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), protracted bottlenecks (proportional size of diverging population) and migration rate between two populations with ancestral and derived 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 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 demographic treatment, and that outliers across different measures (particularly 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

46 quantifying patterns of variation in genome-wide SNPs, which can be used to highlight
47 regions or genes having experienced selection relative to the neutral backdrop of the rest of
48 the genome. These analyses, which we refer to here as outlier scans, have become a
49 common tool in population genetics and have seen great success across diverse, natural
50 systems in identifying candidate genes associated with the evolution of a range of adaptive
51 traits (reviewed recently by (Ahrens et al. 2018)). More recently, the method of overlapping
52 outlier scans across independent lineages has been employed to test whether the same
53 regions are involved in independent adaptation events (i.e. genetic convergent evolution)
54 (reviewed by (Fraser and Whiting 2019)). This study seeks to investigate how different
55 outlier scan methods are influenced by demographic variation in natural populations, and
56 how this may lead to overlapping false-positives.
57
58 Recent discussions have highlighted the propensity of outlier scans to yield false-positives,
59 given that outliers caused by heterogeneous genomic landscapes are commonplace
60 irrespective of selection (Ellegren and Wolf 2017). For example, background selection (BGS),
61 whereby linkage between neutral and deleterious variants reduces local diversity (B.
62 Charlesworth, Morgan, and Charlesworth 1993; Bank et al. 2014; Burri 2017), has been
63 invoked to offer alternative explanations for patterns at first attributed to directional
64 selection (Cruickshank and Hahn 2014). Furthermore, the influence of neutral processes
65 such as genetic drift (B. Charlesworth 2009) can similarly produce elevated genetic
66 divergence and false-positive signatures of selection.
67
68 The strength of these processes is dependent on demography. For example, the influence of
69 neutral processes and BGS should be more pronounced in smaller populations (low N_e) (D.

Charlesworth, Charlesworth, and Morgan 1995; Cutter and Payseur 2013; B. Charlesworth 2009; Yeaman and Otto 2011); as has been demonstrated in humans (Torres, Szpiech, and Hernandez 2018) and *Drosophila* (Sella et al. 2009; B. Charlesworth 1996). This relationship, however, may not be simply linear, with additional simulation evidence suggesting the effects of BGS are strongest at intermediate-low N_e, and weaker at very low or high N_e (Zeng 2013).

Lack of connectivity among populations may also elevate measures of genetic differentiation, if a large global population is composed of smaller, isolated, sub-divided populations that each experience the effects of reduced N_e (B. Charlesworth, Nordborg, and Charlesworth 1997; Hoban et al. 2016). In an attempt to mitigate the likelihood of false-positives, some have advocated using multiple measures of population differentiation and divergence to identify regions of the genome likely to be under selection (B. Charlesworth 1998; Cruickshank and Hahn 2014). However, how these measures are correlated with each other, and with selection under different demographic scenarios, has not been explored.

A diverse array of measures of genetic differentiation and divergence have been employed for outlier scans, but here we focus on two of the most common measures of relative differentiation (F_{ST} and changes in nucleotide diversity [$\Delta\pi$]) and a measure of absolute divergence (D_{XY}). The most commonly used measure in outlier scans is the fixation-index F_{ST} (Weir and Cockerham 1984; Hudson, Slatkin, and Maddison 1992), which measures the relative amount of within- and between-population variance. This measure is therefore maximised when genomic regions exhibit the lowest within- and highest between-population variance. F_{ST} outliers at the right-tail of the distribution are considered

candidates for adaptation because they reflect regions with large differences in allele frequency or high substitution rate (large between-population variance) and/or low nucleotide variation (low within-population variance) relative to the rest of the genome. Changes in nucleotide diversity ($\Delta\pi$) are another indicator of adaptation, as selection on a beneficial allele limits variation within a population resulting in selective sweeps (Smith and Haigh 1974). Comparisons of the ratio of π between diverging populations reveal regions under selection, as local π is reduced in one population in comparison to the other. Whilst similar to F_{ST}, $\Delta\pi$ does not discriminate between which copy of a polymorphism is fixed, such that a substitution between populations is equivalent to a common non-polymorphic site. $\Delta\pi$ outliers therefore represent regions of the genome with reduced π in either population relative to the rest of the genome. As a measure of absolute genetic divergence, D_{XY} (Nei 1987) does not consider the relative frequencies of polymorphisms within populations (B. Charlesworth 1998; Cruickshank and Hahn 2014). D_{XY} can be quantified as the average number of pairwise differences between sequence comparisons between two populations. This measure is therefore influenced by ancestral π and the substitution rate, so D_{XY} outliers highlight regions that are highly variable ancestrally, or in either population (large π), or exhibit many substitutions and thus increased sequence divergence.

Because each measure of differentiation/divergence (hereafter referred to collectively as measures of divergence) quantifies genetic variation between populations in a slightly different way, each has a unique relationship with demography. Whilst we can predict how individual measures are influenced by demography, and subsequently neutral processes, we know little about how different relationships between measures of divergence and

117 demography affect their combined usage. We expect then that the utility of using multiple
118 measures of divergence to detect selection may vary with demography.
119
120 This complex interplay between divergence measures and demography may be further
121 exacerbated in studies that compare scans from multiple populations to identify convergent
122 genomic evolution. Here, researchers use measures of population divergence across
123 independent pairs of evolutionary replicates with outlier loci compared across results. This
124 strategy has been employed extensively across diverse taxa, including: birds (Cooper and Uy
125 2017), fish (Hohenlohe et al. 2010; Jones, Chan, et al. 2012; Rougemont et al. 2017; Fraser
126 et al. 2015; Reid et al. 2016; Meier et al. 2018), insects (Soria-Carrasco et al. 2014; Van
127 Belleghem et al. 2018), mammals (Waterhouse et al. 2018), molluscs (Ravinet et al. 2016;
128 Westram et al. 2014) and plants (Trucchi et al. 2017; Roda et al. 2013). For the sake of
129 consistency, it is common to infer outlier loci within each replicate pair through a common
130 method, but if replicates differ in their demographic histories then the applicability/power
131 of that common method will also vary accordingly. This begs the question, can demographic
132 variation among replicates alone explain signals of, or lack of, convergence?
133
134 Here, we used forward-simulations to investigate the effects of different demographic
135 histories on the relationships between measures of divergence with selection. We varied
136 demography through manipulating the number of founding individuals (founding
137 bottlenecks), the population size of the diverging population (protracted bottlenecks) and
138 the presence/absence of migration. We simulated the effects of selection on 25kb windows,
139 each with a gene designed from features taken from the guppy (*Poecilia reticulata*) genome
140 assembly, a prominent model system for studies of convergent evolution (Fraser et al. 2015;

Reznick and Endler 1982). 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 demographic history 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 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 phenotypically-ancestral population (AP) of N = 1000 and phenotypically-derived population (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 (B. Charlesworth 2009). The main text reflects results for the intermediate mutation rate $4.89e^{-6}$, with others presented in supplementary figures. SLiM

165 employs a classic Wright-Fisher model to simulate populations, in which a population of
166 diploid hermaphrodites proceeds through generations such that an individual's contribution
167 towards the next generation is proportional to its relative fitness.
168
169 Demographic treatments included reducing DP size relative to the AP size (N = 1000) (0.01,
170 0.1, 0.5, 1.0), migration as a proportion of individuals exchanged between populations (0.0,
171 0.002) and founding bottlenecks as the number of individuals sampled from the burn-in
172 population to construct DP genomes (100, 1000). For example, a demographic history with
173 founding bottleneck = 100 individuals, DP size = 0.01, and migration = 0.002 would
174 represent the following scenario: 1) At generation 1, following a burn-in period of 10,000
175 generations to reach mutation-selection balance, populations split as 100 genomes are
176 sampled from the burn-in population of 1000 to form DP, whilst AP is formed by sampling
177 all 1000 burn-in genomes; 2) At the next generation (and for the remainder of the
178 simulation), DP size reflects the protracted bottleneck treatment, in this case 10 (0.01 of
179 1000); 3) AP and DP experience migration in both directions for the remainder of the
180 simulation. Thus, each of our treatments can be thought of as a manipulation of the
181 following: Founding bottlenecks limit the amount of variation within the burn-in population
182 that is available to found DP; Protracted bottlenecks limit the size of DP, reducing
183 mutational input and moderating the strength of neutral evolution and efficacy of selection;
184 and migration dictates the presence of migration between AP and DP (Figure 1). The total N
185 of individuals (AP + DP) within simulations varies between 1010 and 2000 depending on DP
186 size parameter, however this potential expansion does not impact our results (Supporting
187 Information).
188

Combining these treatment levels in a fully factorial experimental design generated a total of 16 different, independent demographic histories that all 25kb gene windows experienced. This factorial design allows us to examine the relative influence of founding bottlenecks, protracted 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 protracted 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 windows with a central gene that varied in length and exon content; with exon number, lengths, and intron lengths drawn at random from the guppy genome gene annotation file (gff). Recombination rate was set at a constant of $1e^{-8}$ across all windows.

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 gene, 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, Goddard, and Visscher 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.

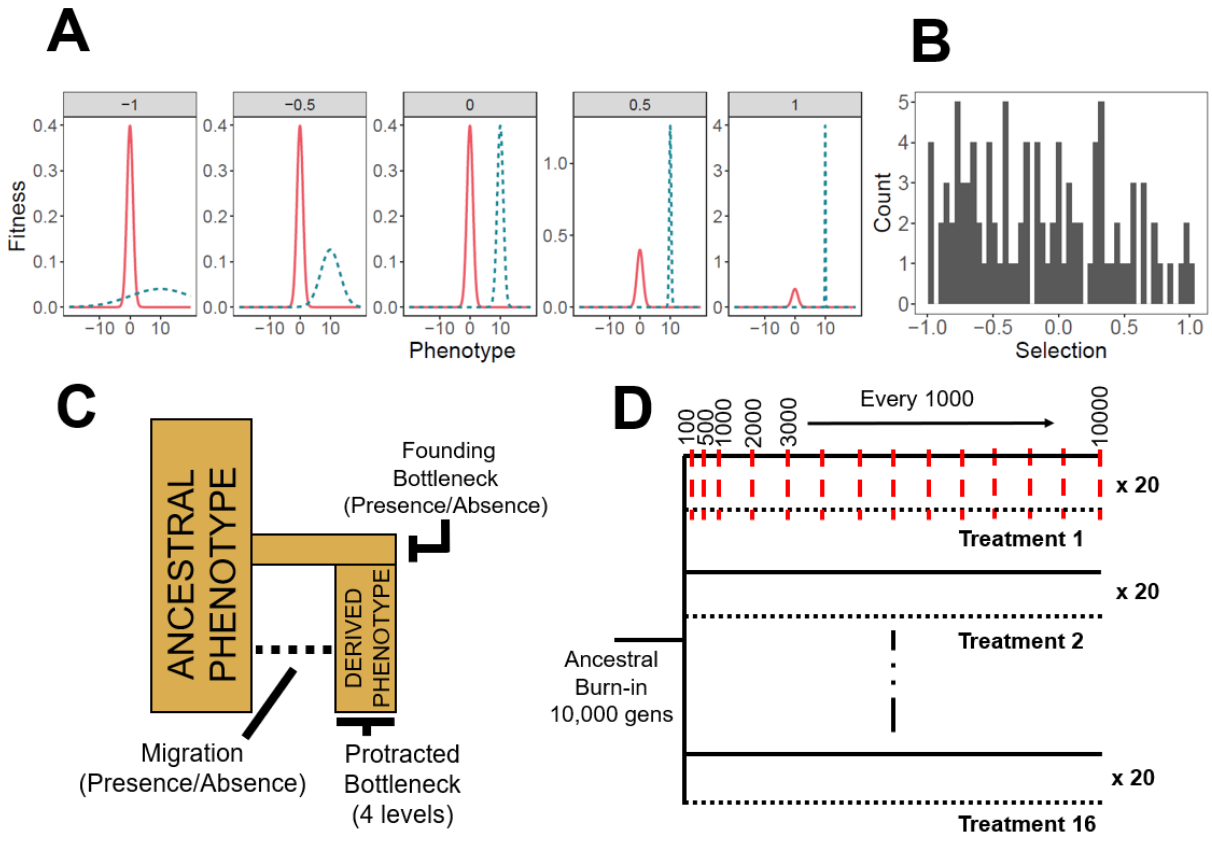


Figure 1: Experimental design for simulations. **A)** Examples of selection treatments experienced by genes from across the range, illustrated as the relationship between relative fitness of an individual and its phenotype. Each facet represents a different fitness landscape modified through editing 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.

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).

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	Derived 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
$BN_{Founding}$	(100, 1000)	Founding bottleneck (N burn-in individual 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 mutation rates with 10-fold increase and reduction)
μ_σ	1.00	Mutation effect size (σ of distribution of effect sizes)
r	$1.00e^{-8}$	Recombination rate (bp^{-1})
C_{max}	1.00	Maximum fitness cost through competition
C_σ	0.40	Local phenotypic competition (σ of fitness reductions between local individuals)
C_{Dist}	1.20	Maximum phenotypic distance between competitive individuals

Each gene was then run across all demographic treatment levels in parallel. Simulations were sampled at 100, 500, 1000, and then every 1000 generations up to 10,000 (Figure 1C). F_{ST} was calculated across the window as one minus the proportion of subpopulation heterozygosity (H_S) relative to total heterozygosity (H_T) (according to Hudson, Slatkin and Maddison (1992)). 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)). Mean heterozygosity (π) for each population was calculated as a single measure across the 25kb window. 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 $\log_{10}$-transformed ratio of π_{AP} to π_{DP}, such that reduced diversity in the DP population increases the value of Δπ. Statistics were calculated over all monomorphic and polymorphic sites of 25kb windows. This has

been designed to replicate genome scans that use sliding-approaches, with each 25kb region representing an independent window.

$$F_{ST} = 1 - \frac{H_S}{H_T}$$

$$D_{XY} = \sum_{ij} AP_i DP_j d_{ij}$$

$$\Delta\pi = \log_{10} \frac{\pi_{AP}}{\pi_{DP}}$$

In total, 100 unique 25kb windows 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 window 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 library of 100 25kb windows 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 windows (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 gene 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

319 also calculated and compared across demographic treatment groups to examine
320 convergence of results across treatments. Overlap was calculated within each permutation,
321 averaged over, and visualised using heatmaps with hierarchical clustering of axes.
322
323 To examine how simulated gene features influenced patterns of genetic divergence, we
324 used a linear mixed modelling (LMM) approach with gene ID and treatment group as
325 random factors. Gene features modelled as independent factors were: number of exons
326 (Exon N), gene size, the proportion of gene that is coding (selection target %), selection
327 applied to each gene (*S*), and the generation at which the optimum phenotype was reached
328 (Pheno Gen).
329
330 **RESULTS**
331 ***Demography modifies measures of divergence***
332 Founding bottlenecks, simulated by reducing the number of possible founding genomes to a
333 random 10% of the ancestral population, had little measurable effect on F_{ST}, D_{XY} or $\Delta\pi$,
334 producing minimal variance between treatments over all sampling points (Figure 2A).
335
336 Protracted bottlenecks, simulated by modifying the stable number of individuals within DP,
337 had pronounced and variable effects on all measures. F_{ST} increased with reductions in DP
338 size. This effect was generally consistent through time, although variance between
339 treatments increased gradually through time (Figure 2B). A similar effect was observed for
340 $\Delta\pi$; however, this measure was particularly susceptible to inflation under the most extreme
341 reductions in DP size, with substantially more elevated values observed between DP size
342 proportions of 0.01 and 0.1 compared with 0.1 and either 0.5 or 1.0. Whilst this pattern was

343 broadly consistent through sampling points, variance among the different DP sizes for Δπ
344 was most pronounced at the initial (100 generations) and final sampling points (10,000
345 generations). Such observations are unsurprising given that both F_{ST} and Δπ increase with
346 processes that reduce within-population genetic variance. Except for the most extreme
347 reduction in DP size, D_{XY} was generally robust to protracted bottlenecks. Following this, in
348 contrast to F_{ST} and Δπ, D_{XY} decreased when DP sizes were reduced (Figure 2B).
349
350 The inclusion of migration reduced absolute values of all measures of divergence. For F_{ST}
351 and D_{XY}, variance between migration treatments increased across the simulation, however
352 D_{XY} was generally more consistent across migration treatments. Δπ was also reduced in the
353 presence of migration, although the effect of migration on Δπ was generally consistent
354 across all generations and did not increase through time, as was the case for F_{ST} and D_{XY}.
355
356 Whilst F_{ST} and D_{XY} increased generally over time, Δπ peaked around generation 500 and
357 declined thereafter. This peak corresponded to the median generation that DP replicates
358 reached their optimum phenotype (median across all data = 519), suggesting this peak
359 reflects selective sweeps. The majority of these patterns were apparent in the $Pheno_{Null}$
360 (Figure S1) and neutral (Figure S2) simulations, however divergence for all measures was
361 reduced and ultimately negligible by the removal of divergent phenotypes when migration
362 was present.

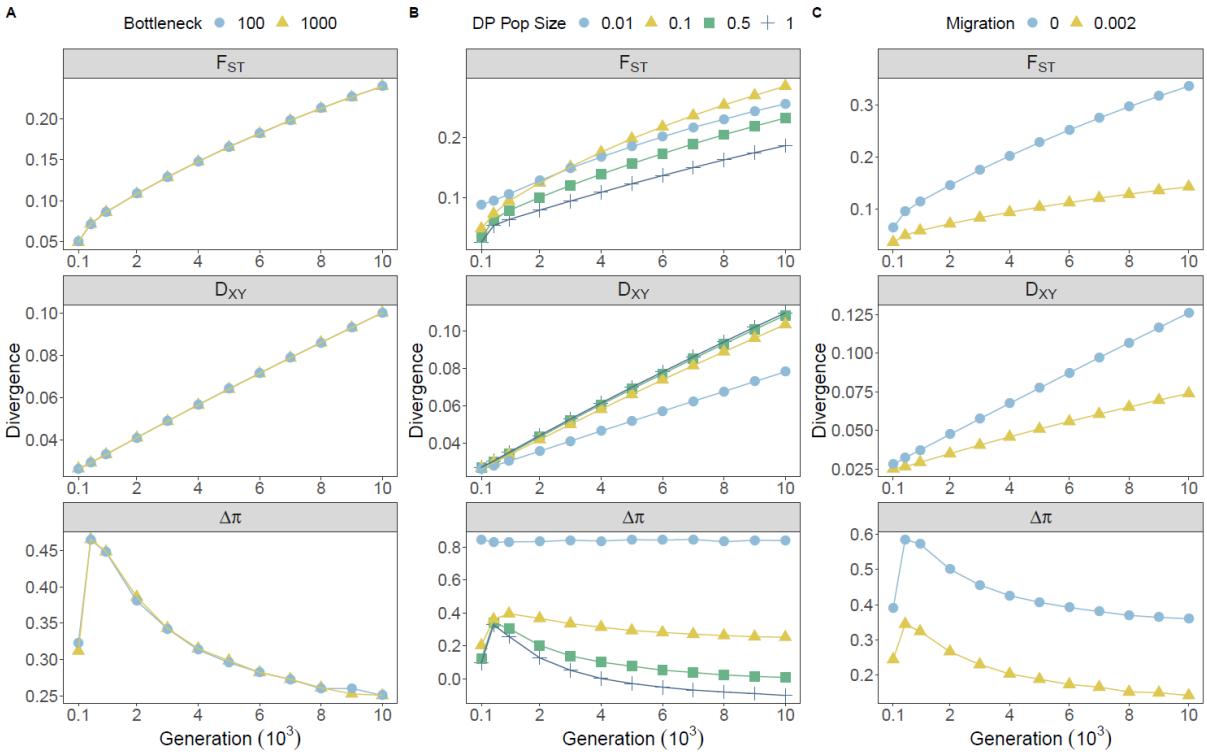


Figure 2: Effects of demographic treatments on measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for founding bottlenecks (**A**), protracted bottlenecks (**B**) and migration (**C**). Each point represents values of divergence averaged across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

Demography moderates the association between measures of divergence and selection

Again, founding bottlenecks had a minimal effect on the correlations observed between strength of selection and measures of divergence with minimal variance observed between bottleneck treatments in all comparisons (Figure 3A).

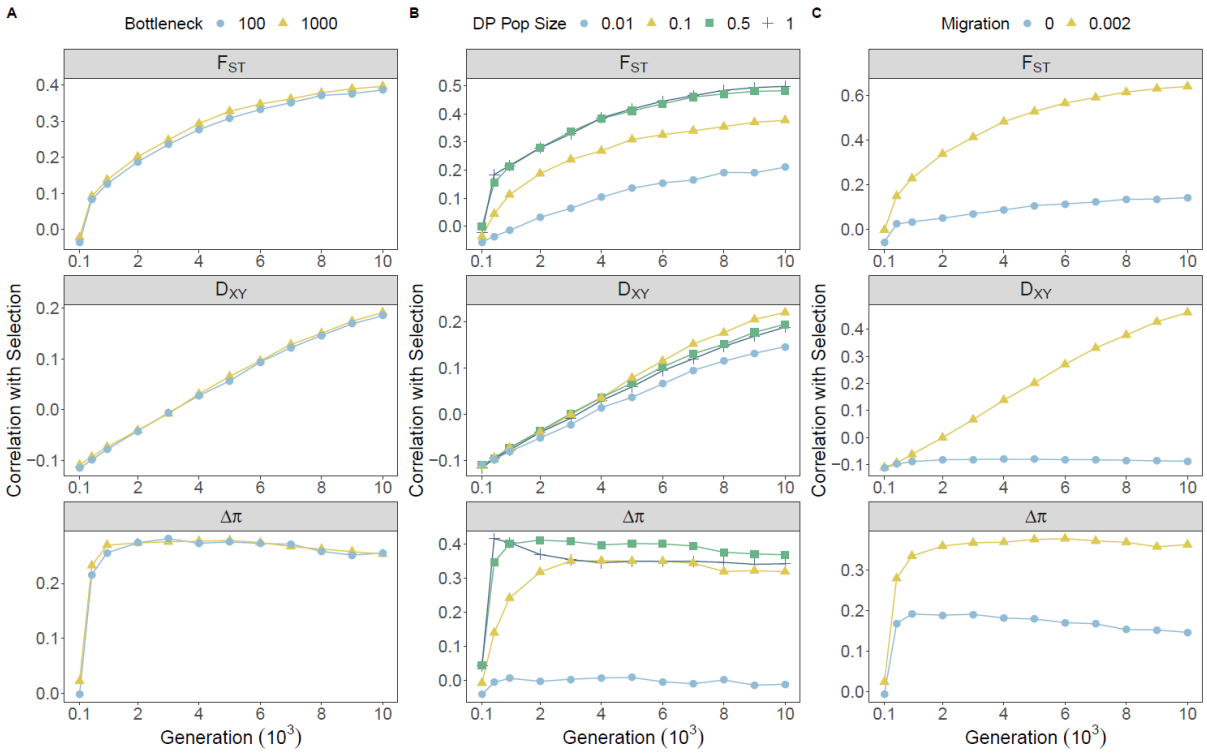
Protracted bottlenecks had substantial effects on relative (F_{ST} and $\Delta\pi$) measures and marginal effects on absolute (D_{XY}) measures of divergence (Figure 3B). Relative measures were less correlated with selection when DP size was reduced (0.01 or 0.1) than with minimal (0.5) or no (1.0) reductions in DP size, which were generally similar. For F_{ST}, the

variance in correlation coefficients between DP size treatments was generally stable across time, whereas variance between treatments for Δπ peaked at 500 generations. For D_{XY}, correlations with selection were broadly consistent across DP size reductions up until 10,000 generations, in which associations with selection were reduced by the strongest protracted bottlenecks, but generally robust (Figure 3B).

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 Δπ across simulations. This observation again highlights the significance of temporal differences between measures. Interestingly, correlations between Δπ and selection persisted, albeit weakly, in the absence of migration, which was not the case for F_{ST} and D_{XY}.

Protracted 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 associations between selection and D_{XY} in $Pheno_{Div}$ simulations are likely more dependent on adaptive substitutions in order to overcome this effect. Δπ was generally positively associated with selection in $Pheno_{Null}$ simulations regardless of migration, but was slightly reduced when migration was absent.

405



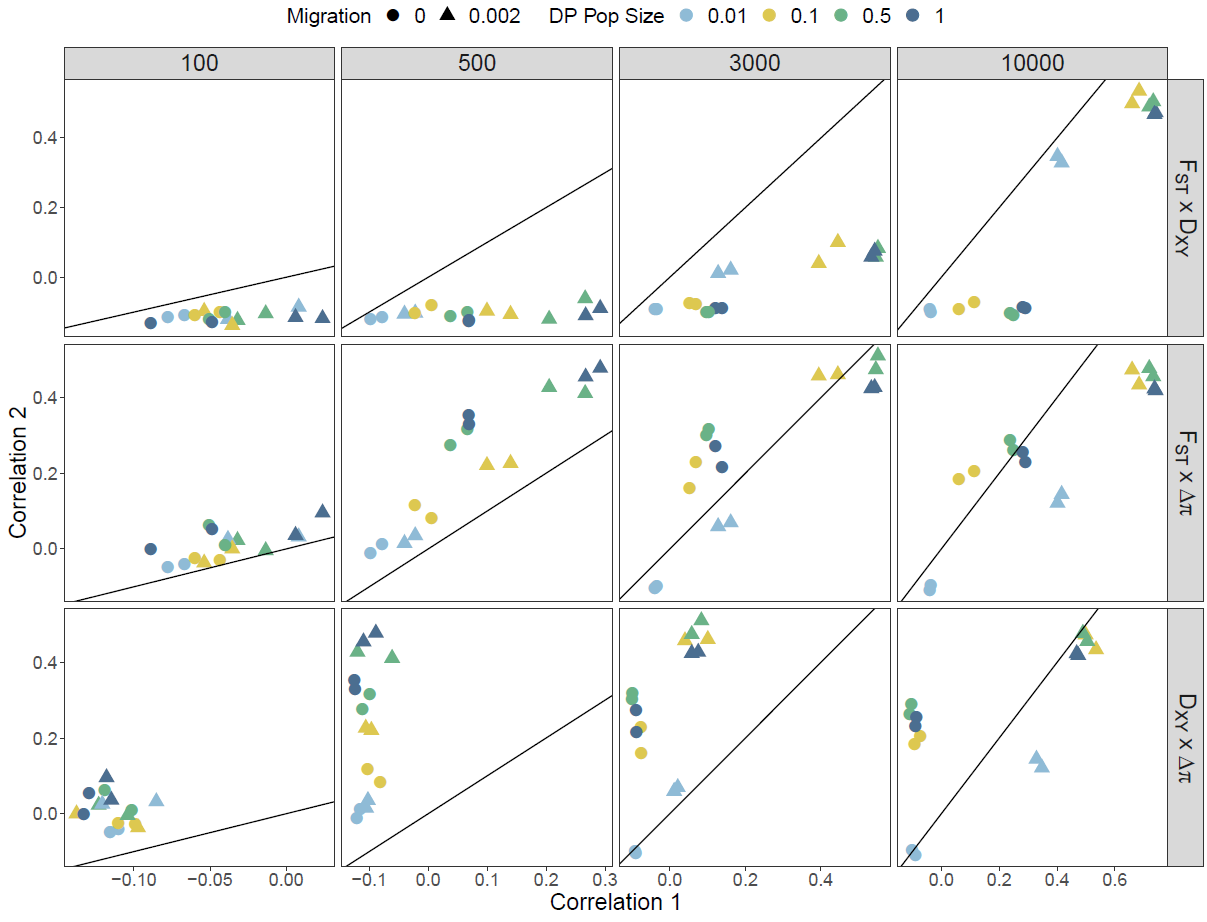
406

407 **Figure 3:** Effects of demographic treatments on the relationship between selection and measures of
408 genetic divergence across all sampling generations. Point colour and shape denote treatment groups
409 for founding bottlenecks (**A**), protracted bottlenecks (**B**) and migration (**C**). Each point represents
410 correlation coefficients calculated across all genes within individual treatments groups, averaged
411 within treatment levels, and averaged over 20 iterations.
412

413

414 By examining the correlation coefficients of all 16 unique demographic histories, we can
415 investigate the combined effect of migration and protracted bottlenecks and directly
416 compare effectiveness of individual measures across time (Figure 4). F_{ST} consistently
417 outperforms D_{XY} in terms of associating with selection under most demographic treatments,
418 particularly when DP sizes are larger and migration is present. By 10,000 generations
419 however, the relative dominance of F_{ST} appears to subside, with the trend through time
420 suggesting a relative improvement in D_{XY} in treatments with gene flow (Figure 4). Δπ is
421 similarly more informative than D_{XY} across sampling generations under most demographic
422 treatments. Interestingly at sampling generation 10,000, protracted bottlenecks produce

423 the biggest bias towards Δπ (Figure 4). The resilience of Δπ under no-migration treatments
424 is also apparent in F_{ST} - Δπ comparisons, such that at 3,000 generations Δπ is more
425 informative than F_{ST} in the absence of migration. Consistent with its rapid response to
426 sweeps around 500 generation, Δπ outperforms F_{ST} under most demographic scenarios in
427 early generations. By 10,000 generations, however, F_{ST} performs as well as Δπ without
428 migration and outperforms Δπ with migration.
429



430

431 **Figure 4:** Pairwise comparisons between the correlation coefficients of selection with F_{ST}, D_{XY} and Δπ
432 across four sampling points. Each data point represents a unique demographic treatment with
433 points coloured according to DP population size and shaped according to migration level. Correlation
434 1 refers to the first measure listed in the comparison and Correlation 2 to the second. The y=x line is
435 plotted within each facet to illustrate biases towards one measure. Points below the line are biased
436 towards Correlation 1, whilst points above the line are biased towards Correlation 2. Each point
437 represents correlation coefficients calculated across all genes within individual treatments groups
438 and averaged over 20 iterations.
439

440

441 ***Demography moderates the shape and tail end of divergence distributions***

442 By comparing distributions across simulations with divergent ($Pheno_{Div}$) and stabilising
443 ($Pheno_{Null}$) selection with neutral runs, we can examine the effect of demographic
444 treatments on the ability of each measure of divergence to discriminate between them
445 (Figure S11-13; Figure 5). There are few differences between distributions of F_{ST} for the
446 three simulation types when migration is absent between AP and DP replicates, highlighting
447 an increased likelihood of false-positives. The exceptions occur around 500 generations
448 (Figure S12) when sweeps are most common, and towards the end of simulations. Following
449 probable sweeps, $Pheno_{Div}$ F_{ST} is slightly elevated, and towards the end of simulations
450 $Pheno_{Div}$ F_{ST} is marginally lower than $Pheno_{Null}$ F_{ST}, which is marginally lower than neutral F_{ST}
451 (Figure 5). With migration, we see few differences between $Pheno_{Null}$ and neutral F_{ST}.
452 $Pheno_{Div}$ F_{ST} distributions however become more positive and flattened, according to
453 variable selection, with the majority of $Pheno_{Div}$ F_{ST} above the 95% quantiles of $Pheno_{Div}$ and
454 neutral F_{ST} by 10,000 generations (Figure 5).

455

456 Similar patterns were observed for D_{XY} distributions (Figure S14-17), with little to
457 discriminate between in treatments without migration. However, without migration, neutral
458 D_{XY} was generally reduced relative to $Pheno_{Null}$ and $Pheno_{Div}$ D_{XY}. With gene flow, like F_{ST},
459 $Pheno_{Div}$ D_{XY} was readily distinguishable from $Pheno_{Null}$ and neutral distributions, but
460 $Pheno_{Null}$ D_{XY} was also generally more positive than neutral D_{XY}. These patterns also emerged
461 more slowly than for F_{ST}. In contrast, $Pheno_{Div}$ $\Delta\pi$ (Figure S18-21) was elevated according to
462 DP size, such that at 500 generations the majority of windows under divergent selection
463 exhibited $\Delta\pi$ above neutral and $Pheno_{Null}$ 95% cut-offs for all treatments with DP size ≥ 0.5.

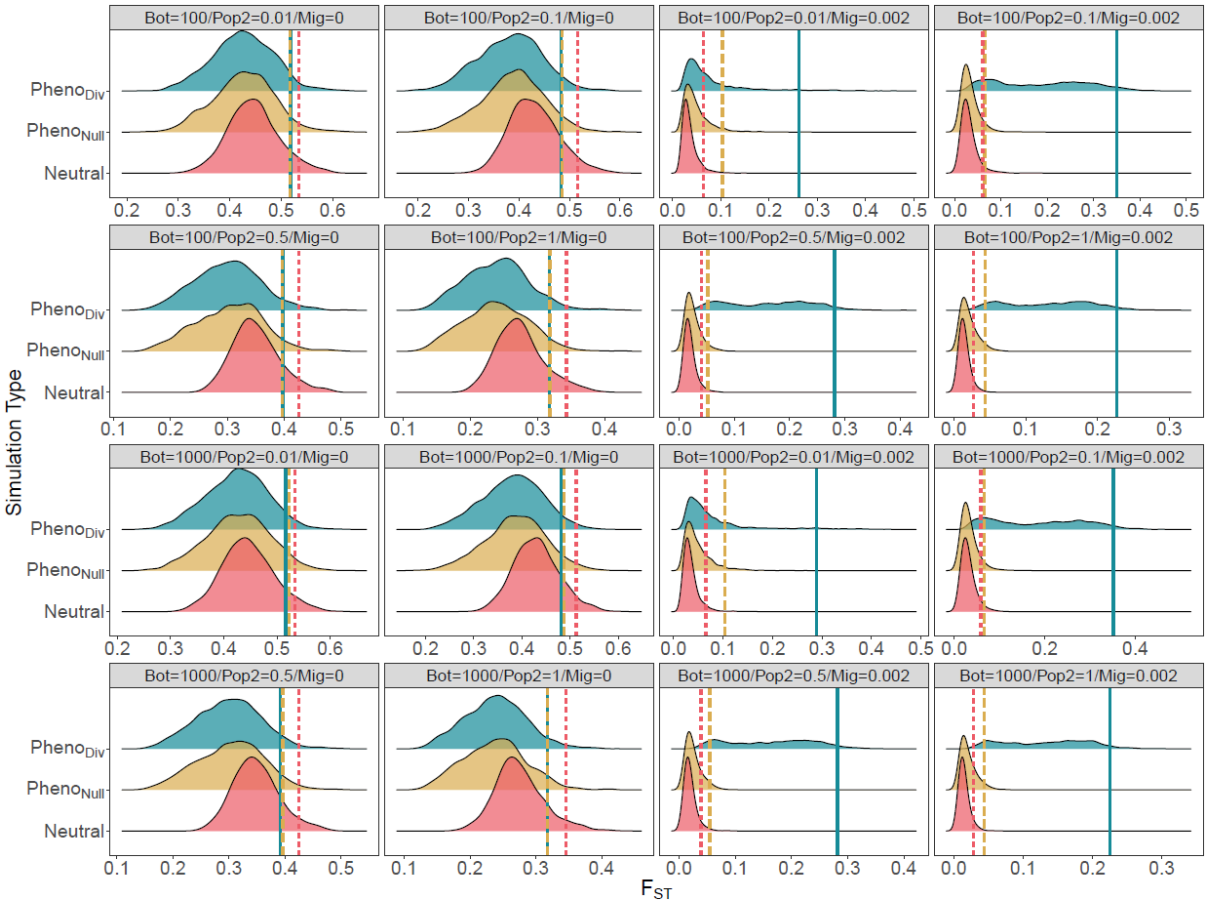


Figure 5: Distributions of F_{ST} under each of the 16 unique demographic treatments under three selection regimes: $Pheno_{Div}$ (divergent selection), $Pheno_{Null}$ (stabilising selection) and Neutral, after 10,000 generations. Upper 5% quantiles are highlighted for each distribution, with linetype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

We quantified false-positive rates (FPR) by permuting over merged data comprised of randomly sampled 5% $Pheno_{Div}$ windows and 95% neutral windows and observing the upper 5% quantile (Table S1). By 100 generations, F_{ST} FPR ranged between 0.35 and 0.93, and was lower with increased DP size and lower in treatments without migration (Table S1). By 500 generations (Table 2), FPR rates were lower with migration and higher without, but only for treatments with DP sizes of 0.5 and 1.0. F_{ST} FPR remained high (> 0.85) for all treatments with smaller DP sizes. By 3,000 generations, migration was the most important demographic

480 factor for F_{ST} FPR. With migration, FPR ranged from 0.14 – 0.70, and decreased with
481 increasing DP size. Without migration, FPR were high (0.93-0.97), in line with the random
482 proportion of neutral (0.95) data. By the end of simulations, F_{ST} FPR was as low as 0.08, and
483 was no greater than 0.26 with migration and DP size ≥ 0.1. Without migration, FPR was >
484 0.97.
485
486 Initial D_{XY} FPR were generally high irrespective of demography, ranging between 0.81 and
487 0.87. FPR were largely similar after 500 generations, but by 3,000 generations there was a
488 distinction between treatments with ($0.33 \leq FPR \leq 0.65$) and without ($0.87 \leq FPR \leq 0.90$)
489 migration. Interestingly, here FPR rates were lower ($0.33 \leq FPR \leq 0.46$) when DP were
490 smaller (size = 0.01, 0.1) rather than larger ($0.57 \leq FPR \leq 0.65$). This pattern was also
491 observed at the end of simulations, with FPR lower without migration ($0.07 \leq FPR \leq 0.36$)
492 and lowest with DP size = 0.1.
493
494 $\Delta\pi$ FPR rates were generally higher across all sampling generations, with FPR not falling
495 below 0.35. There was a clear distinction based on DP size in earlier (100 and 500) sampling
496 points, with FPR lower with larger DP size. However, at later sampling generations (3,000
497 and 10,000), FPR were generally high (≥ 0.70) regardless of demography.
498
499 Taken together, there are clear effects of DP size and migration on distributions of genetic
500 variation and upper quantiles of interest. DP size appears initially most important in the first
501 few hundred generations for F_{ST} and $\Delta\pi$, but these effects are later swamped by the effect
502 of migration for F_{ST} and are simply eroded for $\Delta\pi$. D_{XY} exhibits similar patterns to F_{ST}, but
503 these develop after many more generations, and whilst gene flow increases the

504 informativeness of D_{XY} for detecting divergent selection, as it does for F_{ST}, D_{XY} and F_{ST}

505 experience opposing effects of increases to DP size.

506

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. Data represent means calculated over 100 downsampled permutations.

Generation	Demographic Treatments		Single Measures			Combined Measures											
			F_{ST}	D_{XY}	$\Delta\pi$	F_{ST} - D_{XY}			F_{ST} - $\Delta\pi$			D_{XY} - $\Delta\pi$			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	0.89	0.87	0.96	67.30	0.88	0.92	13.64	0.94	0.99	35.26	0.95	0.98	13.64	0.94	0.99
500		0.1	0.93	0.86	0.99	64.78	0.91	0.94	14.23	0.97	1.00	24.68	0.98	0.99	13.80	0.97	1.00
500		0.5	0.76	0.83	0.62	52.37	0.76	0.87	30.88	0.54	0.86	16.72	0.58	0.93	16.63	0.58	0.93
500		1	0.65	0.82	0.43	44.40	0.67	0.85	34.83	0.28	0.75	15.54	0.37	0.90	14.51	0.35	0.90
500	0.002	0.01	0.92	0.76	0.97	64.81	0.90	0.93	39.14	0.94	0.98	37.16	0.94	0.98	32.55	0.93	0.98
500		0.1	0.88	0.80	0.98	36.35	0.85	0.95	32.22	0.95	0.98	9.19	0.93	0.99	9.19	0.93	0.99
500		0.5	0.49	0.84	0.52	25.47	0.57	0.89	51.96	0.28	0.63	9.78	0.41	0.94	9.78	0.41	0.94
500		1	0.33	0.84	0.36	20.41	0.30	0.86	57.03	0.08	0.47	9.48	0.00	0.91	9.42	0.00	0.91
10000	0	0.01	0.97	0.89	0.93	60.27	0.97	0.98	0.00	0.00	1.00	5.55	0.85	0.99	0.00	0.00	1.00
10000		0.1	0.99	0.88	0.99	60.02	0.98	0.99	0.00	0.00	1.00	5.09	0.96	1.00	0.00	0.00	1.00
10000		0.5	0.98	0.88	0.99	56.39	0.98	0.99	0.99	0.95	1.00	2.01	0.98	1.00	0.99	0.95	1.00
10000		1	0.98	0.88	0.97	64.80	0.98	0.99	4.86	0.99	1.00	7.04	0.96	1.00	4.86	0.99	1.00
10000	0.002	0.01	0.65	0.34	0.84	62.76	0.46	0.66	38.95	0.67	0.87	29.95	0.55	0.86	28.85	0.57	0.87
10000		0.1	0.26	0.07	0.98	78.70	0.06	0.26	12.71	0.85	0.98	4.75	0.63	0.98	4.75	0.63	0.98
10000		0.5	0.13	0.14	0.92	81.99	0.01	0.19	9.62	0.20	0.92	7.38	0.00	0.93	7.38	0.00	0.93
10000		1	0.09	0.19	0.90	80.70	0.03	0.22	10.29	0.00	0.90	9.75	0.00	0.90	9.75	0.00	0.90

507

508
509 ***Demography moderates relationships between measures of divergence***

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

515

516 Positive correlations between F_{ST} and D_{XY} emerged rapidly irrespective of DP size, but
517 smaller DP sizes generally increased the correlation (Figure 6B), with variance between
518 treatments generally decreasing over time. Similarly, F_{ST} and $\Delta\pi$ were generally positively
519 correlated, however reductions in DP size reduced correlation coefficients. By 4,000
520 generations, F_{ST} - $\Delta\pi$ correlations for DP sizes ≥ 0.1 stabilised, but correlations under
521 extreme protracted bottlenecks continued to decline to a low of 0.08 (Figure 6B). D_{XY} - $\Delta\pi$
522 correlations were generally weaker than other comparisons across the course of
523 simulations, but were minimally affected by protracted bottlenecks. Interestingly,
524 correlations with D_{XY} and both F_{ST} and $\Delta\pi$ appear to peak for treatments with minimal
525 population size reductions and without migration at 500-2,000 generations (Figure 6B),
526 suggesting positive correlations with D_{XY} are in part driven by the generation in which
527 sweeps are taking are place.

528

529 Migration induced substantial variance between correlations of divergence measures, with
530 effects dependent on sampling point (Figure 6C). In the absence of migration, F_{ST} and D_{XY}

were more strongly correlated for the first 6,000 generations than in treatments with gene flow. However, from here until 10,000 generations this pattern reversed and F_{ST} - D_{XY} correlations increased with gene flow and deteriorated in allopatry. F_{ST} - $\Delta\pi$ correlations were strong initially, but a lack of migration weakened the correlation over time until measures were uncorrelated by around 3,000 generations. In contrast, in the presence of migration, positive correlations between F_{ST} and $\Delta\pi$ were relatively stable (R^2 = 0.51 – 0.38 over the whole simulation period). Interestingly, without migration, D_{XY} and $\Delta\pi$ were largely uncorrelated, but migration induced a positive correlation between D_{XY} and $\Delta\pi$ that emerged after 1,000 generations and continued to increase through time.

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

linked to selection (drift with ineffective selection) likely drive increased F_{ST} and D_{XY} but not $\Delta\pi$.

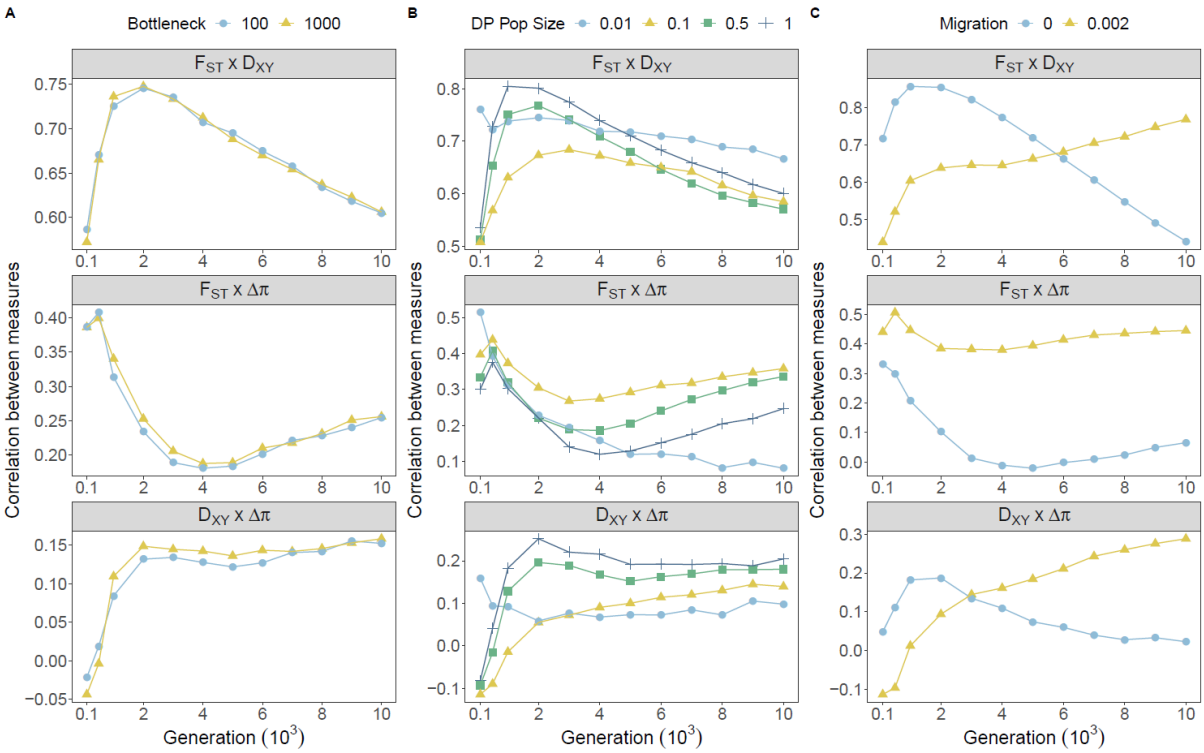


Figure 6: Effects of demographic treatments on the relationship between measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for founding bottlenecks (**A**), protracted bottlenecks (**B**) and migration (**C**). Each point represents correlation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

In $Pheno_{Null}$ simulations, F_{ST} and D_{XY} were positively correlated in all treatments, but stronger associations linked with demographic treatments with effective selection did not emerge (Figure S22). This was also true for neutral simulations (Figure S23), but F_{ST} - D_{XY} correlations were larger than for $Pheno_{Div}$ and $Pheno_{Null}$ when selection was ineffective, reaching a maximum of $R^2 = 0.89$ without migration (Figure S23C). This demonstrates that the positive correlations in $Pheno_{Div}$ simulations are driven in part by divergently adaptive allele frequency shifts and substitutions when selection is effective, but these correlations also

571 emerge under stabilising selection or neutrality. Specifically, strong positive correlations
572 with Δπ were dependent on divergent selection, and F_{ST} – Δπ correlations became negative
573 over time in treatments without migration under neutrality. D_{XY} – Δπ correlations became
574 negative in $Pheno_{Null}$ data with migration and were unassociated without. Thus, these
575 results highlight that the relationships between measures of divergence are highly
576 dependent on demography, time, and selection experienced over a genomic region.
577
578 We then examined FPR and false negative rates (FNR) when combining outliers across F_{ST},
579 D_{XY} and Δπ (Table S1; summaries from 500 and 10,000 generations in Table 2). Combined F_{ST}
580 - D_{XY} outliers exhibited FPR rates that were reasonably high (0.30 - 0.93) and similar to F_{ST}
581 alone at generations 100 and 500, suggesting little improvement in reducing FPR. During
582 this period, FNRs were also high (≥ 0.84), suggesting most windows with divergent selection
583 could not be detected on a neutral backdrop. It was not until 10,000 generations that
584 reductions in FPR were evident for combined F_{ST} - D_{XY} outlier sets for treatments with
585 migration, in some cases dropping to 0.01, although FNR were highly variable (0.19 ≤ FNR ≤
586 0.66). High FPR (0.96 ≤ FPR ≤ 0.98) and high FNR (0.98 ≤ FPNR ≤ 0.99) were observed
587 without migration, highlighting most common outliers between F_{ST} and D_{XY} to be neutral,
588 and a failure to detect almost all divergent windows.
589
590 Combined F_{ST} - Δπ outliers did marginally outperform F_{ST} and Δπ outliers according to FPR
591 with DP sizes of 0.5 or 1.0 and in earlier (100 and 500) sampling generations. Here, FPR
592 dropped to a low of 0.08, but FNR were reasonable through this time (≥ 0.43). Beyond this
593 (sampling generations 3,000 and 10,000), FPR did drop to 0, however these were generally
594 alongside high FNR of up to 1.0, highlighting a failure to detect any common outliers at all. A

595 good example of improvement over singular measures was observed at generation 500,
596 with no founding bottleneck, equal sized populations, and migration. Here, an average of
597 ~53% of divergent windows were detected with an average FPR of 0.08. This is compared
598 with: singular F_{ST}, where FPR/FNR = 0.33; and singular $\Delta\pi$ where FPR/FNR = 0.36. By 10,000
599 generations, combined outlier sets of F_{ST} - $\Delta\pi$ performed poorer than singular F_{ST} with
600 migration present, but generally returned low numbers of false positives, unlike F_{ST} - D_{XY}.
601
602 Combined D_{XY} - $\Delta\pi$ outliers performed poorly across all treatments and all sampling
603 generations, with FNR failing to fall below 0.86. There were, however, initial benefits in
604 terms of reduced FPR in earlier sampling generations for treatments with larger DP sizes
605 (0.5 and 1.0). This discordance between D_{XY} and $\Delta\pi$ and large FNR also limited the combined
606 usage of all three measures together, with similarly high FNR precluding their combined
607 usage.
608
609 These results therefore highlight that combining measures can help reduce FPR, but usually
610 at the cost of increased FNR (expectedly), and only under certain demographies. Our
611 observation that high FPR are prevalent among combined outlier sets from statistics,
612 particularly in the absence of migration, suggests their usage must be dependent on a
613 knowledge of the underlying demographic history of populations. These findings also
614 highlight that the strong correlations that emerge between measures of divergence under
615 scenarios with ineffective selection or even under neutrality do extend to the tail-ends of
616 distributions.
617
618 ***Demography drives signals of convergence irrespective of selection***

619 Clusters of overlapping outliers developed steadily over time (Figure S26-28). By sampling
620 generation 10,000, significant proportions of overlapping outliers were recovered across
621 different demographic treatment groups (Figure 7). For F_{ST} and D_{XY}, clustering of treatments
622 was driven by the presence/absence of migration, with the highest proportions of
623 convergent outliers observed between treatments with migration. Interestingly, for both
624 measures, reasonable proportions of convergent outliers were recovered across no
625 migration treatments, which we have demonstrated to have ineffective selection.
626 Importantly, there was little overlap between these clusters, suggesting different
627 convergent outliers within each cluster. Combining outliers from F_{ST} and D_{XY} did not alter the
628 patterns observed in either, and importantly did not remove convergent outliers across no-
629 migration demographies. There were minimal convergent outliers observed for Δπ outliers,
630 however combining Δπ with F_{ST} and D_{XY} did appear somewhat effective in removing
631 convergent outliers found between no-migration treatments. However, for both
632 combination of Δπ with F_{ST} and D_{XY}, the highest proportional overlap was observed for
633 treatments with the smallest DP size.
634
635 Interestingly, the clustering of F_{ST} - D_{XY} outliers in the presence of migration was greatly
636 reduced when divergent selection was removed in both $Pheno_{Null}$ (Figure S29-32) and
637 neutral data (Figure S33-36), but clusters of outliers in the absence of migration were still
638 apparent.
639

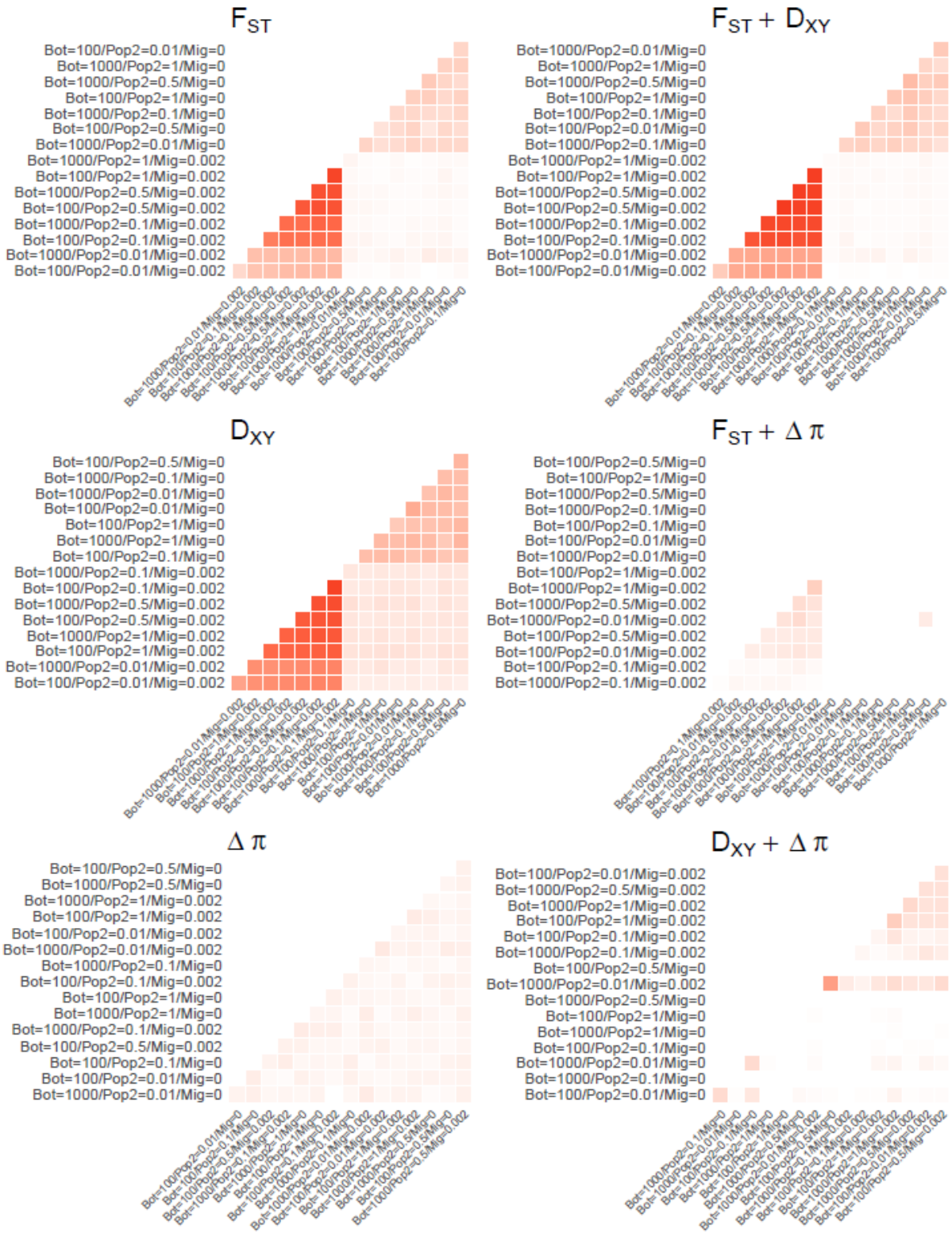


Figure 7: The proportional overlap of outliers above the 95% quantile, averaged across 100 downsampled datasets consisting of 95% Neutral and 5% $Pheno_{Div}$ data for each of the 16 demographic treatments after 10,000 generations. Axis orderings were determined through hierarchical clustering. Heatmaps 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.

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 39.6% of variance in F_{ST} with migration. Conversely, F_{ST} in treatments without migration was strongly positively associated with selection target % (Table 3), with a much weaker positive association with selection (Table 3). However, fixed effects here explained only 6.6% variance in F_{ST} without migration.

Table 3: LMM results for models of measures of divergence explained by features of simulated genes. For all models, random variables included gene ID and demographic treatment. Fixed effects with the largest effect are highlighted in bold. Selection Target and Pheno Gen were log_{10}-transformed in models.

		Var. explained (%)							
Measure	Migration	Fixed	Random	Fixed Effect	Estimate	Std. Error	df	*T*	*P*
F_{ST}	0.002	39.60	60.90	**Selection**	**0.100**	**0.004**	**98**	**25.48**	**<2.00E-16**
				Selection Target	0.009	0.002	103	3.89	0.000177
				Pheno Gen	-0.007	0.001	14530	-9.45	<2.00E-16
F_{ST}	0	6.60	75.40	**Selection Target**	**0.025**	**0.002**	**97**	**11.44**	**<2.00E-16**
				Selection	0.011	0.004	97	3.04	0.00302
D_{XY}	0.002	18.60	69.40	**Selection**	**0.024**	**0.001**	**97**	**16.42**	**<2.00E-16**
				Selection Target	-0.010	0.001	102	-11.70	<2.00E-16
				Pheno Gen	-0.003	0.000	14570	-14.70	<2.00E-16
D_{XY}	0	13.50	33.30	**Selection Target**	**-0.008**	**0.001**	**99**	**-8.00**	**2.39E-12**
				Pheno Gen	0.000	0.000	14480	2.37	0.018
Δπ	0.002	5.00	60.90	**Selection**	**0.168**	**0.011**	**99**	**15.67**	**<2.00E-16**
Δπ	0	0.70	76.70	**Selection**	**0.086**	**0.018**	**99**	**4.73**	**7.63E-06**
				Pheno Gen	-0.009	0.003	12870	-2.70	0.00694

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 selection target % (Table 3), contrary to F_{ST}. In treatments without migration, selection target % was strongly negatively associated with D_{XY} (Table 3), with a weak positive effect of Pheno Gen (Table 3).

Selection was the most important fixed effect for Δπ regardless of migration (Table 3), however both models explained minimal variance (5% with migration, 0.7% without

682 migration). This is consistent with the previous demonstration of the erosion of Δπ over
683 time (Figure 1).
684
685 Removing divergent selection (i.e. in $Pheno_{Null}$ simulations) modified models for F_{ST} and D_{XY}
686 for treatments with migration. Selection and selection target remained the most important
687 model effects, but had similar sized effects, and ultimately variance explained dropped from
688 39.6% to 5.5%. Conversely, the model for $Pheno_{Null}$ D_{XY} with migration increased variance
689 explained from 18.6% to 32.9% compared with $Pheno_{Div}$ D_{XY}. Selection target had a greater
690 effect in this model than strength of selection. Δπ models were unchanged. Together, these
691 results confirm that divergent selection, and not stabilising selection within DP, drive F_{ST} and
692 D_{XY} variation in $Pheno_{Div}$ simulations.
693
694 **DISCUSSION**
695 ***Summary of results***
696 Here, we show that different demographies can have dramatic effects on how measures of
697 population divergence identify regions of the genome under selection. Effects are also
698 strongly time-dependent. Using simulated populations, we have demonstrated the relative
699 influences of founding bottlenecks, protracted bottlenecks, and migration on three
700 commonly used measures of genetic divergence (F_{ST}, D_{XY}, Δπ), whilst demonstrating the
701 relative usefulness of each measure for informing on selection under different demographic
702 conditions.
703
704 We find that founding bottlenecks have little effect on population divergence measures,
705 potentially either because populations quickly recover (before first sampling after 100

706 generations), or founding bottlenecks of 10% (100 individuals) were not extreme enough.
707 Protracted bottlenecks (reductions in DP size) however, artificially inflate F_{ST} and $\Delta\pi$ but
708 reduce D_{XY}, and can erase the relationship of F_{ST} and D_{XY} with selection under the most
709 extreme reductions in population size. Relative measures are, in part, driven by intra-
710 population changes in allele frequency, which become exaggerated in smaller populations
711 as a product of drift (B. Charlesworth 2009; Ellegren and Galtier 2016). As a consequence,
712 we observe inflated measures of relative divergence as allele frequencies drift in smaller DP
713 replicates.
714
715 In contrast, D_{XY} increases with larger DP size, as a product of the relationship between the
716 number of segregating sites and the population-level mutation rate ($4N_e\mu$) (Hartl, Clark, and
717 Clark 1997). D_{XY} is a measure of sequence divergence and is averaged across all sites
718 (although similar statistics limit averaging to segregating sites only), which results in higher
719 D_{XY} as segregating sites are introduced into either population at a rate of $4N_e\mu$. This
720 relationship with the number of segregating sites can be observed by examining the positive
721 relationships between D_{XY} and DP_π (Figure S37). Overall, the relationship between D_{XY} and
722 selection is less affected by protracted bottlenecks than F_{ST} and $\Delta\pi$, likely due to the lack of
723 allele frequency relevance. Consider for example, two SNPs with minor allele frequencies of
724 0/0.5 (SNP 1) and 0.5/0.5 (SNP 2) in AP/DP. Each locus contributes equally towards D_{xy} (SNP
725 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
726 within-population variance observed for SNP 1 inflates F_{ST} and $\Delta\pi$. However, we also see
727 evidence of D_{XY} FPR increasing with increased DP size (Table 2), suggesting this relationship
728 with increased acquisition of segregating sites may conflict with increased efficacy of
729 selection.

730

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

740

741 Migration also exhibited interesting temporal patterns that are useful for discussing the
742 discrepancies observed between $\Delta\pi$ and both F_{ST} and D_{XY}. Of the measures considered here,
743 $\Delta\pi$ uniquely disregards SNP substitutions in its calculation, as fixed substitutions have no
744 influence on intra-population heterozygosity beyond reducing the heterozygosity of
745 proximate SNPs in the aftermath of a hard sweep. This characteristic explains why $\Delta\pi$
746 responds to selection more rapidly than F_{ST} and D_{xy} and is influenced by migration in a
747 constant manner across time. In addition, the inability of $\Delta\pi$ to account for adaptive
748 substitutions is likely why the informativeness of $\Delta\pi$ peaked at around 500 generations
749 (approximate median for sweeps), decayed, and then stabilised by 10,000 generations. In
750 contrast, the contributions of adaptive substitutions to F_{ST} and D_{xy} over time increases their
751 predictive power. This characteristic of $\Delta\pi$, however, also makes it unable to differentiate
752 between divergent and stabilising selection. This temporal observation has important
753 implications for studies of rapid adaptation on the scale of 10s to 100s of generations, in

754 which our simulations suggest Δπ may be the most informative measure of divergence,
755 whilst D_{xy} is initially uninformative for several thousand generations.
756
757 Examining the effects of demographic variation on the associations between measures of
758 divergence themselves is a novel element to this study. The relative weaknesses of
759 individual measures of divergences has prompted a recent movement within the literature
760 to employ multiple measures of divergence to avoid false-positives (for e.g. Tine et al. 2014;
761 Malinsky et al. 2015; Hämälä and Savolainen 2019). Our results provide some support for
762 this strategy, with reductions in FPR in treatments with gene flow generally, and low FPR
763 when combining either F_{ST} or D_{XY} with Δπ (albeit at a reasonably high FNR). However, we
764 also find large number of overlapping outliers across combined F_{ST} - D_{XY} with a large FPR
765 across treatments without migration. Further, these false-positives persisted in $Pheno_{Null}$
766 and neutral data, whereas clusters of treatments with overlapping outliers (Figure 7) and
767 with low FPR were restricted to simulations with divergent phenotypes. It is clear then that
768 combining outlier sets of F_{ST} and D_{XY} only improves analyses under certain demographic
769 histories.
770
771 Cruickshank and Hahn (2014) suggested that a disagreement between F_{ST} and D_{XY} outliers in
772 genomic-islands-of-divergence tests highlights a particular susceptibility of F_{ST} to BGS
773 (although see (Matthey-Doret and Whitlock 2019)). Our results are in line with the notion
774 that D_{XY} may be more resistant to false positives due to BGS, given that reductions in DP size
775 resulted in minimal variance in D_{XY} (assuming reductions in population size are analogous to
776 different rates of BGS across the genome exhibit reduced N_e). However, BGS was not
777 specifically manipulated in these simulations, and results are difficult to disentangle with

778 variation in efficacy of removing deleterious mutations. Further, this minimal variance may
779 be explained by the opposing forces of selection efficacy and acquisition of segregating sites
780 discussed previously. We also found that a greater proportion of variance could be
781 explained by the size of selection target for D_{XY} in the absence of migration than for F_{ST}. We
782 found that F_{ST} and D_{XY} are positively correlated under several demographic scenarios, but
783 this strong association only reflects selection when gene-flow is present and only after
784 several thousand generations, consistent with previous simulation work (Ravinet et al.
785 2017). When migration is absent, and selection ineffective, F_{ST} and D_{XY} are also positively
786 correlated; but this relationship decays over time (Figure 6C), with empirical evidence (see
787 below) suggesting a negative relationship is likely to emerge without gene flow. No-
788 migration treatments are consistent with Isolation-By-Distance demographic histories and,
789 similarly, comparisons between reproductively-isolated populations or species. Indeed,
790 negative correlations between F_{ST} and D_{XY} within and between clades of birds (Vijay et al.
791 2017; Irwin et al. 2016), within a radiation event of monkeyflowers (Stankowski et al. 2018)
792 and between speciating orca populations (Foote et al. 2016) support a declining relationship
793 between these measures over long periods of time in isolation. In agreement, we find that
794 without migration F_{ST} and D_{XY} exhibit opposite associations with the proportion of windows
795 made up of coding elements. Mechanistically, a larger selection target increases the rate of
796 deleterious mutation, reducing local π directly through loss of polymorphic deleterious sites,
797 or indirectly through the loss of linked neutral variants under a BGS model. Reductions in
798 local π increase F_{ST} and decrease D_{XY}. Over time, associations between F_{ST} and D_{XY} should
799 stabilise, given π is generally well-conserved in stable populations even across long time
800 periods (Romiguier et al. 2014; Van Doren et al. 2017; Dutoit et al. 2017).
801

802 By comparing our results to a second dataset that lacked divergent selection ($Pheno_{Null}$), we
803 found consistent support that positive associations between D_{XY} and selection are strongly
804 dependent on the inclusion of a divergent phenotype. Positive associations with F_{ST} are
805 attainable only in the absence of migration, and are weakly negative without, and $\Delta\pi$
806 patterns are primarily driven by variable stabilising selection and made weaker by the
807 inclusion of phenotypic divergence. These comparisons are useful in highlighting the relative
808 roles of adaptive allele frequency changes and substitutions in driving patterns of genetic
809 differentiation. It is also of interest to note that overlapping outliers are readily attainable
810 across no-migration treatments in $Pheno_{Null}$ (Figure S32) and neutral simulations (Figure
811 S36), but not for treatments with migration. This confirms that overlapping outliers in no-
812 migration treatments occur due to common neutral processes within genomic windows,
813 whereas overlapping outliers in treatments with migration are driven by the effects of
814 divergent selection. The discrepancies between our $Pheno_{Div}$ and other simulated datasets
815 highlight the necessity in quantifying phenotypic differences or environmental selection
816 pressure when interpreting patterns of variation across the genome.
817
818 ***Detecting genetic convergence***
819 In addition to understanding how outlier detection in individual pairs were affected by
820 demography, we wanted to explore how studies looking at overlapping outliers in multiple
821 pairs (i.e. detecting genetic convergence) were affected by demography. Our simulation
822 design, in which an ancestral burn-in population is used to found 16 independent AP-DP
823 pairs is analogous to replicated ecotype population pairs in model systems, such as various
824 ecotype pairs of the three-spined stickleback, *Gasterosteus aculeatus* (Hohenlohe et al.
825 2010; Jones, Chan, et al. 2012); high/low predation Trinidadian guppies, *Poecilia reticulata*

826 (Fraser et al. 2015); crab/wave ecotype periwinkles, *Littorina saxitilis* (Ravinet et al. 2016;
827 Kess, Galindo, and Boulding 2018; Westram et al. 2014), and alpine/montane ecotypes of
828 *Heliosperma pusillum* (Trucchi et al. 2017). We found overlapping outliers between
829 demographic treatments and thus, that signals of convergent outliers are attainable for
830 singularly used F_{ST} and D_{XY}, and combined outlier sets for $F_{ST} - \Delta\pi$, $D_{XY} - \Delta\pi$ and F_{ST} - D_{XY}.
831 Clustering of outliers was predominantly driven by the presence or absence of migration
832 (with minimal overlap between clusters) (Figure 7).
833
834 The overlap of quantile-based outliers between demographic treatments without migration
835 and ineffective selection may be explained in part probabilistically. With migration
836 restricted, AP and DP exhibit significantly elevated measures of divergence, as seen in Figure
837 1. This increase results in a normal distribution of F_{ST} and D_{XY} across neutral simulations
838 without migration. In contrast, with migration, drift is limited and random recombination
839 influences gene flow and subsequently variation in divergence. Distributions of divergence
840 with migration under neutrality are therefore are heavily right tail-skewed (Figure S38).
841 Spread of data increases with longer right tails, and density of data in each overlapping
842 distribution is limited to the median and lower quantiles.
843 We also expect some effect of different amounts of starting variation among genome
844 windows following burn-ins. With gene flow restricted, this variation may promote overlap
845 under neutral conditions given all demographic treatments are founded from a common
846 burn-in. However, this feature of our simulations is analogous to the conserved landscapes
847 of variation observed in natural genomes (Burri 2017; Vijay et al. 2017; Stankowski et al.
848 2019).
849

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 windows within 16 unique demographic treatments, assuming the effects of reductions in population size and migration are equivalent for all windows. This may be unrepresentative of genomes sampled from the wild, in which gene flow and effective population size can vary across genomic regions through structural variation, variable recombination rate and BGS (Gossmann, Woolfit, and Eyre-Walker 2011). A prominent example of such a mechanism is the well-characterised consequence of reduced gene flow within inversions that carry locally adapted alleles. Assuming inversion variants are fixed between populations, gene flow across the locus is limited by the prevention of recombinant haplotypes and resistance to introgression (Kirkpatrick and Barton 2006; Ravinet et al. 2017). Recent genome scan studies have highlighted convergent outliers within inversions (Jones, Grabherr, et al. 2012; Nishikawa et al. 2015; Morales et al. 2018), confirming theoretical models regarding their role in shielding adaptive haplotypes from introgression during the adaptation process. Our simulations suggest that genes with reduced migration, and reduced N_e as a consequence, will have inflated measures of divergence and differentiation relative to other genomic regions. Therefore, we predict this may have led to their over-representation in genome scan outliers, and increased potential

874 to overlap across replicate populations. Thus, caution should be taken regarding the
875 adaptive significance of these outliers relative to absolutely lower values of genetic
876 divergence attained from regions outside of chromosomal rearrangements.
877
878 By choosing to use a factorial design here, we have increased our understanding of the
879 interplay between individual features of demographic history and multiple measures of
880 population divergence. However, computational limitations constrained absolute
881 population sizes to a maximum of 1,000 individuals and generations to a maximum of
882 10,000 (with 10,000 generation burn-in). To mitigate these constraints, we repeated the
883 analysis over multiple mutation rates to illustrate patterns over 100-fold variation of θ. In
884 general, most trends were consistent, suggesting that results should be consistent across
885 taxa of variable effective population size. However, certain patterns were exaggerated or
886 dampened with increased or decreased mutation rate respectively. For instance, it is clear
887 that the time lag on informativeness of D_{XY} on selection is shortest when θ is largest, which
888 results in stronger positive associations between F_{ST} and D_{XY} by generation 10,000 (Figure
889 S24). It is well-documented that both N_e (Frankham 1995) and mutation rate (μ)
890 (Hodgkinson and Eyre-Walker 2011) are highly variable across taxa, which suggests that
891 applying knowledge of the relationships between measures of population differentiation
892 will vary in nature.
893
894 Furthermore, temporal variation within our simulations is confounded by the time at which
895 there was a major shift in the DP phenotype (Figure S39), and by extension when selective
896 sweeps occur. This variation in timing is random with respect to adaptive mutations arising
897 *de novo*, but is also influenced by demography. For example, treatments that experience

898 founding bottlenecks are less likely to evolve using variation in the founder, increasing
899 dependence on *de novo* mutations for adaptation. Additionally, variation in our DP size
900 parameter modifies the per population number of new mutations per generation. This is
901 also true for features of simulated genes, such as size of selection target. Predictable
902 temporal variation in the time at which adaptation is likely to occur is a probable source of
903 variance between measures of divergence and is particularly clear for $\Delta\pi$, but this was
904 controlled for in later modelling analyses as a fixed effect.
905
906 A further consideration for these simulations concerns the architecture of the phenotype.
907 Results reported here pertain to mutation effect sizes drawn from a distribution centred at
908 0 with $\sigma = 1$. This produces mutations of typically large effect, but was selected on the basis
909 of phenotypic optima being reasonably distant, with a difference of 10. Thus, 99% of
910 mutations in simulations produce phenotypic differences of less than a third the divergence
911 distance of AP and DP phenotypes. There are numerous factors that influence the
912 distribution of effect sizes in nature, including; selection, mutation, drift, gene flow, extent
913 of pleiotropy and distance to phenotypic optima (Dittmar et al. 2016), with no single
914 expectation for natural systems as a result. The relatively large distance between optima in
915 our simulations, as well as the rapid change in optima implemented in simulations (Collins
916 and De Meaux 2009), likely gives increased importance to mutations of large effect. The
917 interactions between mutation effect size and the results presented here are beyond the
918 scope of the current study, but we did investigate the effect of reducing σ of mutation effect
919 distributions to 0.1 (Supporting Information; Figures S40-45). Briefly, we see reductions in
920 the strength of correlations and associations with selection with decreasing phenotypic
921 effects of mutation, consistent with the notion of softer sweeps on small-effect loci. We also

see increased variance in the amount of time taken for simulations to reach the $Pheno_{Div}$ optimum, which agrees with the probable importance of large effect loci in our standard dataset. We, however, retain strong positive associations between measures such as F_{ST} and D_{XY}, as we see in neutral simulations, as well as overlapping outliers linked to 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 windows 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 indicate regions under selection are dependent on demography and variable through time, with a notable lag in D_{XY}. Furthermore, we provided novel comparisons

946 between measures of genetic divergence that call into question the use of multiple
947 measures to rule out false-positives. We also demonstrated that signals of convergent
948 evolution across independent replicates can be driven by similar demographic histories with
949 minimal influences of selection. Therefore, we strongly advise those using overlapping
950 outlier scans to carefully consider the demographic context of their system to avoid false-
951 positives. In particular, the presence or absence of migration between diverging populations
952 is a key factor determining the informativeness of genetic variation for selection, and
953 importantly shapes our expectations of outlier overlap among replicate population pairs. It
954 is tempting to assume that replication in study design or analysis in the form of taking
955 multiple measures of genetic divergence can reduce the risk of attaining false-positives. We
956 hope to emphasise in this study that this is not always the case, as false-positives (i.e.
957 genome scan outliers that are not associated with regions under divergent selection) can be
958 driven by non-random genomic or common demographic features that cannot be bypassed
959 through replication. Moreover, many of the patterns we observe are variable through time,
960 such that the relevant pitfalls of analyses will depend on the age of the populations being
961 considered. It should thus also be important to estimate population splits, as a young
962 replicate pair and older replicate pair with similar demographic histories should be expected
963 to exhibit potentially different patterns of genetic variation.
964
965 Recent simulation work by Quilodrán et al. (2019), has also emphasised the influence of
966 genomic features and demography, and includes simulation software for estimating the
967 distribution of genetic variation over user-defined chromosomes. Such an approach is
968 particularly useful for systems with chromosome-level genome assemblies in order to gain a
969 sense for how features such as recombination, gene density, and selection targets may

970 produce false-positives under certain demographies. The software employed here, SLiM
971 (Haller and Messer 2019), may also be used to this end, and the scripts accompanying this
972 study will facilitate similar analyses over system-specific genome regions. Further, recent
973 work on genomic landscapes of linked selection (Stankowski et al. 2019) has highlighted that
974 much of the total variance of genetic divergence such as F_{ST}, D_{XY}, and π can be explained by
975 the major principal component (PC1) over numerous pairwise comparisons. These
976 population comparisons need not reflect divergent phenotypes, as PC1 reflects genomic
977 features associated with diversity landscapes. Adopting this approach may be useful for
978 systems lacking a chromosome-level genome assembly by estimating SNPs or regions with
979 non-random elevated measures of divergence associated with genome features. Such SNPs
980 or regions may be particularly prone to false-positives under certain demographic histories.
981

982 **ACKNOWLEDGEMENTS**

983 The authors wish to thank all members of the Fraser lab for insightful discussions on this
984 project, as well as members of the University of Exeter's EB theme and the University of
985 Sussex's EBE group. The authors would also like to thank Kai Zeng and Adam Eyre-Walker for
986 feedback on earlier copies of this manuscript, and two anonymous reviewers for inciteful
987 and helpful comments.
988

989 **AUTHOR CONTRIBUTIONS**

990 JRW and BAF conceived the project, developed the experimental design and contributed
991 towards the writing of the manuscript. JRW wrote and implemented the pipeline for
992 simulations and analysis.
993

DATA ACCESSIBILITY

A full pipeline 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 (S1-48) have been uploaded through the GSA figshare portal. Supplementary figures include results for different mutation rates, $Pheno_{Null}$ and neutral data across sampling generations along with additional figures.

FUNDING INFORMATION

This work was funded as part of a European Research Council grant (758382 GUPPYCon) awarded to BAF.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest

REFERENCES

Ahrens, Collin W, Paul D Rymer, Adam Stow, Jason Bragg, Shannon Dillon, Kate D L Umbers, and Rachael Y Dudaniec. 2018. "The Search for Loci under Selection: Trends, Biases and Progress." *Molecular Ecology* 27 (6): 1342–56.

Bank, Claudia, Gregory B. Ewing, Anna Ferrer-Admettla, Matthieu Foll, and Jeffrey D. Jensen. 2014. "Thinking Too Positive? Revisiting Current Methods of Population Genetic Selection Inference." *Trends in Genetics* 30 (12): 540–46. https://doi.org/10.1016/j.tig.2014.09.010.

Belleghem, Steven M Van, Carl Vangestel, Katrien De Wolf, Zoe De Corte, Markus Most, Pasi Rastas, Luc De Meester, and Frederik Hendrickx. 2018. "Evolution at Two Time Frames." *PLoS Genetics*.

Burri, Reto. 2017. "Linked Selection, Demography and the Evolution of Correlated Genomic Landscapes in Birds and Beyond." *Molecular Ecology* 26 (15): 3853–56. https://doi.org/10.1111/mec.14167.

Charlesworth, Brian. 1996. "Background Selection and Patterns of Genetic Diversity in *Drosophila Melanogaster*." *Genetics Research* 68 (2): 131–49.

———. 1998. “Measures of Divergence between Populations and the Effect of Forces That Reduce Variability.” *Molecular Biology and Evolution* 15 (5): 538–43.

———. 2009. “Effective Population Size and Patterns of Molecular Evolution and Variation.” *Nature Reviews Genetics* 10 (3): 195.

Charlesworth, Brian, M. T. Morgan, and Deborah Charlesworth. 1993. “The Effect of Deleterious Mutations on Neutral Molecular Variation.” *Genetics* 134 (4): 1289–1303. https://doi.org/10.1111/j.0014-3820.2002.tb00188.x.

Charlesworth, Brian, Magnus Nordborg, and Deborah Charlesworth. 1997. “The Effects of Local Selection, Balanced Polymorphism and Background Selection on Equilibrium Patterns of Genetic Diversity in Subdivided Populations.” *Genetics Research* 70 (2): 155–74.

Charlesworth, Deborah, Brian Charlesworth, and M. T. Morgan. 1995. “The Pattern of Neutral Molecular Variation under the Background Selection Model.” *Genetics* 141 (4): 1619–32.

Collins, Sinéad, and Juliette De Meaux. 2009. “Adaptation to Different Rates of Environmental Change in Chlamydomonas.” *Evolution: International Journal of Organic Evolution* 63 (11): 2952–65.

Cooper, Elizabeth A., and J. Albert C. Uy. 2017. “Genomic Evidence for Convergent Evolution of a Key Trait Underlying Divergence in Island Birds.” *Molecular Ecology* 26 (14): 3760–74. https://doi.org/10.1111/mec.14116.

Cruickshank, Tami E., and Matthew W. Hahn. 2014. “Reanalysis Suggests That Genomic Islands of Speciation Are Due to Reduced Diversity, Not Reduced Gene Flow.” *Molecular Ecology* 23 (13): 3133–57. https://doi.org/10.1111/mec.12796.

Cutter, Asher D., and Bret A. Payseur. 2013. “Genomic Signatures of Selection at Linked Sites: Unifying the Disparity among Species.” *Nature Reviews Genetics* 14 (4): 262–74. https://doi.org/10.1007/s10450-013-9579-3.

Dittmar, Emily L, Christopher G Oakley, Jeffrey K Conner, Billie A Gould, and Douglas W Schemske. 2016. “Factors Influencing the Effect Size Distribution of Adaptive Substitutions.” *Proceedings of the Royal Society B: Biological Sciences* 283 (1828): 20153065.

Doren, Benjamin M. Van, Leonardo Campagna, Barbara Helm, Juan Carlos Illera, Irby J. Lovette, and Miriam Liedvogel. 2017. “Correlated Patterns of Genetic Diversity and Differentiation across an Avian Family.” *Molecular Ecology* 26 (15): 3982–97. https://doi.org/10.1111/mec.14083.

Dutoit, Ludovic, Nagarjun Vijay, Carina F Mugal, Christen M Bossu, Reto Burri, Jochen Wolf, and Hans Ellegren. 2017. “Covariation in Levels of Nucleotide Diversity in Homologous Regions of the Avian Genome Long after Completion of Lineage Sorting.” *Proc. R. Soc. B* 284 (1849): 20162756.

Ellegren, Hans, and Nicolas Galtier. 2016. “Determinants of Genetic Diversity.” *Nature Reviews Genetics* 17 (7): 422–33. https://doi.org/10.1038/nrg.2016.58.

Ellegren, Hans, and J. B. W. Wolf. 2017. “Parallelism in Genomic Landscapes of Differentiation, Conserved Genomic Features and the Role of Linked Selection.” *Journal of Evolutionary Biology* 30 (8): 1516–18. https://doi.org/10.1111/jeb.13113.

Foote, Andrew D, Nagarjun Vijay, María C Ávila-Arcos, Robin W Baird, John W Durban, Matteo Fumagalli, Richard A Gibbs, M Bradley Hanson, Thorfinn S Korneliussen, and Michael D Martin. 2016. “Genome-Culture Coevolution Promotes Rapid Divergence of Killer Whale Ecotypes.” *Nature Communications* 7: 11693.

Frankham, Richard. 1995. "Effective Population Size/Adult Population Size Ratios in Wildlife: A Review." *Genetics Research* 66 (2): 95–107.

Fraser, Bonnie A, Axel Künstner, David N Reznick, Christine Dreyer, and Detlef Weigel. 2015. "Population Genomics of Natural and Experimental Populations of Guppies (*Poecilia Reticulata*)." *Molecular Ecology* 24 (2): 389–408.

Fraser, Bonnie A, and James R Whiting. 2019. "What Can Be Learned by Scanning the Genome for Molecular Convergence in Wild Populations?"

Gossmann, Toni I, Megan Woolfit, and Adam Eyre-Walker. 2011. "Quantifying the Variation in the Effective Population Size within a Genome." *Genetics* 189 (4): 1389–1402.

Haller, Benjamin C, and Philipp W Messer. 2019. "SLiM 3: Forward Genetic Simulations beyond the Wright–Fisher Model." *Molecular Biology and Evolution* 36 (3): 632–37.

Hämälä, Tuomas, and Outi Savolainen. 2019. "Genomic Patterns of Local Adaptation under Gene Flow in Arabidopsis Lyrata." *Molecular Biology and Evolution*.

Hartl, Daniel L, Andrew G Clark, and Andrew G Clark. 1997. *Principles of Population Genetics*. Vol. 116. Sinauer associates Sunderland.

Hill, William G, Michael E Goddard, and Peter M Visscher. 2008. "Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits." *PLoS Genetics* 4 (2): e1000008.

Hoban, Sean, Joanna L. Kelley, Katie E. Lotterhos, Michael F. Antolin, Gideon Bradburd, David B. Lowry, Mary L. Poss, Laura K. Reed, Andrew Storfer, and Michael C. Whitlock. 2016. "Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions." *The American Naturalist* 188 (4): 379–97. https://doi.org/10.1086/688018.

Hodgkinson, Alan, and Adam Eyre-Walker. 2011. "Variation in the Mutation Rate across Mammalian Genomes." *Nature Reviews Genetics* 12 (11): 756.

Hohenlohe, Paul A, Susan Bassham, Paul D Etter, Nicholas Stiffler, Eric A Johnson, and William A. Cresko. 2010. "Population Genomics of Parallel Adaptation in Threespine Stickleback Using Sequenced RAD Tags." *PLoS Genetics* 6 (2). https://doi.org/10.1371/journal.pgen.1000862.

Hudson, Richard R, M Slatkin, and W P Maddison. 1992. "Estimation of Levels of Gene Flow from DNA Sequence Data." *Genetics* 132 (2): 583–89.

Irwin, Darren E., Miguel Alcaide, Kira E. Delmore, Jessica H. Irwin, and Gregory L. Owens. 2016. "Recurrent Selection Explains Parallel Evolution of Genomic Regions of High Relative but Low Absolute Differentiation in a Ring Species." *Molecular Ecology* 25 (18): 4488–4507. https://doi.org/10.1111/mec.13792.

Jones, Felicity C., Yingguang Frank Chan, Jeremy Schmutz, Jane Grimwood, Shannon D. Brady, Audrey M. Southwick, Devin M. Absher, et al. 2012. "A Genome-Wide SNP Genotyping Array Reveals Patterns of Global and Repeated Species-Pair Divergence in Sticklebacks." *Current Biology* 22 (1): 83–90. https://doi.org/10.1016/j.cub.2011.11.045.

Jones, Felicity C., Manfred G. Grabherr, Yingguang Frank Chan, Pamela Russell, Evan Mauceli, Jeremy Johnson, Ross Swofford, et al. 2012. "The Genomic Basis of Adaptive Evolution in Threespine Sticklebacks." *Nature* 484 (7392): 55–61. https://doi.org/10.1038/nature10944.

Kess, Tony, Juan Galindo, and Elizabeth G. Boulding. 2018. "Genomic Divergence between Spanish *Littorina Saxatilis* Ecotypes Unravels Limited Admixture and Extensive Parallelism Associated with Population History." *International Journal of Business Innovation and Research* 17 (3): 8311–27. https://doi.org/10.1002/ece3.4304.

Kirkpatrick, Mark, and Nick Barton. 2006. “Chromosome Inversions, Local Adaptation and Speciation.” *Genetics* 173 (1): 419–34. https://doi.org/10.1534/genetics.105.047985.

Künstner, Axel, Margarete Hoffmann, Bonnie A. Fraser, Verena A. Kottler, Eshita Sharma, Detlef Weigel, and Christine Dreyer. 2016. “The Genome of the Trinidadian Guppy, Poecilia Reticulata, and Variation in the Guanapo Population.” *PLoS ONE* 11 (12): 1–25. https://doi.org/10.1371/journal.pone.0169087.

Malinsky, Milan, Richard J Challis, Alexandra M Tyers, Stephan Schiffels, Yohey Terai, Benjamin P Ngatunga, Eric A Miska, Richard Durbin, Martin J Genner, and George F Turner. 2015. “Genomic Islands of Speciation Separate Cichlid Ecomorphs in an East African Crater Lake.” *Science* 350 (6267): 1493–98.

Matthey-Doret, Remi, and Michael C Whitlock. 2019. “Background Selection and FST: Consequences for Detecting Local Adaptation.” *Molecular Ecology*.

Meier, Joana Isabel, David Alexander Marques, Catherine Elise Wagner, Laurent Excoffier, and Ole Seehausen. 2018. “Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids.” *Molecular Biology and Evolution*, no. April 2018: 1–37. https://doi.org/10.1093/molbev/msy051.

Morales, Hernan E, Rui Faria, Kerstin Johannesson, Tomas Larsson, Marina Panova, Anja M Westram, and Roger Butlin. 2018. “Genomic Architecture of Parallel Ecological Divergence: Beyond a Single Environmental Contrast.” *Unpublished Data*, 447854. https://www.biorxiv.org/content/10.1101/447854v1.abstract.

Nei, Masatoshi. 1987. *Molecular Evolutionary Genetics*. Columbia university press.

Nishikawa, Hideki, Takuro Iijima, Rei Kajitani, Junichi Yamaguchi, Toshiya Ando, Yutaka Suzuki, Sumio Sugano, Asao Fujiyama, Shunichi Kosugi, and Hideki Hirakawa. 2015. “A Genetic Mechanism for Female-Limited Batesian Mimicry in *Papilio* Butterfly.” *Nature Genetics* 47 (4): 405.

Nosil, Patrik, Daniel J Funk, and Daniel Ortiz-Barrientos. 2009. “Divergent Selection and Heterogeneous Genomic Divergence.” *Molecular Ecology* 18 (3): 375–402.

Quilodran, Claudio Sebastian, Kristen Ruegg, Ashley T Sendell-Price, Eric Anderson, Tim Coulson, and Sonya Clegg. 2019. “The Multiple Population Genetic and Demographic Routes to Islands of Genomic Divergence.” *BioRxiv*, 673483.

R Core Team. 2016. “R: A Language and Environment for Statistical Computing.” *R Foundation for Statistical Computing Vienna Austria*. https://doi.org/ISBN 3-900051-07-0.

Ravinet, Mark, R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. A.F. Noor, B. Mehlig, and A. M. Westram. 2017. “Interpreting the Genomic Landscape of Speciation: A Road Map for Finding Barriers to Gene Flow.” *Journal of Evolutionary Biology* 30 (8): 1450–77. https://doi.org/10.1111/jeb.13047.

Ravinet, Mark, Anja Westram, Kerstin Johannesson, Roger Butlin, Carl André, and Marina Panova. 2016. “Shared and Nonshared Genomic Divergence in Parallel Ecotypes of *Littorina Saxatilis* at a Local Scale.” *Molecular Ecology* 25 (1): 287–305. https://doi.org/10.1111/mec.13332.

Reid, Noah M, Dina A Proestou, Bryan W Clark, Wesley C Warren, John K Colbourne, Joseph R Shaw, Sibel I Karchner, Mark E Hahn, Diane Nacci, and Marjorie F Oleksiak. 2016. “The Genomic Landscape of Rapid Repeated Evolutionary Adaptation to Toxic Pollution in Wild Fish.” *Science* 354 (6317): 1305–8.

Reznick, David, and John A Endler. 1982. “The Impact of Predation on Life History Evolution in Trinidadian Guppies (*Poecilia Reticulata*).” *Evolution* 36 (1): 160–77.

Roda, Federico, Luke Ambrose, Gregory M. Walter, Huanle L. Liu, Andrea Schaul, Andrew Lowe, Pieter B. Pelser, Peter Prentis, Loren H. Rieseberg, and Daniel Ortiz-Barrientos. 2013. "Genomic Evidence for the Parallel Evolution of Coastal Forms in the *Senecio Lautus* Complex." *Molecular Ecology* 22 (11): 2941–52. https://doi.org/10.1111/mec.12311.

Romiguier, J, Philippe Gayral, Marion Ballenghien, Arnaud Bernard, Vincent Cahais, A Chenuil, Ylenia Chiari, R Dernat, L Duret, and Nicolas Faivre. 2014. "Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity." *Nature* 515 (7526): 261.

Rougemont, Quentin, Pierre-Alexandre Gagnaire, Charles Perrier, Clémence Genthon, Anne-Laure Besnard, Sophie Launey, and Guillaume Evanno. 2017. "Inferring the Demographic History Underlying Parallel Genomic Divergence among Pairs of Parasitic and Nonparasitic Lamprey Ecotypes." *Molecular Ecology* 26 (1): 142–62.

Sella, Guy, Dmitri A Petrov, Molly Przeworski, and Peter Andolfatto. 2009. "Pervasive Natural Selection in the *Drosophila* Genome?" *PLoS Genetics* 5 (6): e1000495.

Smith, John Maynard, and John Haigh. 1974. "The Hitch-Hiking Effect of a Favourable Gene." *Genetics Research* 23 (1): 23–35.

Soria-Carrasco, Víctor, Zachariah Gompert, Aaron A Comeault, Timothy E Farkas, Thomas L Parchman, J Spencer Johnston, C Alex Buerkle, Jeffrey L Feder, Jens Bast, and Tanja Schwander. 2014. "Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation." *Science* 344 (6185): 738–42.

Stankowski, Sean, Madeline A Chase, Allison M Fuiten, Peter L Ralph, and Matthew A Streisfeld. 2018. "The Tempo of Linked Selection: Rapid Emergence of a Heterogeneous Genomic Landscape during a Radiation of Monkeyflowers." *BioRxiv*, 342352. https://doi.org/10.1101/342352.

Stankowski, Sean, Madeline A Chase, Allison M Fuiten, Murillo F Rodrigues, Peter L Ralph, and Matthew A Streisfeld. 2019. "Widespread Selection and Gene Flow Shape the Genomic Landscape during a Radiation of Monkeyflowers." *PLoS Biology* 17 (7): e3000391.

Tine, Mbaye, Heiner Kuhl, Pierre-Alexandre Gagnaire, Bruno Louro, Erick Desmarais, Rute S T Martins, Jochen Hecht, Florian Knaust, Khalid Belkhir, and Sven Klages. 2014. "European Sea Bass Genome and Its Variation Provide Insights into Adaptation to Euryhalinity and Speciation." *Nature Communications* 5: 5770.

Torres, Raul, Zachary A. Szpiech, and Ryan D. Hernandez. 2018. "Human Demographic History Has Amplified the Effects of Background Selection across the Genome." *PLoS Genetics* 14 (6): 1–27. https://doi.org/10.1371/journal.pgen.1007387.

Trucchi, Emiliano, Božo Frajman, Thomas H.A. Haverkamp, Peter Schönswetter, and Ovidiu Paun. 2017. "Genomic Analyses Suggest Parallel Ecological Divergence in *Heliosperma Pusillum* (Caryophyllaceae)." *New Phytologist* 216 (1): 267–78. https://doi.org/10.1111/nph.14722.

Turner, Thomas L, Matthew W Hahn, and Sergey V Nuzhdin. 2005. "Genomic Islands of Speciation in *Anopheles Gambiae*." *PLoS Biology* 3 (9): e285.

Vijay, Nagarjun, Matthias Weissensteiner, Reto Burri, Takeshi Kawakami, Hans Ellegren, and Jochen B.W. Wolf. 2017. "Genomewide Patterns of Variation in Genetic Diversity Are Shared among Populations, Species and Higher-Order Taxa." *Molecular Ecology* 26 (16): 4284–95. https://doi.org/10.1111/mec.14195.

Waterhouse, Matthew D, Liesl P Erb, Erik A Beever, and Michael A Russello. 2018. "Adaptive

1213 Population Divergence and Directional Gene Flow across Steep Elevational Gradients in
1214 a Climate-sensitive Mammal." *Molecular Ecology* 27 (11): 2512–28.
1215 Weir, Bruce S, and C. Clark Cockerham. 1984. "Estimating F-statistics for the Analysis of
1216 Population Structure." *Evolution* 38 (6): 1358–70.
1217 Westram, A. M., J. Galindo, M. Alm Rosenblad, J. W. Grahame, M. Panova, and R. K. Butlin.
1218 2014. "Do the Same Genes Underlie Parallel Phenotypic Divergence in Different
1219 *Littorina Saxatilis* Populations?" *Molecular Ecology* 23 (18): 4603–16.
1220 https://doi.org/10.1111/mec.12883.
1221 Yeaman, Sam, and Sarah P Otto. 2011. "Establishment and Maintenance of Adaptive
1222 Genetic Divergence under Migration, Selection, and Drift." *Evolution: International*
1223 *Journal of Organic Evolution* 65 (7): 2123–29.
1224 Zeng, K. 2013. "A Coalescent Model of Background Selection with Recombination,
1225 Demography and Variation in Selection Coefficients." *Heredity* 110 (4): 363.
1226