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 5, 2019

1   **Abstract**

2

3 # 1 Introduction

4 A long-standing problem in evolutionary genetics is quantifying the roles drift and selection have
5 in shaping genome-wide allele frequency changes. Previous work on this question has focused
6 on teasing apart the impacts of drift and selection on genome-wide diversity using population
7 samples from a single contemporary timepoint, usually by modeling the correlation between re-
8 gional recombination rate and diversity created in the presence of linked selection (**Begun1992-ey**;
9 **Elyashiv2016-vt**). While this approach has dramatically informed our present view of the role
10 of linked selection in shaping patterns of genome-wide diversity across a variety of organisms
11 (**Begun2007-bg**; **Beissinger2016-cm**; **Sattath2011-dr**; **Williamson2014-oy**; **Andersen2012-bj**;
12 **Cutter2010-gi**), and has allowed us to quantity the relative roles positive selection (hitchhik-
13 ing) and negative selection (background selection) have shaping patterns of genome-wide diversity
14 (**Nordborg2005-dc**; **McVicker2009-ax**; **Hernandez2011-gs**; **Elyashiv2016-vt**), we fully lack
15 an understanding of how linked selection acts over an important dimension: time.

16   Presently, there are numerous examples of phenotypes rapidly evolving (**Grant2011-wk**; **Grant2006-hj**;
17 **Reznick1997-mh**; **Franks2007-dr**) yet the polygenic nature of fitness makes detecting the impact
18 selection has on genome-wide variation over short timescales remarkably difficult. This is because
19 the effect of selection on a polygenic trait (such as fitness) is distributed across loci proportional
20 to their effect sizes, leading to subtle allele frequency shifts that are difficult to distinguish from
21 background levels of genetic drift and sampling variance. However, increasingly genomic studies
22 with multiple timepoints and, in some cases, multiple replicate populations are becoming available,
23 are being used to detect selected loci (**Turner2011-sx**; **Turner2012-bm**) and differentiate modes
24 of selection (**Burke2010-tz**; **Barghi2019-qy**). In a previous paper, we proposed that one signal
25 of polygenic linked selection that could be detected from these temporal genomic studies is the tem-
26 poral autocovariance in allele frequency changes (**Buffalo2019-io**). These covariances are directly
27 estimable from temporal genomic data and are created when the loci that underly heritable fitness
28 variation perturbed linked neutral alleles; when genetic drift acts alone in a closed population, these
29 covariances are zero in expectation.

30   Here, we provide the first empirical analyses quantify the impact of linked selection acting over
31 short timescales (tens of generations) across two evolve and re-sequence studies (**Barghi2019-qy**;

Kelly2019-dc), and one artificial selection experiment (**Castro2019-uk**). 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
Kelly2019-dc	*D. simulans*	lab	3	∼1100	14	2
Barghi2019-qy	*D. simulans*	lab	10	∼1000	60	7
Castro2019-uk	*M. musculus*	tibiae length	2	32	20	2
		control	1	28		

We first analyzed **Barghi2019-qy**, 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 genome-wide 6×6 temporal covariance matrix $\mathbf{Q}$ for each of the ten replicates. Each row of these matrices represent the temporal covariance $\text{Cov}(\Delta_{10}p_s, \Delta_{10}p_t)$, between the allele frequency change $(\Delta_{10}p_t)$ 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 Material Sections 5.2 and 5.4 for details on the bias correction). The covariances are expected to be zero when only drift is acting in a closed population, as only heritable variation for fitness can create covariance between allele frequency changes (**Buffalo2019-io**). 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 additive genetic variance for fitness decay (**Buffalo2019-io**). The temporal covariances per replicate are noisier but this general pattern holds; see Supplementary Figures XXX. **Barghi2019-qy**'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 may well be higher yet (see Supplementary Material Section 5.5 for more details).

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 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, $\text{Var}(p_t - p_0) = \sum_{i \neq j} \text{Cov}(\Delta p_i, \Delta p_j)/\text{Var}(p_t - p_0)$, 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
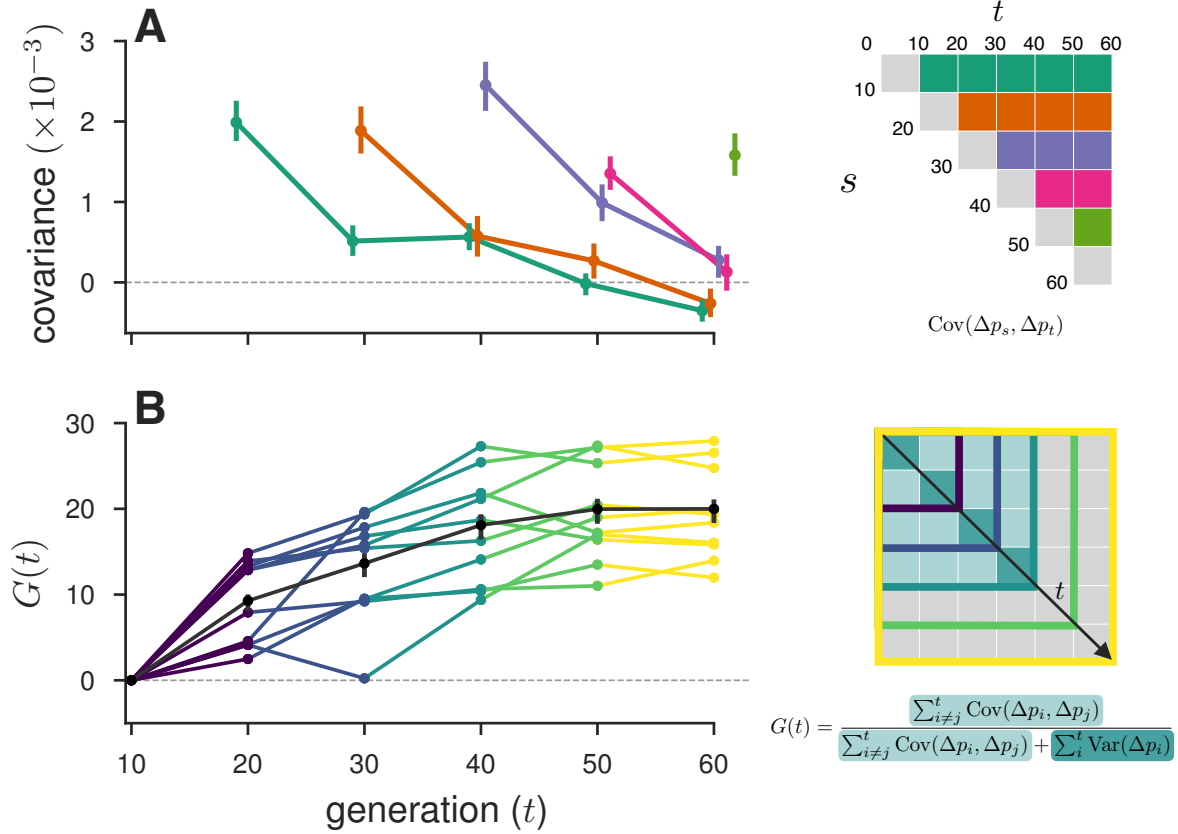
Figure 1: A: Temporal covariance, averaged across all ten replicate populations, through time from the **Barghi2019-qy** study. Each line depicts the temporal covariance $\text{Cov}(\Delta p_s, \Delta p_t)$ from some reference generation s to a later time t which varies along the x-axis; each line corresponds 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 $G(t)$.

selection (**Buffalo2019-io**). 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-qy** 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 **Barghi2019-qy** also allows us to quantify another covariance: the co-variance in allele frequency change between replicate populations experiencing convergent selection pressures. These between-replicate covariances are created in the same way as temporal covari-

ances 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,

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

where A and B here are two replicate labels, and for the **Barghi2019-qy** 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 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 convergent covariances 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.

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 **Kelly2019-dc**, 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 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 $[0.31, 0.40]$). *GIVE conversion % var due to linked selection.*

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 (**Castro2019-uk**). 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.

Finally, we observed that in the *longest study* we analyzed (**Barghi2019-qy**), 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 no 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 changing epistatic*
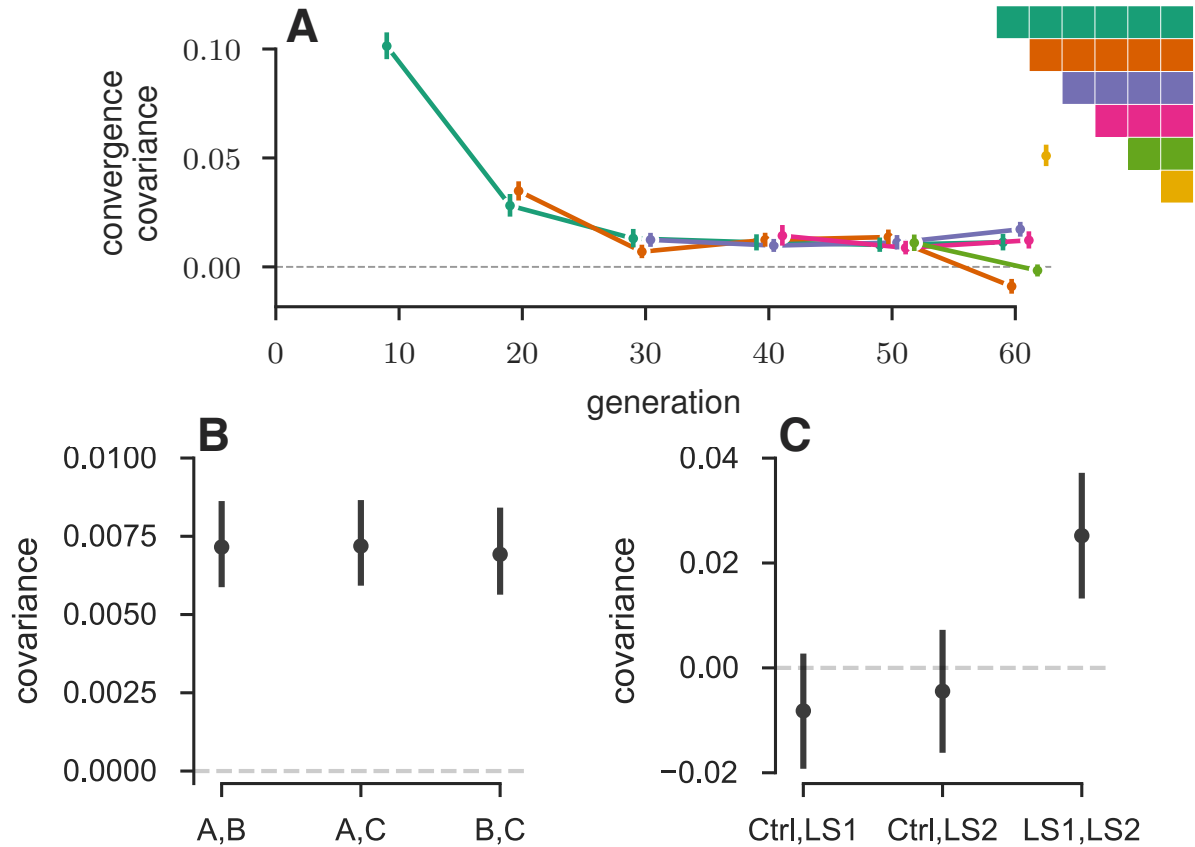
4

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.

5

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 9). *We* see a striking pattern that *the windowed* covariances not only decay towards zero, but in fact become negative through time, consistent with some *many* regions in the genome having had a reversed fitness effect at later timepoints.

f

t

3 Acknowledgments

4 Appendix

5 Estimator Bias Correction

5.1 Correcting variance bias with a single depth sampling process

Following **Waples1989-sj**, 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,

$$\mathrm{Var}(\widetilde{p}_t) = \mathbb{E}(\mathrm{Var}(\widetilde{p}_t|p_t)) + \mathrm{Var}(\mathbb{E}(\widetilde{p}_t|p_t)) \tag{2}$$

$$= \underbrace{\frac{p_t(1-p_t)}{d_t}}_{\text{generation } t \text{ sampling noise}} + \underbrace{\mathrm{Var}(p_t)}_{\text{variance due to evolutionary process}} \tag{3}$$

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 (**Buffalo2019-io**). We can write

$$V(p_t) = V\left(p_0 + (p_1 - p_0) + (p_2 - p_1) + \ldots + (p_t - p_{t-1})\right) \tag{4}$$

$$= V\left(p_0 + \Delta p_0 + \Delta p_1 + \ldots + \Delta p_{t-1}\right) \tag{5}$$

$$= V(p_0) + \sum_{i=0}^{t-1} \mathrm{Cov}(p_0, \Delta p_i) + \sum_{i=0}^{t-1} \mathrm{Var}(\Delta p_i) + \sum_{0 \leq i < j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j). \tag{6}$$

6

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,

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

The second term, the cumulative impact of variance in allele frequency change can be partitioned into heritable fitness and drift components (**Santiago1995-hx**; **Buffalo2019-io**)

$$V(p_t) = \sum_{i=0}^{t-1} \text{Var}(\Delta_D p_i) + \sum_{i=0}^{t-1} \text{Var}(\Delta_H p_i) + \sum_{0 \leq i < j}^{t-1} \text{Cov}(\Delta p_i, \Delta p_j). \tag{8}$$

where $\Delta_H p_t$ and $\Delta_D p_t$ indicate the allele frequency change due to heritable fitness variation and drift respectively. Then, sum of drift variances in allele frequency change is

$$\sum_{i=0}^{t-1} \text{Var}(\Delta_D p_i) = \sum_{i=0}^{t-1} \frac{p_i(1 - p_i)}{2N} \tag{9}$$

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

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

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

which is the usual variance in allele frequency change due to drift. Then, the total allele frequency change from generations 0 to t is $\text{Var}(\widetilde{p}_t - \widetilde{p}_0) = \text{Var}(\widetilde{p}_t) + \text{Var}(\widetilde{p}_0) - 2\,\text{Cov}(\widetilde{p}_t, \widetilde{p}_0)$, where the covariance depends on the nature of the sampling plan (see **Nei1981-oy**; **Waples1989-sj**). In the case where there is heritable variation for fitness, and using the fact that $\text{Cov}(\widetilde{p}_t, \widetilde{p}_0) = {p_0(1-p_0)}/{2N}$ for Plan I sampling procedures (**Waples1989-sj**), we write,

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

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

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

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

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

7

where $C = 1$ if Plan I is used, and $C = 0$ if Plan II is used (see **Waples1989-sj**, 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 **Jonas2016-ia**). 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

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

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

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

$$- \frac{Cp_0(1-p_0)}{N} + p_0(1-p_0)\left[1 - \left(1 - \frac{1}{2N}\right)^t\right] + \sum_{i=0}^{t-1}\mathrm{Var}(\Delta_H p_i) \tag{21}$$

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

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

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

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

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

$$(25)$$

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

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 $\text{Cov}(\Delta \widetilde{p}_t, \Delta \widetilde{p}_{t+1}) = \text{Cov}(\widetilde{p}_{t+1} - \widetilde{p}_t, \widetilde{p}_{t+2} - \widetilde{p}_{t+1})$, and thus shares the sampling noise of timepoint $t + 1$. Thus acts to bias the covariance by subtracting off the noise variance term of $\text{Var}(\widetilde{p}_{t+1})$, 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

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

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,

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

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

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

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 Barghi2019-qy 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 (**Buffalo2019-io**); 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} (**Bhatia2013-zy**). 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 **Wasserman2006-jl**, p. 194 **Davison2013-oy**) 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 Barghi2019-qy

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 Barghi2019-qy Empirical Null and Windowed Covariance Distributions

6.3 Barghi2019-qy Tail Probabilities for Windowed Covariances Distributions

6.4 Bergland2014-ij Re-Analysis

We also applied our temporal covariance approach to **Bergland2014-ij**, which found evidence of genome-wide fluctuating selection between Spring and Fall seasons across three years *Drosophila*

melanogaster. As described in **Buffalo2019-io**, we might 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 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 12) 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 an genome-wide excess of loci showing evidence of fluctuating selection we re-analyzed the data of (**Bergland2014-ij**) 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 **Bergland2014-ij**, 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 **Bergland2014-ij** 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). **Bergland2014-ij** find loci with a significant association with season after a Benjamini and Hochberg FDR p-value adjustment (**Benjamini1995-jy**), *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.*

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 (**Machado2018-cs**). We find, regardless of whether we permute at the locus-level or the permutation replicate-level, that the observed seasonal p-value distribution **Bergland2014-ij** 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 **Bergland2014-ij** 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 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

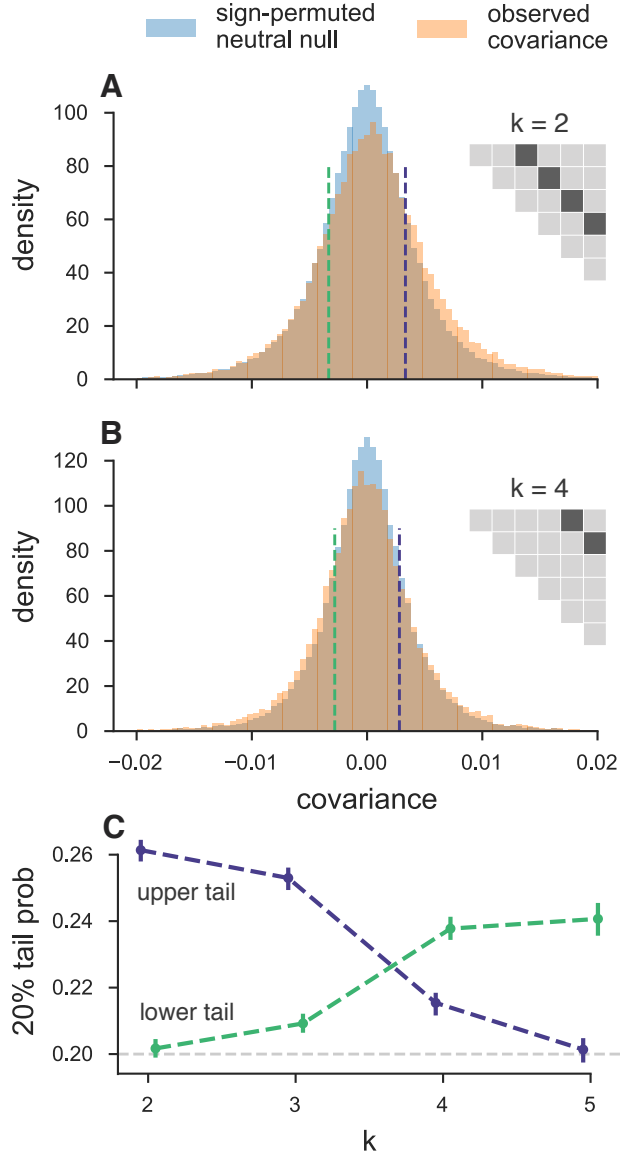316   does not follow that of seasonal fluctuating selection. *ADD MORE HERE*

Figure 3: A, B: The distribution of temporal covariances calculated in 100kb genomic windows from the **Barghi2019-qy** study, plotted alongside an empirical neutral null distribution created by recalculating the windowed covariances on *a 1000 sign permutions* 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.
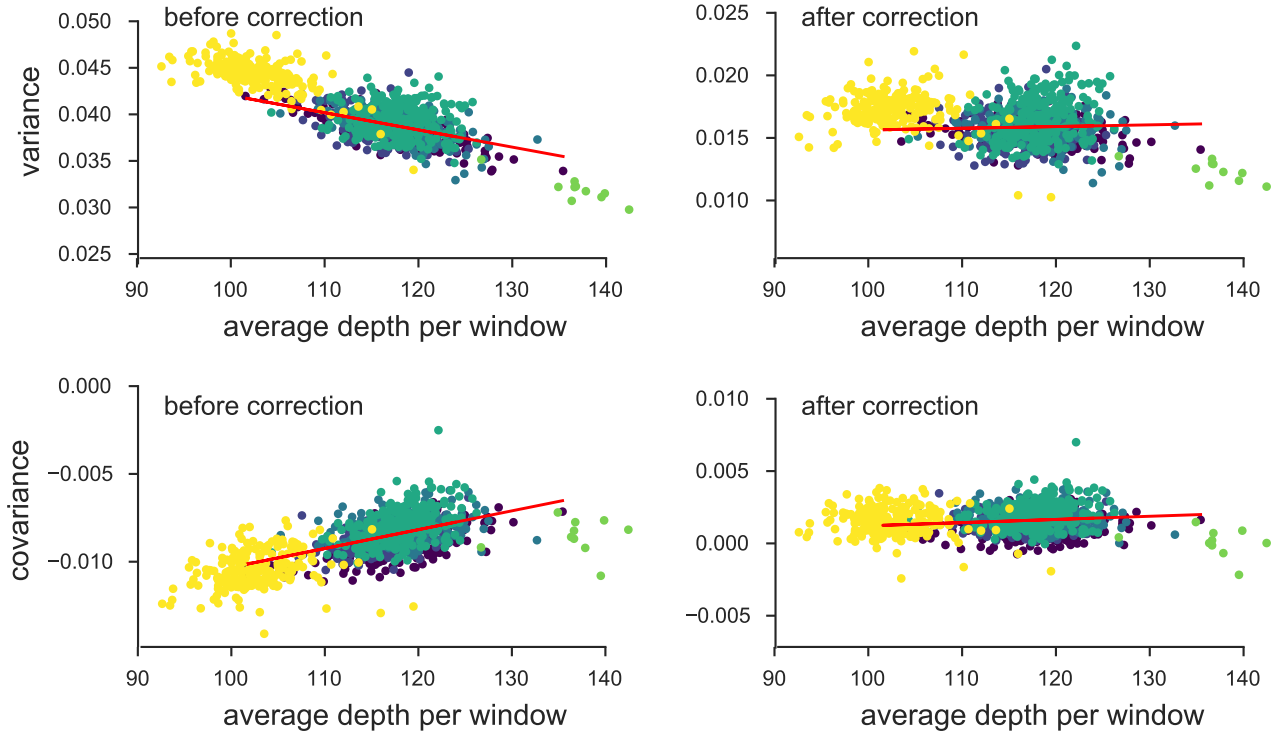
14

Figure 4: The variance and covariances from the **Barghi2019-qy** 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.
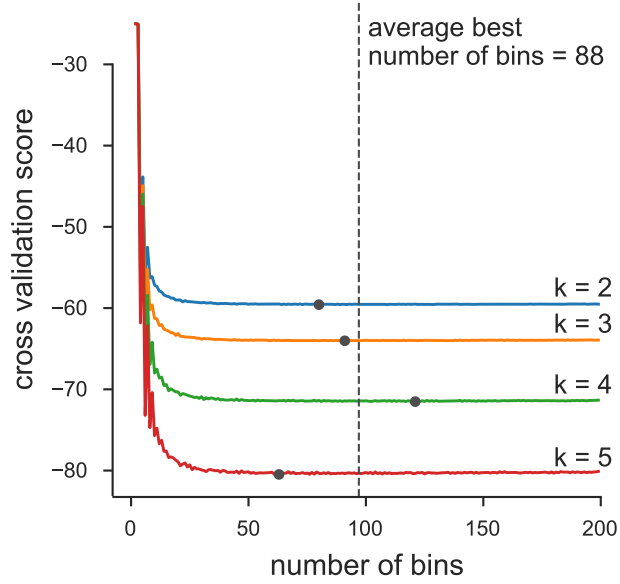
Figure 5: We chose number of bins used in the histograms of Figure 3 via an analytic expression for the cross-validation risk, based on the equation 6.16 of (**Wasserman2006-jl**, p. 129). Above, we plot the cross-validation risk for various numbers of bins, for each of the four off-diagonals of the 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.
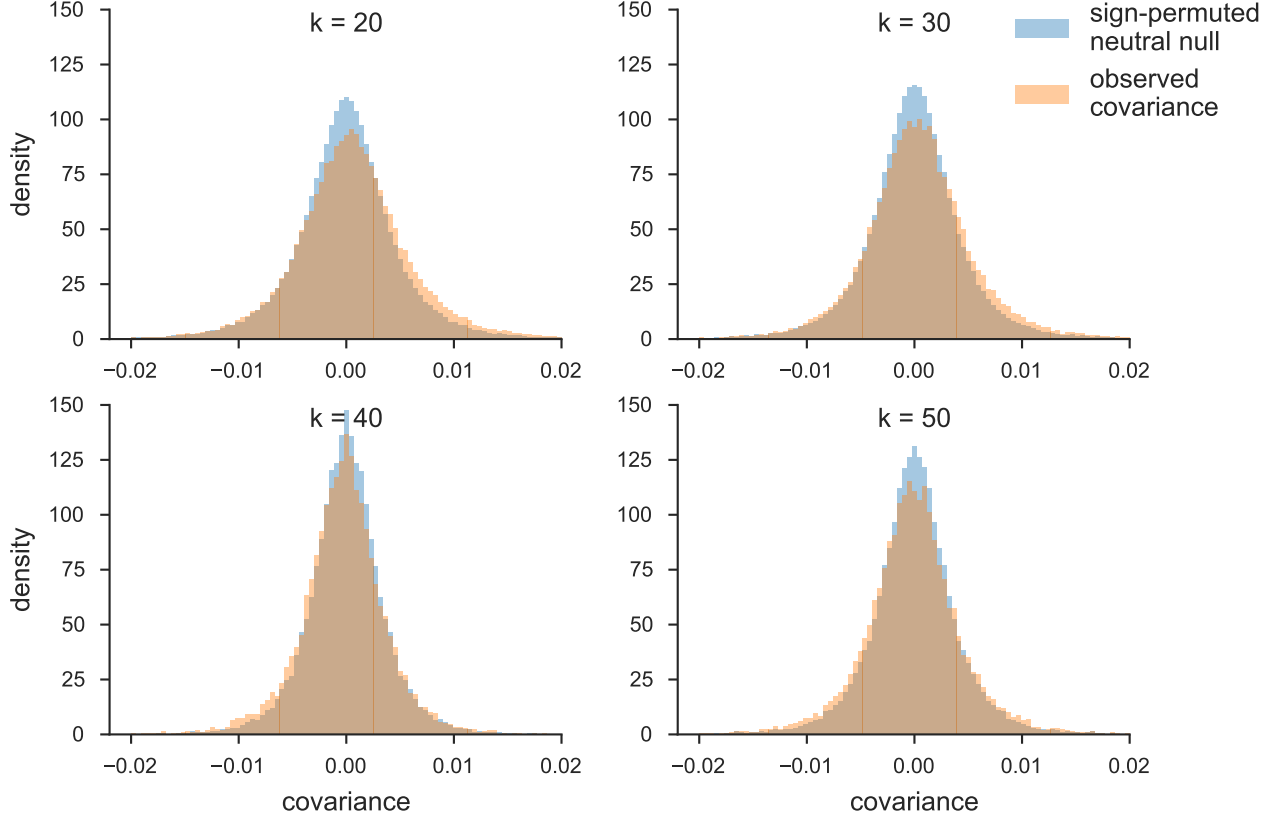
Figure 6: The distribution of temporal covariances calculated across 100kb genomic windows from **Barghi2019-qy**'s study (orange) and the block sign permuted empirical neutral null distribution of the windowed covariances (blue). Each panel shows these windowed covariances and the empirical null distribution for covariances $\text{Cov}(\Delta p_t, \Delta p_{t+k})$, k is the number of generations between allele frequency changes.
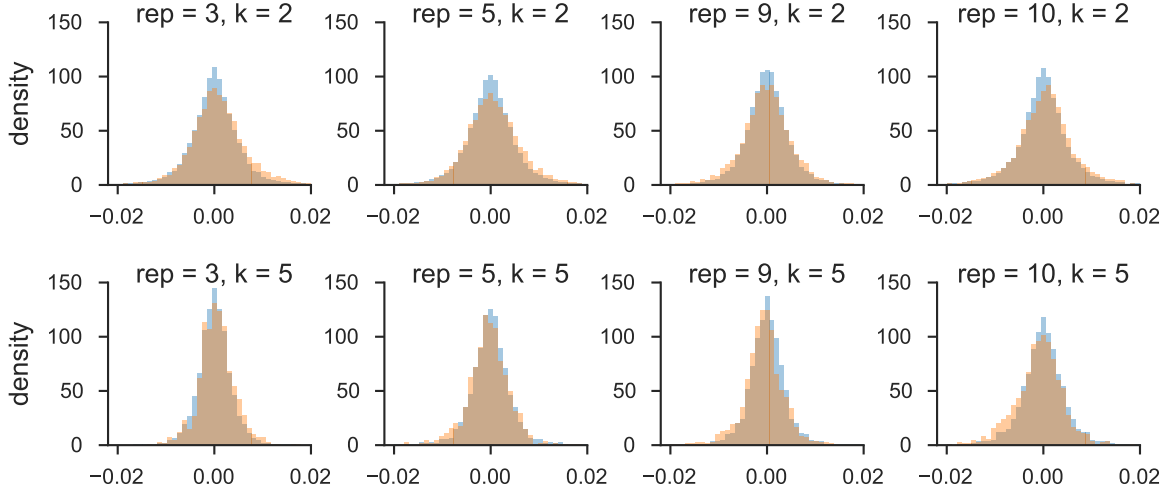
Figure 7: The distribution of windowed temporal covariances alongside the empirical neutral null for five randomly sampled replicates (columns), for $k = 2$ (first row) and $k = 5$ (second row). The main figure of the paper pools all replicate window and empirical neutral null covariances; we show here 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.
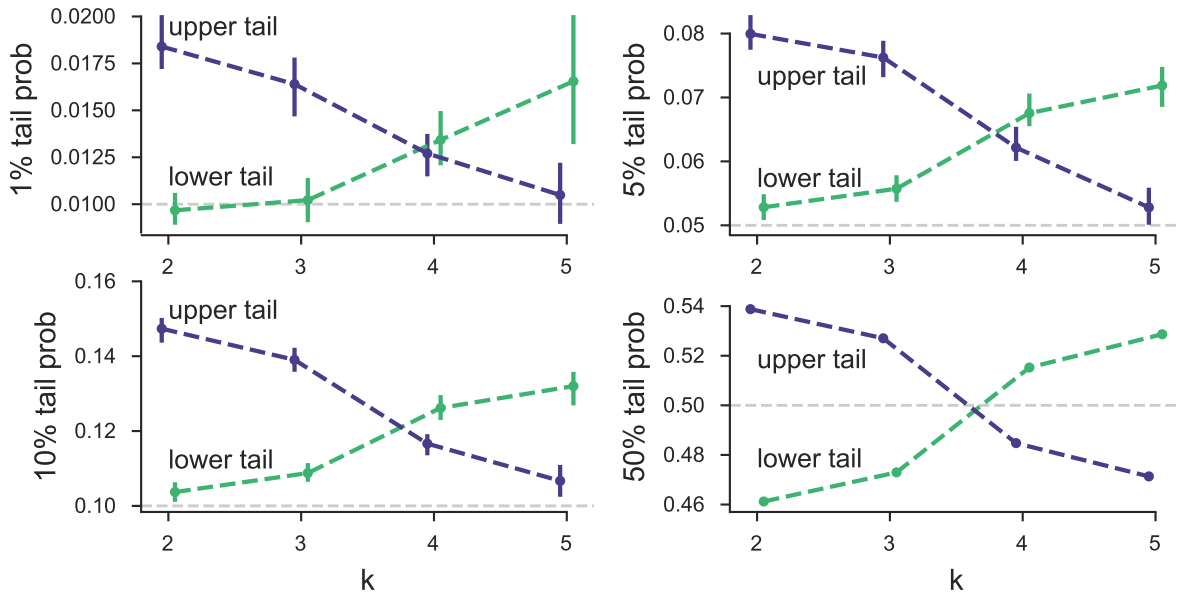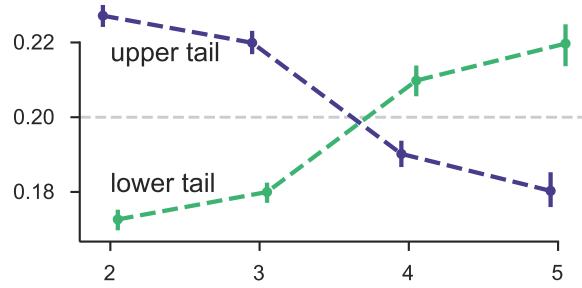

Figure 8

Figure 9: The 20% lower and upper tail probabilities for the observed windowed covariances from the **Barghi2019-qy** study, based on sign-permuting at the chromosome level. This permutation empirical null is robust to long-range linkage disequilibrium acting over entire chromosomes.
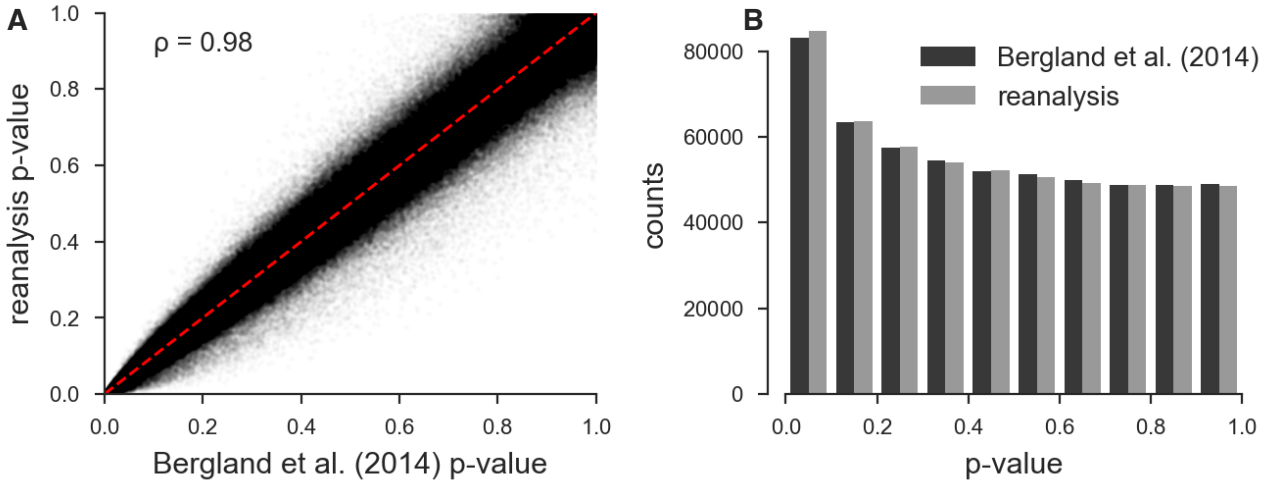

Figure 10: A: Scatterplot of the original unadjusted p-values from **Bergland2014-ij** and the p-values from our reanalysis of the same data using the same statistical methods; the minor discrepancy is likely due to software version differences. B: The histograms of the p-values of our reanalysis and the original **Bergland2014-ij** 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.
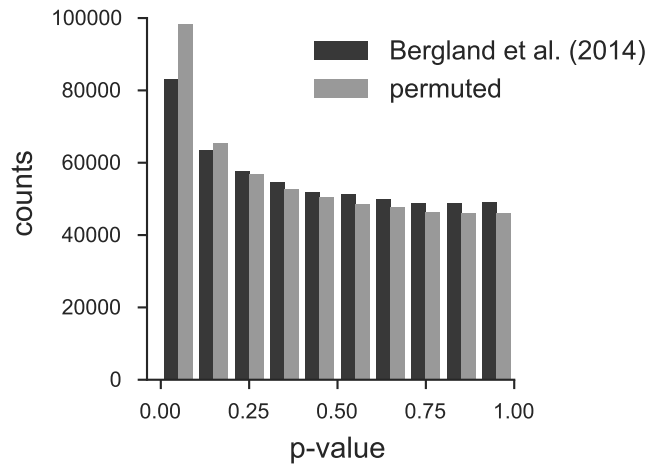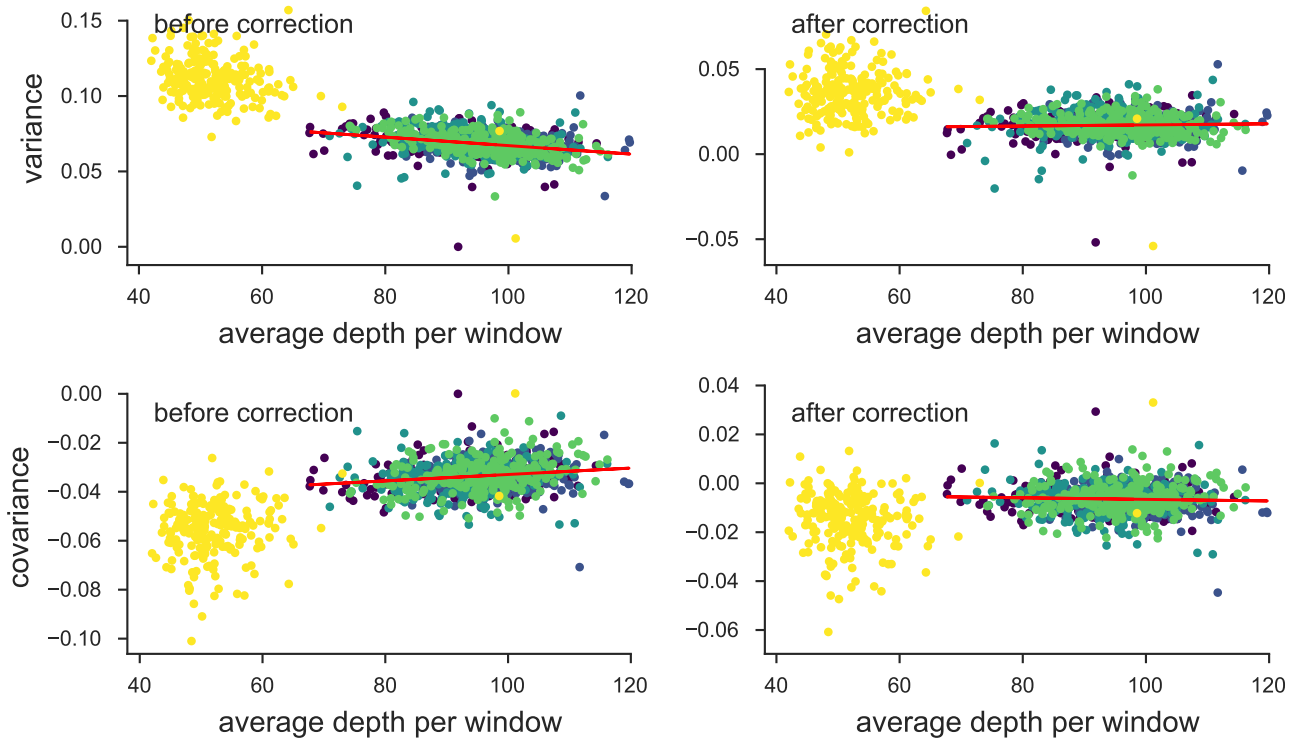
Figure 11

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.