

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

Vince Buffalo<sup>\*,†,1</sup> and Graham Coop<sup>†</sup>

<sup>\*</sup>Population Biology Graduate Group

<sup>†</sup>Center for Population Biology, Department of Evolution and Ecology, University of California, Davis, CA 95616

<sup>1</sup>Email for correspondence: vsbuffalo@ucdavis.edu

September 30, 2019

## Abstract

Rapid phenotypic adaptation is often observed in natural populations and selection experiments. However, detecting the impact of this selection genome-wide 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 over 20% of allele frequency change in these experiments is driven by selection in some 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\_Smith1974-lc**, **Charlesworth1993-gb**; **Nordborg1996-nq**; see **Barton2000-zg** 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 (**Begun1992-ey**; **Elyashiv2016-vt**). This approach has shown linked selection has a major role in shaping patterns of genome-wide diversity across the genomes of some of organisms (**Begun2007-bg**; **Beissinger2016-cm**; **Sattath2011-dr**; **Williamson2014-oy**; **Andersen2012-bj**; **Cutter2010-gi**), and has allowed us to quantify the relative influence of positive selection (hitchhiking) and negative selection (background selection; **Nordborg2005-dc**; **McVicker2009-ax**; **Hernandez2011-gs**; **Elyashiv2016-vt**). However, we lack an understanding of how genome-wide linked selection acts over time.

There are numerous examples of rapid phenotypic adaptation (**Grant2011-wk**; **Grant2006-hj**; **Reznick1997-mh**; **Franks2007-dr**) and rapid genomic evolution in asexual populations (**Good2017-om**; **Bennett1990-bc**; **Baym2016-kh**). Yet the polygenic nature of fitness makes detecting the impact 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 selected loci (**Turner2011-sx**; **Turner2012-bm**) and differentiate modes of selection (**Burke2010-tz**; **Barghi2019-qy**; **Therkildsen2019-zy**). 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 (**Bergland2014-ij**; **Machado2018-cs**). In a previous paper, we proposed that one useful signal for understanding the genome-wide impact of polygenic linked selection detectable from temporal genomic studies is the temporal autocovariance in allele frequency changes (**Buffalo2019-io**). 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 allele frequency changes in one generation predict changes in another due to a shared fitness background and shared selection pressures.

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 (**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 on 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

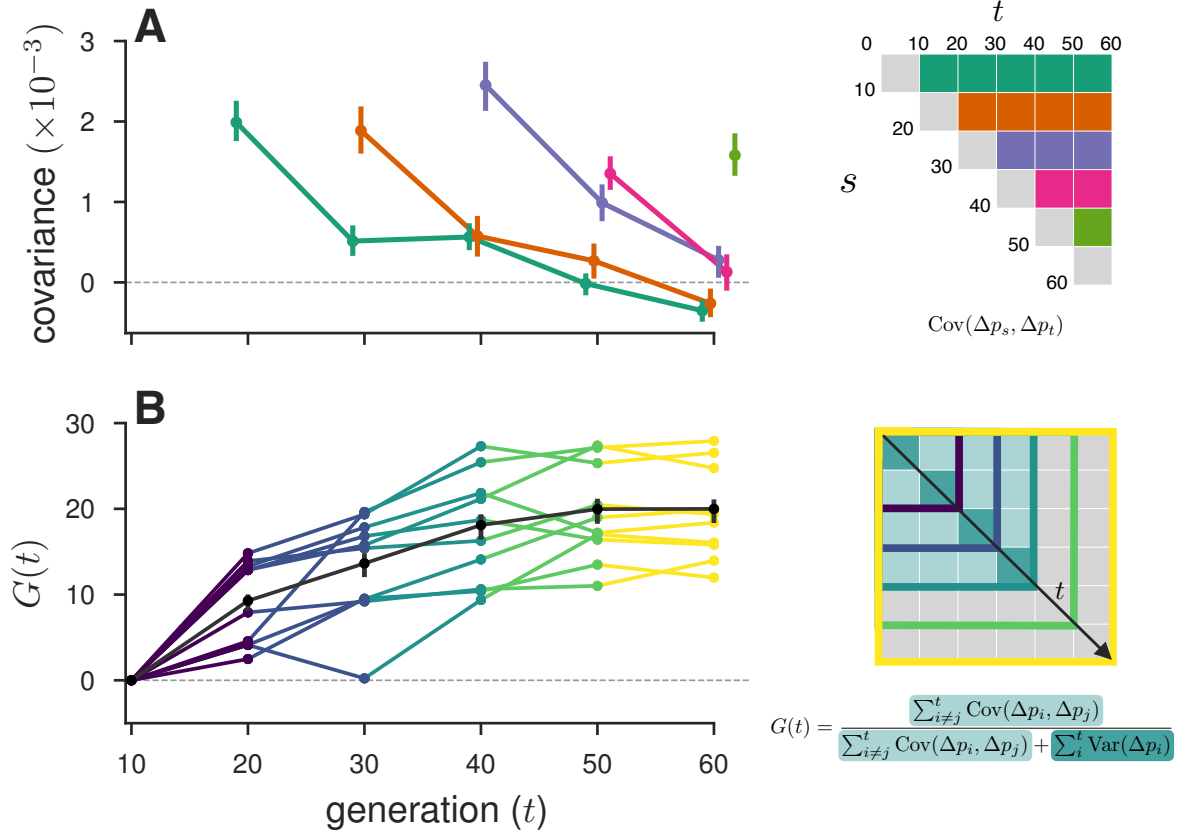
Study	Species	Selection	Replicates	Pop. Size	Gens.	Timepoints
<b>Kelly2019-dc</b>	<i>D. simulans</i>	lab	3	~1100	14	2
<b>Barghi2019-qy</b>	<i>D. simulans</i>	lab	10	~1000	60	7
<b>Castro2019-uk</b>	<i>M. musculus</i>	tibiae length	2	32	20	2
		control	1	28		

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 the genome-wide  $6 \times 6$  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 (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 Material Sections ?? and ?? 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 the additive genetic variance for fitness associated with neutral alleles decays (**Buffalo2019-io**). The temporal covariances per replicate are noisier but this general pattern holds; see Supplementary Figures ??. **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 (e.g. if adjacent generations had been sequenced) may well be higher yet (see Supplementary Material Section ?? 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. 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,  $\text{Var}(p_t - p_0) = \sum_i \text{Var}(\Delta p_i) + \sum_{i \neq j} \text{Cov}(\Delta p_i, \Delta p_j)$ , 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 (**Buffalo2019-io**). We call this measure  $G(t)$ , defined as

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

which estimates the 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



**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  $G(t)$ .

conservative, since the temporal covariances within 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. Greater than 20% of total variation in allele frequency change over 60 generations is the result of linked selection.

Additionally, we looked for a signal of temporal autocovariance in **Bergland2014-ij**, 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 ??.

The replicate design of **Barghi2019-qy** 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 experience 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,

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

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 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 seems to confirm a primary finding of the original **Barghi2019-qy** study: that alternative loci contribute to adaptation in 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 **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 statistically 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 to 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 (gray confidence intervals in Figure 2 C),



**Figure 2:** **A:** The convergence correlation, averaged across replicate pairs, through time. Each line represents the convergence correlation  $\text{cor}(\Delta p_s, \Delta p_t)$  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. **B:** The convergence covariance between individual pairs of replicates in the **Kelly2019-dc** data. **C:** The convergence covariance between individual pairs of replicates in (**Castro2019-uk**) 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 **Castro2019-uk** found to have signals of parallel selection between LS1 and LS2.

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 [0.013, 0.037]).

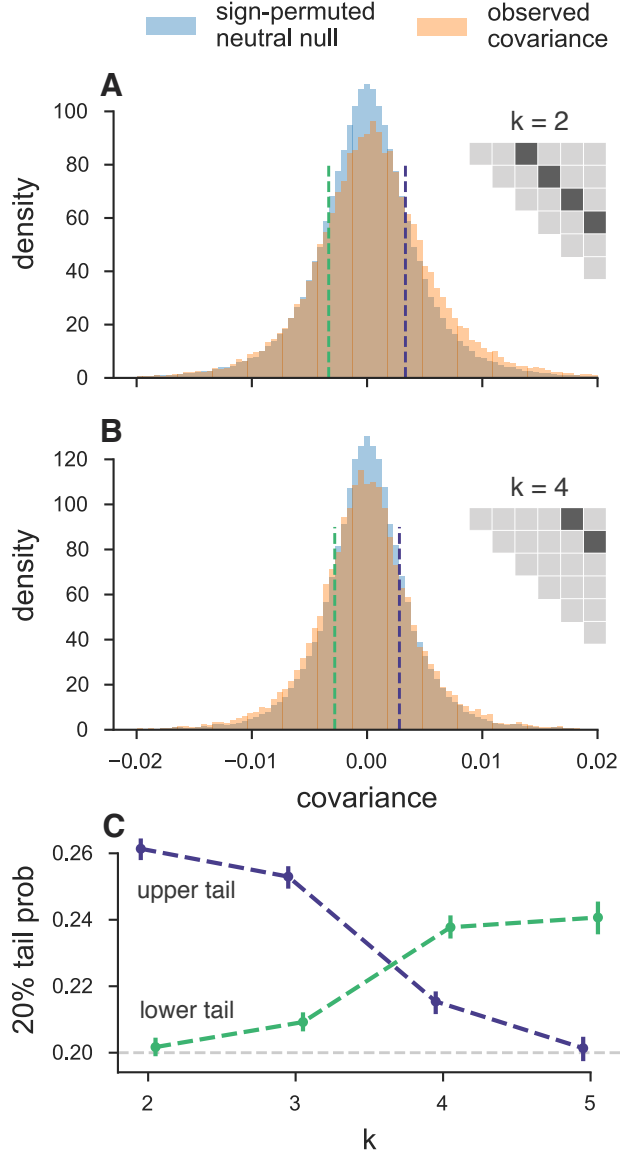
A major finding in the Longshanks study were two major-effect loci that showed parallel frequency shifts between the two selection lines: a region harboring the gene *Nkx3-2* known to be involved in limb development, and another region harboring six other candidate genes. We were curious to what extent 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), indicating that these

across-replicate covariances are indeed driven by a polygenic signal.

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 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 genomic windows which we refer to as *windowed covariances* (Figure 3 shows these distributions pooling across all replicates; see Supplementary Figure ?? for individuals replicates). The covariance of each tile 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, **Nuzhdin2013-gf**; **Baldwin-Brown2014-cl**). To address this, we have developed a permutation-based procedure that constructs a 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 ??). 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\_Smith1974-lc** 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 (**Robertson1961-ho**) and its later developments (**Santiago1995-hx**; **Santiago1998-bs**; see also **Barton2000-zg** 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 (**Buffalo2019-io**). Depending on how many loci affect fitness, such a strong effect



**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 1,000 sign permutations of allele frequency changes within tiles. The histogram bin number is 88, chosen by cross validation (Supplementary Materials ??). In subfigure **A**, windowed covariances  $\text{Cov}(\Delta p_t, \Delta p_{t+k})$  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 intervals, and the light gray dashed line indicates the 20% tail probability expected under the neutral null.



of linked selection may not be differentiable from genetic drift using only single contemporary population samples or looking 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 (**Buffalo2019-io**).

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 reasonably proportion of the response is shared across short timescales. These likely represent substantial under-estimates of the contribution of linked selection because ....

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 (**Buffalo2019-io**). 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. There are LD data for a natural population of *D. simulans* (**Signor2018-wg**), 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 (**Nuzhdin2013-gf**; **Kelly2019-dc**). 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 parameteric model to estimate population parameters such as the additive genetic variation for fitness directly from temporal genomic data alone (**Buffalo2019-io**).

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 than in natural populations. Conversely, these lab populations were maintained at very small population sizes, small effective population sizes 300, 450, and 45 for the **Kelly2019-dc**, and **Castro2019-uk** 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. To 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 **Dmitri\_others**. Efforts to estimate the quantify the impact of linked selection in have suggested that XXX to XXX of allele frequency change over long time periods is driven by selection (**McVicker; Comeron; Elyashiv2016-vt; Russ; Coop**). Thus there is reason to be hopeful that we could detect this from temporal data, which would help to resolve the time-scales that linked selection act over. In our reanalysis of the **Barghi2019-qy** 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 tyo have a much richer picture of linked selection as temporal sequencing becomes more common.

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.

## 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 for helpful discussions.

## Appendix

### Estimator Bias Correction

#### 6.1 Correcting variance bias with a single depth sampling process

Following **Waples1989-sj**, 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  $\text{Var}(p_0) = p_0(1-p_0)/d_0$  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,

$$\text{Var}(\tilde{p}_t) = \mathbb{E}(\text{Var}(\tilde{p}_t|p_t)) + \text{Var}(\mathbb{E}(\tilde{p}_t|p_t)) \quad (3)$$

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

Under a drift-only process,  $\text{Var}(p_t) = p_0(1-p_0) \left[1 - \left(1 - \frac{1}{2N}\right)^t\right]$ . However, with heritable variation in fitness, we need to consider the covariance in allele frequency changes across generations (**Buffalo2019-io**). We can write

$$\text{Var}(p_t) = \text{Var}(p_0 + (p_1 - p_0) + (p_2 - p_1) + \dots + (p_t - p_{t-1})) \quad (5)$$

$$= \text{Var}(p_0 + \Delta p_0 + \Delta p_1 + \dots + \Delta p_{t-1}) \quad (6)$$

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

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  $\text{Var}(p_0) = 0$ , as we treat  $p_0$  as a fixed initial frequency. We can write,

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

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**)

$$\text{Var}(p_t) = \sum_{i=0}^{t-1} \text{Var}(\Delta_D p_i) + \sum_{i=0}^{t-1} \text{Var}(\Delta_H p_i) + \sum_{0 \leq i < j}^{t-1} \text{Cov}(\Delta p_i, \Delta p_j). \quad (9)$$

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

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

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

$$\sum_{i=0}^{t-1} \text{Var}(\Delta_D p_i) = p_0(1-p_0) \sum_{i=0}^{t-1} \frac{1}{2N} \left(1 - \frac{1}{2N}\right)^i \quad (11)$$

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

which is the usual variance in allele frequency change due to drift. Then, the total allele frequency change from generations 0 to  $t$  is  $\text{Var}(\tilde{p}_t - \tilde{p}_0) = \text{Var}(\tilde{p}_t) + \text{Var}(\tilde{p}_0) - 2\text{Cov}(\tilde{p}_t, \tilde{p}_0)$ , 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  $\text{Cov}(\tilde{p}_t, \tilde{p}_0) = p_0(1-p_0)/2N$  for Plan I sampling procedures (**Waples1989-sj**), we write,

$$\text{Var}(\tilde{p}_t - \tilde{p}_0) = \text{Var}(\tilde{p}_t) + \text{Var}(\tilde{p}_0) - 2C \text{Cov}(\tilde{p}_t, \tilde{p}_0) \quad (13)$$

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

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

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

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

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; 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.  $\frac{d_t}{d_t-1} \tilde{p}_t(1-\tilde{p}_t)$ ,

$$\frac{d_0-1}{d_0} \frac{\text{Var}(\tilde{p}_1 - \tilde{p}_0)}{\tilde{p}_0(1-\tilde{p}_0)} - \frac{(d_0-1)}{d_0(d_1-1)} \frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{p}_0(1-\tilde{p}_0)} - \frac{1}{d_0} + \frac{C}{N} = \frac{\text{Var}(\Delta_H p_0)}{p_0(1-p_0)} + \frac{1}{2N} \quad (18)$$

## 6.2 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  $X_t \sim \text{Binom}(n_t, p_t)$  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  $Y_t \sim \text{Binom}(d_t, X_t/n_t)$  reads have the tracked allele. We let  $\tilde{p}_t = Y_t/d_t$  be the observed sample allele frequency. Then, the sampling noise is

$$\text{Var}(\tilde{p}_t | p_t) = \mathbb{E}(\text{Var}(\tilde{p}_t | X_t)) + \text{Var}(\mathbb{E}(\tilde{p}_t | X_t)) \quad (19)$$

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

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

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

$$+ \sum_{0 \leq i < j}^{t-1} \text{Cov}(\Delta p_i, \Delta p_j) \quad (23)$$

Through the law of total expectation (see **Kolaczowski2011-ee** Supplementary File 1 for a sample proof), one can find that an unbiased estimator of the half the heterozygosity is

$$\frac{n_t d_t}{(n_t - 1)(d_t - 1)} \tilde{p}_t (1 - \tilde{p}_t). \quad (24)$$

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

$$\begin{aligned} \text{Var}(\tilde{p}_t - \tilde{p}_0) = & \frac{n_t d_t \tilde{p}_t (1 - \tilde{p}_t)}{(n_t - 1)(d_t - 1)} \left( \frac{1}{n_t} + \frac{1}{d_t} - \frac{1}{n_t d_t} \right) + \frac{n_0 d_0 \tilde{p}_0 (1 - \tilde{p}_0)}{(n_0 - 1)(d_0 - 1)} \left( \frac{1}{n_0} + \frac{1}{d_0} - \frac{1}{n_0 d_0} \right) + \\ & \frac{n_0 d_0 \tilde{p}_0 (1 - \tilde{p}_0)}{(n_0 - 1)(d_0 - 1)} \left[ 1 - \left( 1 - \frac{1}{2N} \right)^t \right] - \frac{C}{N} \frac{n_0 d_0 \tilde{p}_0 (1 - \tilde{p}_0)}{(n_0 - 1)(d_0 - 1)} + \\ & \sum_{i=0}^{t-1} \text{Var}(\Delta_H p_i) + \sum_{0 \leq i < j}^{t-1} \text{Cov}(\Delta p_i, \Delta p_j). \end{aligned} \quad (25)$$

(26)

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

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

### 6.4 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 ??, ??, and ?? 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  $\Delta \mathbf{F}$  which contains the allele frequency changes between adjacent generations, and has dimension  $R \times T \times L$ . The operation  $\text{flat}(\Delta \mathbf{F})$  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

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

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  $\text{flat}(\Delta \mathbf{F})$  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 in the upper and lower triangles of the matrix,

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

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,

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

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

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

where division here is elementwise. Thus,  $\mathbf{B}$  is a  $R \times (T + 1)$  matrix. As explained in Supplementary Material Section ?? and ??, 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

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

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

## 6.5 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,  $\Delta p_{t_0}, \Delta p_{t_1}, \dots, \Delta p_{t_6}$ . Consequently, the ten temporal covariance matrices for each of the ten replicate populations have off-diagonal elements of the form  $\text{Cov}(\Delta p_{t_0}, \Delta p_{t_1}) = \text{Cov}(p_{t_1} - p_{t_0}, p_{t_2} - p_{t_1}) = \sum_{i=0}^{10} \sum_{j=10}^{20} \text{Cov}(\Delta p_i, \Delta p_j)$ . Each diagonal element has the form  $\text{Var}(\Delta p_{t_0}) = \sum_{i=0}^{t_0} \text{Var}(\Delta p_i) + \sum_{i \neq j}^{t_0} \text{Cov}(\Delta p_i, \Delta p_j)$ , 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.

## 6.6 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  $i \in \{1, 2, \dots, W\}$  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,

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

Note that computing the ratio of averages rather than the average of a ratio is a practice common for population genetic statistics like  $F_{ST}$  (**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,

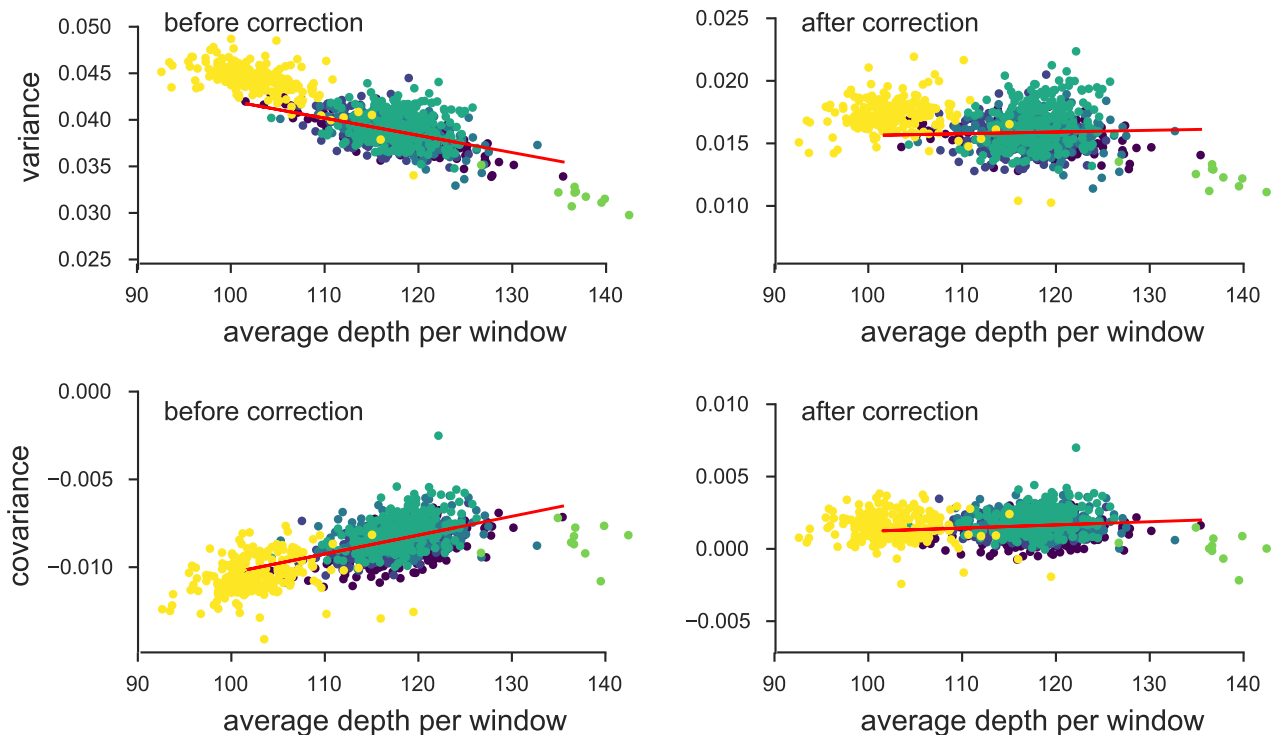
$$C_\alpha = \left( 2\hat{\theta} - q_{1-\alpha/2}, 2\hat{\theta} - q_{\alpha/2} \right). \quad (33)$$

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

## 7 Supplementary Figures

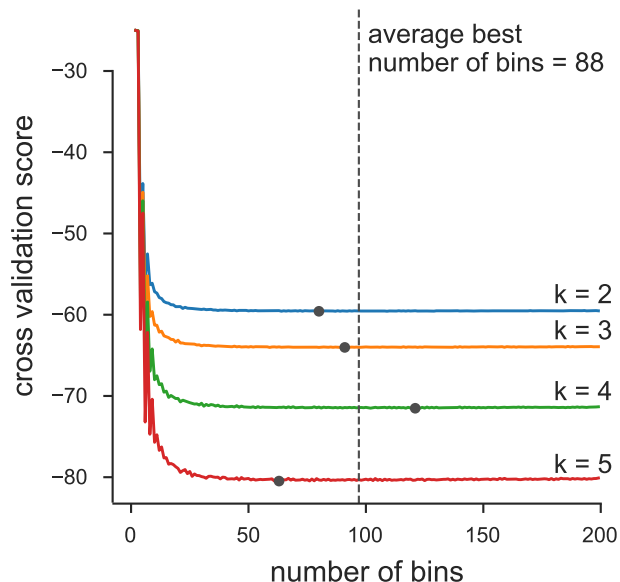
### 7.1 Bias Correction for Barghi2019-qy

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 ???. Overall, we find the biased-correction procedure removes the relationship between variance and covariance and depth, indicating it is working adequately.

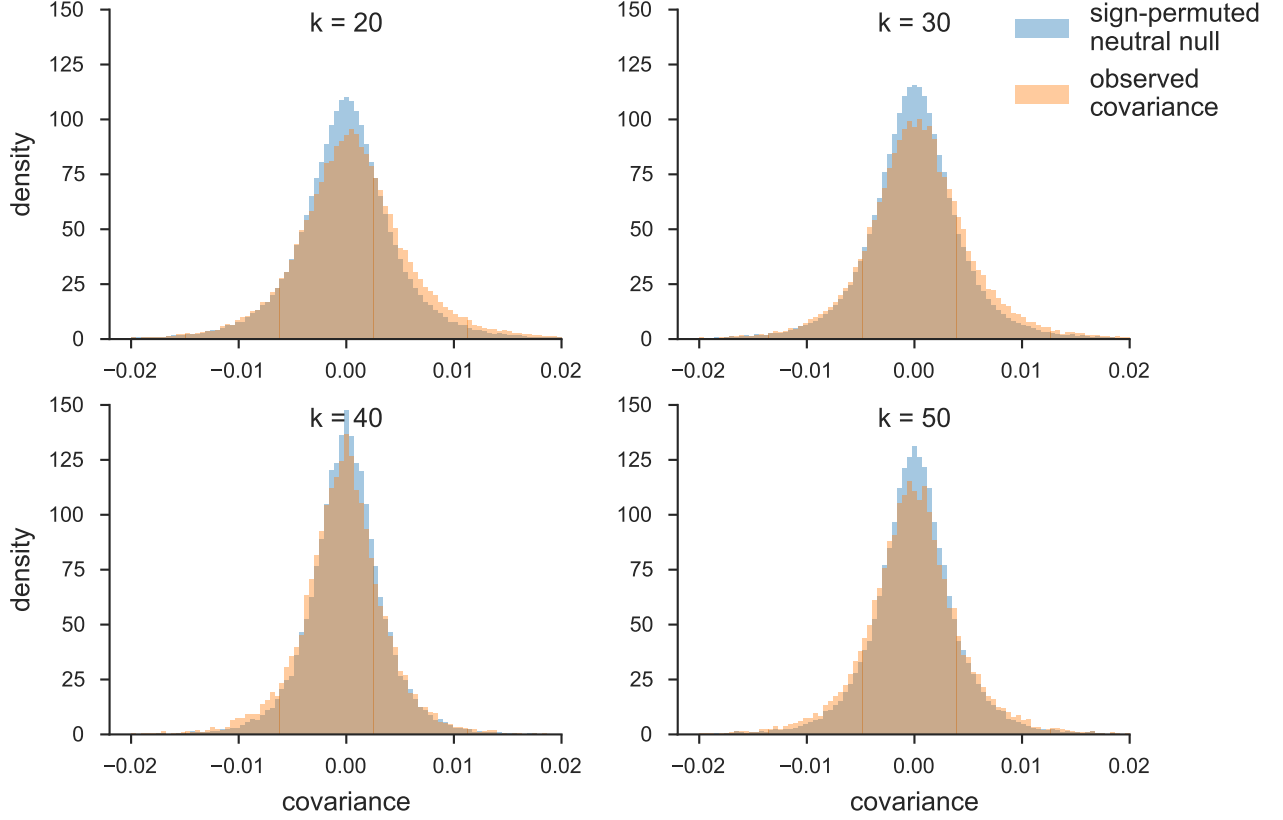


**Figure 4:** The variance and covariances from the **Barghi2019-qy** study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. Overall, the 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.

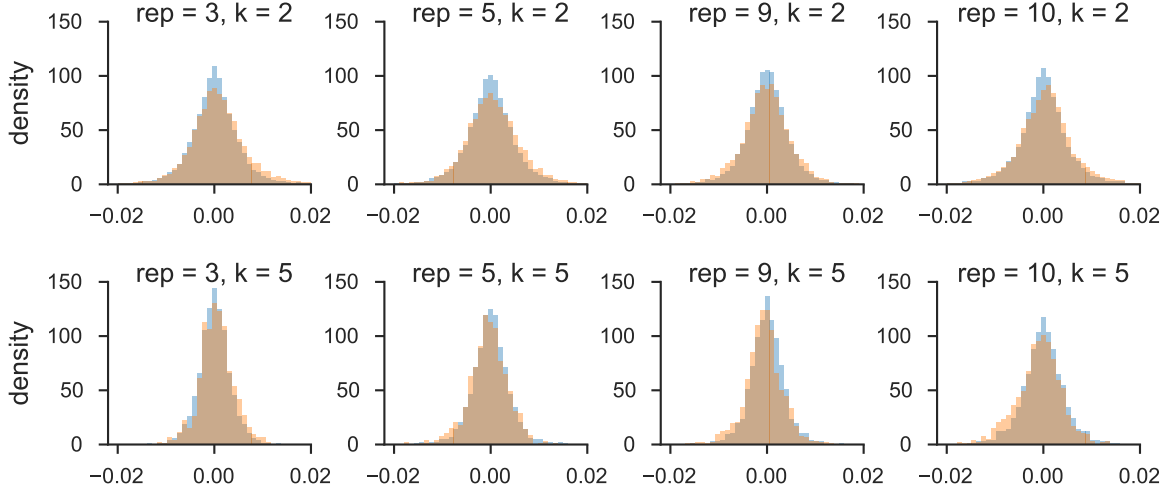




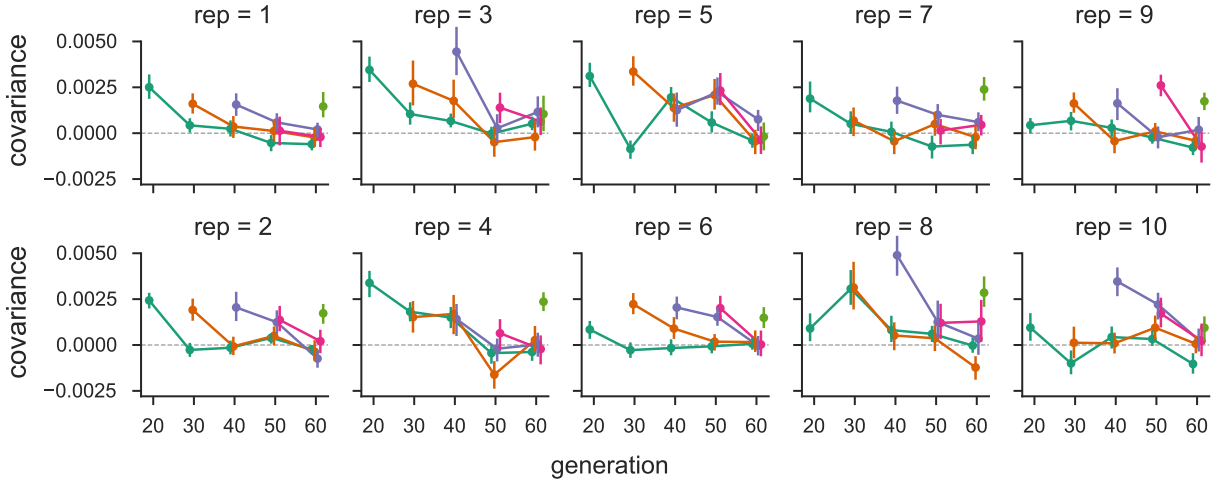
**Figure 5:** We chose number of bins used in the histograms of Figure 3 via an analytic expression for the cross-validation risk, based on the equation 6.16 of (Wasserman2006-jl, p. 129). Above, we plot the cross-validation risk for various numbers of bins, for each of the four off-diagonals of the 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.

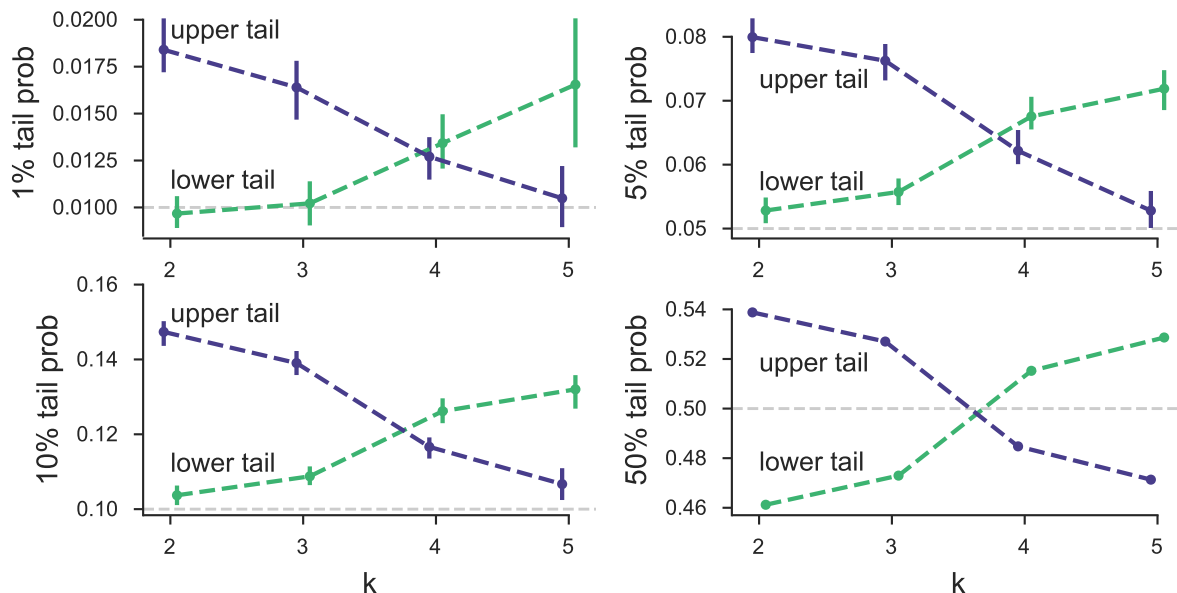


**Figure 7:** The distribution of windowed temporal covariances alongside the empirical neutral null for five randomly sampled replicates (columns), for  $k = 2$  (first row) and  $k = 5$  (second row). The main figure of the paper pools all replicate window and empirical neutral null covariances; we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

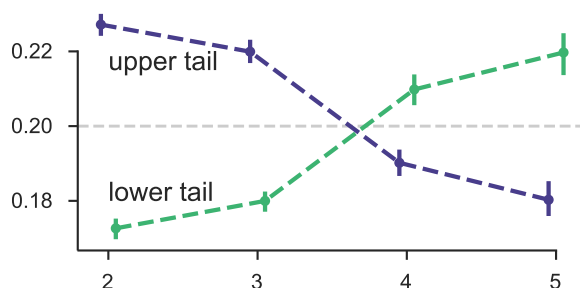


**Figure 8:** The temporal covariances from the **Barghi2019-qy** study, for each replicate individually. As in Figure 1, each line follows the temporal covariances from some initial reference generation through time, which represent the rows of temporal covariance matrix.

### 399 7.3 Barghi2019-qy Tail Probabilities for Windowed Covariances Distributions



**Figure 9:** Barghi2019-qy tail probabilities compared to sign-permuted empirical null distribution for various  $\alpha$  levels.

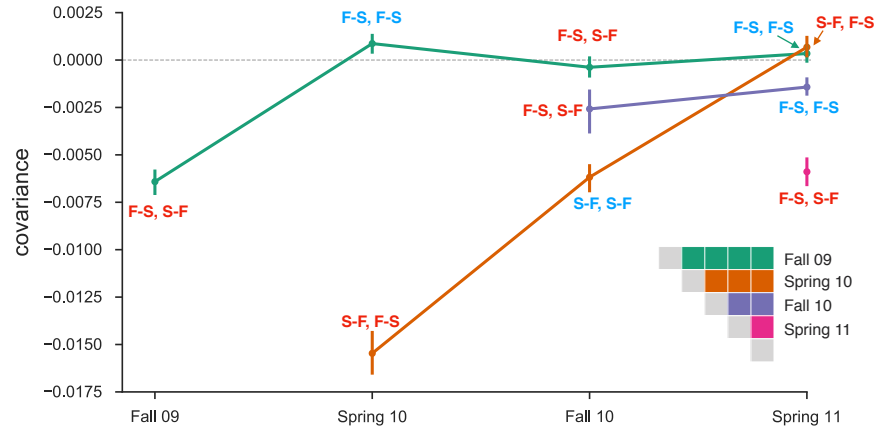


**Figure 10:** The 20% lower and upper tail probabilities for the observed windowed covariances from the **Barghi2019-qy** study, based on sign-permuting at the chromosome level. This permutation empirical null is robust to long-range linkage disequilibrium acting over entire chromosomes.

### 400 7.4 Bergland2014-ij Re-Analysis

401 We also applied our temporal covariance approach to **Bergland2014-ij**, which found evidence of  
 402 genome-wide fluctuating selection between Spring and Fall seasons across three years *Drosophila*  
 403 *melanogaster*. As described in **Buffalo2019-io**, if fluctuating selection pressure among time-periods  
 404 are the dominant genome-wide pattern, we might expect positive covariances between like seasons  
 405 changes (e.g. Spring 2010 to Fall 2010 and Spring 2011 to Fall 2011), and negative covariances

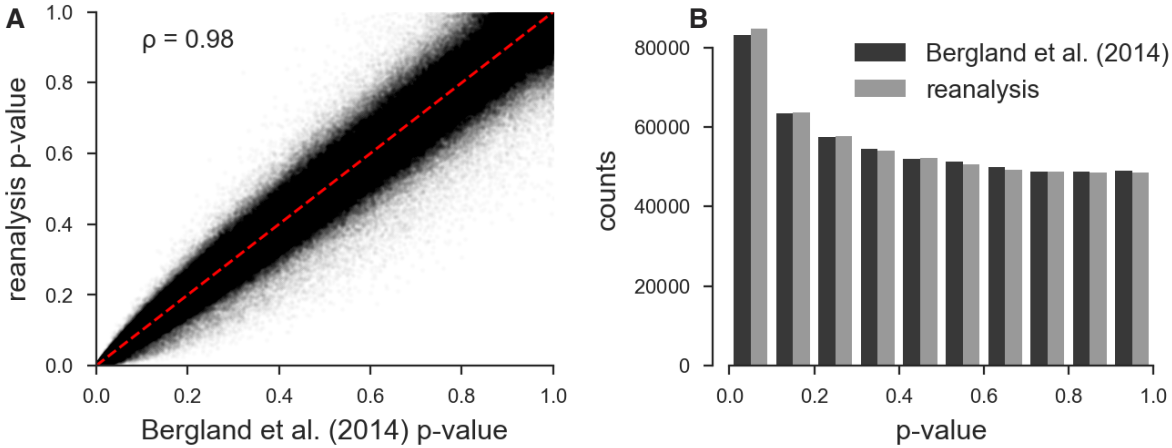
between dislike seasonal changes (e.g. Fall 2009 to Spring 2010 and Fall 2010 to Spring 2011). However, while we find temporal covariances that are non-zero, we find only weak support for a seasonal fluctuating model driving these covariances. In Supplementary Figure ??, we show the temporal covariances 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, the all other temporal covariances do not fit the pattern we would expect under genome-wide fluctuating selection.



**Figure 11:** Temporal covariances from the **Bergland2014-ij** 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 ??) as expected.

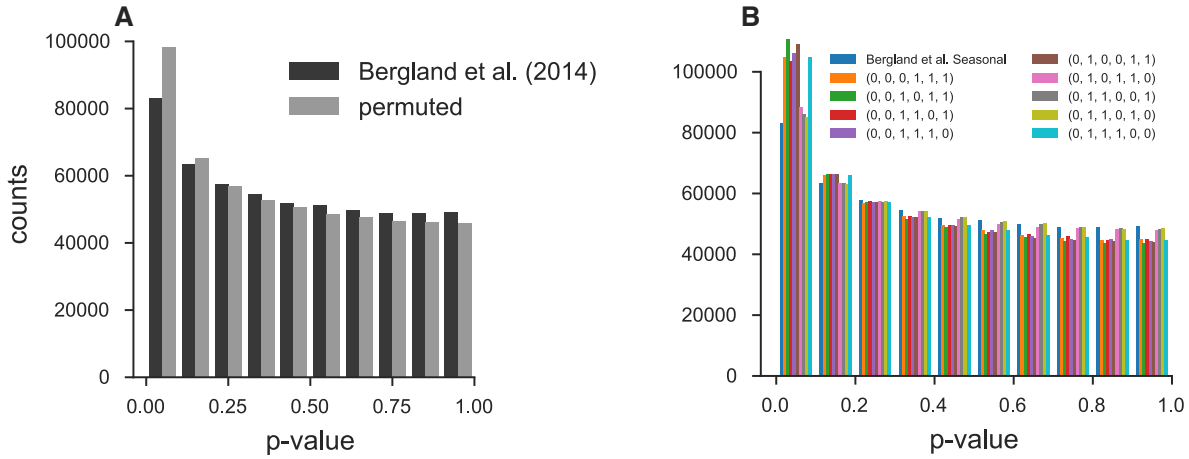
It is unclear how strong the fluctuations would have to be to generate a genome-wide average signal of fluctuating selection from temporal covariances. For example, many loci could still show a signal of fluctuating selection, but the average signal could be overwhelmed by other signals of other selection. To investigate whether there was an genome-wide excess of loci showing evidence of fluctuating selection we reanalyzed the data of (**Bergland2014-ij**) using the same seasonal fluctuating



**Figure 12:** **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.

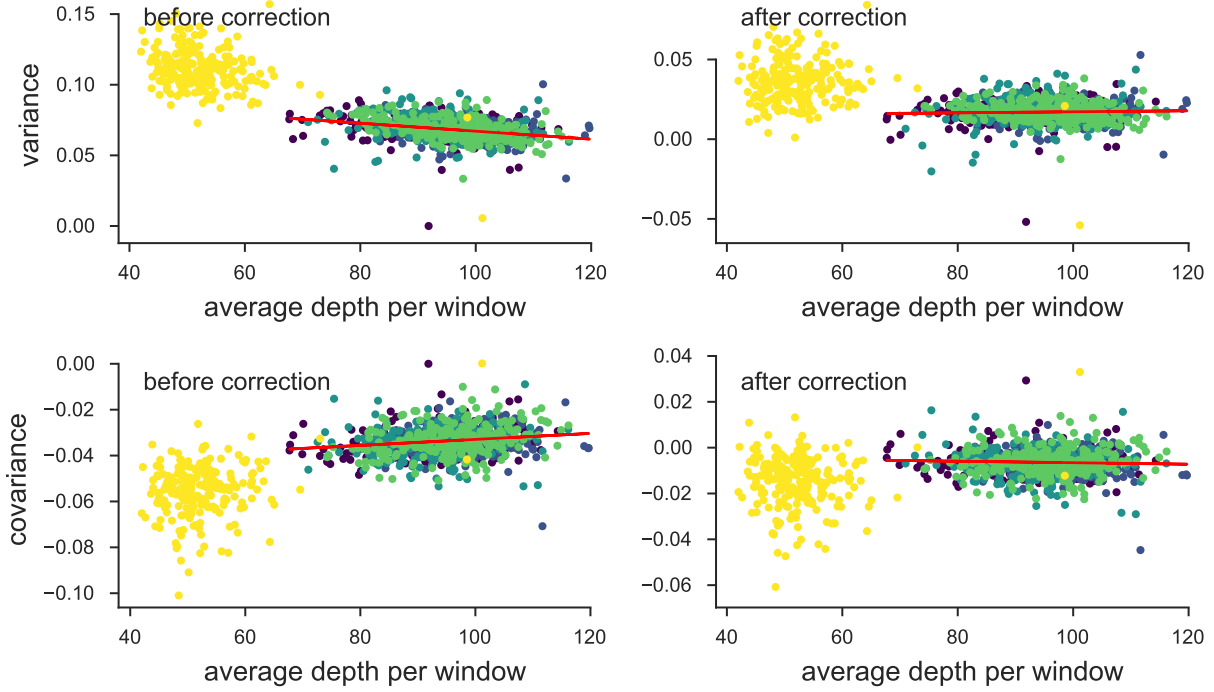
model as the original paper. This model is a Binomial logit-linked GLM fit per-locus, where the Spring/Fall seasons are encoded as a dummy variable, and are regressed on the frequency data. We use the same binomial weighting procedure as **Bergland2014-ij**, where the weights are determined by the effective 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 ?? A), and the histograms of the p-values are nearly identical (Supplementary Figure ?? 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 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 **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 ??, 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



**Figure 13: A:** Histogram of original **Bergland2014-ij** seasonal p-values and p-values creating by randomly permuting the seasons at each locus. **B:** Histogram of original **Bergland2014-ij** p-values alongside all unique permutations (ignoring symmetries that lead to identical p-values).

451 ??), where we see strong evidence of selection (non-zero temporal covariances), yet the pattern  
 452 does not follow that of seasonal fluctuating selection.



**Figure 14:** 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.