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 9, 2019

Abstract

Adaptation can proceed in natural populations and *selection experiments over* very short timescales. However, detecting these changes genome-wide is difficult, as rapid adaptation often proceeds from standing variation or is the result of polygenic selection, 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 their random selected background over time.* Since genetic drift does not lead to temporal covariance, we can *use these covariances to* begin to directly estimate what fraction of the variation in allele frequency change through time is drift 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. Additionally, taking advantage of selection on replicated populations, we can directly estimate the covariance in allele frequency between replicates due to convergent selection pressure. *We also identify signals of negative temporal covariance that show that selection reverses its direction 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 remarkable impact on short term allele frequency dynamics that previously may have been attributed to genetic drift

1 Introduction

A long-standing problem in evolutionary genetics is quantifying the roles genetic drift and selection have in shaping genome-wide allele frequency changes. Selection can both directly and indirectly affect allele frequency change, with the indirect effect coming from the action of selection on correlated loci elsewhere in genome e.g. linked selection (Maynard Smith and Haigh 1974, Charlesworth et al. 1993; Nordborg et al. 1996; see Barton 2000 for a review). Previous work on this question has mostly focused on teasing apart the impacts of drift and selection on genome-wide diversity using population samples from a single contemporary timepoint, often by modeling the correlation between regional recombination rate, gene density, and diversity created in the presence of linked selection (Begun and Aquadro 1992; Elyashiv et al. 2016). This approach *has shown linked selection has a major role in shaping patterns of genome-wide diversity across the genomes of some of organisms* (Andersen et al. 2012; Begun et al. 2007; Beissinger et al. 2016; Cutter and Choi 2010; Sattath et al. 2011; Williamson et al. 2014), and has allowed us to quantify the relative

influence of positive selection (hitchhiking) and *negative selection (background selection; Elyashiv et al. 2016; Hernandez et al. 2011; McVicker et al. 2009; Nordborg et al. 2005*. However, we lack an understanding of how genome-wide linked selection acts over an important dimension: time.

Presently, there are numerous examples of rapid phenotypic adaptation (Franks et al. 2007; Grant and Grant 2006, 2011; Reznick et al. 1997) and rapid genomic evolution in asexual populations (Baym et al. 2016; Bennett et al. 1990; Good et al. 2017). Yet the polygenic nature of fitness makes detecting the impact selection has 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 proportional to their effect sizes, which 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 (Turner and Miller 2012; Turner et al. 2011) and differentiate modes of selection (Barghi et al. 2019; Burke et al. 2010; Therkildsen et al. 2019). In addition these temporal-genomic studies have begun in wild populations, *with the goal of finding variants that exhibit frequency changes consistent with fluctuating selection* (Bergland et al. 2014; Machado et al. 2018). *In a previous paper, we proposed that one useful signal for understanding the genome-wide impact of polygenic linked selection detectable from temporal genomic studies* is the temporal autocovariance in allele frequency changes (Buffalo and Coop 2019). These covariances are directly estimable from temporal genomic data and are created when the loci that underly heritable fitness variation perturb linked neutral alleles; *in contrast*, when genetic drift acts alone in a closed population, these covariances are zero in expectation. *Mathematically, temporal covariances occur 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 (Barghi et al. 2019; Kelly and Hughes 2019), and one artificial selection experiment (Castro et al. 2019). We repeatedly find a signal of temporal covariance, consistent with linked selection acting to significantly perturb genome-wide allele frequency changes across the genome in a manner that other approaches would not be able differentiate from genetic drift. We estimate the lower bound of the proportion of total variation in allele frequency change caused by linked selection, and the correlation between allele frequency changes between replicate populations caused by response to convergent selection pressures. Overall, we demonstrate that linked selection has a powerful role in shaping genome-wide allele frequency changes over very short timescales.

2 Results

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

We first analyzed Barghi et al. (2019), an evolve-and-resequence study with ten replicate populations exposed to a high temperature lab environment and evolved for 60 generations, and sequenced every ten generations. Using the seven timepoints and ten replicate populations, we estimated the genome-wide 6×6 temporal covariance matrix $\mathbf{Q}$ for each of the ten replicates. Each row of these matrices represent the temporal covariance $\mathrm{Cov}(\Delta_{10} p_s, \Delta_{10} p_t)$, between the allele frequency change (*in ten generation intervals, denoted* $\Delta_{10} p_t$) in some initial reference generation s (the row of the matrix), and some later timepoint t (the column of the matrix). We corrected these matrices for biases created due to sampling noise, and normalize the entries for heterozygosity (see Supplementary Material Sections 6.2 and 6.4 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 (Buffalo and Coop 2019). Averaging across the ten replicate temporal covariances matrices, we find temporal covariances that are statistically significant (95% block bootstraps CIs do not contain zero), consistent with linked selection perturbing genome-wide allele frequency changes over very short time periods. The covariances between all adjacent time intervals are positive and then decay towards zero as we look at more distant time intervals, as expected when directional selection affects linked variants' frequency trajectories until ultimately linkage disequilibrium and the additive genetic variance for fitness *associated with neutral alleles* decays (Buffalo and Coop 2019). The temporal covariances per replicate are noisier but this general pattern holds; see Supplementary Figures 8. Barghi et al. (2019)'s design means that the covariances we see in adjacent time intervals are on average ten generations apart, and given the temporal decay in covariance we see, the covariances on shorter time-scales (*e.g. if adjacent generations had been sequenced*) may well be higher yet (see Supplementary Material Section 6.5 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 *partitioned* into variance and covariance components, $\mathrm{Var}(p_t - p_0) = \sum_{i \neq j} \mathrm{Cov}(\Delta p_i, \Delta p_j) / \mathrm{Var}(p_t - p_0)$, and the covariances are zero when drift acts alone, this is a lower bound on how much of the variance in allele frequency change is caused by linked selection (Buffalo and Coop 2019). We call this measure $G(t)$, which is *this measure of* 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 Barghi et al. (2019) leads our measure $G(t)$ to be even more conservative, since the temporal covariances with each ten-generation block between sequenced timepoints are not directly observable, and are not included in the numerator of $G(t)$. Still, we find a remarkably strong signal that greater than 20% of total variation in allele frequency change over 60 generations is directly the result of linked selection.

The replicate design of Barghi et al. (2019) also allows us to quantify another covariance: the covariance in allele frequency change between replicate populations experiencing convergent selection pressures. These between-replicate covariances are created in the same way as temporal covariances are: neutral alleles linked to a particular fitness background experience are expected to have allele frequency changes in the same direction if the selection pressures are similar. *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*

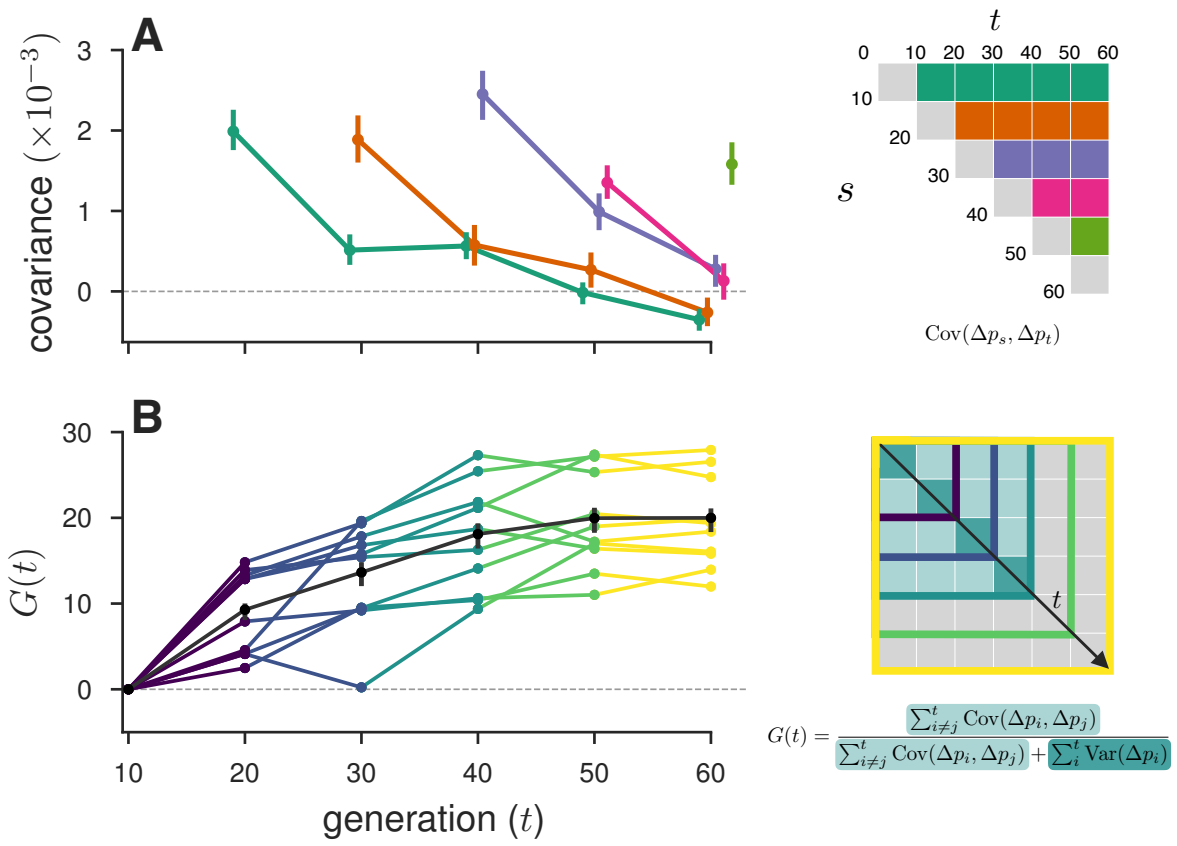

Figure 1: A: Temporal covariance, averaged across all ten replicate populations, through time from the Barghi et al. (2019) study. Each line depicts the temporal covariance $\mathrm{Cov}(\Delta p_s, \Delta p_t)$ from some reference generation s to a later time t which varies along the x-axis; each line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

fitness backgrounds common to each replicate predict (under the same selection pressures) frequency changes between replicates. We measure this through a statistic similar to a correlation, which we call the convergent correlation: the ratio of average between-replicate covariance across all pairs to the average standard deviation across all pairs of replicates,

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

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

We've calculated the convergent correlation for all rows of the replicate covariance matrices. 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.

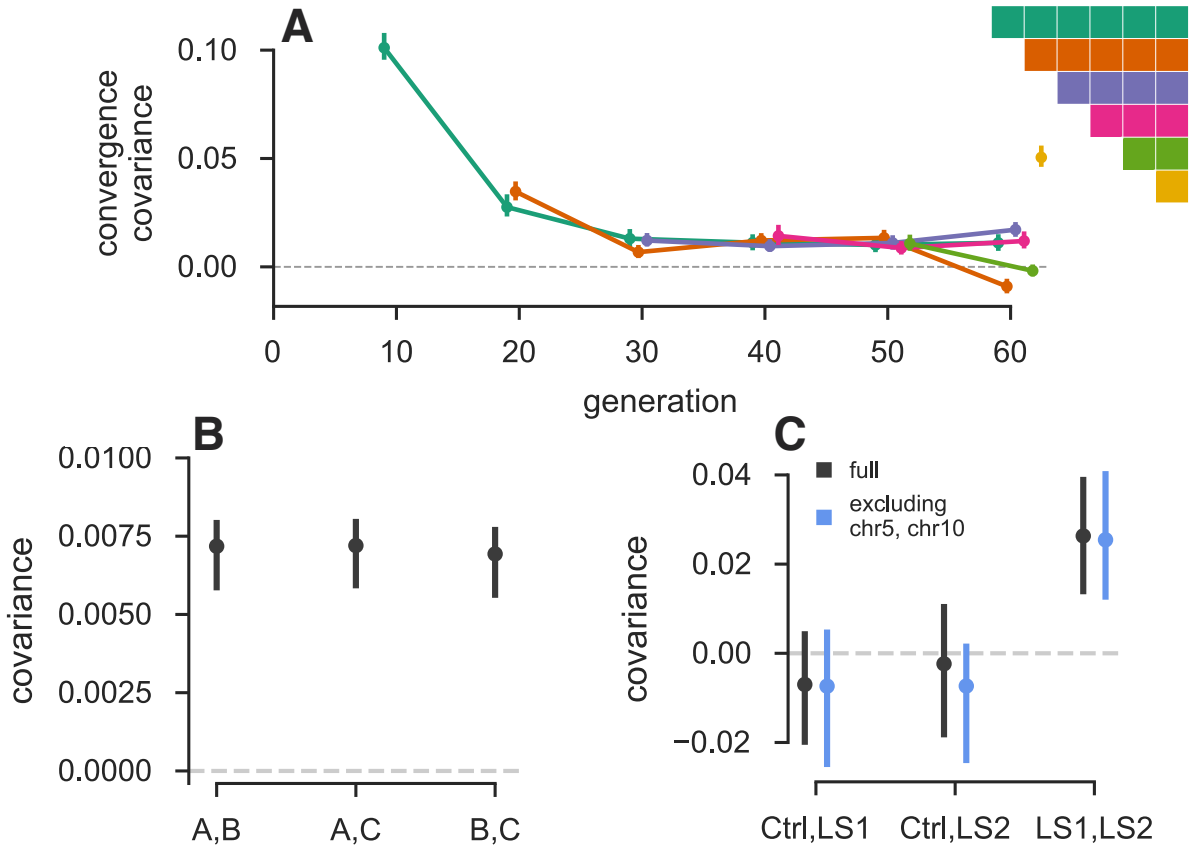

Figure 2: **A**: The convergence correlation, averaged across replicate pairs, through time. Each line represents the convergence correlation $\text{cor}(\Delta p_s, \Delta p_s)$ from a starting reference generation s to a later time t, which varies along the x-axis; each line corresponds to a row of the temporal convergence correlation matrix depicted to the right. **B**: The convergence covariance between individual pairs of replicates in the Kelly and Hughes (2019) data. **C**: The convergence covariance between individual pairs of replicates in (Castro et al. 2019) data, for the two selection lines (LS1 and LS2) and the control (Ctrl); gray CIs are those using the complete dataset, blue CIs exclude chromosomes 5 and 10 which harbor the two regions Castro et al. (2019) found to have signals of parallel selection between LS1 and LS2.

Finally, we observed that in the longest study we analyzed (Barghi et al. 2019), some genome-wide temporal covariances become negative at future timepoints (see the first two rows in Figure 1 A). This shows that alleles that were on average going up initially are later going down in frequency, i.e. that the average direction of selection experienced by alleles has flipped. This must reflect either a change in the environment or the genetic background, due to changing epistatic relationships among alleles 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 (Figure 3, pooling across all replicates; see Supplementary Figure 7 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, Baldwin-Brown et al. 2014; Nuzhdin and Turner 2013). *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 (see Supplementary Material Section 6.7 for more details on this approach). 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 10). 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.

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 us to assess the impact of linked selection in driving convergent patterns of allele frequency change across replicate populations in two other studies. First, we reanalyzed the selection experiment of Kelly and Hughes (2019), which evolved three replicate 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 $[0.31, 0.40]$).

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

3 Discussion

Since the seminal theoretic model of Maynard Smith and Haigh (1974) demonstrating that diversity is reduced at linked neutral variation as an advantageous polymorphism sweeps to fixation, over four decades of theoretical and empirical research has enriched our understanding of linked selection. This work has led to the discovery of selected genes from genomic signals of hitchhiking (Nair et al. 2003; Voight et al. 2006), found genome-wide signals of linked selection (Aguade et al. 1989; Andersen et al. 2012; Begun and Aquadro 1992; Cutter and Choi 2010; Cutter and Payseur 2003) leading to the development of recurrent hitchhiking (Stephan et al. 1992) and background selection (Charlesworth et al. 1993) models that attempt explain this pattern. Since, much attention has focused on quantifying the impact of background selection (McVicker et al. 2009) and developing theoretic machinery (Coop and Ralph 2012) to differentiate the regional reductions in diversity into those caused by full and partial sweeps, and those caused by background selection (Elyashiv et al. 2016). However, as other theoretic models of soft sweeps from standing variation emerged (Hermisson and Pennings 2005; Pennings and Hermisson 2006), attention shifted towards detecting other modes of linked selection (Pritchard et al. 2010), a more difficult endeavor (Przeworski et al. 2005) met with new coalescent models (Berg and Coop 2015) and machine learning methods (Schrider and Kern 2017). An undervalued theoretic approach to understanding the genome-wide effects of selection on standing variation, e.g. selection on an infinitesimal polygenic trait, stems from an early quantitative genetic model of linked selection (Robertson 1961) and its later developments (Santiago and Caballero 1995, 1998). 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 can be readily detected from temporal genomic data (Buffalo and Coop 2019). Here, we provide the first empirical evidence of linked selection acting on standing variation over very short timescales.

We reanalyzed three published temporal genomic datasets and detected the genome-wide signature of linked selection acting over tens of generations. Furthermore, we find the dynamics of temporal autocovariance are consistent with theory; the temporal covariance in allele frequencies decays *as both the fitness variation associated with particular haplotypes and the linkage disequilibria between selected and neutral loci decay through time*. In reanalyzing one study (Barghi et al. 2019), we find that after sixty generations, greater than 20% of the variation in allele frequency change is directly due to the action linked selection. Capitalizing on replicated evolve-and-resequence study designs, we characterized the extent to which convergent selection pressures lead to parallel changes

in allele frequency across replicate populations, finding moderate correlations in allele frequency changes.

While we must keep in mind that the studies analyzed here were all laboratory populations and natural selection in the wild will likely differ, overall we have found that linked selection has a remarkably strong effect for such short study durations. Depending on how polygenic fitness variation is, such a strong effect of linked selection may not be differentiable from genetic drift using only single contemporary population samples. In this way, 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 an infinitesimal trait. In our reanalysis of the Castro et al. (2019) data, by showing the covariance signal remains even after excluding the chromosomes containing two large-effect loci, we showed the covariance signal we detect is indeed driven by polygenic fitness variation (and confirm a point in the original study that tibiae length, aside from these two large effect loci, is an infinitesimal trait).

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 two levels of sampling noise (as individuals are sampled from the population and reads are sampled from amplified DNA), which at any locus swamps the temporal covariance signal and creates spurious covariances. However, these spurious covariances do not share a common sign across timepoints whereas the covariances created by linked selection do; consequently, averaging across the entire genome, the temporal signal exceeds sampling noise.

One limitation of our current work here is that none of the studies we reanalyzed had linkage disequilibria data for the evolved populations. Our theory of temporal autocovariance tells us that the temporal autocovariances a neutral site experiences is determined by the product of the expected linkage disequilibrium and additive fitness variation that persists through the generations (Buffalo and Coop 2019). This leads to a clear prediction: regions of higher linkage disequilibrium and lower recombination should have greater temporal autocovariance than regions with lower LD and higher recombination. Unfortunately, we lack a high-resolution recombination map for *D. simulans* and while there are LD data for *D. simulans* (Signor et al. 2018) we did not find a relationship between temporal covariance and LD.

We believe this is driven by the idiosyncratic nature of LD in evolve-and-resequence populations (Kelly and Hughes 2019; Nuzhdin and Turner 2013), and that we might find such a relationship if LD data from the evolved population were available. Unfortunately, sequencing multiple timepoints is expensive and the low-coverage and/or pooled-sequencing approaches common in these studies prohibit estimating linkage disequilibrium. 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 us better understand sites in high recombination regions are free from the effects of linked selection. Furthermore, such additional data would allow for localizing the effects of temporal covariance, which would allow us to estimate not just the genome-wide fraction of variation in frequency change ($G(t)$), but also the fraction of genome impacted by linked selection, allowing us to synthesize the temporal approach with single-timepoint studies like that of Elyashiv et al. (2016).

Thus far, the most comprehensive studies implicating linked selection in affecting genome-wide diversity have been in *Drosophila* (Begun and Aquadro 1992; Elyashiv et al. 2016; Sattath et al.

2011) and *Caenorhabditis* (Andersen et al. 2012; Cutter and Payseur 2003), whereas the evidence for linked selection in plant and yeast species is weak (Cutter and Payseur 2013). Both *Drosophila* and *Caenorhabditi* have short genetic maps (or in the case of selfing *C. elegans*, small effective recombination rates), and large effect population sizes, two factors predicted to increase the degree to which hitchhiking impacts genome-wide diversity (Barton 2000). However, our results here show that even with small effective population sizes (300, 450, and 45 for the Kelly and Hughes (2019), and Castro et al. (2019) studies respectively) around and, in the case of mice, moderately-sized genetic maps (around 14 Morgans; Cox et al. 2009), we still see a fairly strong effect of linked selection. This suggests that while these factors may govern the dynamics of classic hitchhiking events, they could perhaps play less of a role in polygenic linked selection. However, with only three cases here, further studies on different taxa are needed.

Finally, in reanalyzing the Barghi et al. (2019) study, we find evidence of complex linked selection dynamics. This is due either to a change in the environment and its consequences for the fitness of regions in the genome, recombination breaking up selected alleles, or epistatic interactions between loci. It is important to note that the original study did not intentionally alter the environment in any way; this signal is attained without intentionally seeking it out. Discerning which of these is creating negative temporal autocovariance is a sizable challenge, requiring sequencing each generation and LD data.

Overall, we hope this will encourage more temporal genomic studies in both laboratory and natural populations. Understanding the dynamics of linked selection over short timescales will help to unite phenotypic studies of rapid adaptation of polygenic adaptations with a detectable genomic signature, to address long-standing questions on linked selection, evolutionary quantitative genetics, and the overall impact selection has on genetic variation.

4 Acknowledgments

We would like to thank the authors of the original studies we've analyzed, including 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.

5 Appendix

6 Estimator Bias Correction

6.1 Correcting variance bias with a single depth sampling process

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

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

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

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 (Buffalo and Coop 2019). We can write

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

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

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

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,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 Jónás et al. 2016). Let $X_t \sim \mathrm{Binom}(n_t, p_t)$ where X_t is the count of alleles and n_t is the number of diploids sampled at time t. Then, these individuals are sequenced at a depth of d_t, and $Y_t \sim \mathrm{Binom}(d_t, {}^{X_t}/_{n_t})$ reads have the tracked allele. We let $\widetilde{p}_t = {}^{Y_t}/_{d_t}$ be the observed sample allele frequency. Then, the sampling noise is

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

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

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

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

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

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

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

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

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

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

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

(25)

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

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 $\mathrm{Cov}(\Delta\widetilde{p}_t, \Delta\widetilde{p}_{t+1}) = \mathrm{Cov}(\widetilde{p}_{t+1} - \widetilde{p}_t, \widetilde{p}_{t+2} - \widetilde{p}_{t+1})$, and thus shares the sampling noise of timepoint $t+1$. Thus acts to bias the covariance by subtracting off the noise variance term of $\mathrm{Var}(\widetilde{p}_{t+1})$, so we add the expectation of this bias, derived above, back in. We discuss this in more detail below in deriving the bias correction for the temporal-replicate variance covariance matrix.

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 6.1, 6.2, and 6.3 to multiple sampled loci and the temporal-replicate covariance matrix here. With frequency data collected at $T+1$ timepoints across R replicate populations at L loci, we have multidimensional arrays $\mathbf{F}$ of allele frequencies, $\mathbf{D}$ of sequencing depths, and $\mathbf{N}$ of the number of individuals sequenced, each of dimension $R \times (T+1) \times L$. We calculate the array $\mathbf{\Delta F}$ which contains the allele frequency changes between adjacent generations, and has dimension $R \times T \times L$. The operation flat($\mathbf{\Delta F}$) flattens this array to a $(R \cdot T) \times L$ matrix, such that rows are grouped by replicate, e.g. for timepoint t, replicate r, and locus l such that for allele frequencies $p_{t,r,l}$, the frequency change entries are

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

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

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

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

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

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

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 6.2 and 6.3, 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 $\mathbf{B}$.

6.5 Barghi et al. (2019) Temporal Covariances

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

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 (1)). In these cases, we can exploit the linearity of the expectation to make the bootstrap procedure more computationally efficient, by pre-calculating the statistics of the ratio's numerator and denominator, $N(\mathbf{x}_i)$ and $D(\mathbf{x}_i)$, on the data $\mathbf{x}_i$ for all blocks $i \in \{1, 2, \ldots, W\}$ in the genome. Then we draw W bootstrap samples with replacement, and compute the estimate for bootstrap sample b with an average weighted by the number of loci in all sampled blocks,

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

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

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

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

6.7 The Empirical Neutral Null Windowed Covariance Distribution

7 Supplementary Figures

7.1 Bias Correction for Barghi et al. (2019)

We have investigated the effectiveness of our correction on real data by exploiting the relationship between sampling depth and the magnitude of the variance and covariance biases, and comparing the observed variances and covariances before and after correction. We plot the variance and covariance (between adjacent timepoints) before and after the bias correction against the average sample depth in 100kb genomic windows in Figure 4. Overall, we find the biased-correction procedure removes the relationship between variance and covariance and depth, indicating it is working adequately.

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

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

7.4 Bergland et al. (2014) Re-Analysis

We also applied our temporal covariance approach to Bergland et al. (2014), which found evidence of genome-wide fluctuating selection between Spring and Fall seasons across three years *Drosophila melanogaster*. As described in Buffalo and Coop (2019), we might expect positive covariances between like seasons changes (i.e. Spring 2010 to Fall 2010 and Spring 2011 to Fall 2011), and negative covariances between dislike seasonal changes (i.e. Fall 2009 to Spring 2010 and Fall 2010 to Spring 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 14, 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.

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 13) 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 (Bergland et al. 2014) using the same seasonal fluctuating model as the original paper. This model is a Binomial logit-linked GLM fit per-locus, where the Spring/Fall seasons are encoded as a dummy variable, and are regressed on the frequency data. We use the same binomial weighting procedure as Bergland et al. (2014), where the weights are determined by the effective number of chromosomes, $N_{eff} = (2n_t d_t - 1)/(2n_t + d_t)$ (n_t and d_t are the number of diploid individuals and the read depth at timepoint t, respectively). We fit this model on all loci marked as used in the VCF provided with the Bergland et al. (2014) study (doi:10.5061/dryad.v883p). Overall, our p-values for the Wald test for each locus closely match those of the original paper (Pearson correlation coefficient 0.98, p-value $< 2.2 \times 10^{-16}$; see Supplementary Figure 11 A), and the histograms of the p-values are nearly identical (Supplementary Figure 11 B). Bergland et al. (2014) find loci with a significant association with season after a Benjamini and Hochberg FDR p-value adjustment (Benjamini and Hochberg 1995), *however, the null hypothesis of the Wald test does not give us an idea of the expected number of variants that may spuriously fit the pattern of seasonal fluctuating selection.*

To investigate whether there is a genome-wide evidence of an enrichment of fluctuating selection we created an empirical null distribution by randomly permuting the season labels and re-running the per-locus seasonal GLM model, as proposed by Machado et al. (2018). We find, regardless of whether we permute at the locus-level or the permutation replicate-level, that the observed seasonal p-value distribution Bergland et al. (2014) is not enriched for significant p-values beyond what we would expect from the permutation null. In fact, there appears there is more enrichment for low p-values when seasonal labels are randomly permuted (Supplementary Figure 12, suggesting by random chance we might expect more variants with a seasonal fluctuating pattern than found in the original Bergland et al. (2014) study. While surprising, this could be explained by the presence of temporal structure across the samples not consistent with seasonal fluctuating selection. Some fraction of the 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 14), where we see strong evidence of selection (non-zero temporal covariances), yet the pattern does not follow that of seasonal fluctuating selection.

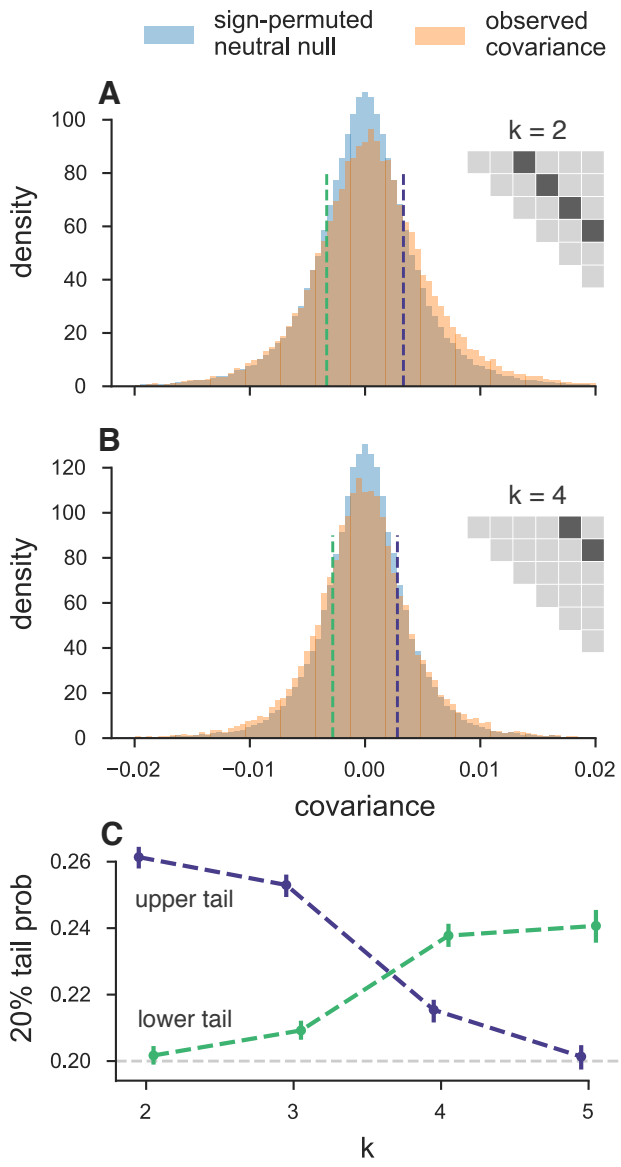

Figure 3: **A**, **B**: The distribution of temporal covariances calculated in 100kb genomic windows from the Barghi et al. (2019) study, plotted alongside an empirical neutral null distribution created by recalculating the windowed covariances on a 1,000 sign permutations of allele frequency changes within tiles. The histogram 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 **A** the covariances are separated by $k = 4 \times 10$ generations; each k is an off-diagonal from the variance diagonal of the temporal covariance matrix (see cartoon of upper-triangle of covariance matrix in subfigures **A** and **B**, where the first diagonal is the variance, and the dark gray indicates which off-diagonal of the covariance matrix is plotted in the histograms). **C**: The lower and upper tail probabilities of the observed windowed covariances, at 20% and 80% quintiles of the empirical neutral null distribution, for varying time between allele frequency changes (i.e. which off-diagonal k). The confidence intervals are 95% block-bootstrap confidence interval, and the light gray dashed line indicates the 20% tail probability expected under the neutral null.

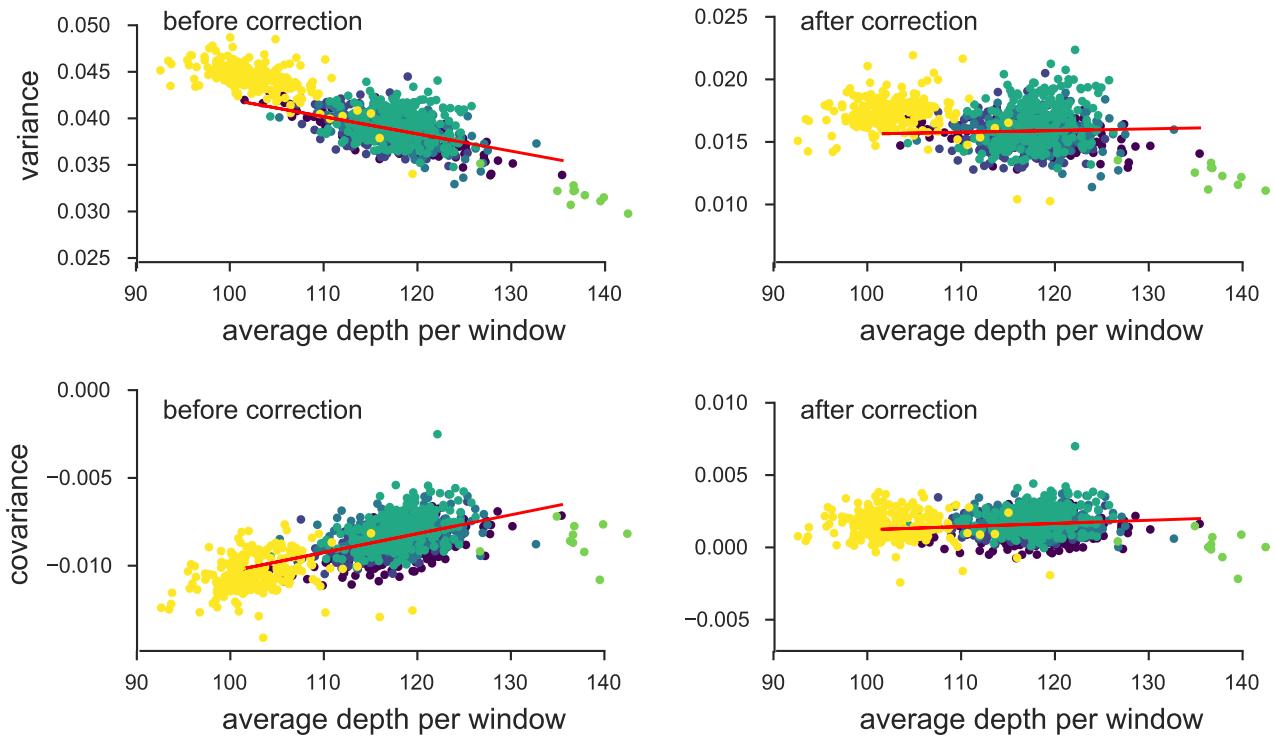

Figure 4: The variance and covariances from the Barghi et al. (2019) study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. Overall, the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

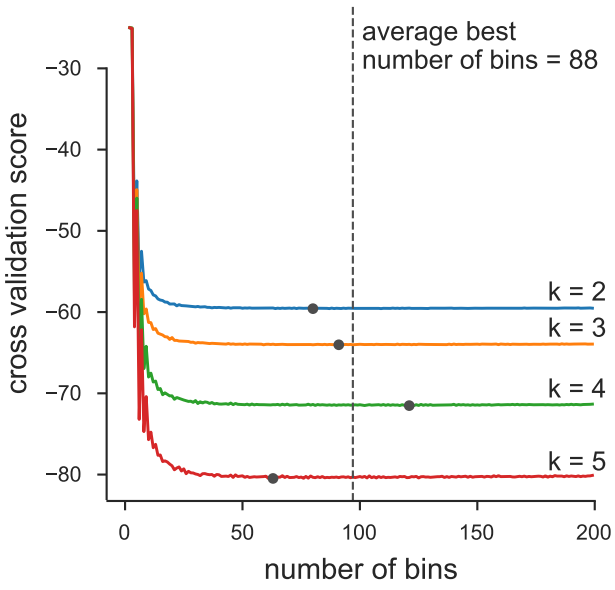

Figure 5: We chose number of bins used in the histograms of Figure 3 via an analytic expression for the cross-validation risk, based on the equation 6.16 of (Wasserman 2006, p. 129). Above, we plot the cross-validation risk for various numbers of bins, for each of the four off-diagonals of the temporal covariance matrix that we analyze. Overall, because the number of data points is large, oversmoothing is less of a problem, leading the cross-validation risk to be relatively flat across a large number of bins. Each gray point indicates the minimal risk for a particular off-diagonal, and the dashed line indicates the best average binwidth across off-diagonals.

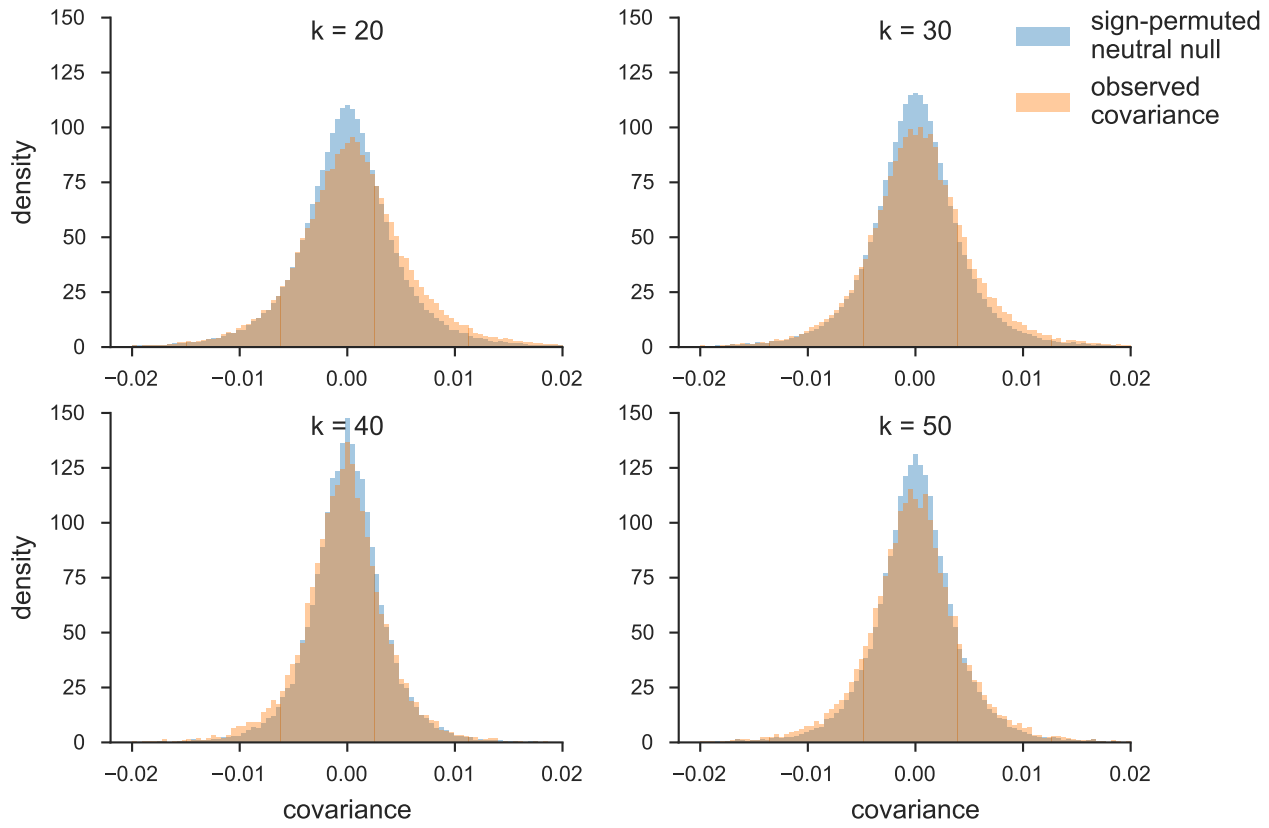

Figure 6: The distribution of temporal covariances calculated across 100kb genomic windows from Barghi et al. (2019)'s study (orange) and the block sign permuted empirical neutral null distribution of the windowed covariances (blue). Each panel shows these windowed covariances and the empirical null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

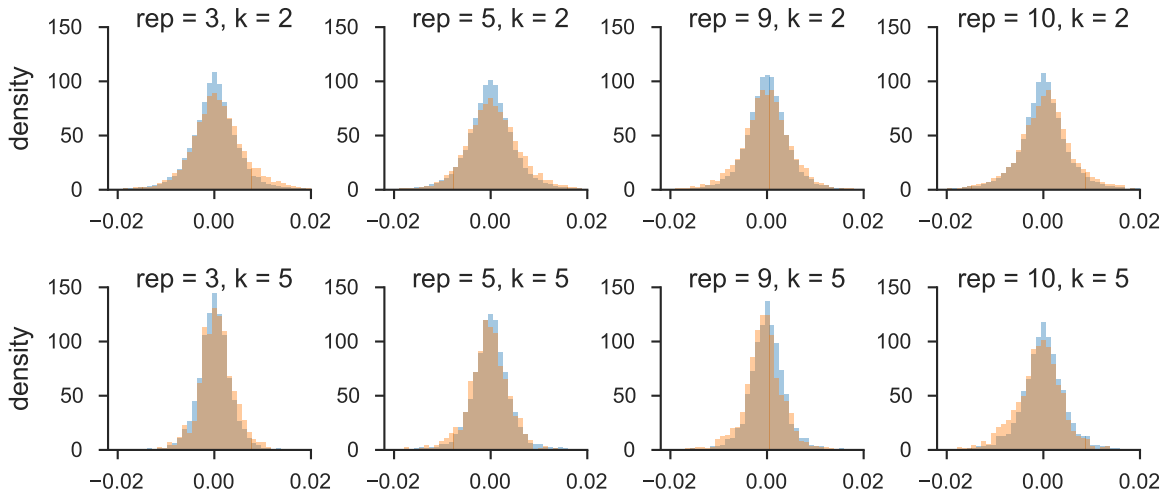

Figure 7: The distribution of windowed temporal covariances alongside the empirical neutral null for five randomly sampled replicates (columns), for $k = 2$ (first row) and $k = 5$ (second row). The main figure of the paper pools all replicate window and empirical neutral null covariances; we show here the windowed temporal covariances tend to shift from being positive (a heavier right tail) to become more negative (a heavier left tail) through time within particular replicates.

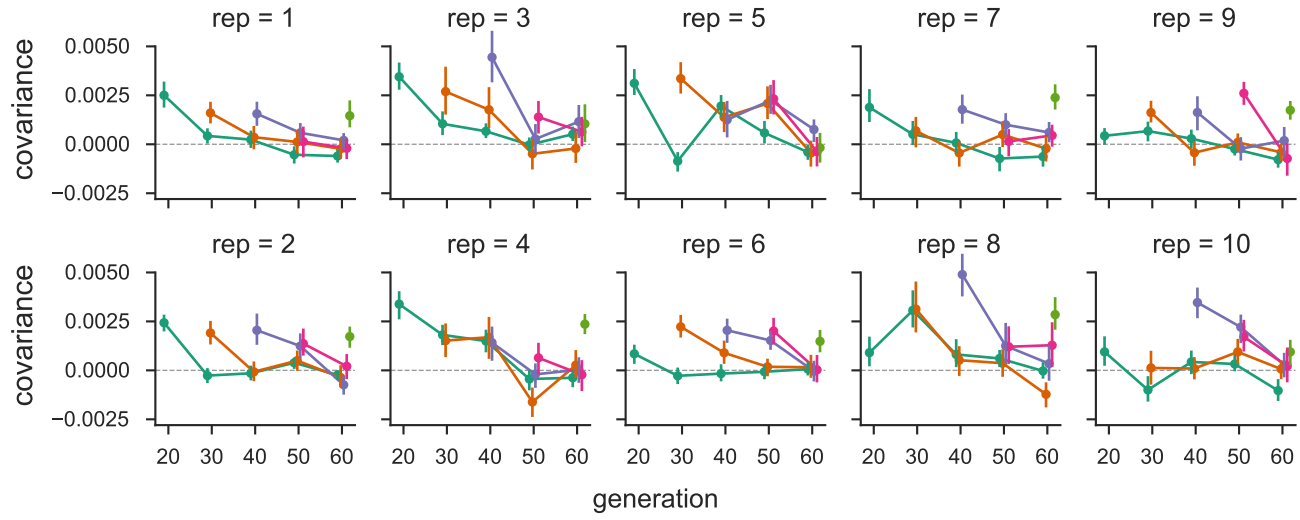

Figure 8: The temporal covariances from the Barghi et al. (2019) study, for each replicate individually. As in Figure 1, each line follows the temporal covariances from some initial reference generation through time, which represent the rows of temporal covariance matrix.

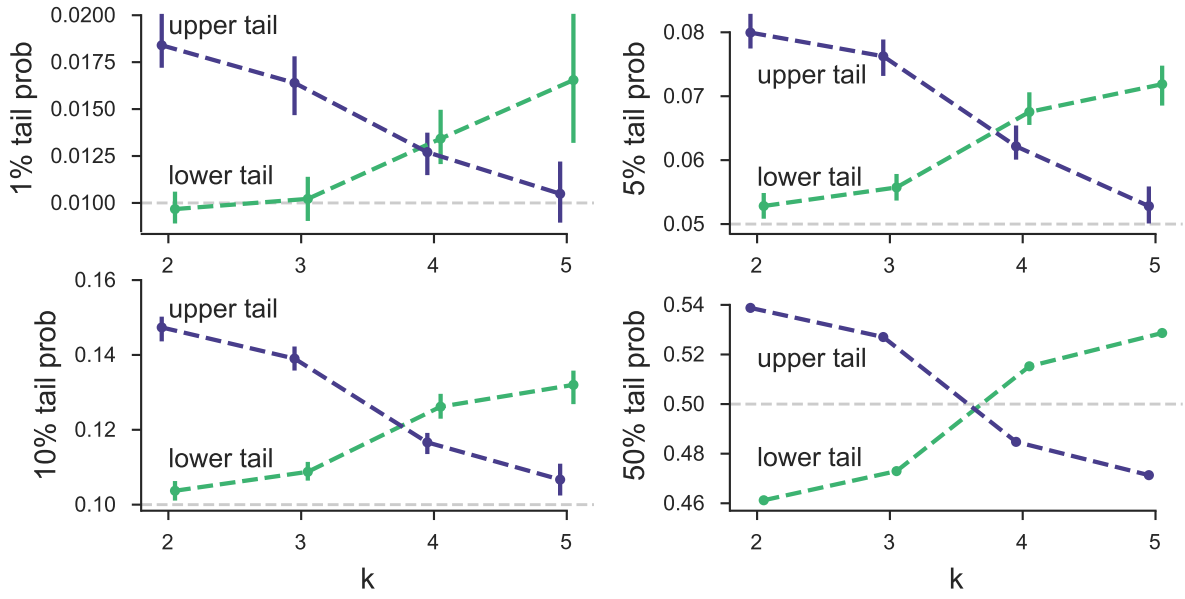

Figure 9

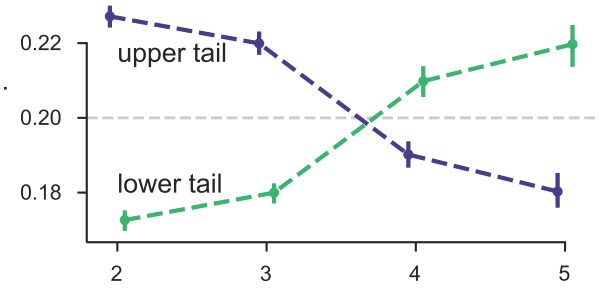

Figure 10: The 20% lower and upper tail probabilities for the observed windowed covariances from the Barghi et al. (2019) study, based on sign-permuting at the chromosome level. This permutation empirical null is robust to long-range linkage disequilibrium acting over entire chromosomes.

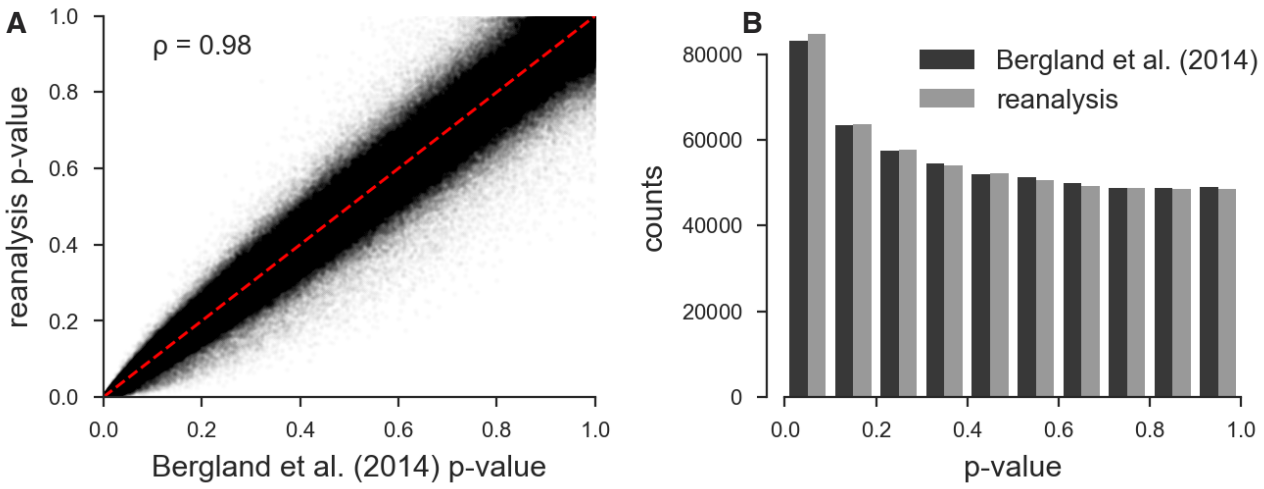

Figure 11: A: Scatterplot of the original unadjusted p-values from Bergland et al. (2014) and the p-values from our reanalysis of the same data using the same statistical methods; the minor discrepancy is likely due to software version differences. B: The histograms of the p-values of our reanalysis and the original Bergland et al. (2014) data; again the minor discrepancy is likely due to software differences. Overall, our implementation of Bergland et al.'s statistical methods produces results very close to the original analysis.

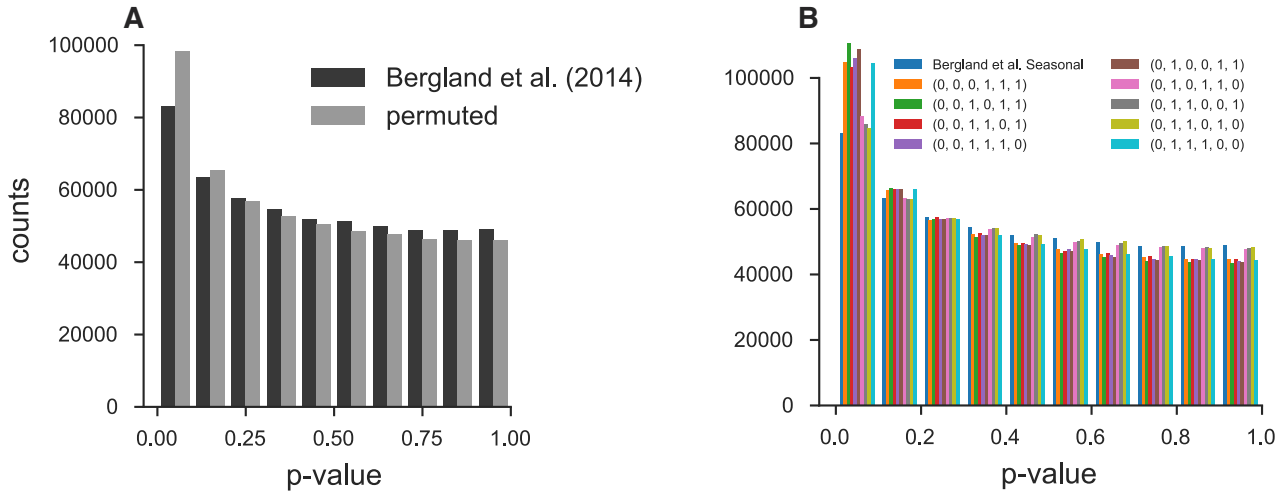

Figure 12

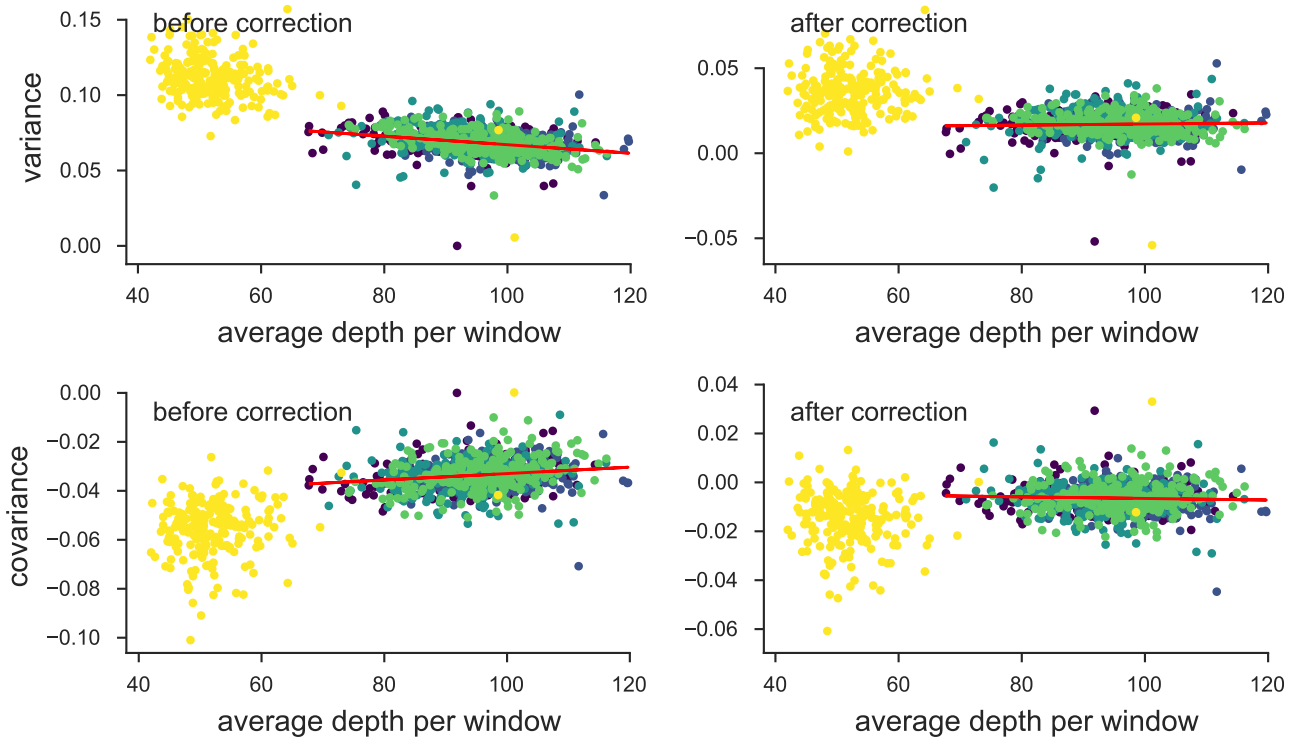

Figure 13: The variance and covariances from the Bergland et al. (2014) study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. The bias correction procedure is correcting sampling bias in both the variance and covariance such that the relationship with depth is constant. Colors indicate the different chromosomes of *D. melanogaster*; we have excluded the X chromosome (yellow points; chromosome 4 was not in the original study) from the regression due to large differences in average coverage.

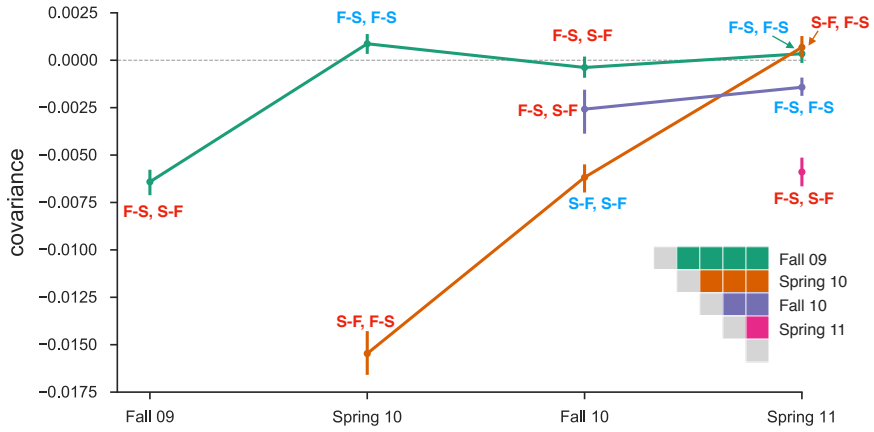

Figure 14: Temporal covariances from the Bergland et al. (2014) study, from varying reference generations (e.g. rows along the temporal covariance matrix). Each covariance is labeled indicating whether the covariance is between two like seasonal transitions (e.g. the covariance between allele frequency changes from fall to spring in one year, and fall to spring in another or two dislike seasons (e.g. the covariance between fall to spring in one year, and spring to fall in another year). Covariances between like transitions are expected to be positive when there is a genome-wide effect of fluctuating selection (and these labels are colored blue), while covariances between dislike transitions are expected to be negative (and these labels are colored red). 95% confidence intervals were constructed by a block-bootstrapping procedure where the blocks are megabase tiles.