Contingent Convergence: The ability to detect convergent genomic evolution is dependent on demographic similarity

AUTHORS

James R. Whiting[1], Bonnie A. Fraser[1]

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

KEYWORDS

Genome scans, population genetics, F_{ST}, D_{XY}, outlier scans, convergence

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, these methods have the potential to be confounded by features of demographic history, which are considerably variable 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 time points. In a factorial design with 32 treatments, we manipulate the strength of bottlenecks (N of founding individuals), contractions (proportional size of diverging population) and migration rate between an ancestral phenotype population and a diverging phenotype population. Our results illustrate known constraints of individual measures associated with strong contractions and a lack of gene-flow, 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 genes marked as outliers in independent treatments) are attainable as a product of similar demographic treatment, and that convergent outliers across different measures (particularly F_{ST} and D_{XY}) are achievable with little influence of selection. The data presented act as an illustrative guide of how underappreciated yet quantifiable measures of demographic history influence commonly employed methods for detecting selection.

INTRODUCTION

The advent of modern sequencing technologies has facilitated big-data studies of adaptation and evolution across the genome. These studies often centre around quantifying patterns of variation in genome-wide SNPs, which can be used to highlight regions or genes having experienced selection

relative to the neutral backdrop of the rest of the genome. These analyses, which we refer to here as outlier scans, have become a common tool in population genetics and have seen great success across diverse, natural systems in identifying candidate genes associated with the evolution of a range of adaptive traits (reviewed recently by (1)). In addition, the method of overlapping outlier scans across independent lineages has been employed extensively in recent studies seeking to test whether the same regions are involved in independent adaptation events (genetic convergent evolution) across diverse taxa including birds (2), fish (3–8), insects (9,10), mammals (11), molluscs (12,13) and plants (14,15).

Following initial excitement in the ability of outlier scans to identify locally adapted genomic regions, the last 10 years have seen a healthy scepticism develop regarding how indicative genetically differentiated regions are of natural selection (16). Other genomic processes that illicit heterogeneity in diversity across the genome, such as background selection (BGS) (17,18), have been invoked to offer alternative explanations for patterns of genome-wide variation at first attributed to selection (19). Furthermore, the influence of neutral processes such as genetic drift are similarly predicted to preclude the informativeness of measures of genetic differentiation if populations are subject to random differentiation through drift (20).

A major concern arises therefore when applying outlier scans to studies of natural populations, given the role of often unknown demographic parameters in influencing both BGS and neutral evolution. For example, the efficacy of selection is dependent on the effective population size (N_e) (21–24), such that the effects of neutral processes and BGS should be more pronounced in smaller populations; as has been demonstrated in humans (25) and *Drosophila* (26,27). Additionally, evidence suggests the effects of BGS are strongest when the interaction between N_e and selection is intermediate (28). Meanwhile, population sub-division and migration rates influence global population variance, which has implications for elevating measures of genetic differentiation (29,30) as sub-populations experience the effects of reduced N_e and increased linkage as discussed above. In an attempt to mitigate the confounding effects of these processes, some studies advocate the corroboration of multiple measures of population differentiation and divergence (19,31).

A diverse array of measures of genetic differentiation and divergence have been employed in the literature, however here we focus on two of the most common measures of relative differentiation (F_{ST} and π) and a third measure of absolute divergence (D_{XY}) that represents an alternative strategy. Undeniably, the most commonly used measure is the fixation-index F_{ST} (32,33), which measures the

relative importance of within- and between-population variance. This measure is thus 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 dramatically different allele frequencies or high substitution rate (large between-population variance) and/or abnormally low nucleotide variation (low within-population variance) relative to the rest of the genome. Changes in nucleotide diversity (π) are another indicator of adaptation, as selection on a beneficial allele limits variation within a population resulting in selective sweeps (34). 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 broadly similar to F_{ST}, changes in π ($\Delta\pi$) do not discriminate between which copy of a polymorphism is fixed, such that a substitution between populations is equivalent to a shared non-polymorphic site. $\Delta\pi$ outliers therefore represent regions of the genome with abnormally reduced π in either population relative to the rest of the genome. As a measure of absolute genetic divergence, D_{XY} (35) does not consider the relative frequencies of polymorphisms within populations (19,31). D_{XY} can be approximated as the likelihood of drawing the same allele from two populations, 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 outliers highlight regions that are highly variable in the ancestral population (large ancestral π) or exhibit many substitutions and thus increased sequence divergence.

Because each measure of differentiation quantifies genetic variation between populations in a slightly different way, each varies in a unique way with demographic variation. Whilst we possess good predictions for how individual measures are influenced by demography, and subsequently neutral processes, we know little about how the relationships between measures are affected. Because different measures are affected differently by demographic history, through its relationships with neutral processes and BGS (19,36,37), we should expect that their combined predictive power for regions of the genome experiencing selection is similarly variable across different demographic histories. In addition, the combining of scans using different measures may provide unwarranted confidence in results if different measures are susceptible to demography in a similar fashion.

This complex interplay between selection measures and demography is further exasperated where studies compare multiple populations to identify convergent genomic evolution. Here measures of population divergence are used across independent pairs of evolutionary replicates with ‘outlier’ loci

compared across results (6,14,38). Whilst this strategy has had some success in identifying adaptive convergent loci, the differential effects of contrasting demographic histories in independent evolution events, and the subsequent effects on measures of population divergence, have rarely been considered in how they affect our ability to detect convergent genetic evolution.

We sought to address these concerns by using forward-simulations to investigate the effects of different demographic histories on the relationships between measures of divergence with selection. Whilst we focus on the commonly employed method of examining genome-wide SNP variation between ancestral and derived populations, we should note that other methods are employed in the literature (e.g. replicated clines, phylogenomics, single population measures e.g. Tajima's D). We also investigated how demography influenced the overlap of outliers when each demographic treatment represents an independent evolutionary replicate, as is the case in studies of convergent genetic evolution. We varied demography through manipulating the number of founding individuals (bottlenecks), the population size of the derived population (contractions) and the presence/absence of migration. We simulated the effects of selection on realistic genes designed from features taken from the guppy genome assembly, a prominent model system for studies of convergent evolution (6,39). Moreover, we measured genetic divergence at four set time points through F_{ST}, D_{XY} and $\Delta\pi$ between a founding population and the diverging population 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 1) Type-I error for individual measures (outliers under weak/no selection) and 2) Type-III error when employing multiple measures. Type-III error (accepting a correct result for an incorrect reason) here would occur if common outliers across measures are attainable irrespective of selection; thus, using evidence from multiple measures to validate a claim of adaptive significance is prone to error. 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* (40) was used to simulate population divergence under contrasting demographic treatments in a fully factorial design across a founding population of N = 1000 across with a mutation rate based on the guppy genome (41) 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, realistic values of θ ($4N_e\mu$). SLiM employs a classic Wright-Fisher model to simulate populations, in which a population of diploid hermaphrodites proceeds through generations such that an individual's contribution towards the next generation is proportional its relative fitness.

Demographic treatments included varying degrees of population contraction relative to the founding population size (0.01, 0.1, 0.5, 1.0), migration as a proportion of individuals exchanged between populations (0.0, 0.002) and founding bottlenecks as a number of individuals (2, 20, 100, 1000). Combining these treatment levels in a fully factorial experimental design generated a total of 32 different independent demographic histories that all genes experienced.

SLiM runs were performed over simulated genes of varying length (in order to ensure robustness across different genomic elements); with exon number (max = 10), exon lengths, and intron lengths drawn at random from the guppy genome gene annotation file. Genes were also given 1kb flanking regions and a divergence selection coefficient (S_σ) drawn at random.

Selection in our model was the fitness consequences incurred through distance from a phenotypic optimum in a one-dimensional landscape. Intensity of selection was manipulated by modifying the standard deviation of the normal distribution curve from which the density distribution was calculated through the "dnorm" function in SLiM. Standard deviations for fitness distributions (S_σ) were 0.1, 0.5, 1.0, 5.0, and 10.0, with 0.1 representing the steepest fitness peak and 10.0 the shallowest. All subsequent analyses' standard deviations were 1/x-transformed so that $S_{1/\sigma} = 10$ represented the strongest selection treatment (Figure 1). Phenotypes were calculated per individual 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). Effect sizes for mutations with phenotypic effects were drawn from a Gaussian distribution with mean = 0 and $\sigma = 1$. The remaining synonymous exon mutations, and those in introns and flanks, had no effect on fitness. Recombination rate was set at a constant of $1e^{-8}$ across all genomic elements.

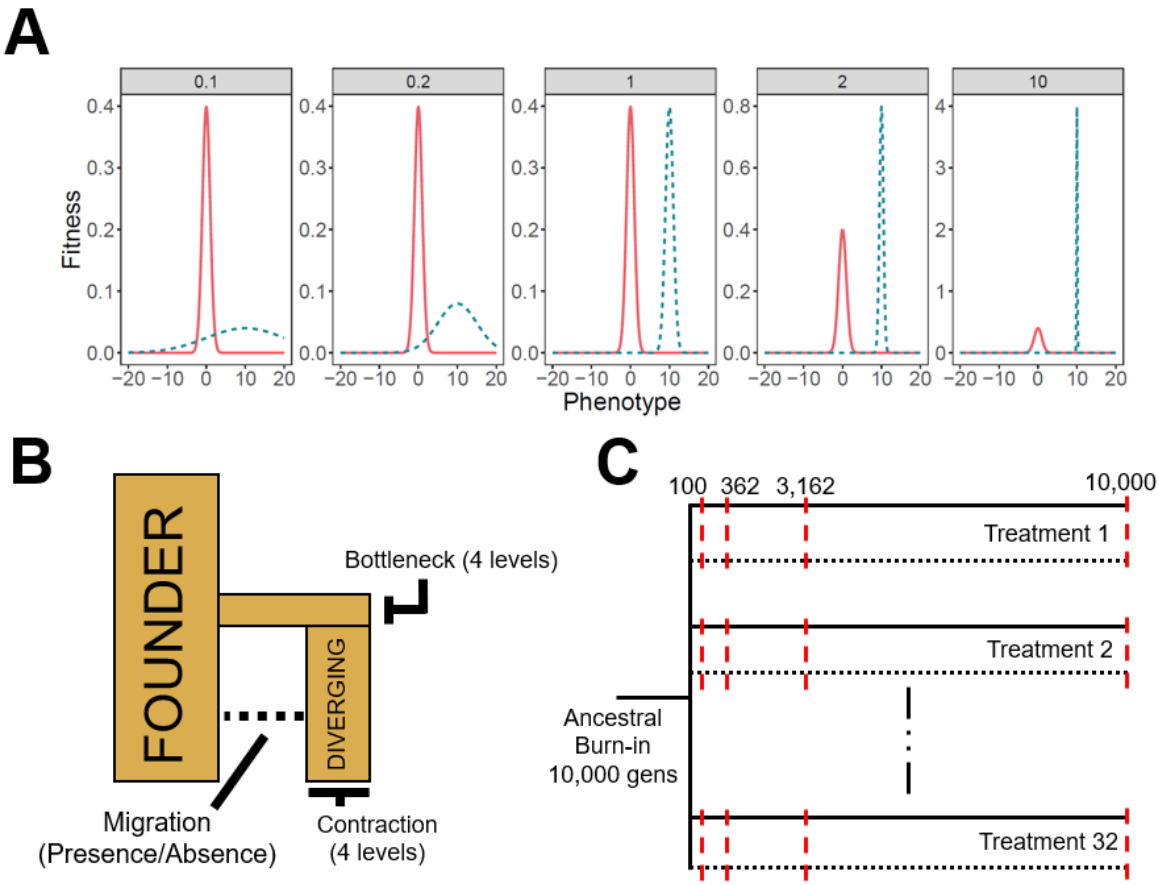


Figure 1: Experimental design for simulations. **A)** Selection treatments assigned to genes, illustrated as the relationship between relative fitness of an individual and its phenotype. Each facet represents a different fitness landscape modified through editing the standard deviation of the normal distribution of fitness consequences in the diverging (blue, dashed) population. Facet labels constitute 1/standard deviation applied. The x-axes represent phenotype as calculated through non-synonymous mutations and the y-axes represent relative fitness of individuals. The phenotype of the ancestral population (red, solid) is the same in all treatments (mean = 0, sd = 1), whilst diverging populations have a diverged phenotypic optimum (mean = 10, sd = S_σ). **B)** Demographic representation of treatment factors. **C)** Representation of simulation timeline for treatments, illustrating that all treatments share an ancestral burn-in population before splitting into 32 replicated "ancestral" (solid) and "derived" (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 50 generations. The purpose of this averaging was to achieve a general sense of population differentiation at four discrete sampling points, such that values represent stable patterns rather than more 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*2*N (10,000) generations to develop stable levels of natural variation under stabilising selection, adapting the methodology of (36). During this period, burn-in populations evolved towards a phenotypic optimum of 0, defined as a normal distribution with mean = 0 and σ = 1.0 (Figure 1). 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', the founding population continued to evolve around a phenotypic optimum of 0, whilst the diverging population's phenotypic optimum was centred around 10 with fitness consequences defined according to the selection treatment. 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 drawn from a Gaussian distribution with mean = 1, σ = 0.4 and occurred reciprocally between local individuals with phenotypes with a difference of ≥ 1.2 (3 * 0.4).

Each gene was then run across all demographic treatment levels in parallel. Simulations were sampled at 0.05*2*N (100), 0.158*2*N (316), 1.581*2*N (3162) and 5*2*N (10,000) generations (Figure 1C). At each sampling point, F_{ST} (calculated according to (33)), D_{XY} (calculated according to (35)), and mean heterozygosity (π) of each population were calculated as the average across the preceding 50 generations. Averaging was performed to gauge a general sense of differentiation around the four sampling points, such that measures would not be dramatically biased by events occurring within individual generations. The change in mean heterozygosity between the founder and divergent population (Δπ) was calculated at each sampling point as the $\log_{10}$-transformed ratio of founder π to divergent π, such that reduced diversity in the diverging population increases the value of Δπ. Statistics were calculated over all sites, akin to assessing differentiation over a gene or chromosomal interval.

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

$$D_{XY} = \sum_{ij} x_i y_j d_{ij}$$

$$\Delta\pi = \log_{10} \frac{\pi_{ancestral}}{\pi_{diverged}}$$

220 In total, ~1,000 genes were simulated across 32 demographic treatments across three mutation
221 rates. In addition, a fourth set of simulations were performed in which the phenotypic optimum for
222 both populations was kept at 0 ($Pheno_{Null}$). This dataset was used to assess whether patterns
223 associated with selection were driven by phenotypic divergence or variable stabilising selection
224 within populations. Sets of genes were generated randomly for each mutation rate. Outliers for gene
225 sets were taken as upper quantiles (either the upper 1% or 5% depending on analysis, see below).
226
227

228 **RESULTS**

229 ***Demography modifies measures of divergence***

230 Bottlenecks, simulated by modifying the number of founding individuals, had little effect on F_{ST} and
231 D_{XY}, producing minimal variance between treatments over all sampling points (Figure 2A).
232 Conversely, there was a relatively pronounced effect on Δπ that persisted for several hundred
233 generations. Genes that were simulated with the most extreme bottlenecks displayed inflated
234 measures of divergence because of a sharp reduction in π in the diverging population.
235

236 Contractions, simulated by modifying the stable number of individuals within the diverging
237 population, had pronounced and variable effects on all measures. F_{ST} was negatively associated with
238 contractions, with values becoming inflated with any reduction in diverging population size. This
239 effect was generally consistent through time, however early sampling points exhibited the largest
240 variance between the most extreme contraction treatments, whereas later on the strongest
241 variance was observed between weaker treatments (Figure 2B). A similar effect was observed for
242 Δπ; however, this measure was particularly susceptible to inflation under the most extreme
243 contraction scenarios, with substantially more elevated values observed between contraction
244 proportions of 0.01 and 0.1 compared with 0.1 and either 0.5 or 1.0. Whilst this pattern was broadly
245 consistent through sampling points, variance was most pronounced at the initial sampling point (100
246 generations). Such observations are unsurprising given that both F_{ST} and Δπ increase with any
247 process that reduces within-population genetic variance. D_{XY} was largely unaffected by contraction
248 treatments up until the third sampling point (3162 generations). Following this, in contrast to F_{ST} and
249 Δπ, D_{XY} decreased when diverging population sizes were reduced (Figure 2B).
250

251 The inclusion of migration expectedly reduced absolute values of all measures of divergence. For F_{ST}
252 and D_{XY}, variance between migration treatments increased across the scope of the simulation,
253 however D_{XY} was generally more resilient to variance induced through migration. Δπ was also

reduced in the presence of migration, although the effect of migration on Δπ was generally consistent across all generations and did not increase through time, as was the case for F_{ST} and D_{XY}.

Unsurprisingly, there was a consistent trend regardless of demographic treatment for F_{ST} and D_{XY} to increase generally over time, however this was not the case for Δπ, which was generally constant through time whilst marginally declining. The majority of these patterns were apparent in the $Pheno_{Null}$ simulations (Figure S3), however genetic differentiation for all measures was reduced by the removal of divergent phenotypes when gene flow was present.

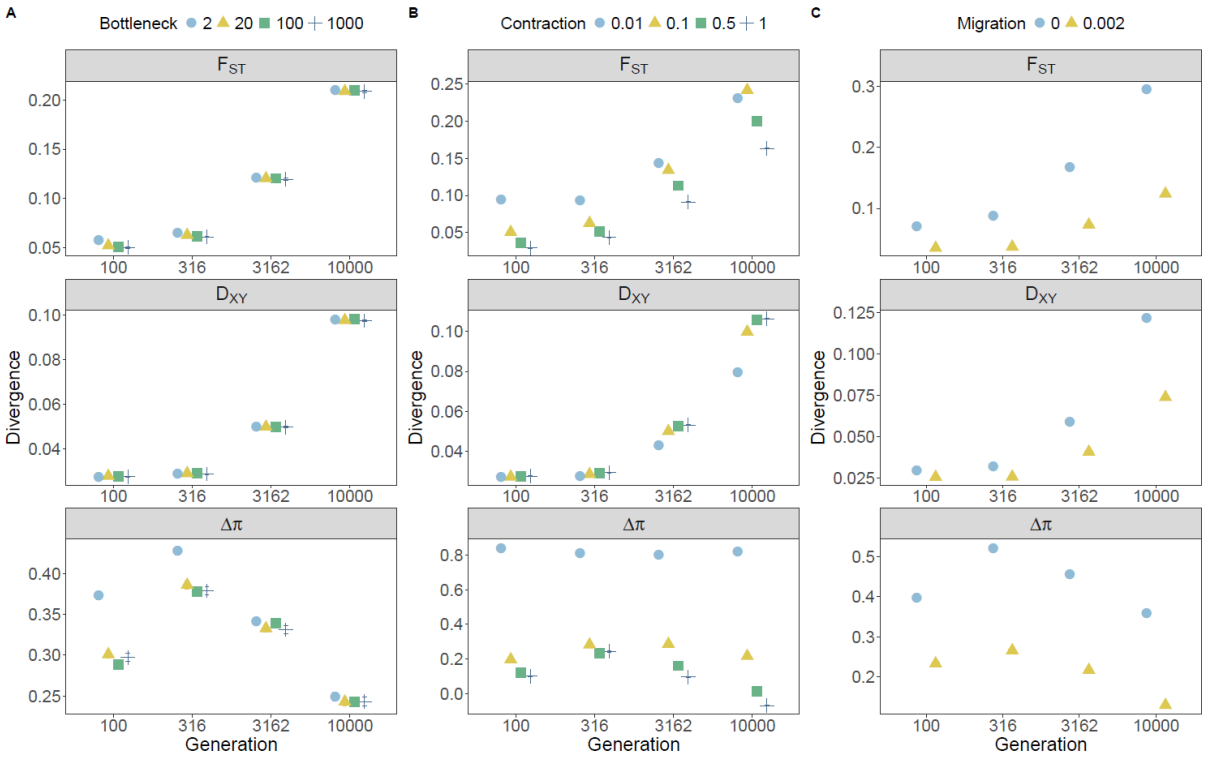


Figure 2: Effects of demographic treatments on measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for bottlenecks (**A**), population contraction (**B**) and migration (**C**). Each point represents values of divergence averaged across all genes, and all other treatment levels beyond those of the treatment group of interest.

Demography moderates the association between measures of divergence and selection

To assess the effect of demographic treatments on correlations with $S_{1/\sigma}$, we subsetted our data within each treatment level for each demographic factor (e.g. subsetting contraction groups of 0.01,

275 0.1, 0.5 and 1.0) and calculated Pearson's correlation coefficients across all treatments with the
276 treatment level of interest (such that for each population contraction level, 8 treatment groups were
277 included and averaged over). Bottlenecks had a minimal effect on the correlations observed
278 between strength of selection and measures of divergence with minimal variance observed between
279 bottleneck treatments in all comparisons (Figure 3A).
280
281 Population contractions had significant effects on relative (F_{ST} and $\Delta\pi$) measures and marginal
282 effects on absolute (D_{XY}) measures of divergence (Figure 3B). Relative measures in smaller diverging
283 populations (0.01, 0.1) were less informative, and indeed largely uninformative, of selection than
284 those observed in treatments with minimal (0.5) or no (1.0) contractions (Table 1). For F_{ST} the
285 variance in correlation coefficients between contraction treatment groups generally increased
286 across time whereas variance between treatments for $\Delta\pi$ peaked at 3,162 generations. For D_{XY},
287 correlations with selection were broadly resistant to contraction-induced variance up until 10,000
288 generations, in which associations with selection were reduced by stronger contractions (Figure 3B).
289
290 Expectedly, the absence of migration largely precluded the ability of relative measures of divergence
291 to predict strength of selection, with notable variance observed between no migration (M= 0.0) and
292 minimal migration (M = 0.002) treatments by 3,162 generations for all divergence measures (Figure
293 3C) (Table 1). Both F_{ST} and D_{XY} variance between migration treatments was greatest at the 10,000
294 generations sampling point, whereas similar variance was observed between migration treatments
295 for $\Delta\pi$ at both 3,162 and 10,000 generations. This observation again highlights the significance of
296 temporal effects on F_{ST} and $\Delta\pi$.
297
298 $Pheno_{Null}$ simulations resulted in weaker correlations with $S_{1/\sigma}$ for F_{ST}, whilst D_{XY} became negatively
299 correlated with $S_{1/\sigma}$ over time (Figure S6B) rather than positively correlated. $\Delta\pi$ became more
300 positively correlated with $S_{1/\sigma}$ over time for all demographic treatments (Figure S6C). These results
301 demonstrate that increases in F_{ST} are partly dependent on divergent selection, increases in D_{XY} are
302 exclusively dependent on divergent selection (and adaptive substitutions), and $\Delta\pi$ increases are less
303 indicative of divergent selection and more of stabilising selection and within-population variance.

Table 1: Correlation coefficients between measures of divergence and selection, separated according to individual treatment levels across all sampling generations. For each measure, the largest correlation coefficient is highlighted in bold.

Treatment	Level	F_{ST}				D_{XY}				$\Delta\pi$			
		100	316	3,162	10,000	100	316	3,162	10,000	100	316	3,162	10,000
Bottleneck	2	-0.006	0.040	0.129	0.169	0.016	0.019	0.033	0.095	0.006	0.083	0.148	0.072
	20	0.013	0.051	0.140	0.165	0.031	0.026	0.043	0.093	0.010	0.097	0.137	0.071
	100	0.010	0.043	0.147	0.177	0.024	0.022	0.050	0.102	0.009	0.084	0.138	0.074
	1000	0.007	0.040	0.133	0.165	0.020	0.021	0.041	0.097	0.019	0.079	0.149	0.070
Contraction	0.01	0.015	0.019	0.033	0.053	0.030	0.034	0.039	0.052	0.004	0.008	0.006	-0.024
	0.1	0.000	-0.001	0.087	0.157	0.015	0.006	0.020	0.110	0.034	0.040	0.189	0.122
	0.5	0.009	0.078	0.269	0.304	0.025	0.027	0.063	0.147	0.013	0.168	0.378	0.214
	1	0.000	0.136	0.306	0.355	0.022	0.020	0.051	0.138	0.031	0.260	**0.406**	0.245
Migration	0	0.003	0.029	0.074	0.062	0.020	0.014	0.006	-0.015	0.018	0.081	0.094	0.007
	0.002	0.012	0.101	0.335	**0.429**	0.027	0.033	0.092	**0.223**	0.003	0.105	0.241	0.191

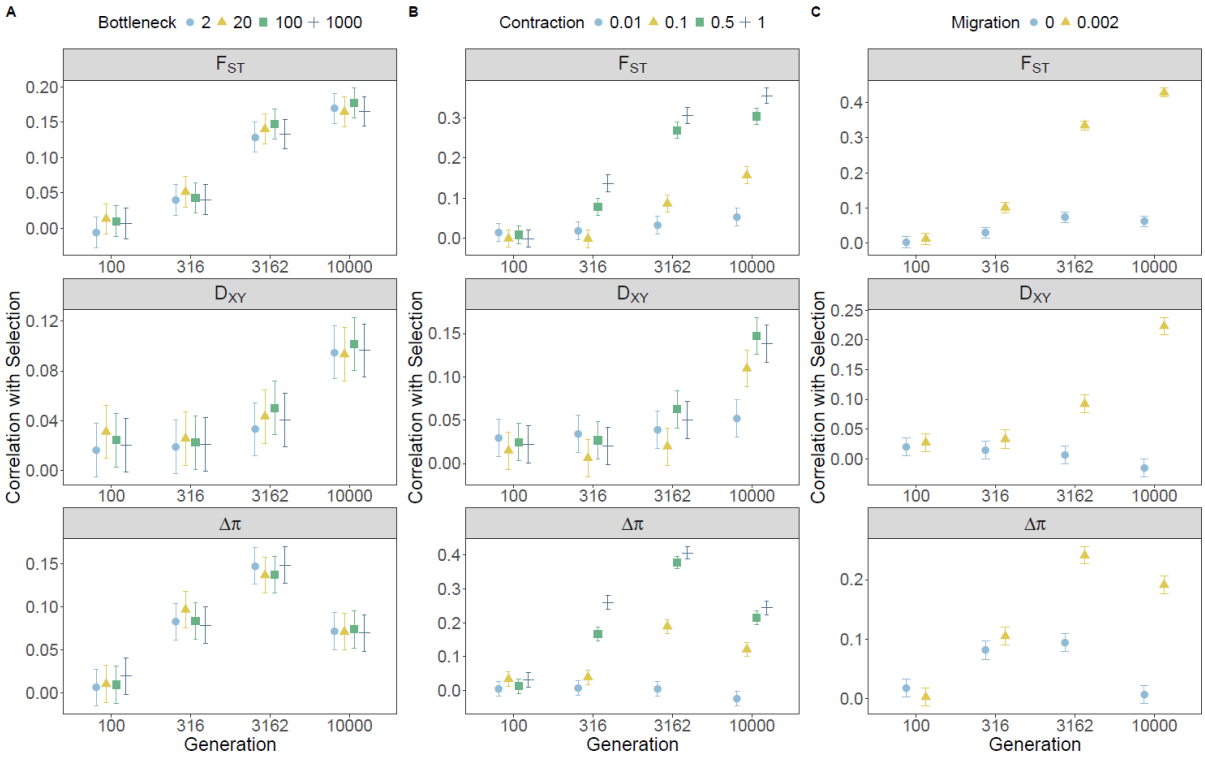


Figure 3: Effects of demographic treatments on the relationship between selection ($S_{1/\sigma}$) and measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for bottlenecks (**A**), population contraction (**B**) and migration (**C**). Each point represents correlation coefficients averaged across all genes, and all other treatment levels beyond those of the treatment group of interest.

By examining the correlation coefficients of all 32 unique demographic histories, we can investigate the combined effect of migration and contraction and directly compare effectiveness of individual measures across time (Figure 4). F_{ST} consistently outperforms D_{XY} in terms of associating with $S_{1/\sigma}$ under most demographic treatments, particularly when contractions are weak and migration is present. By 10,000 generations however, the relative dominance of FST appears to subside, with both F_{ST} and D_{XY} performing as well as each other in the presence of migration (Figure 4). Δπ is similarly more informative than D_{XY} across sampling generations under most demographic treatments. Interestingly, at sampling points 316 and 3,162 contractions produce the biggest bias towards Δπ, however by 10,000 generations zero migration produces the largest deviance away from the y=x relationship (Figure 4). The resilience of Δπ under no-migration treatments is also apparent in F_{ST} - Δπ comparisons, such that at 316 and 3,162 generations Δπ is more informative

than F_{ST} in the absence of migration. By 10,000 generations however F_{ST} outperforms Δπ under most scenarios, indicative of the decaying association between Δπ and $S_{1/\sigma}$ discussed previously.

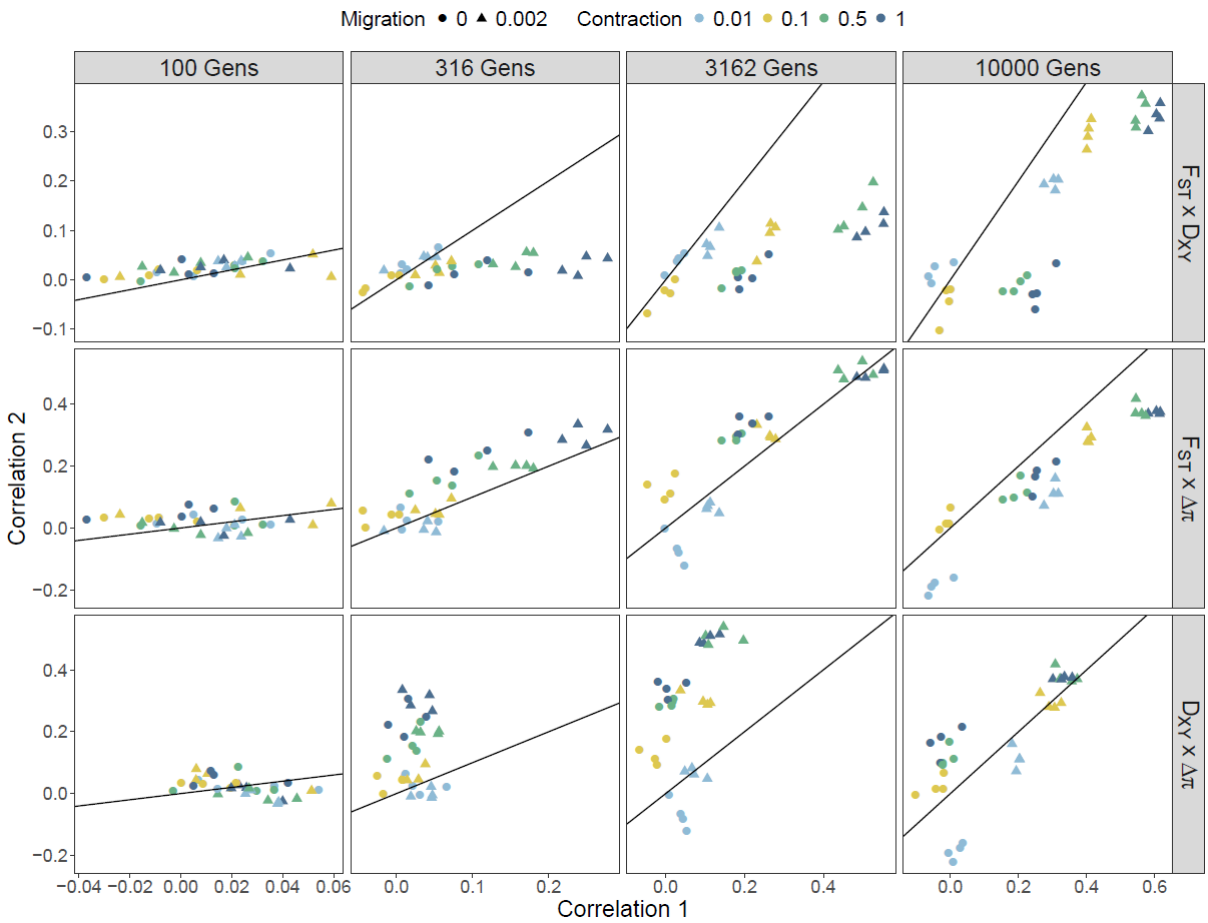


Figure 4: Pairwise comparisons between the correlation coefficients of $S_{1/\sigma}$ with F_{ST}, D_{XY} and Δπ across all sampling points. Each data point represents a unique demographic treatment with points coloured according to contraction level and shaped according to migration level. Correlation 1 refers to the first measure listed in the comparison and Correlation 2 to the second. The y=x line is plotted within each facet to illustrate biases towards one measure. Points below the line are biased towards Correlation 1, whilst points above the line are biased towards Correlation 2.

Demography moderates the tail end of divergence distributions

Ultimately, outlier tests are concerned with the extreme tail end of the distribution, and as such correlation coefficients across the whole dataset may fail to capture consistency in the tail ends. To counteract this, we asked at each sampling point what the relative proportions of outliers above the 99th quantile were per $S_{1/\sigma}$ treatment, and how these proportions varied according to individual

demographic treatments (Figure 5) and collectively (Figure 6) after 10,000 generations (results for other sampling points and mutation rates Figure S10-20). Each simulated 'genome' consisted of ~1000 genes evenly distributed across the 5 treatment levels of $S_{1/\sigma}$, meaning expected null proportions of each $S_{1/\sigma}$ value are 0.2. This analysis assumes that if measures of differentiation are good at highlighting genes under the strongest selection coefficients, genes within the upper 1% of the distribution should be disproportionately associated with stronger selection ($S_{1/\sigma}$ = 2 or 10). We found that migration and contraction treatments influenced the relative proportions of $S_{1/\sigma}$ treatments in the upper F_{ST} quantile. Under the weakest contraction treatment, the proportion of $S_{1/\sigma}$ = 10 outliers increased from 0.19 (contraction = 0.01) to 0.33 (contraction = 1.0) (null expectation of 0.20 given 5 selection treatments). Migration also dramatically increased the proportion of genes under the strongest selection from 0.19 to 0.63. The extreme effect of migration is evident when combining contraction and migration treatments (Figure 6). Whilst weaker contractions yield a greater proportion of strong selection outliers (moving left to right in the figure), migration is the dominant factor given that the treatment with the most extreme contraction (0.01) and migration present has a stronger selection bias than the treatment with no contraction and migration absent (Figure 6).

The upper quantile of D_{XY} outliers was unaffected by bottlenecks and contractions but was heavily biased towards $S_{1/\sigma}$ = 10 genes in the presence of migration (Figure 5). Proportions of $S_{1/\sigma}$ = 10 outliers doubled in the presence of migration from 0.27 to 0.54. Proportions across the 16 unique contraction/migration patterns produced results analogous to F_{ST}, with migration the dominant factor in biasing proportions towards strong selection genes (Figure 6).

Except for the most extreme population contraction, $\Delta\pi$ outlier proportions were strongly indicative of genes under strong selection, yielding outlier proportions for $S_{1/\sigma}$ = 10 genes of 0.72 (contraction = 0.1), 0.61 (contraction = 0.5) and 0.58 (contraction = 1.0) at sampling generation 10,000. Effects of contractions were visible for $\Delta\pi$ outliers by generation 100 at the most extreme treatments (Figure S10) and clear across all 3 weaker contraction treatments by 3162 generations (Figure S12). Migration had a more modest effect on outlier proportions, increasing the proportion of genes under the strongest selection from 0.16 to 0.41 at sampling generation 10,000. This modest effect of migration was not as persistent through time however, emerging only by generation 3,162 with $S_{1/\sigma}$ =10 proportions of 0.18 (migration = 0) and 0.24 (migration = 0.002). These results are corroborated across the 16 contraction/migration treatments, as proportions acquired from treatments with extreme contractions (contraction = 0.01) are relatively uniform with only a small biasing effect of

migration (Figure 6). For all other treatments, outliers are heavily biased towards strong selection genes.

Taken together, these results highlight that outlier detection using F_{ST} and D_{XY} are heavily dependent on the presence of migration, and to a lesser extent the absence or minimising of population contractions for F_{ST}. Δπ outliers appear rapidly and are generally robust in relation to population contractions up until the most extreme treatment. Whilst migration also influences Δπ, the biasing of outliers towards genes under the strongest selection treatment is less extreme.

Removing divergent selection ($Pheno_{Null}$) moderated outlier proportions in line with expectations given correlations with $S_{1/\sigma}$ (Figures S21-24). Interestingly, over time both F_{ST} and Δπ outliers were skewed further towards stronger selection treatments when migration was absent (stronger patterns of reduced within-population variance) in comparison to simulations with phenotypic divergence, whilst D_{XY} outliers became biased towards weaker selection treatments in the presence of gene flow.

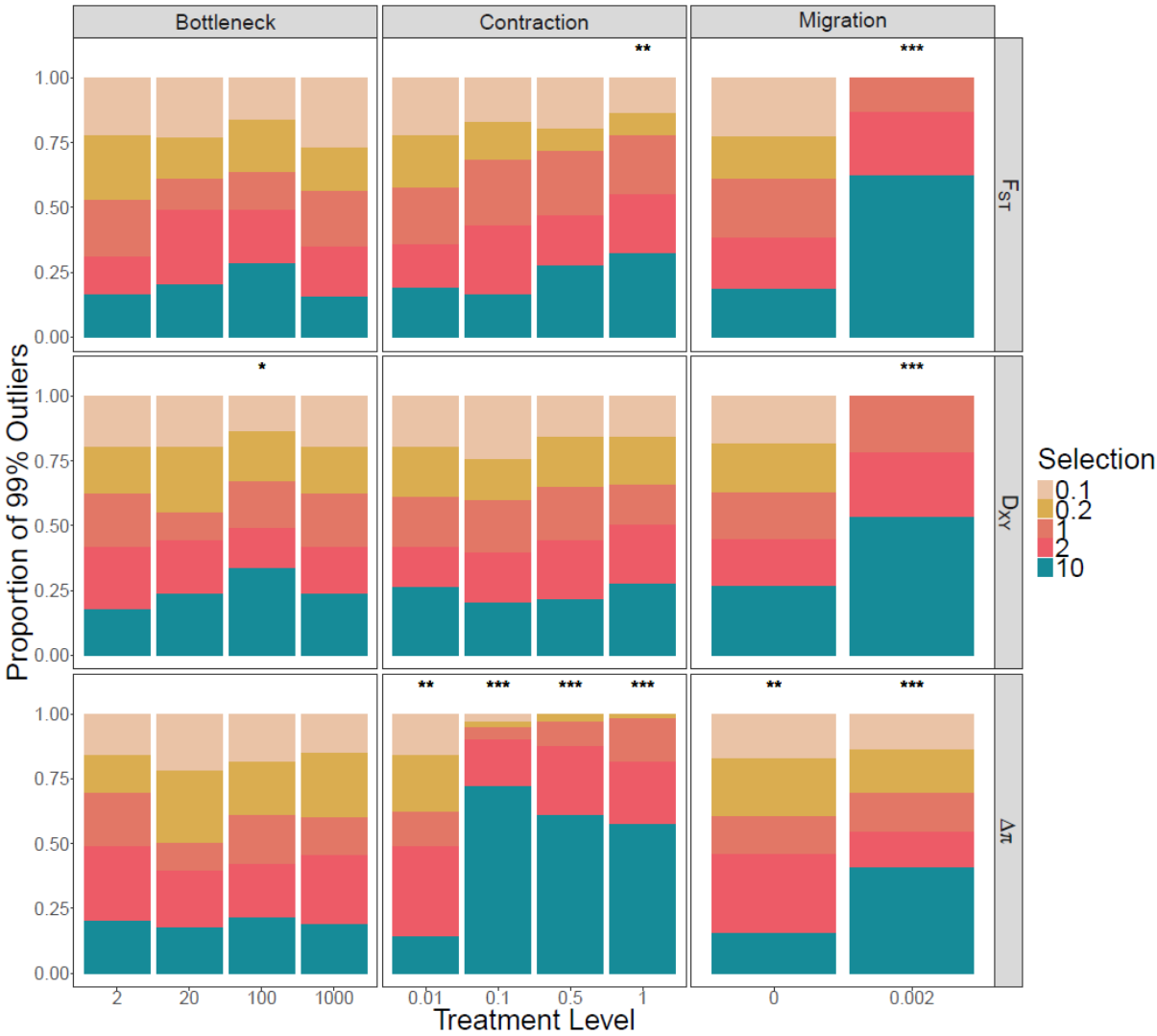


Figure 5: The effect of demography on representation of genes under different selection treatments as a proportion of outliers above the 99% quantile. Colour of bars within stacks denotes different selection treatments. Vertical facets represent different demographic treatments and horizontal facets show effect per measure of genetic divergence. Results shown are from sampling generation 10,000. Results for other sampling points can be found in Figures S10-12.

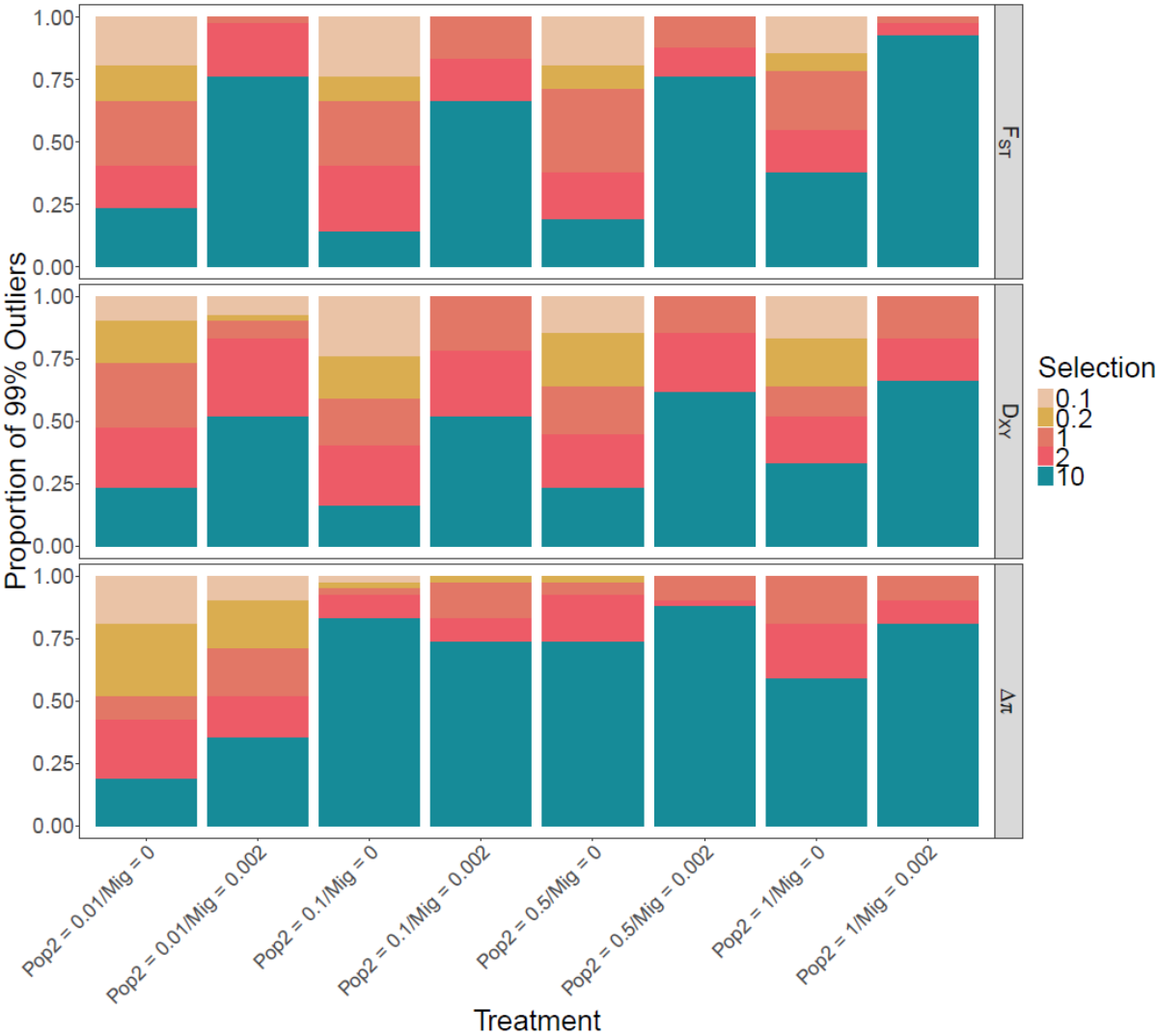


Figure 6: The combined effects of contractions (proportion relative to founding population size = 1,000) and migration in 16 unique demographic treatments (all combinations of contraction and migration treatments) on representation of genes under different selection treatments as a proportion of outliers above the 99% quantile. Full information regarding demographic treatment is described on the x-axis. Colour of bars within stacks denotes different selection treatments. Horizontal facets show effect per measure of genetic divergence. Results shown are from sampling generation 10,000.

Demography moderates associations between measures of divergence

Given F_{ST}, D_{XY} and Δπ are all measures of population genetic divergence, there is an assumption that positive associations should exist between them. We employed the same analysis as above for correlations with $S_{1/\sigma}$, but instead examined correlations between individual measures. Bottlenecks

had minor effects on the correlations observed between measures of divergence (Figure 7A). Expectedly, effects were strongest at the first sampling point, with the most extreme bottlenecks increasing the correlation coefficients between F_{ST} & D_{XY} and D_{XY} & Δπ and decreasing the correlation coefficient between F_{ST} & Δπ. Variation in correlation coefficients was generally minimal however compared to the effects of other demographic treatments.

F_{ST} & D_{XY} were positively associated irrespective of demography, but smaller population sizes generally increased the association (Figure 7B) with variance between treatments generally decreasing through time. Similarly, F_{ST} & Δπ were typically positively associated, however contractions had a more pronounced effect on the association. Extreme contractions increased the association between F_{ST} & Δπ, with contraction-induced variance increasing across sampling points (Figure 7B). At 10,000 generations when variance was largest, F_{ST} & Δπ ranged from being generally well associated under extreme contractions (R^2 = 0.60) to unassociated when populations were equally sized (R^2 = 0.08). D_{XY} & Δπ correlations were initially weak early in the simulations, however increased generally over time, with populations contractions similarly inflating associations between measures. Interestingly, associations with D_{XY} and both F_{ST} and Δπ appear to peak for weak contraction treatments at 3,162 generations, suggesting a probable temporal factor within the simulations driving positive associations with D_{XY} at some point between 316 and 3,162 generations.

Migration induced substantial variance between correlations of divergence measures, with effects dependent on sampling point (Figure 7C). In the absence of migration, F_{ST} & D_{XY} were strongly associated at 100 and 316 generations, whilst associations were weaker in migration-present treatments. However, by 3,162 generations F_{ST} & D_{XY} associations increased substantially in the presence of migration, and by 10,000 generations the effect of migration had reversed entirely. F_{ST} & Δπ associations were strong initially, but migration weakened the association over time so migration-present associations ranged from R^2 = 0.67 (100 gens) to R^2 = 0.14 (10,000 gens). In contrast, in the absence of migration positive associations between F_{ST} & Δπ were relatively stable (R^2 = 0.55 – 0.51 over the same time period). Interestingly, without migration D_{XY} & Δπ were largely unassociated, but migration-present treatments induced a negative association between D_{XY} & Δπ over time from R^2 = -0.04 (100 gens) to R^2 = -0.28 (1000 gens).

Contextualised by our previous demonstrations of associations with selection, these results highlight that selection can induce positive associations between measures. Namely, F_{ST} & D_{XY} become strongly associated by generation 10,000 with migration present, which we know is when F_{ST} and D_{XY} are most strongly linked to selection. Crucially, however this relationship only emerges after several thousand generations, before which we observe positive associations in migration-absent

treatments when associations with selection are weak for both measures. Extreme contractions also induce positive associations between F_{ST} & D_{XY} that are likely indicative of false-positives given the poor association with selection in these treatments. These results highlight that positive associations between statistics are also achievable under neutrality. It is also interesting to note the decay of associations with Δπ and both F_{ST} and D_{XY} under demographies in which selection should be effective. These observations are most likely driven by the substitution rate. Substitution rates are increased under stronger selection (42) and drive increased F_{ST} and D_{XY} but not Δπ.

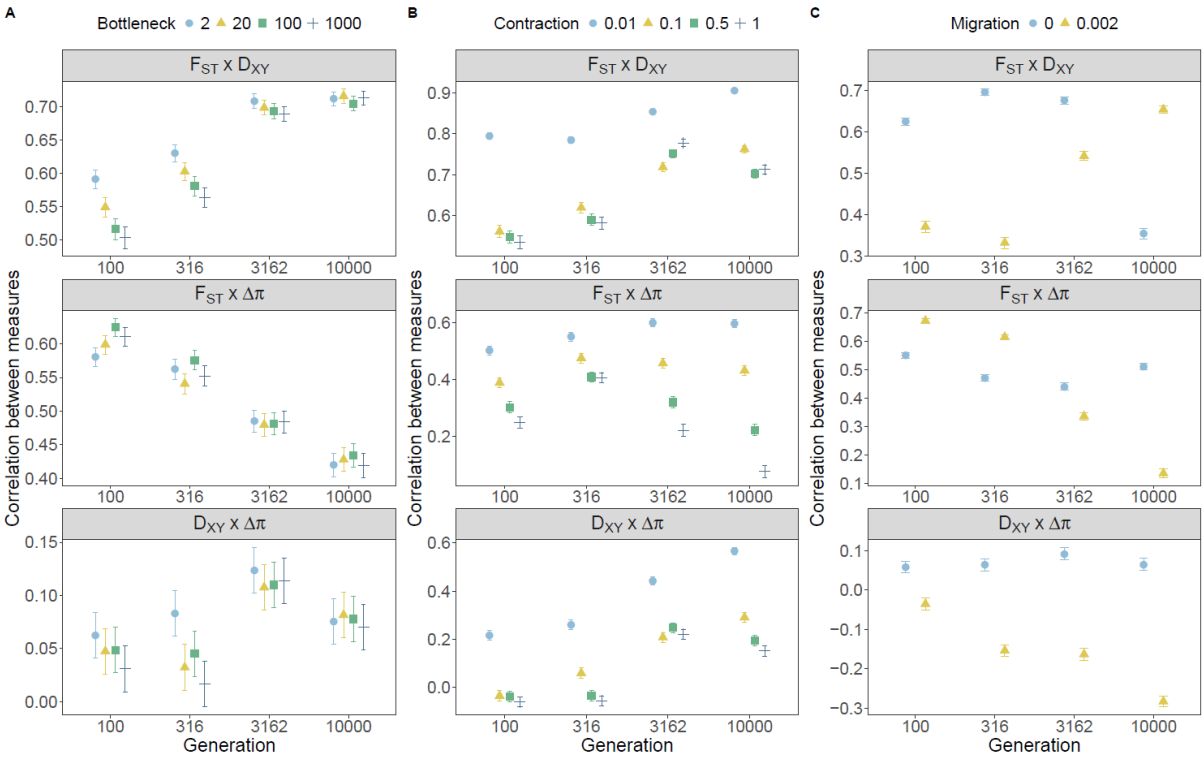


Figure 7: Effects of demographic treatments on the relationship between measures of genetic divergence across all sampling generations. Point colour and shape denote treatment groups for bottlenecks (**A**), population contraction (**B**) and migration (**C**). Each point represents correlation coefficients averaged across all genes, and all other treatment levels beyond those of the treatment group of interest.

In $Pheno_{Null}$ simulations the influence of bottleneck treatments and contraction treatments was broadly unchanged, however the notable emergence of a correlation between measures of F_{ST} and D_{XY} in simulations with gene flow failed to emerge by generation 10,000 (Figure S27). This demonstrates that this positive correlation is driven by divergently adaptive allele frequency shifts and substitutions. In addition, the decay of the relationship between F_{ST} and Δπ in the face of

migration did not occur without divergent selection. Rather, this relationship remained strongly positive ($R^2 > 0.5$) up to 10,000 generations regardless of migration treatment (Figure S27C).

To examine the overlap of outliers across measures of divergence, we again sub-divided our treatments into treatment-specific levels and compared the three-way and all pairwise upper 95% quantile gene overlaps across measures of divergence at sampling point 10,000. Three-way agreement was generally greater than expected by-chance under most conditions, with the largest overlap observed under extreme contractions (Figure 8). Our results demonstrate large overlap between F_{ST} and D_{XY} outliers across all treatments, which was marginally increased by contraction treatments that favour the effectiveness of selection. This association suggests stronger selection may improve the positive association between F_{ST} and D_{XY}, however a causative effect is not clear. Whilst overlap increased by selection-favouring conditions, all treatment levels exhibited above-expected levels of F_{ST} - D_{XY} overlap, suggesting a strong potential for false-positives. Overlap between $\Delta\pi$ and both F_{ST} and D_{XY} was substantially less but had contrary associations with contractions. More extreme contractions tended to increase the overlap between $\Delta\pi$ and F_{ST} outliers, suggesting a shared risk of false-positives. In contrast, $\Delta\pi$ and D_{XY} outlier overlap tended to increase under weaker contractions, suggesting a beneficial effect of combining these measures to remove false-positives. Migration increased outlier overlap between $\Delta\pi$ and both F_{ST} and D_{XY}.

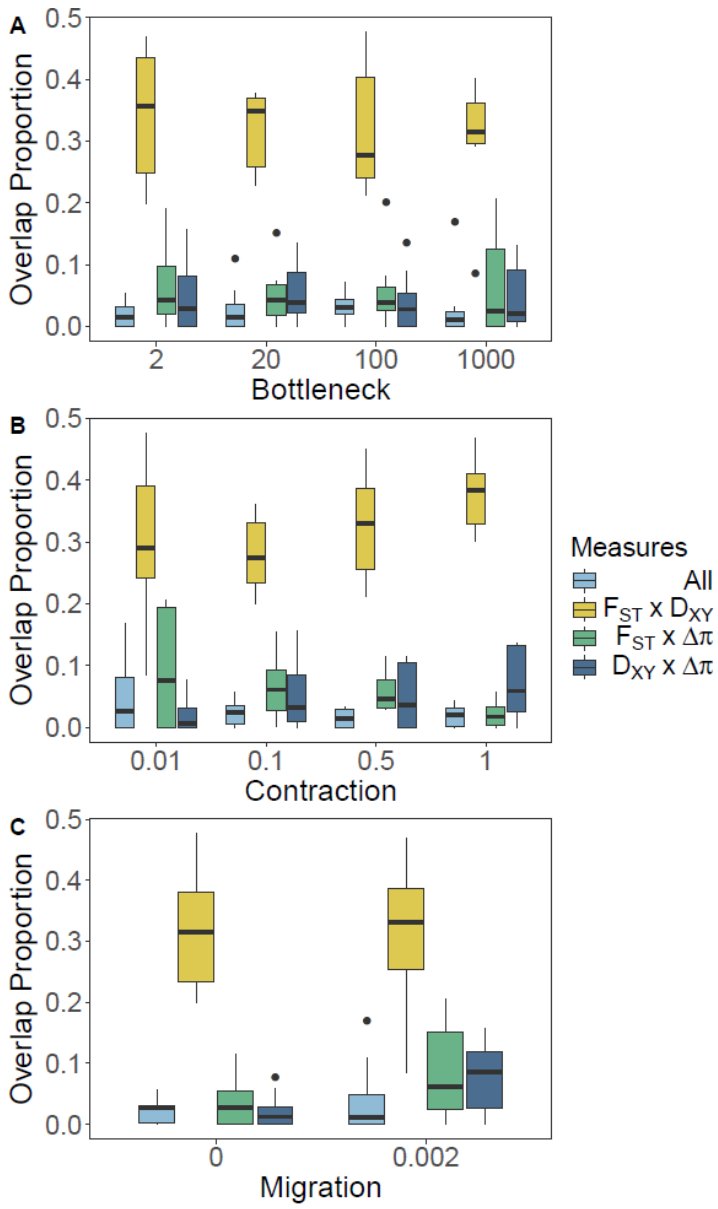
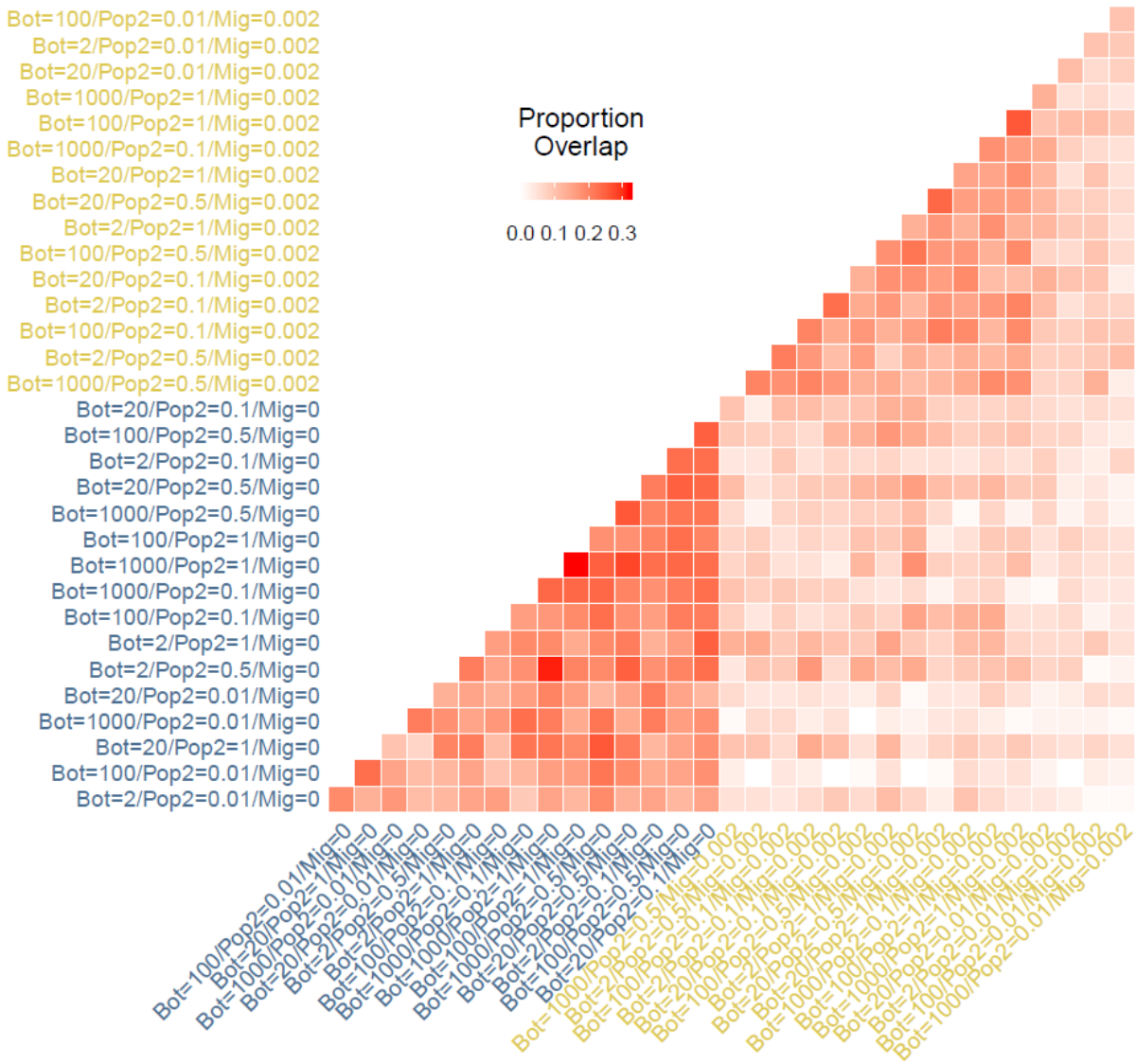


Figure 8: Proportional overlap of 95% outliers across measures of divergence, separated according to treatment levels of bottlenecks (**A**), contractions (**B**) and migration (**C**).

Demography drives signals of convergence irrespective of selection

To assess the convergence of outliers across demographic treatments, we downsampled the dataset such that we retained a random 10% of $S_{1/\sigma}$ = 10 and a random 50% of $S_{1/\sigma}$ = 2 genes per iteration. This step was taken to try to reduce the proportions of genes under the strongest selection treatment within the dataset from their initial ~20% proportions to a more realistic ~2% and ~10% respectively. Downsampling was bootstrapped 100 times, with the proportional overlap between demographic treatments recorded per iteration and averaged across all runs. Following

502 downsampling, we examined proportional overlap of the two 95% outlier sets with an expected
503 proportional overlap of 0.05 under a null hypothesis.
504
505 By sampling generation 10,000, proportion of overlap in outlier genes was apparent in F_{ST} results
506 across all demographic treatments that had no migration (Figure 9). Within this cluster, the overlap
507 of outliers above the 95% ranged between 0.07 and 0.32, suggesting a 1.4- to 6.4-fold enrichment
508 above the null expectation. These minimum and maximum values were observed between extremes
509 of contraction (minimum) or when contractions were absent (maximum). Given the poor association
510 between F_{ST} and selection in the absence of migration, convergent F_{ST} outliers here are likely driven
511 by neutral processes and would constitute a false-positive signal. Outside of this cluster, there was
512 also reasonable clustering of results across treatments that included migration and weaker
513 contractions (0.1, 0.5 or 1.0). Here, overlap of outliers ranged between 0.08 and 0.26, representing a
514 1.6- to 5.2-fold enrichment above the expected. In the presence of migration and minimal
515 contractions F_{ST} is a good predictor of selection in our dataset, suggesting this convergence may be
516 driven by inferred selection. Beyond these clusters, overlap between treatments of contrasting
517 migration treatment failed to exceed 0.18. Interestingly, the clustering of F_{ST} outliers in the presence
518 of migration was heavily reduced when divergent selection was removed ($Pheno_{Null}$) but clustering of
519 outliers in the absence of migration was still strong (Figure S39).
520



521

522

523 **Figure 9:** The proportional overlap of outliers above the 95% quantile for F_{ST} values, averaged across
524 100 randomly downsampled datasets for each of the 32 demographic treatments. Colouring of axis
525 labels denotes migration treatment level.

526

527 Clustering of D_{XY} outliers followed a broadly similar pattern to F_{ST} with migration treatments
528 dominating the clustering of treatments and inducing proportional overlaps of 0.11-0.34 (2.2- to 6.8-
529 fold enrichment) in the absence of migration (Figure S31). In general, proportional overlap of Δπ
530 outliers was lower than for F_{ST} and D_{XY}, with a maximum overlap observed of 0.21 (fold-enrichment
531 4.2) (Figure S32). This value was observed in two treatments with minimal contractions (0.5-1.0) and
532 migration present. Generally, however, clustering by demography was weaker in Δπ than for F_{ST} and
533 D_{XY}, corroborating previous results suggesting Δπ may be less susceptible to variation in migration
534 and contraction (with the exception of extreme contraction).

535

536 Because migration was the dominant factor in clustering of treatments with convergent overlaps, we
537 sought to investigate what features of simulated genes drove variance in measures of divergence
538 with treatments separated by migration factor using a Random Forest approach (Figure 10). For F_{ST},
539 we observe that the best predictor of variance in the absence of migration is diversity (π) of the
540 diverging population whilst treatments with migration are most strongly influenced by $S_{1/\sigma}$. This is
541 consistent with the notion of F_{ST} relying on within-population genetic variance in the absence of
542 efficient selection. In contrast, D_{XY} in the absence of migration is most strongly driven by gene
543 features, predominantly the proportion of simulated gene that is coding and by the length of the
544 simulated gene. With migration, D_{XY} variance is driven by selection and diverging π. Unsurprisingly,
545 Δπ is also strongly influenced by diverging π (the denominator in its calculation), but we can see a
546 secondary role for selection in treatments with migration and little influence of gene features.
547

548 Removing divergent selection ($Pheno_{Null}$) did not modify the relative importance of any features of
549 simulations for F_{ST}, but did affect D_{XY} and Δπ (Figure S44). With migration, the relative importance
550 for diverging population π in explaining D_{XY} variance increased substantially. For Δπ, selection was as
551 important as diverging population π in explaining variance.
552

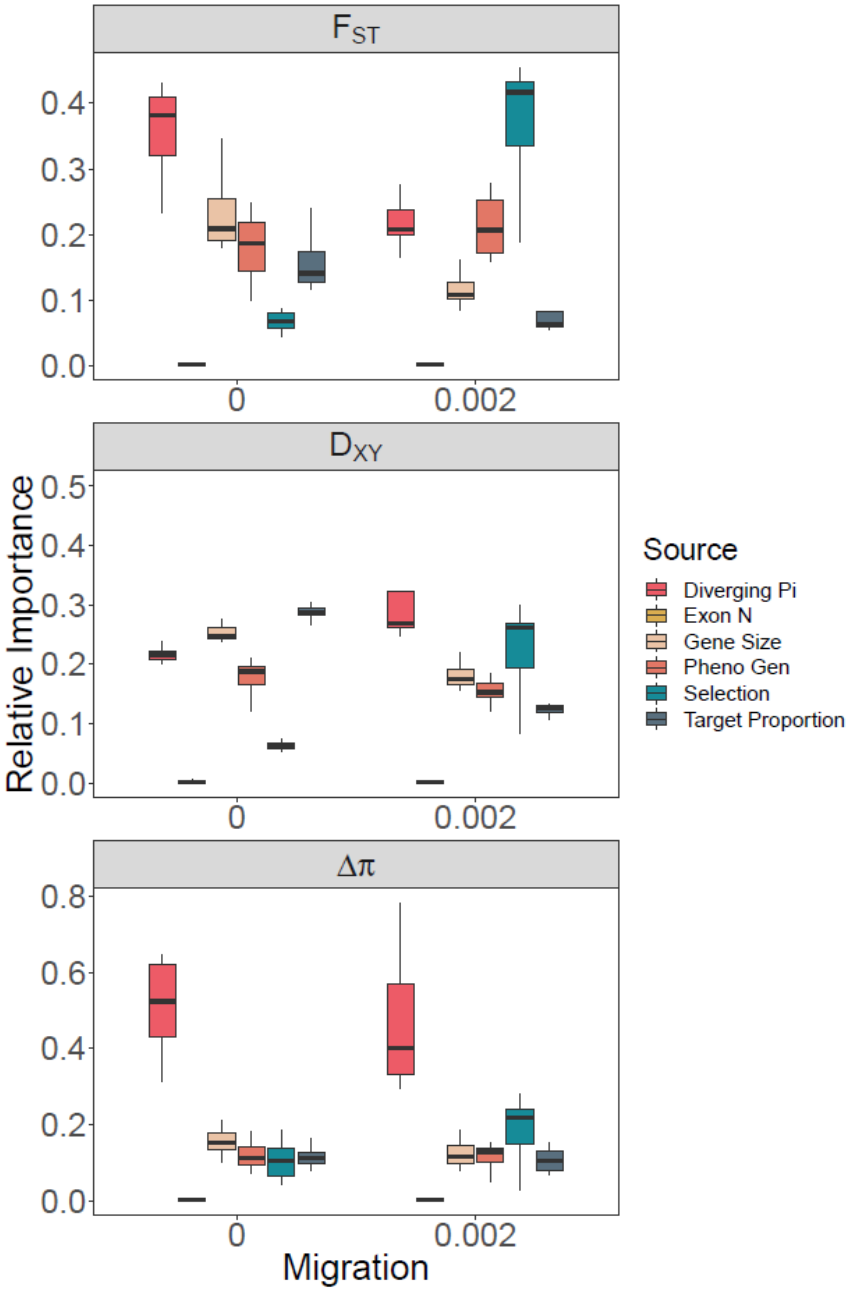


Figure 10: The relative importance of simulated gene features in eliciting variance in measures of genetic variance. Boxplots denote the relative importance of each explanatory variable across 16 individual treatments with/without migration taken at sampling generation 10,000. Variables considered are within-population π of the diverging population and gene features including: Number of Exons (Exon N), Gene Size, and the proportion of gene that is coding (Target Proportion). Also included are the selection ($S_{1/\sigma}$) imposed on the gene and the generation at which the optimum phenotype is reached (Pheno Gen).

DISCUSSION

Summary of results

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

We find that founding bottlenecks have little effect on population divergence measures, with only minor influence over differences in nucleotide diversity ($\Delta\pi$) shortly after founding. Population contractions artificially inflate F_{ST} and $\Delta\pi$ but reduce D_{XY} and can erase the relationship of F_{ST} and D_{XY} with selection under the most extreme contractions. These results can be explained in part due to the relationship between founding population size and relative (F_{ST} and $\Delta\pi$) and absolute (D_{xy}) divergence measures. Relative measures are, in part, driven by intra-population changes in allele frequency, which become exaggerated in smaller populations as a product of drift (23,43). As a consequence, we observe inflated measures of relative divergence as allele frequencies drift in our smaller diverging population.

D_{xy} in contrast increases under weaker population contractions as a product of the relationship between the number of segregating sites and the population-level mutation rate ($4N_e\mu$) (44). D_{xy} is a measure of sequence divergence and is averaged across all sites (although averaging across all segregating sites is possible), which results in higher D_{XY} as segregating sites are introduced into either population at rate of $4N_e\mu$. This relationship with the number of segregating sites can be observed by examining the positive relationships between D_{xy} and π in our diverging population (Figure S45). The absence of a strong contraction effect in moderating D_{xy} associations with $S_{1/\sigma}$ likely stems from the lack of allele frequency relevance. Consider for example, two SNPs with minor allele frequencies of 0/0.5 (SNP 1) and 0.5/0.5 (SNP 2) in the founding/diverging population. Each locus contributes equally towards D_{xy} (SNP 1 = [0 x 0.5] + [1 x 0.5] = 0.5; SNP 2 = [0.5 x 0.5] + [0.5 x 0.5] = 0.5), whereas the reduction of within-population variance observed for SNP 1 inflates F_{ST} and $\Delta\pi$.

Migration, even at the relatively modest rate of 0.2% employed here, reduced absolute values for all measures of divergence substantially. However, whilst absolute values were reduced, their informativeness of selection coefficients increased dramatically. Such a result is expected given the

role of migration and gene-flow in homogenising neutral loci (reducing measures of divergence), whilst retaining population divergence around adaptive loci (increasing informativeness). This dynamic is often invoked within the population 'genomic islands of divergence' model to explain patterns of heterogenous genomic divergence in studies of speciation-with-gene-flow (45,46).

Migration exhibited interesting temporal patterns that are useful for discussing the discrepancies observed between $\Delta\pi$ and both F_{ST} and D_{xy}. Of the measures considered here, $\Delta\pi$ uniquely disregards SNP substitutions in its calculation, as fixed substitutions have no influence on intra-population heterozygosity beyond reducing the heterozygosity of proximate SNPs in the aftermath of a hard sweep. This characteristic explains why $\Delta\pi$ responds to selection more rapidly than F_{ST} and D_{xy} and is influenced by migration in a constant manner across time. In addition, a failure to account for adaptive substitutions in $\Delta\pi$ is the most probable explanation for $\Delta\pi$ informativeness peaking at 3,162 generations and decaying by 10,000 generations. In contrast, the contributions of adaptive substitutions to F_{ST} and D_{xy} over time increases their predictive power. This temporal observation has important implications for studies of rapid adaptation on the scale of 10s to 100s of generations, in which our simulations suggest $\Delta\pi$ may be the most informative measure of divergence, whilst D_{xy} was largely uninformative until > 3,162 generations.

A novel element to our simulations involved examining the effects of demographic variation on the associations between measures of divergence themselves. The relative weaknesses of individual measures of divergences has prompted a recent movement within the literature to employ multiple measures of divergence to avoid false-positives (19). Our results provide some support for this recommendation, with outlier overlap across measures generally increasing under conditions that favour informativeness of selection, suggesting their combined use may be useful in removing false-positives. Importantly, we also find that F_{ST} and D_{xy} (and to a lesser extent F_{ST} and $\Delta\pi$) highlight convergent outliers across demographic treatments regardless of the effectiveness of selection.

Cruickshank and Hahn (19) suggested that a failure of F_{ST} to exhibit similar patterns to D_{xy} in genomic-islands-of-divergence tests highlighted a particular susceptibility of F_{ST} to BGS (although see (36)). Our results are in line with the notion that D_{xy} is more resistant to BGS given that contraction treatment levels resulted in minimal variance in D_{xy} (assuming treatment levels of population contraction are analogous to different rates of BGS across the genome in exhibiting reduced N_e). Our results add support to the emergence of a positive association between F_{ST} and D_{XY}, but only in the presence of gene flow between populations and only after several thousand generations, as has

632 been previously simulated (47). The latter dependence on time is due to the delayed predictive
633 power of D_{xy}.
634
635 However, our results also demonstrate that F_{ST} - D_{xy} associations emerge under demographic
636 conditions where neither is well correlated with selection, and that this relationship decays over
637 time in the absence of migration (Figure 7C). No-migration treatments are consistent with Isolation-
638 By-Distance demographic histories and similarly inter-specific comparisons. With regards to the
639 latter, empirical observations of negative correlations between F_{ST} and D_{XY} within and between
640 clades of birds (48,49) and within a radiation of monkeyflowers (50) support the simulated
641 prediction of an eventual negative relationship between these measures. This negative association
642 in the absence of effective selection is predictable, given F_{ST} is negatively associated with local π
643 (increases within-population variance), D_{XY} is positively associated with local π (reduces likelihood of
644 common sequence) and π is generally well-conserved in stable populations even across long time
645 periods (51–53).
646
647 Similarly, F_{ST} – $\Delta\pi$ outliers were found under most demographic scenarios and marginally increased
648 under extreme contractions. These results support the notion that whilst the use of multiple
649 measures may be useful in reducing Type-I error, there is an inherent risk associated with Type-III
650 error (accepting true-positives for the wrong reason) given the risk of Type-I error under certain
651 demographic histories.
652
653 In comparing our results to a second dataset that lacked divergent selection ($Pheno_{Null}$), we find
654 consistent support that patterns associated with D_{XY} are strongly dependent on divergent selection;
655 F_{ST} patterns are attainable with only stabilising selection (but are weaker than when divergent
656 selection is included); and $\Delta\pi$ patterns are primarily driven by variable stabilising selection and made
657 weaker by the inclusion of phenotypic divergence. These comparisons are useful in informing
658 specifically on the role of adaptive allele frequency changes and substitutions in driving patterns of
659 genetic differentiation. In particular, our observation that positive F_{ST} - D_{XY} correlations emerge only
660 with effective, divergent selection in the face of gene-flow will be informative in assessing the
661 usefulness of this combination of statistics in studies of adaptation.
662
663
664 ***Application of results***

665 We sought to utilise our experimental design to investigate the effects of variable demography in
666 studies of convergence. Our simulation design, in which an ancestral burn-in population is used to
667 found 32 independent pairs of ancestral – diverging populations is analogous to the commonly
668 employed sampling method of examining replicated ecotype population pairs in model systems,
669 such as various ecotype pairs of the three-spined stickleback, *Gasterosteus aculeatus* (3,4); high/low
670 predation Trinidadian guppies, *Poecilia reticulata* (6); crab/wave ecotype periwinkles, *Littorina*
671 *saxitilis* (12,13,54), and alpine/montane ecotypes of *Heliosperma pusillum* (14). By overlapping
672 outliers called from one demographic treatment with others, we demonstrate that signals of
673 convergent genes (proportions above a random null expectation) are attainable for F_{ST} and D_{XY}
674 across multiple replicates, with clustering predominantly driven by the presence or absence of
675 migration (with minimal convergence between clusters). Given that within our simulations selection
676 is expectedly ineffective when migration is absent, convergence of outliers across migration-free
677 treatments suggests a different subset of genes within our analysis, which are not under strong
678 selection, are repeatedly reporting the strongest signals of genetic divergence for both F_{ST} and D_{XY}.
679
680 In seeking to explain convergence patterns associated with migration, we found that for both F_{ST} and
681 D_{XY} there were different sources of variance depending on migration treatment. F_{ST} was heavily
682 influenced by within-population genetic diversity (π) in the absence of migration. This is consistent
683 with previous theoretical and empirical work highlighting the susceptibility of F_{ST} to processes that
684 reduce local population diversity disproportionately across the genome, such as heterogeneity in
685 recombination landscape (22,47), structural variation (55), and gene density (22,56) in the absence
686 of effective selection. With migration present, and effective selection, we find that convergence is
687 most probably driven by selection in our simulations. These results therefore demonstrate
688 convergence within demographically similar pairs, and by extension non-convergence across
689 demographically dissimilar pairs, is only selection-dependent under certain conditions. Without
690 migration, D_{XY} convergence was achieved through features of simulated genes, predominantly
691 length of gene and proportion of gene that was coding. Such an effect may be explained by the
692 expectation of reduced neutrality of mutations within coding regions, producing a consistent effect
693 across demographic treatments for genes with increased coding content to exhibit reduced neutral
694 polymorphism (across both the ancestral and diverging population) and lower D_{XY} as a result. The
695 secondary effect of gene length is less clear, given that D_{XY} is averaged across all sites, but may stem
696 from potentially lower levels of linked selection in larger genes (43) or the increased accumulation of
697 segregating sites in larger genes (given mutation occurs with a fixed probability per site). With
698 effective selection in the presence of migration, we observe the expected effect of simulated $S_{1/\sigma}$ in

699 producing convergent D_{XY} outliers. Interestingly however, we also observe a fairly strong effect of
700 diverging population π that hints towards a susceptibility to BGS even in the presence of effective
701 selection. This dependence on the efficacy of selection may explain the discrepancies within the
702 current literature regarding the susceptibility of both F_{ST} (19,29,57) and D_{XY} (36) to BGS.
703
704 It is perhaps surprising that population contractions (with the exception of the most extreme factor
705 level) did not have a dominant influence on our convergence matrix, given the strong effect on
706 population contractions in reducing the efficacy of selection. However, this finding is consistent with
707 the increased influence of random drift in populations with reduced N_e, thereby limiting
708 repeatability of outliers.
709
710 In addition, in our analyses we group genes within 32 unique demographic treatments, assuming the
711 effects of contraction and migration are equivalent for all genes. This may be unrepresentative of
712 genomes sampled from the wild, in which gene flow and effective population size can vary across
713 genomic regions through structural variation, variable recombination rate and BGS as previously
714 discussed. A prominent example of such a mechanism is the well-characterised consequence of
715 reduced gene-flow within inversions that carry locally adapted alleles. Assuming inversion variants
716 are fixed between populations, migration across the locus is limited by the prevention of
717 recombinant haplotypes and resistance to introgression (47,55). Recent genome scan studies have
718 highlighted convergent outliers within inversions (58–60), confirming theoretical models regarding
719 the role for chromosomal rearrangements in shielding adaptive haplotypes from introgression
720 during the adaptation process. Our simulations highlight that measures of divergence across
721 inversions carrying locally adapted variants should behave according to reduced migration, and
722 reduced N_e as a consequence, with inflated measures of divergence and differentiation relative to
723 other genomic regions. This is predicted to lead to their over-representation in genome scan
724 outliers, thus caution should be taken regarding the adaptive significance of these outliers relative to
725 absolutely lower values of genetic divergence attained from regions outside of chromosomal
726 rearrangements.
727
728
729 ***Limitations of simulations***
730 The factorial design of our study represents a substantial progression in understanding the interplay
731 between individual features of demographic history and multiple measures of population
732 divergence. However, like all models it is limited. Computational limitations constrained absolute

population sizes to a maximum of 1,000 individuals and generations to a maximum of 10,000. To mitigate these constraints, we repeated the analysis over multiple mutation rates to illustrate patterns over 100-fold variation of θ. In general, most trends were consistent, suggesting that results should be consistent across taxa of variable effective population size. However certain patterns are exaggerated, for instance it is clear that the time lag on informativeness of D_{XY} on selection is shortest when θ is largest, which results in stronger positive associations between F_{ST} and D_{XY} by generation 10,000. It is well documented that both N_e (61) and mutation rate (μ) (62) are highly variable across taxa, which suggests that applying knowledge of the relationships between measures of population differentiation will vary by system.

Furthermore, temporal variation within our simulations is confounded by the time at which major shifts in the diverging population's phenotype (Figure S46-49), and by extension selective sweeps, occur. This variation is random with respect to adaptive mutations arising *de novo*, but is also influenced by demography (Figure S50). For example, treatments that are founded with strong bottlenecks are less likely to evolve using variation in the founder, increasing dependence on *de novo* mutations. Additionally, variation in our population contraction parameter modifies the per population number of new mutations per generation. This is also true for features of simulated genes, such as absolute length of coding sequence. Predictable temporal variation in the time at which adaptation is likely to occur is a probable source of variance between measures of differentiation and is particularly clear for Δπ.

Concluding remarks

In conclusion, we have used forward-in-time simulations to perform a factorial experiment in which we demonstrate the relationships between three measures of genetic differentiation, selection, and features of demographic history that are commonly variable in natural populations. In agreement with previous theoretical work, we find 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 provide novel comparisons between measures of genetic differentiation that call into question the perceived wisdom that using multiple measures of genetic differentiation can rule out false-positives. We also demonstrate that signals of convergent evolution across independent replicates can be driven by similar demographic histories with minimal influences of selection. Therefore, we strongly advise those using overlapping outlier scans to carefully consider the demographic context of their system to avoid false-positives.

Ultimately, the need to quantify and correct for variable demographic histories when interpreting measures of genetic differentiation and divergence is not new advice. It is however still relatively uncommon to see demography taken into consideration in published genome scan studies. It is tempting to assume that replication in study design or analysis in the form of taking multiple measures of genetic divergence can reduce the risk of attaining false-positives. We hope to emphasise in this study that this is not the case, as false-positives (i.e. genome scan outliers that are not associated with regions under natural selection) can be driven by non-random genomic or common demographic features that cannot be bypassed through replication. Moreover, many of the patterns we observe are variable through time, such that the relevant pitfalls of analyses will depend on the age of the populations being considered.

Finally, we hope that researchers interested in studying adaptation in the genomes of natural populations will find this study informative in aiding interpretations of outlier scans and choosing appropriate measures suitable to their demographic context.

ACKNOWLEDGEMENTS

The authors wish to thank all members of the Fraser lab for insightful discussions on this project, as well as members of the University of Exeter's EB theme and the University of Sussex's EBE group. The authors would also like to thank Dr Kai Zeng for feedback on an earlier copy of this manuscript.

AUTHOR CONTRIBUTIONS

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

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).

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

1. Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL, et al. The search for loci under selection: trends, biases and progress. Mol Ecol. 2018;27(6):1342–56.
2. Cooper EA, Uy JAC. Genomic evidence for convergent evolution of a key trait underlying divergence in island birds. Mol Ecol. 2017;26(14):3760–74.
3. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 2010;6(2).
4. Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, et al. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Curr Biol. 2012;22(1):83–90.
5. Rougemont Q, Gagnaire P, Perrier C, Genthon C, Besnard A, Launey S, et al. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Mol Ecol. 2017;26(1):142–62.
6. Fraser BA, Künstner A, Reznick DN, Dreyer C, Weigel D. Population genomics of natural and experimental populations of guppies (Poecilia reticulata). Mol Ecol. 2015;24(2):389–408.
7. Reid NM, Proestou DA, Clark BW, Warren WC, Colbourne JK, Shaw JR, et al. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science (80-). 2016;354(6317):1305–8.
8. Meier JI, Marques DA, Wagner CE, Excoffier L, Seehausen O. Genomics of parallel ecological speciation in Lake Victoria cichlids. Mol Biol Evol [Internet]. 2018;(April 2018):1–37. Available from: https://academic.oup.com/mbe/advance-article/doi/10.1093/molbev/msy051/4958411
9. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, et al. Stick insect genomes reveal natural selection’s role in parallel speciation. Science (80-). 2014;344(6185):738–42.
10. Van Belleghem SM, Vangestel C, De Wolf K, De Corte Z, Most M, Rastas P, et al. Evolution at two time frames. PLoS Genet. 2018;
11. Waterhouse MD, Erb LP, Beever EA, Russello MA. Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. Mol

833 Ecol. 2018;27(11):2512–28.
834 12. Ravinet M, Westram A, Johannesson K, Butlin R, André C, Panova M. Shared and nonshared
835 genomic divergence in parallel ecotypes of Littorina saxatilis at a local scale. Mol Ecol.
836 2016;25(1):287–305.
837 13. Westram AM, Galindo J, Alm Rosenblad M, Grahame JW, Panova M, Butlin RK. Do the same
838 genes underlie parallel phenotypic divergence in different Littorina saxatilis populations? Mol
839 Ecol. 2014;23(18):4603–16.
840 14. Trucchi E, Frajman B, Haverkamp THA, Schönswetter P, Paun O. Genomic analyses suggest
841 parallel ecological divergence in heliosperma pusillum (caryophyllaceae). New Phytol.
842 2017;216(1):267–78.
843 15. Roda F, Ambrose L, Walter GM, Liu HL, Schaul A, Lowe A, et al. Genomic evidence for the
844 parallel evolution of coastal forms in the *Senecio lautus* complex. Mol Ecol.
845 2013;22(11):2941–52.
846 16. Ellegren H, Wolf JBW. Parallelism in genomic landscapes of differentiation, conserved
847 genomic features and the role of linked selection. J Evol Biol [Internet]. 2017;30(8):1516–8.
848 Available from: http://doi.wiley.com/10.1111/jeb.13113
849 17. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral
850 molecular variation. Genetics. 1993;134(4):1289–303.
851 18. Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD. Thinking too positive? Revisiting
852 current methods of population genetic selection inference. Trends Genet. 2014;30(12):540–
853 6.
854 19. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to
855 reduced diversity, not reduced gene flow. Mol Ecol. 2014;23(13):3133–57.
856 20. Kimura M. The neutral theory of molecular evolution. Cambridge University Press; 1983.
857 21. Charlesworth D, Charlesworth B, Morgan MT. The pattern of neutral molecular variation
858 under the background selection model. Genetics. 1995;141(4):1619–32.
859 22. Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: Unifying the disparity
860 among species. Nat Rev Genet [Internet]. 2013;14(4):262–74. Available from:
861 http://dx.doi.org/10.1038/nrg3425
862 23. Charlesworth B. Effective population size and patterns of molecular evolution and variation.
863 Nat Rev Genet. 2009;10(3):195.
864 24. Yeaman S, Otto SP. Establishment and maintenance of adaptive genetic divergence under
865 migration, selection, and drift. Evol Int J Org Evol. 2011;65(7):2123–9.
866 25. Torres R, Szpiech ZA, Hernandez RD. Human demographic history has amplified the effects of

867 background selection across the genome. PLoS Genet. 2018;14(6):1–27.
868 26. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the Drosophila
869 genome? PLoS Genet. 2009;5(6):e1000495.
870 27. Charlesworth B. Background selection and patterns of genetic diversity in Drosophila
871 melanogaster. Genet Res (Camb). 1996;68(2):131–49.
872 28. Zeng K. A coalescent model of background selection with recombination, demography and
873 variation in selection coefficients. Heredity (Edinb). 2013;110(4):363.
874 29. Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced
875 polymorphism and background selection on equilibrium patterns of genetic diversity in
876 subdivided populations. Genet Res (Camb). 1997;70(2):155–74.
877 30. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the
878 Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am Nat
879 [Internet]. 2016;188(4):379–97. Available from:
880 http://www.journals.uchicago.edu/doi/10.1086/688018
881 31. Charlesworth B. Measures of divergence between populations and the effect of forces that
882 reduce variability. Mol Biol Evol. 1998;15(5):538–43.
883 32. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure
884 Author (s): B . S . Weir and C . Clark Cockerham Published by : Society for the Study of
885 Evolution Stable URL : http://www.jstor.org/stable/2408641 Accessed : 01-08-2016 16 : 07
886 UTC Your u. Evolution (N Y). 1984;38(6):1358–70.
887 33. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence
888 data. Genetics. 1992;132(2):583–9.
889 34. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res (Camb).
890 1974;23(1):23–35.
891 35. Nei M. Molecular evolutionary genetics. Columbia university press; 1987.
892 36. Matthey-Doret R, Whitlock MC. Background selection and the statistics of population
893 differentiation: consequences for detecting local adaptation. bioRxiv. 2018;326256.
894 37. Whitlock MC, Lotterhos KE. Reliable detection of loci responsible for local adaptation:
895 inference of a null model through trimming the distribution of F ST. Am Nat.
896 2015;186(S1):S24–36.
897 38. Hohenlohe PA, Bassham S, Currey M, Cresko WA. Extensive linkage disequilibrium and
898 parallel adaptive divergence across threespine stickleback genomes. Philos Trans R Soc L B
899 Biol Sci [Internet]. 2012;367(1587):395–408. Available from:
900 http://www.ncbi.nlm.nih.gov/pubmed/22201169

39. Reznick D, Endler JA. The impact of predation on life history evolution in Trinidadian guppies (Poecilia reticulata). Evolution (N Y). 1982;36(1):160–77.

40. Haller BC, Messer PW. SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol Biol Evol. 2019;36(3):632–7.

41. Künstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, et al. The genome of the trinidadian guppy, Poecilia reticulata, and variation in the Guanapo population. PLoS One. 2016;11(12):1–25.

42. Lanfear R, Kokko H, Eyre-Walker A. Population size and the rate of evolution. Trends Ecol Evol. 2014;29(1):33–41.

43. Ellegren H, Galtier N. Determinants of genetic diversity. Nat Rev Genet. 2016;17(7):422–33.

44. Hartl DL, Clark AG, Clark AG. Principles of population genetics. Vol. 116. Sinauer associates Sunderland; 1997.

45. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. Mol Ecol. 2009;18(3):375–402.

46. Turner TL, Hahn MW, Nuzhdin S V. Genomic islands of speciation in Anopheles gambiae. PLoS Biol. 2005;3(9):e285.

47. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J Evol Biol. 2017;30(8):1450–77.

48. Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JBW. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. Mol Ecol. 2017;26(16):4284–95.

49. Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Mol Ecol. 2016;25(18):4488–507.

50. Stankowski S, Chase MA, Fuiten AM, Ralph PL, Streisfeld MA. The tempo of linked selection: Rapid emergence of a heterogeneous genomic landscape during a radiation of monkeyflowers [Internet]. bioRxiv. 2018. 342352 p. Available from: https://www.biorxiv.org/content/early/2018/06/21/342352

51. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515(7526):261.

52. Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M. Correlated patterns of genetic diversity and differentiation across an avian family. Mol Ecol. 2017;26(15):3982–97.

935 53. Dutoit L, Vijay N, Mugal CF, Bossu CM, Burri R, Wolf J, et al. Covariation in levels of nucleotide
936 diversity in homologous regions of the avian genome long after completion of lineage sorting.
937 Proc R Soc B. 2017;284(1849):20162756.
938 54. Kess T, Galindo J, Boulding EG. Genomic divergence between Spanish Littorina saxatilis
939 ecotypes unravels limited admixture and extensive parallelism associated with population
940 history. Int J Bus Innov Res. 2018;17(3):8311–27.
941 55. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics.
942 2006;173(1):419–34.
943 56. Enard D, Messer PW, Petrov DA. Genome-wide signals of positive selection in human
944 evolution. Genome Res. 2014;
945 57. Noor MAF, Bennett SM. Islands of speciation or mirages in the desert? Examining the role of
946 restricted recombination in maintaining species. Heredity (Edinb). 2009;103(6):439.
947 58. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of
948 adaptive evolution in threespine sticklebacks. Nature [Internet]. 2012;484(7392):55–61.
949 Available from: http://www.nature.com/doifinder/10.1038/nature10944
950 59. Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, et al. A genetic mechanism for
951 female-limited Batesian mimicry in Papilio butterfly. Nat Genet. 2015;47(4):405.
952 60. Morales HE, Faria R, Johannesson K, Larsson T, Panova M, Westram AM, et al. Genomic
953 architecture of parallel ecological divergence: beyond a single environmental contrast.
954 bioRxiv. 2018;447854.
955 61. Frankham R. Effective population size/adult population size ratios in wildlife: a review. Genet
956 Res (Camb). 1995;66(2):95–107.
957 62. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes.
958 Nat Rev Genet. 2011;12(11):756.
959
960
961
962
963
964