Biases in demographic modelling affect our understanding of the process of speciation

Paolo Momigliano[1,✉], **Ann-Britt Florin**[2], **and Juha Merilä**[1]

[1]Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland
[2]Department of Aquatic Resources, Institute of Coastal Research, Swedish University of Agricultural Sciences, Öregrund, Swedend

Estimating patterns of gene flow during the early stages of speciation is central to understanding whether reproductive isolation arises via the gradual erosion of gene flow or through successive stages of strict isolation and secondary contact. Such scenarios can be explicitly tested by comparing the joint allele frequency spectrum (jAFS) of a set of populations to jAFS simulated under scenarios of isolation with migration (IM) and secondary contact (SC). However, the potential effect of unaccounted demographic events (such as population expansions and bottlenecks) on model choice and parameter estimation remains largely unexplored. Using simulations, we demonstrate that under realistic scenarios of population divergence with constant gene flow, failure to account for population size changes in either a daughter population or the ancestral population leads to overestimated divergence time and to a bias towards the choice of SC models. On the other hand, when the simulations included long periods of strict isolation the correct gene flow scenario was usually retrieved. We illustrate these issues reconstructing the demographic history of North Sea and Baltic Sea turbots (*Schopthalmus maximus*) by testing 16 IM and 16 SC scenarios, modelling changes in effective population sizes as well as the effects of linked selection and heterogeneous migration rates across the genome. As in the simulated data, failure to account for changes in effective population sizes resulted in selecting SC models with a long period of strict isolation and divergence times preceding the formation of the Baltic Sea. In contrast, models accounting for population size changes suggest that the Baltic Sea turbot population originated from a very recent (<6 kya) invasion and has diverged with constant gene flow from the North Sea. The results have broad implications for the study of speciation by high-lighting the potential effects of ancestral size changes and bottlenecks followed by growth on choices between competing scenarios of speciation. In general, extreme caution should be exercised when interpreting results of demographic model comparisons.

Speciation | Demographic modelling | Allele Frequency Spectrum | Secondary Contact | Isolation with Migration

Correspondence: ***paolo.momigliano@helsinki.fi***

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*" (1), 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 (2). The study of population pairs where reproductive isolation is incomplete—the "grey zone" of speciation *sensu* Roux *et al* (3)—can therefore provide insight into the processes leading to the evolution of different species. According to the genic view of speciation (4), in the early stages of speciation 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 of higher differentiation due to selection, while unimpeded gene-flow homogenizes the rest of the genome (3, 5–7). Partial reproductive isolation could arise via the gradual erosion of gene flow (primary divergence), for example because of multifarious selection across environments (6, 8, 9). 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 (e.g. 10–13). Both processes can lead to nearly identical 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 (7). Therefore, distinguishing whether divergence initiated in the presence or absence of gene flow, while central to our understanding of the process of speciation, is not a trivial task.

Demographic modelling provides a framework to reconstruct how gene flow has changed through the evolutionary history of incipient species. Within the past two decades, several computational methods have been developed to reconstruct the demographic history of population divergence from genomic data. Such approaches usually rely on comparing summary statistics obtained from empirical data to simulations performed under competing divergence models, of which the most commonly tested ones include isolation with continuous migration (IM), secondary contact (SC), strict isolation (SI) and ancient migration (AM) (10). Approximate Bayesian computation (ABC) approaches (14, 15), as well as methods based on the diffusion approximation of the joint site frequency spectrum (jAFS) (DaDi 16) or its direct computation using a model of ordinary differential equations (*moments* 17), have been broadly applied to test among competing demographic models of divergence. The models usually

assume that an ancestral population of size N_{REF} 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 modelling the effect of barrier loci—modelled as heterogeneous migration rates across the genome (10, 18)—as well as the effect of selective sweeps and background selection—modelled as heterogeneous effective population size (N_e) across the genome (3, 19). Modelling heterogeneous N_e and migration rates is important, as ignoring them may lead to biases in model choice (3, 11).

Recent studies of several population pairs in the "grey zone" of speciation have revealed that in most cases the speciation process started during long periods of strict isolation, and that contemporary heterogenous gene flow is a result of recent secondary contact (e.g. 3, 10, 11, 13, 18–21). These studies provide support to the hypothesis that initial allopatric phases of differentiation play a central role in the evolution of reproductive isolation (11), and hence primary divergence with gene flow due to ecological selection is much rarer than suggested by some authors (8). However, while these models can provide great insight into the demographic history of incipient species, they also show significant limitations. If the models tested are not close enough to the real divergence scenario, both model choice (for example, the choice bewteen a SC and IM model) and parameter estimation may be strongly affected. While great effort has been placed recentely to overcome potential biases due to barrier loci and linked selection (3, 10, 13, 18–22), it remains unclear how unmodelled 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 10). If, for example, there was an unmodelled 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 towards the time of ancestral population size change. Similarly, unmodelled bottlenecks followed by exponential growth in one of the daughter populations may lead to overestimation of T_S, as small populations experience faster genetic drift. It is unclear how changes in N_e in ancestral and daughter populations affect model choice, as to the best of our knowledge, no study yet addressed this question. This 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.

Here, we used both simulations and empirical data to demonstrate that unmodelled demographic events in both ancestral and daughter populations can strongly affect both model choice and parameter estimation. Using coalescent simulation 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 overestimation of T_S and to a strong bias towards the choice of SC models. We then reconstruct the demographic history of the Atlantic and Baltic Sea populations of the turbot *Schoptalmus maximus*, which a recent study (23) suggested have diverged at the start of the last glacial maximum (>50 kya) and experienced secondary contact following the end of the last glaciation. We demonstrate that these inferences were biased because of the failure to account for a past demographic expansion (which leads to overestimate T_S), and because of the failure to model a bottleneck during the invasion of the Baltic Sea (leading to the erroneous choice of a SC model). We discuss the potential implications of our findings for speciation studies of non-model organisms in general.

Methods

Analyses of simulated data.

Simulations. To test the potential effect of unmodelled demographic events on model choice and parameter estimation, we used the software *ms* (24) to simulate the divergence of two populations under four gene-flow scenarios: SC and IM, each with symmetric ($N_e m$ = 1) and asymmetric gene flow ($N_e m_{12}$=0.25 and $N_e m_{21}$=1, where $N_e m_{12}$ is the absolute number of individuals migrating in a given generation from population one to population two). For each of these gene flow scenarios, we modelled 64 combinations of demographic events in a fully orthogonal design (Fig. 1A). The simplest models represent IM and SC scenarios where an ancestral population of size N_{REF} 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_0) and N_2 ($N_2 = 0.25 \times N_1$, i.e. 5000 individuals). We then included seven scenarios of population expansion at time T_{AE} (T_{AE}=0.5, in units of $4 \times N_1$ 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_0) for parameter scaling is always N_1, ancestral expansions were modelled 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 modelled 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 2000 generations following T_S, after which the N_e of population two reached N_2. Bottlenecks followed by growth in one of the daughter populations may 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 4000 generations ago ($T_S = 0.05 \times 4N_0$). In SC models, the time of divergence is set at 8000 generations ago ($T_S = 0.1 \times 4N_0$), while secondary contact is established at time T_{SC} ($T_{SC} = 0.25 \times T_S$), i.e. 2000 generations ago).
Fully orthogonal combinations of all the demographic scenarios outlined above (SC and IM, symmetric and asymmet-

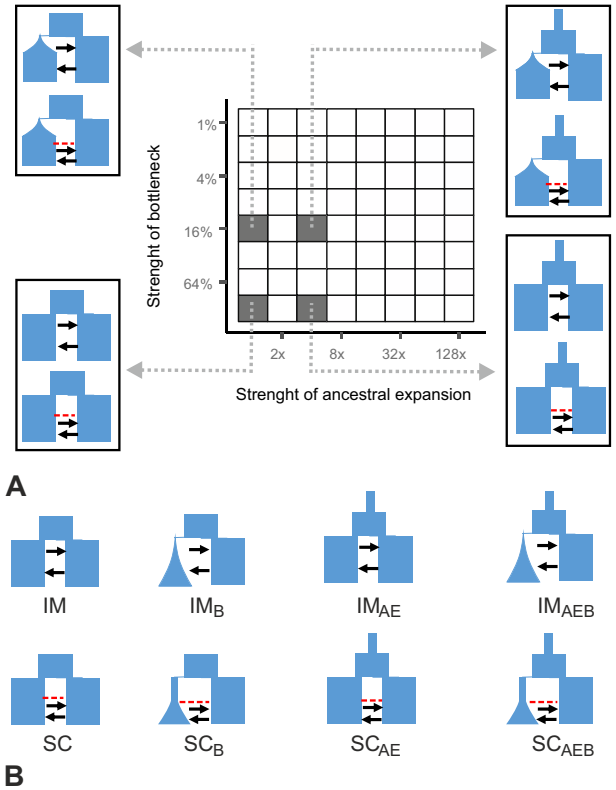


Fig. 1. Schematic representation of the coalescent simulations **A** and models used for demographic inference **B**. For both the smaller (100 000 loci) and larger (1 million loci) data sets, we simulated data under the IM and SC scenarios with 64 possible combinations of ancestral expansion and bottlenecks followed by growth. All simulations were ran both under symmetric and asymmetric migration scenarios, giving a total of 256 demographic scenarios. In panel **B** are shown the demographic models used for inference, i.e. the basic IM and SC models as well as modifications that include ancestral population size changes as well as bottlenecks followed by growth in one of the daughter populations. Notice that IM$_B$, SC$_B$, IM$_{AEB}$ and SC$_{AEB}$ models do not match exactly coalescent simulations.

ric 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 populations and of 1 million unlinked loci of a length of 36 bp, which, when using a standard germ-line mutation rate (μ) of 1×10^{-8} and a N_0 of 20 000 individuals, results in roughly 80-120 thousand unlinked SNPs (depending on the specific model). We then estimated the folded and unfolded jAFS from the simulated data. These data sets represent a standard scenario when working with high quality whole genome data, assuming both scenarios whereby the genome of a closely related species is not and is available to polarize the jAFS. Secondly, for each of the 256 scenarios we simulated sampling of 10 individuals per populations and of 100 000 unlinked loci of a length of 36 bp, resulting in roughly 8-12 thousand 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.

Demographic modelling of simulated data. Demographic modelling of simulated data was carried out using the software package *moments* (17), which is a development of the DaDi (16) 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 many of its limitations (17). For each of the 256 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. 1B). The models consist of an ancestral population of size N_{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_{SC}. For each of these basic 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. 1B). 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 modelled in previously utilized models in speciation research (e.g. 12). In IM$_B$ models exponential growth starts at time T_S and continues until the present while in SC$_B$ models exponential growth starts at time T_{SC} and continues until the present (Fig. 1 B).

Models were optimized for three rounds following an approach similar to Portik et al (25). In the first round, a Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm optimization (function "optimize.log"; max-iter=10) was run for 10 sets of threefold 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 10 sets of twofold randomly perturbed parameters, increasing max-iter to 20. We repeated the same procedure for round 3, but using onefold perturbed parameters and setting max-iter to 30. We ran the entire optimization procedure 5 times to check for convergence among independent optimizations. We selected the replicate with the highest likelihood for each of eight two-populations models and calculated the Akaike Information Criterion as $-2(log-likelihood)+2K$ where K is the number of model parameters, and the $\Delta AIC_i = AIC_i - AIC_{min}$. We then calculated Akaike weight of evidence (W_{AIC}) as outlined in Rougeux et al (19). The equation is oulined below, and R represents the total number of models conpared.

$$W_{AIC} = \frac{e^{\frac{(-\Delta AIC_i)}{2}}}{\sum_{i=1}^{R} e^{\frac{(-\Delta AIC_i)}{2}}}$$

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 or not accounting for ancestral expansion and bottlenecks affected parameter estimation. Second, we looked at whether including models with ancestral population expansions and bottlenecks improved model choice and parameter estimation. When comparing the 8 full models, we calculated the W_{AIC} for each gene flow scenario (IM vs SC) as the sum of W_{AIC} for all IM and SC models. For all 512 simulations, model optimization and parameter estimation were carried out using both the folded and unfolded jAFS, to test whether the unpolarized and polarized jAFS are less 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. For these analyses, we performed model choice for the simulations of 1 million loci under the IM scenario and both symmetric and asymmetric gene flow using the IM and SC basic models only. We used exactly the same optimization strategy, but only ran 3 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 *Scophtalmus maximus* from seven locations and three biogeographic regions (Supplementary Table 1 and Fig. 5A): the North Sea (1 location, N=20), the transition zone separating the North Sea from the Baltic Sea (2 sampling locations, Vendelsö: N=27 and Öresund: N=35), and the Baltic Sea (4 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 (26), while samples from the North Sea were originally collected by Nielsen et al (27).

Library preparation. We built 2b-RAD libraries following the approach described by Wang et al (28), but with degenerate adaptors to allow the identification of PCR duplicates. The protocol is described in detail by Momigliano et al (29). 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 BcgI (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 10 cycles of PCR as described in Momigliano et al (29). Fragments of the expected size were isolated using a PluePippin machine (Sage Science). Libraries were sequenced on Illumina machines (NextSeq500 and Hiseq 4000) to achieve a mean coverage of approximately 20x.

Bioinformatics. Raw reads were demultiplexed and PCR duplicates were removed as per Momigliano et al (29) , then mapped to the latest version of *S. maximus* reference genome (30, Assembly ASM318616v1, GenBank accession: GCA003186165) using Bowtie2 (31). SAM files were converted to BAM files and indexed using SAMTOOLS (32). A genotype likelihood file in beagle format was produced in the software ANGSD (33), 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 Fig. 1 for summary statistics. A principal component analysis (PCA) based on genotype likelihoods was performed using the software PCAngds (34) 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 jAFSs where the minor allele is computed separately for each AFS while *moments* expects minor alleles to be estimate for the jAFS. Thus, for the jAFS for NS and BS, 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*. Bioinformatic steps, scripts for analyses, and the jAFS's used are publicly available (see Data Access section).

Modelling 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 one-dimensional (for one-population models) and joint folded allele frequency spectra (henceforth 1d-AFS and jAFS, respectively). 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 (35), we first determined the demographic history of each population independently using 1d-AFS, using both *moments* and Stairway plots (36). We then proceeded to compare 24 two-population models to determine the demographic history from the jAFS We used a μ of (1×10^{-8}) and a generation time of 3.5 years for scaling demographic events to make results directly comparable to Le Moan et al (23). All models tested in this study are available from GitHub.

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 (36). Stairway plots use composite likelihood estimations of Θ at different epochs, which is then scaled using the mutation rate. For Baltic Sea populations, we estimated past demographic changes from the 1d-SFS inferred both from pooling all samples from different locations (since there is no sign of population structure) and 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 times 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 1-population models using *moments*: a Standard Neutral Model (SNM, assuming constant population size at equilibrium), a 1 Epoch model (1EP, assuming a sudden population change at time T1) and a 2 Epochs model including a sudden demographic change at time T_{AE} followed by a bottleneck at time T_B followed by exponential growth ($2EP_B$). The 1EP model represents a single demographic change and a scenario where a demographic expansion/contraction occurred either in the ancestral population from which the NS and BS populations are derived or in the NS and BS populations themselves. The $2EP_B$ model represent 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 ky.

Two-population models. Given the existence of a well-known hybrid-zone between the Baltic Sea and the North Sea (27), 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_{REF} 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 (12). We therefore had four possible variations of each of the basic model (for example: SC, SC_{2M}, SC_{2N}, SC_{2M2N}), for a total of eight basic divergence models.

Both Stairway Plots and *moments* analyses of the empirical data suggest (see Results) a demographic expansions 20-100 kya, and that the BS 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_{REF} to N_1 and N_2 at the 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 remain 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.

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 (25). In the first round we generated 30 sets of threefold 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 threefold randomly perturbed parameters, followed by the same optimization strategy. We repeated the same procedure for round three,four and five, but using twofold (round three) and onefold (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 (37), 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_{REF} as outlined in DaDi's manual. N_{REF} 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 modelling 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 (23).

Results

Analyses of simulated data. Analyses based on simulations of 1 million or 100 000 loci gave similar results (compare Figs. 2 & 4 with Supplementary Figs. 15 & 16), 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, while the results for the smaller numbers of loci are presented in the supplementary materials. Models generally converged very well with the optimization scheme utilized (Supplementary Figs. 2-9).

Biases in IM and SC models. Parameter estimation and model choice from basic IM and SC models were severely biased

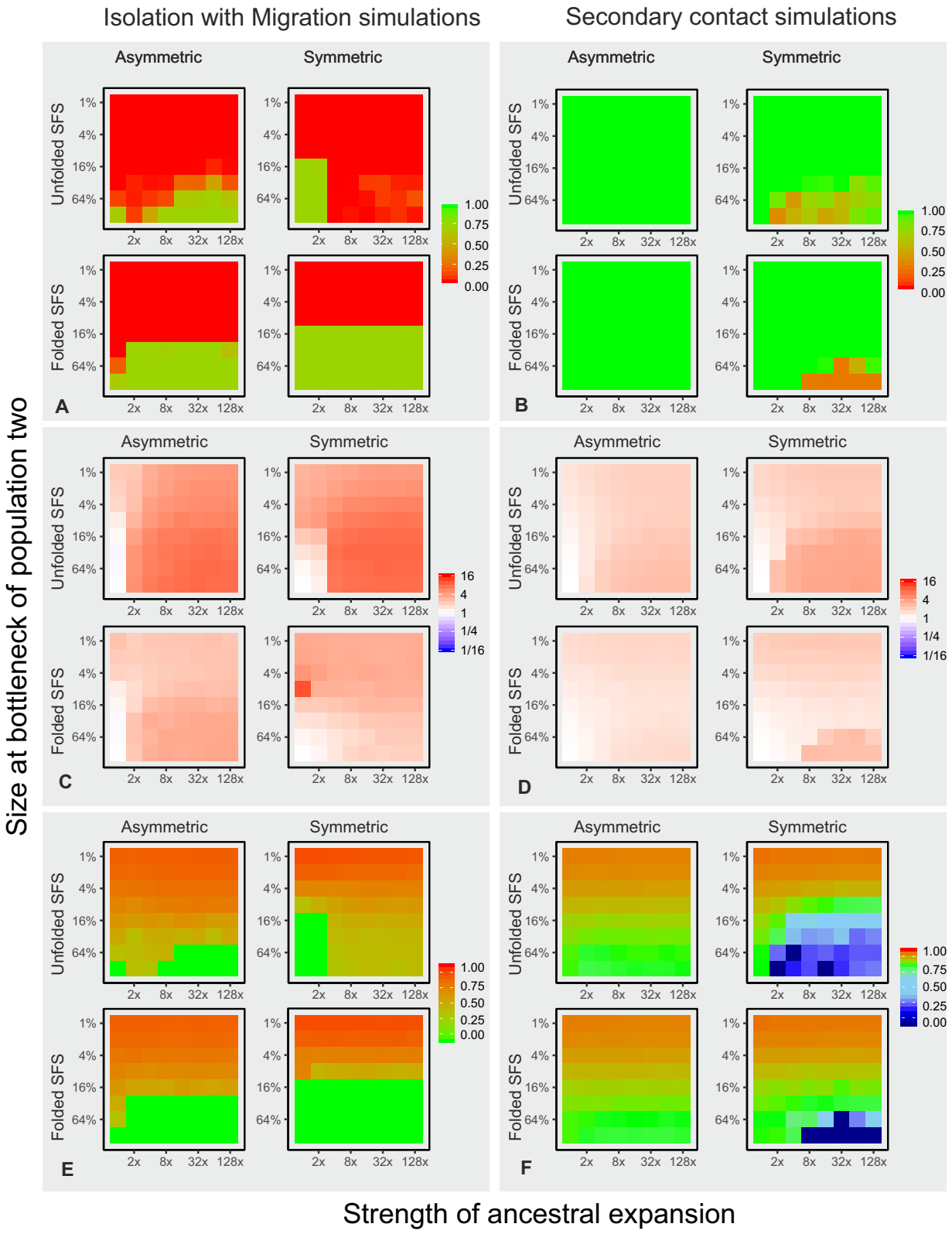


Fig. 2. Model choice and parameter misestimation for the simple IM and SC models for all simulations of the larger data sets (1 million loci). Left panels (**A**, **C**, **E**) show results from simulations with constant migration (IM), right panels (**B**, **D**, **F**) show results from simulation with a period of strict isolation (SC). Within each panel, results are shown for simulations with symmetric and asymmetric migration, and for model estimates using the folded or unfolded jAFS. Within each panel, each graph represent the values for all 64 simulations as per Fig. 1. Panel **A** and **B** show weight of evidence for the correct gene flow scenario (0-1). Panel **C** and **D** show misestemation 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).

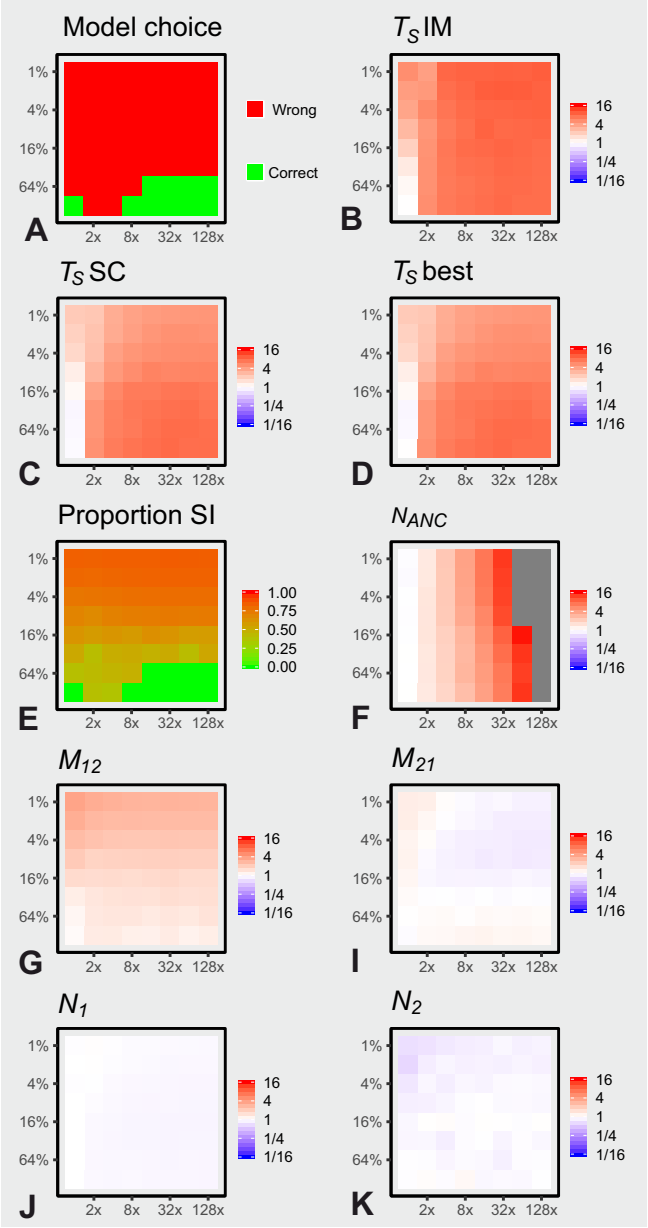


Fig. 3. Model choice and parameter estimation of simple IM and SC models from IM scenarios with asymmetric migration estimated using the unfolded jAFS. **A** cells are coded either as red or green depending on whether the correct model yielded the lowest AIC. **B** misestimation of T_S from the IM models **C** Misestimation of T_S from SC models. **D** Misestimation of T_S by the model with lowest AIC. **E** estimated proportion of the divergence time for which the model inferred strict isolation (correct is 0, shown in green). **F**, **G**, **I**, **J** and **K** show, respectively, misestimation of N_{ANC}, M_{12}, M_{21}, N_1 and N_2 by the model with lowest AIC. Squares in grey represent values off the scale.

when the demographic scenario of the simulations deviated from the models. While for SC simulations the underlying gene flow scenario was almost always identified (Fig. 2 B), when the true divergence scenario of the simulations had continuous gene flow even moderate unmodelled bottlenecks resulted in strong support for SC (Fig. 2 A). Stronger bottlenecks were associated with an overestimation of divergence times, and the bias was much more severe for simulations with constant gene flow (IM; Fig. 2 C&D, Supplementary Fig. 10 A&B). There was also a clear relationship between the length of the estimated period of strict isolation (SI) in IM simulations and the strength of the unmodelled bottleneck (Fig. 2 E). Hence, unmodelled bottlenecks generally led to choice of SC models with long periods of strict isolation even when the simulation scenario had constant gene flow. Unmodelled size changes in the ancestral population led to severe overestimates of divergence time, particularly in IM scenarios (Fig. 2 B) and to a lesser degree in SC scenarios. The stronger the unmodelled ancestral expansion, the more estimates of T_S approached the true value of T_{AE} (which is $10 \times T_S$ under IM scenarios, $5 \times T_S$ under SC scenarios, see methods; Fig. 2 C&D, Fig. 3 B&C, Supplementary Fig. 10 A&B). Unmodelled ancestral expansions also biased model choice towards SC when the true scenario had constant migration, but only when the unfolded jAFS were used for demographic inference. However, when an unmodelled ancestral expansion led to incorrectly choose the SC model, the proportion of time of SI inferred was low (< 50% of the total divergence time) unless the simulation included also a strong bottleneck (Supplementary Fig. 11 A&B). Interestingly, in SC simulations without strong bottlenecks, an unmodelled ancestral expansion led in some cases to a bias towards IM models (Fig. 2 B), and in general to an underestimate of the proportion of SI (Supplementary Fig. 11 A&B). Not surprisingly, unmodelled ancestral expansions led also to strong overestimates of N_{ANC} (Fig. 3F and Supplementary Fig. 12).

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 to the folded jAFS (Fig. 2 C&D, Supplementary Fig. 10 A&B). Similarly, bias in model choice towards SC models was stronger when using the unfolded jAFS (Fig. 2A and Fig. 4, Supplementary Fig. 10). It should be noted, however, that estimates of contemporary N_e and migration rates were always fairly accurate (see Fig. 3 and Supplementary Figs. 13 & 14). Furthermore, in general these biases were less severe when the smaller data set was used for inference (Supplementary Fig. 15). 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. 17).

Biases in 8-model comparisons. 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

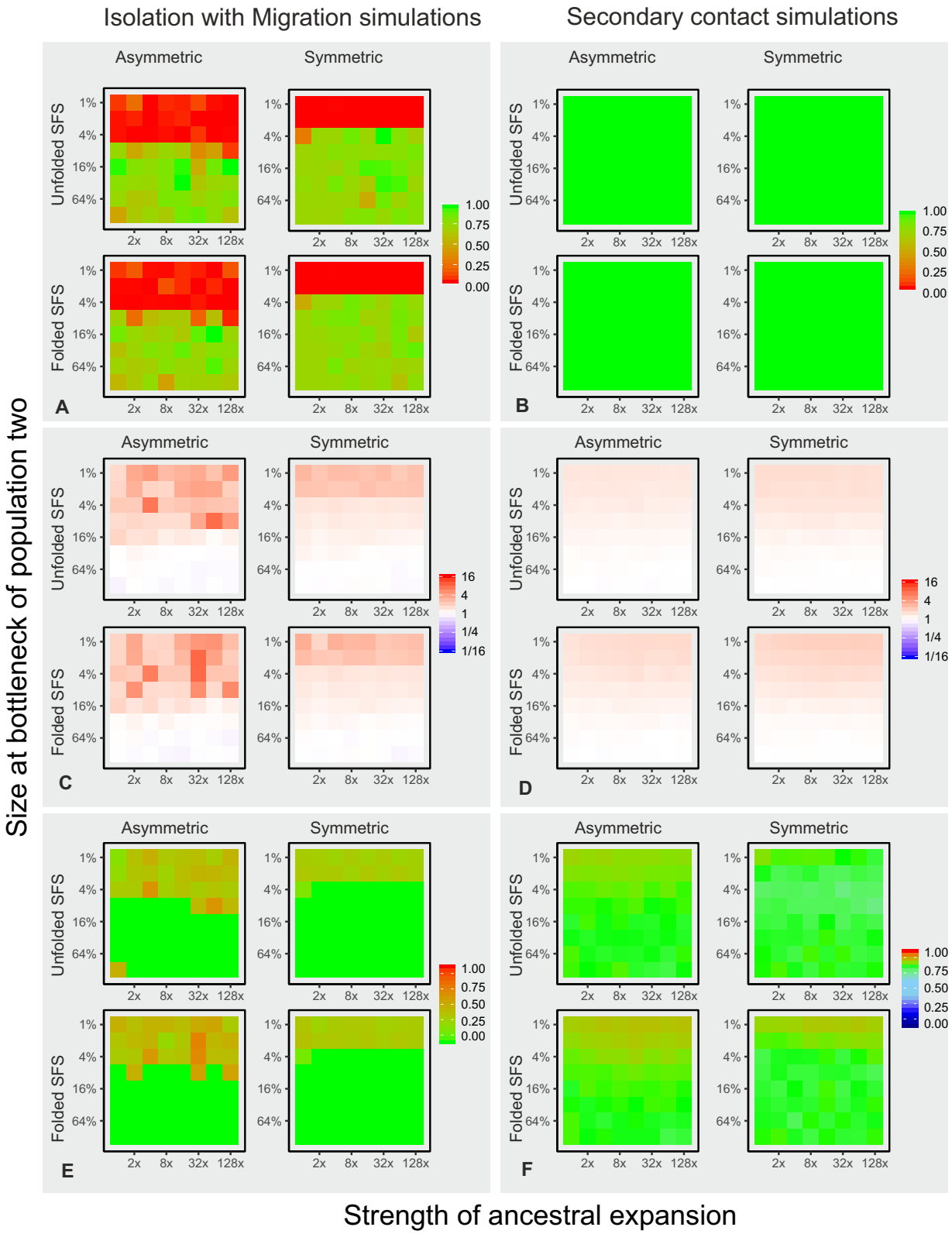


Fig. 4. Model choice and parameter misestimation for all eight divergence models for all simulations of the larger data sets (1 million loci). Left panels **A, C & C** show results from simulations with constant migration (IM), right panels **B, D & F** show results from simulation with a period of strict isolation (SC). Within each panel, results are shown for simulations with symmetric and asymmetric migration, and for model estimates using the folded or unfolded jAFS. Within each panel, each graph represent the values for all 64 simulations as per Fig. 1. Panel **A** and **B** show weight of evidence for the correct gene flow scenario (0-1). **C & D** show misestemation 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 represent the correct time, i.e. 0 for IM model and 0.75 for SC models).

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 (Fig. 4 A). 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 towards the same demographic history (Fig. 4C, Supplementary Fig. 11 C&D). While ancestral expansions and bottlenecks still caused a systematic bias towards overestimating divergence times, this bias was negligible (Fig. 4B, Supplementary Figs. 10 C&D).

Analyses of empirical data. A principal component analysis based on genotype likelihoods from 9063 biallelic SNPs with MAF > 0.02 shows a clear partitioning along the 2nd PC between the North Sea and Baltic Sea individuals (Figs. 5 A&B), with several individuals from the transition zone (locations VD and OS in Fig. 5 B) showing intermediate genotypes. Based on these results, individuals from admixed populations (VD and OS) where excluded from further analyses. Stairway plots revealed that all populations shared a similar demographic history until 10-20 kya, with evidence of a population expansion between 20 and 100 kya (Fig. 5 C). Populations from the Baltic Sea (GDY, DAB and GOT) show signs of population contraction followed by growth following the end of the last glaciation, and lower contemporary N_e compared to NS. One-population models based on all individuals from the NS and BS separately give strong support for population size change, rejecting the SNM. In the North Sea , the 1EP model had stronger support than the $2EP_B$ model (W_{AIC} of 0.88 and 0.12, respectively) while for the BS 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 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 BS should be formally tested in two-population models.

The folded jAFS used for testing two-population models included 16270 biallelic SNPs. Model choice and parameter estimates in two-population models were strongly affected by the inclusion or exclusion of ancestral expansions and bottlenecks (Fig. 6). In general, both in SC and IM models, scenarios including ancestral population expansions (AE) and bottlenecks (B) had lower AIC than simple IM and SC models, with AEB models performing best (Fig. 6 A). 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 Fig. 6 E we show the AIC of IM and SC models (and their AE, B, and

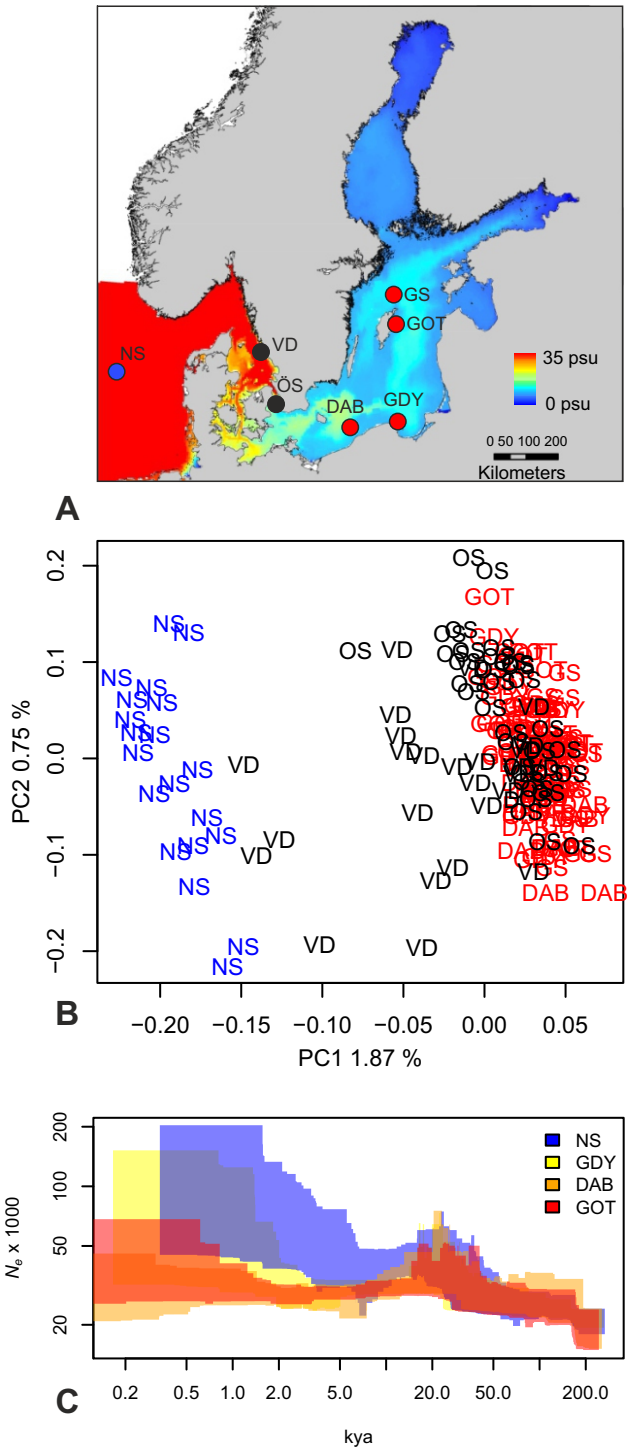


Fig. 5. **A** Map of sampling locations of turbots (*Scophthalmus maximus*) showing the modelled bottom salinity of the Baltic Sea. **B** PCA performed from genotype likelihoods of 9063 bi-allelic SNPs, color and population codes as per **A**. **C** Changes in N_e across time inferred for four populations (NS and three BS populations) via Stairway Plots. The polygons represent 95% confidence intervals.

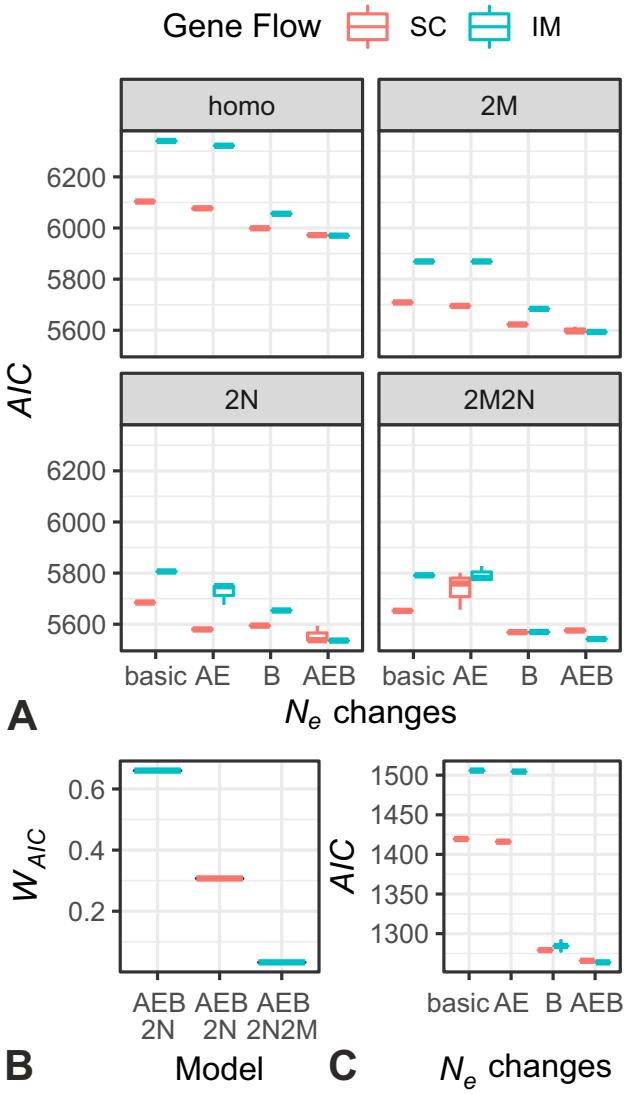


Fig. 6. Results from the 32 models optimized for the *S. maximus* data (**A** & **B**) and data simulated with similar parameters (**C**). **A** shows AIC from the best 3 replicates of SC (red boxplots) and IM (blue boxplots) models for any combination of ancestral expansion/bottleneck (AE, B and AEB models). Top right panel (Homo): homogenous gene flow and N_e. Panel 2M: heterogeneous migration rates across the genome. Panel 2N: heterogenous 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

AEB variations) for a simulation with an ancestral expansion ($N_{AE} = 2 \times N_{REF}$) and bottleneck (where N_2 at time of split is 4% of current N_2). Models accounting for heterogenous migration rates (2M) and N_e (2N) across the genome fitted the data better (Fig. 6A), 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_{AEB2N} and SC_{AEB2N}, with W_{AIC} of 0.66 and 0.31 (Fig. 6C), respectively. However, these two models converged to approximately the same scenarios (Fig. 7). Both models suggest an ancestral population expansion approximately 30-100 kya, and a colonization of the Baltic Sea < 5.5 kya. The estimate of the period of strict isolation (T_{SI}) from the SC model is very small (640 years) and its confidence intervals estimated using the Fisher Information Matrix include zero (Fig. 7G), again suggesting there is no clear evidence for a period of strict isolation.

Discussion

Reconstructing the demographic history of diverging populations is of central importance to understanding 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, unmodelled expansions and bottlenecks can lead to greatly overestimate divergence times and to favor scenarios that include long periods of strict isolation. These biases can be minimized by comparing more realistic models, but a small systematic bias towards overestimation of divergence times and choice of secondary contact models always remained, possibly because intentionally our demographic inference models do not match exactly our coalescent simulations. Using data from turbot populations from the Baltic Sea and North Sea, we demonstrate that a previous study (23) that did not take potential demographic changes in both ancestral and daughter populations into account, likely overestimated divergence times (ca. 50 kya) by an order of magnitude, and concluded that these populations diverged during a long allopatric phase whereas our results indicate a very recent divergence (<6 kya).

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 unmodelled demographic events. Surprisingly, until now this assumption was never formally tested, though it has been shown that in the original IM program (38) departures from assumptions can lead to strong biases in parameter estimation (39). Using extensive simulations, we demonstrate that under a very broad range of realistic divergence scenarios this assumption does not hold. One clear pattern emerging from our simulations is that failure to model a change in N_e in the ancestral population leads to overestimate divergence time. 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 unmodelled ancestral expansion or contraction will push estimates of T_S back to the time of the ancestral change in N_e. Exactly as expected, for our IM coalescent simulations that included an ancestral expansion, demographic models that did not account for ancestral expansion tended to overestimate T_S by up to a factor of ten (as $T_{AE} = 0.5$, i.e. $10 \times T_S$). In our SC coalescent simulations T_S was set to 0.1 and T_{AE} to 0.5, and as expected, the overestimation of T_S was up to a factor of five. In IM simulations, unmodelled ancestral expansions also led to a bias towards SC models, albeit this bias was observed only when demographic modelling was carried out using the unfolded jAFS.

When the real divergence scenario includes a long period of strict isolation as well as an ancestral population expansion, there may be a slight bias towards choosing IM models if the

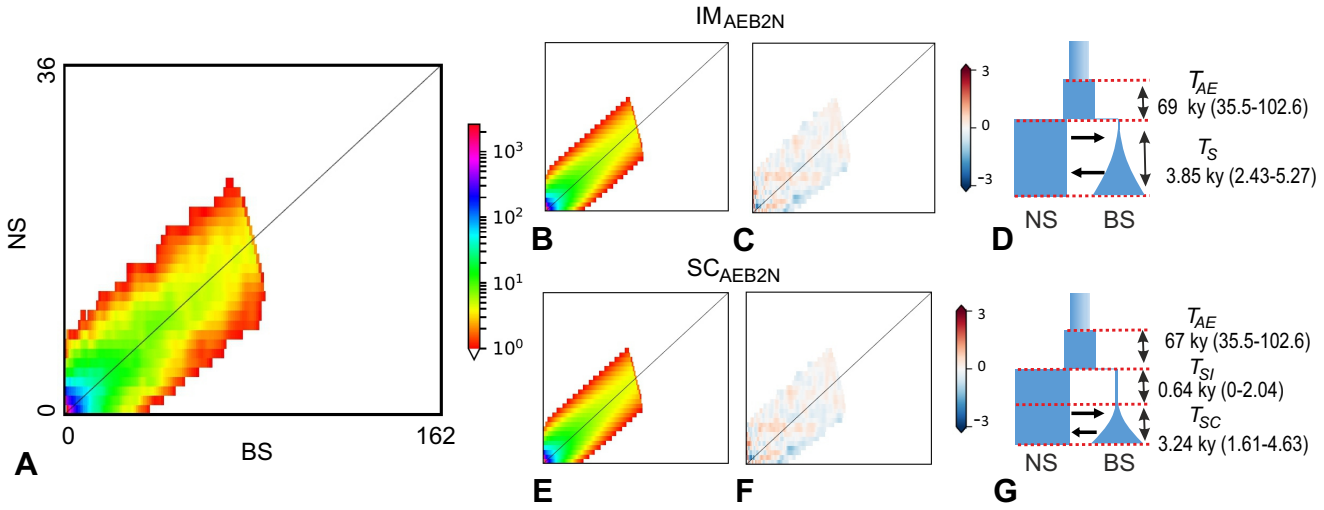


Fig. 7. Observed jAFS (**A**), modelled jAFS (**B & E**), model residuals (**C & F**) and parameters (**D & G**) for the two best models. **B** modelled jAFS, **C** residuals and **D** graphical representation of the IM$_{AEB2N}$ model as well as estimated time parameters with 95% Confidence Intervals. **E** modelled jAFS, **F** residuals and **G** graphical representation of the SC$_{AEB2N}$ model as well as estimated time parameters with 95% Confidence Intervals.

ancestral expansion is not included in the model parameters. This is most likely because the effect of not modelling 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 (3) 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. Another observation is that when the true divergence scenario is a basic SC model, IM models that included very abrupt population size changes performed better than IM models that did not (but always worse than SC models). For example, IM$_B$ models outperformed IM models, and IM$_{AEB}$ models performed best (Supplementary Figs 2-9). 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 (40). In reality, when comparing more biologically realistic models, a unique function producing the expected AFS is identifiable (22, 41). 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 modelling: models are always extreme simplification of complex biological processes, and we often do not know what complexities can be safely excluded. Roux et al (3, 11) 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 towards choice of SC models when the real scenario generating the data had continuous gene flow, and the general overestimation of the proportion of 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 recent speciation research; 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 (e.g. 42–44). Furthermore, as testing on 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 utilize approach to estimate demographic parameters based on the jAFS of multiple populations, DaDi (16, ; Supplementay Fig. 17). It is unclear at this stage what the effect of unmodelled 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 Θ, π, 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 (e.g. 11, 13, 45). It is interesting to notice that when reconstructing human demographic history, it is common to test more realistic demographic scenarios modelling past demographic changes in ancestral and daughter populations as well as bottlenecks followed by population growth (16, 17, 35). 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 unmodelled demographic events. Here we showed that neither of these assumptions are correct.

It is not our intention to suggest that most studies that com-

pared IM and SC models and found strong support for SC have likely chosen the wrong scenario. Some studies have indeed modelled the potential effects of bottlenecks (e.g. 12), and formal model testing is often only one of several lines of evidence suggesting secondary contact (e.g. 11, 18–20). Furthermore, in this study we focused on relatively shallow demographic scenarios (thousands or tens of thousands generations). Nevertheless, the results clearly show that realistic demographic scenarios that could be expected during the past glacial cycles can generate strong false support for SC models with long periods of strict isolation. This has clear implication 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.

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 marine 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 population and species in the Baltic Sea being the result of both primary and secondary divergence (e.g. 44, 46–48). The evolutionary origin of some specific taxa, such as the Baltic Sea populations of Pleuronectiformes, remains controversial (23, 29, 49, 50) . In our empirical study our model fit suggests that *S. maximus* from the Baltic Sea originated from a very recent invasion (< 6kya) 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 (51). At this time, as noted by Momigliano et al (49), 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 (23) 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 (23). 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 (23). As there is no other evidence for such scenario apart from testing of IM and SC demographic models, we suggest Le Moan et al (23) results are more likely 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 towards the same demographic scenario (Fig. 7) giving strong support for postglacial divergence between flatfish populations in the North Sea and the Baltic Sea. Interestingly, Momigliano et al (49) 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 modelling 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 modelling approaches will provide in the future support for an alternative evolutionary scenario.

Conclusion.

In conclusion, using very extensive simulations as well as empirical data on turbots 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 unmodelled demographic events, and warn that given the centrality of formal testing between competing divergence scenarios in current speciation research, 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.

Data Access

All raw sequence data will be available from GenBank Short Read Archive after publication. All codes and pipelines used for data analyses are accessible via GitHub (https://github.com/Nopaoli/Demographic-Modelling).

ACKNOWLEDGEMENTS

We thank Roger Butlin, Ryan Gutenkunst, Paul Blischak, Xin Huang, Petri Kemppainen, Bohao Fang 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 316294 to PM and 218343 to JM) and the Helsinki Institute of Life Science (HiLIFE).

AUTHOR CONTRIBUTIONS

PM conceived the study, performed laboratory work, analyzed the data and wrote the manuscript. JM contributed to study design, provided constructive feedback on the manuscript and funded sequencing. ABF contributed most of the samples and provided constructive feedback on the manuscript.

Bibliography

1. Alfred R. Wallace. On the law which has regulated the introduction of new species. *The Annals and Magazine of Natural History*, 15:185–196, 1855.
2. Daniel I Bolnick and Benjamin M Fitzpatrick. Sympatric speciation: models and empirical evidence. *Annual Review of Ecology and Systematics*, 38:459–487, 2007.
3. Camille Roux, Christelle Fraisse, Jonathan Romiguier, Yoann Anciaux, Nicolas Galtier, and Nicolas Bierne. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS biology*, 14(12):e2000234, 2016.
4. Chung-I Wu. The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6):851–865, 2001.
5. Patrik Nosil. *Ecological speciation*. Oxford University Press, 2012.
6. Patrik Nosil and Jeffrey L. Feder. Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 367(1587):332–342, 2012.
7. Mark Ravinet, R Faria, RK Butlin, J Galindo, N Bierne, M Rafajlović, MAF Noor, B Mehlig, and AM Westram. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of evolutionary biology*, 30(8):1450–1477, 2017.
8. Patrik Nosil. Speciation with gene flow could be common. *Molecular Ecology*, 17(9):2103–2106, 2008.
9. Patrik Nosil, Luke J. Harmon, and Ole Seehausen. Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution*, 24(3):145–156, 2009.
10. Camille Roux, Georgia Tsagkogeorga, Nicolas Bierne, and Nicolas Galtier. Crossing the species barrier: Genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, 30(7):1574–1587, 2013.
11. Camille. Roux, Christelle Fraïsse, V. Castric, X. Vekemans, G. H. Pogson, and Nicolas Bierne. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? a case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, 27(8):1662–1675, 2014.
12. Clement Rougeux, Louis Bernatchez, and Pierre-Alexandre Gagnaire. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9(8):2057–2074, 2017.
13. Quentin Rougemont and Louis Bernatchez. The demographic history of atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate bayesian computations. *Evolution*, 72(6):1261–1277, 2018.
14. Mark A Beaumont, Wenyang Zhang, and David J Balding. Approximate bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.
15. Laurent Excoffier, Isabelle Dupanloup, Emilia Huerta-Sánchez, Vitor C Sousa, and Matthieu Foll. Robust demographic inference from genomic and snp data. *PLoS genetics*, 9(10): e1003905, 2013.
16. Ryan N. Gutenkunst, Ryan D. Hernandez, Scott H. Williamson, and Carlos D. Bustamante. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLOS Genetics*, 5(10):e1000695, 2009.
17. Julien Jouganous, Will Long, Aaron P. Ragsdale, and Simon Gravel. Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics*, 206(3):1549–1567, 2017.
18. Mbaye Tine, Heiner Kuhl, Pierre-Alexandre Gagnaire, Bruno Louro, Erick Desmarais, Rute ST Martins, Jochen Hecht, Florian Knaust, Khalid Belkhir, and Sven Klages. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications*, 5:5770, 2014.
19. Quentin Rougemont, Pierre-Alexandre Gagnaire, Charles Perrier, Clémence Genthon, Anne-Laure Besnard, Sophie Launey, and Guillaume Evanno. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*, 26(1):142–162, 2017.
20. A. Le Moan, P. A. Gagnaire, and F. Bonhomme. Parallel genetic divergence among coastal-marine ecotype pairs of european anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, 25(13):3187–3202, 2016.
21. Pierre-Alexandre Gagnaire, Jean-Baptiste Lamy, Florence Cornette, Serge Heurtebise, Lionel Dégremont, Emilie Flahauw, Pierre Boudry, Nicolas Bierne, Sylvie Lapegue, and evolution. Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome Biology and Evolution*, 10(9):2518–2534, 2018.
22. Anand Bhaskar and Yun S Song. Descartes' rule of signs and the identifiability of population demographic models from genomic variation data. *Annals of statistics*, 42(6):2469, 2014.
23. Alan Le Moan, Oscar Gaggiotti, Romina Henriques, Paulino Martinez, Dorte Bekkevold, and Jakob Hemmer-Hansen. Beyond parallel evolution: when several species colonize the same environmental gradient. *bioRxiv*, page 662569, 2019.
24. Richard R Hudson. Generating samples under a wright–fisher neutral model of genetic variation. *Bioinformatics*, 18(2):337–338, 2002.
25. Daniel M. Portik, Adam D. Leaché, Danielle Rivera, Michael F. Barej, Marius Burger, Mareike Hirschfeld, Mark-Oliver Rödel, David C. Blackburn, and Matthew K. Fujita. Evaluating mechanisms of diversification in a guineo-congolian tropical forest frog using demographic model selection. *Molecular Ecology*, 26(19):5245–5263, 2017.
26. Ann-Britt Florin and Jacob Höglund. Absence of population structure of turbot (*Psetta maxima*) in the baltic sea. *Molecular Ecology*, 16(1):115–126, 2007.
27. Einar E Nielsen, Peter H Nielsen, Dorte Meldrup, and Michael M Hansen. 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. *Molecular Ecology*, 13(3):585–595, 2004.
28. Shi Wang, Eli Meyer, John K. McKay, and Mikhail V. Matz. 2b-rad: a simple and flexible method for genome-wide genotyping. *Nat Meth*, 9(8):808–810, 2012.
29. Paolo Momigliano, Gaël P. J. Denys, Henri Jokinen, and Juha Merilä. *Platichthys solemdali* sp. nov.(actinopterygii, pleuronectiformes): a new flounder species from the baltic sea. *Frontiers in Marine Science*, 5:225, 2018.
30. Antonio Figueras, Diego Robledo, André Corvelo, Miguel Hermida, Patricia Pereiro, Juan A Rubiolo, Jèssica Gómez-Garrido, Laia Carreté, Xabier Bello, Marta Gut, et al. Whole genome sequencing of turbot (*Scophthalmus maximus*; pleuronectiformes): a fish adapted to demersal life. *DNA research*, 23(3):181–192, 2016.
31. Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357, 2012.
32. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.
33. Thorfinn Sand Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. Angsd: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1):356, 2014.
34. Jonas Meisner and Anders Albrechtsen. Inferring population structure and admixture proportions in low-depth ngs data. *Genetics*, 210(2):719, 2018.
35. Simon Gravel, Brenna M. Henn, Ryan N. Gutenkunst, Amit R. Indap, Gabor T. Marth, Andrew G. Clark, Fuli Yu, Richard A. Gibbs, and Carlos D. Bustamante. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29):11983–11988, 2011.
36. Xiaoming Liu and Yun-Xin Fu. Exploring population size changes using snp frequency spectra. *Nature genetics*, 47(5):555–559, 2015.
37. Alec J Coffman, Ping Hsun Hsieh, Simon Gravel, and Ryan N Gutenkunst. Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33(2):591–593, 2015.
38. Jody Hey and Rasmus Nielsen. 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, 2004.
39. Céline Becquet and Molly Przeworski. Learning about modes of speciation by computational approaches. *Evolution*, 63(10):2547–2562, 2009.
40. Simon Myers, Charles Fefferman, and Nick Patterson. Can one learn history from the allelic spectrum? *Theoretical Population Biology*, 73(3):342–348, 2008.
41. Zvi Rosen, Anand Bhaskar, Sebastien Roch, and Yun S Song. Geometry of the sample frequency spectrum and the perils of demographic inference. *Genetics*, 210(2):665–682, 2018.
42. Godfrey M. Hewitt. Post-glacial re-colonization of european biota. *Biological Journal of the Linnean Society*, 68(1-2):87–112, 2008.
43. Shenglin Liu, Michael M Hansen, and Magnus W Jacobsen. Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, 25(20):5187–5202, 2016.
44. Xueyun Feng, Juha Merilä, and Ari Löytynoja. We shall meet again-genomics of historical admixture in the sea. *bioRxiv*, 2020.
45. Miguel Navascués, Delphine Legrand, Cécile Campagne, Marie-Louise Cariou, and Frantz Depaulis. Distinguishing migration from isolation using genes with intragenic recombination: detecting introgression in the *Drosophila simulans* species complex. *BMC Evolutionary Biology*, 14(1):89, 2014.
46. R Väinölä and MM Hvilsom. Genetic divergence and a hybrid zone between baltic and north sea *Mytilus* populations (mytilidae: Mollusca). *Biological Journal of the Linnean Society*, 43 (2):127–148, 1991.
47. R Vainölä. Repeated trans-arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. *Marine Biology*, 143(5):935–946, 2003.
48. Ricardo T. Pereyra, Lena Bergström, Lena Kautsky, and Kerstin Johannesson. Rapid speciation in a newly opened postglacial marine environment, the baltic sea. *BMC Evolutionary Biology*, 9(1):70, 2009.
49. Paolo Momigliano, Henri Jokinen, Antoine Fraimout, Ann-Britt Florin, Alf Norkko, and Juha Merilä. Extraordinarily rapid speciation in a marine fish. *Proceedings of the National Academy of Sciences*, 114(23):6074–6079, 2017.
50. Henri Jokinen, Paolo Momigliano, and Juha Merilä. From ecology to genetics and back: the tale of two flounder species in the baltic sea. *ICES Journal of Marine Science*, 2019.
51. Bo G Gustafsson and Per Westman. On the causes for salinity variations in the baltic sea during the last 8500 years. *Paleoceanography*, 17(3):12–1–12–14, 2002.

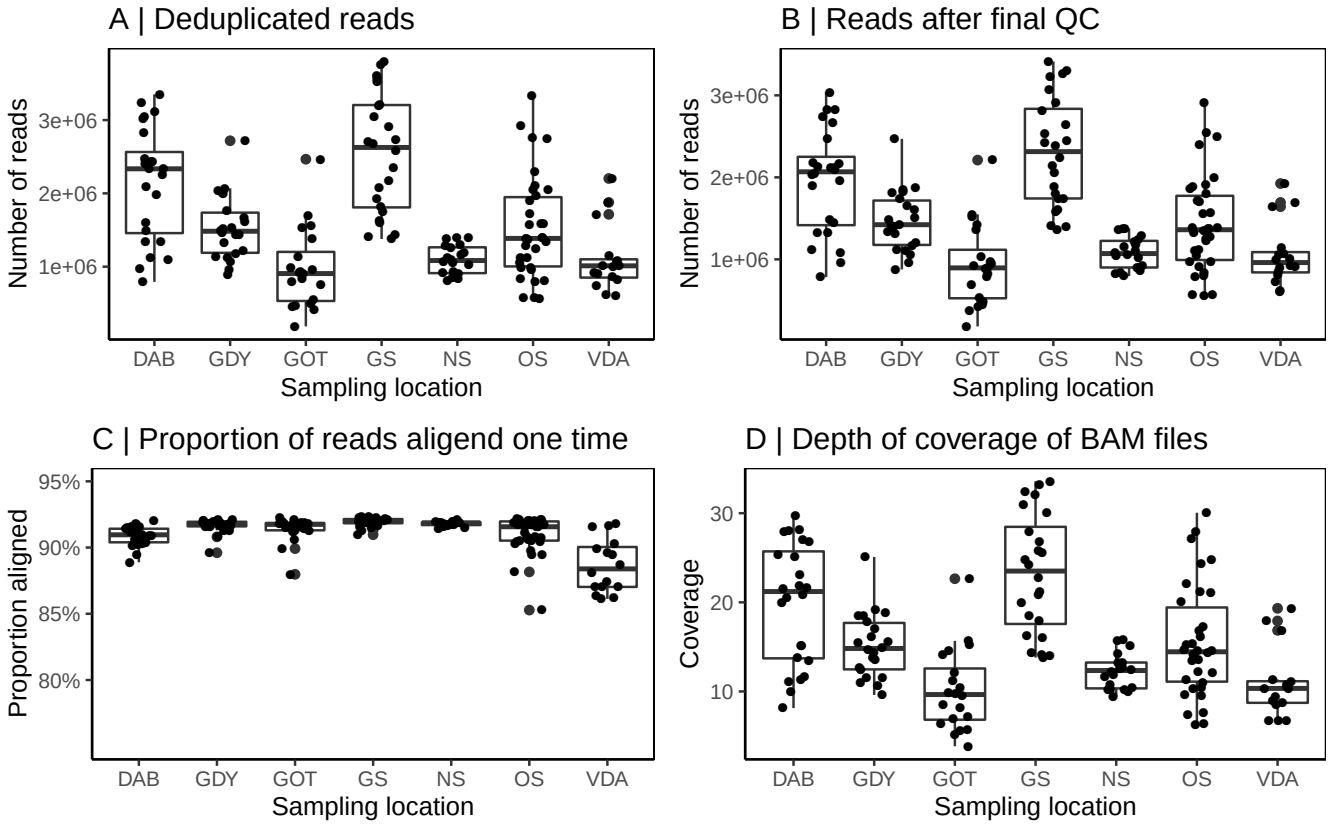
Supplementary Information

Supplementary Table 1: Location name, population codes, sample size (N) and geographic coordinates of all sampling locations of *S. maxixmus* used in this study.

Location	Population code	N	Latitude	Longitude	Reference
North Sea	NS	20	56.58 N	06.12 E	Nielsen et al. (2004)
Vendelsö	VD	27	57.17° N	12.06° E	Florin et al. (2007)
Öresund	OS	35	55.34° N	12.51° E	Florin et al. (2007)
Dabki	DAB	24	54.45° N	6.30° E	Florin et al. (2007)
Gdynia	GDY	23	54.45° N	8.30° E	Florin et al. (2007)
Gotland	GOT	20	57.32° N	18.57° E	Florin et al. (2007)
Gotska Station	GS	24	58.22 °N	19.15° E	Florin et al. (2007)

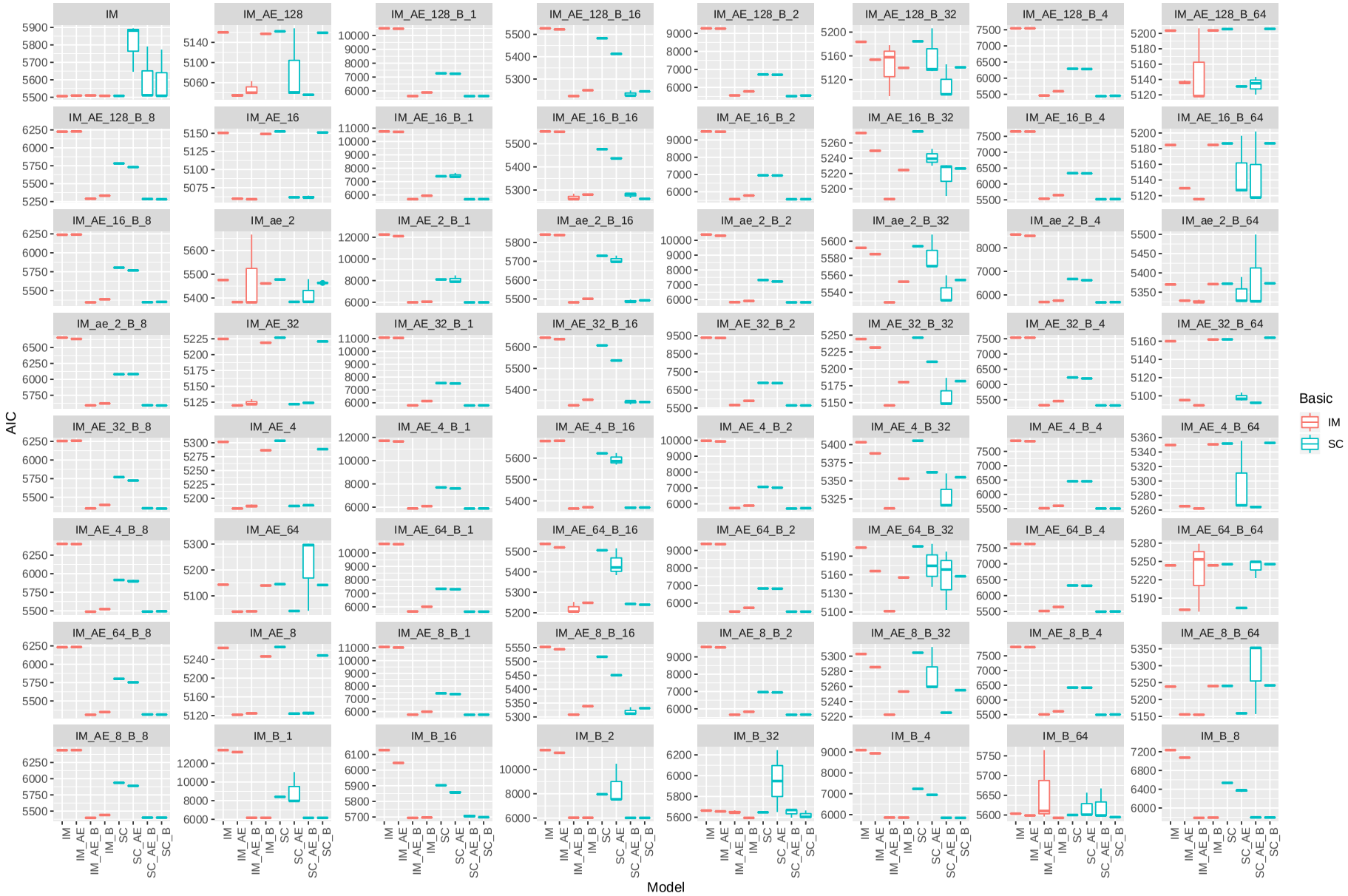
Supplementary Table 2: Parameter estimates ($\pm$sd) of the two best models for the empirical study. AIC= Akaike Information Criterion, W_{AIC}= weight of evidence, $\Theta = 4N_{REF}\mu$, N_{AE}= N_e after population expansion, N_1= contemporary N_e in the North Sea, N_2= contemporary N_e in the Baltic Sea, s=N_e of the Baltic Sea at time of invasion as a proportion of N_1, T_{AE}= time of ancestral expasion, T_S = time of population divergence, T_{SC}= time of secondary contact, M_{12}= proportion of North Sea population that is made up of migrants from the Baltic Sea, M_{21}= proportion of Baltic Sea population that is made up of migrants from the North Sea, Bf= background factor, i.e. N_e of the part of the genome with reduced recombination as a proportion of N_e in the rest of the genome, Q= proportion of the genome with reduced N_e. Population sizes are given in units relative to N_{REF}. Times are given in units of $2N_{REF}$ generations, and need to be added backward in time (the real time of ancestral expansion for the IM_{AEB2n} model is T_{AE} + T_S). Migration rates are given in units of $2N_{REF} \times M_{ij}$.

Model	*AIC*	W_{AIC}	Θ	N_{AE}	N_1	N_2	*s*	T_{AE}	T_S	T_{SC}	M_{12}	M_{21}	*Bf*	*Q*
IM_{AEB2N}	5536	0.66	2398 ±104	1.68 ±0.069	2.78 ±0.56	2.77 ±0.73	0.0356 ±0.006	0.37 ±00.076	0.02 ±0.003	NA	4.66 ±2.35	28.15 ±2	0.1 ±0.013	0.22 ±0.03
SC_{AEB2N}	5537.5	0.31	2383 ±105	1.675 ±0.07	2.898 ±0.62	2.75 ±0.7	0.054 ±0.023	0.38 ±0.08	0.003 ±0.003	0.017 ±0.037	4.53 ±2.54	28.15 ±2.26	0.1 ±0.012	0.22 ±0.03

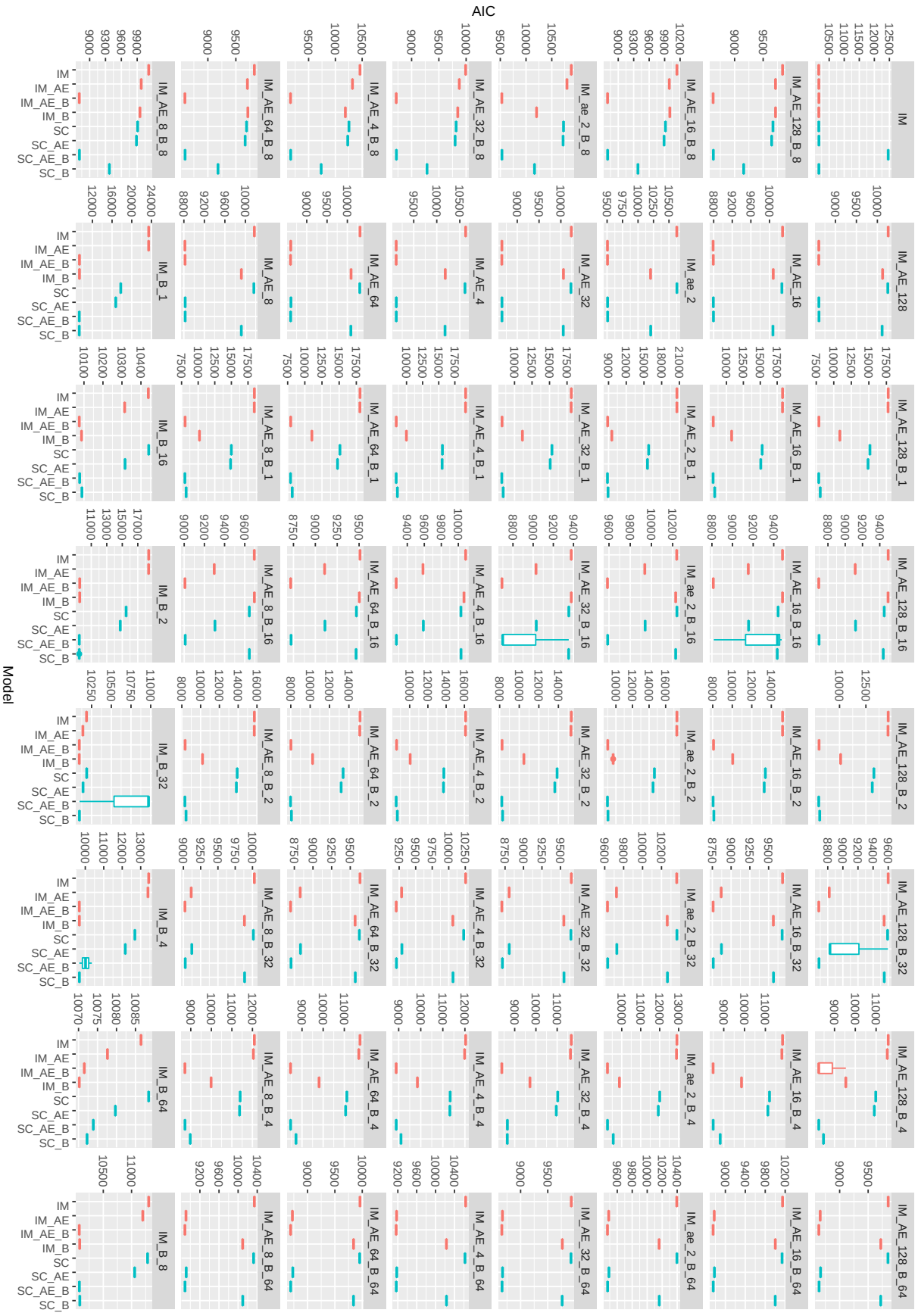


Supplementary Fig. 1: summary statistics from 2b-RAD data. Number of deduplicated reads (**A**) reads after quality check (**B**), proportion of reads which mapped one time to the reference genome (**C**) and mean per individual depth of coverage calculated from BAM files (**D**).

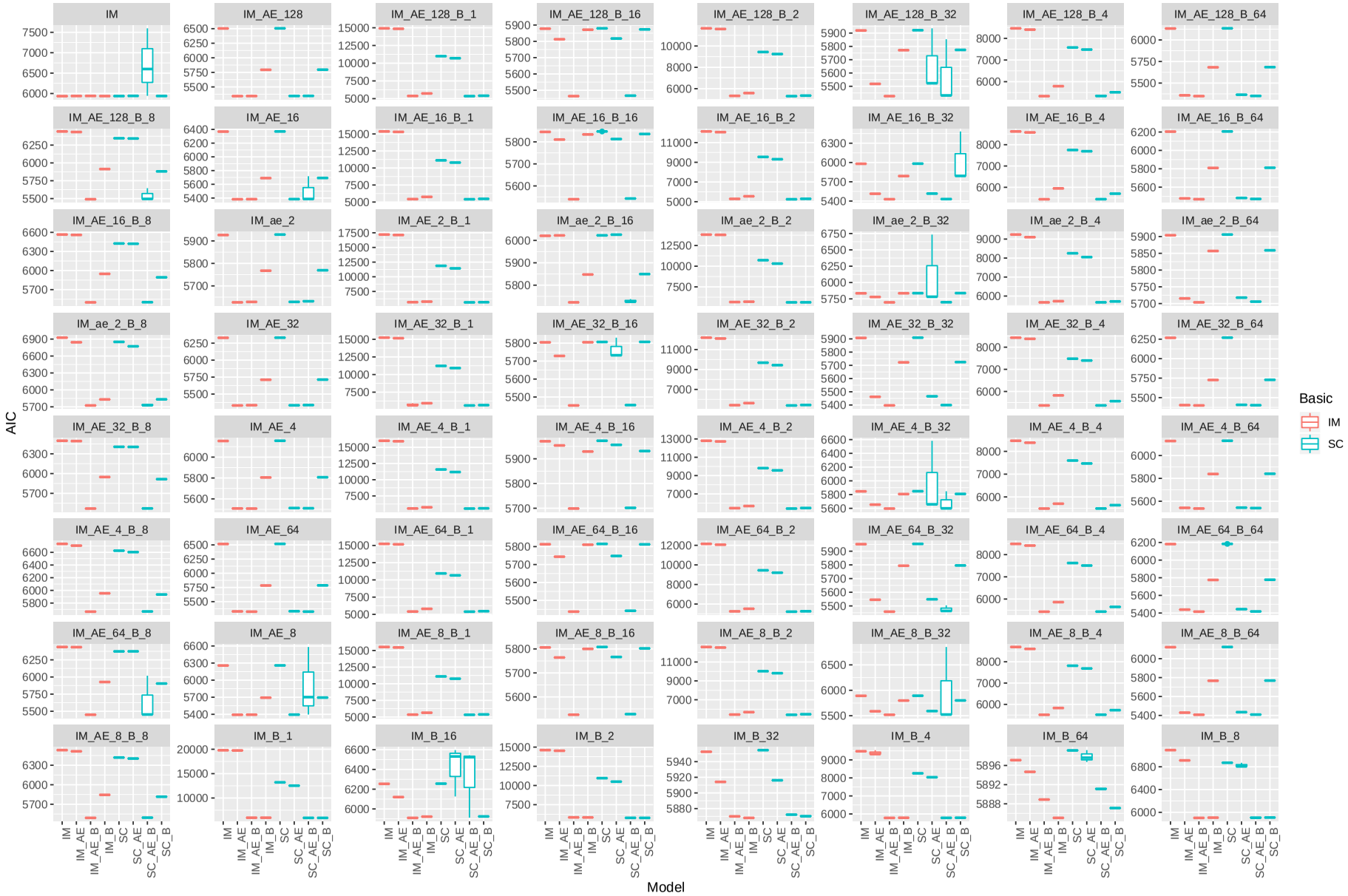
Supplementary Fig. 2: Boxplot showing AIC estimates for IM and SC models for the unfolded jAFS of 64 simulations (1 million loci) under the asymmetric IM model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.



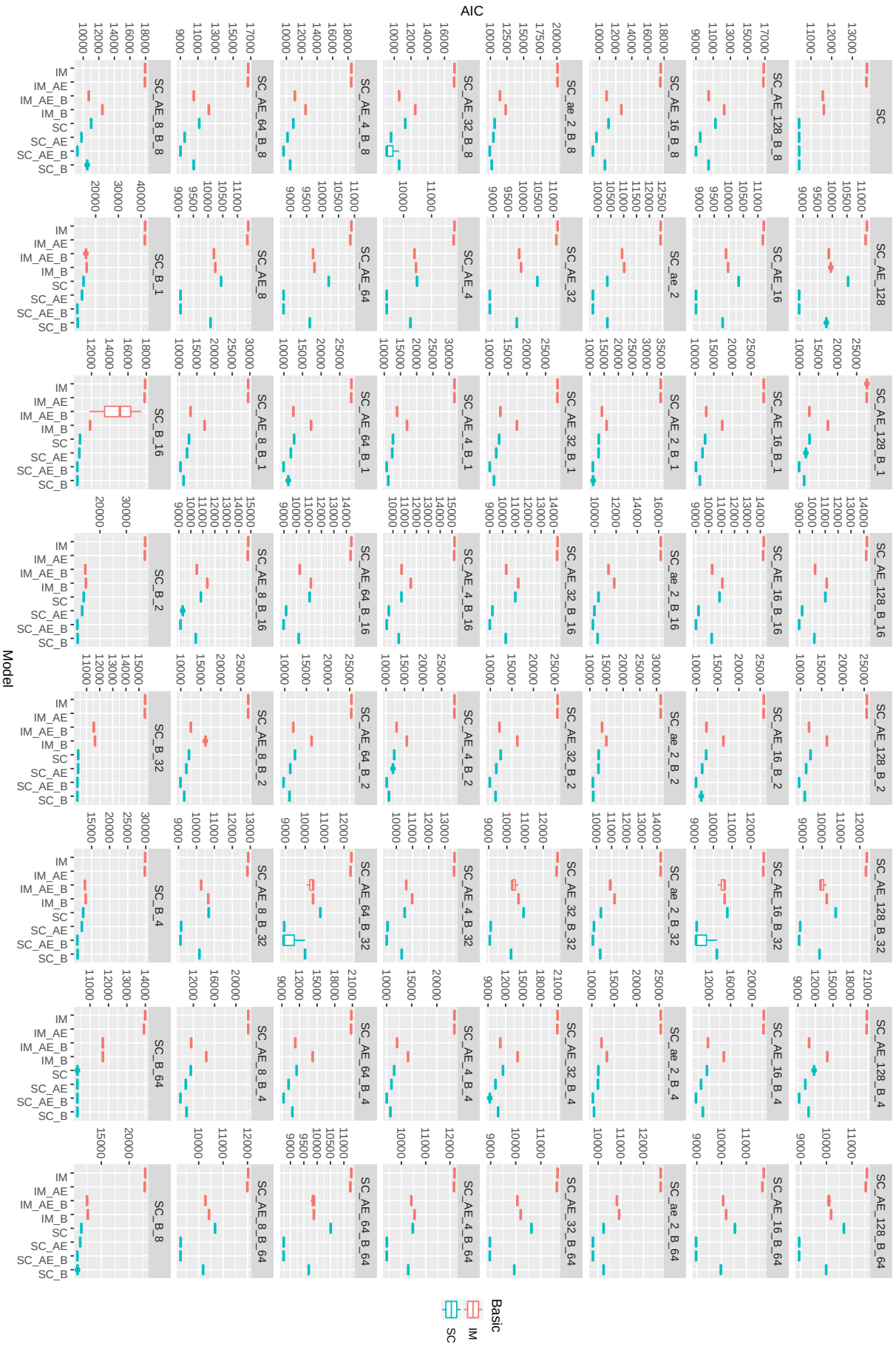
Supplementary Fig. 3: Boxplot showing AIC estimates for IM and SC models for the folded jAFS of 64 simulations (1 million loci) under the asymmetric IM model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.



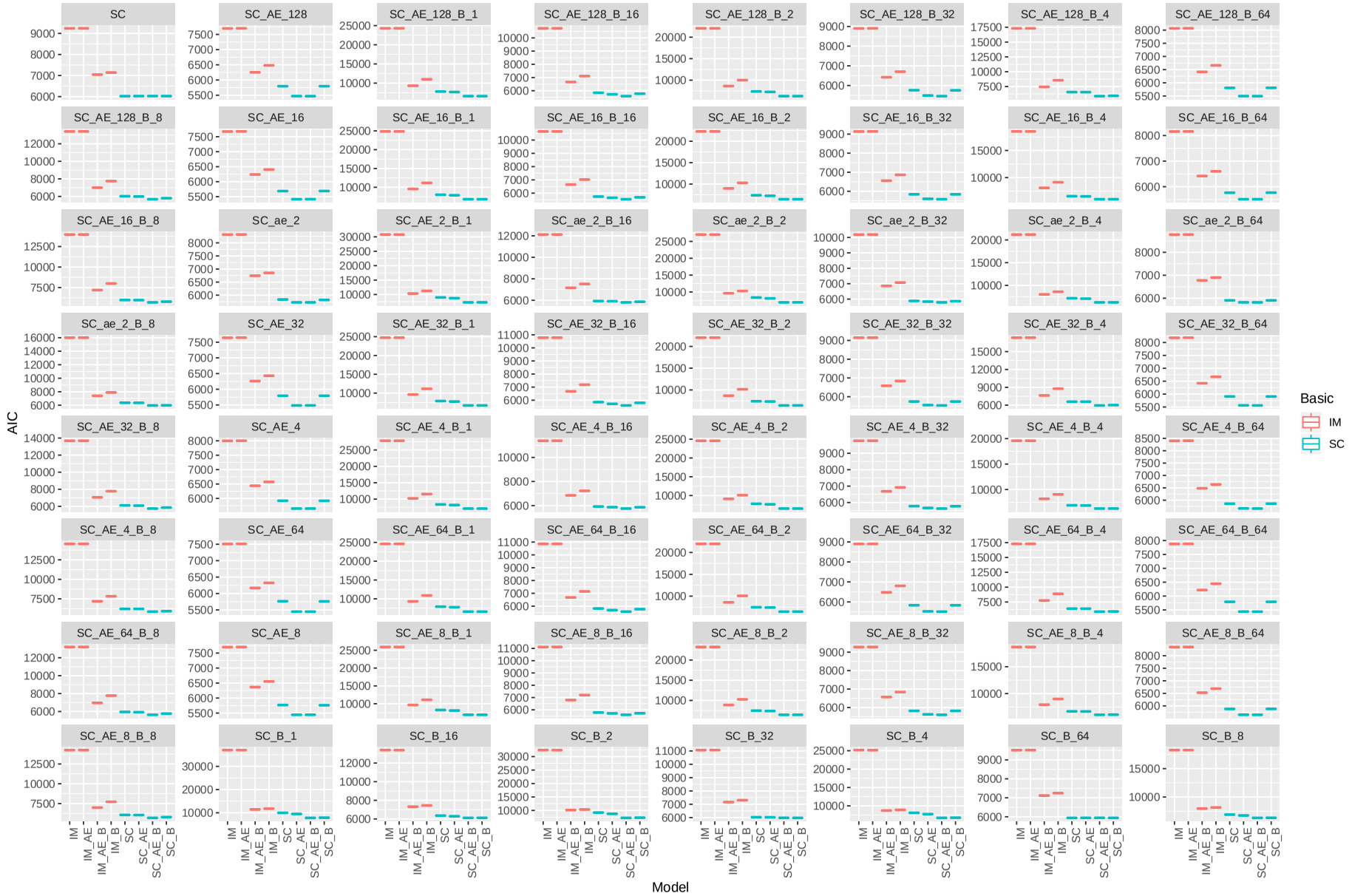
Supplementary Fig. 4: Boxplot showing AIC estimates for IM and SC models for the unfolded jAFS of 64 simulations (1 million loci) under the symmetric IM model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.



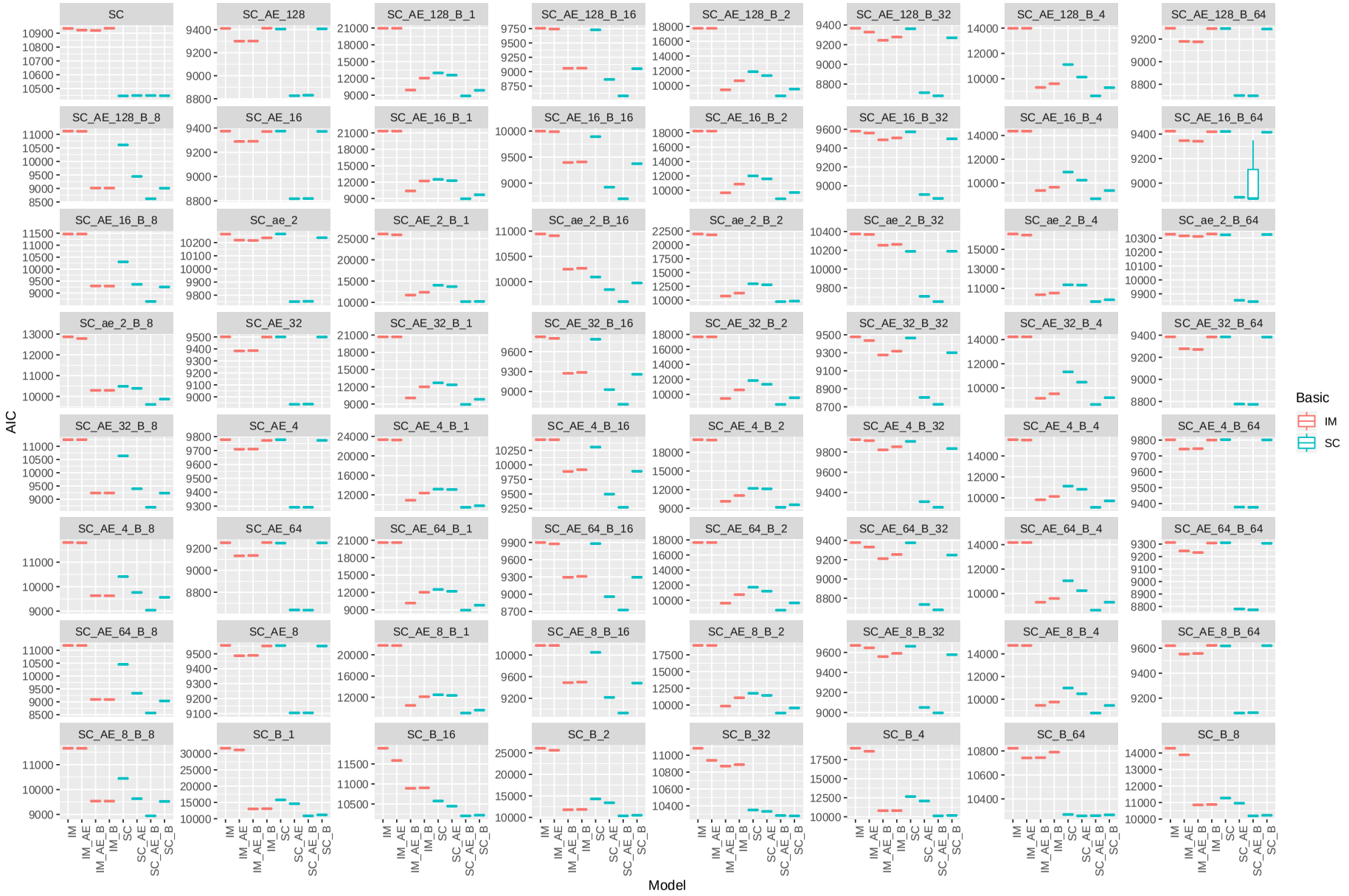
Supplementary Fig. 5: Boxplot showing AIC estimates for IM and SC models for the folded jAFS of 64 simulations (1 million loci) under the symmetric IM model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.



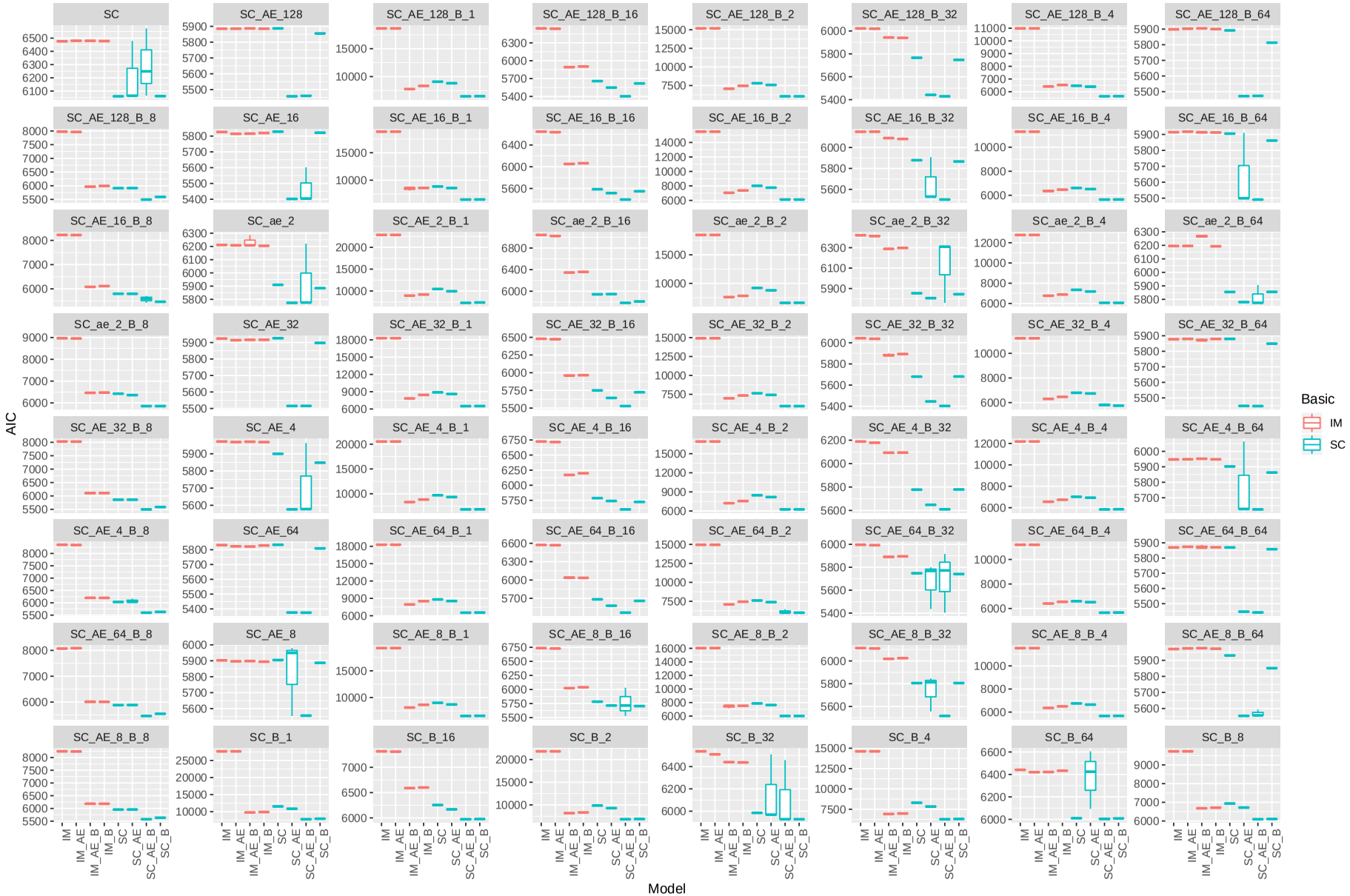
Supplementary Fig. 6: Boxplot showing AIC estimates for IM and SC models for the unfolded jAFS of 64 simulations (1 million loci) under the asymmetric SC model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.



Supplementary Fig. 7: Boxplot showing AIC estimates for IM and SC models for the folded jAFS of 64 simulations (1 million loci) under the asymmetric SC model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.

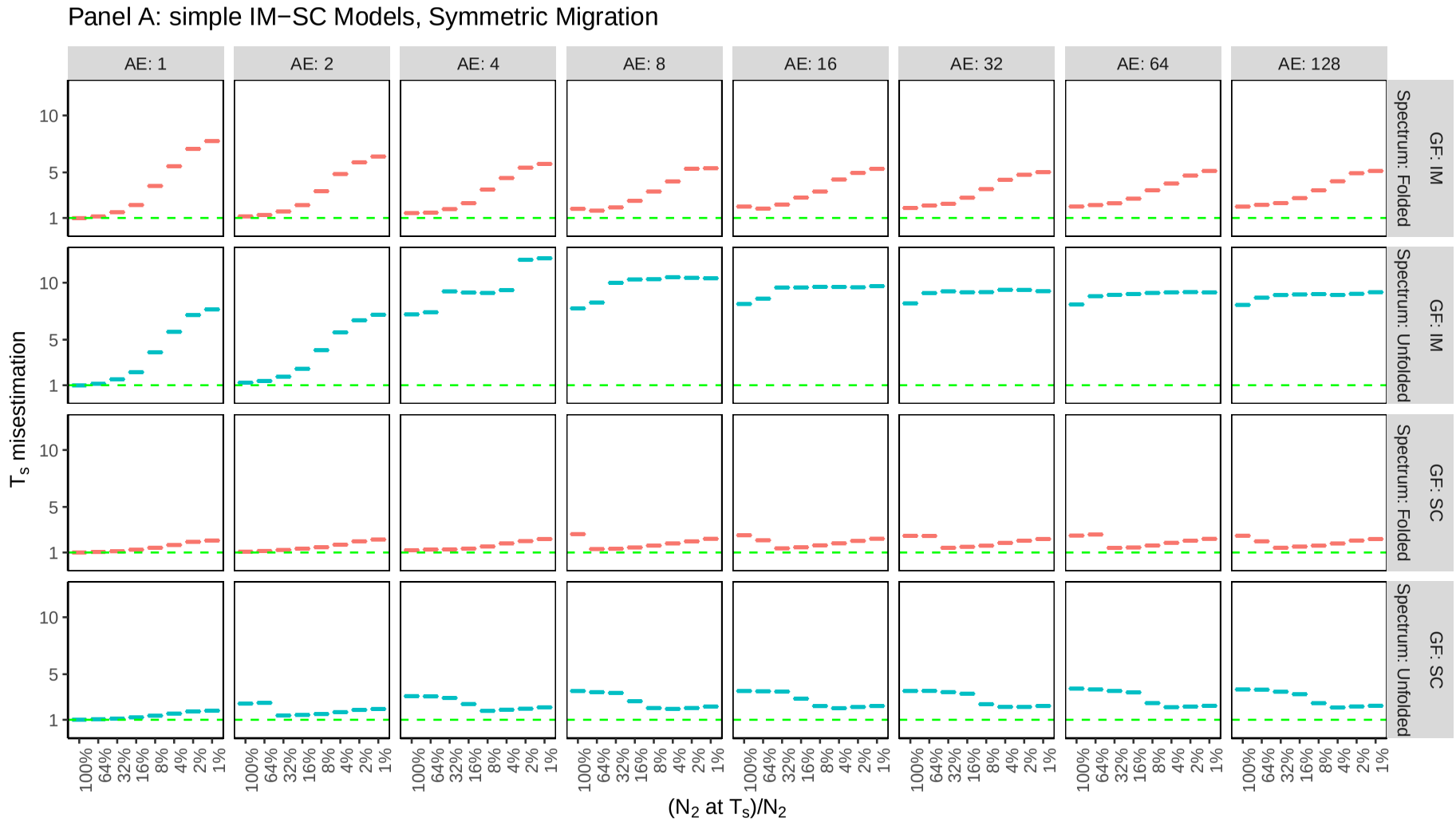


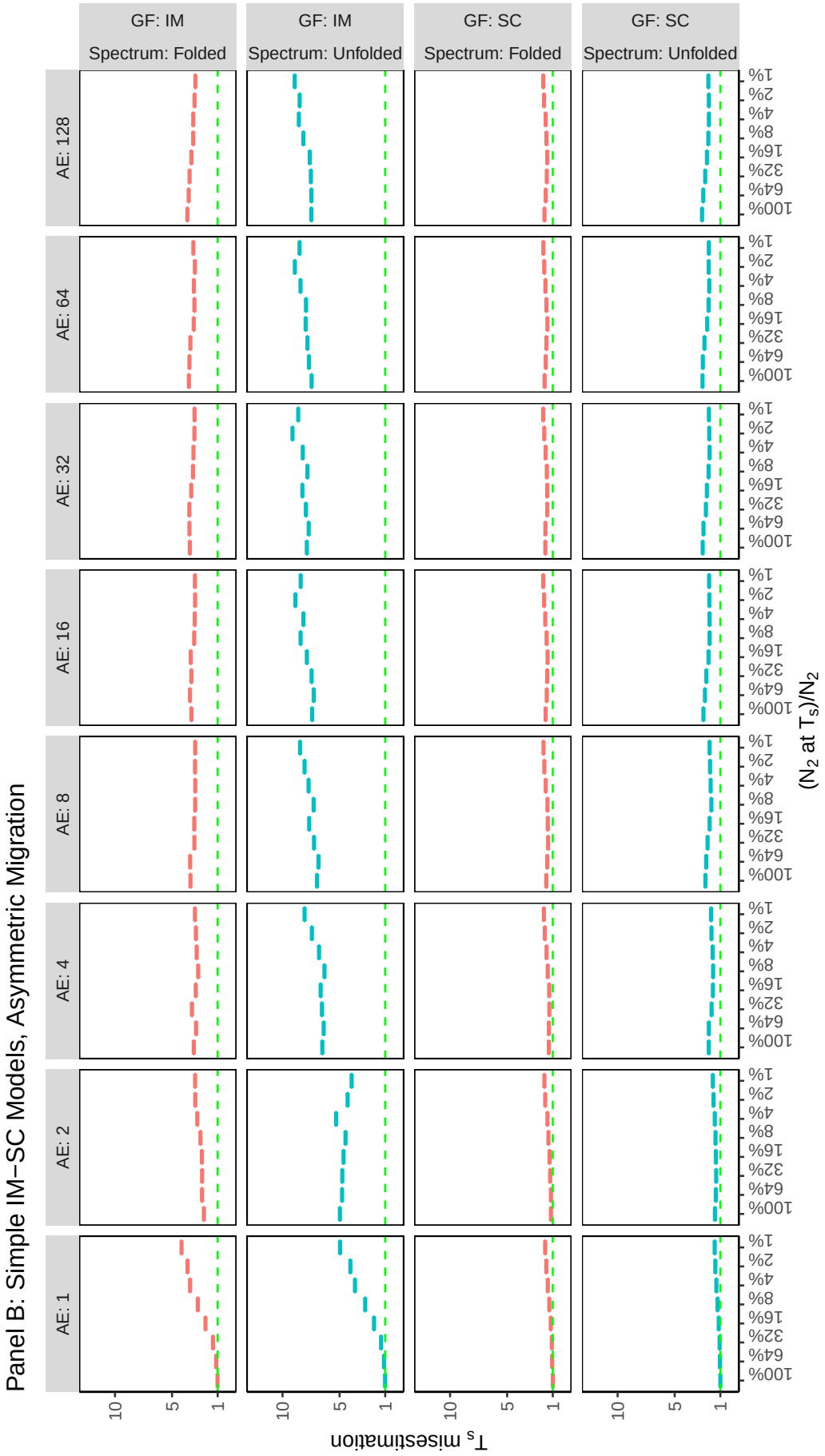
Supplementary Fig. 8: Boxplot showing AIC estimates for IM and SC models for the unfolded jAFS of 64 simulations (1 million loci) under the symmetric SC model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.

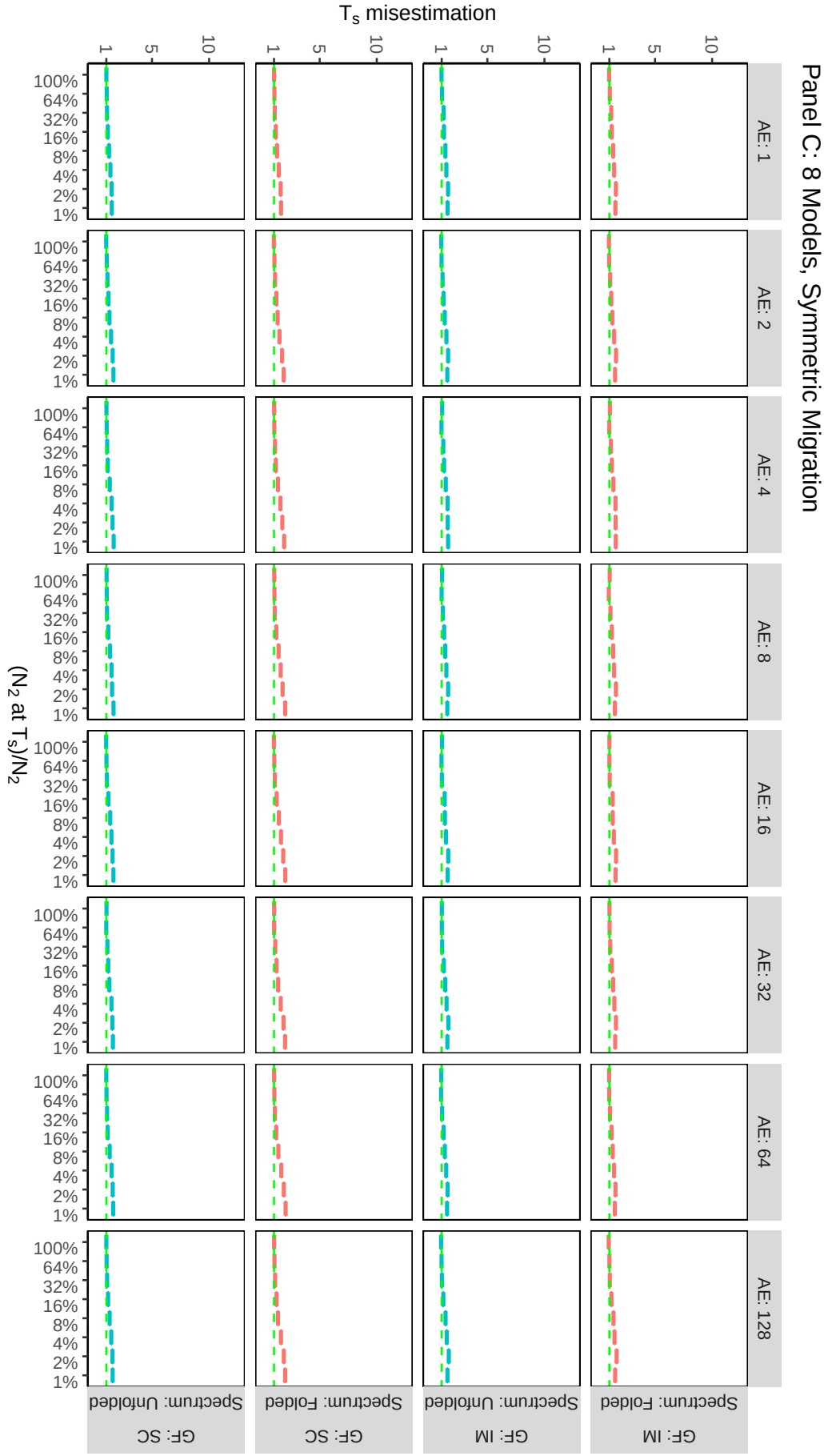


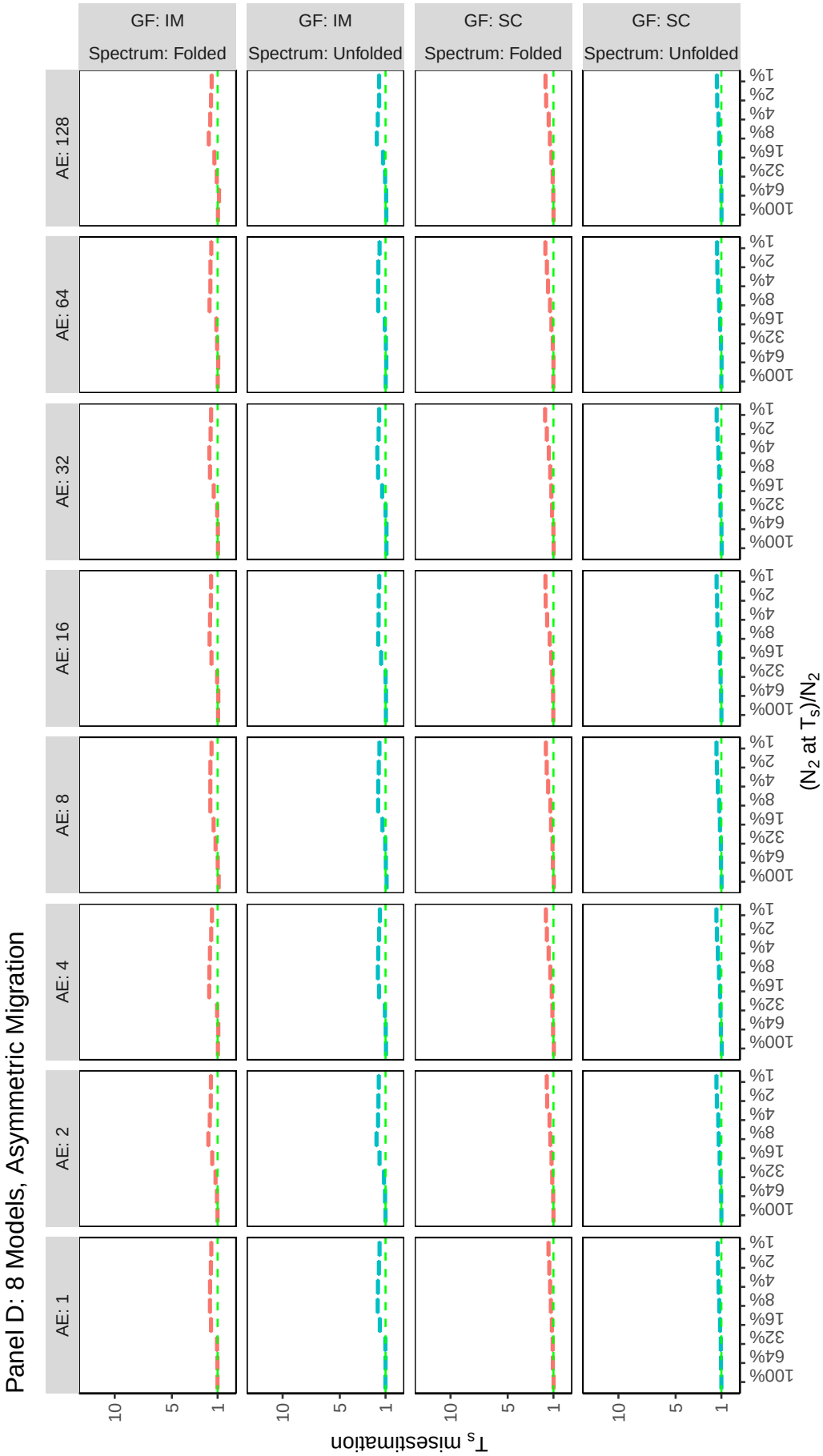
Supplementary Fig. 9: Boxplot showing AIC estimates for IM and SC models for the folded jAFS of 64 simulations (1 million loci) under the symmetric SC model. The best three replicates for each model/simulation combination are plotted. Note the very narrow boxplots showing models have converged to the same likelihhods, with a few exception.

Supplementary Fig. 10: misestimation of T_S. Columns is the grids group by simulation Ancestral Expansion (from 1, i.e. no expansion, to 128 x). The x axes represent the size of population two (N_2) at time of split (T_S) as a proportion of current N_2. Rows of the plot grids group by gene flow scenario of the simulation (IM and SC) and by whether the unfolded or folded spectrum was used for inference. Panels A and B show T_S misestimation when the basic IM and SC models are used for inference, and under a symmetric and asymmetric migration scenario respectively. Panels C and D show (T_S) misestimation when the full 8 models (IM,SC, IMAE, SC_{AE}, IM_B, SC_B, IM_{AEB}, SC_{AEB} are used. For each model, T_S misestimation was calculated as $T_S^{model}/T_S^{simulation}$.

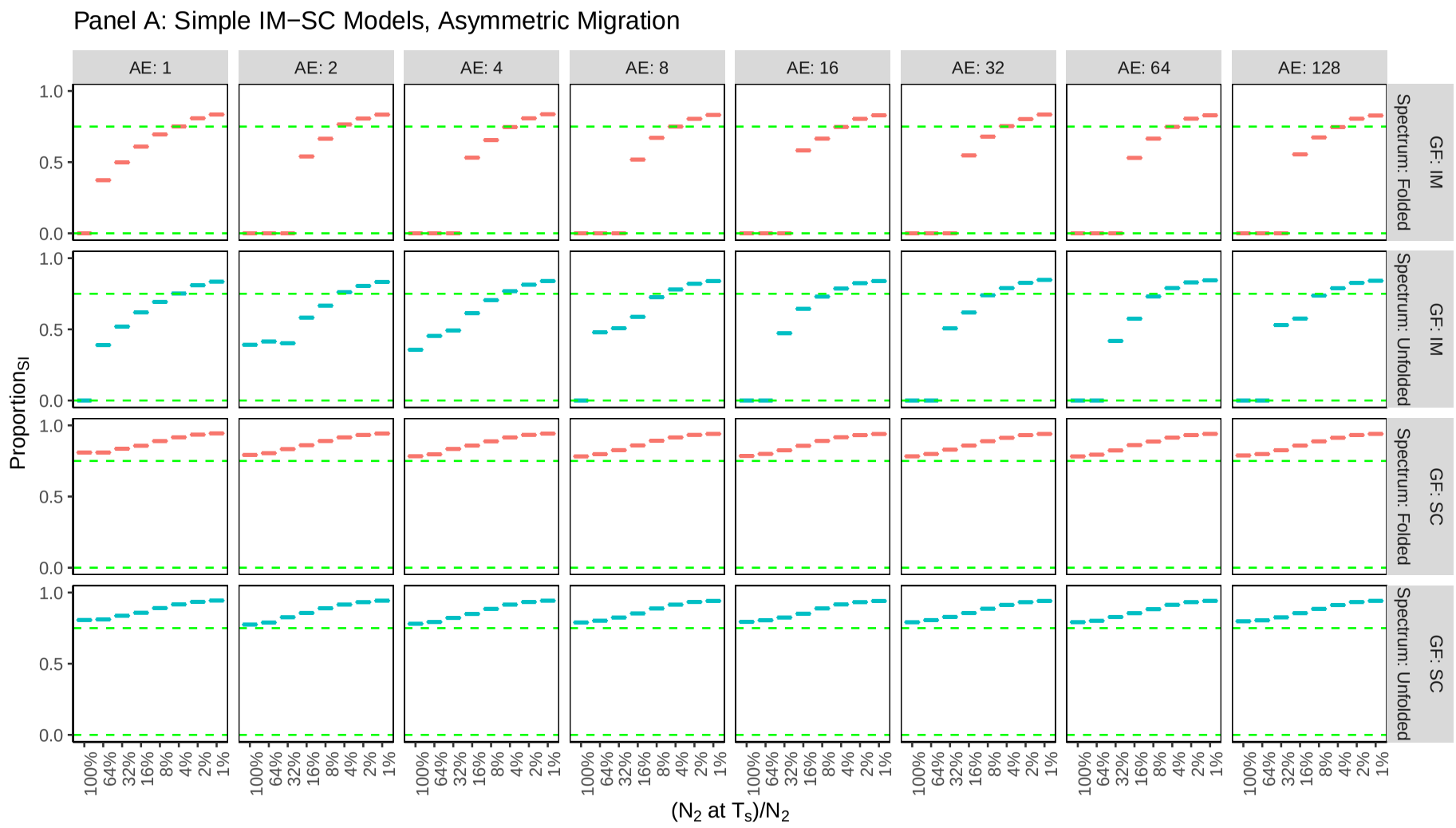


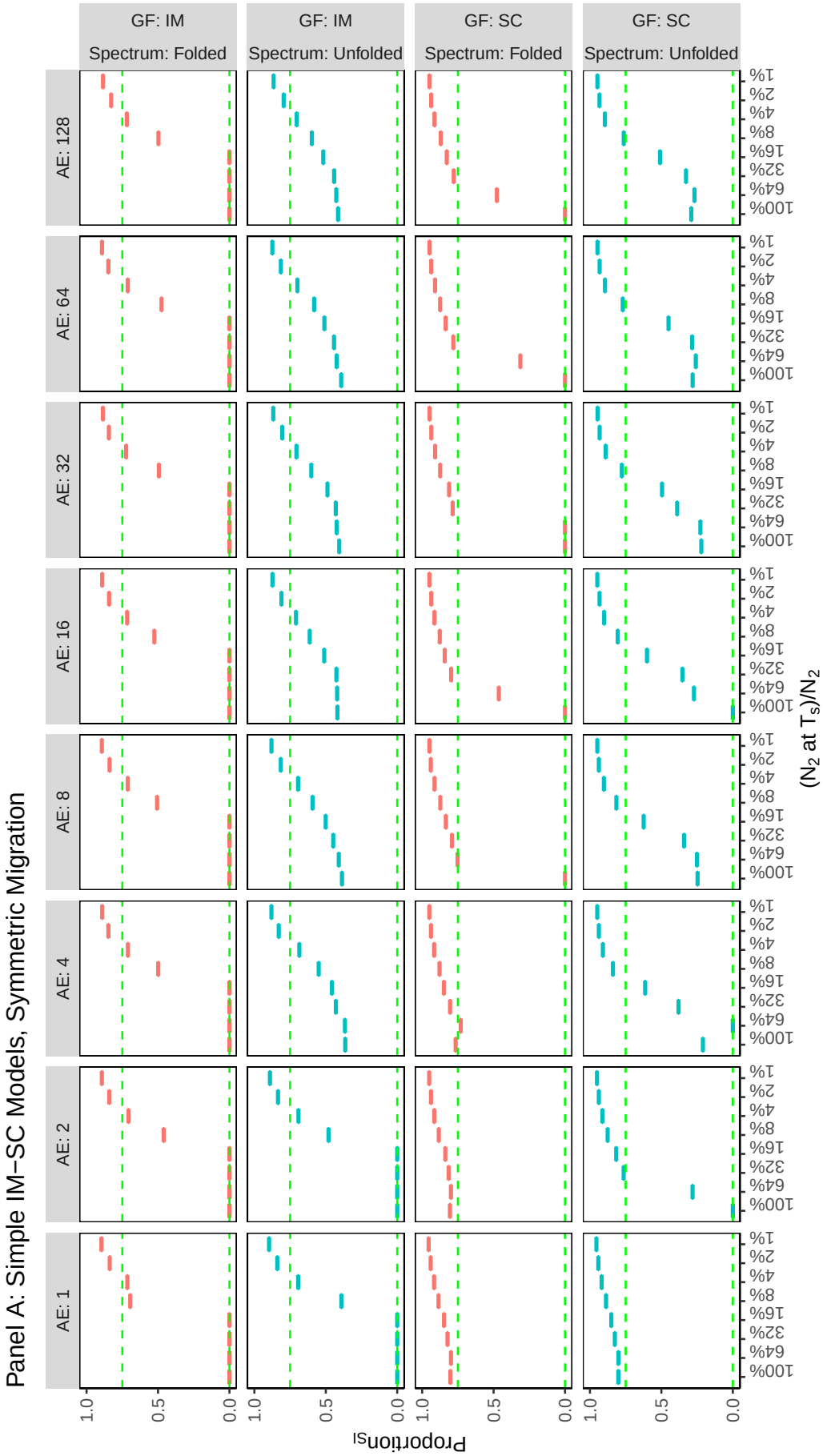

Panel B: Simple IM−SC Models, Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
10
5
1
100%
64%
32%
16%
8%
4%
2%
1%
T_s misestimation
$(N_2 \text{ at } T_s)/N_2$

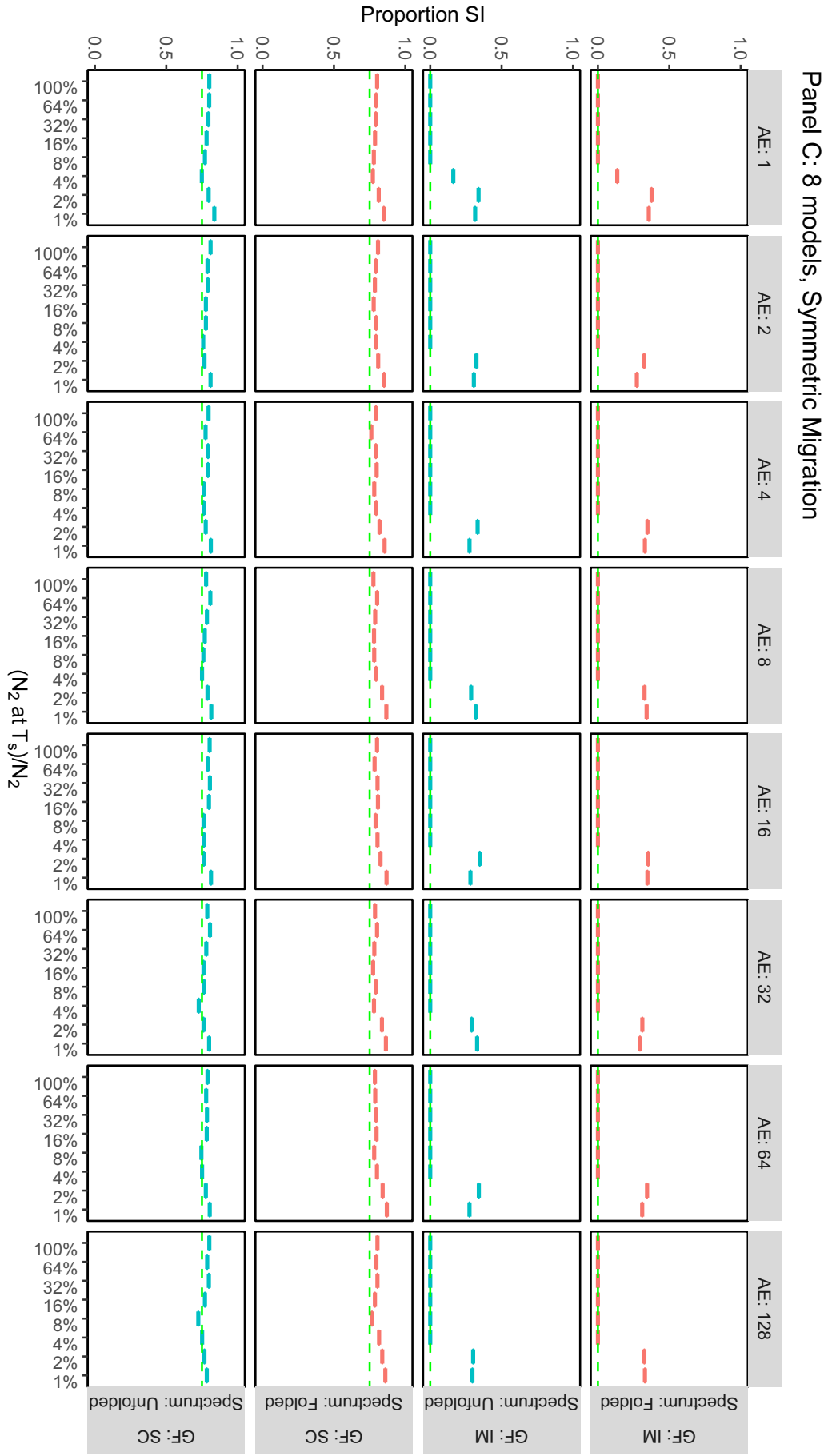

Panel C: 8 Models, Symmetric Migration
T_s misestimation
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
GF: IM
GF: SC
Spectrum: Folded
Spectrum: Unfolded
1
5
10
100%
64%
32%
16%
8%
4%
2%
1%
(N_2 at T_s)/N_2

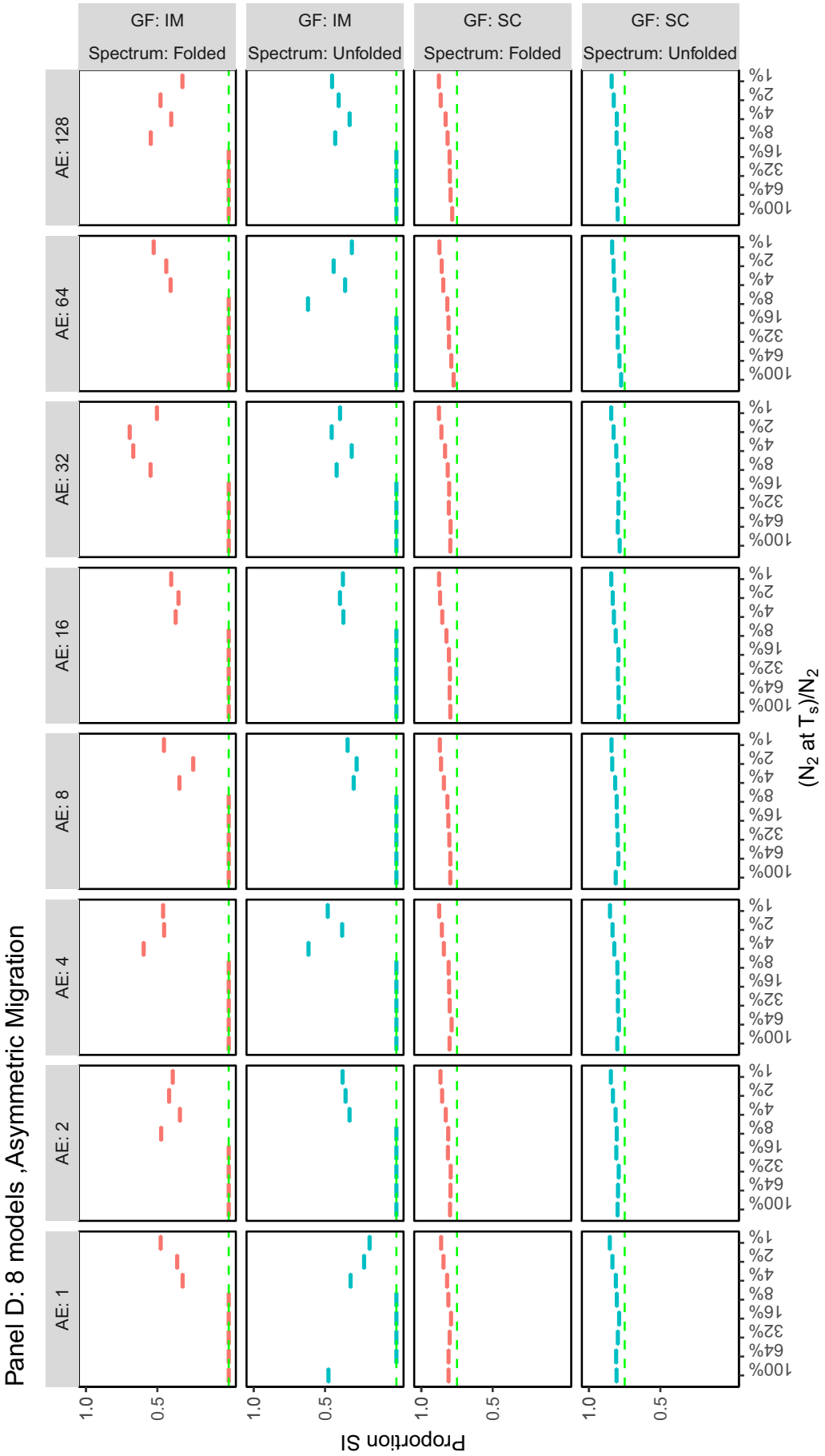

Panel D: 8 Models, Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
10
5
1
T_s misestimation
100%
64%
32%
16%
8%
4%
2%
1%
$(N_2 \text{ at } T_s)/N_2$

Supplementary Fig. 11: Inferred proportion of SI during divergence by the best model. Columns and rows in the grids group by simulations as for Supplementary Fig. 10. Green dotted lines represent the true proportion of SI for the simulation scenarios. For each simulation the value reported is the value from the model replicate with the lowest AIC.

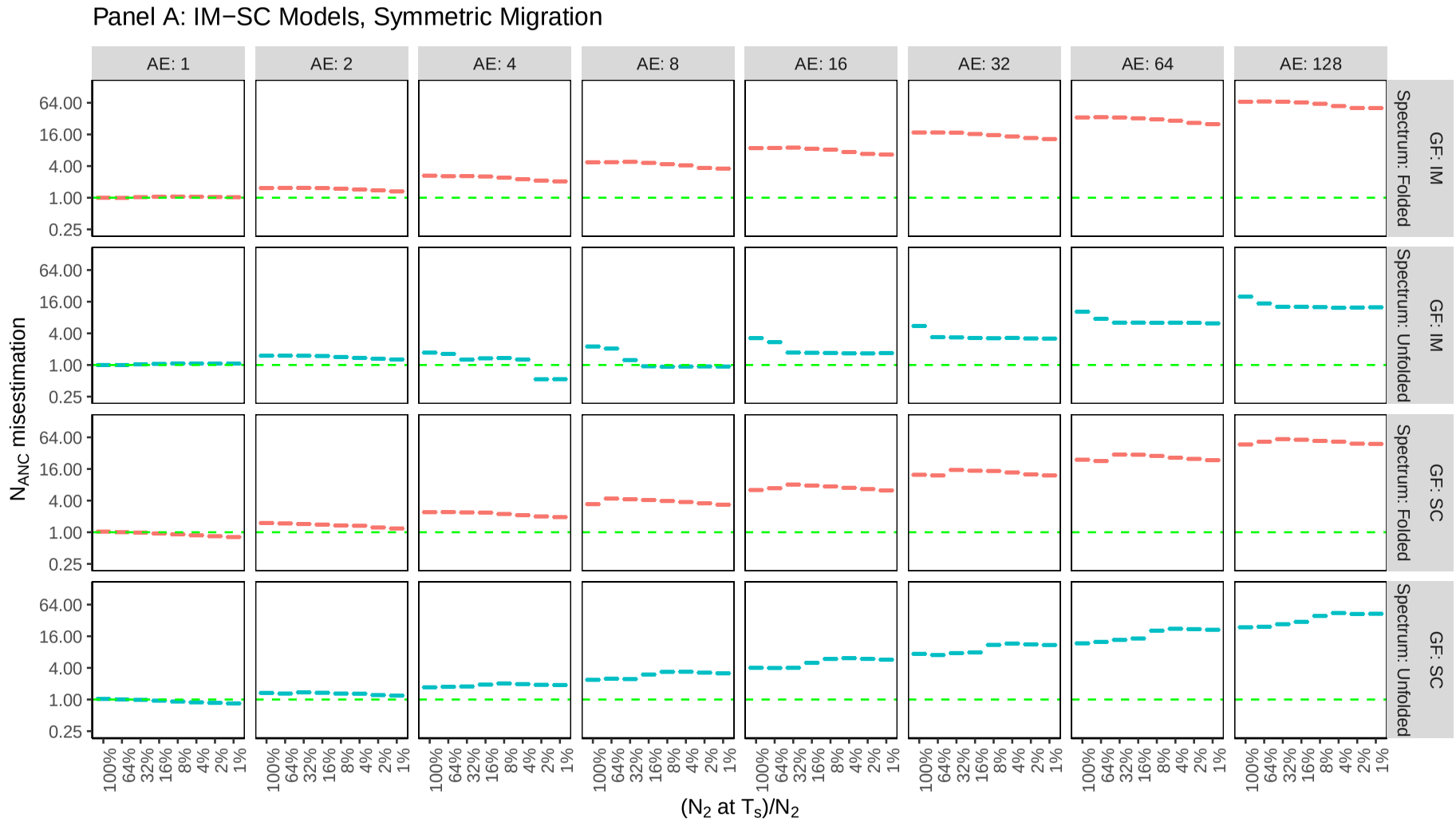


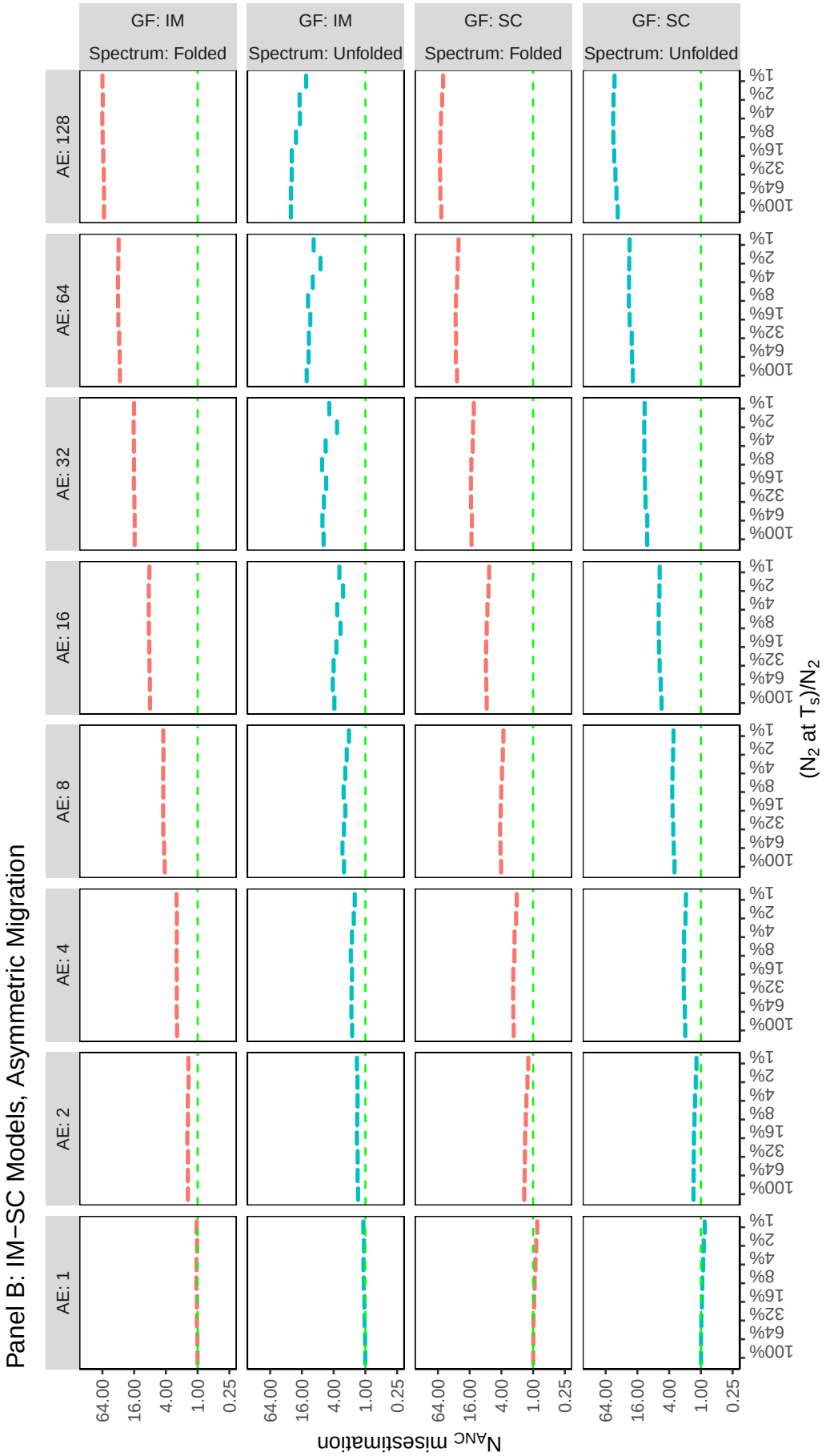

Panel A: Simple IM−SC Models, Symmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
1.0
0.5
0.0
Proportion$_{SI}$
100%
64%
32%
16%
8%
4%
2%
1%
(N_2 at T_S)/N_2

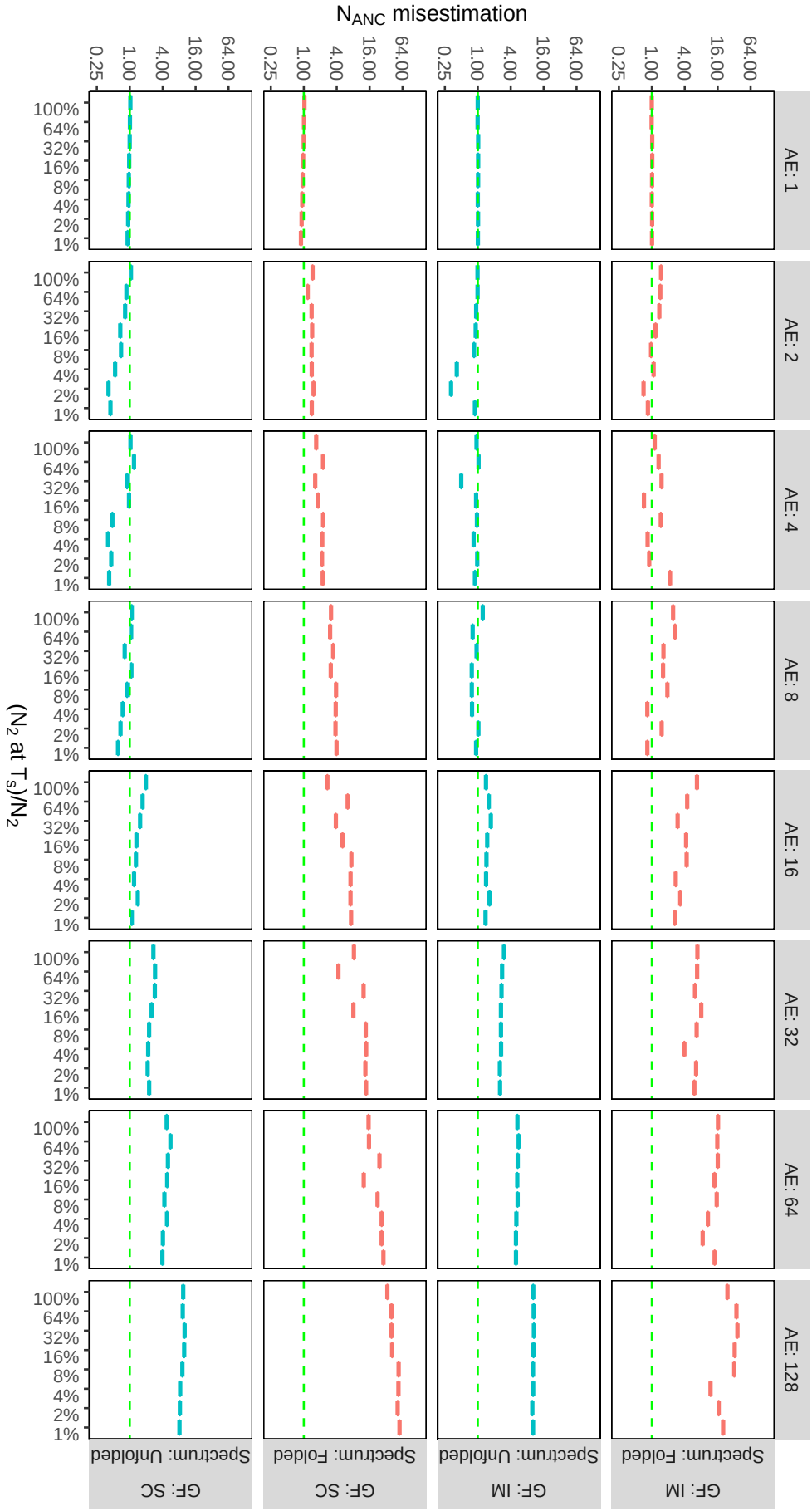

Panel C: 8 models, Symmetric Migration
Proportion SI
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
GF: IM
GF: SC
Spectrum: Folded
Spectrum: Unfolded
0.0
0.5
1.0
100%
64%
32%
16%
8%
4%
2%
1%
(N2 at Ts)/N2

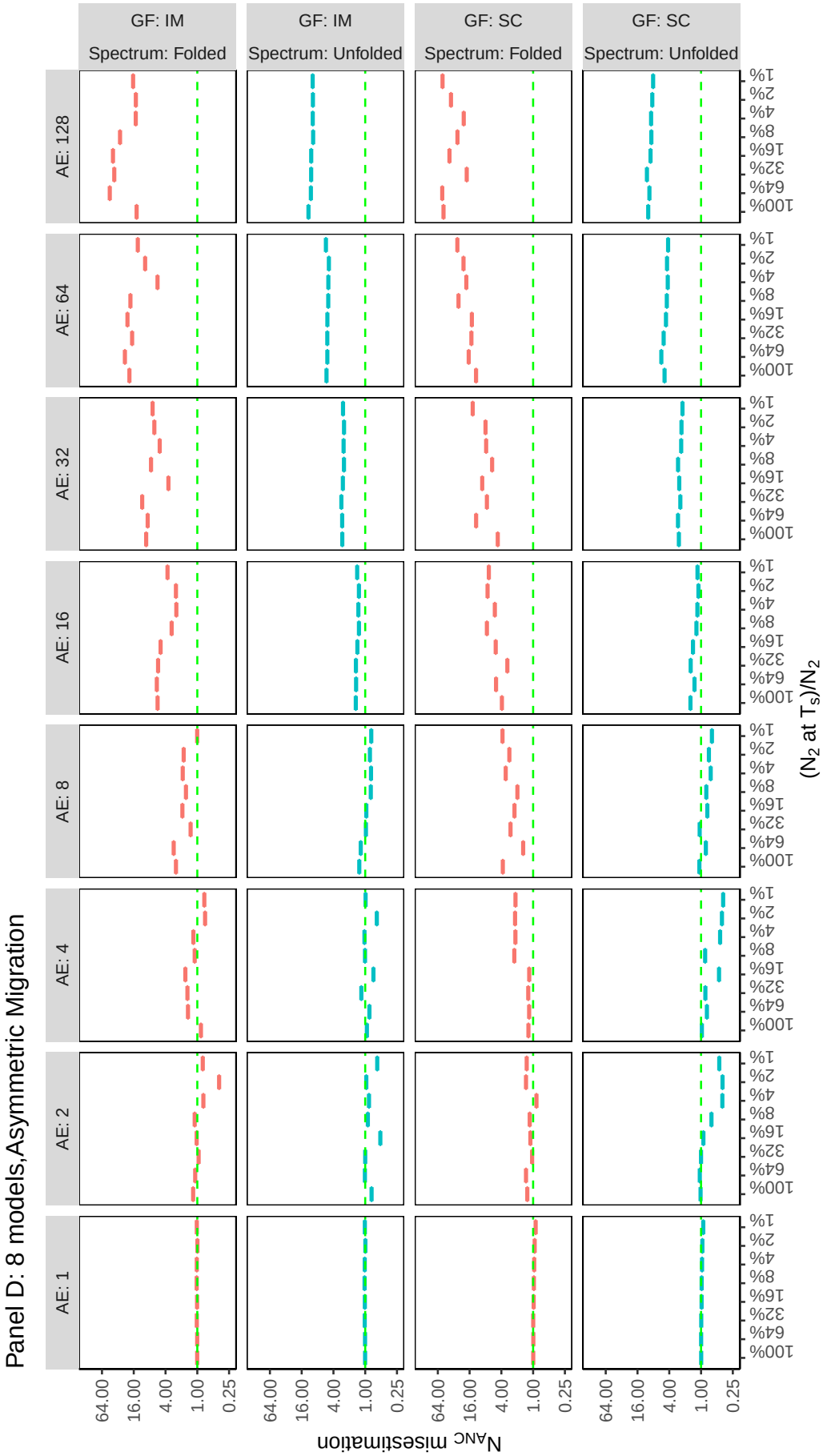

Panel D: 8 models ,Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 128
AE: 64
AE: 32
AE: 16
AE: 8
AE: 4
AE: 2
AE: 1
1%
2%
4%
8%
16%
32%
64%
100%
1.0
0.5
Proportion SI
(N2 at Ts)/N2

Supplementary Fig. 12: misestimation of N_{ANC}. Columns in the grids group by simulation Ancestral Expansion (from 1, i.e. no expansion, to 128 x). The x axes represent the size of population 2 (N_2) at time of split (T_S) as a proportion of current (N_2). Rows of the plot grids group by gene flow scenario of the simulation (IM and SC) and by whether the unfolded or folded spectrum was used for inference. Panels **A** and **B** show N_{ANC} misestimation when the basic IM and SC models are used for inference, and under a symmetric and asymmetric migration scenario respectively. Panels **C** and **D** show N_{ANC} misestimation when the full 8 models (IM,SC,IM_{AE}, SC_{AE}, IM_B, SC_B, IM_{AEB}, SC_{AEB} are used. For each model,N_{ANC} misestimation was calculated as N_{ANC}model/ N_{ANC} simulation.

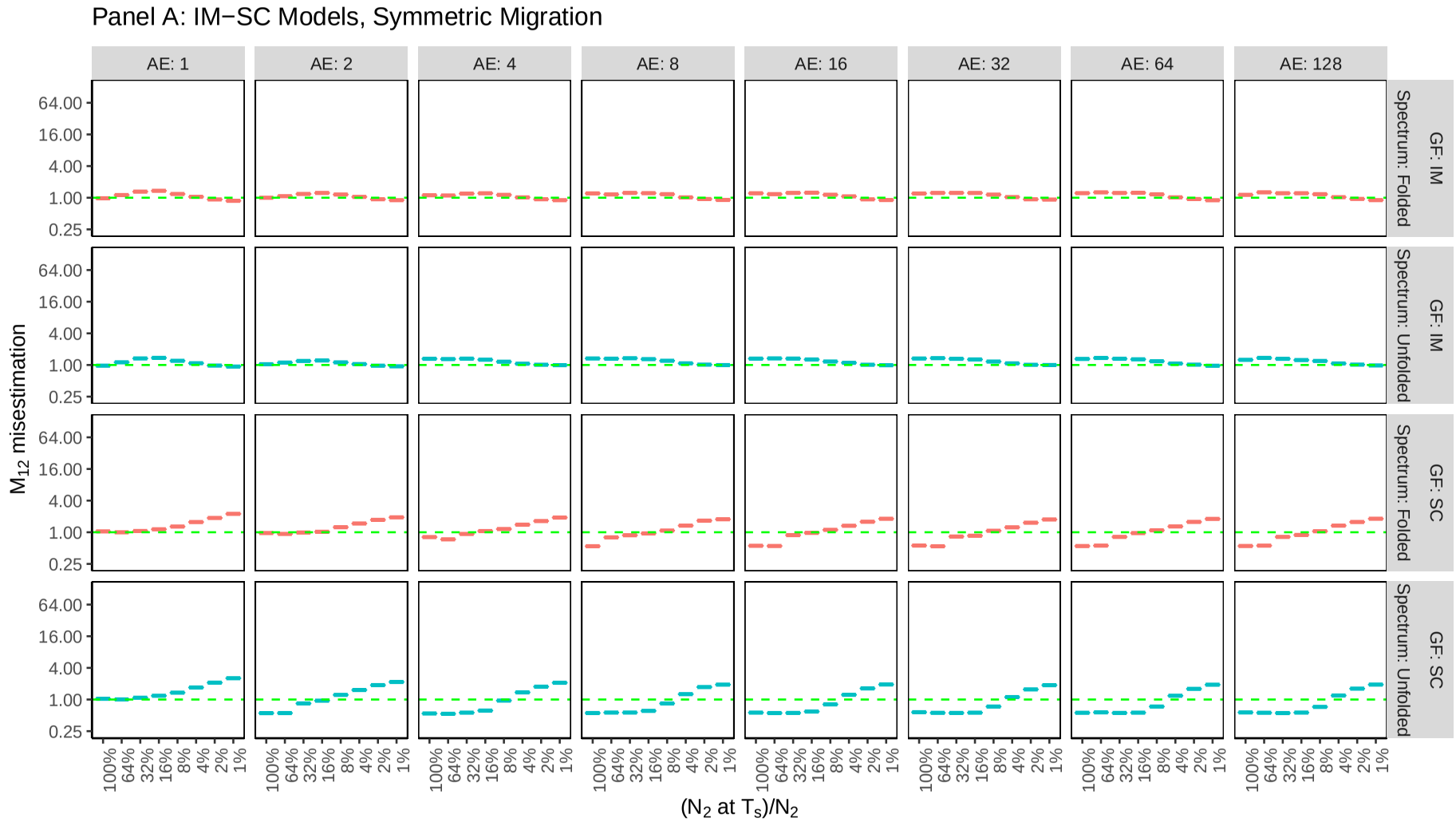


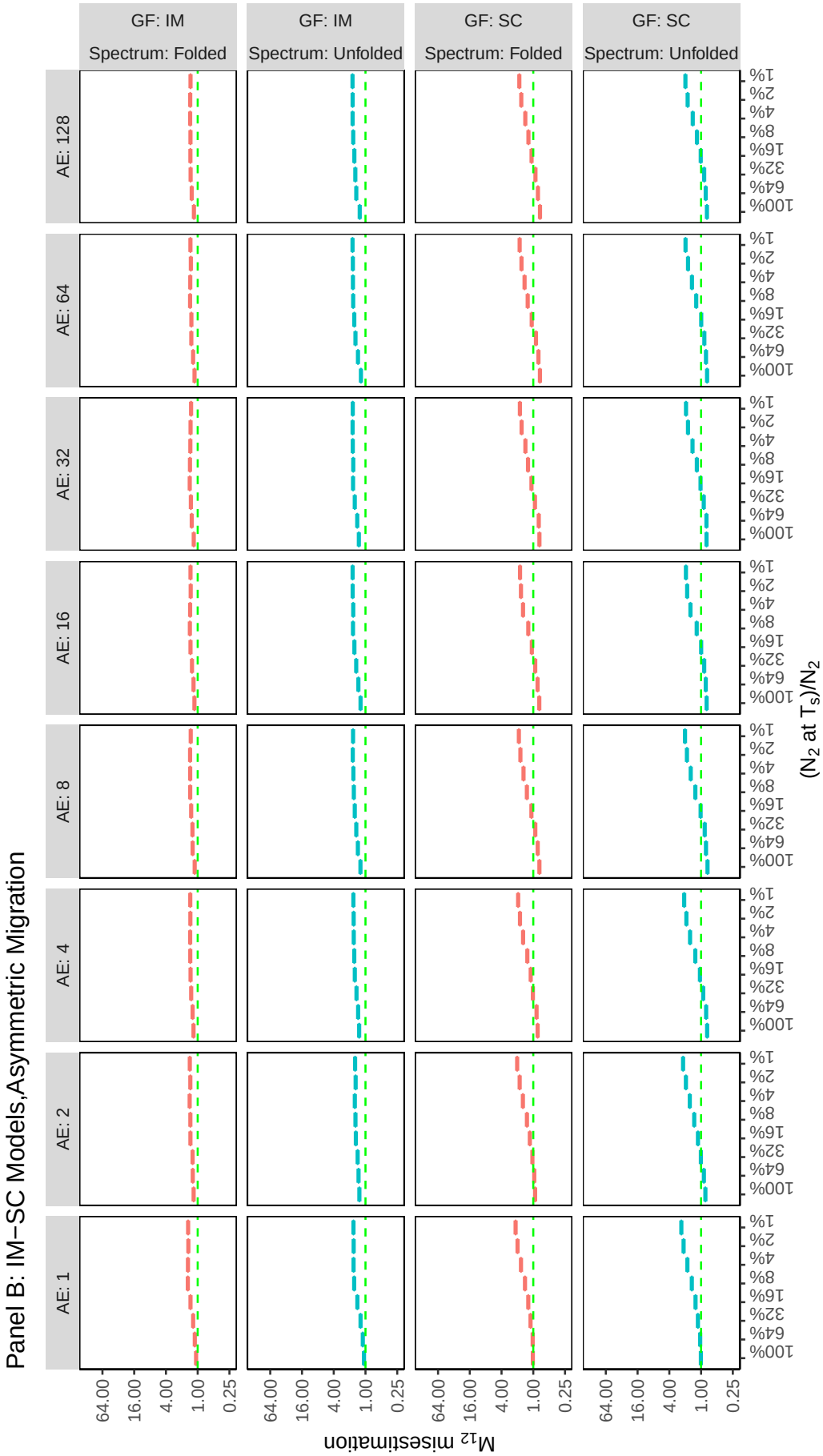

Panel B: IM−SC Models, Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
64.00
16.00
4.00
1.00
0.25
100%
64%
32%
16%
8%
4%
2%
1%
N_{ANC} misestimation
$(N_2 \text{ at } T_s)/N_2$

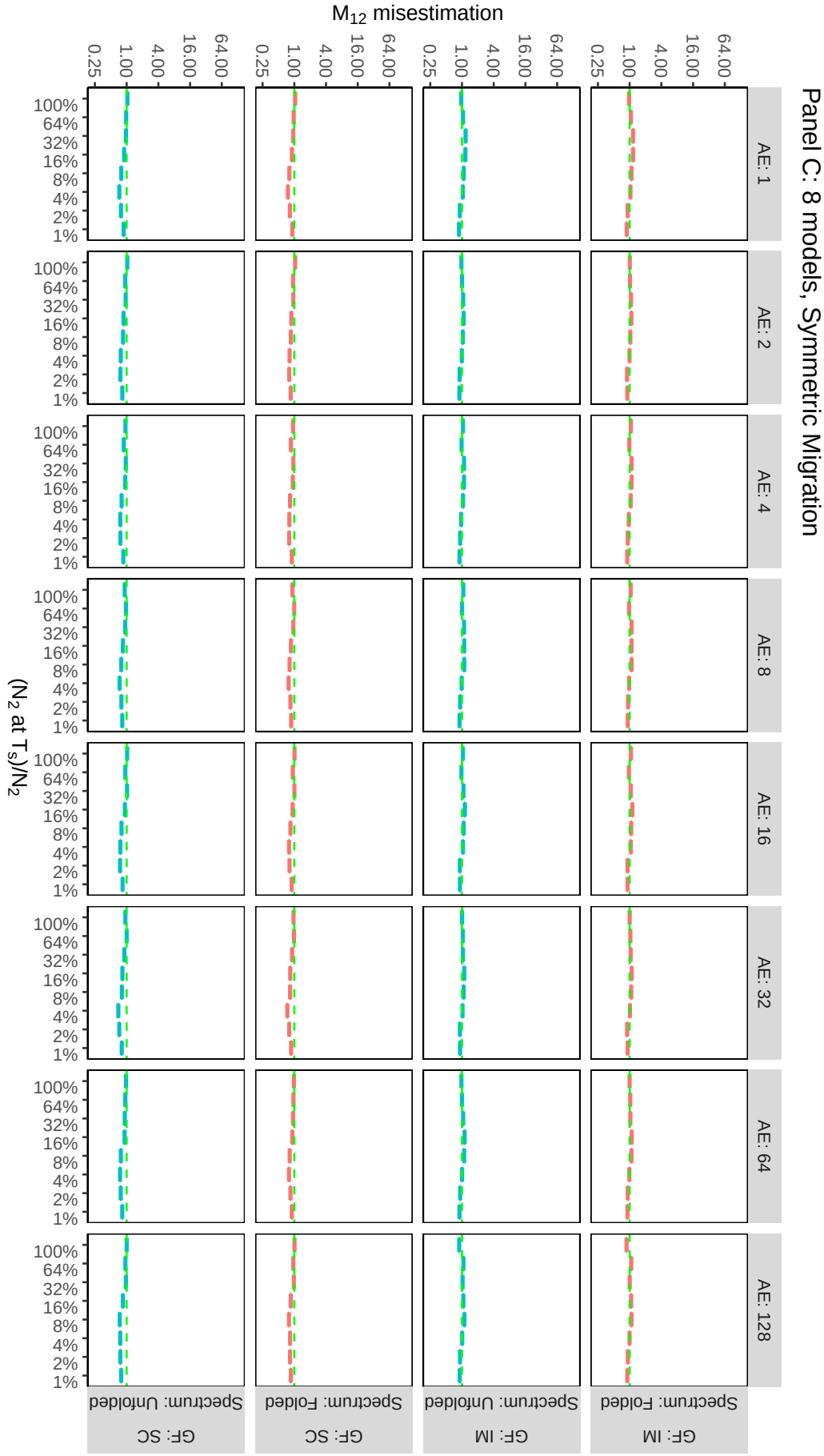

Panel C: 8 models, Symmetric Migration
N_ANC misestimation
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
GF: IM
GF: SC
Spectrum: Folded
Spectrum: Unfolded
0.25
1.00
4.00
16.00
64.00
1%
2%
4%
8%
16%
32%
64%
100%
(N2 at Ts)/N2

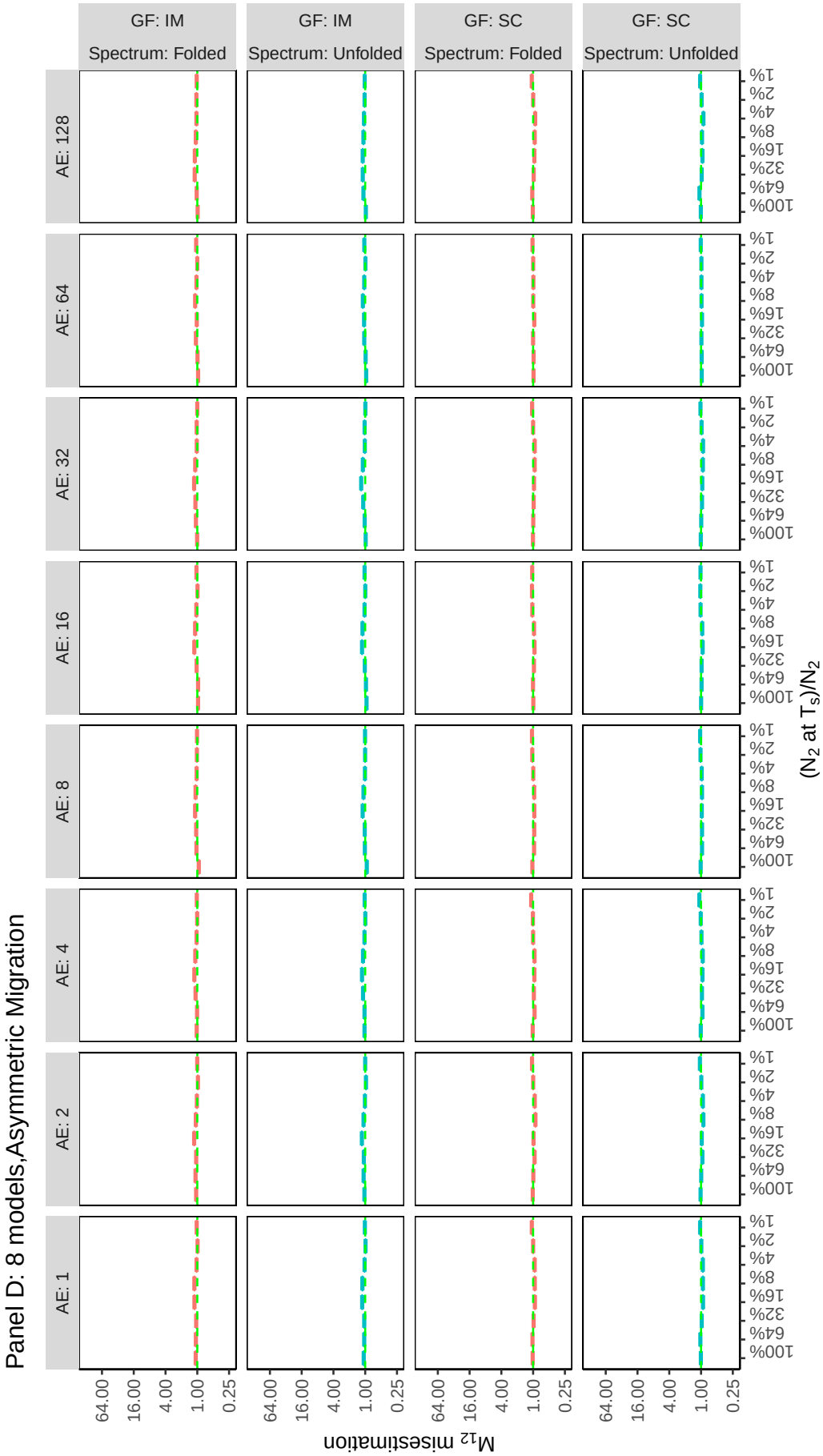

Panel D: 8 models,Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
64.00
16.00
4.00
1.00
0.25
100%
64%
32%
16%
8%
4%
2%
1%
N_{ANC} misestimation
$(N_2 \text{ at } T_s)/N_2$

Supplementary Fig. 13: misestimation of M_{12}. Columns is the grids group by simulation Ancestral Expansion (from 1, i.e. no expansion, to 128 x). The x axes represent the size of population two (N_2) at time of split (T_S) as a proportion of current (N_2). Rows of the plot grids group by gene flow scenario of the simulation (IM and SC) and by whether the unfolded or folded spectrum was used for inference. Panels **A** and **B** show M_{12} misestimation when the basic IM and SC models are used for inference, and under a symmetric and asymmetric migration scenario respectively. Panels **C** and **D** show M_{12} misestimation when the full 8 models (IM,SC,IM_{AE}, SC_{AE}, IM_B, SC_B, IM_{AEB}, SC_{AEB} are used. For each model, M_{12} misestimation was calculated as $M_{12}^{model}/M_{12}^{simulation}$

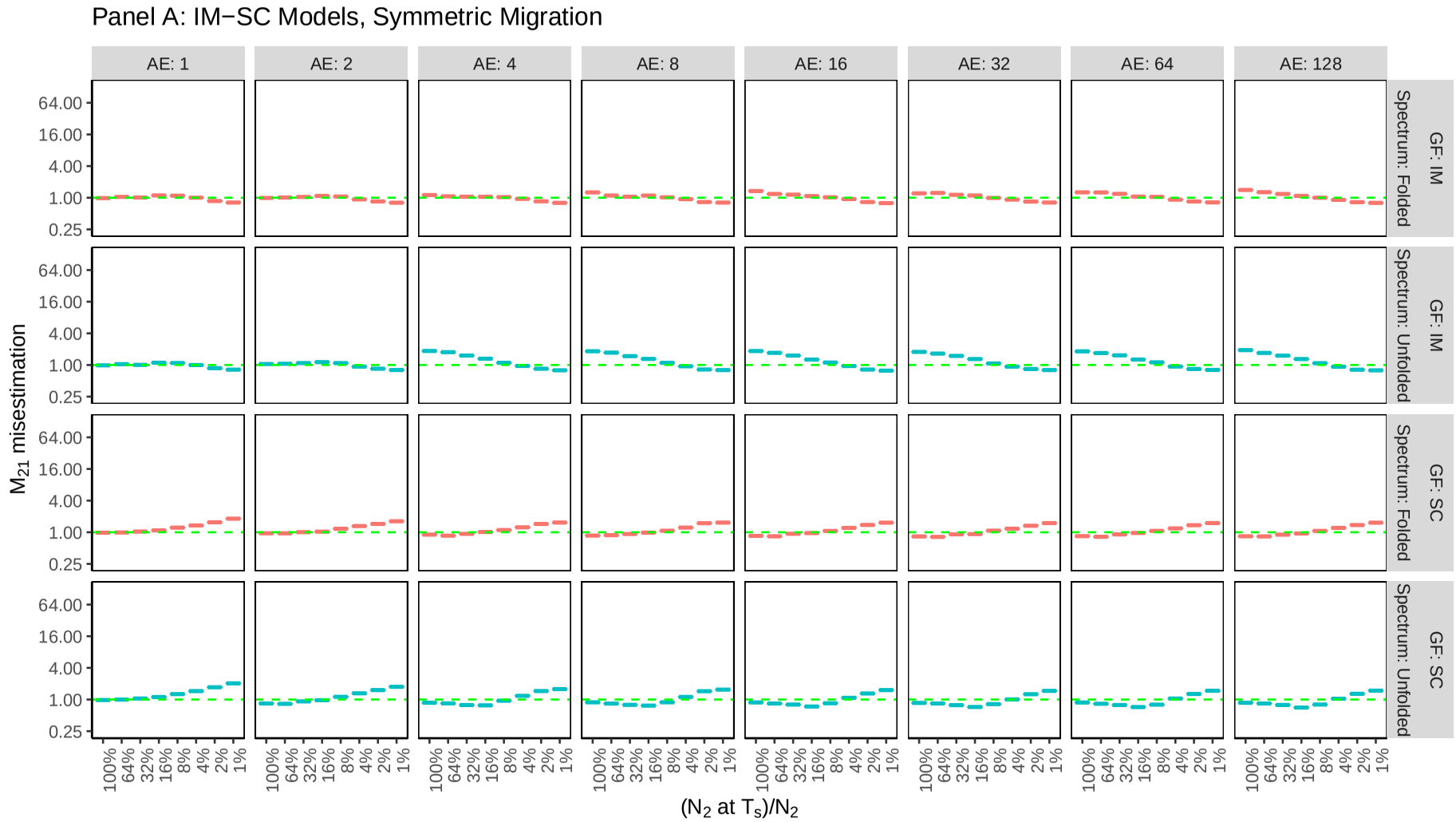


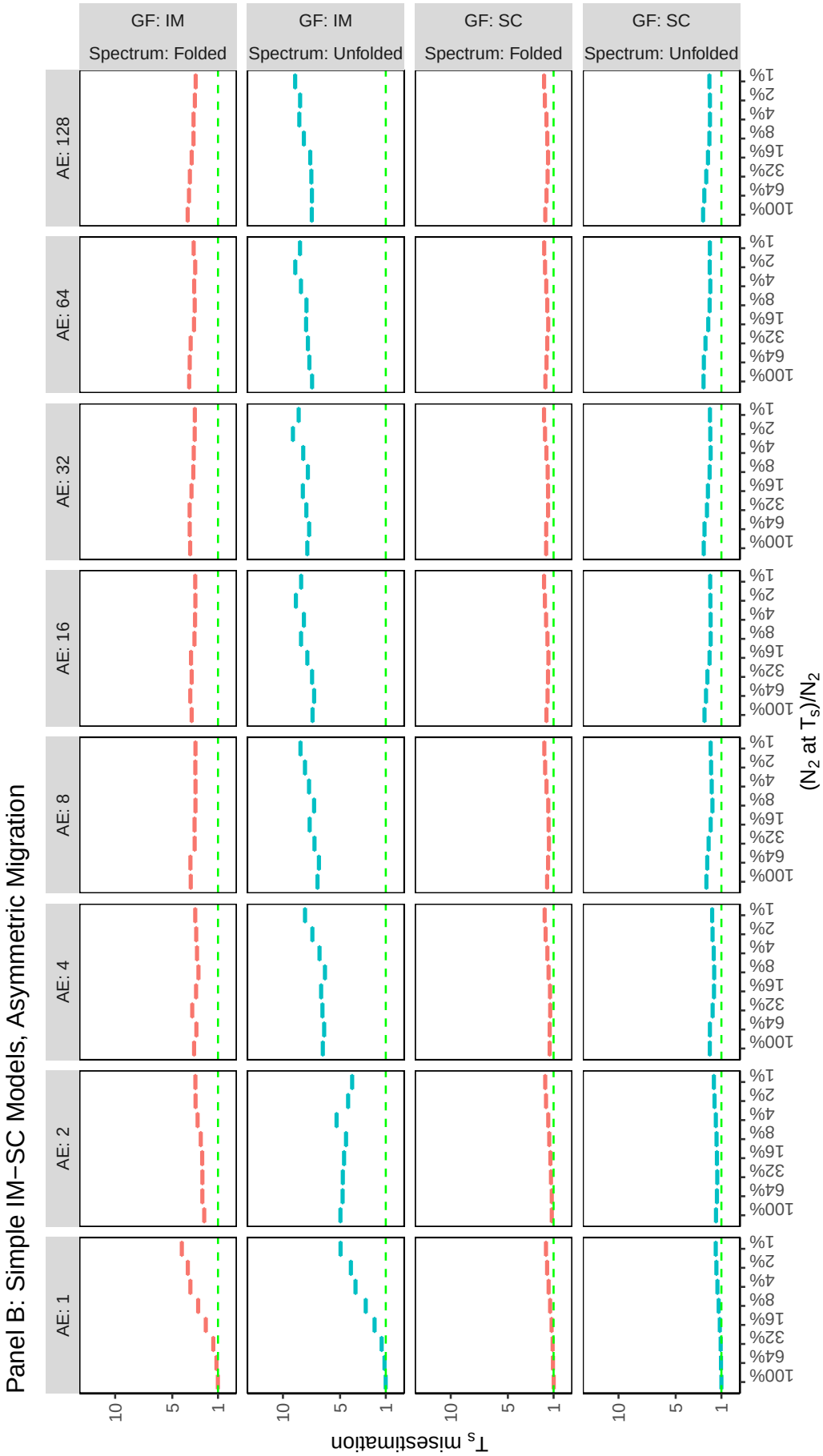

Panel B: IM−SC Models,Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
64.00
16.00
4.00
1.00
0.25
M12 misestimation
100%
64%
32%
16%
8%
4%
2%
1%
(N2 at Ts)/N2

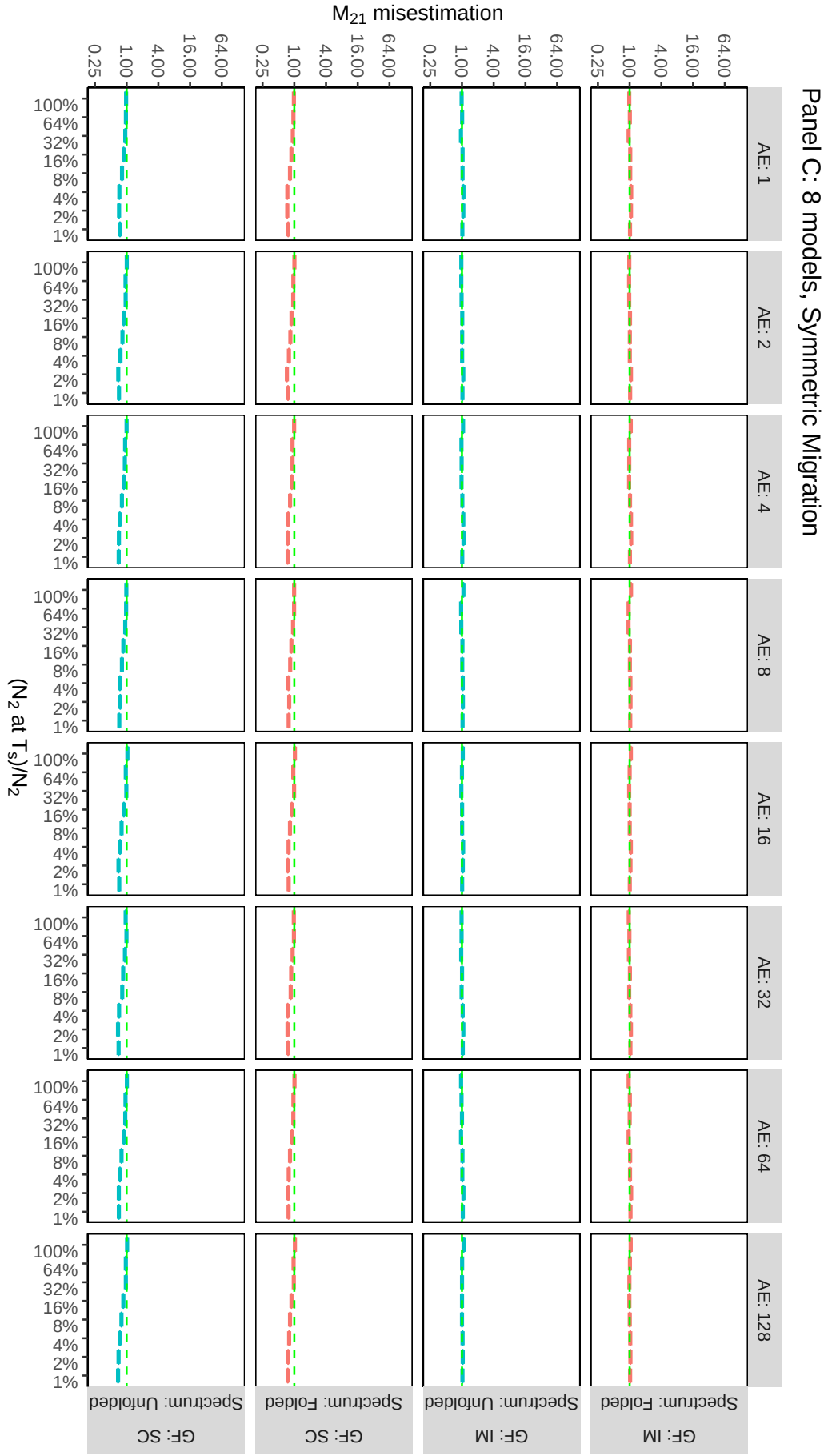

Panel C: 8 models, Symmetric Migration
M12 misestimation
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
GF: IM
GF: SC
Spectrum: Folded
Spectrum: Unfolded
0.25
1.00
4.00
16.00
64.00
1%
2%
4%
8%
16%
32%
64%
100%
(N2 at Ts)/N2

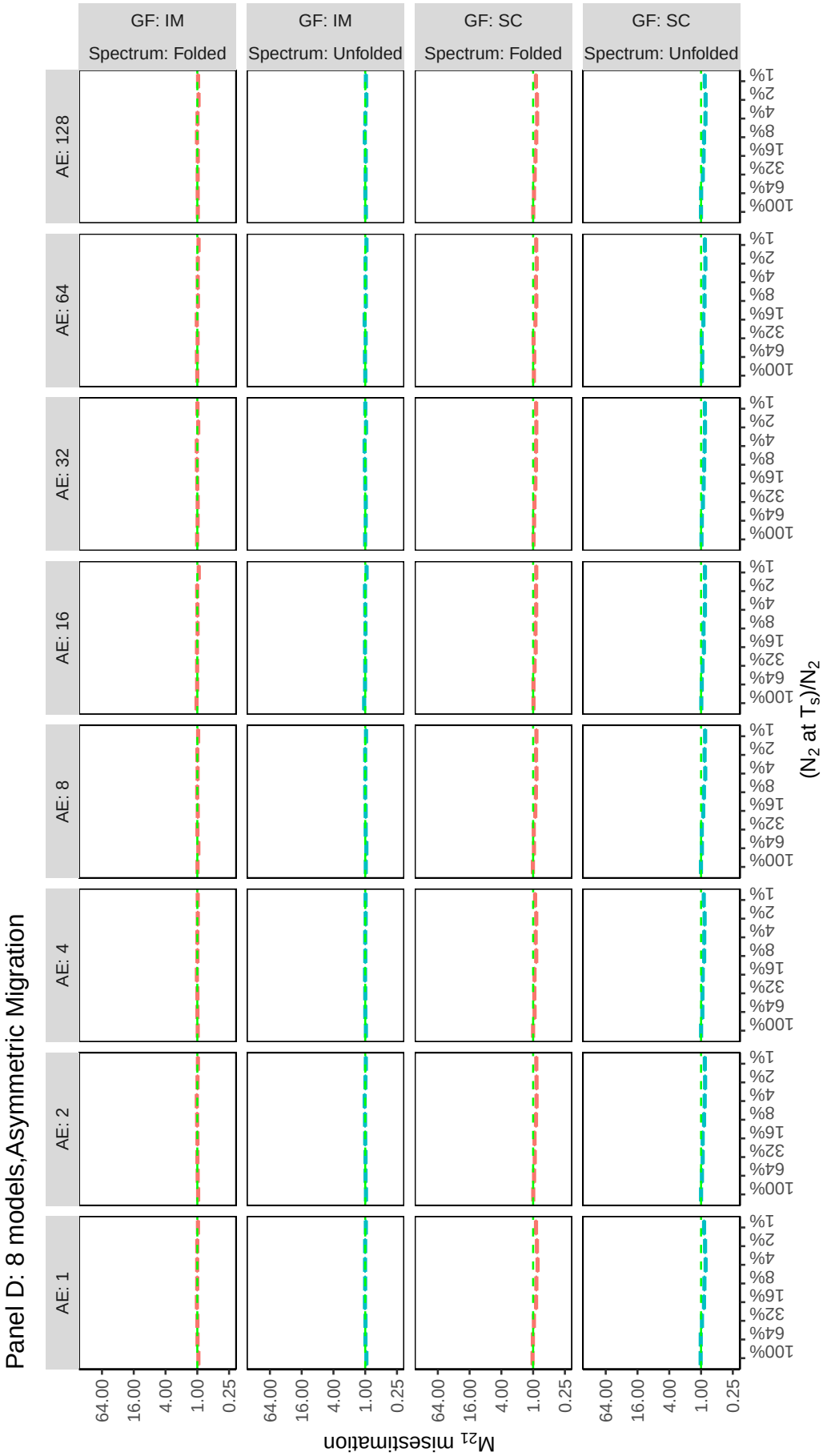

Panel D: 8 models,Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
64.00
16.00
4.00
1.00
0.25
M12 misestimation
100%
64%
32%
16%
8%
4%
2%
1%
(N2 at Ts)/N2

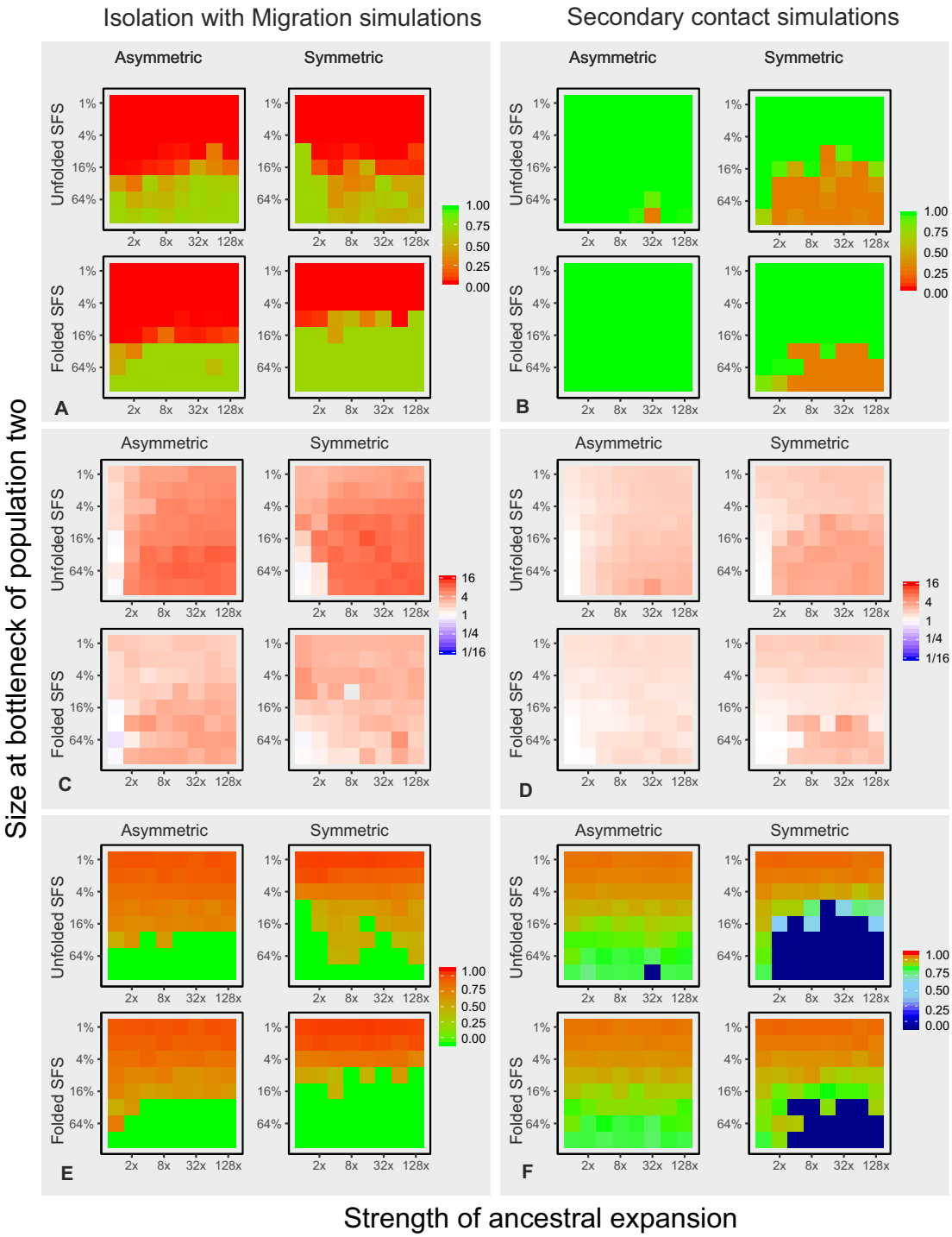
Supplementary Fig. 14: misestimation of M_{21}. Columns is the grids group by simulation Ancestral Expansion (from 1, i.e. no expansion, to 128 x). The x axes represent the size of population two (N_2) at time of split (T_S) as a proportion of current (N_2). Rows of the plot grids group by gene flow scenario of the simulation (IM and SC) and by whether the unfolded or folded spectrum was used for inference. Panels **A** and **B** show M_{21} misestimation when the basic IM and SC models are used for inference, and under a symmetric and asymmetric migration scenario respectively. Panels **C** and **D** show (N_2) misestimation when the full 8 models (IM,SC,IM_{AE}, SC_{AE}, IM_B, SC_B, IM_{AEB}, SC_{AEB} are used. For each model, M_{21} misestimation was calculated as $M_{21}^{model}/M_{21}^{simulation}$



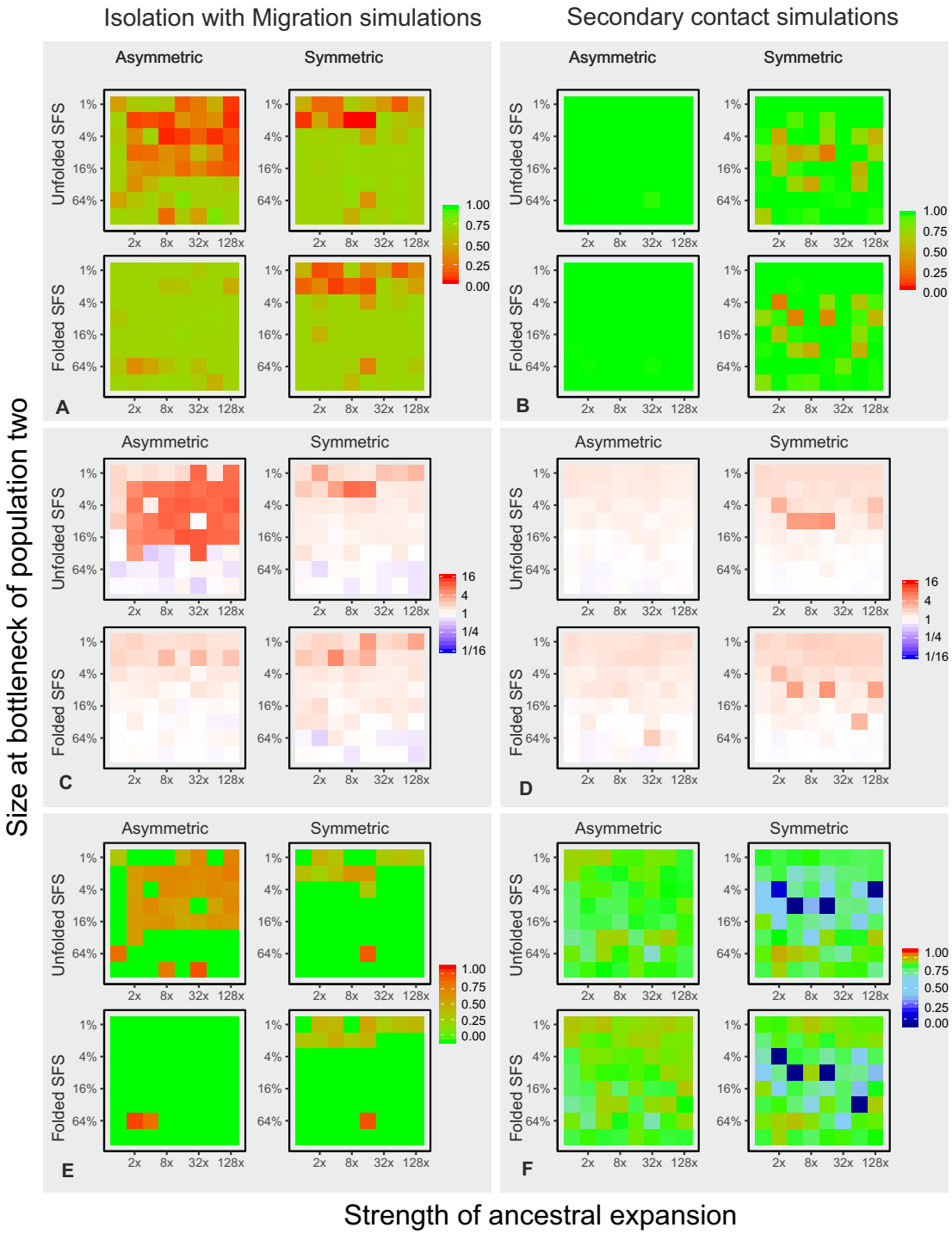

Panel B: Simple IM−SC Models, Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
10
5
1
100%
64%
32%
16%
8%
4%
2%
1%
T_s misestimation
$(N_2 \text{ at } T_s)/N_2$


Panel C: 8 models, Symmetric Migration
M21 misestimation
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
0.25
1.00
4.00
16.00
64.00
100%
64%
32%
16%
8%
4%
2%
1%
(N2 at Ts)/N2
GF: IM
GF: SC
Spectrum: Folded
Spectrum: Unfolded

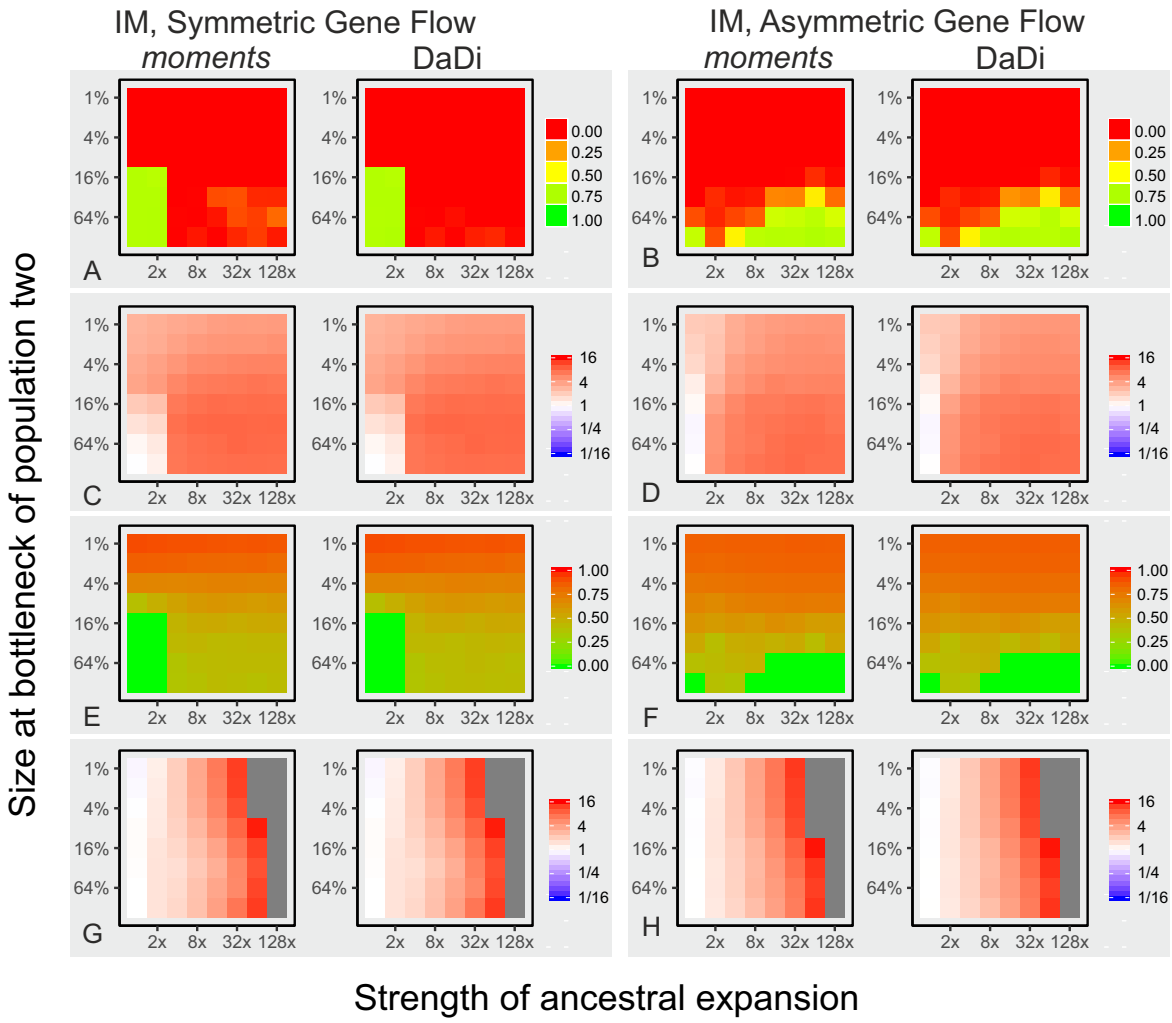

Panel D: 8 models,Asymmetric Migration
GF: IM
Spectrum: Folded
GF: IM
Spectrum: Unfolded
GF: SC
Spectrum: Folded
GF: SC
Spectrum: Unfolded
AE: 1
AE: 2
AE: 4
AE: 8
AE: 16
AE: 32
AE: 64
AE: 128
64.00
16.00
4.00
1.00
0.25
M21 misestimation
100%
64%
32%
16%
8%
4%
2%
1%
(N2 at Ts)/N2



Supplementary Fig. 15: Model choice and parameter misestimation for the simple IM-SC models for all simulations of the smaller data sets (100 000 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 model estimates using the folded or unfolded jAFS. Within each panel, each graph represents the values for all 64 simulations as per Fig. 1. Panel **A** & **B** show weight of evidence for the correct gene flow scenario (0-1). Panel **C** & **D** show misestimation of the parameter T_S as T_S model / T_S 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). This figure is the equivalent of Fig. 2 in the main manuscript, but representing values for the smaller simulated datatsets.



Supplementary Fig. 16: Model choice and parameter misestimation for all 8 divergence models for all simulations of the smaller data sets (100 000 loci). Left panels (**A, C, E**) show results from simulations with constant migration (IM), left 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 model estimates using the folded or unfolded jAFS. Within each panel, each graph represents the values for all 64 simulations as per Fig. 1. Panel **A** & **B** show weight of evidence for the correct gene flow scenario (0-1). Panel **C** & **D** show misestimation of the parameter T_S as T_S model / T_S 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). This figure is the equivalent of Fig. 4 in the main manuscript, but representing values for the smaller simulated data sets.



Supplementary Fig. 17: Comparison of model choice and parameter estimates of the IM and SC simple models in *moments* and DaDi for IM simulations (1 million loci) under a scenario of symmetric (left panels: **A**, **C**, **E**, **G**) and asymmetric (right panels: **B**,**D**,**F**,**H**) gene flow. Panel **A** and **B** show weight of evidence for the correct gene flow scenario (0-1). Panel **C** and **D** show misesistemation of the parameter T_S (calculated as T_S model / T_S simulation) for the best fitting model. Panels **E** and **F** show the estimated proportion of the divergence time for which the best fitting model inferred strict isolation (green represent the correct time, i.e. 0 for IM model). Panels **G** and **H** show the misestimaton of the ancestral population size N_{ANC} (N_{ANC} model / N_{ANC} simulations). Notice that *moments* and DaDi gave nearly identical results. Grey squares represent off-scale values.