, Ma W-J, Hood ME, Giraud T. 2020. Higher gene flow in sex-related chromosomes than in autosomes during fungal divergence. *Mol Biol Evol.* 37(3):668–682.
- Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biol J Linn Soc.* 68(1–2):87–112.
- Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167(2):747–760.
- Hofer T, Foll M, Excoffier L. 2012. Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics* 13(1):107.
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18(2):337–338.
- Jacobs A, Carruthers M, Yurchenko A, Gordeeva NV, Alekseyev SS, Hooker O, Leong JS, Minkley DR, Rondeau EB, Koop BF, et al. 2020. Parallelism in eco-morphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. *PLoS Genet.* 16(4):e1008658.
- Jay F, Boitard S, Austerlitz F. 2019. An ABC method for whole-genome sequence data: inferring paleolithic and neolithic human expansions. *Mol Biol Evol.* 36(7):1565–1579.
- Johannesson K, Le Moan A, Perini S, André C. 2020. A Darwinian laboratory of multiple contact zones. *Trends Ecol Evol.* 35(11):1021–1036.
- Jokinen H, Momigliano P, Merilä J. 2019. From ecology to genetics and back: the tale of two flounder species in the Baltic Sea. *ICES J Mar Sci.* 76(7):2267–2275.
- Jouganous J, Long W, Ragsdale AP, Gravel S. 2017. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics* 206(3):1549–1567.
- Kimura M, Ohta T. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61(3):763–771.
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15(1):356.
- Lande R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. *Am Nat.* 116(4):463–479.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 9(4):357–359.
- 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*. Available from: <https://doi.org/10.1101/662569>.
- Le Moan A, Gagnaire PA, Bonhomme F. 2016. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol Ecol.* 25(13):3187–3202.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Liu S, Hansen MM, Jacobsen MM. 2016. Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Mol Ecol.* 25(20):5187–5202.

- Liu X, Fu Y-X. 2015. Exploring population size changes using SNP frequency spectra. *Nat Genet.* 47(5):555–559.
- Luikart G, Allendorf F, Cornuet J, Sherwin W. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J Hered.* 89(3):238–247.
- Marques DA, Lucek K, Meier JJ, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet.* 12(2):e1005887.
- Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210(2):719–731.
- Momigliano P, Denys GPJ, Jokinen H, Merilä J. 2018. *Platichthys solemdali* sp. nov. (Actinopterygii, Pleuronectiformes): a new flounder species from the Baltic Sea. *Front Mar Sci.* 5:225.
- Momigliano P, Jokinen H, Fraimout A, Florin A-B, Norkko A, Merilä J. 2017. Extraordinarily rapid speciation in a marine fish. *Proc Natl Acad Sci U S A.* 114(23):6074–6079.
- Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, Moodley Y, Jensen JD. 2015. Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. *Genetics* 200(3):947–963.
- Myers S, Fefferman C, Patterson N. 2008. Can one learn history from the allelic spectrum? *Theor Popul Biol.* 73(3):342–348.
- Nelson TC, Cresko WA. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evol Lett.* 2(1):9–21.
- Nielsen EE, Nielsen PH, Meldrup D, Hansen MM. 2004. Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. *Mol Ecol.* 13(3):585–595.
- Nosil P. 2008. Speciation with gene flow could be common. *Mol Ecol.* 17(9):2103–2106.
- Nosil P. 2012. Ecological speciation. New York: Oxford University Press.
- Nosil P, Feder JL. 2012. Genomic divergence during speciation: causes and consequences. *Philos Trans R Soc Lond B Biol Sci.* 367(1587):332–342.
- Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol.* 24(3):145–156.
- Noskova E, Ulyantsev V, Klaus-Peter K, Stephen J, Dobrynin P. 2020. GADMA: genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. *GigaScience* 9(3):giaa005.
- Ota R, Waddell PJ, Hasegawa M, Shimodaira H, Kishino H. 2000. Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. *Mol Biol Evol.* 17(5):798–803.
- Portik DM, Leaché AD, Rivera D, Barej MF, Burger M, Hirschfeld M, Rödel M-O, Blackburn DC, Fujita MK. 2017. Evaluating mechanisms of diversification in a Guineo-Congolian tropical forest frog using demographic model selection. *Mol Ecol.* 26(19):5245–5263.
- Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife* 7:e36317.
- Quinlan AR, Hall IM. 2010. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.
- Raj A, Stephens M, Pritchard JK. 2014. faststructure: variational inference of population structure in large SNP data sets. *Genetics* 197(2):573–589.
- Ravinet M, Faria R, Butlin R, Galindo J, Bierne N, Rafajlović M, Noor M, Mehlig B, Westram A. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol.* 30(8):1450–1477.
- Rosen Z, Bhaskar A, Roch S, Song YS. 2018. Geometry of the sample frequency spectrum and the perils of demographic inference. *Genetics* 210(2):665–682.
- Rougemont Q, Bernatchez L. 2018. The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations. *Evolution* 72(6):1261–1277.
- Rougemont Q, Gagnaire P, Perrier C, Genthon C, Besnard A, Launey S, Evanno G. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Mol Ecol.* 26(1):142–162.
- Rougemont Q, Moore J-S, Leroy T, Normandeau E, Rondeau EB, Withler RE, Van Doornik DM, Crane PA, Naish KA, Garza JC, et al. 2020. Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific salmon. *PLoS Genet.* 16(8):e1008348.
- Rougeux C, Bernatchez L, Gagnaire P-A. 2017. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biol Evol.* 9(8):2057–2074.
- Rougeux C, Gagnaire P-A, Bernatchez L. 2019. Model-based demographic inference of introgression history in European whitefish species pairs. *J Evol Biol.* 32(8):806–817.
- Roux C, Fraïsse C, Castric V, Vekemans X, Pogson GH, Bierne N. 2014. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *J Evol Biol.* 27(8):1662–1675.
- Roux C, Fraïsse C, Romiguier J, Anciaux Y, Galtier N, Bierne N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* 14(12):e2000234.
- Roux C, Tsagkogeorga G, Bierne N, Galtier N. 2013. Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Mol Biol Evol.* 30(7):1574–1587.
- Schridder DR, Shanku AG, Kern AD. 2016. Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 204(3):1207–1223.
- Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.* 23(1):23–35.
- Stuglik MT, Babik W. 2016. Genomic heterogeneity of historical gene flow between two species of newts inferred from transcriptome data. *Ecol Evol.* 6(13):4513–4525.
- Tavares H, Whibley A, Field DL, Bradley D, Couchman M, Copsey L, Elleouet J, Burrus M, Andalo C, Li M, et al. 2018. Selection and gene flow shape genomic islands that control floral guides. *Proc Natl Acad Sci U S A.* 115(43):11006–11011.
- Tine M, Kuhl H, Gagnaire P-A, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun.* 5:5770.
- Van Belleghem SM, Vangestel C, De Wolf K, De Corte Z, Möst M, Rastas P, De Meester L, Hendrickx F. 2018. Evolution at two time frames: polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genet.* 14(11):e1007796.
- Wallace AR. 1855. On the law which has regulated the introduction of new species. *Ann Mag Nat Hist.* 15:185–196.
- Wang S, Meyer E, McKay JK, Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods.* 9(8):808–810.
- Wu C-I. 2001. The genic view of the process of speciation. *J Evol Biol.* 14(6):851–865.