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

September 4, 2019

Abstract

1 Introduction

A long-standing problem in evolutionary genetics is quantifying the roles drift and selection have in shaping genome-wide allele frequency changes. Previous work on this question has focused on teasing apart the impacts of drift and selection on genome-wide diversity using population samples from a single contemporary timepoint, usually by modeling the correlation between regional recombination rate and diversity created in the presence of linked selection (XXX). While this approach has dramatically informed our present view of the role of linked selection in shaping genome-wide diversity across a variety of organisms (XXX), and has allowed us to quantity the relative roles positive selection (hitchhiking) and negative selection (background selection) have shaping patterns of genome-wide diversity (XXX), we fully lack an understanding of how linked selection acts over an important dimension: time.

Presently, there are numerous examples of phenotypes rapidly evolving (XXX), yet the polygenic nature of fitness makes detecting the impact selection has on genome-wide variation over short timescales remarkably difficult. This is because the effect of selection on a polygenic trait (such as fitness) is distributed across loci proportional to their effect sizes, leading to subtle allele frequency shifts that are difficult to distinguish from background levels of genetic drift and sampling variance. However, increasingly genomic studies with multiple timepoints and, in some cases, multiple replicate populations are becoming available, used to detect selected loci (XXX) and differentiate models of selection (XXX). In a previous paper, we proposed that one signal of polygenic linked selection that could be detected from these temporal genomic studies is the temporal autocovariance in allele frequency changes (Buffalo and Coop 2019). These covariances are directly estimable from temporal genomic data and are created only in the presence of linked selection, but not genetic drift in a closed population.

Here, we provide the first empirical analyses 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 one 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 of the proportion of

total variation in allele frequency change caused by linked selection, and the correlation between allele frequency changes between replicate populations caused by 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.

2 Results

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

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 a bias-corrected genome-wide 6×6 temporal covariance matrix $\mathbf{Q}$ for each of the ten replicates (see Supplementary Material Sections 5.2 and 5.4 for details on the bias correction). Since the Barghi et al. (2019) study sequenced each replicate population every ten generations, we calculate the temporal covariance matrix over ten-generation allele frequency changes, which we write as $\mathrm{Cov}(\Delta_{10}p_s, \Delta_{10}p_t)$. The diagonal elements of this temporal covariance matrix are the sum of all variances in allele frequency changes and pairwise covariances, e.g. element $\mathbf{Q}_{t,t} = \sum_{i=t}^{9} \mathrm{Var}(\Delta p_i) + 2\sum_{i\neq j} \mathrm{Cov}(\Delta p_i, \Delta p_j)$. Each off-diagonal element $\mathbf{Q}_{t,s},\ t \neq s$ contains the sum of pairwise covariances, $\mathbf{Q}_{t,s} = \sum_{i=t}^{t+9}\sum_{j=s}^{s+9} \mathrm{Cov}(\Delta p_i, \Delta p_j)$ (see Supplementary Material Section 5.5 for more details). These off-diagonal covariances are expected to be zero when only drift is acting in the population, as only linked selection can create covariance between allele frequency changes (Buffalo and Coop 2019). Furthermore, since we cannot directly observe the temporal covariances between adjacent timepoints, but only the sums of temporal covariances, we expect the off-diagonal elements of $\mathbf{Q}$ to be weak. This is because the magnitude of a temporal covariance $\mathrm{Cov}(\Delta p_i, \Delta p_j)$ is governed by the level of additive genetic variance and linkage disequilibrium that persists after $|i-j|$ generations of decay, which in Barghi et al. (2019)'s study design is on average 10 generations.

Still, averaging across the ten temporal covariances matrices, we find positive temporal covariances across time that are statistically significant (95% block bootstraps for all first off-diagonal elements do not contain zero), consistent with linked selection perturbing allele frequency changes over very short time periods. We visualize these covariances in Figure (A), which depicts the temporal covariances through time, for each of the five off-diagonal rows temporal covariance matrix. Each row represents the temporal covariance $\mathrm{Cov}(\Delta_{10}p_s, \Delta_{10}p_t)$, between some initial reference generation s (the row of the matrix), and some later timepoint t (the column of the matrix). For each row, the covariances at first are positive, and then decay towards zero as expected when directional selection affects linked variants' frequency trajectories until ultimately linkage disequilibrium and additive genetic variance for decay (Buffalo and Coop 2019). Note that per replicate, the signal is a bit noisier; see Supplementary Figures XXX.

While the presence of positive temporal covariances is consistent with linked selection affecting allele frequencies over time, this measure not easily interpretable. Additionally, we can quantify

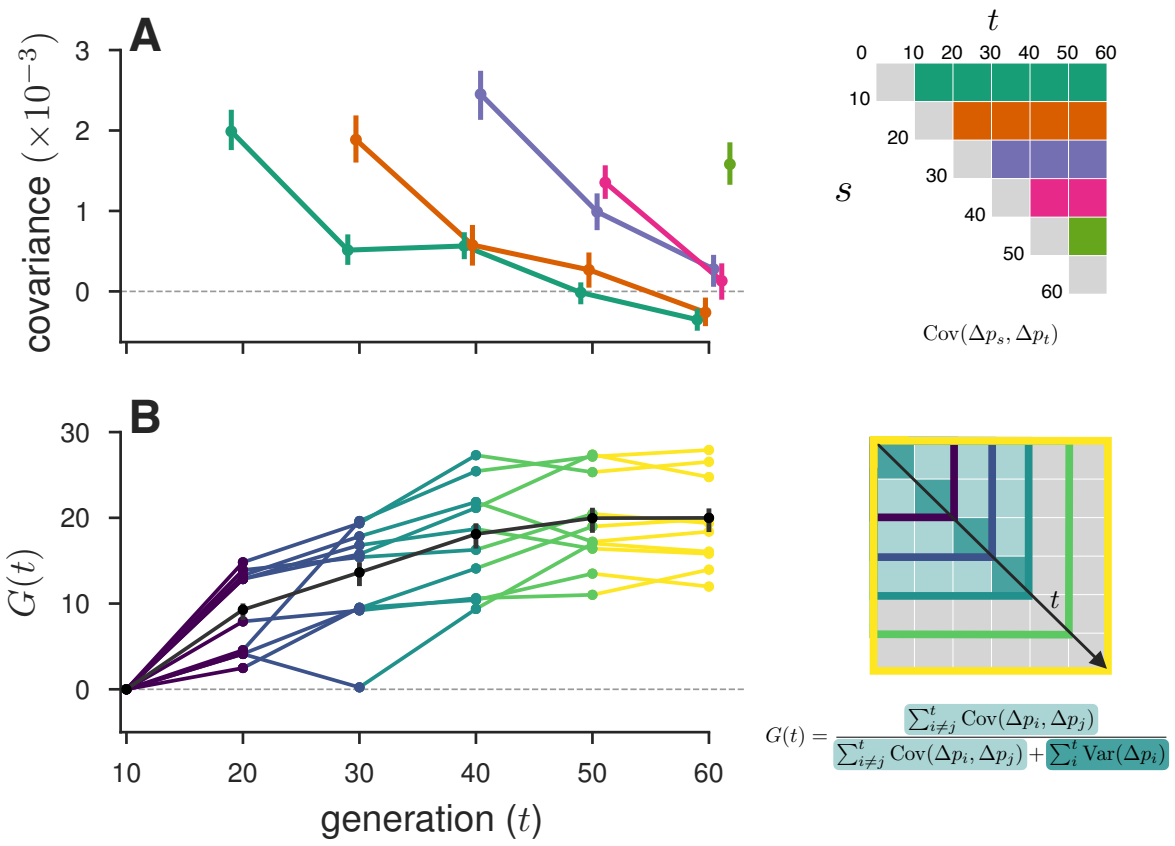

Figure 1: A: Temporal covariance, averaged across all ten replicate populations, through time from the Barghi et al. (2019) study. Each line depicts the temporal covariance $\mathrm{Cov}(\Delta p_s, \Delta p_t)$ from some reference generation s to a later time t which varies along the x-axis; each line 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 $G(t)$.

the impact of linked selection on allele frequency change as the ratio of total covariance in allele frequency change to the total variance in allele frequency change. Since the total variation in allele frequency change can be decomposed into variance and covariance components, $\mathrm{Var}(p_t - p_0) = \sum_{i \neq j} \mathrm{Cov}(\Delta p_i, \Delta p_j) / \mathrm{Var}(p_t - p_0)$, 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 $G(t)$, which is a lower bound on the total effect of linked selection between the initial generation 0 and some later generation t, which can be varied to see how this quantity grows through time. As with the temporal covariances, the study design of **Barghi2019** leads our measure $G(t)$ to be even more conservative, since the temporal covariances with each ten-generation block between sequenced timepoints are not directly observable, and are not included in the numerator of $G(t)$. Still, we find a remarkably strong signal that greater than

20% of total variation in allele frequency change over 60 generations is directly the result of linked selection.

The replicate design of Barghi et al. (2019) also allows us to quantify another covariance: the covariance in allele frequency change between replicate populations experiencing convergent selection pressure. These between-replicate covariances are created in the same way as temporal covariances are: neutral alleles linked to a particular fitness background experience are expected to have allele frequency changes in the same direction if the selection pressures are similar. We measure this through a statistic similar to a correlation, which we call the convergent correlation; this is the ratio of average between-replicate covariance across all pairs to the average standard deviation across all pairs of replicates,

$$\mathrm{cor}(\Delta p_s, \Delta p_t) = \frac{\mathbb{E}_{A\neq B}\left(\mathrm{Cov}(\Delta p_{s,A}, \Delta p_{t,B})\right)}{\mathbb{E}_{A\neq B}\left(\sqrt{\mathrm{Var}(\Delta p_{s,A})\,\mathrm{Var}(\Delta p_{t,B})}\right)} \tag{1}$$

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, which unlike temporal covariance matrices have diagonal elements that are also covariances (e.g. $\mathrm{Cov}(\Delta_{10}p_{t,A}, \Delta_{10}p_{t,B})$). Like temporal covariances, we visualize these through time (Figure 2 A), 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 2 A). We find these decay very quickly, from an initial convergence correlation coefficient 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.

A benefit of between-replicate covariances is unlike temporal covariances, these can be calculated with only two sequenced timepoints and a replicated study design. This allowed 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 and Hughes (2019), which evolved three replicate populations of *Drosophila simulans* for 14 generations in 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 statistical significant (95% CIs, [0.0063, 0.0086], [0.0064, 0.0086], [0.0061, 0.0083]). We estimate the convergent correlation coefficient across these replicates as 0.36 (95% block-bootstrap confidence interval $[0.31, 0.40]$).

Second, we reanalyzed the Longshanks selection experiment, which selected for longer tibiae length relative to body size in mice, leading to a response of selection of about 5 standard deviations over the course of twenty generations (Castro et al. 2019). 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 the two Longshanks lines. We find that this is the case (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 line is statistically significantly different from zero (CIs XXX).

Paragraph about large effect loci.

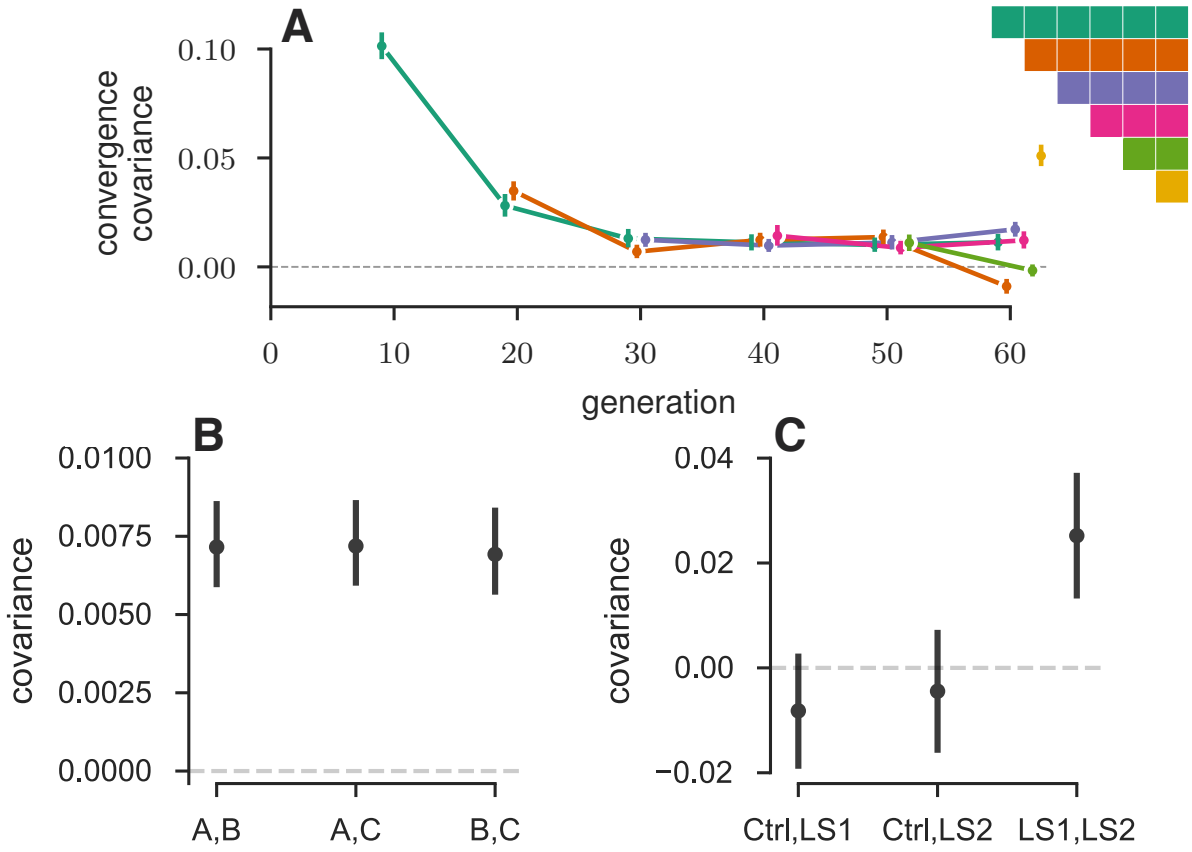

Figure 2: A: The convergence correlation, averaged across replicate pairs, through time. Each line represents the convergence correlation $\text{cor}(\Delta p_s, \Delta p_s)$ from a starting reference generation s to a later time t, which varies along the x-axis; each line corresponds to a row of the temporal convergence correlation matrix depicted to the right.

Finally, we observed that in the longest run evolve-and-resequence study we analyzed (Barghi et al. 2019), some temporal covariances appeared to become negative at future timepoints (see the first two rows in Figure 1 A). This could be caused by advantageous fitness backgrounds later becoming disadvantageous, either due to frequency-dependent selection, recombination breaking up coupled beneficial and deleterious alleles, or an environmental change in the direction of selection. However, the signal of a change in the direction of selection localized to a particular region of the genome may be washed out when we calculate genome-wide temporal covariances. Averaging over such large numbers of loci is necessary to detect a signal of covariance in allele frequency changes when both drift (inflated in evolve-and-sequence studies which use small population sizes) and sampling variance can create spurious covariances. To address this limitation, we calculated temporal covariances over 100kb genomic windows (mean loci per window 4,606, median 4,830). While the covariance of each tile is noisy, we can analyze the genome-wide distribution of tile covariances for changes through time. We would like to compare the distribution of these tile covariances to a neutral expectation, but using a theoretic neutral null distribution would be inappropriate given complex dependencies across the genome 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).

Instead, we have developed a permutation-based procedure that constructs a null distribution of the genome-wide covariances as they would look if only neutral genetic drift was acting. We do this by randomly flipping the signs of the allele frequency changes per-genomic window 1,000 times, which acts to destroy the systematic covariances created by linked selection but creates a sampling distribution of the covariances spuriously created by neutral genetic drift. This empirical neutral null distribution is conservative in the sense that the variances of the covariances are wider than expected under drift alone, since drift also acts to inflate the magnitude of allele frequency change. By flipping the sign at the block-level, we preserve the complex dependencies between adjacent loci created by linkage disequilibrium; see Supplementary Material Section 5.7 for more details on this approach.

Using the Barghi et al. (2019) study's data, we used the empirical neutral null to investigate whether there was a general trend for the temporal covariance $\mathrm{Cov}(\Delta p_t, \Delta p_{t+k})$ to be positive for small k (as expected under selection), but then become systematically negative for larger k (as would occur under evolutionary processes that reverse the fitness of regions across the genome). Such a change in the fitness direction of haplotypes would not be discernible from genome-wide covariances like those in Figure 1. We find (Figure 3 A and B), pooling across all replicates (see Supplementary Figure 7 for individuals replicates), 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 additionally quantified the degree to which the left and right tails change as we look at covariances between timepoints further apart (Figure 3 C). We find that for covariances 2 timepoints apart ($k = 2 \times 10$ generations), the 20% right tail is inflated by a of 1.3 (95% CI [1.29, 1.32]), consistent with linked selection creating positive covariance in allele frequency change. Under constant directional selection, we expect these positive covariances between increasingly distant timepoints to decay to zero as additive genetic variance is exhausted by selection and linkage disequilibrium decays. Instead we find that for covariances 4 timepoints apart ($k = 4 \times 10$ generations), the left tail of negative covariances is inflated by a factor of 1.19 (95% CI [1.17, 1.21]). 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 9). Overall, we see a striking pattern that covariances not only decay towards zero, but in fact become negative through time, consistent with some regions in the genome having had a reversed fitness effect at later timepoints.

3 Acknowledgments

Doc Edge.

4 Appendix

5 Estimator Bias Correction

5.1 Correcting variance bias with a single depth sampling process

Following Waples (1989), we have that that the variance 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,

$$\begin{aligned} \mathrm{Var}(\widetilde{p_t}) &= \mathbb{E}(\mathrm{Var}(\widetilde{p_t}|p_t)) + \mathrm{Var}(\mathbb{E}(\widetilde{p_t}|p_t)) && (2) \\ &= \underbrace{\frac{p_t(1-p_t)}{d_t}}_{\text{generation } t \text{ sampling noise}} + \underbrace{\mathrm{Var}(p_t)}_{\text{variance due to evolutionary process}} && (3) \end{aligned}$$

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

$$\begin{aligned} V(p_t) &= V\left(p_0 + (p_1 - p_0) + (p_2 - p_1) + \ldots + (p_t - p_{t-1})\right) && (4) \\ &= V\left(p_0 + \Delta p_0 + \Delta p_1 + \ldots + \Delta p_{t-1}\right) && (5) \\ &= V(p_0) + \sum_{i=0}^{t-1} \mathrm{Cov}(p_0, \Delta p_i) + \sum_{i=0}^{t-1} \mathrm{Var}(\Delta p_i) + \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j). && (6) \end{aligned}$$

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 $V(p_0) = 0$, as we treat p_0 as a fixed initial frequency. We can write,

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

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)

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

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

$$\sum_{i=0}^{t-1} \mathrm{Var}(\Delta_{\scriptscriptstyle D} p_i) = \sum_{i=0}^{t-1} \frac{p_i(1-p_i)}{2N} \quad (9)$$

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

$$\sum_{i=0}^{t-1} \mathrm{Var}(\Delta_D p_i) = p_0(1-p_0) \sum_{i=0}^{t-1} \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^i \tag{10}$$

$$= p_0(1-p_0) \left[1 - \left(1 - \frac{1}{2N}\right)^t\right] \tag{11}$$

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,

$$\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = \mathrm{Var}(\widetilde{p}_t) + \mathrm{Var}(\widetilde{p}_0) - 2C\,\mathrm{Cov}(\widetilde{p}_t, \widetilde{p}_0) \tag{12}$$

$$= \frac{p_t(1-p_t)}{d_t} + \frac{p_0(1-p_0)}{d_0} + p_0(1-p_0)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] + \tag{13}$$

$$\sum_{i=0}^{t-1} \mathrm{Var}(\Delta_H p_i) + \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j) - \frac{C p_0(1-p_0)}{2N} \tag{14}$$

$$\frac{\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0)}{p_0(1-p_0)} = 1 + \frac{p_t(1-p_t)}{p_0(1-p_0)d_t} + \frac{1}{d_0} - \left(1 - \frac{1}{2N}\right)^t + \tag{15}$$

$$\sum_{i=0}^{t-1} \frac{\mathrm{Var}(\Delta_H p_i)}{p_0(1-p_0)} + \sum_{0 \le i < j}^{t-1} \frac{\mathrm{Cov}(\Delta p_i, \Delta p_j)}{p_0(1-p_0)} - \frac{C}{N} \tag{16}$$

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). We move terms creating a corrected estimator for the population variance in allele frequency change, and replace all population heterozygosity terms with the unbiased sample estimators, e.g. $\frac{d_t}{d_t - 1}\widetilde{p}_t(1-\widetilde{p}_t)$,

$$\frac{d_0 - 1}{d_0} \frac{\mathrm{Var}(\widetilde{p}_1 - \widetilde{p}_0)}{\widetilde{p}_0(1-\widetilde{p}_0)} - \frac{(d_0 - 1)}{d_0(d_1 - 1)} \frac{\widetilde{p}_1(1-\widetilde{p}_1)}{\widetilde{p}_0(1-\widetilde{p}_0)} - \frac{1}{d_0} + \frac{C}{N} = \frac{\mathrm{Var}(\Delta_H p_0)}{p_0(1-p_0)} + \frac{1}{2N} \tag{17}$$

5.2 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

$$\mathrm{Var}(\widetilde{p}_t | p_t) = \mathbb{E}(\mathrm{Var}(\widetilde{p}_t | X_t)) + \mathrm{Var}(\mathbb{E}(\widetilde{p}_t | X_t)) \tag{18}$$

$$= p_t(1-p_t)\left(\frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t}\right) \tag{19}$$

$$\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = p_t(1-p_t)\left(\frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t}\right) + p_0(1-p_0)\left(\frac{1}{n_0} + \frac{1}{d_0} - \frac{1}{n_0 d_0}\right) \tag{20}$$

$$- \frac{C p_0(1-p_0)}{N} + p_0(1-p_0)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] + \sum_{i=0}^{t-1} \mathrm{Var}(\Delta_H p_i) \tag{21}$$

$$+ \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j) \tag{22}$$

Through the law of total expectation, one can find that an unbiased estimator of the heterozygosity is

$$\frac{n_t d_t}{(n_t - 1)(d_t - 1)} \widetilde{p}_t(1 - \widetilde{p}_t). \tag{23}$$

Replacing this unbiased heterozygosity into our expression above, the total sample variance is

$$\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = \frac{n_t d_t \widetilde{p}_t(1-\widetilde{p}_t)}{(n_t-1)(d_t-1)}\left(\frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t}\right) + \frac{n_0 d_0 \widetilde{p}_0(1-\widetilde{p}_0)}{(n_0-1)(d_0-1)}\left(\frac{1}{n_0} + \frac{1}{d_0} - \frac{1}{n_0 d_0}\right) + \tag{24}$$

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

$$\sum_{i=0}^{t-1} \mathrm{Var}(\Delta_H p_i) + \sum_{0 \le i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j). \tag{25}$$

As with equation (17), we can rearrange this to get a biased-correct estimate of the variance in allele frequency change between adjacent generations, $\mathrm{Var}(\Delta p_t)$.

5.3 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 that back in.

5.4 Variance-Covariance Matrix Correction

Now, we extend the bias corrections for single locus variance and covariance described in Supplementary Material Sections 5.1, 5.2, and 5.3 to multiple sampled loci. With frequency collected at $T+1$ timepoints across R replicate populations at L loci, we have $\mathbf{F}$ of allele frequencies, $\mathbf{D}$ multidimensional array of sequencing depths, and a $\mathbf{N}$ multidimensional array 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 and allele frequencies $p_{t,r}$, for a single locus the entries are

$$\text{flat}(\mathbf{\Delta F}) = \begin{bmatrix} \Delta p_{1,0,0} & \Delta p_{2,0,0} & \dots & \Delta p_{1,1,0} & \Delta p_{2,1,0} & \dots & \Delta p_{T,R,0} \\ \Delta p_{1,0,1} & \Delta p_{2,0,1} & \dots & \Delta p_{1,1,1} & \Delta p_{2,1,1} & \dots & \Delta p_{T,R,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Delta p_{1,0,L} & \Delta p_{2,0,L} & \dots & \Delta p_{1,1,L} & \Delta p_{2,1,L} & \dots & \Delta p_{T,R,L} \end{bmatrix} \tag{26}$$

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,

$$\mathbf{Q}' = \begin{bmatrix} \mathbf{Q}'_{1,1} & \mathbf{Q}'_{1,2} & \dots & \mathbf{Q}'_{1,R} \\ \mathbf{Q}'_{2,1} & \mathbf{Q}'_{2,2} & \dots & \mathbf{Q}'_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}'_{R,1} & \mathbf{Q}'_{R,2} & \dots & \mathbf{Q}'_{R,R} \end{bmatrix} \tag{27}$$

where each submatrix $\mathbf{Q}'_{i,j}$ is the $T \times T$ sample covariance matrix for replicates i and j.

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

$$\mathbf{B} = \frac{1}{L} \sum_{l=1}^{L} \frac{\mathbf{h}_l}{2} \circ \left(\frac{1}{\mathbf{d}_l} + \frac{1}{2\mathbf{n}_l} + \frac{1}{2\mathbf{d}_l \circ \mathbf{n}_l} \right) \tag{28}$$

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

$$\mathbf{H} = \frac{2\mathbf{D} \circ \mathbf{N}}{(\mathbf{D} - 1) \circ (\mathbf{N} - 1)} \circ \mathbf{F} \circ (1 - \mathbf{F}) \tag{29}$$

where division here is elementwise. Thus, $\mathbf{B}$ is a $R \times (T+1)$ matrix. As explained in Supplementary Material Section 5.2 and 5.3, the temporal variances and covariances require bias corrections, meaning each temporal covariance submatrix $\mathbf{Q}_{r,r}$ requires two corrections. For an element $Q_{r,t,s} = \text{Cov}(\Delta p_t, \Delta p_s)$ of the temporal covariance submatrix for replicate r, $\mathbf{Q}_{r,r}$, we apply the following correction

$$Q_{r,t,s} = \begin{cases} Q'_{r,t,s} - b_{r,t} - b_{r,t+1}, & \text{if } t = s \\ Q'_{r,t,s} + b_{r,\max(t,s)}, & \text{if } |t - s| = 1 \end{cases} \tag{30}$$

where $b_{r,t}$ is element in row r and column t of $\mathbf{B}$. Additionally, in some study designs, a single timepoint is shared for the initial generation across replicates. In this case, the sampling noise is shared between

5.5 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.

5.6 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 (XXX), $G(t)$ which is the ratio of covariance to total variance (XXX), and the convergence correlation (equation (1)). 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 number of loci in all sampled blocks,

$$\tilde{\theta}_b = \sum_{i=1}^{W} w_i \frac{N(\mathbf{x}_i)}{D(\mathbf{x}_i)} \tag{31}$$

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,

$$C_\alpha = \left(2\widehat{\theta} - q_{1-\alpha/2}, 2\widehat{\theta} - q_{\alpha/2}\right). \tag{32}$$

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

5.7 The Empirical Neutral Null Windowed Covariance Distribution

6 Supplementary Figures

6.1 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 4. Overall, we find the correction strongly

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

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

6.4 Bergland et al. (2014) Re-Analysis

We also applied our temporal covariance approach to Bergland et al. (2014), a study that finds evidence of genome-wide fluctuating selection between Spring and Fall seasons across three years. As described in Buffalo and Coop (2019), we would expect positive covariances between like seasons (i.e. Spring 2010 and Spring 2011), and negative covariances between dislike seasons (i.e. Spring 2010 and Fall 2011) when fluctuating selection is strong and acts genome-wide, as the original study found. 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 XXX, we show the temporal covariances from varying reference generations, across Spring/Fall, Spring/Spring, and Fall/Fall pairs. 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, the all other temporal covariances do not fit the pattern we would expect under genome-wide fluctuating selection.

We were concerned that this might be due to a flaw in our methods. To address this concern, we first ensured that our temporal-covariance matrix bias correction was working as expected. We find that it corrects the relationship between depth and both variance and covariance (Supplementary Figure 12) as expected. We then re-analyzed 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 Spring/Fall seasons are encoded as a dummy variable, and are regressed on the frequency data. 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 10 A), and the histograms of the p-values are nearly identical (Supplementary Figure 10 B).

While 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), they do not construct a null distribution of p-values as would be expected if fluctuating selection were not acting. Without such a distribution, we do not have a statistical expectation for the number of variants across the genome with frequency trajectories that spuriously fit the pattern of seasonal fluctuating selection. We found that we can create this empirical null distribution by randomly permuting the season labels and re-running the per-locus seasonal GLM model. 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 11, 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 fraction of the ${}^{6!}/_{(6-2)!} = 30$ 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 XXX), where we see strong evidence of selection (non-zero temporal covariances), yet the pattern does not follow that of seasonal fluctuating selection.

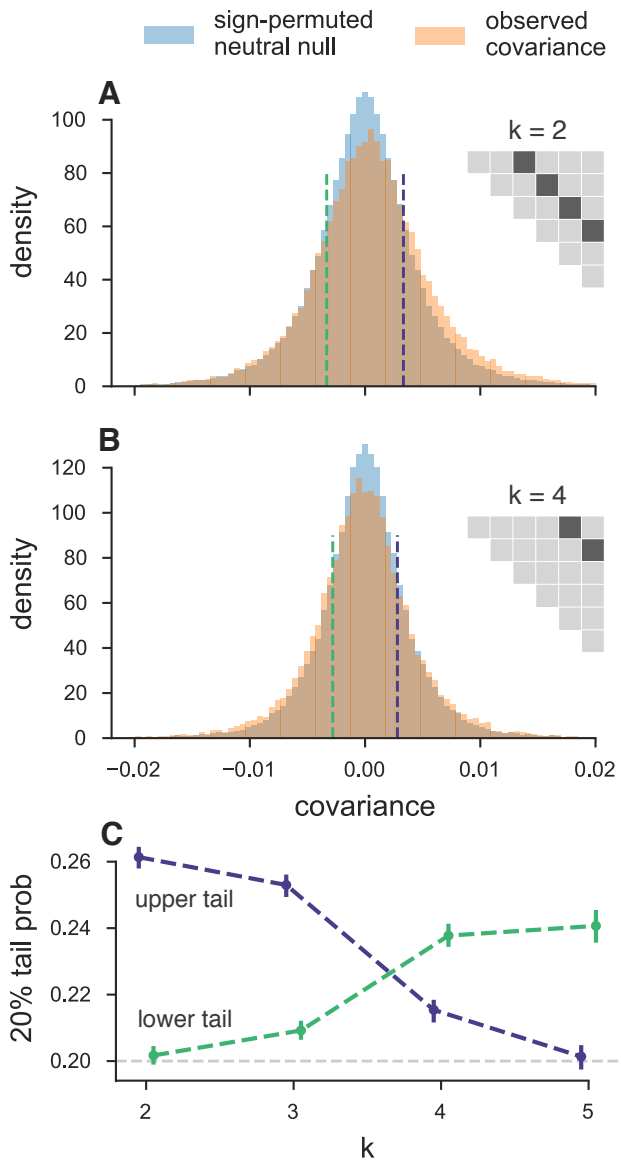

Figure 3: A, B: The distribution of temporal covariances calculated in 100kb genomic windows from the Barghi et al. (2019) study, plotted alongside an empirical neutral null distribution created by recalculating the windowed covariances on sign permuting the allele frequency changes. The bin number is 88, chosen by cross validation (Supplementary Materials 5). 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 B 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 interval, and the light gray dashed line indicates the 20% tail probability expected under the neutral null.

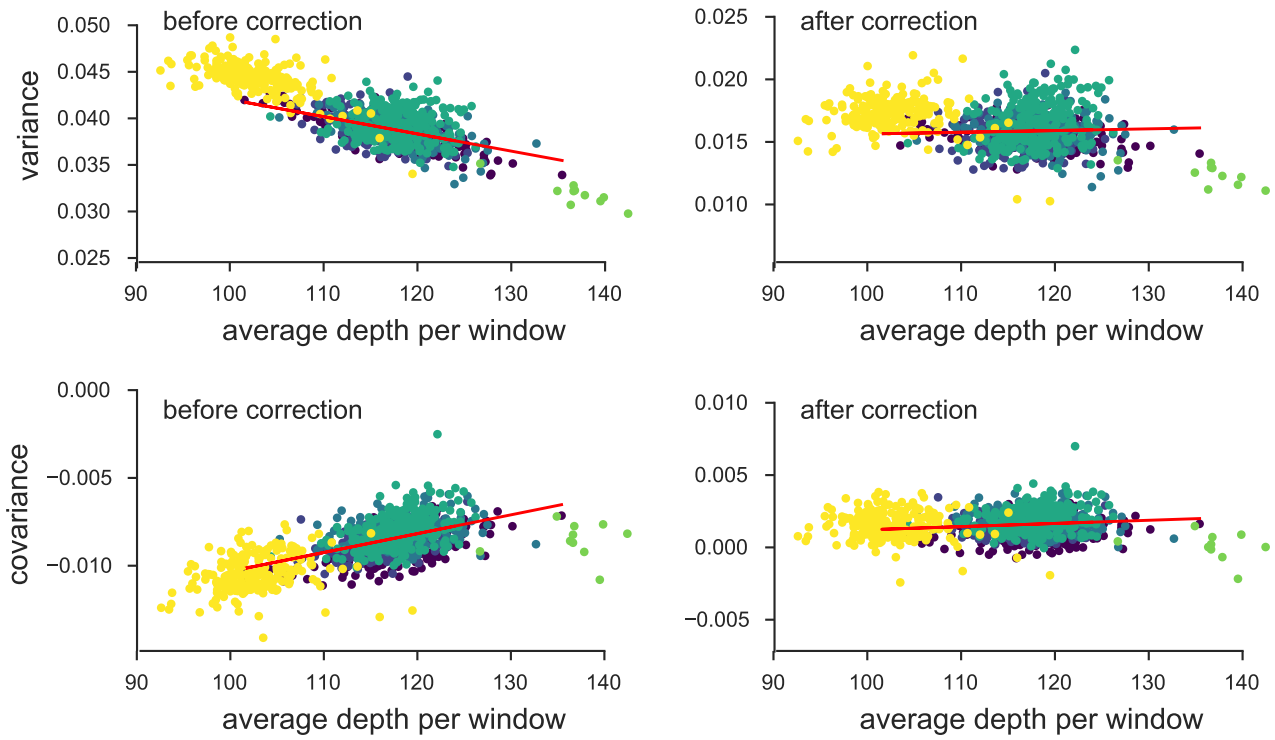

Figure 4: The variance and covariances from the Barghi et al. (2019) study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. Overall, the 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.

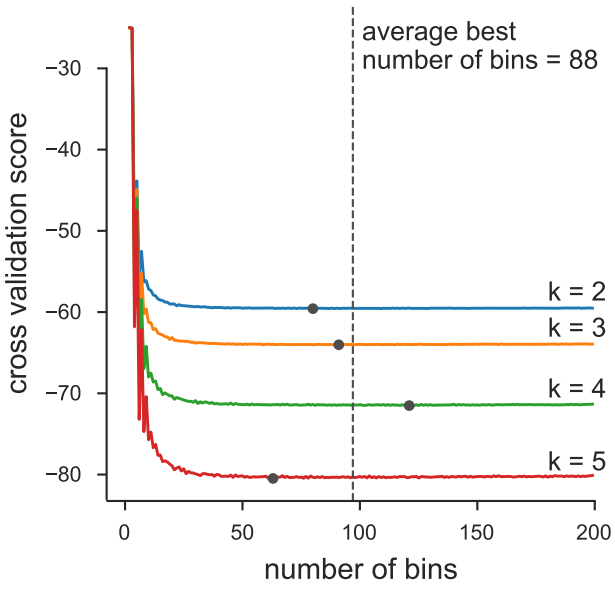

Figure 5: We chose number of bins used in the histograms of Figure 3 via an analytic expression for the cross-validation risk, based on the equation 6.16 of (Wasserman 2006, p. 129). Above, we plot the cross-validation risk for various numbers of bins, for each of the four off-diagonals of the temporal 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.

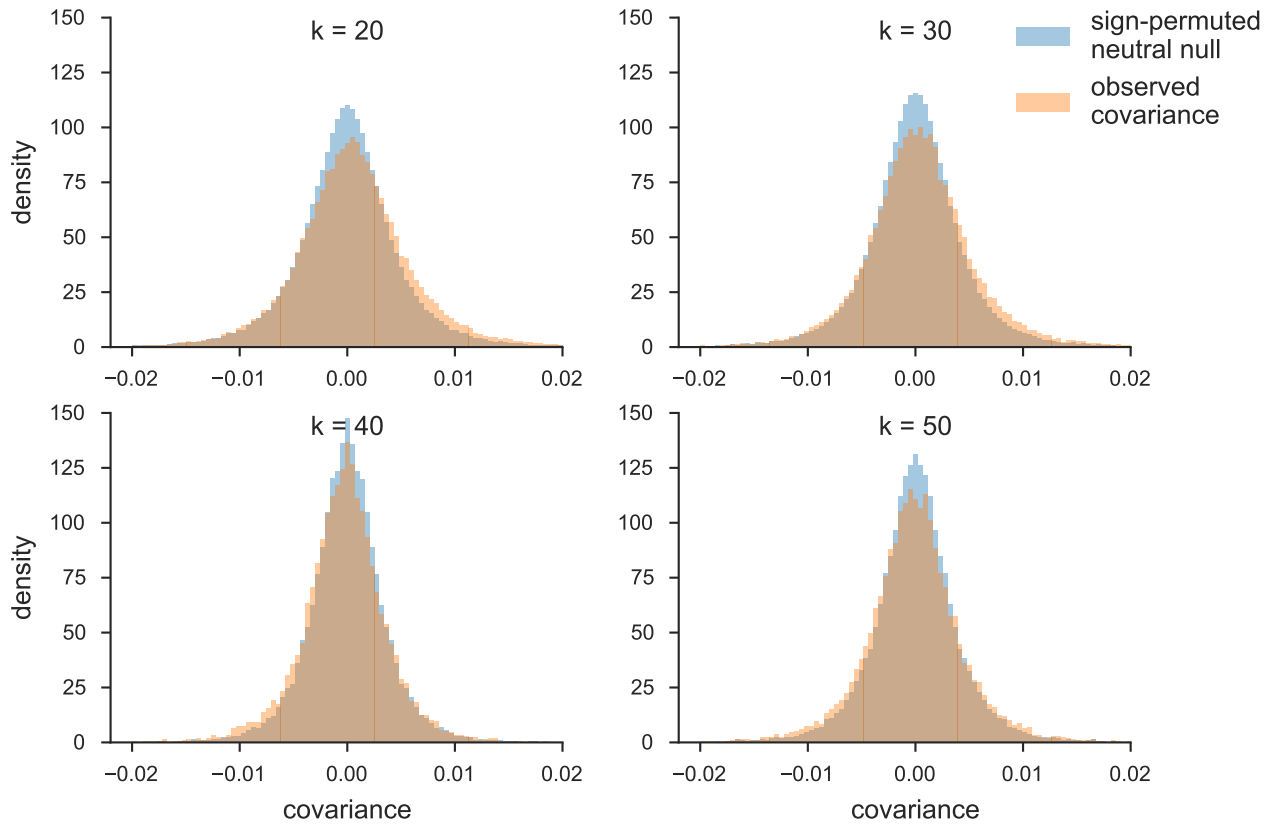

Figure 6: The distribution of temporal covariances calculated across 100kb genomic windows from Barghi et al. (2019)'s study (orange) and the block sign permuted empirical neutral null distribution of the windowed covariances (blue). Each panel shows these windowed covariances and the empirical null distribution for covariances $\mathrm{Cov}(\Delta p_t, \Delta p_{t+k})$, k is the number of generations between allele frequency changes.

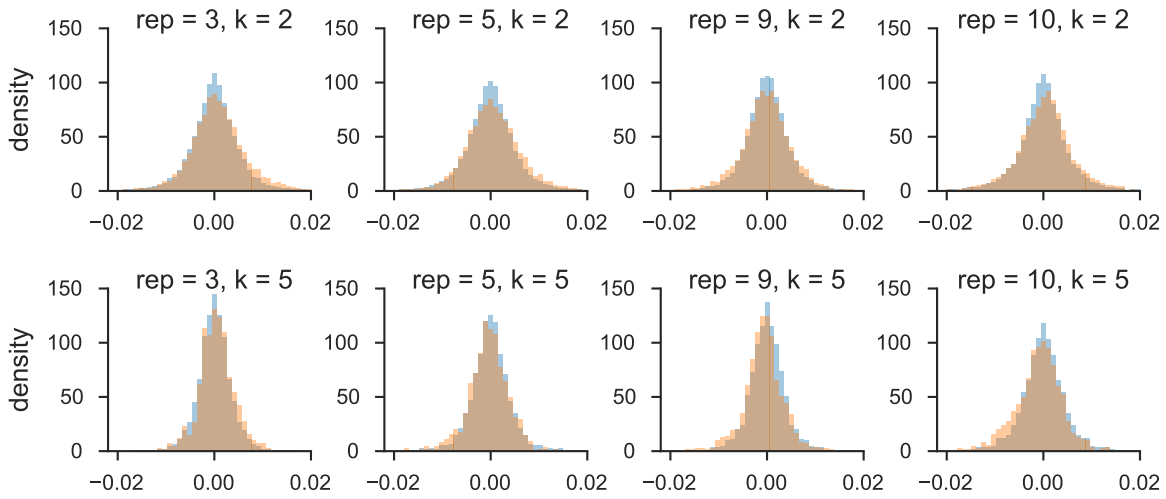

Figure 7: The distribution of windowed temporal covariances alongside the empirical neutral null for five randomly sampled replicates (columns), for $k = 2$ (first row) and $k = 5$ (second row). The main figure of the paper pools all replicate window and empirical neutral null covariances; we show here 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.

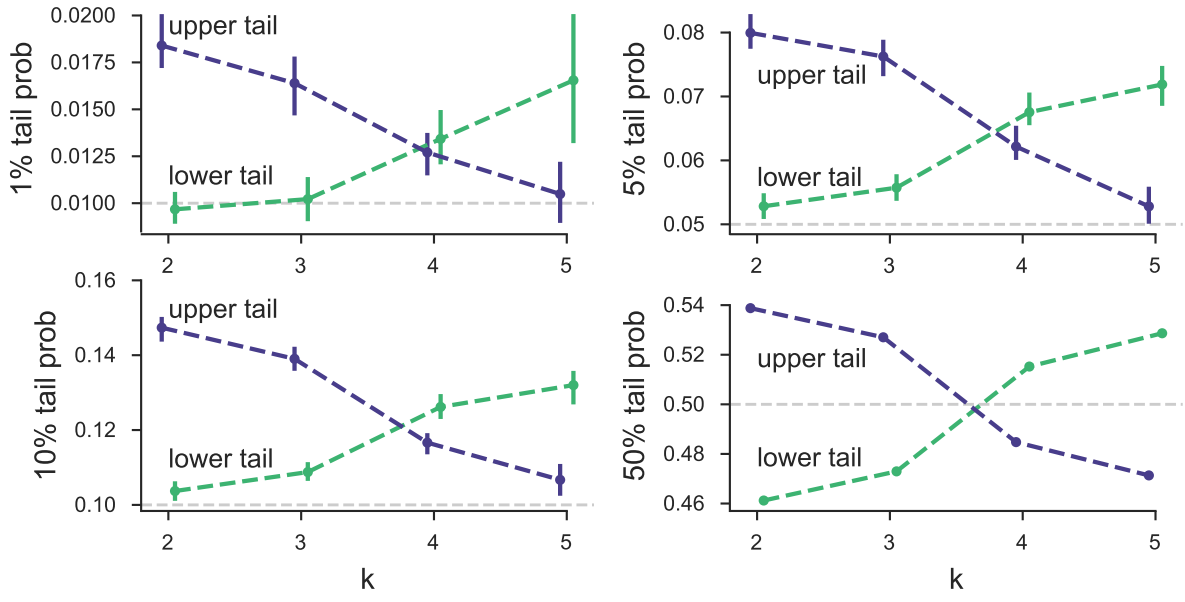

Figure 8

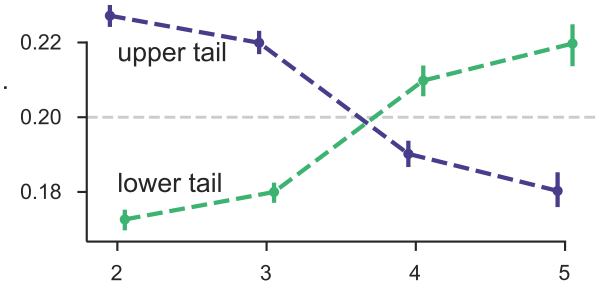

Figure 9: The 20% lower and upper tail probabilities for the observed windowed covariances from the Barghi et al. (2019) study, based on sign-permuting at the chromosome level. This permutation empirical null is robust to long-range linkage disequilibrium acting over entire chromosomes.

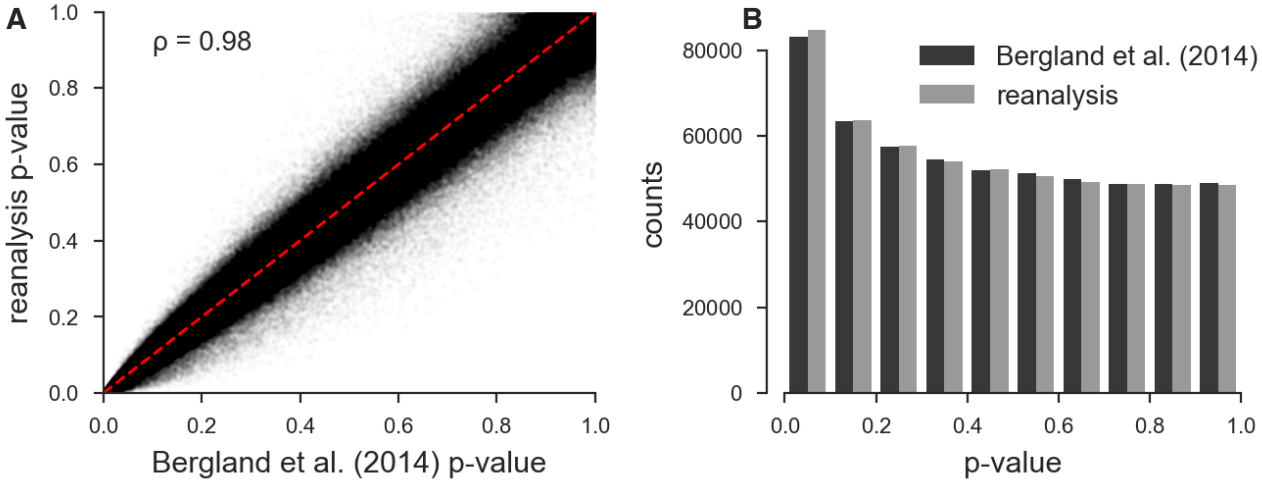

Figure 10: A: Scatterplot of the original unadjusted p-values from Bergland et al. (2014) and the p-values from our reanalysis of the same data using the same statistical methods; the minor discrepancy is likely due to software version differences. B: The histograms of the p-values of our reanalysis 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.

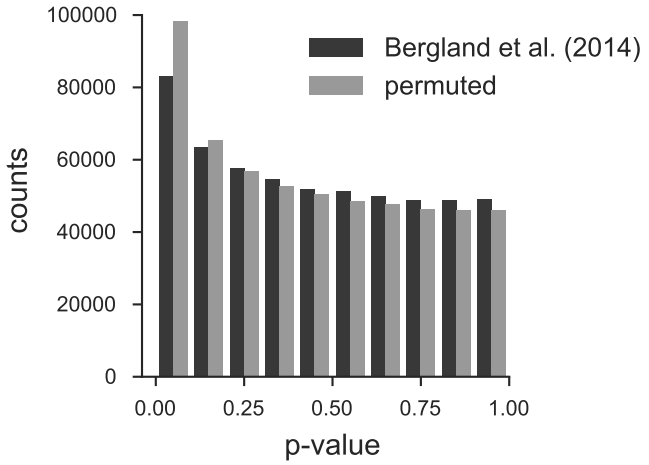

Figure 11

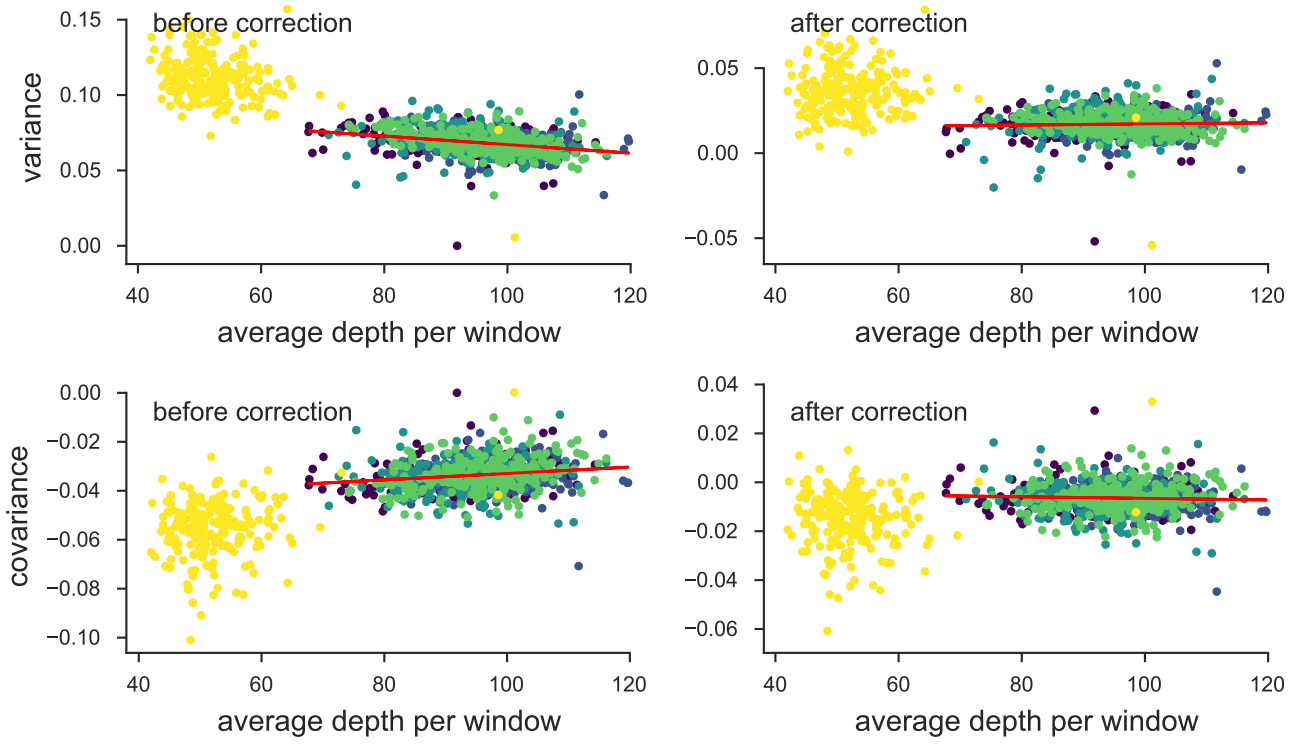

Figure 12: The variance and covariances from the Bergland2014-ij study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. The bias correction 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.