Estimating the genome-wide contribution of selection to temporal allele frequency change

Vince Buffalo[*,†,1] and Graham Coop[†]

[*]Population Biology Graduate Group
[†]Center for Population Biology, Department of Evolution and Ecology, University of California, Davis, CA 95616
[1]Email for correspondence: vsbuffalo@ucdavis.edu

October 8, 2019

Abstract

Rapid phenotypic adaptation is often observed in natural populations and selection experiments. However, detecting the genome-wide impact of this selection is difficult, since adaptation often proceeds from standing variation and selection on highly polygenic traits, both of which may leave faint genomic signals indistinguishable from a noisy background of genetic drift. One promising signal comes from the genome-wide covariance between allele frequency changes observable from temporal genomic data, e.g. evolve-and-resequence studies. These temporal covariances reflect how the change in neutral allele frequency at one timepoint is predictive of the changes at later timepoints when there is heritable fitness variation in the population, as neutral alleles can remain associated with selected alleles over time. Since genetic drift does not lead to temporal covariance, we can use these covariances to estimate what fraction of the variation in allele frequency change through time is driven by linked selection. Here, we reanalyze two *Drosophila simulans* evolve-and-resequence studies, and one artificial selection experiment in mice, to quantify the effects of linked selection over short timescales using covariance among time-points and across replicates. We estimate that at least 17% to 37% of allele frequency change is driven by selection in these experiments. Against this background of positive genome-wide temporal covariances we also identify signals of negative temporal covariance corresponding to reversals in the direction of selection for a reasonable proportion of loci over the time course of a selection experiment. Overall, we find that in the three studies we analyzed, linked selection has a large impact on short-term allele frequency dynamics that is readily distinguishable from genetic drift.

1 Introduction

A long-standing problem in evolutionary genetics is quantifying the roles of genetic drift and selection in shaping genome-wide allele frequency changes. Selection can both directly and indirectly affect allele frequencies, with the indirect effect coming from the action of selection on correlated loci elsewhere in genome e.g. linked selection (Maynard Smith and Haigh 1974, Charlesworth et al. 1993; Nordborg et al. 1996; see Barton 2000 for a review). Previous work on this question has mostly focused on teasing apart the impacts of drift and selection on genome-wide diversity using population samples from a single contemporary timepoint, often by modeling the correlation

between regional recombination rate, gene density, and diversity created in the presence of linked selection (Cutter and Payseur 2013; Sella et al. 2009). This approach has shown linked selection has a major role in shaping patterns of genome-wide diversity across the genomes of a range of sexual species (Andersen et al. 2012; Andolfatto 2007; Begun et al. 2007; Beissinger et al. 2016; Cutter and Choi 2010; Elyashiv et al. 2016; Macpherson et al. 2007; Sattath et al. 2011; Williamson et al. 2014), and has allowed us to quantify the relative influence of positive selection (hitchhiking) and negative selection (background selection; Andolfatto 2007; Elyashiv et al. 2016; Hernandez et al. 2011; Macpherson et al. 2007; McVicker et al. 2009; Nordborg et al. 2005). However, we lack an understanding of how genome-wide linked selection acts over time.

There are numerous examples of rapid phenotypic adaptation (Franks et al. 2007; Grant and Grant 2006, 2011; Reznick et al. 1997) and rapid, selection-driven genomic evolution in asexual populations (Baym et al. 2016; Bennett et al. 1990; Good et al. 2017). Yet the polygenic nature of fitness makes detecting the impact of selection on genome-wide variation over short timescales in sexual populations remarkably difficult. This is because the effect of selection on a polygenic trait (such as fitness) is distributed across loci in proportion to their effect sizes. This can lead to subtle allele frequency shifts on standing variation that are difficult to distinguish from background levels of genetic drift and sampling variance. However, increasingly genomic experimental evolution studies with multiple timepoints, and in some cases multiple replicate populations, are being used to detect large effect selected loci (Turner and Miller 2012; Turner et al. 2011) and differentiate modes of selection (Barghi et al. 2019; Burke et al. 2010; Therkildsen et al. 2019). In addition these temporal-genomic studies have begun in wild populations, some with the goal of finding variants that exhibit frequency changes consistent with fluctuating selection (Bergland et al. 2014; Machado et al. 2018). In a previous paper, we proposed that one useful signal for understanding the genome-wide impact of polygenic linked selection detectable from genomic studies with multiple timepoints is the temporal autocovariance in allele frequency changes (Buffalo and Coop 2019). These covariances are directly estimable from temporal genomic data and are created when the loci that underly heritable fitness variation perturb the frequencies of linked neutral alleles; in contrast, when genetic drift acts alone in a closed population, these covariances are zero in expectation. Mathematically, temporal covariances are useful because it is natural to decompose the total variance in allele frequency change across a set of time intervals into the variances and covariances in allele frequency change among time intervals. Furthermore, biologically, these covariances reflect the extent to which neutral allele frequency changes in one generation predict changes in another due to a shared selection pressures and associations to selected loci.

Here, we provide the first empirical analyses to quantify the impact of linked selection acting over short timescales (tens of generations) across two evolve and re-sequence studies (Barghi et al. 2019; Kelly and Hughes 2019), and an artificial selection experiment (Castro et al. 2019). We repeatedly find a signal of temporal covariance, consistent with linked selection acting to significantly perturb genome-wide allele frequency changes across the genome in a manner that other approaches would not be able differentiate from genetic drift. We estimate the lower bound on the proportion of total variation in allele frequency change caused by selection, and the correlation between allele frequency changes between replicate populations caused by the response to convergent selection pressures. Overall, we demonstrate that linked selection has a powerful role in shaping genome-wide allele frequency changes over very short timescales.

Study	Species	Selection	Replicates	Pop. Size	Gens.	Timepoints
Kelly and Hughes (2019)	*D. simulans*	lab adaptation	3	∼1100	14	2
Barghi et al. (2019)	*D. simulans*	lab adaptation	10	∼1000	60	7
Castro et al. (2019)	*M. musculus*	tibiae length	2	32	20	2
		control	1	28		

Table 1: A summary of the main selection studies we analyzed.

2 Results

We first analyzed Barghi et al. (2019), an evolve-and-resequence study with ten replicate populations exposed to a high temperature lab environment and evolved for 60 generations, and sequenced every ten generations. Using the seven timepoints and ten replicate populations, we estimated the genome-wide 6×6 temporal covariance matrix $\mathbf{Q}$ for each of the ten replicates. Each row of these matrices represent the temporal covariance $\mathrm{Cov}(\Delta_{10}p_s, \Delta_{10}p_t)$, between the allele frequency change (in ten-generation intervals, denoted $\Delta_{10}p_t$) in some initial reference generation s (the row of the matrix), and some later timepoint t (the column of the matrix). We corrected these matrices for biases created due to sampling noise, and normalize the entries for heterozygosity (see Supplementary Materials Sections 1.1.2 and 1.1.4). *These* covariances are expected to be zero when only drift is acting, as only heritable variation for fitness can create covariance between allele frequency changes in a closed population (Buffalo and Coop 2019). Averaging across the ten replicate temporal covariances matrices, we find temporal covariances that are statistically significant (95% block bootstraps CIs do not contain zero), consistent with linked selection perturbing genome-wide allele frequency changes over very short time periods. The covariances between all adjacent time intervals are positive and then decay towards zero as we look at more distant time intervals, as expected when directional selection affects linked variants' frequency trajectories until ultimately linkage disequilibrium and the additive genetic variance for fitness associated with neutral alleles decays (Buffalo and Coop 2019). The temporal covariances per replicate are noisier but this general pattern holds; see Supplementary Figure S2. Barghi et al. (2019)'s design means that the covariances we see in adjacent time intervals are on average ten generations apart, and given the temporal decay in covariance we see, the covariances on shorter time-scales (e.g. if adjacent generations had been sequenced) may well be higher yet (see Supplementary Material Section 1.1.5 for more details).

One concern is that these covariances reflect the localized impact of a few large-effect loci rather than selection on a polygenic trait. Since our covariances are essentially averages over loci, the covariance estimate can be strongly affected by a single outlier region. To test whether large outlier regions drive the genome-wide signal we see in the Barghi et al. (2019) data, we calculate the covariances in 100kb windows along the genome (we refer to these as windowed covariances throughout) and take the median windowed covariance, and trimmed-mean windowed covariance, as a measure of the genome-wide covariance robust to large-effect loci. These robust estimates (Supplementary Table S1 and Supplementary Figure S3) confirm the patterns we see using the mean covariance, confirming that genomic temporal covariances are non-zero due to the impact of selection *acting across* many genomic windows.

While the presence of positive temporal covariances is consistent with selection affecting allele frequencies over time, this measure is not easily interpretable. We can calculate a more intuitive measure from the temporal covariances to quantify the impact of selection on allele frequency change: the ratio of total covariance in allele frequency change to the total variance in allele

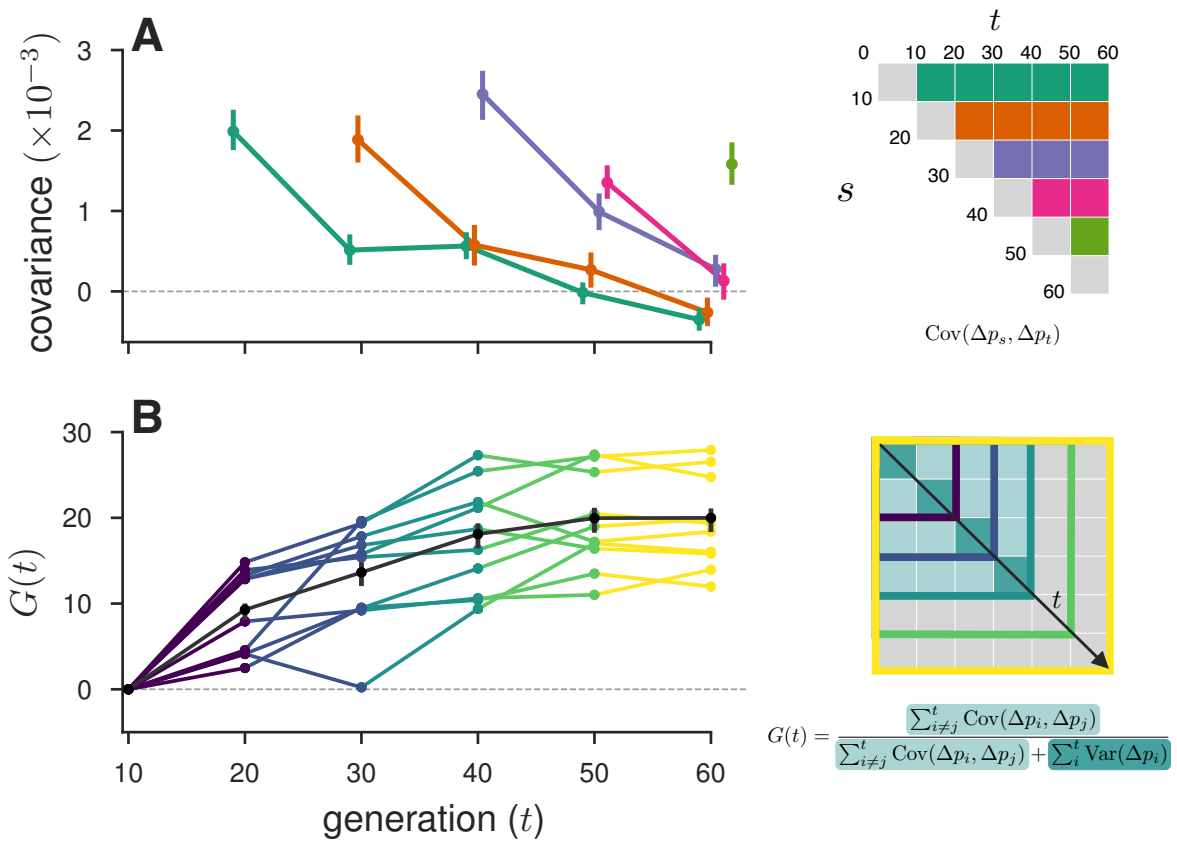

Figure 1: A: Temporal covariance, averaged across all ten replicate populations, through time from the Barghi et al. (2019) study. Each line depicts the temporal covariance $\mathrm{Cov}(\Delta p_s, \Delta p_t)$ from some reference generation s to a later time t which varies along the x-axis; each line corresponds to a row of the upper-triangle of the temporal covariance matrix with the same color (upper right). The ranges around each point are 95% block-bootstrap confidence intervals. B: The proportion of the total variance in allele frequency change explained by linked selection, $G(t)$, as it varies through time t along the x-axis. The black line is the $G(t)$ averaged across replicates, with the 95% block-bootstrap confidence interval. The other lines are the $G(t)$ for each individual replicate, with colors indicating what subset of the temporal-covariance matrix to the right is being included in the calculation of $G(t)$.

frequency change. We denote the change in allele frequency as $\Delta p_t = p_{t+1} - p_t$, where p_t is the allele frequency in generation t. Since the total variation in allele frequency change can be partitioned into variance and covariance components, $\mathrm{Var}(p_t - p_0) = \sum_{i=0}^{t-1} \mathrm{Var}(\Delta p_i) + \sum_{i \neq j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j)$ (we bias correct these for sequencing depth), and the covariances are zero when drift acts alone, this is a lower bound on how much of the variance in allele frequency change is caused by linked selection (Buffalo and Coop 2019). We call this measure $G(t)$, defined as

$$G(t) = \frac{\sum_{i \neq j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j)}{\mathrm{Var}(p_t - p_0)} \tag{1}$$

which estimates the effect of selection on allele frequency change between the initial generation 0 and some later generation t, which can be varied to see how this quantity grows through time.

Since Barghi et al. (2019) experiment is sequenced every ten generations, in the numerator for the covariance we use the allele frequency changes between adjacent timepoints, which are ten generations apart. Consequently, this leads our measure $G(t)$ to be strongly conservative, since the temporal covariances within each ten-generation block are not directly observable, and thus are not included in the numerator of $G(t)$. Still, we find a remarkably strong signal. Greater than 20% of total, genome-wide allele frequency change over 60 generations is the result of selection (Figure 1B).

Additionally, we looked for a signal of temporal autocovariance in Bergland et al. (2014), a study that collected *Drosophila melanogaster* through Spring-Fall season pairs across three years. If there was a strong pattern of genome-wide fluctuating selection, we might expect a pattern of positive covariances between similar seasonal changes, e.g. Spring-Fall in two adjacent years, and negative covariances between dissimilar seasonal changes, e.g. Spring-Fall and Fall-Spring in two adjacent years. However, we find no such signal over years; we discuss this in more depth in Supplementary Materials Section 1.2.3.

The replicate design of Barghi et al. (2019) allows us to quantify another covariance: the covariance in allele frequency change between replicate populations experiencing convergent selection pressures. These between-replicate covariances are created in the same way as temporal covariances: neutral alleles linked to a particular fitness background are expected to have allele frequency changes in the same direction if the selection pressures are similar. Intuitively, where temporal covariances reflect that neutral alleles associated with heritable fitness backgrounds are predictive of frequency changes between generations, replicate covariances reflect that heritable fitness backgrounds common to each replicate predict (under the same selection pressures) frequency changes between replicates. We measure this through a statistic similar to a correlation, which we call the convergent correlation: the ratio of average between-replicate covariance across all pairs to the average standard deviation across all pairs of replicates,

$$\mathrm{cor}(\Delta p_s, \Delta p_t) = \frac{\mathbb{E}_{A \neq B}\left(\mathrm{Cov}(\Delta p_{s,A}, \Delta p_{t,B})\right)}{\mathbb{E}_{A \neq B}\left(\sqrt{\mathrm{Var}(\Delta p_{s,A})\,\mathrm{Var}(\Delta p_{t,B})}\right)} \tag{2}$$

where A and B here are two replicate labels, and for the Barghi et al. (2019) data, we use $\Delta_{10} p_t$.

We've calculated the convergent correlation for all rows of the replicate covariance matrices. Like temporal covariances, we visualize these through time (Figure 2A), with each line representing the convergent correlation from a particular reference generation s as it varies with t (shown on the x-axis). In other words, each of the colored lines corresponds to the like-colored row of the convergence correlation matrix (upper left in Figure 2A). We find these convergent covariances are relatively weak, and decay very quickly from an initial value of about 0.1 (95% block bootstrap confidence intervals $[0.094, 0.11]$) to around 0.01 (95% CIs [0.0087, 0.015]) within 20 generations. This suggests that while a reasonable fraction of the initial response is shared over the replicates, this is followed by a rapid decay, a result consistent with the primary finding of the original Barghi et al. (2019) study: that alternative loci contribute to longer term adaptation *across* the different replicates.

A benefit of between-replicate covariances is that unlike temporal covariances, these can be calculated with only two sequenced timepoints and a replicated study design. This allowed us to assess the impact of linked selection in driving convergent patterns of allele frequency change across replicate populations in two other studies. First, we reanalyzed the selection experiment of Kelly

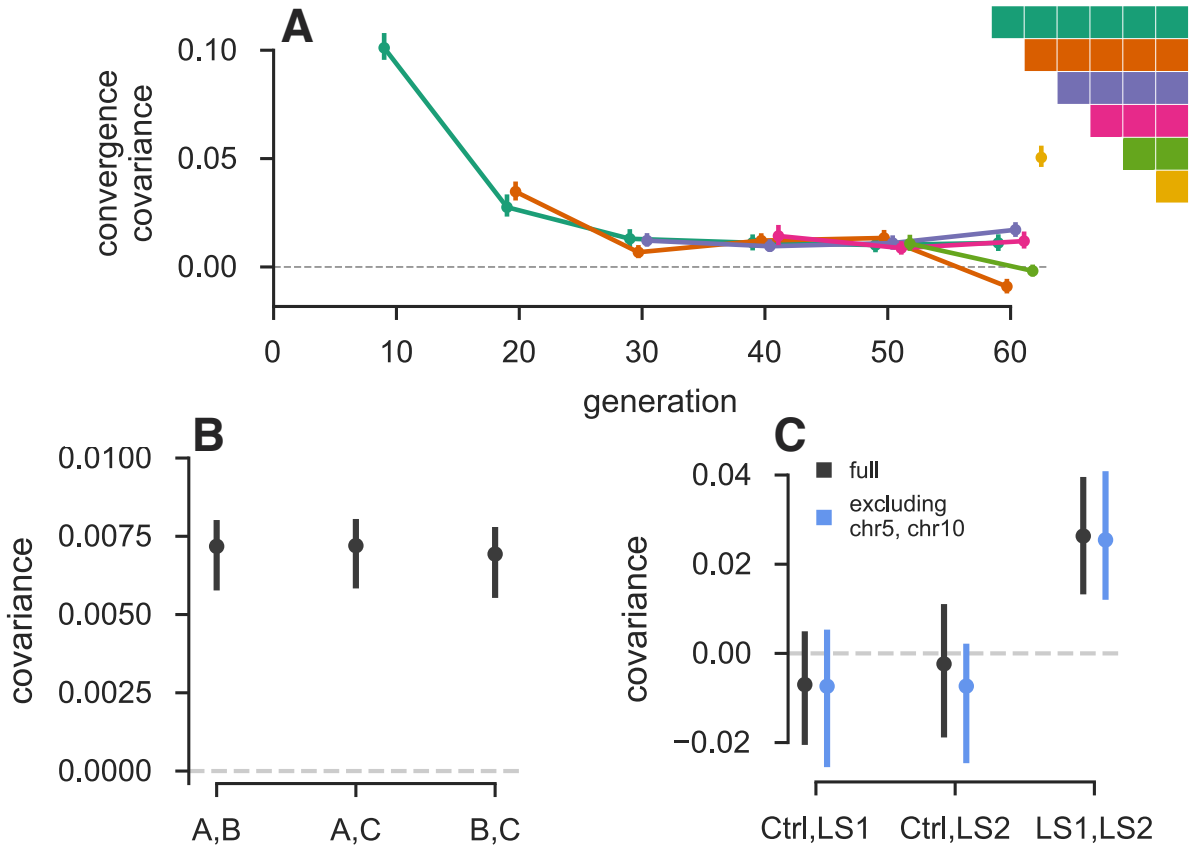

Figure 2: **A**: The convergence correlation, averaged across Barghi et al. (2019) replicate pairs, through time. Each line represents the convergence correlation cor$(\Delta p_s, \Delta p_s)$ from a starting reference generation s to a later time t, which varies along the x-axis; each line corresponds to a row of the temporal convergence correlation matrix depicted to the right. We note that convergent correlation for the last timepoint is an outlier; we are unsure as to the cause of this, e.g. it does not appear to be driven by a single pair of replicates. **B**: The convergence covariance between individual pairs of replicates in the Kelly and Hughes (2019) data. **C**: The convergence covariance between individual pairs of replicates in (Castro et al. 2019) data, for the two selection lines (LS1 and LS2) and the control (Ctrl); gray CIs are those using the complete dataset, blue CIs exclude chromosomes 5 and 10 which harbor the two regions Castro et al. (2019) found to have signals of parallel selection between LS1 and LS2.

and Hughes (2019), which evolved three replicate wild populations of *Drosophila simulans* for 14 generations adapting to a novel laboratory environment. Since each replicate was exposed to the same selection pressure and share linkage disequilibria common to the original natural founding population, we expected each of the three replicate populations to have positive between-replicate covariances. We find all three pairwise between-replicate covariances are positive and statistically significant (Figure 2 B. We estimate the convergent correlation coefficient across these replicates as 0.36 (95% CI $[0.31, 0.40]$). Similarly, we can calculate the proportion of the total variance in allele frequency change from convergent selection pressure analogous to G where the numerator is is the convergent covariance and the denominator is the total variance (see Supplementary Material 1.1.7). We find that 37% of the total variance is due to shared allele frequency changes caused by

selection (95% CI [29%, 41%]; these are similar to the convergence correlation, since the variance is relatively constant across the replicates.

Next, we reanalyzed the Longshanks selection experiment, which selected for longer tibiae length relative to body size in mice, leading to a response to selection of about 5 standard deviations over the course of twenty generations (Castro et al. 2019; Marchini et al. 2014). This study includes two independent selection lines, Longshanks 1 and 2 (LS1 and LS2), and an unselected control line (Ctrl). Consequently, this selection experiment offers a useful control to test our between-replicate covariances: we expect to see positive between-replicate covariance in the comparison between the two Longshanks selection lines, but not between the two pairwise comparisons between the control line and *each of* the two Longshanks lines. We find that this is the case (gray confidence intervals in Figure 2 C), with the two Longshanks comparisons to the control line not being significantly different from zero, while the comparison between the two Longshanks lines is statistically significantly different from zero (CIs [0.0129, 0.0400]).

One finding in the Longshanks study *was that two* major-effect loci showed parallel frequency shifts between the two selection lines: a region harboring the gene Nkx3-2 known to be involved in limb development, and another region containing six other candidate genes. We were curious to what extent our genome-wide covariances were being driven by these two outlier large-effect loci, so we excluded them from the analysis. Since we do not know the extent to which linkage disequilibrium around these large-effect loci affects neighboring loci, we took the conservative precaution of excluding the entire chromosomes these loci reside on (chromosomes 5 and 10), and re-calculating the temporal covariances. We find excluding these large effect loci has little impact on the confidence intervals (blue confidence intervals in Figure 2 C), indicating that these across-replicate covariances are indeed driven by a large number of loci. This is consistent with a signal of selection on a polygenic trait driving genome-wide change, although we note that large-effect loci can contribute to the indirect change at unlinked loci (Robertson 1961; Santiago and Caballero 1995).

The presence of an unselected control line provides an alternative way to partition the effects of linked selection and genetic drift: we can compare the total variance in allele frequency change of the control line (which excludes the effect of artificial selection on allele frequencies) to the total variance in frequency change of the Longshanks selection lines. We can partition the allele frequency change between the two timepoints (20 generations apart) for a Longshanks line as $\Delta p_{t,\mathrm{LS1}} = \Delta_D p_{t,\mathrm{LS1}} + \Delta_U p_{t,\mathrm{LS1}} + \Delta_S p_{t,\mathrm{LS}}$ where these terms are the decomposition in the allele frequency change due to drift in *Longshanks replicate 1 ($\Delta_D p_{t,\mathrm{LS1}}$), selection unique to the LS1 replicate ($\Delta_U p_{t,\mathrm{LS1}}$), and selection response shared between the two Longshanks replicates ($\Delta_S p_{t,\mathrm{LS}}$) respectively (and similarly for the Longshanks two line, LS2)*. By construction we will assume that each of these terms are uncorrelated within replicates, and that only the shared term covaries between the replicates. Assuming that we can approximate the contribution of genetic drift in the LS lines as the variance in allele frequency change observed in the control, i.e. $\mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}}) = \mathrm{Var}(\Delta_D p_{t,\mathrm{LS2}}) = \mathrm{Var}(\Delta_D p_{t,\mathrm{LS2}})$, then we can estimate the increase in variance in allele frequency change due to selection as $(\mathrm{Var}(\Delta p_{t,\mathrm{LS1}}) + \mathrm{Var}(\Delta p_{t,\mathrm{LS2}}))/2 - \mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}})$ and the shared effect of selection across selected lines as $\mathrm{Cov}(\Delta p_{t,\mathrm{LS1}}, \Delta p_{t,\mathrm{LS2}})$ (see Supplementary Material Section 1.1.7 for more details). We estimate at least 32% (95% CI [21%, 48%]) of the variance in allele frequency change is driven by the effects of selection, of which 14% (95% CI [3%, 33%]) is estimated to be unique to a selection line, and 17% (95% CI [9%, 23%]) is the effect of shared selection between the two Longshanks selection lines.

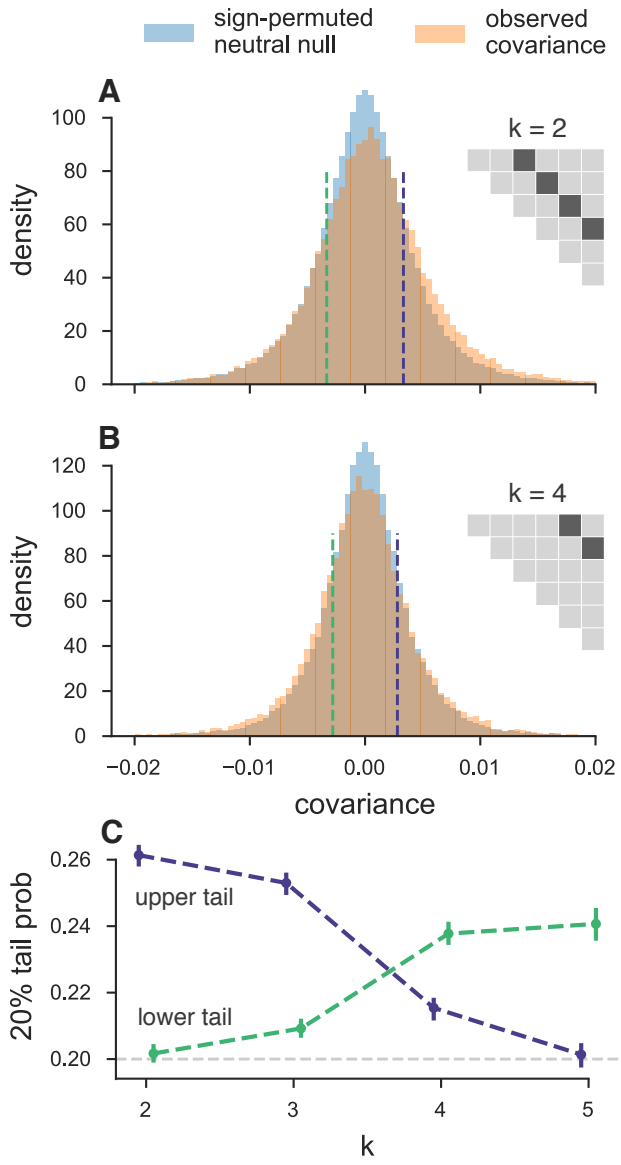

Figure 3: **A**, **B**: The distribution of temporal covariances calculated in 100kb genomic windows from the Barghi et al. (2019) study, plotted alongside an empirical neutral null distribution created by recalculating the windowed covariances on 1,000 sign permutations of allele frequency changes within tiles. The histogram bin number is 88, chosen by cross validation (Supplementary Materials S4). In subfigure **A**, windowed covariances $\mathrm{Cov}(\Delta p_t, \Delta p_{t+k})$ are separated by $k = 2 \times 10$ generations and in subfigure **A** the covariances are separated by $k = 4 \times 10$ generations; each k is an off-diagonal from the variance diagonal of the temporal covariance matrix (see cartoon of upper-triangle of covariance matrix in subfigures **A** and **B**, where the first diagonal is the variance, and the dark gray indicates which off-diagonal of the covariance matrix is plotted in the histograms). **C**: The lower and upper tail probabilities of the observed windowed covariances, at 20% and 80% quintiles of the empirical neutral null distribution, for varying time between allele frequency changes (i.e. which off-diagonal k). The confidence intervals are 95% block-bootstrap confidence intervals, and the light gray dashed line indicates the 20% tail probability expected under the neutral null. Similar figures for different values of k are in Supplementary Figures S6.

Finally, we observed that in the longest study we analyzed (Barghi et al. 2019), some genome-wide temporal covariances become negative at future timepoints (see the first two rows in Figure 1 A). This shows that alleles that were on average going up initially are later going down in frequency, i.e. that the average direction of selection experienced by alleles has flipped. This must reflect either a change in the environment or the genetic background, due to epistatic relationships among alleles altered by frequency changes or recombination breaking up selective alleles. Such reversals of selective dynamics could be occurring at other timepoints but the signal of a change in the direction of selection at particular loci may be washed out when we calculate our genome-wide average temporal covariances. To address this limitation, we calculated the distribution of the temporal covariances over *100kb windowed covariances* (Figure 3 shows these distributions pooling across all replicates; see Supplementary Figure S5 for individuals replicates). The covariance estimate of each genomic window will be noisy, due to sampling and genetic drift, and the neutral distribution of the covariance is complicated due to linkage disequilibria (which can occur over long physical distances in E&R and selection studies, Baldwin-Brown et al. 2014; Nuzhdin and Turner 2013). To address this, we have developed a permutation-based procedure that constructs an empirical null distribution by randomly flipping the signs of the allele frequency changes per-genomic window. This destroys the systematic covariances created by linked selection and creates a sampling distribution of the covariances spuriously created by neutral genetic drift while preserving the complex dependencies between adjacent loci created by linkage disequilibrium. This empirical neutral null distribution is conservative in the sense that the variances of the covariances are wider than expected under drift alone as they include the effect of selection on the allele frequency change within a time-interval, just not between time-intervals. We see (Figure 3 A and B) that windowed temporal covariances between close timepoints are skewed positive (a heavy right tail), while between more distant timepoints these windowed temporal covariances tend to shift to become more negative (a heavy left tail). We quantified the degree to which the left and right tails are inflated compared to the null distribution as a function of time, and see excesses in both tails in Figure 3 C. This finding is also robust to sign-permuting allele frequency changes on a chromosome-level, the longest extent that gametic linkage disequilibria can extend (Supplementary Figure S8). We see a striking pattern that the windowed covariances not only decay towards zero, but in fact become negative through time, consistent with many regions in the genome having had a reversed fitness effect at later timepoints.

3 Discussion

Since the seminal analysis of Maynard Smith and Haigh (1974) demonstrating that linked neutral diversity is reduced as an advantageous polymorphism arises and sweeps to fixation, over four decades of theoretical and empirical research has enriched our understanding of linked selection. One under-used approach to understand the genome-wide effects of selection on standing variation, e.g. selection on an infinitesimal polygenic trait, stems from an early quantitative genetic model of linked selection (Robertson 1961) and its later developments (Santiago and Caballero 1995, 1998; Woolliams et al. 1993; Wray and Thompson 1990; see also Barton 2000 for a comparison of these models with classic hitchhiking models). Implicit in these models is that autocovariance between allele frequency change is created when there is heritable fitness variation in the population, a signal that may be readily detected from temporal genomic data (Buffalo and Coop 2019). Depending on how many loci affect fitness, such a strong effect of linked selection may not be differentiable

from genetic drift using only single contemporary population samples or looking at at temporal allele frequency change at each locus in isolation. In this way, averaging summaries of temporal data allows us to sidestep the key problem of detecting selection from standing variation: that the genomic footprint leaves too soft of a signature to differentiate from a background of genetic drift. In fact we find that the temporal covariance signal is detectable even in the most extremely difficult to detect soft sweep case: polygenic selection on highly polygenic traits (Buffalo and Coop 2019).

It is worth building some intuition why temporal covariance allows us to detect such faint signals of polygenic linked selection from temporal genomic data. Each variant is subject to both variance in allele frequency due to drift and sampling noise, which at any locus may swamp the temporal covariance signal and creates spurious covariances. However, these spurious covariances do not share a directional signal whereas the covariances created by linked selection do; consequently, averaging across the entire genome, the temporal signal exceeds sampling noise.

Our analyses reveal that a sizable proportion of allele frequency change in these populations is due to the (indirect) action of selection. Capitalizing on replicated designs, we characterized the extent to which convergent selection pressures lead to parallel changes in allele frequencies across replicate populations, and found that a reasonable proportion of the response is shared across short timescales. These likely represent substantial under-estimates of the contribution of linked selection because the studies we have reanalyzed do not sequence the population each generation, preventing us from including the effects of stronger correlations between adjacent generations. Furthermore, our estimation methods are intentionally conservative, for example they exclude the contribution of selection that does not persist across generations and selection that reverses sign; thus they can be seen as a strong lower bound of the effects of selection.

These estimates of the contribution of selection could be refined by using patterns of LD and recombination which would allow us to more fully parameterize a linked-selection model of temporal allele frequency change (Buffalo and Coop 2019). The basic prediction is that regions of higher linkage disequilibrium and lower recombination should have greater temporal autocovariance than regions with lower LD and higher recombination. However, one limitation of these pooled sequence datasets is that none of the studies we reanalyzed estimated linkage disequilibria data for the evolved populations. *While* there are LD data for a natural population of *D. simulans* (Howie et al. 2018; Signor et al. 2018), we did not find a relationship between temporal covariance and LD. We believe this is driven by the idiosyncratic nature of LD in evolve-and-resequence populations, which often extends over large genomic distances (Kelly and Hughes 2019; Nuzhdin and Turner 2013). Future studies complete with LD data and recombination maps would allow one to disentangle the influence of closely linked sites from more distant sites in causing temporal autocovariance, and allow the fitting of more parametric models to estimate population parameters such as the additive genetic variation for fitness directly from temporal genomic data alone (Buffalo and Coop 2019).

Our primary focus here has been on evolution in laboratory populations. It is unclear whether we should expect a similar impact of selection in natural populations. In some of these experiments, selection pressures may have been stronger or more sustained that in natural populations *(Hairston et al. 2005; Hendry and Kinnison 1999)*. Conversely, these lab populations were maintained at very small effective population sizes, estimated at 300, 450, and 45 for the Barghi et al. (2019), Kelly and Hughes (2019), and Castro et al. (2019) studies respectively, which will amplify the role of genetic drift. The advantage of lab experiments is that they are closed populations, in natural populations temporal covariance could also arise from the systematic migration of alleles from differentiated populations. *Adapting* these methods to natural populations will require either populations that

are reasonably closed to migration, or for the effect of migration to be accounted for possibly either by knowledge of allele frequencies in source populations or the identification of migrant individuals.

While it challenging to apply temporal methods to natural populations there is a lot of promise for these approaches (Bergland et al. 2014; Machado et al. 2018). Efforts to quantify the impact of linked selection have found obligately sexual organisms have up to an 89% reduction in genome-wide diversity over long time periods (Comeron 2014; Coop 2016; Corbett-Detig et al. 2015; Elyashiv et al. 2016; McVicker et al. 2009) *Thus linked selection makes a sizeable contribution to long-term allele frequency change in some species,* and there is reason to be hopeful that we could detect this from temporal data, which would help to resolve the timescales that linked selection act over. In our reanalysis of the Barghi et al. (2019) study, we find evidence of complex linked selection dynamics, with selection pressures flipping over time due to either environmental change, the breakup of epistatic combinations or advantageous haplotypes. Such patterns would be completely obscured in samples only from contemporary populations. *Thus,* we can hope to have a much richer picture of the*impact of selection* as temporal sequencing becomes more common, *allowing* us to observe the effects of ecological dynamics in genomic data (Hairston et al. 2005).

Furthermore, understanding the dynamics of linked selection over short timescales will help to unite phenotypic studies of rapid adaptation with a detectable genomic signature, to address long-standing questions concerning linked selection, evolutionary quantitative genetics, and the overall impact selection has on genetic variation.

4 Acknowledgments

We would like to thank the authors of the original studies we've analyzed, including Neda Barghi, Christian Schlötterer, John Kelly, Kimberly Hughes, Frank Chan, Campbell Rolian, Nick Barton, Alan Bergland, and Dmitri Petrov. We would also like to thank Doc Edge for helpful statistical advice, and Matt Osmond, Erin Calfee, and Sivan Yair for helpful discussions. This research was supported by an NSF Graduate Research Fellowship grant awarded to VB (1650042), and NIH (R01-GM108779) and NSF (1353380) awarded to GC.

References

Andersen, Erik C et al. (2012). "Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity". en. In: *Nat. Genet.* 44.3, pp. 285–290.

Andolfatto, Peter (2007). "Hitchhiking effects of recurrent beneficial amino acid substitutions in the Drosophila melanogaster genome". en. In: *Genome Res.* 17.12, pp. 1755–1762.

Baldwin-Brown, J G, A D Long, and K R Thornton (2014). "The Power to Detect Quantitative Trait Loci Using Resequenced, Experimentally Evolved Populations of Diploid, Sexual Organisms". In: *Mol. Biol. Evol.* 31.4, pp. 1040–1055.

Barghi, Neda et al. (2019). "Genetic redundancy fuels polygenic adaptation in Drosophila". en. In: *PLoS Biol.* 17.2, e3000128.

Barton, N H (2000). "Genetic hitchhiking". In: *Proceedings of the Royal Society of London B: Biological Sciences* 355.1403, pp. 1553–1562.

Baym, Michael et al. (2016). "Spatiotemporal microbial evolution on antibiotic landscapes". en. In: *Science* 353.6304, pp. 1147–1151.

Begun, David J et al. (2007). "Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans". en. In: *PLoS Biol.* 5.11, e310.

Beissinger, Timothy M et al. (2016). "Recent demography drives changes in linked selection across the maize genome". en. In: *Nat Plants* 2, p. 16084.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *J. R. Stat. Soc. Series B Stat. Methodol.* 57.1, pp. 289–300.

Bennett, A F, K M Dao, and R E Lenski (1990). "Rapid evolution in response to high-temperature selection". en. In: *Nature* 346.6279, pp. 79–81.

Bergland, Alan O, Emily L Behrman, Katherine R O'Brien, Paul S Schmidt, and Dmitri A Petrov (2014). "Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila". In: *PLoS Genet.* 10.11, e1004775–19.

Bhatia, Gaurav, Nick Patterson, Sriram Sankararaman, and Alkes L Price (2013). "Estimating and interpreting FST: the impact of rare variants". en. In: *Genome Res.* 23.9, pp. 1514–1521.

Buffalo, Vince and Graham Coop (2019). "The Linked Selection Signature of Rapid Adaptation in Temporal Genomic Data". en.

Burke, Molly K et al. (2010). "Genome-wide analysis of a long-term evolution experiment with Drosophila". In: *Nature* 467.7315, pp. 587–590.

Castro, João Pl et al. (2019). "An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice". en. In: *Elife* 8.

Charlesworth, Brian, M T Morgan, and D Charlesworth (1993). "The effect of deleterious mutations on neutral molecular variation". In: *Genetics* 134.4, pp. 1289–1303.

Comeron, Josep M (2014). "Background selection as baseline for nucleotide variation across the Drosophila genome". en. In: *PLoS Genet.* 10.6, e1004434.

Coop, Graham (2016). *Does linked selection explain the narrow range of genetic diversity across species?* Tech. rep. Cold Spring Harbor Labs Journals.

Corbett-Detig, Russell B, Daniel L Hartl, and Timothy B Sackton (2015). "Natural Selection Constrains Neutral Diversity across A Wide Range of Species". In: *PLoS Biol.* 13.4, e1002112–25.

Cutter, A D and J Y Choi (2010). "Natural selection shapes nucleotide polymorphism across the genome of the nematode Caenorhabditis briggsae". In: *Genome Res.* 20.8, pp. 1103–1111.

Cutter, Asher D and Bret A Payseur (2013). "Genomic signatures of selection at linked sites: unifying the disparity among species". In: *Nature Publishing Group* 14.4, pp. 262–274.

Davison, A C and D V Hinkley (2013). *Bootstrap Methods and Their Application*. New York, NY, USA: Cambridge University Press.

Elyashiv, Eyal et al. (2016). "A Genomic Map of the Effects of Linked Selection in Drosophila". In: *PLoS Genet.* 12.8, e1006130–24.

Franks, Steven J, Sheina Sim, and Arthur E Weis (2007). "Rapid evolution of flowering time by an annual plant in response to a climate fluctuation". In: *Proceedings of the National Academy of Sciences* 104.4, pp. 1278–1282.

Good, Benjamin H, Michael J McDonald, Jeffrey E Barrick, Richard E Lenski, and Michael M Desai (2017). "The dynamics of molecular evolution over 60,000 generations". en. In: *Nature* 551.7678, pp. 45–50.

Grant, Peter R and B Rosemary Grant (2006). "Evolution of Character Displacement in Darwin's Finches". In: *Science* 313.5784, pp. 224–226.

Grant, Peter R and B Rosemary Grant (2011). "Causes of lifetime fitness of Darwin's finches in a fluctuating environment". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 108.2, pp. 674–679.

Hairston Jr, Nelson G, Stephen P Ellner, Monica A Geber, Takehito Yoshida, and Jennifer A Fox (2005). "Rapid evolution and the convergence of ecological and evolutionary time". In: *Ecol. Lett.* 8.10, pp. 1114–1127.

Hendry, A P and M T Kinnison (1999). "Perspective: the pace of modern life: measuring rates of contemporary microevolution". In: *Evolution* 53.6, p. 1637.

Hernandez, R D, J L Kelley, E Elyashiv, and S C Melton (2011). "Classic selective sweeps were rare in recent human evolution". In: *Science*.

Howie, James Malcolm, Rupert Mazzucco, Thomas Taus, Viola Nolte, and Christian Schlotterer (2018). "DNA motifs are not general predictors of recombination in two Drosophila sister species". en.

Jónás, Á, T Taus, C Kosiol, and C Schlötterer (2016). "Estimating the Effective Population Size from Temporal Allele Frequency Changes in Experimental Evolution". In: *Genetics*.

Kelly, John K and Kimberly A Hughes (2019). "Pervasive Linked Selection and Intermediate-Frequency Alleles Are Implicated in an Evolve-and-Resequencing Experiment of Drosophila simulans". en. In: *Genetics* 211.3, pp. 943–961.

Kolaczkowski, Bryan, Andrew D Kern, Alisha K Holloway, and David J Begun (2011). "Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster". en. In: *Genetics* 187.1, pp. 245–260.

Machado, Heather E et al. (2018). "Broad geographic sampling reveals predictable and pervasive seasonal adaptation in Drosophila". en.

Macpherson, J Michael, Guy Sella, Jerel C Davis, and Dmitri A Petrov (2007). "Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in Drosophila". en. In: *Genetics* 177.4, pp. 2083–2099.

Marchini, Marta et al. (2014). "Impacts of genetic correlation on the independent evolution of body mass and skeletal size in mammals". en. In: *BMC Evol. Biol.* 14, p. 258.

Maynard Smith, John and John Haigh (1974). "The hitch-hiking effect of a favourable gene". In: *Genet. Res.* 23.01, pp. 23–35.

McVicker, Graham, David Gordon, Colleen Davis, and Phil Green (2009). "Widespread Genomic Signatures of Natural Selection in Hominid Evolution". In: *PLoS Genet.* 5.5, e1000471–16.

Nei, Masatoshi and Fumio Tajima (1981). "Genetic Drift And Estimation Of Effective Population Size". In: *Genetics* 98.3, pp. 625–640.

Nordborg, Magnus, Brian Charlesworth, and Deborah Charlesworth (1996). "The effect of recombination on background selection*". In: *Genet. Res.* 67.02, pp. 159–174.

Nordborg, Magnus et al. (2005). "The Pattern of Polymorphism in Arabidopsis thaliana". In: *PLoS Biol.* 3.7, e196–11.

Nuzhdin, Sergey V and Thomas L Turner (2013). "Promises and limitations of hitchhiking mapping". en. In: *Curr. Opin. Genet. Dev.* 23.6, pp. 694–699.

Reznick, David N, Frank H Shaw, F Helen Rodd, and Ruth G Shaw (1997). "Evaluation of the Rate of Evolution in Natural Populations of Guppies (Poecilia reticulata)". In: *Science* 275.5308, pp. 1934–1937.

Robertson, Alan (1961). "Inbreeding in artificial selection programmes". In: *Genet. Res.* 2.02, pp. 189–194.

Santiago, Enrique and Armando Caballero (1995). "Effective size of populations under selection". In: *Genetics* 139.2, pp. 1013–1030.
— (1998). "Effective Size and Polymorphism of Linked Neutral Loci in Populations Under Directional Selection". In: *Genetics* 149.4, pp. 2105–2117.
Sattath, Shmuel, Eyal Elyashiv, Oren Kolodny, Yosef Rinott, and Guy Sella (2011). "Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in Drosophila simulans". en. In: *PLoS Genet.* 7.2, e1001302.
Sella, Guy, Dmitri A Petrov, Molly Przeworski, and Peter Andolfatto (2009). "Pervasive natural selection in the Drosophila genome?" en. In: *PLoS Genet.* 5.6, e1000495.
Signor, Sarah A, Felicia N New, and Sergey Nuzhdin (2018). "A Large Panel of Drosophila simulans Reveals an Abundance of Common Variants". en. In: *Genome Biol. Evol.* 10.1, pp. 189–206.
Therkildsen, Nina O et al. (2019). "Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing". en. In: *Science* 365.6452, pp. 487–490.
Turner, T L and P M Miller (2012). "Investigating Natural Variation in Drosophila Courtship Song by the Evolve and Resequence Approach". In: *Genetics* 191.2, pp. 633–642.
Turner, Thomas L, Andrew D Stewart, Andrew T Fields, William R Rice, and Aaron M Tarone (2011). "Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila melanogaster". In: *PLoS Genet.* 7.3, e1001336–10.
Waples, R S (1989). "A generalized approach for estimating effective population size from temporal changes in allele frequency". In: *Genetics* 121.2, pp. 379–391.
Wasserman, Larry (2006). *All of Nonparametric Statistics*. en. Springer Science & Business Media.
Williamson, Robert J et al. (2014). "Evidence for widespread positive and negative selection in coding and conserved noncoding regions of Capsella grandiflora". en. In: *PLoS Genet.* 10.9, e1004622.
Woolliams, J A, N R Wray, and R Thompson (1993). "Prediction of long-term contributions and inbreeding in populations undergoing mass selection". In: *Genet. Res.* 62.3, pp. 231–242.
Wray, Naomi R and Robin Thompson (1990). "Prediction of rates of inbreeding in selected populations". In: *Genet. Res.* 55.01, pp. 41–54.

Supplementary Material

1.1 Estimator Bias Correction

1.1.1 Correcting variance bias with a single depth sampling process

Following Waples (1989), we have that the variance in allele frequency change at a locus in the initial generation, which is entirely due to the binomial sampling process, is $\mathrm{Var}(p_0) = {}^{p_0(1-p_0)}/_{d_0}$ where d_0 is the number of binomial draws (e.g. read depth). At a later timepoint, the variance in allele frequency is a result of both the binomial sampling process at time t and the evolutionary process. Using the law of total variation we can partition the variation from each process,

$$\mathrm{Var}(\widetilde{p_t}) = \mathbb{E}(\mathrm{Var}(\widetilde{p_t}|p_t)) + \mathrm{Var}(\mathbb{E}(\widetilde{p_t}|p_t)) \tag{3}$$

$$= \underbrace{\frac{p_t(1-p_t)}{d_t}}_{\text{generation } t \text{ sampling noise}} + \underbrace{\mathrm{Var}(p_t)}_{\text{variance due to evolutionary process}} . \tag{4}$$

Under a drift-only process, $\mathrm{Var}(p_t) = p_0(1-p_0)\left[1-\left(1-\frac{1}{2N}\right)^t\right]$. However, with heritable variation in fitness, we need to consider the covariance in allele frequency changes across generations (Buffalo and Coop 2019). We can write

$$\mathrm{Var}(p_t) = \mathrm{Var}\left(p_0 + (p_1 - p_0) + (p_2 - p_1) + \ldots + (p_t - p_{t-1})\right) \tag{5}$$

$$= \mathrm{Var}\left(p_0 + \Delta p_0 + \Delta p_1 + \ldots + \Delta p_{t-1}\right) \tag{6}$$

$$= \mathrm{Var}(p_0) + \sum_{i=0}^{t-1}\mathrm{Cov}(p_0, \Delta p_i) + \sum_{i=0}^{t-1}\mathrm{Var}(\Delta p_i) + \sum_{0\le i<j}^{t-1}\mathrm{Cov}(\Delta p_i, \Delta p_j). \tag{7}$$

Each allele frequency change is equally like to be positive as it is to be negative; thus by symmetry this second term is zero. Additionally $\mathrm{Var}(p_0) = 0$, as we treat p_0 as a fixed initial frequency. We can write,

$$\mathrm{Var}(p_t) = \sum_{i=0}^{t-1}\mathrm{Var}(\Delta p_i) + \sum_{0\le i<j}^{t-1}\mathrm{Cov}(\Delta p_i, \Delta p_j). \tag{8}$$

The second term, the cumulative impact of variance in allele frequency change can be partitioned into heritable fitness and drift components (Buffalo and Coop 2019; Santiago and Caballero 1995)

$$\mathrm{Var}(p_t) = \sum_{i=0}^{t-1}\mathrm{Var}(\Delta_{\scriptscriptstyle D} p_i) + \sum_{i=0}^{t-1}\mathrm{Var}(\Delta_{\scriptscriptstyle H} p_i) + \sum_{0\le i<j}^{t-1}\mathrm{Cov}(\Delta p_i, \Delta p_j). \tag{9}$$

where $\Delta_{\scriptscriptstyle H} p_t$ and $\Delta_{\scriptscriptstyle D} p_t$ indicate the allele frequency change due to heritable fitness variation and drift respectively. Then, sum of drift variances in allele frequency change is

$$\sum_{i=0}^{t-1}\mathrm{Var}(\Delta_{\scriptscriptstyle D} p_i) = \sum_{i=0}^{t-1}\frac{p_i(1-p_i)}{2N} \tag{10}$$

replacing the heterozygosity in generation i with its expectation, we have

$$\sum_{i=0}^{t-1}\mathrm{Var}(\Delta_{\scriptscriptstyle D} p_i) = p_0(1-p_0)\sum_{i=0}^{t-1}\frac{1}{2N}\left(1-\frac{1}{2N}\right)^i \tag{11}$$

$$= p_0(1-p_0)\left[1-\left(1-\frac{1}{2N}\right)^t\right] \tag{12}$$

which is the usual variance in allele frequency change due to drift. Then, the total allele frequency change from generations 0 to t is $\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = \mathrm{Var}(\widetilde{p}_t) + \mathrm{Var}(\widetilde{p}_0) - 2\,\mathrm{Cov}(\widetilde{p}_t, \widetilde{p}_0)$, where the covariance depends on the nature of the sampling plan (see Nei and Tajima 1981; Waples 1989). In the case where there is heritable variation for fitness, and using the fact that $\mathrm{Cov}(\widetilde{p}_t, \widetilde{p}_0) = {}^{p_0(1-p_0)}/_{2N}$ for Plan I sampling procedures (Waples 1989), we write,

$$\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = \mathrm{Var}(\widetilde{p}_t) + \mathrm{Var}(\widetilde{p}_0) - 2C\,\mathrm{Cov}(\widetilde{p}_t, \widetilde{p}_0) \tag{13}$$

$$= \frac{p_t(1-p_t)}{d_t} + \frac{p_0(1-p_0)}{d_0} + p_0(1-p_0)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] + \tag{14}$$

$$\sum_{i=0}^{t-1} \mathrm{Var}(\Delta_H p_i) + \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j) - \frac{C p_0(1-p_0)}{2N} \tag{15}$$

$$\frac{\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0)}{p_0(1-p_0)} = 1 + \frac{p_t(1-p_t)}{p_0(1-p_0)d_t} + \frac{1}{d_0} - \left(1 - \frac{1}{2N}\right)^t + \tag{16}$$

$$\sum_{i=0}^{t-1} \frac{\mathrm{Var}(\Delta_H p_i)}{p_0(1-p_0)} + \sum_{0 \le i < j}^{t-1} \frac{\mathrm{Cov}(\Delta p_i, \Delta p_j)}{p_0(1-p_0)} - \frac{C}{N} \tag{17}$$

where $C = 1$ if Plan I is used, and $C = 0$ if Plan II is used (see Waples 1989, p. 380 and Figure 1 for a description of these sampling procedures; throughout the paper we use sampling Plan II). Rearranging, we can create a bias-corrected estimator for the population variance in allele frequency change, and replace all population heterozygosity terms with the unbiased sample estimators, e.g. $\frac{d_t}{d_t - 1}\widetilde{p}_t(1-\widetilde{p}_t)$,

$$\frac{d_0 - 1}{d_0}\frac{\mathrm{Var}(\widetilde{p}_1 - \widetilde{p}_0)}{\widetilde{p}_0(1-\widetilde{p}_0)} - \frac{(d_0 - 1)}{d_0(d_1 - 1)}\frac{\widetilde{p}_1(1-\widetilde{p}_1)}{\widetilde{p}_0(1-\widetilde{p}_0)} - \frac{1}{d_0} + \frac{C}{N} = \frac{\mathrm{Var}(\Delta_H p_0)}{p_0(1-p_0)} + \frac{1}{2N} \tag{18}$$

1.1.2 Correcting variance bias with individual and depth sampling processes

Here, we extend the sampling bias correction described above to handle two binomial sampling processes: one as individuals are binomially sampled from the population, and another as reads are binomially sampled during sequencing. (see also Jónás et al. 2016). Let $X_t \sim \mathrm{Binom}(n_t, p_t)$ where X_t is the count of alleles and n_t is the number of diploids sampled at time t. Then, these individuals are sequenced at a depth of d_t, and $Y_t \sim \mathrm{Binom}(d_t, {}^{X_t}/_{n_t})$ reads have the tracked allele. We let $\widetilde{p}_t = {}^{Y_t}/_{d_t}$ be the observed sample allele frequency. Then, the sampling noise is

$$\mathrm{Var}(\widetilde{p}_t | p_t) = \mathbb{E}(\mathrm{Var}(\widetilde{p}_t | X_t)) + \mathrm{Var}(\mathbb{E}(\widetilde{p}_t | X_t)) \tag{19}$$

$$= p_t(1-p_t)\left(\frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t}\right) \tag{20}$$

$$\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = p_t(1-p_t)\left(\frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t}\right) + p_0(1-p_0)\left(\frac{1}{n_0} + \frac{1}{d_0} - \frac{1}{n_0 d_0}\right) \tag{21}$$

$$- \frac{Cp_0(1-p_0)}{N} + p_0(1-p_0)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] + \sum_{i=0}^{t-1} \mathrm{Var}(\Delta_H p_i) \tag{22}$$

$$+ \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j) \tag{23}$$

Through the law of total expectation (see Kolaczkowski et al. 2011 Supplementary File 1 for a sample proof), one can find that an unbiased estimator of the half the heterozygosity is

$$\frac{n_t d_t}{(n_t - 1)(d_t - 1)} \widetilde{p}_t(1 - \widetilde{p}_t). \tag{24}$$

Replacing this unbiased estimator for half of the heterozygosity into our expression above, the total sample variance is

$$\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = \frac{n_t d_t \widetilde{p}_t(1-\widetilde{p}_t)}{(n_t-1)(d_t-1)}\left(\frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t}\right) + \frac{n_0 d_0 \widetilde{p}_0(1-\widetilde{p}_0)}{(n_0-1)(d_0-1)}\left(\frac{1}{n_0} + \frac{1}{d_0} - \frac{1}{n_0 d_0}\right) + \tag{25}$$

$$\frac{n_0 d_0 \widetilde{p}_0(1-\widetilde{p}_0)}{(n_0-1)(d_0-1)}\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] - \frac{C}{N}\frac{n_0 d_0 \widetilde{p}_0(1-\widetilde{p}_0)}{(n_0-1)(d_0-1)} +$$

$$\sum_{i=0}^{t-1} \mathrm{Var}(\Delta_H p_i) + \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j). \tag{26}$$

As with equation (18), we can rearrange this to get a biased-corrected estimate of the variance in allele frequency change between adjacent generations, $\mathrm{Var}(\Delta p_t)$.

1.1.3 Covariance Correction

We also need to apply a bias correction to the temporal covariances (and possibly the replicate covariances if the initial sample frequencies are all shared). The basic issue is that $\mathrm{Cov}(\Delta\widetilde{p}_t, \Delta\widetilde{p}_{t+1}) = \mathrm{Cov}(\widetilde{p}_{t+1} - \widetilde{p}_t, \widetilde{p}_{t+2} - \widetilde{p}_{t+1})$, and thus shares the sampling noise of timepoint $t+1$. Thus acts to bias the covariance by subtracting off the noise variance term of $\mathrm{Var}(\widetilde{p}_{t+1})$, so we add the expectation of this bias, derived above, back in. We discuss this in more detail below in deriving the bias correction for the temporal-replicate variance covariance matrix.

1.1.4 Temporal-Replicate Covariance Matrix Correction

In practice, we simultaneously estimate the temporal and replicate covariance matrices for each replicate, which we call the temporal-replicate covariance matrix. This needs a bias correction; we extend the bias corrections for single locus variance and covariance described in Supplementary

Material Sections 1.1.1, 1.1.2, and 1.1.3 to multiple sampled loci and the temporal-replicate covariance matrix here. With frequency data collected at $T+1$ timepoints across R replicate populations at L loci, we have multidimensional arrays $\mathbf{F}$ of allele frequencies, $\mathbf{D}$ of sequencing depths, and $\mathbf{N}$ of the number of individuals sequenced, each of dimension $R \times (T+1) \times L$. We calculate the array $\mathbf{\Delta F}$ which contains the allele frequency changes between adjacent generations, and has dimension $R \times T \times L$. The operation flat($\mathbf{\Delta F}$) flattens this array to a $(R \cdot T) \times L$ matrix, such that rows are grouped by replicate, e.g. for timepoint t, replicate r, and locus l such that for allele frequencies $p_{t,r,l}$, the frequency change entries are

$$\text{flat}(\mathbf{\Delta F}) = \begin{bmatrix} \Delta p_{1,0,0} & \Delta p_{2,0,0} & \dots & \Delta p_{1,1,0} & \Delta p_{2,1,0} & \dots & \Delta p_{T,R,0} \\ \Delta p_{1,0,1} & \Delta p_{2,0,1} & \dots & \Delta p_{1,1,1} & \Delta p_{2,1,1} & \dots & \Delta p_{T,R,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Delta p_{1,0,L} & \Delta p_{2,0,L} & \dots & \Delta p_{1,1,L} & \Delta p_{2,1,L} & \dots & \Delta p_{T,R,L} \end{bmatrix} \tag{27}$$

where each $\Delta p_{t,r,l} = p_{t+1,r,l} - p_{t,r,l}$. Then, the sample temporal-replicate covariance matrix $\mathbf{Q}'$ calculated on flat($\mathbf{\Delta F}$) is a $(R \cdot T) \times (R \cdot T)$ matrix, with the R temporal-covariance block submatrices along the diagonal, and the $R(R-1)$ replicate-covariance submatrices matrices in the upper and lower triangles of the matrix,

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{Q}'_{1,1} & \mathbf{Q}'_{1,2} & \dots & \mathbf{Q}'_{1,R} \\ \mathbf{Q}'_{2,1} & \mathbf{Q}'_{2,2} & \dots & \mathbf{Q}'_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}'_{R,1} & \mathbf{Q}'_{R,2} & \dots & \mathbf{Q}'_{R,R} \end{bmatrix} \tag{28}$$

where each submatrix $\mathbf{Q}'_{i,j}$ $(i \neq j)$ is the $T \times T$ sample replicate covariance matrix for replicates i and j, and the submatrices along the diagonal $\mathbf{Q}'_{r,r}$ are the temporal covariance matrices for replicate r.

Given the bias of the sample covariance of allele frequency changes, we calculated an expected bias matrix $\mathbf{B}$, averaging over loci,

$$\mathbf{B} = \frac{1}{L} \sum_{l=1}^{L} \frac{\mathbf{h}_l}{2} \circ \left(\frac{1}{\mathbf{d}_l} + \frac{1}{2\mathbf{n}_l} + \frac{1}{2\mathbf{d}_l \circ \mathbf{n}_l} \right) \tag{29}$$

where $\circ$ denotes elementwise product, and $\mathbf{h}_l$, $\mathbf{d}_l$, and $\mathbf{n}_l$, are rows corresponding to locus l of the unbiased heterozygosity arrays $\mathbf{H}$, depth matrix $\mathbf{D}$, and number of diploids matrix $\mathbf{N}$. The unbiased $R \times (T+1) \times L$ heterozygosity array can be calculated as

$$\mathbf{H} = \frac{2\mathbf{D} \circ \mathbf{N}}{(\mathbf{D}-1) \circ (\mathbf{N}-1)} \circ \mathbf{F} \circ (1-\mathbf{F}) \tag{30}$$

where division here is elementwise. Thus, $\mathbf{B}$ is a $R \times (T+1)$ matrix. As explained in Supplementary Material Section 1.1.2 and 1.1.3, the temporal variances and covariances require bias corrections, meaning each temporal covariance submatrix $\mathbf{Q}_{r,r}$ requires two corrections. For an element $Q_{r,t,s} =$

$\mathrm{Cov}(\Delta p_t, \Delta p_s)$ of the temporal covariance submatrix for replicate r, $\mathbf{Q}_{r,r}$, we apply the following correction

$$Q_{r,t,s} = \begin{cases} Q'_{r,t,s} - b_{r,t} - b_{r,t+1}, & \text{if } t = s \\ Q'_{r,t,s} + b_{r,\max(t,s)}, & \text{if } |t-s| = 1 \end{cases} \tag{31}$$

where $b_{r,t}$ is element in row r and column t of $\mathbf{B}$.

1.1.5 Barghi et al. (2019) Temporal Covariances

Since each replicate population was sequenced every ten generations, the timepoints $t_0 = 0$ generations, $t_1 = 10$ generations, $t_2 = 20$ generations, etc., lead to observed allele frequency changes across ten generation blocks, $\Delta p_{t_0}, \Delta p_{t_1}, \ldots, \Delta p_{t_6}$. Consequently, the ten temporal covariance matrices for each of the ten replicate populations have off-diagonal elements of the form $\mathrm{Cov}(\Delta p_{t_0}, \Delta p_{t_1}) = \mathrm{Cov}(p_{t_1} - p_{t_0}, p_{t_2} - p_{t_1}) = \sum_{i=0}^{10} \sum_{j=10}^{20} \mathrm{Cov}(\Delta p_i, \Delta p_j)$. Each diagonal element has the form $\mathrm{Var}(\Delta p_{t_0}) = \sum_{i=0}^{t_0} \mathrm{Var}(\Delta p_i) + \sum_{i \neq j}^{t_0} \mathrm{Cov}(\Delta p_i, \Delta p_j)$, and is thus a combination of the effects of drift and selection, as both the variance in allele frequency changes and cumulative temporal autocovariances terms increase the variance in allele frequency. With sampling each generation, one could more accurately partition the total variance in allele frequency change (Buffalo and Coop 2019); while we cannot directly estimate the contribution of linked selection to the variance in allele frequency change here, the presence of a positive observed covariance between allele frequency change can only be caused linked selection.

1.1.6 Block Bootstrap Procedure

To infer the uncertainty of covariance, convergence correlation, and $G(t)$ estimates, we used a block bootstrap procedure. This is a version of the bootstrap that resamples blocks of data points, rather than individual data points, to infer the uncertainty of an statistic in the presence of unknown correlation structure between data. With genome-wide data, linkage disequilibria between sites creates complex and unknown dependencies between variants. The estimators used in this paper are predominantly ratios, e.g. temporal-replicate covariance standardized by half the heterozygosity, $G(t)$ which is the ratio of covariance to total variance, and the convergence correlation (equation (2)). In these cases, we can exploit the linearity of the expectation to make the bootstrap procedure more computationally efficient, by pre-calculating the statistics of the ratio's numerator and denominator, $N(\mathbf{x}_i)$ and $D(\mathbf{x}_i)$, on the data $\mathbf{x}_i$ for all blocks $i \in \{1, 2, \ldots, W\}$ in the genome. Then we draw W bootstrap samples with replacement, and compute the estimate for bootstrap sample b with an average weighted by the fraction w_i of total loci contained in each block,

$$\tilde{\theta}_b = \frac{\sum_{i=1}^{W} w_i N(\mathbf{x}_i)}{\sum_{i=1}^{W} w_i D(\mathbf{x}_i)} \tag{32}$$

Note that computing the ratio of averages rather than the average of a ratio is a practice common for population genetic statistics like F_{ST} (Bhatia et al. 2013). With these B bootstrap estimates, we calculate the $\alpha/2$ and $1-\alpha/2$ quantiles, which we use to estimate the $1-\alpha = 95\%$ pivot confidence intervals (p. 33 Wasserman 2006, p. 194 Davison and Hinkley 2013) throughout the paper,

$$C_\alpha = \left(2\widehat{\theta} - q_{1-\alpha/2}, 2\widehat{\theta} - q_{\alpha/2}\right). \tag{33}$$

where $\widehat{\theta}$ is the estimate, and q_x is bootstrap quantile for probability x.

1.1.7 Replicate G and Partitioning the Variance in Allele Frequency

We define a statistic similar to G for estimating the proportion of allele frequency change common between two replicate populations due to linked selection. Covariance in allele frequency change between two replicate populations is due to convergent selection pressure selecting haplotypes shared between the two replicate populations, which acts to perturb linked neutral variation in parallel way.

$$G_R(t) = \frac{\mathbb{E}_{A\neq B}(\sum_{i\neq j}^{t} \mathrm{Cov}(\Delta p_{i,A}, \Delta p_{j,B}))}{\mathbb{E}_R(\mathrm{Var}(p_{t,R} - p_{0,R}))} \tag{34}$$

where $\mathbb{E}_{A\neq B}$ indicates that the expectation is taken over all ordered pairs of replicates (e.g. summing all off-diagonal elements replicate covariances), and $\mathbb{E}_R$ indicates taking expectation over all replicates. This measures the fraction of variance in allele frequency change (averaged across replicates) due to shared selection pressure.

Extending our theoretic work in Buffalo and Coop (2019), we can partition the allele frequency change in two replicates into drift, and shared selection and replicate-specific selection components of allele frequency change. For two replicates, A and B,

$$\Delta p_{t,A} = \Delta_{D} p_{t,A} + \Delta_{U} p_{t,A} + \Delta_{S} p_t \tag{35}$$

$$\Delta p_{t,B} = \Delta_{D} p_{t,B} + \Delta_{U} p_{t,B} + \Delta_{S} p_t \tag{36}$$

where $\Delta_{D} p_{t,A}$ is allele frequency change due to drift (this is specific to a replicate, and equal to $\Delta_{N} p_{t,A} + \Delta_{M} p_{t,A}$ in the notation of Buffalo and Coop 2019), $\Delta_{U} p_{t,A}$ is the allele frequency change from indirect selection specific to replicate A (and likewise with $\Delta_{U} p_{t,A}$ for replicate B), and $\Delta_{S} p_t$ is the allele frequency change from indirect selection shared across the replicates A and B (this term lacks a replicate subscript since by construction it is identical between replicates). By construction, each of these terms is uncorrelated, so the variance and be written as:

$$\mathrm{Var}(\Delta p_{t,A}) = \mathrm{Var}(\Delta_{D} p_{t,A}) + \mathrm{Var}(\Delta_{U} p_{t,A}) + \mathrm{Var}(\Delta_{S} p_t) \tag{37}$$

$$\tag{38}$$

The shared effects of indirect selection can be quantified from the observed allele frequency changes, since the covariance in allele frequency change across replicates is the covariance of the shared term by construction,

$$\mathrm{Cov}(\Delta p_{t,A}, \Delta p_{t,B}) = \mathrm{Cov}(\Delta_{S} p_t, \Delta_{S} p_t) = \mathrm{Var}(\Delta_{S} p_t) \tag{39}$$

In artificial selection studies with a control (non-selected) line, such as the Castro et al. (2019) study, this allows us to estimate the contribution of the effects of shared and unique indirect selection. In the case of this study, we can estimate the drift, unique selection effect, and shared selection effect terms using the fact that,

$$\Delta p_{t,\mathrm{LS1}} = \Delta_{D} p_{t,\mathrm{LS1}} + \Delta_{U} p_{t,\mathrm{LS1}} + \Delta_{\mathrm{LS}} p_t \tag{40}$$

$$\Delta p_{t,\mathrm{LS2}} = \Delta_{D} p_{t,\mathrm{LS2}} + \Delta_{U} p_{t,\mathrm{LS2}} + \Delta_{\mathrm{LS}} p_t \tag{41}$$

$$\Delta p_{t,\mathrm{Ctrl}} = \Delta_{D} p_{t,\mathrm{Ctrl}}. \tag{42}$$

Note that since the control replicate does not undergo artificial selection, we assume that its allele frequency changes are determined entirely by genetic drift. With free mating individuals (such as in a cage population), this may not be the case, and sequencing adjacent generations would allow one to differentiate the effects of selection and drift.

We assume that we can approximate the contribution of genetic drift in the Longshanks selection lines with the observed variance in the control line, or $\mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}}) = \mathrm{Var}(\Delta_{D p_{t,\mathrm{LS1}}}) = \mathrm{Var}(\Delta_{D p_{t,\mathrm{LS2}}})$. Then, the combined effects of selection can be estimated by averaging the variances of the two Longshanks selection lines, and subtracting the variance in allele frequency change in the control line, which we treat as driven by drift alone (since matings are random). Note that each variance is bias-corrected according to the methods described in Supplementary Materials 1.1.4, and the average sequencing depths between lines are nearly identical. Thus, we have

$$(\mathrm{Var}(\Delta p_{t,\mathrm{LS1}}) + \mathrm{Var}(\Delta p_{t,\mathrm{LS2}}))/2 - \mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}}) = \overline{\mathrm{Var}(\Delta_{U} p_{t,\mathrm{LS}})} + \mathrm{Var}(\Delta_{\mathrm{LS}} p_t) \tag{43}$$

where the bar indicates values averaged both Longshanks selection lines. Additionally, use the fact that

$$\mathrm{Cov}(\Delta p_{t,\mathrm{LS1}}, \Delta p_{t,\mathrm{LS2}}) = \mathrm{Var}(\Delta_{\mathrm{LS}} p_t) \tag{44}$$

we can also separate out the unique and shared components by subtracting off this covariance,

$$\overline{\mathrm{Var}(\Delta_{U} p_{t,\mathrm{LS}})} = (\mathrm{Var}(\Delta p_{t,\mathrm{LS1}}) + \mathrm{Var}(\Delta p_{t,\mathrm{LS2}}))/2 - \mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}}) - \mathrm{Cov}(\Delta p_{t,\mathrm{LS1}}, \Delta p_{t,\mathrm{LS2}}). \tag{45}$$

Finally, we can divide each of these values by the total variance to get the proportion of total variance drift, and unique and shared effects of selection contribute towards the total. To derive confidence intervals for the estimates of unique and shared effects of selection, we use a block bootstrap procedure as described in Supplementary Materials Section 1.1.6.

1.1.8 The Empirical Neutral Null Windowed Covariance Distribution

To detect an excess of genomic regions with unusually high or low covariances, we need to compare the distribution of observed windowed covariances to a null distribution of windowed covariances that we would expect under no selection. While we could construct a theoretic sampling distribution of the spurious covariances created by neutral genetic drift at particular site, the unknown linkage

disequilibrium between sites would mean that this is not an adequate null model for the distribution of windowed covariances in our data.

To address this limitation, we construct a neutral null model by sign-permuting the observed allele frequency changes. This destroys the covariances built up by selection, mimicking a neutral allele's frequency trajectory. This approach is conservative, since selection also acts to increase the magnitude of allele frequency changes (see equation 1 of Buffalo and Coop 2019), but this magnitude is not affected by the sign-permutation procedure. Consequently, the resulting empirical null distribution has higher variance than would be expected under neutrality alone.

Still, we wanted to ensure that LD between sign-permuted blocks, which will affect the variance of the empirical null distribution, does not impact our primary finding that the distribution of temporal covariances becomes increasingly negative in the Barghi et al. (2019) dataset through time. To address this, we also sign-permuted at the whole chromosome level finding we recapitulate the same pattern (Supplementary Figure S8).

Supplementary Figures

1.1.9 Bias Correction for Barghi et al. (2019)

We have investigated the effectiveness of our correction on real data by exploiting the relationship between sampling depth and the magnitude of the variance and covariance biases, and comparing the observed variances and covariances before and after correction. We plot the variance and covariance (between adjacent timepoints) before and after the bias correction against the average sample depth in 100kb genomic windows in Figure S1. Overall, we find the biased-correction procedure removes the relationship between variance and covariance and depth, indicating it is working adequately.

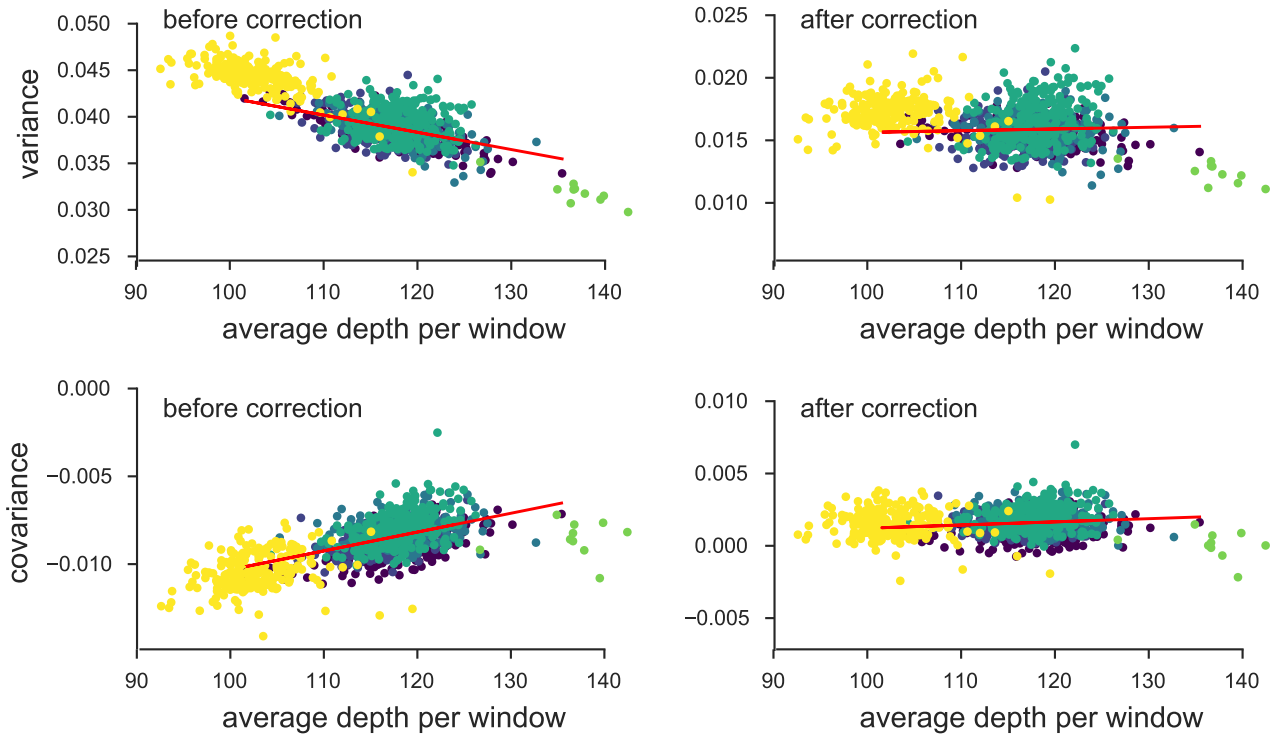

Figure S1: The variance and covariances from the Barghi et al. (2019) study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. Overall, the bias correction corrects sampling bias in both the variance and covariance such that the relationship with depth is constant. Colors indicate the different chromosomes of *D. simulans*; we have excluded the X chromosome (yellow points) and chromosome 4 points (green points to far right) from the regression due to large differences in average coverage.

1.2 Barghi et al. (2019) Trimmed Window Covariances

Here we report median and trimmed mean of the windowed covariances. We note that the median covariance is also limiting result of a trimmed mean that symmetrically excludes the upper and

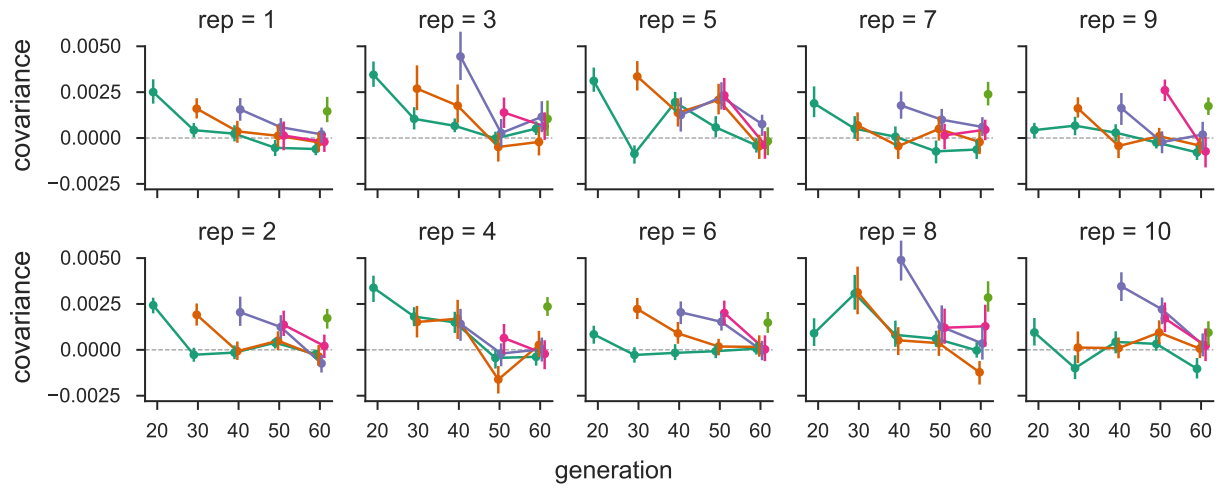

Figure S2: The temporal covariances from the Barghi et al. (2019) study, for each replicate individually. As in Figure 1, each line follows the temporal covariances from some initial reference generation through time, which represent the rows of temporal covariance matrix.

lower α tails to calculate the trimmed average windowed covariance. As α increases to 0.5, the trimmed covariance converges to the median windowed covariance (by the definition of the median; see Supplementary Figure S3). Thus our genomic temporal covariances are non-zero due to the impact of selection on many genomic windows.

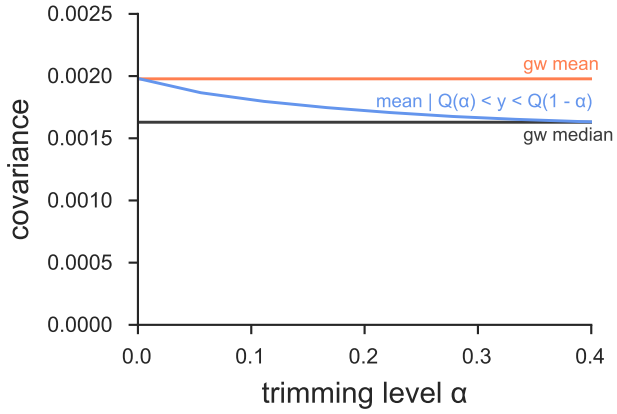

Figure S3: The genome-wide covariance ($\mathrm{Cov}(\Delta p_0, \Delta p_1 0)$ pooling all replicates) averaged (red line) and the median windowed covariance (blue) for the Barghi et al. (2019) dataset. The trimmed average window covariance, excluding the α lower and upper tails, converges to the median windowed covariance. This indicates that genome-wide covariance are not being overly dominated by a large-effect loci in few windows.

s	t	median	median 95% CI	trimmed mean	trimmed mean 95% CI
0	10	1.629	$[1.532, 1.738]$	1.874	$[1.777, 1.969]$
0	20	0.371	$[0.276, 0.465]$	0.491	$[0.403, 0.585]$
0	30	0.479	$[0.4, 0.589]$	0.516	$[0.434, 0.602]$
0	40	0.059	$[-0.012, 0.15]$	0.027	$[-0.05, 0.099]$
0	50	-0.204	$[-0.271, -0.125]$	-0.259	$[-0.329, -0.187]$
10	20	1.549	$[1.427, 1.659]$	1.722	$[1.617, 1.83]$
10	30	0.438	$[0.339, 0.539]$	0.506	$[0.399, 0.609]$
10	40	0.233	$[0.149, 0.328]$	0.254	$[0.159, 0.343]$
10	50	-0.355	$[-0.454, -0.289]$	-0.319	$[-0.401, -0.237]$
20	30	1.981	$[1.856, 2.095]$	2.195	$[2.084, 2.302]$
20	40	0.792	$[0.698, 0.894]$	0.903	$[0.815, 0.999]$
20	50	0.123	$[0.042, 0.207]$	0.221	$[0.141, 0.309]$
30	40	1.296	$[1.208, 1.425]$	1.385	$[1.287, 1.483]$
30	50	0.07	$[-0.037, 0.183]$	0.116	$[0.023, 0.21]$
40	50	1.36	$[1.271, 1.446]$	1.513	$[1.427, 1.601]$

Table S1: Table of median of windowed covariance estimates ($\mathrm{Cov}(\Delta p_s, \Delta p_t) \times 100$) between generations t and s and the trimmed mean windowed covariance which excludes the lower and upper 5% windows with the highest covariance.

1.2.1 Barghi et al. (2019) Empirical Null and Windowed Covariance Distributions

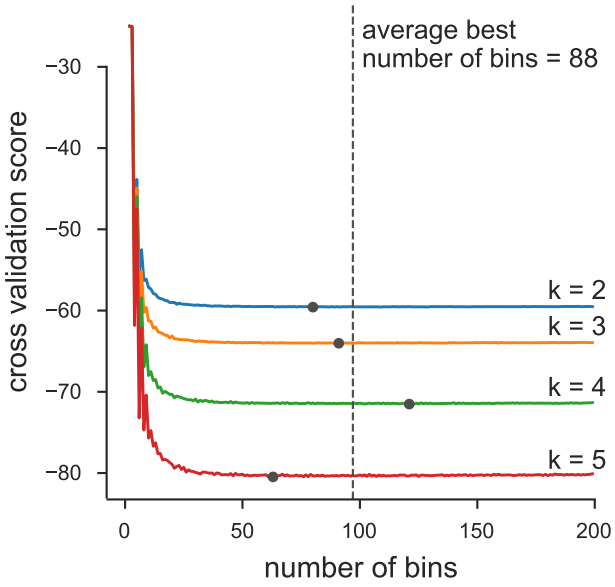

Figure S4: We chose number of bins used in the histograms of Figure 3 via an analytic expression for the cross-validation risk, based on the equation 6.16 of (Wasserman 2006, p. 129). Above, we plot the cross-validation risk for various numbers of bins, for each of the four off-diagonals of the temporal covariance matrix that we analyze. Overall, because the number of data points is large, oversmoothing is less of a problem, leading the cross-validation risk to be relatively flat across a large number of bins. Each gray point indicates the minimal risk for a particular off-diagonal, and the dashed line indicates the best average binwidth across off-diagonals.

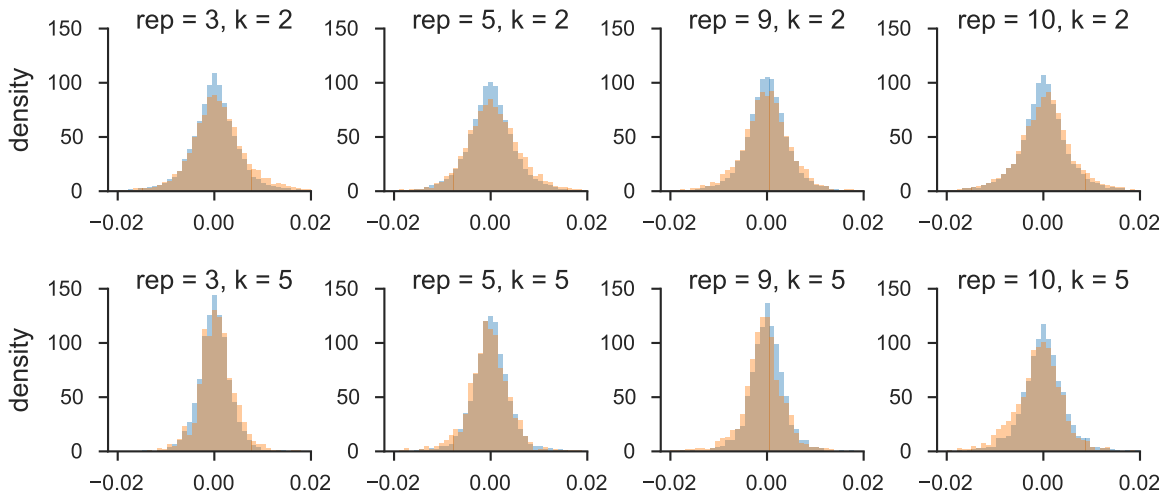

Figure S5: The distribution of windowed temporal covariances alongside the empirical neutral null for five randomly sampled replicates (columns), for $k = 2$ (first row) and $k = 5$ (second row). The main figure of the paper pools all replicate window and empirical neutral null covariances; we show here the windowed temporal covariances tend to shift from being positive (a heavier right tail) to become more negative (a heavier left tail) through time within particular replicates.

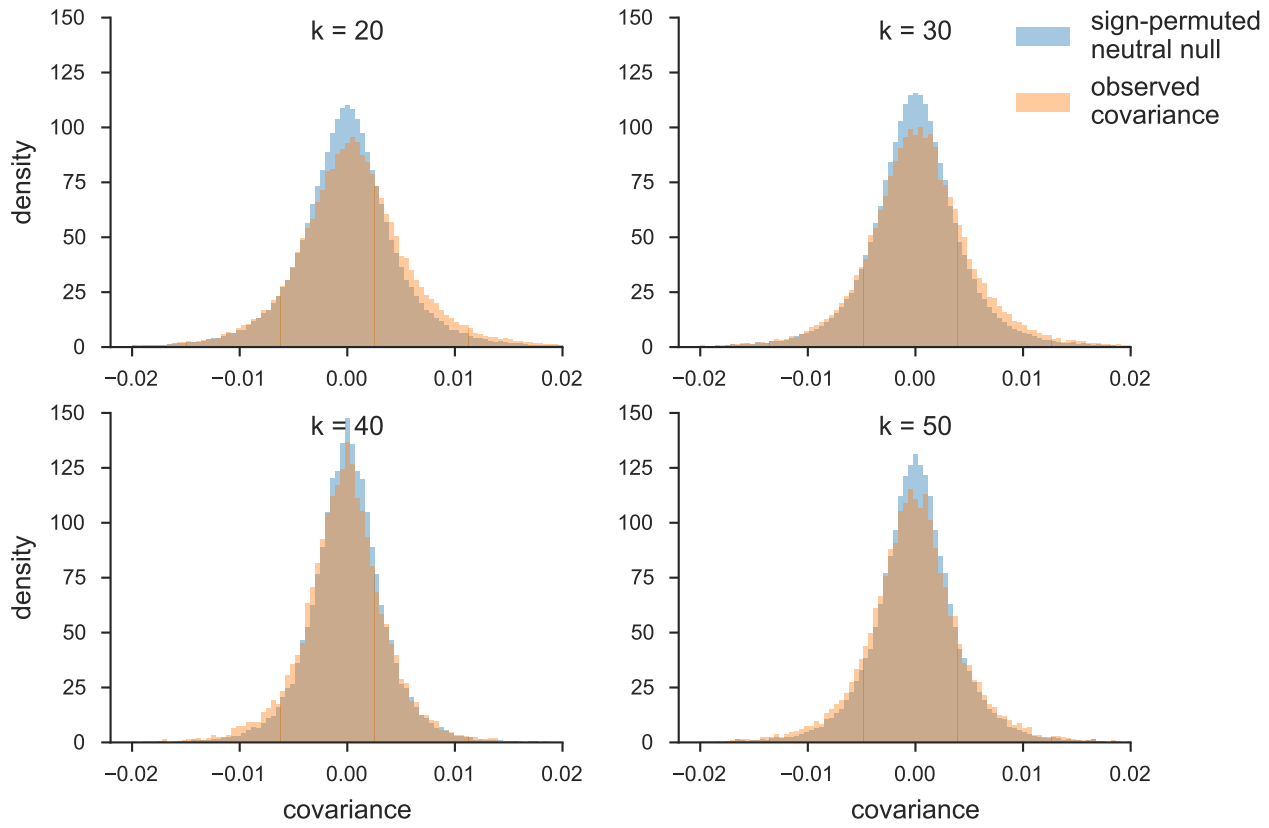

Figure S6: The distribution of temporal covariances calculated across 100kb genomic windows from Barghi et al. (2019)'s study (orange) and the block sign permuted empirical neutral null distribution of the windowed covariances (blue). Each panel shows these windowed covariances and the empirical null distribution for covariances $\mathrm{Cov}(\Delta p_t, \Delta p_{t+k})$, k is the number of generations between allele frequency changes.

1.2.2 Barghi et al. (2019) Tail Probabilities for Windowed Covariances Distributions

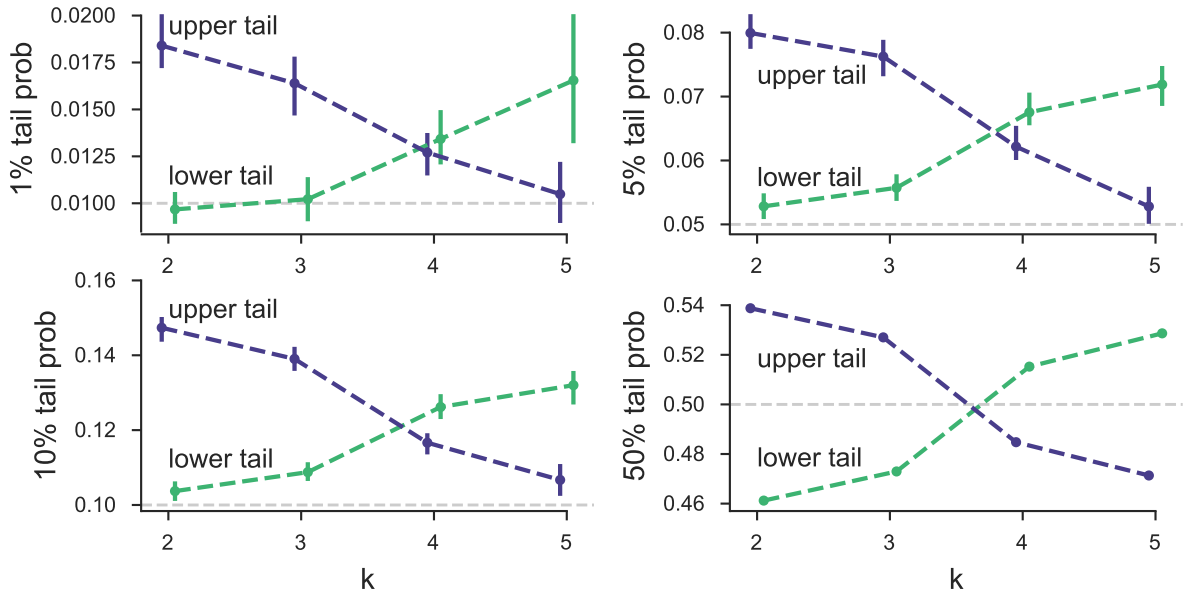

Figure S7: Barghi et al. (2019) tail probabilities compared to sign-permuted empirical null distribution for various α levels.

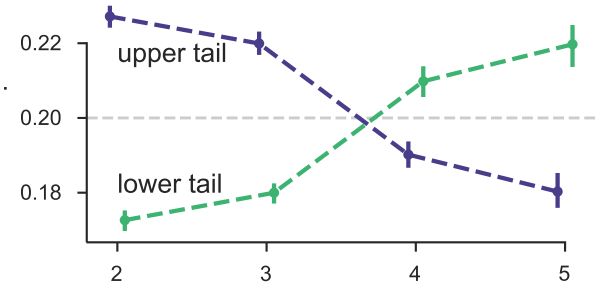

Figure S8: The 20% lower and upper tail probabilities for the observed windowed covariances from the Barghi et al. (2019) study, based on sign-permuting at the chromosome level. This permutation empirical null is robust to long-range linkage disequilibrium acting over entire chromosomes (see Supplementary Material section 1.1.8).

1.2.3 Bergland et al. (2014) Re-Analysis

We also applied our temporal covariance approach to Bergland et al. (2014), which found evidence of genome-wide fluctuating selection between Spring and Fall seasons across three years *Drosophila melanogaster*. As described in Buffalo and Coop (2019), if fluctuating selection pressure among time-periods are the dominant genome-wide pattern, we might expect positive covariances between

like seasons changes (e.g. Spring 2010 to Fall 2010 and Spring 2011 to Fall 2011), and negative covariances between dislike seasonal changes (e.g. Fall 2009 to Spring 2010 and Fall 2010 to Spring 2011). However, while we find temporal covariances that are non-zero, we find only weak support for a seasonal fluctuating model driving these covariances. In Supplementary Figure S9, we show the temporal covariances from varying reference generations, across seasonal transitions that are alike (e.g. the covariance between the allele frequency changes between Fall 2009 and Spring 2009, and frequency changes between Fall 2010 and Spring 2010), and dislike (e.g. the covariance between the allele frequency change between Fall 2009 and Spring 2009, and the frequency changes between Spring 2010 and Fall 2009). The first row of temporal covariance matrix is consistent with fluctuating selection operating for two timepoints, as the first covariance is negative, and the second is positive, and later covariances are not statistically differentiable from zero (which could occur if LD and additive genetic variance decay). However, all other temporal covariances do not fit the pattern we would expect under genome-wide fluctuating selection.

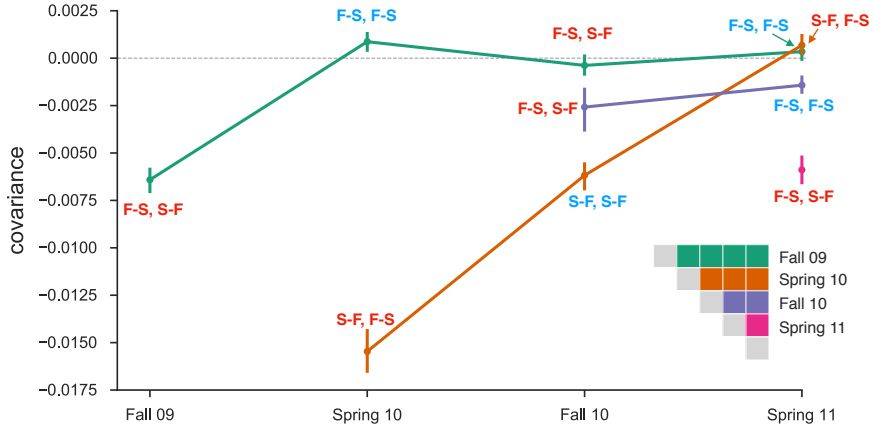

Figure S9: Temporal covariances from the Bergland et al. (2014) study, from varying reference generations (e.g. rows along the temporal covariance matrix). Each covariance is labeled indicating whether the covariance is between two like seasonal transitions (e.g. the covariance between allele frequency changes from fall to spring in one year, and fall to spring in another) or two dislike seasons (e.g. the covariance between fall to spring in one year, and spring to fall in another year). Covariances between like transitions are expected to be positive when there is a genome-wide effect of fluctuating selection (and these labels are colored blue), while covariances between dislike transitions are expected to be negative (and these labels are colored red). 95% confidence intervals were constructed by a block-bootstrapping procedure where the blocks are megabase tiles.

We wanted to establish that our temporal-covariance matrix bias correction was working correctly. We find that it corrects the relationship between depth and both variance and covariance (Supplementary Figure S12) as expected.

It is unclear how strong the fluctuations would have to be to generate a genome-wide average signal of fluctuating selection from temporal covariances. For example, many loci could still show a signal of fluctuating selection, but the average signal could be overwhelmed by other signals of other selection. To investigate whether there was a genome-wide excess of loci showing evidence

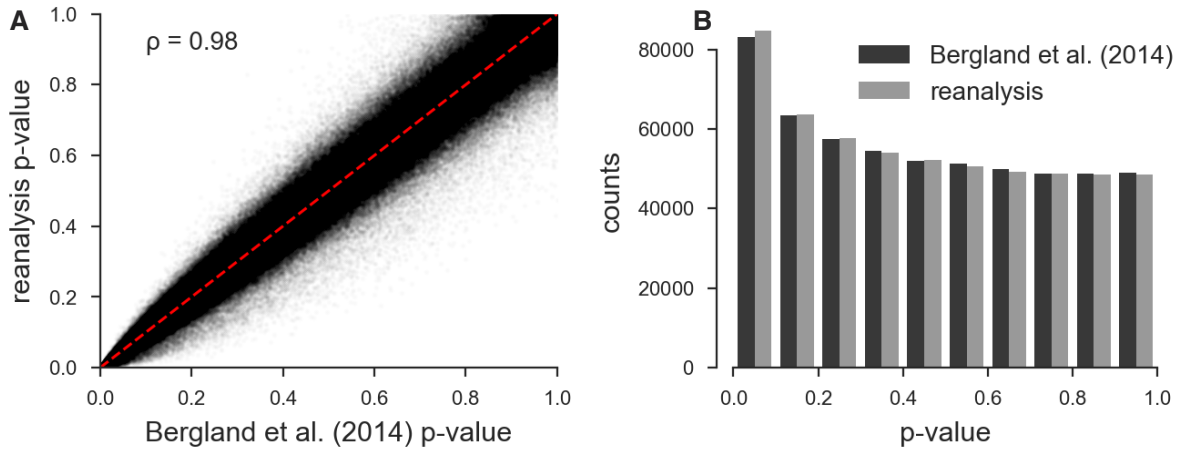

Figure S10: **A**: Scatterplot of the original unadjusted p-values from Bergland et al. (2014) and the p-values from our reanalysis of the same data using the same statistical methods; the minor discrepancy is likely due to software version differences. **B**: The histograms of the p-values of our reanalysis and the original Bergland et al. (2014) data; again the minor discrepancy is likely due to software differences. Overall, our implementation of Bergland et al.'s statistical methods produces results very close to the original analysis.

of fluctuating selection we reanalyzed the data of (Bergland et al. 2014) using the same seasonal fluctuating model as the original paper. This model is a Binomial logit-linked GLM fit per-locus, where the frequencies are regressed on the Spring/Fall seasons are encoded as a dummy variable. We use the same binomial weighting procedure as Bergland et al. (2014), where the weights are determined by the effective number of chromosomes, $N_{eff} = (2n_t d_t - 1)/(2n_t + d_t)$ (n_t and d_t are the number of diploid individuals and the read depth at timepoint t, respectively). We fit this model on all loci marked as used in the VCF provided with the Bergland et al. (2014) study (doi:10.5061/dryad.v883p). Overall, our p-values for the Wald test for each locus closely match those of the original paper (Pearson correlation coefficient 0.98, p-value $< 2.2 \times 10^{-16}$; see Supplementary Figure S10 A), and the histograms of the p-values are nearly identical (Supplementary Figure S10 B). Bergland et al. (2014) find loci with a significant association with season after a Benjamini and Hochberg FDR p-value adjustment (Benjamini and Hochberg 1995), however, the null hypothesis of the Wald test does not give us an idea of the expected number of variants that may spuriously fit the pattern of seasonal fluctuating selection as it does not account for genetic drift or other forms of hitchhiking.

To investigate whether there is a genome-wide evidence of an enrichment of fluctuating selection we created an empirical null distribution by randomly permuting the season labels and re-running the per-locus seasonal GLM model, as proposed by Machado et al. (2018). We find, regardless of whether we permute at the locus-level or the permutation replicate-level, that the observed seasonal p-value distribution Bergland et al. (2014) is not enriched for significant p-values beyond what we would expect from the permutation null. In fact, there appears there is more enrichment for low p-values when seasonal labels are randomly permuted (Supplementary Figure S11, suggesting by random chance we might expect more variants with a seasonal fluctuating pattern than found in the original Bergland et al. (2014) study. While surprising, this could be explained by the presence of temporal structure across the samples not consistent with seasonal fluctuating selection. Some

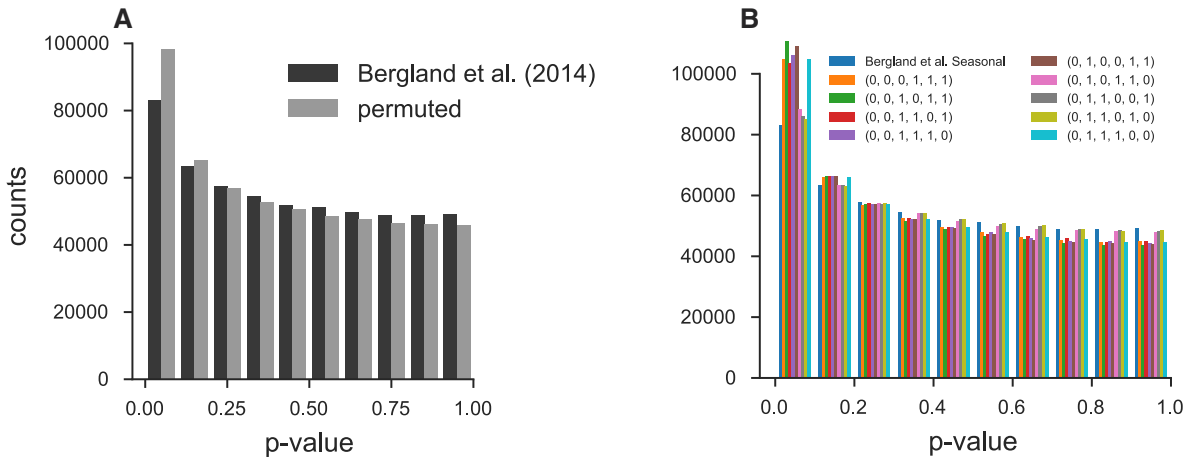

Figure S11: **A**: Histogram of original Bergland et al. (2014) seasonal p-values and p-values creating by randomly permuting the seasons at each locus. **B**: Histogram of original Bergland et al. (2014) p-values alongside all unique permutations (ignoring symmetries that lead to identical p-values).

fraction of the permutations happen to fit this structure well, leading to an enrichment of small p-values. This non-seasonal temporal structure is also evident in our temporal covariances (Supplementary Figure S9), where we see strong evidence of selection (non-zero temporal covariances), yet the pattern does not follow that of seasonal fluctuating selection.

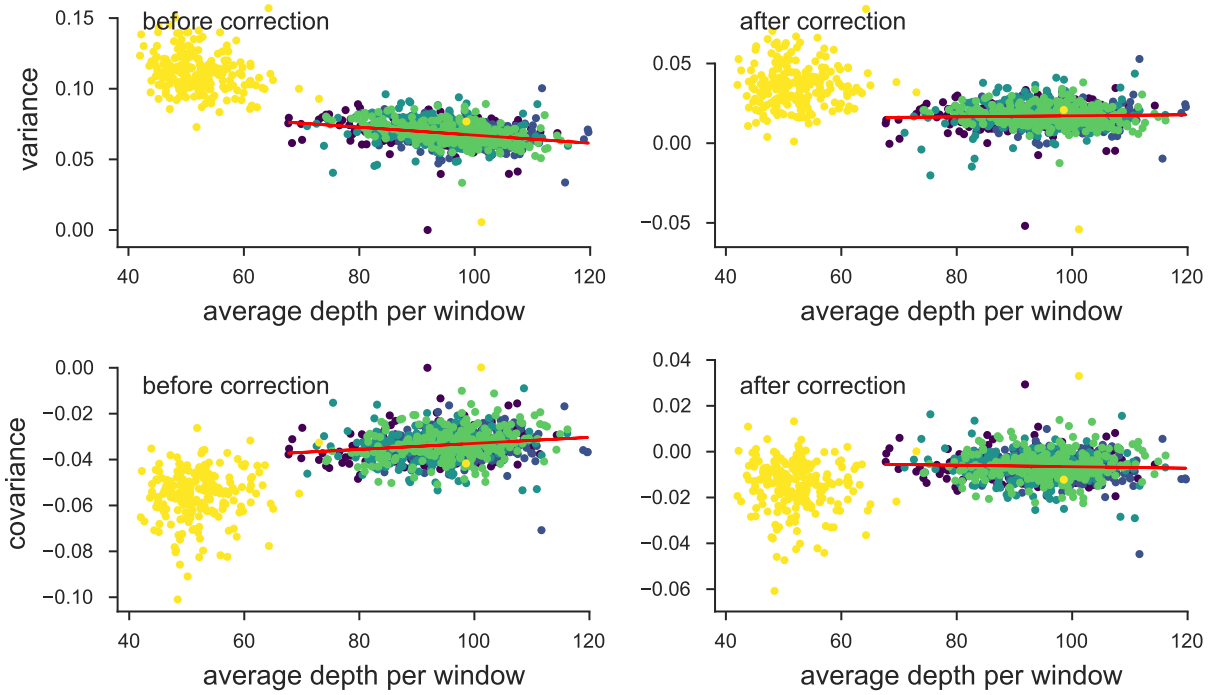

Figure S12: The variance and covariances from the Bergland et al. (2014) study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. The bias correction procedure is correcting sampling bias in both the variance and covariance such that the relationship with depth is constant. Colors indicate the different chromosomes of *D. melanogaster*; we have excluded the X chromosome (yellow points; chromosome 4 was not in the original study) from the regression due to large differences in average coverage.