G3: Genes|Genomes|Genetics Early Online, published on December 23, 2019 as doi:10.1534/g3.119.400970

Contingent Convergence: The ability to detect convergent genomic evolution is dependent on population size and migration

AUTHORS

James R. Whiting*, Bonnie A. Fraser*

*Department of Biosciences, University of Exeter, Geoffrey Pope Building, Exeter, EX4 4QD

Running Title: Demography, outlier scans & convergence

Keywords: Outlier scans, demographic history, F_{ST}, D_{XY}, false-positives, convergence

***Corresponding Author:** James Whiting
Department of Biosciences
University of Exeter
Geoffrey Pope Building
Exeter
EX4 4QD
j.whiting2@exeter.ac.uk

ABSTRACT

Outlier scans, in which the genome is scanned for signatures of selection, have become a prominent tool in studies of local adaptation, and more recently studies of genetic convergence in natural populations. However, such methods have the potential to be confounded by features of demographic history, such as population size and migration, which are considerably varied across natural populations. In this study, we use forward-simulations to investigate and illustrate how several measures of genetic differentiation commonly used in outlier scans (F_{ST}, D_{XY} and $\Delta\pi$) are influenced by demographic variation across multiple sampling generations. In a factorial design with 16 treatments, we manipulate the presence/absence of founding bottlenecks (N of founding individuals), prolonged bottlenecks (proportional size of diverging population) and migration rate between two populations with ancestral and diverged phenotypic optima. Our results illustrate known constraints of individual measures associated with reduced population size and a lack of migration; but notably we demonstrate how relationships between measures are similarly dependent on these features of demography. We find that false-positive signals of convergent evolution (the same simulated outliers detected in independent treatments) are attainable as a product of similar population size and migration treatments (particularly for D_{XY}), and that outliers across different measures (for e.g. F_{ST} and D_{XY}) can occur with little influence of selection. Taken together, we show how underappreciated, yet quantifiable measures of demographic history can influence commonly employed methods for detecting selection.

INTRODUCTION

Studies assessing adaptation and evolution across the genome are increasing in popularity with the availability of modern sequencing technologies. These studies often centre around quantifying patterns of variation in genome-wide SNPs, which can be used to highlight regions or genes having experienced selection relative to the neutral backdrop of the rest of the genome. These analyses, which we refer to here as outlier scans, have become a common tool in population genetics and have been useful across diverse, natural systems in identifying candidate genes associated with the evolution of a range of adaptive traits (reviewed recently by (Ahrens *et al.* 2018)). More recently, the method of overlapping outlier scans across independent lineages has been employed to test whether the same regions are involved in independent adaptation events (i.e. genetic convergent evolution) (reviewed by (Fraser and Whiting 2019)). This study seeks to investigate how different outlier scan methods are influenced by demographic variation in natural populations, and how this may lead to overlapping false-positives.

Recent discussions have highlighted the propensity of outlier scans to yield false-positives, given that outliers caused by heterogeneous genomic landscapes are commonplace irrespective of selection (Ellegren and Wolf 2017). For example, background selection (BGS), whereby linkage between neutral and deleterious variants reduces local diversity (Charlesworth *et al.* 1993; Bank *et al.* 2014; Burri 2017), has been invoked to offer alternative explanations for patterns at first attributed to directional selection (Cruickshank and Hahn 2014). Furthermore, the influence of neutral processes such as genetic drift (Charlesworth 2009) can similarly produce elevated genetic divergence and false-positive signatures of selection.

69 The strength of these processes is dependent on demography. For example, the influence of
70 neutral processes and BGS should be more pronounced in smaller populations (low N_e)
71 (Charlesworth *et al.* 1995; Charlesworth 2009; Yeaman and Otto 2011; Cutter and Payseur
72 2013); as has been demonstrated in humans (Torres *et al.* 2018) and *Drosophila*
73 (Charlesworth 1996; Sella *et al.* 2009). This relationship, however, may not be simply linear,
74 with additional simulation evidence suggesting the effects of BGS are strongest at
75 intermediate-low N_e, and weaker at very low or high N_e (Zeng 2013).
76
77 Lack of connectivity among populations may also elevate measures of genetic
78 differentiation, if a large global population is composed of smaller, isolated, sub-divided
79 populations that each experience the effects of reduced N_e (Charlesworth *et al.* 1997;
80 Hoban *et al.* 2016). In an attempt to mitigate the likelihood of false-positives, some have
81 advocated using multiple measures of population differentiation and divergence to identify
82 regions of the genome likely to be under selection (Charlesworth 1998; Cruickshank and
83 Hahn 2014). However, how these measures are correlated with each other, and with
84 selection under different demographic scenarios, has not been explored.
85
86 A diverse array of measures of genetic differentiation and divergence have been employed
87 for outlier scans, but here we focus on two of the most common measures of relative
88 differentiation (F_{ST} and changes in nucleotide diversity [$\Delta\pi$]) and a measure of absolute
89 divergence (D_{XY}). The fixation-index F_{ST} (Weir and Cockerham 1984; Hudson *et al.* 1992),
90 which measures the relative amount of within- and between-population variance. This
91 measure is therefore maximised when genomic regions exhibit the lowest within- and
92 highest between-population variance. F_{ST} outliers at the right-tail of the distribution are

93 considered candidates for adaptation because they reflect regions with large differences in
94 allele frequency or high substitution rate (large between-population variance) and/or low
95 nucleotide variation (low within-population variance) relative to the rest of the genome.
96 Changes in nucleotide diversity ($\Delta\pi$) are another indicator of adaptation, as selection on a
97 beneficial allele limits variation within a population resulting in selective sweeps (Smith and
98 Haigh 1974). Comparisons of the ratio of π between diverging populations reveal regions
99 under selection, as local π is reduced in one population in comparison to the other. Whilst
100 similar to F_{ST}, $\Delta\pi$ does not discriminate between which copy of a polymorphism is fixed,
101 such that a substitution between populations is equivalent to a common non-polymorphic
102 site. $\Delta\pi$ outliers therefore represent regions of the genome with reduced π in either
103 population relative to the rest of the genome. As a measure of absolute genetic divergence,
104 D_{XY} (Nei 1987) does not consider the relative frequencies of polymorphisms within
105 populations (Charlesworth 1998; Cruickshank and Hahn 2014). D_{XY} can be quantified as the
106 average number of pairwise differences between sequence comparisons between two
107 populations. This measure is therefore influenced by ancestral π and the substitution rate,
108 so D_{XY} outliers highlight regions that are highly variable ancestrally, or in either population
109 (large π), or exhibit many substitutions and thus increased sequence divergence.
110
111 Because each measure of differentiation/divergence (hereafter referred to collectively as
112 measures of divergence) quantifies genetic variation between populations in a slightly
113 different way, each has a unique relationship with demography. Whilst we can predict how
114 individual measures are influenced by demography, and subsequently neutral processes, we
115 know little about how different relationships between measures of divergence and

demography affect their combined usage. We expect then that the utility of using multiple measures of divergence to detect selection may vary with demography.

This complex interplay between divergence measures and demography may be further exacerbated in studies that compare scans from multiple populations to identify convergent genomic evolution. Here, researchers use measures of population divergence across independent pairs of evolutionary replicates with outlier loci compared across results. This strategy has been employed extensively across diverse taxa, including: birds (Cooper and Uy 2017), fish (Hohenlohe *et al.* 2010; Jones *et al.* 2012a; Fraser *et al.* 2015; Reid *et al.* 2016; Rougemont *et al.* 2017; Meier *et al.* 2018), insects (Soria-Carrasco *et al.* 2014; Van Belleghem *et al.* 2018), mammals (Waterhouse *et al.* 2018), molluscs (Westram *et al.* 2014; Ravinet *et al.* 2016) and plants (Roda *et al.* 2013; Trucchi *et al.* 2017). For the sake of consistency, it is common to infer outlier loci within each replicate pair through a common method, but if replicates differ in their demographic histories then the applicability/power of that common method will also vary accordingly. This begs the question, can demographic variation among replicates alone explain signals of, or lack of, convergence?

Here, we used forward-simulations to investigate the effects of different demographic histories on the relationships between measures of divergence with selection. We varied demography through manipulating the number of founding individuals (founding bottlenecks), the population size of the diverging population (prolonged bottlenecks) and the presence/absence of migration. We simulated the effects of selection on 25kb regions, each with a gene designed from features taken from the guppy (*Poecilia reticulata*) genome assembly, a prominent model system for studies of convergent evolution (Reznick and

Endler 1982; Fraser *et al.* 2015). Moreover, we measured genetic divergence at 12 set time points through F_{ST}, D_{XY} and $\Delta\pi$ between diverging populations to examine temporal relationships. Our aim was to investigate the significance of founding/prolonged bottlenecks and migration between populations when employing outlier scans, and highlight demographic scenarios that are susceptible to false-positives. We also aimed to test the occurrence of common outliers across measures, and whether overlapping outliers are consistently good indicators of selection. We sought to answer the following questions: 1) How do these demographic factors influence measures of divergence through time? 2) How well do measures of divergence identify regions of the genome under strong selection through time and across demographic factors? 3) How are the different measures of divergence related through time and across demographic factors? 4) Where do we detect the strongest signals of convergence (i.e. overlapping outliers using single or multiple measures), and are these consistent with selection?

METHODS

The forward-simulation software SLiM 3.0 (Haller and Messer 2019) was used to simulate population divergence under contrasting demographic treatments in a fully factorial design between a population with an ancestral phenotype (AP) of N = 1000 and population with a diverged phenotype (DP), with a mutation rate based on the guppy genome ($4.89e^{-8}$) (Künstner *et al.* 2016) and scaled 100-fold over three raw mutation rates ($4.89e^{-5}$, $4.89e^{-6}$, $4.89e^{-7}$) to ensure robustness of results across different, more realistic values of θ ($4N_e\mu$). These scaled mutation rates therefore represent effective population sizes of 10-, 100-, and 1000-times greater than the number of individuals in our simulations (N = 1000), in line with estimates from other species (Charlesworth 2009). The main text reflects results for the

intermediate mutation rate $4.89e^{-6}$, with others presented in supplementary figures. SLiM employs a classic Wright-Fisher model to simulate populations, in which a population of diploid hermaphrodites proceeds through generations such that an individual's contribution towards the next generation is proportional to its relative fitness.

Demographic treatments included reducing DP size (prolonged bottlenecks) relative to the AP size (N = 1000) (0.01, 0.1, 0.5, 1.0), migration as a proportion of individuals exchanged between populations (0.0, 0.002) and founding bottlenecks as the number of individuals sampled from the burn-in population to construct DP genomes (N = 100 or 1000). For example, a demographic history with founding bottleneck = 100 individuals, DP size = 0.01, and migration = 0.002 would represent the following scenario: 1) At generation 1, following a burn-in period of 10,000 generations to reach mutation-selection balance, populations split as 100 genomes are sampled from the burn-in population of 1000 to form DP, whilst AP is formed by sampling all 1000 burn-in genomes; 2) At the next generation (and for the remainder of the simulation), DP size reflects the prolonged bottleneck treatment, in this case 10 (0.01 of 1000); 3) AP and DP experience migration in both directions for the remainder of the simulation. Thus, each of our treatments can be thought of as a manipulation of the following: Founding bottlenecks limit the amount of variation within the burn-in population that is available to found DP; Prolonged bottlenecks limit the size of DP, reducing mutational input and moderating the strength of neutral evolution and efficacy of selection; and migration dictates the presence of migration between AP and DP (Figure 1). The total N of individuals (AP + DP) within simulations varies between 1010 and 2000 depending on DP size parameter, however this potential expansion does not impact our results (Supporting Information).

Combining these treatment levels in a fully factorial experimental design generated a total of 16 different, independent demographic histories that all 25kb gene regions (see below for architecture) experienced. This factorial design allows us to examine the relative influence of founding bottlenecks, prolonged bottlenecks and migration, but our study is limited to these features of demographic history and does not extend to population expansions, more complex cases of migration, or demographic fluctuations through time. In some cases, the influence of prolonged bottlenecks renders the effects of founding bottlenecks unnecessary. However, when DP size is greater than 100, the inclusion of the founding bottleneck allows us to compare populations with equivalent mutational input but different standing variation.

SLiM runs were performed independently over 25kb genomic regions with a central 'gene' that functioned as a QTL and varied in length and exon content (Figure 1E); with exon number, lengths, and intron lengths drawn at random from the guppy genome gene annotation file (gff). In total, we simulated a dataset of 100 25kb regions (Figure 1F). Recombination rate was scaled alongside mutation rate from a human-derived $r = 1e^{-8}$ to $1e^{-6}$.

Selection (*S*) in our model was represented as the fitness consequences incurred through distance from a phenotypic optimum in a one-dimensional fitness landscape. Selection in Wright-Fisher models is soft, in that low fitness individuals are not removed but have a lower likelihood of contributing to the next generation. Phenotypic optima were maintained over the course of simulations; thus, selection was constant throughout. This setup can be

considered as analogous to two environments with contrasting optimum trait values for a single trait. Intensity of selection was manipulated by modifying the standard deviation (S_σ) of the normal distribution curve from which the density distribution was calculated through the "dnorm" function in SLiM. Values for *S* were drawn from a continuous distribution between -1.00 and 1.00 and transformed such that $S_\sigma = 10^{-S}$, yielding values between 0.1 and 10, with 0.1 representing the steepest fitness peak (S = 1) and 10 the shallowest (S = -1) (Figure 1A). Phenotypes were calculated per individual, per region, as the additive phenotypic effects of exonic non-synonymous mutations, which appeared at a rate of 7/3 relative to synonymous mutations (assuming that most mutations in the third base of a codon do not alter the amino acid). Additive genetic variance was assessed due to its prevalence in complex traits in nature (Hill *et al.* 2008). Effect sizes for mutations with phenotypic effects were drawn from a Gaussian distribution with mean = 0 and σ = 1. The remaining synonymous exon mutations, and mutations in introns and outside of genes, had no effect on fitness.

For each simulation, populations were seeded with 1000 individuals and allowed to proceed for a burn-in period of 5*2N (10,000) generations to reach mutation-selection-migration balance. During this period, burn-in populations evolved towards the ancestral phenotypic optimum of 0, defined as a normal distribution with mean = 0 and σ = 1.0 (Figure 1A). Burn-in populations were then subjected to each demographic treatment to simulate the founding of multiple populations from a shared ancestral state. During this 'divergence period', AP continued to evolve around the ancestral optimum of 0, whilst DP's phenotypic optimum was centred around 10, with fitness consequences defined according to S_σ. Individuals also experienced fitness costs associated with phenotypic proximity to other

individuals within the population as a proxy for competition and to ensure a realistic amount of phenotypic variation persisted within populations. Fitness costs due to competitive proximity were scaled to a maximum value of 1 with σ = 0.4 and occurred reciprocally between local individuals with phenotypes with a difference of ≤ 1.2 (3 * 0.4).

Simulations were sampled at 100, 500, 1000, and then every 1000 generations up to 10,000 (Figure 1C). F_{ST} was calculated across the 25kb region based on the proportion of subpopulation heterozygosity (H_S) relative to total heterozygosity (H_T) (according to Hudson, Slatkin and Maddison (1992)): $F_{ST} = 1 - \frac{H_S}{H_T}$. D_{XY} was calculated as the sum of nucleotide differences (d_{ij}) between the i^{th} haplotype from AP and the j^{th} haplotype from DP (according to Nei (1987)): $D_{XY} = \sum_{ij} AP_i DP_j d_{ij}$. Mean heterozygosity (π) for each population was calculated as a single measure across the 25kb region. At each sampling point, each measure was calculated and averaged across the preceding 20 generations. Averaging was performed such that measures would not be dramatically biased by events occurring within individual generations. The change in mean heterozygosity between AP and DP (Δπ) was calculated as the ratio of π_{AP} to π_{DP}, such that reduced diversity in the DP population increases the value of Δπ: $\Delta\pi = \log_{10} \frac{\pi_{AP}}{\pi_{DP}}$. Statistics were calculated over all monomorphic and polymorphic sites of 25kb regions. This has been designed to replicate genome scans that use window-approaches, with each 25kb region analogous to an independent window. In total, 100 unique 25kb regions were simulated across 16 demographic treatments across three mutation rates, with results for the intermediate mutation rate presented in the main text. To account for stochastic noise in the simulation, each 25kb region was iterated 20 times for each demographic treatment. Simulations with divergent AP and DP phenotypes

are referred to as "$Pheno_{Div}$" simulations. Additional simulations were performed in which both populations shared the ancestral phenotype ("$Pheno_{Null}$" simulations) and in which all sites were neutral with no selection imposed ("Neutral" simulations). The former of these was used to assess whether patterns associated with selection were driven by phenotypic divergence or variable stabilising selection within populations, whilst the latter was used to disentangle effects of selection and neutral processes such as drift. A common set of 100 25kb regions were used for all simulations. Outliers for simulations were taken as upper 95% quantiles.

All data analysis was performed in R (3.5) (R Core Team 2016). To assess relationships between divergence measures and selection, data were grouped by sampling generation within each treatment group (N = 16) and Pearson's correlation coefficients were calculated between measures of divergence and selection (*S*) at each sampling point for all regions (N = 100). Correlation coefficients were then grouped within specific treatment levels (e.g. DP size = 0.5, or migration = 0.002) and averaged to give a coefficient reflecting each specific treatment level. These correlation coefficients were calculated for each iteration (N = 20) and averaged over to give final values.

To assess the effects of treatments on detecting outliers, we compared distributions and 95% cut-offs within each treatment for each measure of divergence for $Pheno_{Div}$, $Pheno_{Null}$, and Neutral simulations. We limited this analysis to early (100, 500), intermediate (3000) and late (10,000) sampling generations. Here, data from all iterations of each 25kb region were pooled. To calculate false positive (FPR) and false negative rates (FNR), we pooled Neutral simulation data within treatment groups (20 iterations of 100 genes, N = 2000),

removed a random set of iterations (N = 100), and replaced it with an assortment of random iterations of each gene (e.g. Gene 1/Iteration 2, Gene 2/Iteration 14... Gene 100/Iteration 4) from $Pheno_{Div}$ data. We calculated FPR as the proportion of data above the 95% quantile for each measure of divergence/differentiation that came from neutral simulations. For single measures, FPR = FNR as 5% of data in each permutation comes from $Pheno_{Div}$. We combined outlier sets across all combinations of F_{ST}, D_{XY} and $\Delta\pi$ and examined neutral proportions within outlier sets to determine FPR. FNR for combined outlier sets corresponded to the proportion of $Pheno_{Div}$ data not recovered in the combined outlier sets. These permutations were performed 100 times with results averaged. Proportional overlap of outlier sets was also calculated and compared across demographic treatment groups to examine convergence of results across treatments. Overlap was calculated within each permutation, averaged over, and visualised using heatmaps with hierarchical clustering of axes.

To examine how simulated gene features influenced patterns of genetic divergence, we used a linear mixed modelling (LMM) approach with gene ID and treatment group as random factors. Gene features modelled as independent factors were: number of exons (Exon N), gene size, the proportion of gene that is coding (selection target %), selection applied to each gene (*S*), and the generation at which the optimum phenotype was reached (Pheno Gen).

DATA AVAILABILITY

A full set of scripts including all bash, Eidos and R scripts necessary to repeat this analysis can be downloaded from Github (https://github.com/JimWhiting91/Contingent_Convergence_Pipeline). Supplementary

figures (S1-49) have been uploaded through the GSA figshare portal. Supplementary figures include results for different mutation/recombination rates, $Pheno_{Null}$ and neutral data across sampling generations along with additional figures.

RESULTS

Demography modifies measures of divergence

Founding bottlenecks, simulated by reducing the number of possible founding genomes to a random 10% of the ancestral population, had little measurable effect on F_{ST}, D_{XY} or $\Delta\pi$, producing minimal variance between treatments over all sampling points (Figure 2A).

Prolonged bottlenecks, simulated by modifying the stable number of individuals within DP, had pronounced and variable effects on all measures. F_{ST} increased with reductions in DP size. This effect was generally consistent through time, although variance between treatments increased gradually through time (Figure 2B). A similar effect was observed for $\Delta\pi$; however, this measure was particularly susceptible to inflation under the most extreme reductions in DP size, with substantially more elevated values observed between DP size proportions of 0.01 and 0.1 compared with 0.1 and either 0.5 or 1.0. This pattern was broadly consistent through sampling points. Such observations are unsurprising given that both F_{ST} and $\Delta\pi$ increase with processes that reduce within-population genetic variance. D_{XY} was generally robust to prolonged bottlenecks, but in contrast to F_{ST} and $\Delta\pi$, D_{XY} decreased when DP sizes were reduced (Figure 2B).

The inclusion of migration reduced absolute values of all measures of divergence. For F_{ST} and D_{XY}, variance between migration treatments increased across the simulation, however

331 D_{XY} was generally more consistent across migration treatments. Δπ was also reduced in the
332 presence of migration, although the effect of migration on Δπ was generally consistent
333 across all generations and did not increase through time, as was the case for F_{ST} and D_{XY}.
334
335 Whilst F_{ST} and D_{XY} increased generally over time, Δπ peaked around generation 500 and
336 declined thereafter. This peak corresponded to the median generation that DP replicates
337 reached their optimum phenotype (median across all data = 493), suggesting this peak
338 reflects selective sweeps. The majority of these patterns were apparent in the $Pheno_{Null}$
339 (Figure S1) and neutral (Figure S2) simulations, however divergence for all measures was
340 reduced and ultimately negligible by the removal of divergent phenotypes when migration
341 was present.
342
343 ***Demography moderates the association between measures of divergence and selection***
344 Again, founding bottlenecks had a minimal effect on the correlations observed between
345 strength of selection and measures of divergence with minimal variance observed between
346 bottleneck treatments in all comparisons (Figure 3A).
347
348 Prolonged bottlenecks had substantial effects on relative (F_{ST} and Δπ) measures and
349 marginal effects on absolute (D_{XY}) measures of divergence (Figure 3B). F_{ST} correlations with
350 selection became consistently weaker as DP size reduced. There was minimal difference
351 between Δπ correlations with selection except for the most extreme reductions in DP size.
352 Effects for both were generally consistent through time. For D_{XY}, correlations with selection
353 were broadly consistent with minimal variance across DP size reductions (Figure 3B).
354

Expectedly, the absence of migration largely precluded the ability of measures of divergence to predict strength of selection, with notable variance observed between no migration (0.0) and minimal migration (0.002) treatments emerging rapidly for all divergence measures (Figure 3C). Both F_{ST} and D_{XY} variance between migration treatments was greatest at the 10,000 generations sampling point, whereas similar variance was observed between migration treatments for $\Delta\pi$ across simulations. This observation again highlights the significance of temporal differences between measures. Interestingly, correlations between $\Delta\pi$ and selection persisted, albeit weakly, in the absence of migration, which was not the case for F_{ST} and D_{XY}. Further, at larger population scaling ($\mu = 4.89e^{-5}$, $r = 1e^{-5}$) D_{XY} correlations with selection were negative (although became more positive over time with migration) (Figure S6), most likely due to a stronger influence of mutational input. In larger populations, positive correlations were observed between F_{ST} and selection without migration, but were weaker than with migration (Figure S6).

Prolonged bottlenecks had a minimal effect on how measures of divergence correlated with selection in $Pheno_{Null}$ data (Figure S5). Both F_{ST} and D_{XY} became negatively correlated with selection over time without divergent selection, likely due to stronger selection on common alleles shared between AP and DP. Negative correlations were stronger for D_{XY}, consistent with reductions in DP_{π} with stronger selection. This suggests positive associations between selection and D_{XY} in $Pheno_{Div}$ simulations are likely more dependent on adaptive substitutions in order to overcome this effect. $\Delta\pi$ was generally positively associated with selection in $Pheno_{Null}$ simulations regardless of migration, but was slightly reduced when migration was absent.

By examining the correlation coefficients of all 16 unique demographic histories, we can investigate the combined effect of migration and prolonged bottlenecks and directly compare effectiveness of individual measures across time (Figure 4). F_{ST} consistently outperforms D_{XY} in terms of associating with selection under most demographic treatments, particularly when DP sizes are larger and migration is present. By 10,000 generations however, the relative dominance of F_{ST} appears to subside, with the trend through time suggesting a relative improvement in D_{XY} in treatments with migration (Figure 4). Δπ is similarly more informative than D_{XY} across sampling generations under most demographic treatments. Interestingly at sampling generation 10,000, reductions in prolonged bottlenecks produce the biggest bias towards Δπ (Figure 4). The resilience of Δπ under no-migration treatments is also apparent in F_{ST} - Δπ comparisons, such that at 3,000 generations Δπ is more informative than F_{ST} in the absence of migration. Consistent with its rapid response to sweeps around 500 generation, Δπ slightly outperforms F_{ST} under most demographic scenarios in early generations. By 10,000 generations, however, F_{ST} performs as well as Δπ without migration and outperforms Δπ with migration.

Demography moderates the shape and tail end of divergence distributions

By comparing distributions across simulations with divergent ($Pheno_{Div}$) and stabilising ($Pheno_{Null}$) selection with neutral runs, we can examine the effect of demographic treatments on the ability of each measure of divergence to discriminate between them (Figure S11-13; Figure 5). There are few differences between distributions of F_{ST} for the three simulation types when migration is absent between AP and DP replicates, highlighting an increased likelihood of false-positives. The exceptions occur in early sampling points at 100 and 500 generations (Figure S12) when sweeps are most common. With migration,

403 $Pheno_{Null}$ F_{ST} is marginally elevated compared with neutral F_{ST}, but distributions are broadly
404 similar. $Pheno_{Div}$ F_{ST} distributions however become more positive and flattened, according to
405 variable selection, with the majority of $Pheno_{Div}$ F_{ST} above the 95% quantiles of $Pheno_{Div}$ and
406 neutral F_{ST} by 10,000 generations (Figure 5).
407
408 Similar patterns were observed for D_{XY} distributions (Figure S14-17), with little to
409 discriminate between in treatments without migration. However, without migration, neutral
410 D_{XY} was generally reduced relative to $Pheno_{Null}$ and $Pheno_{Div}$ D_{XY} at earlier sampling points.
411 With migration, like F_{ST}, $Pheno_{Div}$ D_{XY} was readily distinguishable from $Pheno_{Null}$ and neutral
412 distributions, but $Pheno_{Null}$ D_{XY} was also generally more positive than neutral D_{XY}. These
413 patterns also emerged more slowly than for F_{ST}. In contrast, $Pheno_{Div}$ $\Delta\pi$ (Figure S18-21) was
414 elevated according to DP size, such that at 500 generations the majority of 25kb regions
415 under divergent selection exhibited $\Delta\pi$ above neutral and $Pheno_{Null}$ 95% cut-offs for all
416 treatments with DP size ≥ 0.5.
417
418 We quantified false-positive rates (FPR) by permuting over merged data comprised of
419 randomly sampled 5% $Pheno_{Div}$ regions and 95% neutral regions and observing the upper
420 5% quantile (Table S1). By 100 generations, F_{ST} FPR ranged between 0.06 and 0.91, and was
421 lower with increased DP size and lower in treatments without migration (Table S1). By 500
422 generations (Table 2), FPR rates were lower with migration and higher without, but only for
423 treatments with DP sizes of 0.5 and 1.0. F_{ST} FPR remained high (> 0.81) for all treatments
424 with smaller DP sizes. By 3,000 generations, migration was the most important demographic
425 factor for F_{ST} FPR. With migration, FPR ranged from 0.15 – 0.61, and decreased with
426 increasing DP size. Without migration, FPR were high (0.88 - 0.97), close to the random

proportion of neutral (0.95) data. By the end of simulations, F_{ST} FPR was as low as 0.13, and was no greater than 0.27 with migration and DP size ≥ 0.1. Without migration, FPR was > 0.94.

Initial D_{XY} FPR were generally high irrespective of demographic treatment, ranging between 0.78 and 0.93. FPR were largely similar after 500 generations, but by 3,000 generations there was a distinction between treatments with ($0.07 \leq FPR \leq 0.71$) and without ($0.85 \leq FPR \leq 0.88$) migration. Interestingly, here FPR rates were lower ($0.07 \leq FPR \leq 0.24$) when DP were smaller (size = 0.01, 0.1) rather than larger ($0.47 \leq FPR \leq 0.71$). This pattern was also observed at the end of simulations, with FPR lower without migration ($0.01 \leq FPR \leq 0.34$) and lowest with DP size = 0.1.

Δπ FPR rates were generally higher across all sampling generations, with FPR not falling below 0.28. There was a clear distinction based on DP size in earlier (100 and 500) sampling points, with FPR lower with larger DP size. However, at later sampling generations (3,000 and 10,000), FPR were generally high (≥ 0.68) regardless of treatment.

Taken together, there are clear effects of DP size and migration on distributions of genetic variation and upper quantiles of interest. DP size appears initially most important in the first few hundred generations for F_{ST} and Δπ, but these effects are later swamped by the effect of migration for F_{ST} and are simply eroded for Δπ. D_{XY} exhibits similar patterns to F_{ST}, but these develop after many more generations, and whilst gene flow increases the informativeness of D_{XY} for detecting divergent selection, as it does for F_{ST}, D_{XY} and F_{ST} experience opposing effects of increases to DP size.

Demography moderates relationships between measures of divergence

Given F_{ST}, D_{XY} and Δπ are all measures of population genetic divergence, there is an assumption that positive correlations should exist between them. We employed the same analysis as above for correlations with selection, but instead examined correlations between individual measures. Founding bottlenecks had minimal effects on the correlations observed between all measures of divergence (Figure 6A).

Positive correlations between F_{ST} and D_{XY} emerged rapidly irrespective of DP size, but smaller DP sizes generally increased the correlation (Figure 6B), with variance between treatments generally decreasing over time. Similarly, F_{ST} and Δπ were generally positively correlated, however reductions in DP size reduced correlation coefficients. By 4,000 generations, F_{ST} - Δπ correlations for DP sizes ≥ 0.1 stabilised around 0.4, but correlations under extreme prolonged bottlenecks continued to decline to a low of 0.19 (Figure 6B). D_{XY} - Δπ correlations were generally weaker than other comparisons across the course of simulations, but were minimally affected by prolonged bottlenecks.

Migration induced substantial variance between correlations of divergence measures, with effects dependent on sampling point (Figure 6C). In the absence of migration, F_{ST} and D_{XY} were more strongly correlated for the first 4,000 generations than in treatments with migration. However, from here until 10,000 generations this pattern reversed and F_{ST} - D_{XY} correlations increased with migration and deteriorated in allopatry. F_{ST} - Δπ correlations were strong initially, but a lack of migration weakened the correlation over time until

474 measures were uncorrelated by around 4,000 generations. In contrast, in the presence of
475 migration, positive correlations between F_{ST} and $\Delta\pi$ were strong and relatively stable (R^2 =
476 0.75 – 0.61 over the whole simulation period). Interestingly, without migration, D_{XY} and $\Delta\pi$
477 were largely uncorrelated, but migration induced a positive correlation between D_{XY} and $\Delta\pi$
478 that emerged after 2,000 generations and continued to increase through time.

479

480 Contextualised by our previous demonstrations of associations with selection, these results
481 highlight that selection can induce positive correlations between measures. By 10,000
482 generations, all pairwise comparisons of divergence measures become positively correlated
483 with migration, which we know is when variation is most strongly associated with selection.
484 Crucially however, this relationship only emerges after several thousand generations, before
485 which we observe positive correlations in migration-absent treatments when associations
486 with selection are weak for all measures (F_{ST} - D_{XY} in particular). Extreme reductions in DP
487 size also increase positive correlations between F_{ST} and D_{XY} despite poor associations with
488 selection in these treatments. These results highlight that positive correlations between
489 statistics are also achievable in the absence of divergent selection. It is also interesting to
490 note the decay of correlations with $\Delta\pi$ and both F_{ST} and D_{XY} in the absence of migration.
491 These observations are most likely driven by the substitution rate. Substitutions that are not
492 linked to selection (drift with ineffective selection) likely drive increased F_{ST} and D_{XY} but not
493 $\Delta\pi$.

494

495 In $Pheno_{Null}$ simulations, F_{ST} and D_{XY} were positively correlated in all treatments, but stronger
496 associations linked with demographic treatments with effective selection did not emerge

497 (Figure S22). This was also true for neutral simulations (Figure S23), but F_{ST} - D_{XY} correlations
498 were slightly larger than for $Pheno_{Div}$ and $Pheno_{Null}$ when selection was ineffective, reaching
499 a maximum of R^2 = 0.81 without migration (Figure S23C). This demonstrates that the
500 positive correlations in $Pheno_{Div}$ simulations are driven in part by divergently adaptive allele
501 frequency shifts and substitutions when selection is effective, but these correlations can
502 also emerge under stabilising selection or neutrality. Specifically, strong positive correlations
503 with $\Delta\pi$ were dependent on divergent selection, and F_{ST} – $\Delta\pi$ correlations became negative
504 over time in treatments without migration under neutrality. D_{XY} – $\Delta\pi$ correlations became
505 negative in $Pheno_{Null}$ data with migration and were unassociated without. Thus, these
506 results highlight that the relationships between measures of divergence are highly
507 dependent on migration, population size, time, and selection experienced over a genomic
508 region.
509
510 We then examined FPR and false negative rates (FNR) when combining outliers across F_{ST},
511 D_{XY} and $\Delta\pi$ (Table S1; summaries from 500 and 10,000 generations in Table 2). Combined F_{ST}
512 - D_{XY} outliers exhibited FPR rates that were highly variable (0.00 - 0.95) and similar to or
513 slightly lower than F_{ST} alone at generations 100 and 500, suggesting some improvement in
514 reducing FPR. During this period, FNRs were also high (≥ 0.78), suggesting most regions with
515 divergent selection could not be detected on a neutral backdrop. Combined F_{ST} - D_{XY} outlier
516 sets performed well at generations 3,000 and 10,000 for treatments with migration, in some
517 cases dropping to 0 although FNR were highly variable (0.26 ≤ FNR ≤ 1.00). High FPR (0.84 ≤
518 FPR ≤ 0.99) and high FNR (0.93 ≤ FPNR ≤ 1.00) were observed without migration,
519 highlighting most common outliers between F_{ST} and D_{XY} to be neutral, and a failure to detect
520 almost all divergent regions.

Combined F_{ST} - Δπ outliers tended to outperform F_{ST} and Δπ outliers according to FPR with DP sizes of 0.5 or 1.0 and in earlier (100 and 500) sampling generations. Here, FPR dropped to a low of 0 but FNR were reasonable through this time (≥ 0.35). Beyond this (sampling generations 3,000 and 10,000), FPR again dropped to 0, however these were generally alongside high FNR of up to 1.0 without migration, highlighting a failure to detect any common outliers at all. A good example of improvement over singular measures was observed at generation 500, with a founding bottleneck, equal sized populations, and migration. Here, an average of ~66% of divergent regions were detected with an average FPR of 0. This is compared with: singular F_{ST}, where FPR/FNR = 0.18; and singular Δπ where FPR/FNR = 0.29. By 10,000 generations, combined outlier sets of F_{ST} - Δπ performed poorer than singular F_{ST} with migration present, but generally returned low numbers of false positives when migration was absent, unlike F_{ST} - D_{XY}.

Combined D_{XY} - Δπ outliers performed poorly across all treatments and all sampling generations, with FNR failing to fall below 0.71. There were, however, some benefits in terms of low FPR for treatments with larger DP sizes (0.5 and 1.0) across all sampling points, suggesting high confidence in outliers (although most divergent regions are missed). This discordance between D_{XY} and Δπ and large FNR also limited the combined usage of all three measures together, with similarly high FNR largely precluding their combined usage. All three statistics did exhibit low FPR with larger DP populations at 100, and 500 generations despite migration. However, at later sampling generations performance of all three combined outlier sets was poor (high FPR/FNR) if migration was absent.

These results therefore highlight that combining measures can help reduce FPR, but usually at the cost of increased FNR (expectedly), and only under certain demographies. Our observation that high FPR are prevalent among combined outlier sets from statistics, particularly in the absence of migration, suggests their usage must be dependent on a knowledge of disparity in population size and connectivity of populations. These findings also highlight that the strong correlations that emerge between measures of divergence under scenarios with ineffective selection or even under neutrality do extend to the tail-ends of distributions.

Demography drives signals of convergence irrespective of selection

Clusters of overlapping outliers developed steadily over time (Figure S26-28). By sampling generation 10,000, significant proportions of overlapping outliers were recovered across different demographic treatment groups (Figure 7). For F_{ST} and D_{XY}, clustering of treatments was driven by the presence/absence of migration, with the highest proportions of convergent outliers observed between treatments with migration.

Interestingly, for D_{XY} reasonable proportions of convergent outliers were also recovered across no migration treatments, and the same was true of F_{ST} by 3,000 generations (Figure S28), and for both measures at 10,000 generations in 'smaller' populations with reduced scaling ($\mu = 4.89e^{-7}$, $r = 1e^{-7}$; Figure S37). This is despite these treatment groups lacking effective selection. Importantly, there was little overlap between migration and no-migration clusters, suggesting different convergent outliers within each. Combining outliers from F_{ST} and D_{XY} reduced overlap among no-migration treatments, but did not remove overlapping outliers altogether.

There were minimal convergent outliers observed for Δπ outliers, however combining Δπ with F_{ST} and D_{XY} did appear somewhat effective in removing convergent outliers found between no-migration treatments. However, for both combination of Δπ with F_{ST} and D_{XY}, the highest proportional overlap was observed for treatments with the smallest DP size.

Interestingly, the clustering of F_{ST} - D_{XY} outliers in the presence of migration was greatly reduced when divergent selection was removed in both $Pheno_{Null}$ (Figure S29-32) and neutral data (Figure S33-36), but clusters of outliers in the absence of migration were still apparent.

Because migration was the dominant factor in clustering of treatments with convergent outlier overlap, we sought to investigate what features of simulated genes drove variance in measures of divergence with treatments separated by migration factor using linear mixed models at the final sampling generation. With migration between AP and DP, selection had by far the strongest effect on F_{ST} (Table 3) in the expected positive direction. We also observed a weaker positive association with selection target (% coding) of gene (Table 3), and weaker negative associations with Pheno Gen (generation DP optimum reached) (Table 3). These fixed effects explained 48.96% of variance in F_{ST} with migration.

Conversely, fixed effects explained minimal variance (1.10%) of F_{ST} in treatments without migration. These fixed effects were significant, but had weak positive associations with selection, exon N and selection target %, and a weak negative association with Pheno Gen.

D_{XY} was similarly most strongly positively associated with selection in treatments with migration (Table 3), but the strength of this effect relative to the other model effects was not as large as observed for F_{ST}. D_{XY} was also negatively associated with Pheno Gen (Table 3), as was F_{ST}, but negatively associated with Exon N (Table 3). This model however explained less variance of D_{XY} with migration (9.53%) than F_{ST} with migration. In treatments without migration, selection target % was strongly negatively associated with D_{XY} (Table 3), and this fixed effect explained 30.70% of D_{XY} variance without migration.

Selection was the most important fixed effect for Δπ regardless of migration (Table 3), however both models explained minimal variance (8.27% with migration, 0.59% without migration). This is consistent with the previous demonstration of the erosion of Δπ over time (Figure 1).

Removing divergent selection (i.e. in $Pheno_{Null}$ simulations) modified models for F_{ST} and D_{XY} for treatments with migration. Selection and selection target remained the most important model effects, but had more similar sized effects, and ultimately variance explained dropped from 48.96% to 3.60%. Conversely, the model for $Pheno_{Null}$ D_{XY} with migration increased variance explained from 9.53% to 13.10% compared with $Pheno_{Div}$ D_{XY}. Selection target had a strongly significant negative association alongside strength of selection in this model. Δπ models were largely unchanged. Together, these results confirm that divergent selection, and not stabilising selection within DP, drive F_{ST} and D_{XY} variation in $Pheno_{Div}$ simulations.

DISCUSSION

Summary of results

Here, we show that features of demography can have dramatic effects on how measures of population divergence identify regions of the genome under selection. Effects are also strongly time-dependent. Using simulated populations, we have demonstrated the relative influences of founding bottlenecks, prolonged bottlenecks, and migration on three commonly used measures of genetic divergence (F_{ST}, D_{XY}, $\Delta\pi$), whilst demonstrating the relative usefulness of each measure for informing on selection when population sizes and connectivity vary.

We find that founding bottlenecks have little effect on population divergence measures, potentially either because populations quickly recover (before first sampling after 100 generations), or founding bottlenecks of 10% (100 individuals) were not extreme enough. Prolonged bottlenecks (reductions in DP size) however, artificially inflate F_{ST} and $\Delta\pi$ but reduce $D_{XY,}$ and can erase the relationship of F_{ST} and D_{XY} with selection under the most extreme reductions in population size. Relative measures are, in part, driven by intra-population changes in allele frequency, which become exaggerated in smaller populations as a product of drift (Charlesworth 2009; Ellegren and Galtier 2016). As a consequence, we observe inflated measures of relative divergence as allele frequencies drift in smaller DP replicates.

In contrast, D_{XY} increases with larger DP size, as a product of the relationship between the number of segregating sites and the population-level mutation rate ($4N_e\mu$) (Hartl *et al.* 1997). D_{XY} is a measure of sequence divergence and is averaged across all sites (although similar statistics limit averaging to segregating sites only), which results in higher D_{XY} as

segregating sites are introduced into either population at a rate of $4N_e\mu$. This relationship with the number of segregating sites can be observed by examining the positive relationships between D_{XY} and DP_{π} (Figure S38). Overall, the relationship between D_{XY} and selection is less affected by prolonged bottlenecks than F_{ST} and $\Delta\pi$, likely due to the lack of allele frequency relevance. Consider for example, two SNPs with minor allele frequencies of 0/0.5 (SNP 1) and 0.5/0.5 (SNP 2) in AP/DP. Each locus contributes equally towards D_{xy} (SNP 1 = [0 x 0.5] + [1 x 0.5] = 0.5; SNP 2 = [0.5 x 0.5] + [0.5 x 0.5] = 0.5), whereas the reduction of within-population variance observed for SNP 1 inflates F_{ST} and $\Delta\pi$. However, we also see evidence of D_{XY} FPR increasing with increased DP size (Table 2), suggesting this relationship with increased acquisition of segregating sites may conflict with increased efficacy of selection.

Migration, even at the relatively modest rate of 0.2% employed here, substantially reduced absolute values for all measures of divergence. However, whilst absolute values were reduced, their informativeness of selection coefficients increased dramatically (both in terms of their overall relationship with selection and in identifying outliers). Such a result is expected given the role of gene flow in homogenising neutral loci (reducing measures of divergence), whilst retaining population divergence around adaptive loci (increasing informativeness). The well-known 'genomic islands of divergence' model is often invoked to explain this pattern of heterogenous genomic divergence in studies of speciation-with-gene-flow (Turner *et al.* 2005; Nosil *et al.* 2009).

Migration also exhibited interesting temporal patterns that are useful for discussing the discrepancies observed between $\Delta\pi$ and both F_{ST} and D_{XY}. Of the measures considered here,

665 Δπ uniquely disregards SNP substitutions in its calculation, as fixed substitutions have no
666 influence on intra-population heterozygosity beyond reducing the heterozygosity of
667 proximate SNPs in the aftermath of a hard sweep. This characteristic explains why Δπ
668 responds to selection more rapidly than F_{ST} and D_{xy} and is influenced by migration in a
669 constant manner across time. In addition, the inability of Δπ to account for adaptive
670 substitutions is likely why the informativeness of Δπ peaked at around 500 generations
671 (approximate median for sweeps), decayed, and then stabilised. In contrast, the
672 contributions of adaptive substitutions to F_{ST} and D_{xy} over time increases their predictive
673 power. This characteristic of Δπ, however, also makes it unable to differentiate between
674 divergent and stabilising selection. This temporal observation has important implications for
675 studies of rapid adaptation on the scale of 10s to 100s of generations, in which our
676 simulations suggest Δπ may be the most informative measure of divergence, whilst D_{xy} is
677 initially uninformative for several thousand generations.
678
679 Examining the effects of bottlenecks and migration on the associations between measures
680 of divergence themselves is a novel element to this study. The relative weaknesses of
681 individual measures of divergences has prompted a recent movement within the literature
682 to employ multiple measures of divergence to avoid false-positives (for e.g. Tine et al. 2014;
683 Malinsky et al. 2015; Hämälä and Savolainen 2019). Our results provide some support for
684 this strategy, with reductions in FPR in treatments with gene flow generally, and low FPR
685 when combining either F_{ST} or D_{XY} with Δπ (albeit at a reasonably high FNR). However, we
686 also find large numbers of overlapping outliers across combined F_{ST} - D_{XY} with a large FPR
687 across treatments without migration. Further, these false-positives persisted in $Pheno_{Null}$
688 and neutral data, whereas clusters of treatments with overlapping outliers (Figure 7) and

689 with low FPR were restricted to simulations with divergent phenotypes. It is clear then that
690 combining outlier sets of F_{ST} and D_{XY} only improves analyses under certain demographic
691 histories, in particular when populations are connected by migration.
692
693 Cruickshank and Hahn (2014) suggested that a disagreement between F_{ST} and D_{XY} outliers in
694 genomic-islands-of-divergence tests highlights a particular susceptibility of F_{ST} to BGS
695 (although see (Matthey-Doret and Whitlock 2019)). Our results are in line with the notion
696 that D_{XY} may be more resistant to false positives due to BGS, given that reductions in DP size
697 resulted in minimal variance in D_{XY} (assuming reductions in population size are analogous to
698 different rates of BGS across the genome exhibit reduced N_e). However, BGS was not
699 specifically manipulated in these simulations, and results are difficult to disentangle with
700 variation in efficacy of removing deleterious mutations. Further, this minimal variance may
701 be explained by the opposing forces of selection efficacy and acquisition of segregating sites
702 discussed previously. We also found that a greater proportion of variance could be
703 explained by the size of selection target for D_{XY} (30.7%) in the absence of migration than for
704 F_{ST} (1.1%). We found that F_{ST} and D_{XY} are positively correlated under several demographic
705 scenarios, but this strong association only reflects selection when gene flow is present and
706 only after several thousand generations, consistent with previous simulation work (Ravinet
707 *et al.* 2017). When migration is absent, and selection ineffective, F_{ST} and D_{XY} are also
708 positively correlated; but this relationship decays over time (Figure 6C), with empirical
709 evidence (see below) suggesting a negative relationship is likely to emerge without gene
710 flow. No-migration treatments are consistent with Isolation-By-Distance demographic
711 histories and, similarly, comparisons between reproductively-isolated populations or
712 species. Indeed, negative correlations between F_{ST} and D_{XY} within and between clades of

713 birds (Irwin *et al.* 2016; Vijay *et al.* 2017), within a radiation event of monkeyflowers
714 (Stankowski *et al.* 2019) and between speciating orca populations (Foote *et al.* 2016)
715 support a declining relationship between these measures over long periods of time in
716 isolation. In agreement, we find that without migration F_{ST} and D_{XY} exhibit opposite
717 associations with the proportion of regions made up of coding elements. Mechanistically, a
718 larger selection target increases the rate of deleterious mutation, reducing local π directly
719 through loss of polymorphic deleterious sites, or indirectly through the loss of linked neutral
720 variants under a BGS model. Reductions in local π increase F_{ST} and decrease D_{XY}. Over time,
721 associations between F_{ST} and D_{XY} should stabilise, given π is generally well-conserved in
722 stable populations even across long time periods (Romiguier *et al.* 2014; Dutoit *et al.* 2017;
723 Van Doren *et al.* 2017).
724
725 By comparing our results to a second dataset that lacked divergent selection ($Pheno_{Null}$), we
726 found consistent support that positive associations between D_{XY} and selection are strongly
727 dependent on the inclusion of a divergent phenotype. Positive associations with F_{ST} are
728 attainable only in the absence of migration, and are weakly negative without, and Δπ
729 patterns are primarily driven by variable stabilising selection and made weaker by the
730 inclusion of phenotypic divergence. These comparisons are useful in highlighting the relative
731 roles of adaptive allele frequency changes and substitutions in driving patterns of genetic
732 differentiation. It is also of interest to note that overlapping outliers are readily attainable
733 (most strongly for D_{XY}) across no-migration treatments in $Pheno_{Null}$ (Figure S32) and neutral
734 simulations (Figure S36), but not for treatments with migration. This confirms that
735 overlapping outliers in no-migration treatments occur due to common non-divergent
736 processes within genomic regions, whereas overlapping outliers in treatments with

737 migration are driven by the effects of divergent selection. The discrepancies between our
738 $Pheno_{Div}$ and other simulated datasets highlight the necessity in quantifying phenotypic
739 differences or environmental selection pressure when interpreting patterns of variation
740 across the genome.
741
742 ***Detecting genetic convergence***
743 In addition to understanding how outlier detection in individual pairs were affected by
744 bottlenecks and migration, we wanted to explore how studies looking at overlapping
745 outliers in multiple pairs (i.e. detecting genetic convergence) were affected by these aspects
746 of demography. Our simulation design, in which an ancestral burn-in population is used to
747 found 16 independent AP-DP pairs is analogous to replicated ecotype population pairs in
748 model systems, such as various ecotype pairs of the three-spined stickleback, *Gasterosteus*
749 *aculeatus* (Hohenlohe *et al.* 2010; Jones *et al.* 2012a); high/low predation Trinidadian
750 guppies, *Poecilia reticulata* (Fraser *et al.* 2015); crab/wave ecotype periwinkles, *Littorina*
751 *saxitilis* (Westram *et al.* 2014; Ravinet *et al.* 2016; Kess *et al.* 2018), and alpine/montane
752 ecotypes of *Heliosperma pusillum* (Trucchi *et al.* 2017). We found overlapping outliers
753 between demographic treatments and thus, that signals of convergent outliers are
754 attainable for singularly used F_{ST} and D_{XY}, and combined outlier sets for $F_{ST} - \Delta\pi$, $D_{XY} - \Delta\pi$
755 and F_{ST} - D_{XY}. Clustering of outliers was predominantly driven by the presence or absence of
756 migration (with minimal overlap between clusters) (Figure 7).
757
758 The overlap of quantile-based outliers between demographic treatments without migration
759 and ineffective selection may be explained in part probabilistically. With migration
760 restricted, AP and DP exhibit significantly elevated measures of divergence, as seen in Figure

1. This increase results in a normal distribution of F_{ST} and D_{XY} across neutral simulations without migration. In contrast, with migration, drift is limited and random recombination influences gene flow and subsequently variation in divergence. Distributions of divergence with migration under neutrality are therefore are heavily right tail-skewed (Figure S39). Spread of data increases with longer right tails, and density of data in each overlapping distribution is limited to the median and lower quantiles.

We also expect some effect of different amounts of starting variation among genome regions following burn-ins. With gene flow restricted, this variation may promote overlap under neutral conditions given all demographic treatments are founded from a common burn-in. However, this feature of our simulations is analogous to the conserved landscapes of variation observed in natural genomes (Burri 2017; Vijay *et al.* 2017; Stankowski *et al.* 2019).

Empirically, the influence of migration on outlier overlap has been observed in replicate pairs of parasitic and non-parasitic lampreys. As we show here, when comparing outliers from disconnected and connected parasitic/non-parasitic pairs, Rougemont et al. (2017) recover greater numbers of overlapping outliers among comparisons of disconnected pairs than connected pairs. Overlapping outliers among connected pairs are however better correlated, which the authors suggest reflects selection.

Limitations of simulations

In our analyses, we grouped genomic regions within 16 unique demographic treatments, assuming the effects of reductions in population size and migration are equivalent for all

785 regions. This may be unrepresentative of genomes sampled from the wild, in which gene
786 flow and effective population size can vary across genomic regions through structural
787 variation, variable recombination rate and BGS (Gossmann *et al.* 2011). A prominent
788 example of such a mechanism is the well-characterised consequence of reduced gene flow
789 within inversions that carry locally adapted alleles. Assuming inversion variants are fixed
790 between populations, gene flow across the locus is limited by the prevention of
791 recombinant haplotypes and resistance to introgression (Kirkpatrick and Barton 2006;
792 Ravinet *et al.* 2017). Recent genome scan studies have highlighted convergent outliers
793 within inversions (Jones *et al.* 2012b; Nishikawa *et al.* 2015; Morales *et al.* 2019), confirming
794 theoretical models regarding their role in shielding adaptive haplotypes from introgression
795 during the adaptation process. Our simulations suggest that genes with reduced migration,
796 and reduced N_e as a consequence, will have inflated measures of divergence and
797 differentiation relative to other genomic regions. Therefore, we predict this may have led to
798 their over-representation in genome scan outliers, and increased potential to overlap across
799 replicate populations. Thus, caution should be taken regarding the adaptive significance of
800 these outliers relative to absolutely lower values of genetic divergence attained from
801 regions outside of chromosomal rearrangements.
802
803 By choosing to use a factorial design here, we have increased our understanding of the
804 interplay between specific features of demographic history and multiple measures of
805 population divergence. However, computational limitations constrained absolute
806 population sizes to a maximum of 1,000 individuals and generations to a maximum of
807 10,000 (with 10,000 generation burn-in). To mitigate these constraints, we repeated the
808 analysis over multiple mutation rates (and scaled recombination rate) to illustrate patterns

809 over 100-fold variation of θ. In general, most trends were consistent, suggesting that results
810 should be consistent across taxa of variable effective population size. However, certain
811 patterns were exaggerated or dampened with increased or decreased mutation rate
812 respectively. For instance, the overlap of outliers between no-migration treatments was
813 exaggerated at our smallest level of population scaling (Figure S37), suggesting false-positive
814 signals of convergence may be more likely with reduced N_e. Further, patterns associated
815 with D_{XY} and selection vary according to population size, appearing delayed in smaller
816 populations (Figure S7) and swamped by mutational input in larger populations (Figure S6).
817 This latter effect can induce negative associations between selection and D_{XY} with effective
818 selection, as selection reduces local genetic variation that would otherwise increase D_{XY}. It is
819 well-documented that both N_e (Frankham 1995) and mutation rate (μ) (Hodgkinson and
820 Eyre-Walker 2011) are highly variable across taxa, which suggests that applying knowledge
821 of the relationships between measures of population differentiation will vary in nature.
822
823 In addition, temporal variation within our simulations is confounded by the time at which
824 there was a major shift in the DP population phenotype (Figure S40), and by extension when
825 selective sweeps occur. This variation in timing is random with respect to adaptive
826 mutations arising *de novo*, but is also influenced by demographic treatment. For example,
827 treatments that experience founding bottlenecks are less likely to evolve using variation in
828 the founder, increasing dependence on *de novo* mutations for adaptation. Additionally,
829 variation in our DP size parameter modifies the per population number of new mutations
830 per generation. This is also true for features of simulated genes, such as size of selection
831 target. Predictable temporal variation in the time at which adaptation is likely to occur is a

832 probable source of variance between measures of divergence and is particularly clear for
833 $\Delta\pi$, but this was controlled for in later modelling analyses as a fixed effect.
834
835 A further consideration for these simulations concerns the architecture of the phenotype.
836 Results reported here pertain to mutation effect sizes drawn from a distribution centred at
837 0 with $\sigma = 1$. This produces mutations of typically large effect, but was selected on the basis
838 of phenotypic optima being reasonably distant, with a difference of 10. Thus, 99% of
839 mutations in simulations produce phenotypic differences of less than a third the divergence
840 distance of AP and DP phenotypes. There are numerous factors that influence the
841 distribution of mutation effect sizes in nature, including: selection, mutation, drift, gene
842 flow, extent of pleiotropy and distance to phenotypic optima (Dittmar *et al.* 2016), with no
843 single expectation for natural systems as a result. The relatively large distance between
844 optima in our simulations, as well as the rapid change in optima implemented in simulations
845 (Collins and De Meaux 2009), likely gives increased importance to mutations of large effect.
846 The interactions between mutation effect size and the results presented here are beyond
847 the scope of the current study, but we did investigate the effect of reducing μ_σ of mutation
848 effect distributions to 0.1 (Supporting Information; Figures S40-45). Briefly, we see
849 reductions in the strength of correlations and associations with selection with decreasing
850 phenotypic effects of mutation, consistent with the notion of softer sweeps on small-effect
851 loci. We also see increased variance in the amount of time taken for simulations to reach
852 the $Pheno_{Div}$ optimum, which agrees with the probable importance of large effect loci in our
853 standard dataset. We, however, retain strong positive associations between measures such
854 as F_{ST} and D_{XY}, as we see in neutral simulations, as well as overlapping outliers linked to
855 selection at the tail ends of $Pheno_{Div}$ distributions. Running the simulations in this way

suggests that many of the patterns described here may be robust to scenarios with reduced mutation effect sizes.

It is also important when translating these results to genomic data to consider how correlations between measures of divergence depend on the selection type used in simulations. For example, F_{ST} and D_{XY} are strongly positively correlated under neutrality without migration, but under the same demographic scenarios we observe a decay in the relationship between F_{ST} and D_{XY} when divergent selection is involved. Genomes of natural populations will include regions that are neutral or nearly neutral, under stabilising selection around a common phenotypic optimum, or divergent between populations. Thus, the patterns described here may not apply to all genomic regions pooled together.

Concluding remarks

We have used forward-in-time simulations to perform a factorial experiment in which we explored the relationships between three measures of genetic divergence, selection, and features of demographic history that are commonly variable in natural populations. In agreement with previous theoretical work, we found the reliance of measures of genetic divergence to identify loci under selection are dependent on population size and migration, and variable through time, with a notable lag in D_{XY}. Furthermore, we provided novel comparisons between measures of genetic divergence that call into question the use of multiple measures to rule out false-positives. We also demonstrated that signals of convergent evolution across independent replicates can be driven by similar features of demographic history with minimal influences of selection, specifically, across replicate populations pairs that lack migration. Therefore, we strongly advise those using overlapping

outlier scans to carefully consider the demographic context of their system to avoid false-positives. In particular, the presence or absence of migration between diverging populations is a key factor determining the informativeness of genetic variation for selection, and importantly shapes our expectations of outlier overlap among replicate population pairs. It is tempting to assume that replication in study design or analysis in the form of taking multiple measures of genetic divergence can reduce the risk of attaining false-positives. We hope to emphasise in this study that this is not always the case, as false-positives (i.e. genome scan outliers that are not associated with regions under divergent selection) can be driven by non-random genomic or common demographic features that cannot be bypassed through replication. Moreover, many of the patterns we observe are variable through time, such that the relevant pitfalls of analyses will depend on the age of the populations being considered. It should thus also be important to estimate population splits, as a young replicate pair and older replicate pair with similar demographic histories should be expected to exhibit potentially different patterns of genetic variation.

Recent simulation work by Quilodrán et al. (2019), has also emphasised the influence of genomic features and demography, and includes simulation software for estimating the distribution of genetic variation over user-defined chromosomes. Such an approach is particularly useful for systems with chromosome-level genome assemblies in order to gain a sense for how features such as recombination, gene density, and selection targets may produce false-positives under certain demographies. The software employed here, SLiM (Haller and Messer 2019), may also be used to this end, and the scripts accompanying this study will facilitate similar analyses over system-specific genome regions. Further, recent work on genomic landscapes of linked selection (Stankowski *et al.* 2019) has highlighted that

much of the total variance of genetic divergence such as F_{ST}, $D_{XY,}$ and π can be explained by the major principal component (PC1) over numerous pairwise comparisons. These population comparisons need not reflect divergent phenotypes, as PC1 reflects genomic features associated with diversity landscapes. Adopting this approach may be useful for systems lacking a chromosome-level genome assembly by estimating SNPs or regions with non-random elevated measures of divergence associated with genome features. Such SNPs or regions may be particularly prone to false-positives under certain demographic histories.

ACKNOWLEDGEMENTS

The authors wish to thank all members of the Fraser lab for insightful discussions on this project, as well as members of the University of Exeter's EB theme and the University of Sussex's EBE group. Thanks to the University of Exeter's ISCA HPC infrastructure for performing simulations. The authors would also like to thank Kai Zeng and Adam Eyre-Walker for feedback on earlier copies of this manuscript, and two anonymous reviewers for inciteful and helpful comments. This work was funded as part of a European Research Council grant (758382 GUPPYCon) awarded to BAF.

REFERENCES

Ahrens, C. W., P. D. Rymer, A. Stow, J. Bragg, S. Dillon *et al.*, 2018 The search for loci under selection: trends, biases and progress. Mol. Ecol. 27: 1342–1356.

Bank, C., G. B. Ewing, A. Ferrer-Admettla, M. Foll, and J. D. Jensen, 2014 Thinking too positive? Revisiting current methods of population genetic selection inference. Trends Genet. 30: 540–546.

Van Belleghem, S. M., C. Vangestel, K. De Wolf, Z. De Corte, M. Most *et al.*, 2018 Evolution at two time frames. PLoS Genet.

Burri, R., 2017 Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. Mol. Ecol. 26: 3853–3856.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. (Camb). 68: 131–149.

933 Charlesworth, B., 2009 Effective population size and patterns of molecular evolution and
934 variation. Nat. Rev. Genet. 10: 195.
935 Charlesworth, B., 1998 Measures of divergence between populations and the effect of
936 forces that reduce variability. Mol. Biol. Evol. 15: 538–543.
937 Charlesworth, D., B. Charlesworth, and M. T. Morgan, 1995 The pattern of neutral molecular
938 variation under the background selection model. Genetics 141: 1619–1632.
939 Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious
940 mutations on neutral molecular variation. Genetics 134: 1289–1303.
941 Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection,
942 balanced polymorphism and background selection on equilibrium patterns of genetic
943 diversity in subdivided populations. Genet. Res. (Camb). 70: 155–174.
944 Collins, S., and J. De Meaux, 2009 Adaptation to different rates of environmental change in
945 Chlamydomonas. Evol. Int. J. Org. Evol. 63: 2952–2965.
946 Cooper, E. A., and J. A. C. Uy, 2017 Genomic evidence for convergent evolution of a key trait
947 underlying divergence in island birds. Mol. Ecol. 26: 3760–3774.
948 Cruickshank, T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of
949 speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23: 3133–
950 3157.
951 Cutter, A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites:
952 Unifying the disparity among species. Nat. Rev. Genet. 14: 262–274.
953 Dittmar, E. L., C. G. Oakley, J. K. Conner, B. A. Gould, and D. W. Schemske, 2016 Factors
954 influencing the effect size distribution of adaptive substitutions. Proc. R. Soc. B Biol. Sci.
955 283: 20153065.
956 Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette *et al.*, 2017 Correlated
957 patterns of genetic diversity and differentiation across an avian family. Mol. Ecol. 26:
958 3982–3997.
959 Dutoit, L., N. Vijay, C. F. Mugal, C. M. Bossu, R. Burri *et al.*, 2017 Covariation in levels of
960 nucleotide diversity in homologous regions of the avian genome long after completion
961 of lineage sorting. Proc. R. Soc. B 284: 20162756.
962 Ellegren, H., and N. Galtier, 2016 Determinants of genetic diversity. Nat. Rev. Genet. 17:
963 422–433.
964 Ellegren, H., and J. B. W. Wolf, 2017 Parallelism in genomic landscapes of differentiation,
965 conserved genomic features and the role of linked selection. J. Evol. Biol. 30: 1516–
966 1518.
967 Foote, A. D., N. Vijay, M. C. Ávila-Arcos, R. W. Baird, J. W. Durban *et al.*, 2016 Genome-
968 culture coevolution promotes rapid divergence of killer whale ecotypes. Nat. Commun.
969 7: 11693.
970 Frankham, R., 1995 Effective population size/adult population size ratios in wildlife: a
971 review. Genet. Res. (Camb). 66: 95–107.
972 Fraser, B. A., A. Künstner, D. N. Reznick, C. Dreyer, and D. Weigel, 2015 Population genomics
973 of natural and experimental populations of guppies (*Poecilia reticulata*). Mol. Ecol. 24:
974 389–408.
975 Fraser, B. A., and J. R. Whiting, 2019 What can be learned by scanning the genome for
976 molecular convergence in wild populations?
977 Gossmann, T. I., M. Woolfit, and A. Eyre-Walker, 2011 Quantifying the variation in the
978 effective population size within a genome. Genetics 189: 1389–1402.
979 Haller, B. C., and P. W. Messer, 2019 SLiM 3: Forward genetic simulations beyond the

Wright–Fisher model. Mol. Biol. Evol. 36: 632–637.
Hämälä, T., and O. Savolainen, 2019 Genomic patterns of local adaptation under gene flow in Arabidopsis lyrata. Mol. Biol. Evol.
Hartl, D. L., A. G. Clark, and A. G. Clark, 1997 *Principles of population genetics*. Sinauer associates Sunderland.
Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.
Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd *et al.*, 2016 Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am. Nat. 188: 379–397.
Hodgkinson, A., and A. Eyre-Walker, 2011 Variation in the mutation rate across mammalian genomes. Nat. Rev. Genet. 12: 756.
Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6:.
Hudson, R. R., M. Slatkin, and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583–589.
Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens, 2016 Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Mol. Ecol. 25: 4488–4507.
Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady *et al.*, 2012a A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Curr. Biol. 22: 83–90.
Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012b The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55–61.
Kess, T., J. Galindo, and E. G. Boulding, 2018 Genomic divergence between Spanish *Littorina saxatilis* ecotypes unravels limited admixture and extensive parallelism associated with population history. Int. J. Bus. Innov. Res. 17: 8311–8327.
Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics 173: 419–434.
Künstner, A., M. Hoffmann, B. A. Fraser, V. A. Kottler, E. Sharma *et al.*, 2016 The genome of the trinidadian guppy, Poecilia reticulata, and variation in the Guanapo population. PLoS One 11: 1–25.
Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai *et al.*, 2015 Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science (80-.). 350: 1493–1498.
Matthey-Doret, R., and M. C. Whitlock, 2019 Background selection and FST: consequences for detecting local adaptation. Mol. Ecol.
Meier, J. I., D. A. Marques, C. E. Wagner, L. Excoffier, and O. Seehausen, 2018 Genomics of parallel ecological speciation in Lake Victoria cichlids. Mol. Biol. Evol. 1–37.
Morales, H. E., R. Faria, K. Johannesson, T. Larsson, M. Panova *et al.*, 2019 Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. Sci. Adv. 5: eaav9963.
Nei, M., 1987 *Molecular evolutionary genetics*. Columbia university press.
Nishikawa, H., T. Iijima, R. Kajitani, J. Yamaguchi, T. Ando *et al.*, 2015 A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. Nat. Genet. 47: 405.
Nosil, P., D. J. Funk, and D. Ortiz-Barrientos, 2009 Divergent selection and heterogeneous

genomic divergence. Mol. Ecol. 18: 375–402.
Quilodrán, C. S., K. Ruegg, A. T. Sendell-Price, E. C. Anderson, T. Coulson *et al.*, 2019 The multiple population genetic and demographic routes to islands of genomic divergence. Methods Ecol. Evol.
R Core Team, 2016 R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. Vienna Austria 0: {ISBN} 3-900051-07-0.
Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne *et al.*, 2017 Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30: 1450–1477.
Ravinet, M., A. Westram, K. Johannesson, R. Butlin, C. André *et al.*, 2016 Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. Mol. Ecol. 25: 287–305.
Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne *et al.*, 2016 The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science (80-.). 354: 1305–1308.
Reznick, D., and J. A. Endler, 1982 The impact of predation on life history evolution in Trinidadian guppies (*Poecilia reticulata*). Evolution (N. Y). 36: 160–177.
Roda, F., L. Ambrose, G. M. Walter, H. L. Liu, A. Schaul *et al.*, 2013 Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. Mol. Ecol. 22: 2941–2952.
Romiguier, J., P. Gayral, M. Ballenghien, A. Bernard, V. Cahais *et al.*, 2014 Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515: 261.
Rougemont, Q., P. Gagnaire, C. Perrier, C. Genthon, A. Besnard *et al.*, 2017 Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Mol. Ecol. 26: 142–162.
Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto, 2009 Pervasive natural selection in the *Drosophila* genome? PLoS Genet. 5: e1000495.
Smith, J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. (Camb). 23: 23–35.
Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman *et al.*, 2014 Stick insect genomes reveal natural selection's role in parallel speciation. Science (80-.). 344: 738–742.
Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph *et al.*, 2019 Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. PLoS Biol. 17: e3000391.
Tine, M., H. Kuhl, P.-A. Gagnaire, B. Louro, E. Desmarais *et al.*, 2014 European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5: 5770.
Torres, R., Z. A. Szpiech, and R. D. Hernandez, 2018 Human demographic history has amplified the effects of background selection across the genome. PLoS Genet. 14: 1–27.
Trucchi, E., B. Frajman, T. H. A. Haverkamp, P. Schönswetter, and O. Paun, 2017 Genomic analyses suggest parallel ecological divergence in *Heliosperma pusillum* (Caryophyllaceae). New Phytol. 216: 267–278.
Turner, T. L., M. W. Hahn, and S. V Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol. 3: e285.

Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren *et al.*, 2017 Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. Mol. Ecol. 26: 4284–4295.
Waterhouse, M. D., L. P. Erb, E. A. Beever, and M. A. Russello, 2018 Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. Mol. Ecol. 27: 2512–2528.
Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution (N. Y). 38: 1358–1370.
Westram, A. M., J. Galindo, M. Alm Rosenblad, J. W. Grahame, M. Panova *et al.*, 2014 Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? Mol. Ecol. 23: 4603–4616.
Yeaman, S., and S. P. Otto, 2011 Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evol. Int. J. Org. Evol. 65: 2123–2129.
Zeng, K., 2013 A coalescent model of background selection with recombination, demography and variation in selection coefficients. Heredity (Edinb). 110: 363.

FIGURE LEGENDS

Figure 1: Experimental design for simulations. **A)** Examples of selection treatments experienced by genes from across the range, illustrated as the relationship between relative fitness of an individual and its phenotype. Each facet represents a different fitness landscape modified through editing the standard deviation of the normal distribution of fitness consequences in DP (blue, dashed). Facet labels constitute S, which were transformed as 10^{-S} to give fitness function standard deviations (S_σ) The x-axes represent phenotype as calculated through non-synonymous mutations and the y-axes represent relative fitness of individuals. The ancestral phenotype of AP (red, solid) is the same in all treatments (mean = 0, σ = 1), whilst DPs have a diverged phenotypic optimum (mean = 10, σ = S_σ). **B)** Distribution of S values applied to genes (1 per gene) **C)** Demographic representation of treatment factors. **D)** Representation of simulation timeline for treatments, illustrating that all treatments share an ancestral burn-in population before splitting into 16 replicated "AP" (solid) and "DP" (dotted) population pairs. Red, dashed lines denote sampling generations at which F_{ST}, D_{XY} and $\Delta\pi$ are calculated and averaged across the preceding 20 generations. The purpose of this averaging was to achieve a general sense of population differentiation at sampling points, such that values represent stable patterns rather than stochastic generation to generation variation. **E)** Examples of two simulated 25kb regions, showing central genes that function as QTL, and illustrating how regions vary in gene length, exon N and selection target (%). **F)** Representation of 100 simulated 25kb regions dataset per iteration (N = 20) per treatment group (N = 16).

Figure 2: Effects of demographic treatments on measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for founding bottlenecks (**A**), prolonged bottlenecks (**B**) and migration (**C**). Each point represents values of divergence averaged across all genes within individual treatments groups, averaged within treatment levels, and averaged over 20 iterations.

Figure 3: Effects of demographic treatments on the relationship between selection and measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for founding bottlenecks (**A**), prolonged bottlenecks (**B**) and migration (**C**). Each point represents correlation coefficients calculated across all genes within individual treatments groups, averaged within treatment levels, and averaged over 20 iterations.

Figure 4: Pairwise comparisons between the correlation coefficients of selection with F_{ST}, D_{XY} and $\Delta\pi$ across four sampling points. Each data point represents a unique demographic treatment with points coloured according to DP population size and shaped according to migration level. Correlation 1 refers to the first measure listed in the comparison and Correlation 2 to the second. The y=x line is plotted within each facet to illustrate biases towards one measure. Points below the line are biased towards Correlation 1, whilst points above the line are biased towards Correlation 2. Each point represents correlation coefficients calculated across all genes within individual treatments groups and averaged over 20 iterations.

Figure 5: Distributions of F_{ST} under each of the 16 unique demographic treatments under three selection regimes: $Pheno_{Div}$ (divergent selection), $Pheno_{Null}$ (stabilising selection) and Neutral, after 10,000 generations. Upper 5% quantiles are highlighted for each distribution, with linetype corresponding to selection: Solid = Divergent, Dashed = Stabilising, Dotted = Neutral. Each distribution represents data pooled from 20 iterations of 100 25kb regions (N = 2000).

Figure 6: Effects of demographic treatments on the relationship between measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for founding bottlenecks (**A**), prolonged bottlenecks (**B**) and migration (**C**). Each point represents correlation coefficients calculated across all genes within individual treatments groups, averaged within treatment levels, and averaged over 20 iterations.

Figure 7: The proportional overlap of outliers above the 95% quantile, averaged across 100 downsampled datasets consisting of 95% Neutral and 5% $Pheno_{Div}$ data for each of the 16 demographic treatments after 10,000 generations. Axis orderings were determined through hierarchical clustering. Heatmaps are shown for single measures of F_{ST}, D_{XY} and $\Delta\pi$ in the first column, and combined measures in the second column. Heatmaps are coloured according to a common scale of 0 to 1. Treatments are labelled with founding bottleneck (Bot), DP population size (Pop2), and migration (Mig) values.

TABLE 1: SIMULATION PARAMETERS

Variable	Value	Description
AP	0	Ancestral phenotypic optimum
AP-S_σ	1	AP Selection (σ of fitness around phenotypic optimum, after transforming)
AP_N	1000	AP population size
DP	10	Diverged phenotypic optimum
DP-S_σ	0.1-10.00	DP Selection (σ of fitness around phenotypic optimum, after transforming)
DP_N	(10, 100, 500, 1000)	DP population size
BN	(100, 1000)	Founding bottleneck (N burn-in genomes sampled to populate DP)
m	(0, 0.002)	Migration (Percentage gene flow in both directions)
μ	$4.89e^{-6}$ ($4.89e^{-5}$, $4.89e^{-7}$)	Mutation rate (bp^{-1}) (additional rates for higher/lower scaling)
μ_σ	1.00	Mutation effect size (σ of distribution of effect sizes)
r	$1.00e^{-6}$ ($1.00e^{-5}$, $1.00e^{-7}$)	Recombination rate (bp^{-1}) (additional rates for higher/lower scaling)
C_{max}	1.00	Maximum fitness cost through competition
C_σ	0.40	Local phenotypic competition (σ of fitness reductions between individuals)
C_{Dist}	1.20	Maximum phenotypic distance between competitive individuals

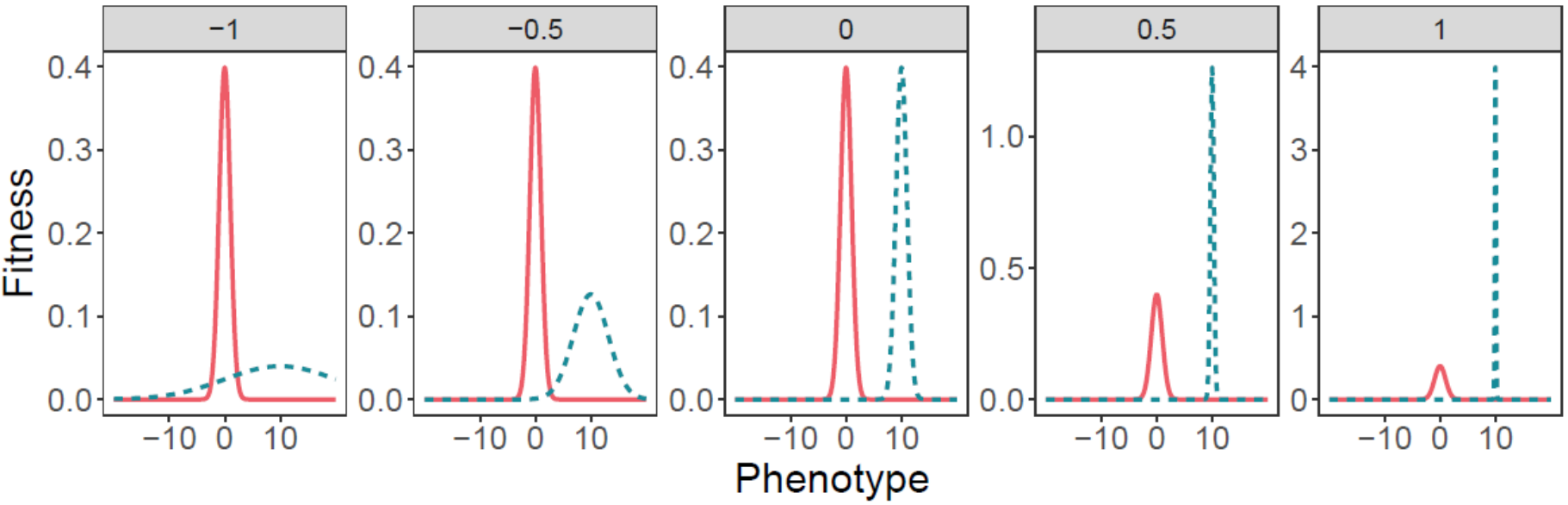
TABLE 2: FALSE POSITIVE (FPR) AND FALSE NEGATIVE RATES (FNR) CALCULATED ACROSS ALL MEASURES OF DIVERGENCE AND THEIR COMBINED USE. ESTIMATES WERE CALCULATED BY COMBINING 5% OF DATA UNDER DIVERGENT SELECTION WITH 95% NEUTRAL DATA AND TAKING UPPER 5% CUT-OFFS. FOR SINGLE MEASURES, OUTLIER N IS ALWAYS 100 (5% OF 2,000) AND FNR = FPR.

Gen	Demographic Treatments		Single Measures			Combined Measures											
			F_{ST}	D_{XY}	Δπ	F_{ST} & D_{XY}			F_{ST} & Δπ			D_{XY} & Δπ			All 3		
	Migration	DP Size	FPR	FPR	FPR	Outlier N	FPR	FNR	Outlier N	FPR	FNR	Outlier N	FPR	FNR	Outlier N	FPR	FNR
500	0	0.01	0.87	0.87	0.99	62.21	0.87	0.92	9.64	0.98	1.00	26.21	0.98	1.00	9.64	0.98	1.00
500		0.1	0.91	0.86	0.99	54.79	0.89	0.94	14.77	0.98	1.00	13.63	0.99	1.00	7.89	0.98	1.00
500		0.5	0.50	0.81	0.47	31.56	0.49	0.84	41.54	0.19	0.66	12.11	0.37	0.92	10.01	0.24	0.92
500		1	0.28	0.80	0.28	26.84	0.28	0.81	59.81	0.04	0.43	14.99	0.00	0.85	14.75	0.00	0.85
500	0.002	0.01	0.93	0.76	0.98	60.15	0.92	0.95	19.57	0.97	0.99	17.07	0.96	0.99	15.21	0.96	0.99
500		0.1	0.85	0.81	0.96	14.83	0.78	0.97	29.09	0.89	0.97	1.48	0.72	0.99	1.48	0.72	0.99
500		0.5	0.35	0.90	0.47	10.56	0.37	0.93	58.81	0.17	0.51	5.20	0.21	0.96	5.20	0.21	0.96
500		1	0.18	0.90	0.30	11.37	0.29	0.92	66.35	0.02	0.35	6.61	0.12	0.94	5.67	0.00	0.94
10000	0	0.01	0.95	0.89	0.97	27.86	0.96	0.99	0.00	0.00	1.00	7.57	0.96	1.00	0.00	0.00	1.00
10000		0.1	0.98	0.89	1.00	31.29	0.99	1.00	0.73	0.73	1.00	1.53	0.98	1.00	0.00	0.00	1.00
10000		0.5	0.98	0.88	0.98	31.21	0.99	1.00	7.60	0.99	1.00	4.98	0.95	1.00	3.79	0.99	1.00
10000		1	0.94	0.88	0.95	31.47	0.95	0.98	5.25	0.82	0.99	5.15	0.87	0.99	2.80	0.89	1.00
10000	0.002	0.01	0.53	0.17	0.79	56.84	0.21	0.55	34.41	0.44	0.81	22.76	0.17	0.81	22.56	0.17	0.81
10000		0.1	0.27	0.04	0.89	72.06	0.00	0.28	24.08	0.54	0.89	11.12	0.00	0.89	11.12	0.00	0.89
10000		0.5	0.17	0.17	0.70	74.68	0.02	0.27	32.26	0.09	0.70	28.57	0.00	0.71	28.57	0.00	0.71
10000		1	0.14	0.34	0.78	67.03	0.07	0.37	25.33	0.11	0.78	20.70	0.05	0.80	20.70	0.05	0.80

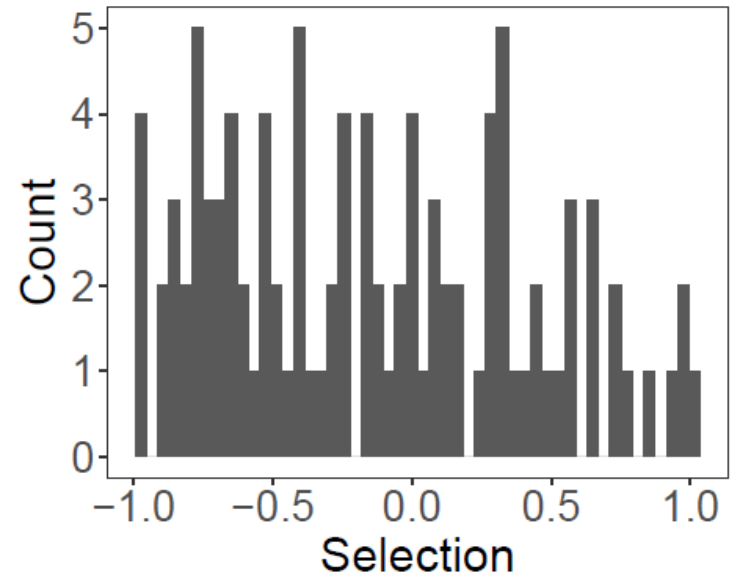
TABLE 3: LMM RESULTS FOR MODELS OF MEASURES OF DIVERGENCE EXPLAINED BY FEATURES OF SIMULATED REGIONS. FOR ALL MODELS, RANDOM VARIABLES INCLUDED GENE ID AND DEMOGRAPHIC TREATMENT.

		Var. explained (%)							
Measure	Migration	Fixed	Total	Fixed Effect	Estimate	Std. Error	df	*t*	*P*
F_{ST}	0.002	48.96	69.12	Selection	0.044	0.002	98	27.27	<2.00E-16
				Selection Target	0.009	0.001	103	9.32	2.47E-15
				Pheno Gen	-0.002	0.000	14530	-7.55	4.60E-14
F_{ST}	0	1.10	96.84	Selection	0.005	0.001	96	4.32	3.78E-05
				Pheno Gen	0.000	0.000	14480	-2.58	0.010*
				Exon N	0.005	0.002	96	2.51	0.014*
				Selection Target	0.004	0.002	97	2.46	0.015*
D_{XY}	0.002	9.53	84.25	Selection	0.007	0.000	98	16.97	<2.00E-16
				Pheno Gen	-0.001	0.000	14560	-10.49	<2.00E-16
				Exon N	-0.001	0.000	101	-4.15	6.89E-05
D_{XY}	0	30.70	53.94	Selection Target	-0.003	0.000	98	-11.57	<2.00E-16
Δπ	0.002	8.27	76.20	Selection	0.184	0.006	99	29.82	<2.00E-16
				Selection Target	0.018	0.004	120	4.71	6.84E-06
				Pheno Gen	0.006	0.002	14250	2.92	0.004
Δπ	0	0.59	95.84	Selection	0.066	0.004	99	16.14	<2.00E-16
				Exon N	-0.018	0.003	111	-6.15	1.24E-08
				Pheno Gen	-0.004	0.001	13900	-3.33	8.67E-04

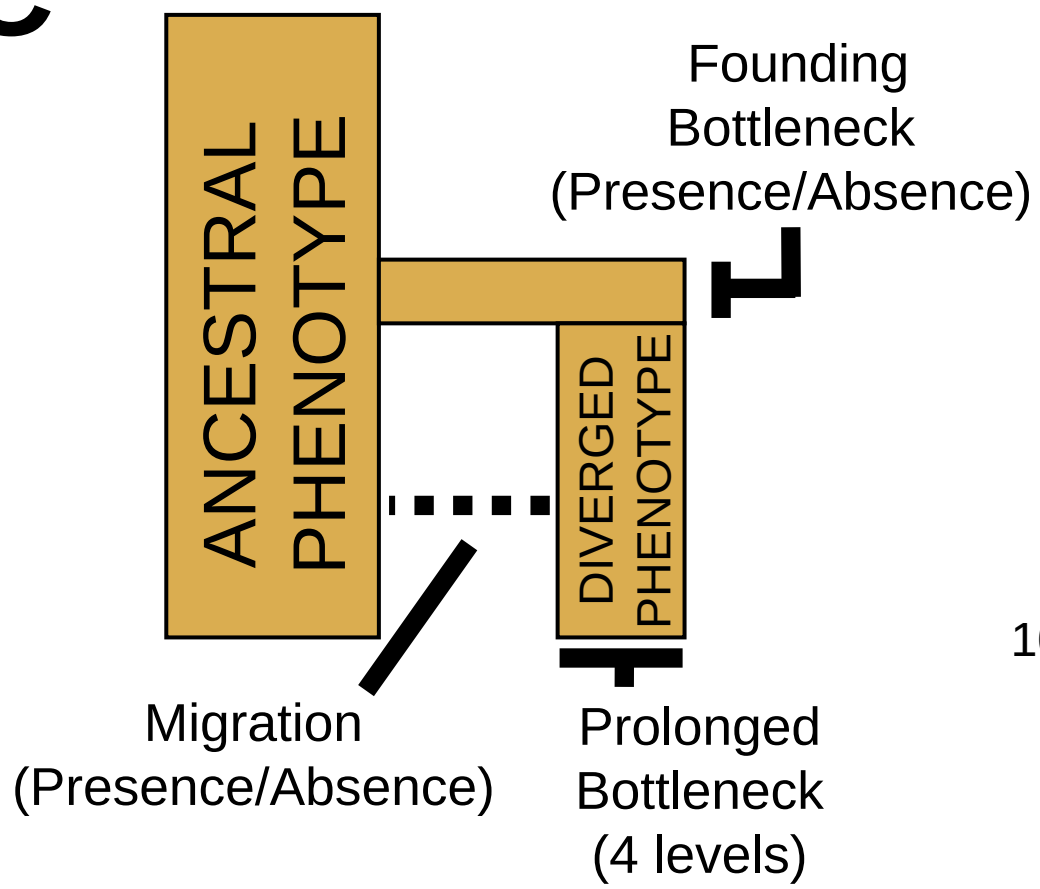
A


−1
−0.5
0
0.5
1
Fitness
Phenotype

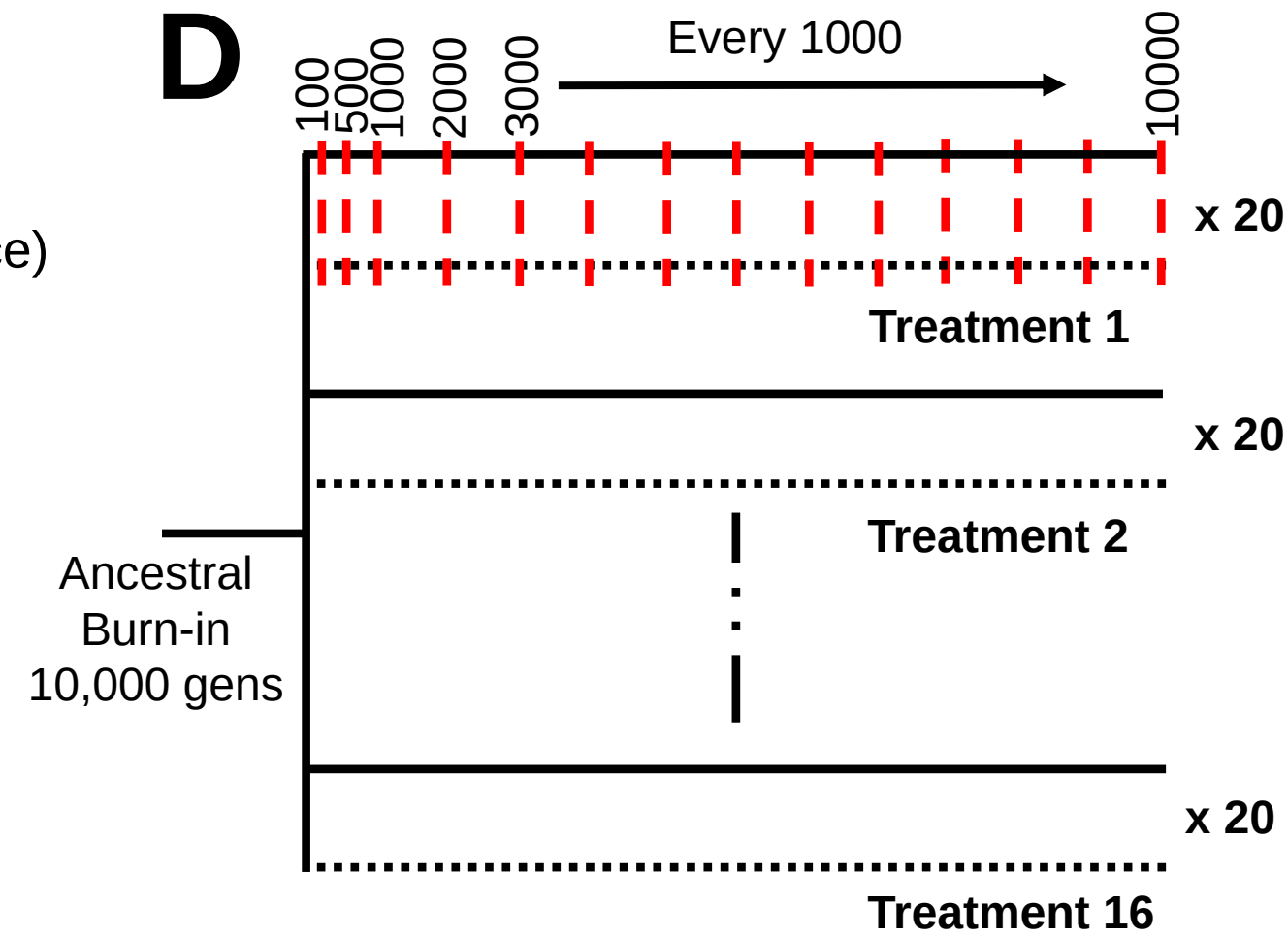
B


Count
Selection

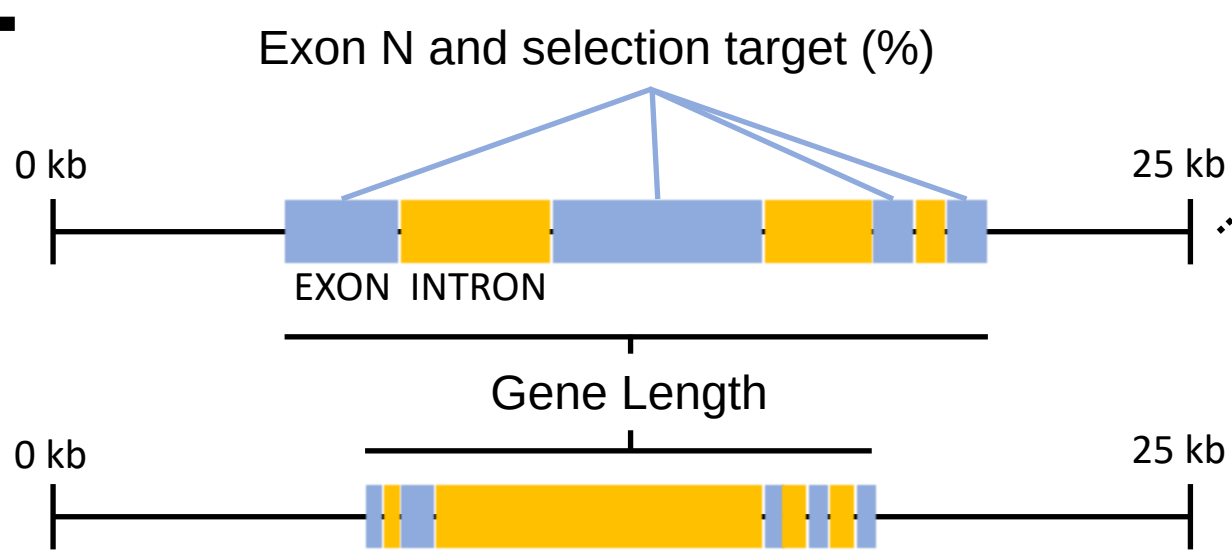
C


ANCESTRAL PHENOTYPE
DIVERGED PHENOTYPE
Founding Bottleneck (Presence/Absence)
Migration (Presence/Absence)
Prolonged Bottleneck (4 levels)

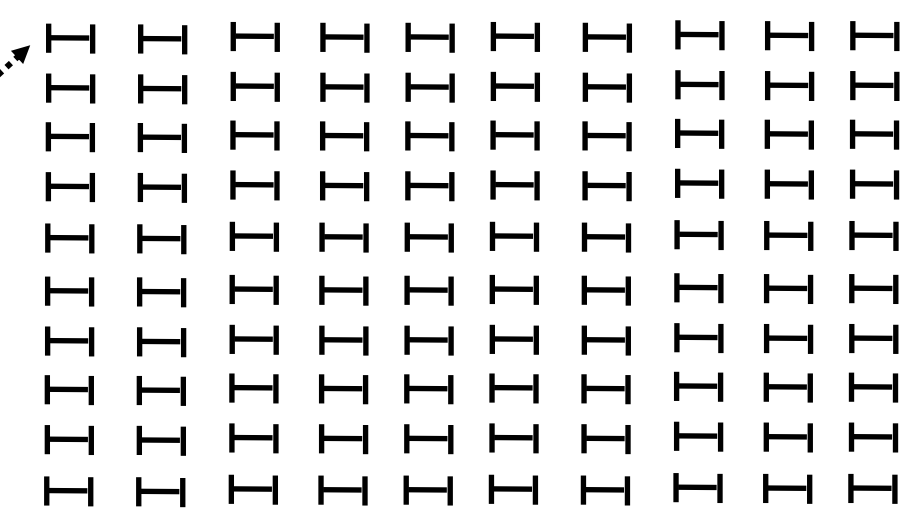
D

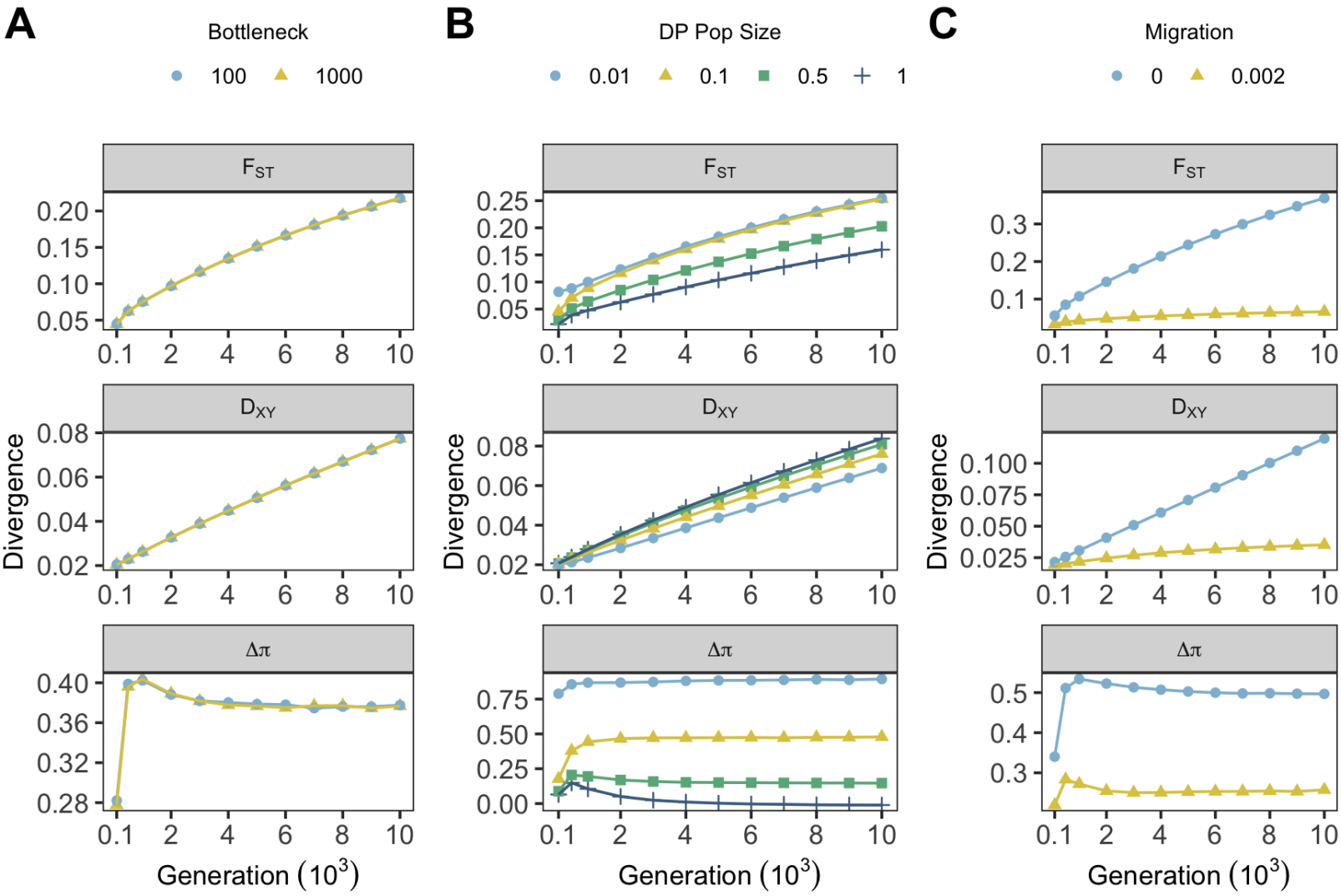

100
500
1000
2000
3000
Every 1000
10000
x 20
Treatment 1
Treatment 2
Ancestral Burn-in 10,000 gens
x 20
Treatment 16

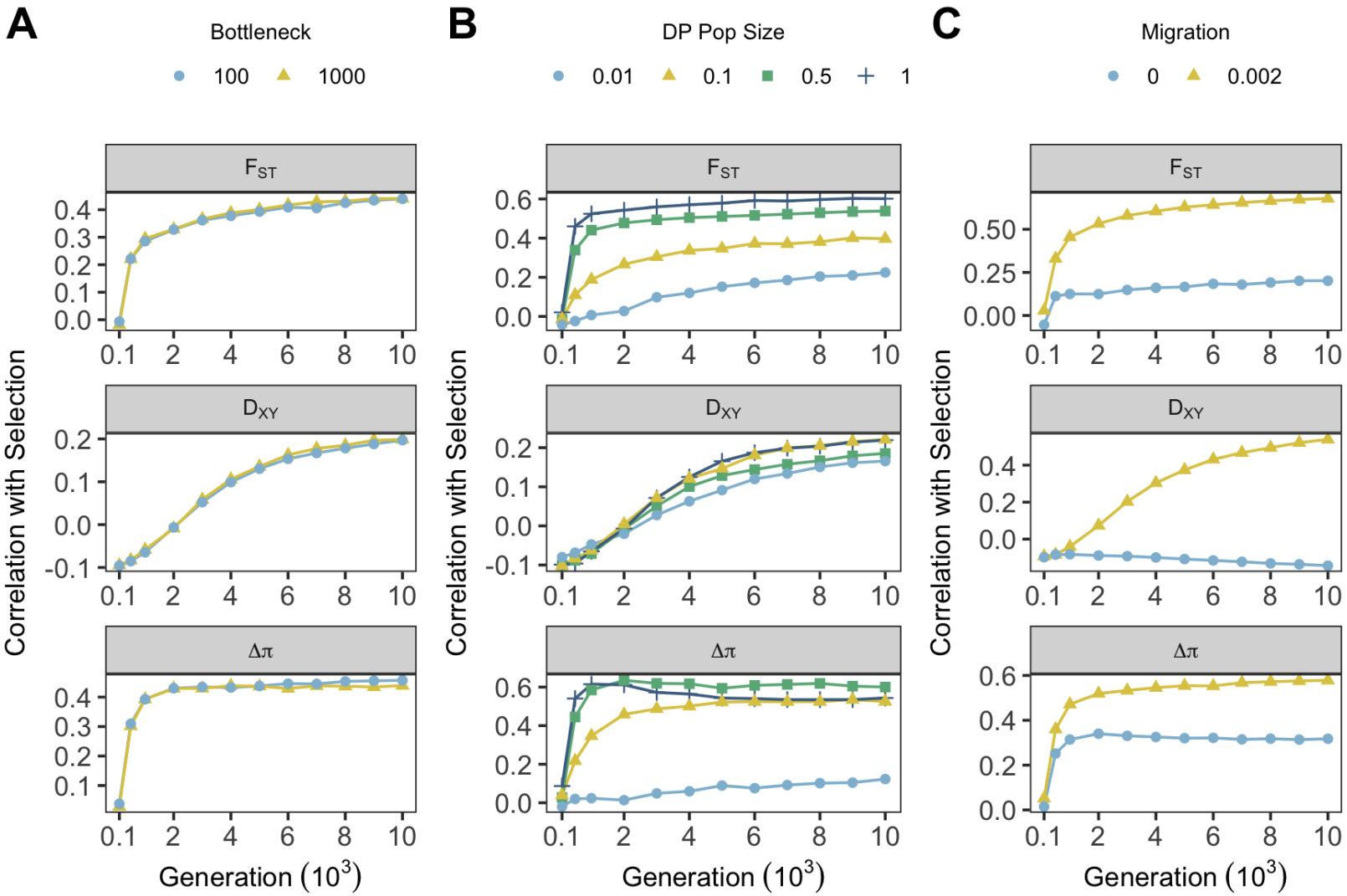
E

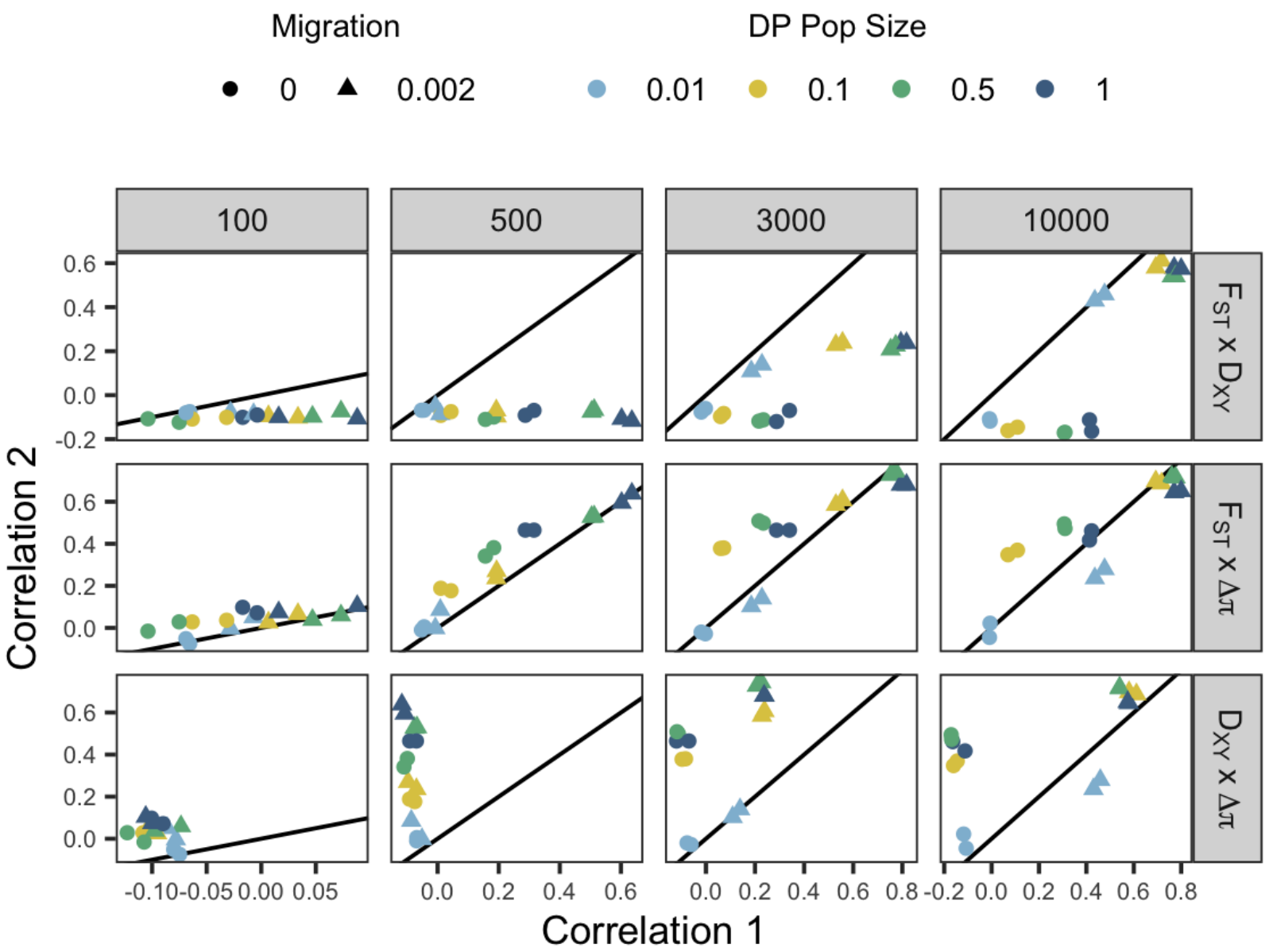

Exon N and selection target (%)
0 kb
25 kb
EXON INTRON
Gene Length
0 kb
25 kb

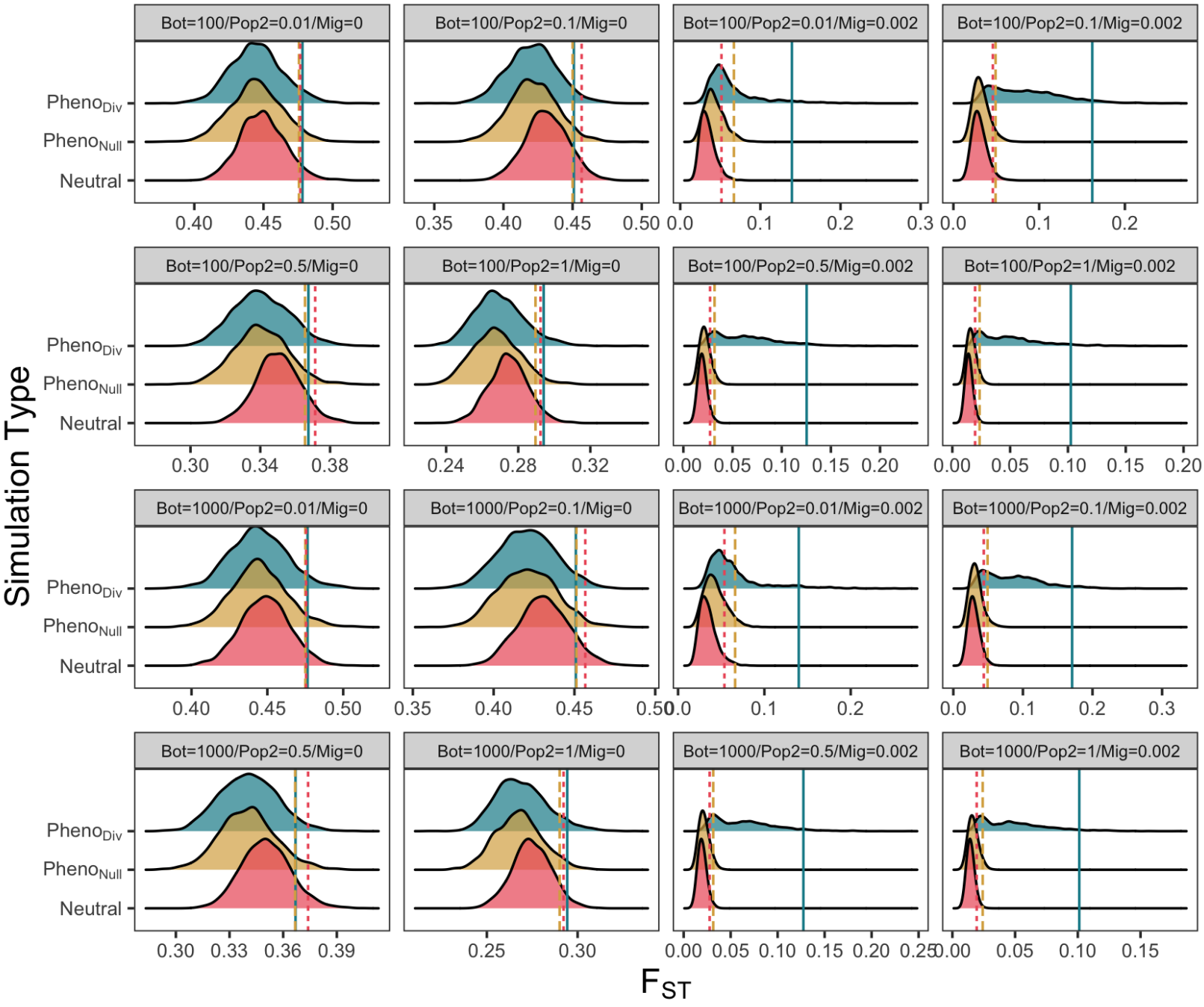
F

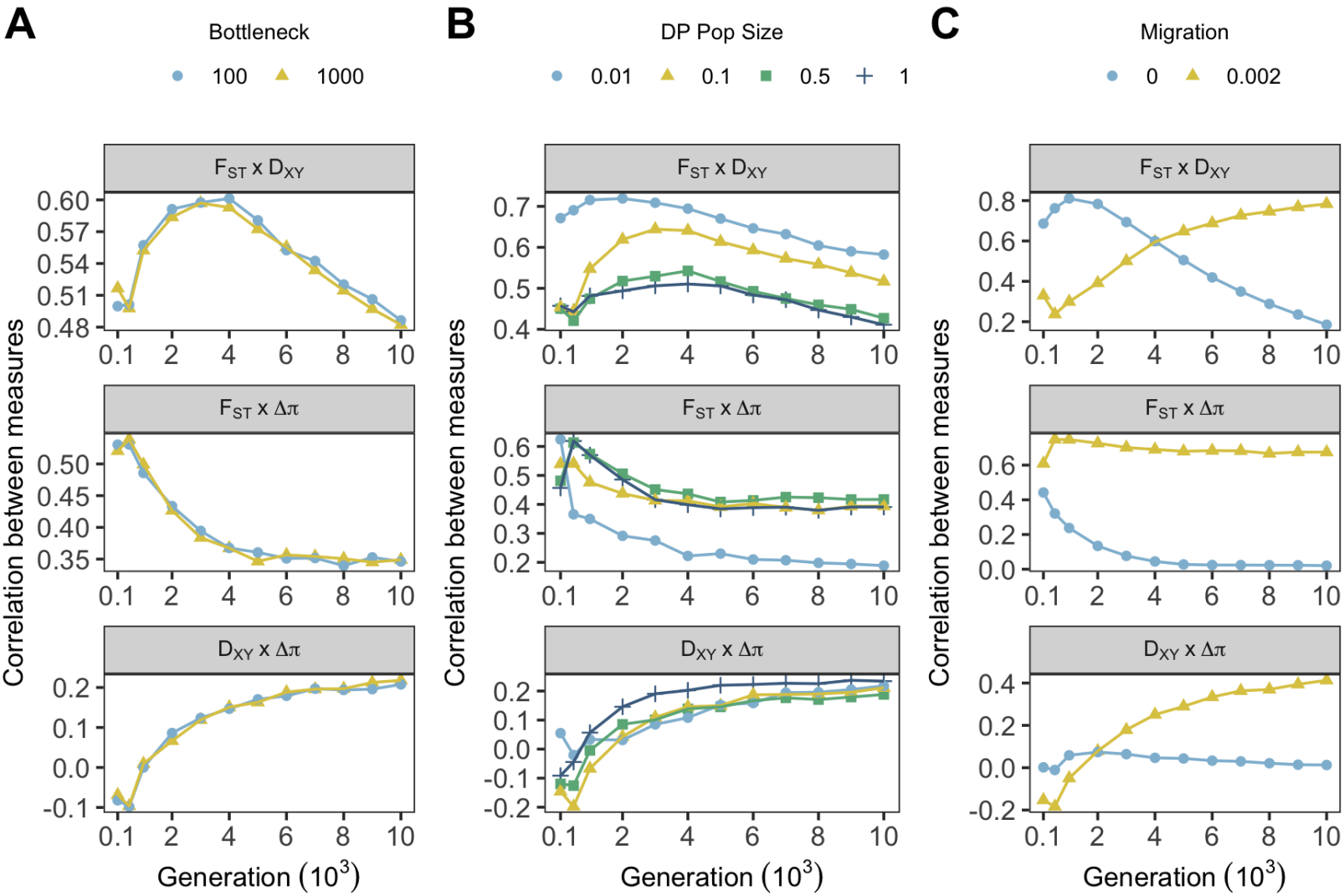



A
Bottleneck
100
1000
B
DP Pop Size
0.01
0.1
0.5
1
C
Migration
0
0.002
F_{ST}
D_{XY}
Δπ
Divergence
Generation (10^3)

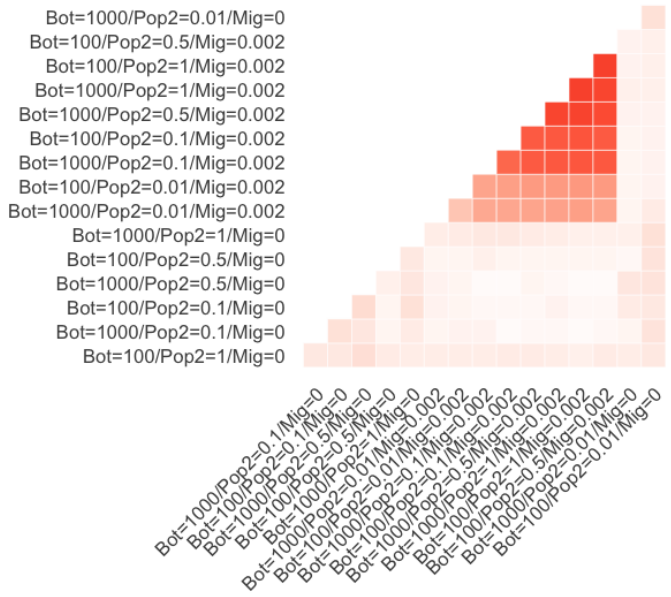

A
Bottleneck
100
1000
B
DP Pop Size
0.01
0.1
0.5
1
C
Migration
0
0.002
F_{ST}
D_{XY}
Δπ
Correlation with Selection
Generation (10^3)


Migration
0
0.002
DP Pop Size
0.01
0.1
0.5
1
100
500
3000
10000
F_{ST} x D_{XY}
F_{ST} x $\Delta\pi$
D_{XY} x $\Delta\pi$
Correlation 2
Correlation 1

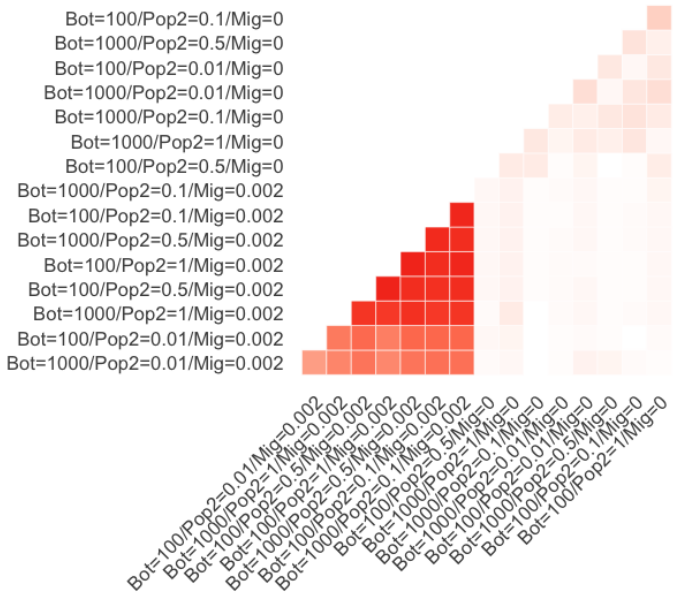

Bot=100/Pop2=0.01/Mig=0
Bot=100/Pop2=0.1/Mig=0
Bot=100/Pop2=0.01/Mig=0.002
Bot=100/Pop2=0.1/Mig=0.002
Bot=100/Pop2=0.5/Mig=0
Bot=100/Pop2=1/Mig=0
Bot=100/Pop2=0.5/Mig=0.002
Bot=100/Pop2=1/Mig=0.002
Bot=1000/Pop2=0.01/Mig=0
Bot=1000/Pop2=0.1/Mig=0
Bot=1000/Pop2=0.01/Mig=0.002
Bot=1000/Pop2=0.1/Mig=0.002
Bot=1000/Pop2=0.5/Mig=0
Bot=1000/Pop2=1/Mig=0
Bot=1000/Pop2=0.5/Mig=0.002
Bot=1000/Pop2=1/Mig=0.002
Pheno$_{Div}$
Pheno$_{Null}$
Neutral
Simulation Type
F_{ST}


A
Bottleneck
100
1000
B
DP Pop Size
0.01
0.1
0.5
1
C
Migration
0
0.002
F_{ST} x D_{XY}
F_{ST} x Δπ
D_{XY} x Δπ
Correlation between measures
Generation (10^3)

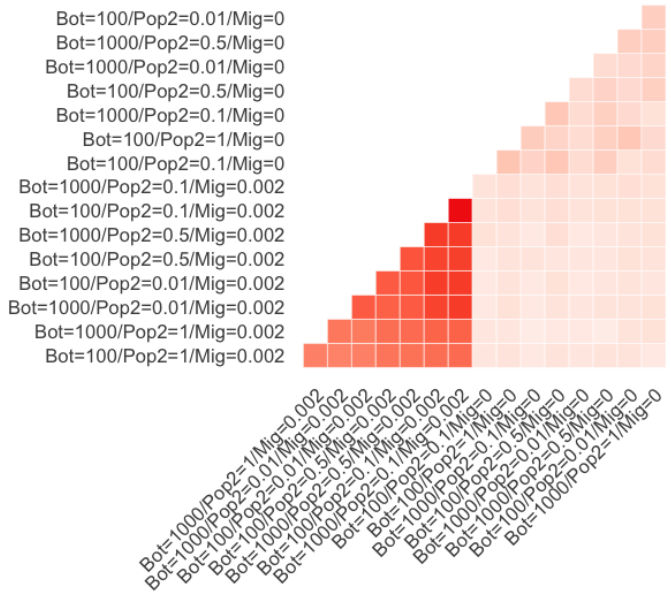
F_{ST}



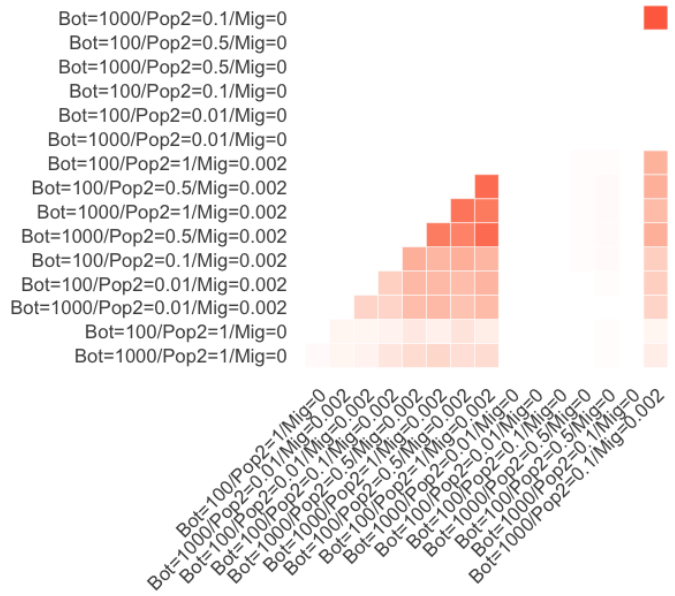
$F_{ST} + D_{XY}$



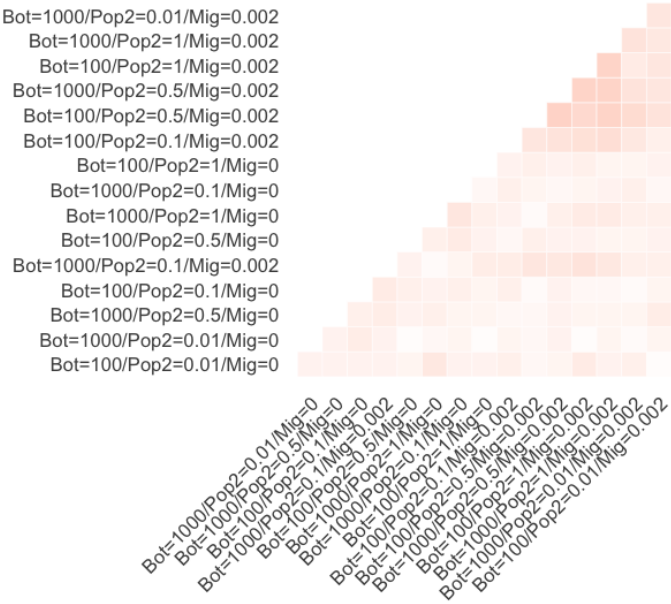
D_{XY}



$F_{ST} + \Delta\pi$



$\Delta\pi$



$D_{XY} + \Delta\pi$