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

Vince Buffalo*,†,1 and Graham Coop†

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

May 6, 2020

Abstract

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

Significant Statement

A long-standing problem in evolution biology is to understand the different processes that shape the genetic composition of populations. In a population closed to migrants, the two processes that change *allele frequencies* are selection, which systematically increases beneficial alleles and removes *deleterious* ones, and genetic drift which stochastically changes allele frequencies as some parents have more or less copies of specific alleles to the next generation. Previous efforts to disentangle these two processes have often used population genomic samples from a single timepoint and specific models of how selection at one allele affects its neighbors (known as linked selection). Here, we use genomic data taken at multiple timepoints to ~~*directly*~~ quantify *the relative contributions of* selection

and drift *to* allele frequency *changes* at a genome-wide level through time. We show selection acts over short timescales in three evolve-and-resequence studies and has a sizeable impact on genome-wide allele frequency changes.

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 affect allele frequencies, *both directly and indirectly,* with the indirect effect coming from the action of selection on *linked* 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 (**Cutter2013-ba**; **Sella2009-nx**). This approach has shown linked selection has a major role in shaping patterns of genome-wide diversity across the genomes of a range of sexual species (**Macpherson2007-qt**; **Andolfatto2007-uy**; **Begun2007-bg**; **Beissinger2016-cm**; **Sattath2011-dr**; **Williamson2014-oy**; **Andersen2012-bj**; **Cutter2010-gi**; **Elyashiv2016-vt**), and has allowed us to quantify the relative influence of positive selection (hitchhiking) and negative selection (background selection; **Nordborg2005-dc**; **McVicker2009-ax**; **Andolfatto2007-uy**; **Macpherson2007-qt**; **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, selection-driven genomic evolution in asexual populations (**Good2017-om**; **Bennett1990-bc**; **Baym2016-kh**). Yet the polygenic nature of fitness makes detecting the impact of selection on genome-wide variation over short timescales in sexual populations remarkably difficult *(***Latta1998-me***;* **Pritchard2010-tk***;* **Kemper2014-bx***)*. This is because the effect of selection on a polygenic trait (such as fitness) is distributed across loci in proportion to their effect sizes. This can lead to subtle allele frequency shifts on standing variation that are difficult to distinguish from background levels of genetic drift and sampling variance. However, increasingly genomic experimental evolution studies with multiple timepoints, and in some cases multiple replicate populations, are being used to detect large effect selected loci (**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 genomic studies with multiple timepoints 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 neutral allele frequency changes

Study	Species	Selection	Replicates	Pop. Size	Gens.	Timepoints
Kelly2019-dc	*D. simulans*	lab adaptation	3	∼1100	14	2
Barghi2019-qy	*D. simulans*	lab adaptation	10	∼1000	60	7
Castro2019-uk	*M. musculus*	tibiae length	2	32	17	2
		control	1	28		

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

in one generation predict changes in another due to a shared selection pressures and associations to selected loci.

Here, we provide the first empirical analyses to quantify the impact of linked selection acting over short timescales (tens of generations) across two evolve and re-sequence studies (**Barghi2019-qy**; **Kelly2019-dc**), and an 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 selection, and the correlation between allele frequency changes between replicate populations caused by the response to convergent selection pressures. Overall, we demonstrate that linked selection has a powerful role in shaping genome-wide allele frequency changes over very short timescales.

2 Results

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 6×6 temporal covariance matrix $\mathbf{Q}$ for each of the ten replicates. Each row of these matrices represent the temporal covariance $\mathrm{Cov}(\Delta_{10}p_s, \Delta_{10}p_t)$, between the allele frequency change (in ten-generation intervals, denoted $\Delta_{10}p_t$) in some initial reference generation s (the row of the matrix), and some later timepoint t (the column of the matrix). We corrected these matrices for biases created due to sampling noise, and normalize the entries for heterozygosity (see Supplementary Materials Sections **??** and **??**). These covariances are expected to be zero when only drift is acting, as only heritable variation for fitness can create covariance between allele frequency changes in a closed population (**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 (Figure 1 A), 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 (*which could occur as a population reaches a new optimum*) (**Buffalo2019-io**). The temporal covariances per replicate are noisier but this general pattern holds; see Supplementary Figure **??**. **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).

~~*One concern is that these covariances reflect the localized impact of a few large-effect loci rather than selection on a polygenic trait.*~~ Since our covariances are essentially averages over loci, the covariance estimate *could* be strongly affected by *a few* outlier regions. To test whether large outlier regions drive the genome-wide signal we see in the **Barghi2019-qy** data, we calculate the covariances in 100kb windows along the genome (we refer to these as windowed covariances throughout) and take the median windowed covariance, and trimmed-mean windowed covariance, as a measure of the genome-wide covariance robust to large-effect loci. These robust estimates (Supplementary Table **??** and Supplementary Figure **??**) confirm the patterns we see using the mean covariance, confirming that genomic temporal covariances are non-zero due to the impact of selection acting across many genomic windows.

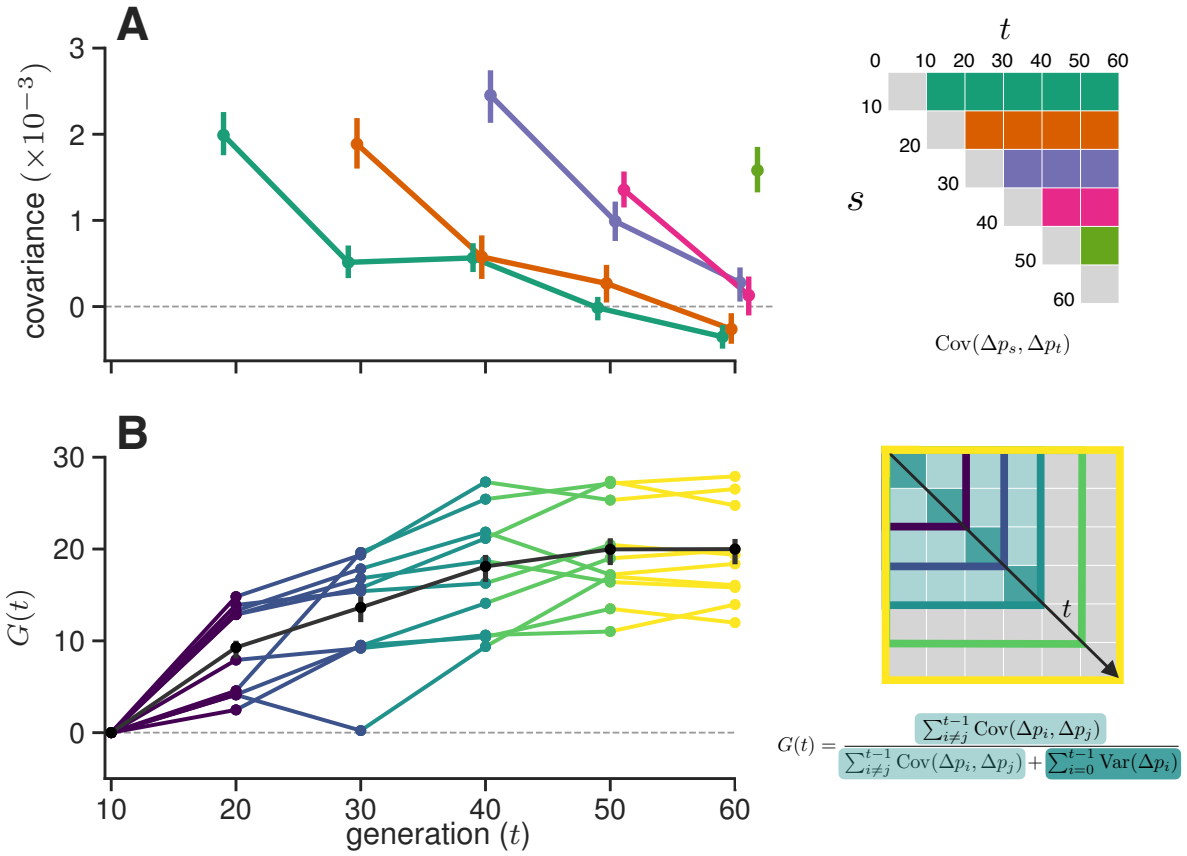

Figure 1: A: Temporal covariance, averaged across all ten replicate populations, through time from the **Barghi2019-qy** study. Each line depicts the temporal covariance $\mathrm{Cov}(\Delta p_s, \Delta p_t)$ from some reference generation s to a later time t which varies along the x-axis; each line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

While the presence of positive temporal covariances is consistent with selection affecting allele

frequencies over time, this measure is not easily interpretable. We can calculate a more intuitive measure from the temporal covariances to quantify the impact of selection on allele frequency change: the ratio of total covariance in allele frequency change to the total variance in allele frequency change. We denote the change in allele frequency as $\Delta p_t = p_{t+1} - p_t$, where p_t is the allele frequency in generation t. Since the total variation in allele frequency change can be partitioned into variance and covariance components, $\mathrm{Var}(p_t - p_0) = \sum_{i=0}^{t-1} \mathrm{Var}(\Delta p_i) + \sum_{i \neq j}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j)$ (we bias correct these for sequencing depth), and the covariances are zero when drift acts alone, this is a lower bound on how much of the variance in allele frequency change is caused by linked selection (**Buffalo2019-io**). We call this measure $G(t)$, defined as

$$G(t) = \frac{\sum_{i=0}^{t-1} \sum_{j \neq i}^{t-1} \mathrm{Cov}(\Delta p_i, \Delta p_j)}{\mathrm{Var}(p_t - p_0)} \tag{1}$$

which estimates the *impact of selection* on allele frequency change between the initial generation 0 and some later generation t, which can be varied to see how this quantity grows through time. *This measure can intuitively be understood as the relative fraction of allele frequency change normally thought of as "drift" that is actually due to selection; the denominator of* $G(t)$ *is* $\mathrm{Var}(p_t - p_0)$*, which under drift alone is proportional to* $1/2N_e$*. By measuring the contribution of temporal covariances caused by selection,* $G(t)$ *looks at the fraction selection contributes to what looks like a faster rate of drift, or lower* N_e*.* Since **Barghi2019-qy** experiment is sequenced every ten generations, in the numerator for the covariance we use the allele frequency changes between adjacent timepoints, which are ten generations apart; *thus the strong covariances between adjacent generations are not observable and do not contribute to the numerator of* $G(t)$*. Additionally, selection also inflates the variance in allele frequency change per generation; however, this effect cannot be easily distinguished from drift. For both these reasons, our measure* $G(t)$ *is strongly conservative* (*we demonstrate this through simulations in Supplementary Material Section* **??**). Still, we find a remarkably strong signal. Greater than 20% of total, genome-wide allele frequency change over 60 generations is the result of selection (Figure 1 B).

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, *and in reproducing their original analysis, we find that their number of statistically significant seasonal SNPs is not enriched compared to an empirical null distribution created by permuting seasonal labels*; 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 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,

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

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 *correlation coefficients* are relatively weak, and decay very quickly from an initial value of about 0.1 (95% block bootstrap confidence intervals $[0.094, 0.11]$) to around 0.01 (95% CIs [0.0087, 0.015]) within 20 generations. This suggests that while a reasonable fraction of the initial response is shared over the replicates, this is followed by a rapid decay, a result consistent with the primary finding of the original **Barghi2019-qy** study: that alternative loci contribute to longer term adaptation across the different replicates.

A benefit of between-replicate covariances is that unlike temporal covariances, these can be calculated with only two sequenced timepoints and a replicated study design. This allowed us to assess the impact of linked selection in driving convergent patterns of allele frequency change across replicate populations in two other studies. First, we reanalyzed the selection experiment of **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 (Figure 2 B. We estimate the convergent correlation coefficient across these replicates as 0.36 (95% CI $[0.31, 0.40]$). Similarly, we can calculate the proportion of the total variance in allele frequency change from convergent selection pressure analogous to G where the numerator is is the convergent covariance and the denominator is the total variance (see Supplementary Material **??**). We find that 37% of the total variance is due to shared allele frequency changes caused by selection (95% CI [29%, 41%]; these are similar to the convergence correlation, since the variance is relatively constant across the replicates.

Next, we reanalyzed the Longshanks selection experiment, which selected for longer tibiae length relative to body size in mice, leading to a response to selection of about 5 standard deviations over the course of twenty generations (**Marchini2014-de**; **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 each of the two Longshanks lines. We find that this is the case (gray confidence intervals in Figure 2 C), with the two Longshanks comparisons to the control line not being significantly different from zero, while the comparison between the two Longshanks lines is statistically significantly different from zero (CIs [0.0129, 0.0400]).

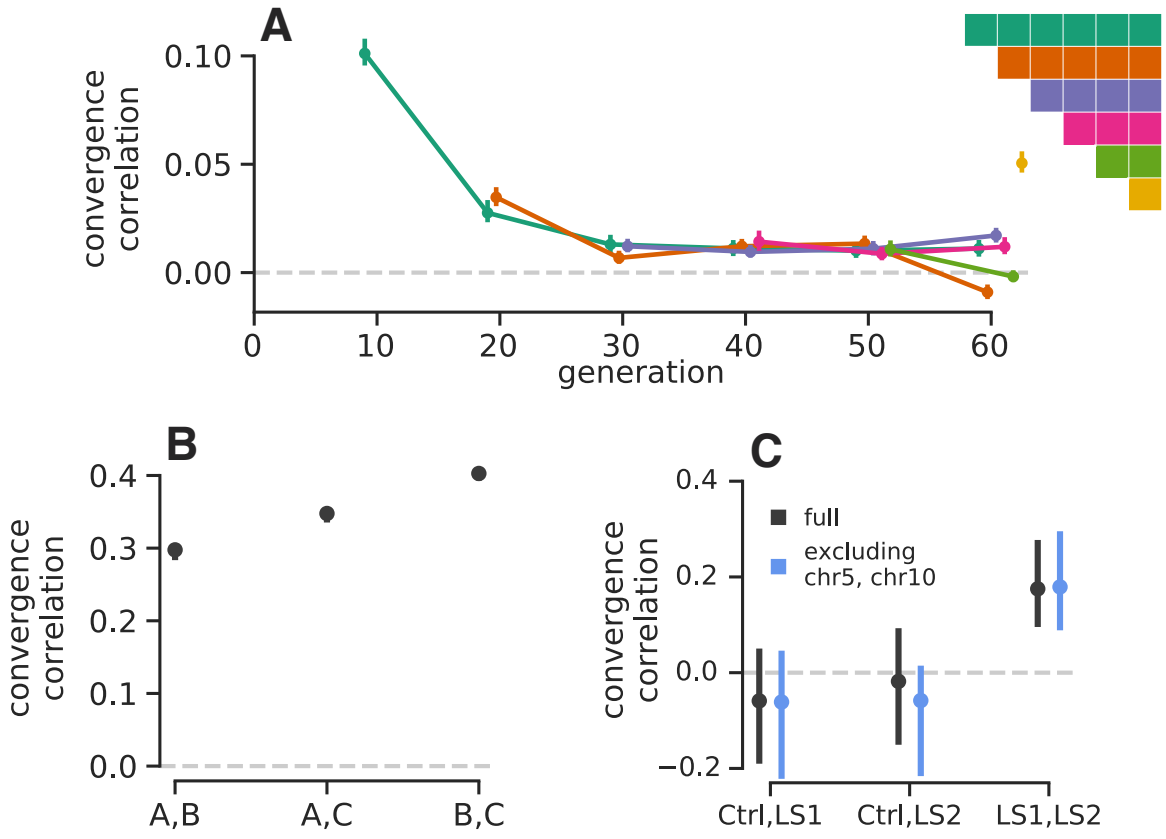

Figure 2: **A**: The convergence *correlations*, averaged across **Barghi2019-qy** replicate pairs, through time. Each line represents the convergence correlation cor($\Delta p_s, \Delta p_s$) from a starting reference generation s to a later time t, which varies along the x-axis; each line corresponds to a row of the temporal convergence correlation matrix depicted to the right *(where the diagonal elements represent the convergence correlations between the same timepoints across replicate populations)*. We note that convergent correlation for the last timepoint is an outlier; we are unsure as to the cause of this, e.g. it does not appear to be driven by a single pair of replicates. **B**: The convergence *correlations* between individual pairs of replicates in the **Kelly2019-dc** data (*note the confidence intervals are plotted, but are small on this y-axis scale*). **C**: The convergence *correlations* 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. *Through simulations, we have found that the differences in convergence correlation confidence interval widths between these* Drosophila *studies and the Longshanks study are due to the differing population sizes.*

One finding in the Longshanks study was that two major-effect loci showed parallel frequency shifts between the two selection lines: a region harboring the gene Nkx3-2 known to be involved in limb development, and another region containing six other candidate genes. We were curious to what extent our genome-wide covariances were being driven by these two outlier large-effect loci, so we excluded them from the analysis. Since we do not know the extent to which linkage disequilibrium around these large-effect loci affects neighboring loci, we took the conservative precaution of excluding the entire chromosomes these loci reside on (chromosomes 5 and 10), and re-calculating the temporal covariances. We find excluding these large effect loci has little impact

on the confidence intervals (blue confidence intervals in Figure 2 C), indicating that these across-replicate covariances are indeed driven by a large number of loci. This is consistent with a signal of selection on a polygenic trait driving genome-wide change, although we note that large-effect loci can contribute to the indirect change at unlinked loci (**Robertson1961-ho**; **Santiago1995-hx**).

The presence of an unselected control line provides an alternative way to partition the effects of linked selection and genetic drift: we can compare the total variance in allele frequency change of the control line (which excludes the effect of artificial selection on allele frequencies) to the total variance in frequency change of the Longshanks selection lines. This allows us to estimate the increase in variance in allele frequency change due to selection, which we can further partition into the effects of selection shared between selection lines and those unique to a selection line by using estimating the shared effect through the observed covariance between replicates (see Materials and Methods **??** and Supplementary Material Section **??** for more details). We estimate at least 32% (95% CI $[21\%, 48\%]$) of the variance in allele frequency change is driven by the effects of selection, of which 14% (95% CI $[3\%, 33\%]$) is estimated to be unique to a selection line, and 17% (95% CI $[9\%, 23\%]$) is the effect of shared selection between the two Longshanks selection lines (and the value of the convergence correlation between the Longshanks lines, a related statistic, is 0.18, 95% CI [0.0743, 0.254]).

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 windowed covariances (Figure 3 shows these distributions pooling across all replicates; see Supplementary Figure **??** for individuals replicates). The covariance estimate of each genomic window will be noisy, due to sampling and genetic drift, and the neutral distribution of the covariance is complicated due to linkage disequilibria (which can occur over long physical distances in E&R and selection studies, **Nuzhdin2013-gf**; **Baldwin-Brown2014-cl**). To address this, we have developed a permutation-based procedure that constructs an empirical null distribution by randomly flipping the signs of the allele frequency changes per-genomic window. This destroys the systematic covariances created by linked selection and creates a sampling distribution of the covariances spuriously created by neutral genetic drift while preserving the complex dependencies between adjacent loci created by linkage disequilibrium. This empirical neutral null distribution is conservative in the sense that the variances of the covariances are wider than expected under drift alone as they include the effect of selection on the allele frequency change within a time-interval, just not between time-intervals. We see (Figure 3 A and B) that windowed temporal covariances between close timepoints are skewed positive (a heavy right tail), while between more distant timepoints these windowed temporal covariances tend to shift to become more negative (a heavy left tail). We quantified the degree to which the left and right tails are inflated compared to the null distribution as a function of time, and see excesses in both tails in Figure 3 C. This finding is also robust to sign-permuting allele frequency changes on a chromosome-level, the longest extent that gametic linkage disequilibria can extend (Supplementary Figure **??**). We see a striking pattern that

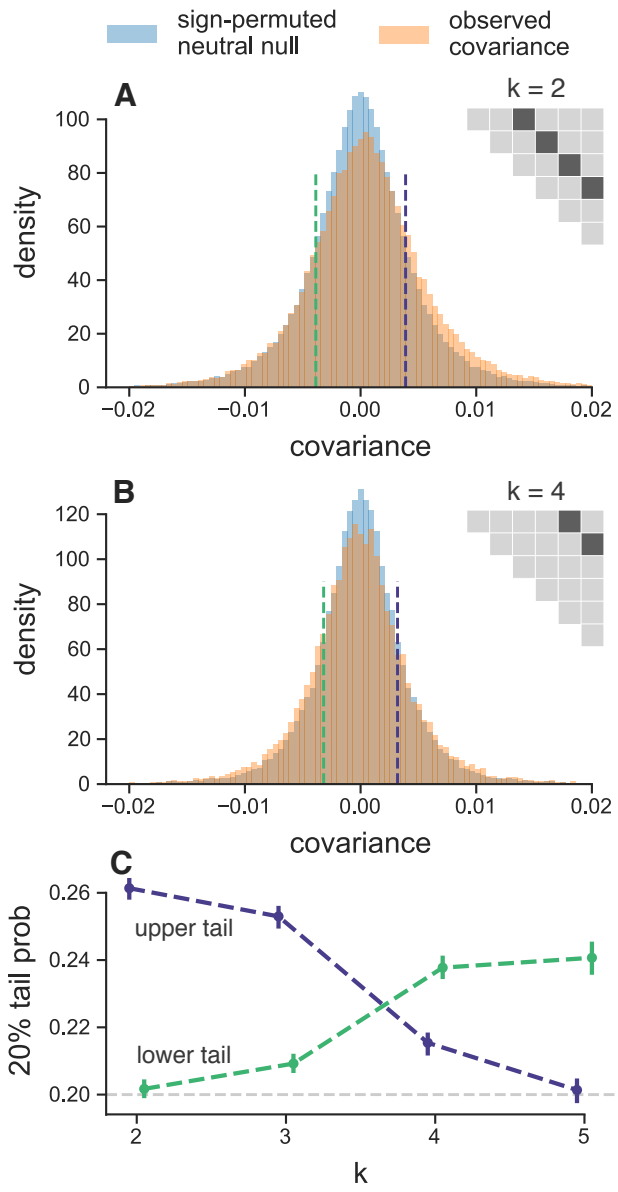

Figure 3: **A**, **B**: The distribution of temporal covariances calculated in 100kb genomic windows from the **Barghi2019-qy** study, plotted alongside an empirical neutral null distribution created by recalculating the windowed covariances on 1,000 sign permutations of allele frequency changes within tiles. The histogram bin number is 88, chosen by cross validation (Supplementary Materials **??**). In subfigure **A**, windowed covariances $\mathrm{Cov}(\Delta p_t, \Delta p_{t+k})$ are separated by $k = 2 \times 10$ generations and in subfigure **A** the covariances are separated by $k = 4 \times 10$ generations; each k is an off-diagonal from the variance diagonal of the temporal covariance matrix (see cartoon of upper-triangle of covariance matrix in subfigures **A** and **B**, where the first diagonal is the variance, and the dark gray indicates which off-diagonal of the covariance matrix is plotted in the histograms). **C**: The lower and upper tail probabilities of the observed windowed covariances, at 20% and 80% quintiles of the empirical neutral null distribution, for varying time between allele frequency changes (i.e. which off-diagonal k). The confidence intervals are 95% block-bootstrap confidence intervals, and the light gray dashed line indicates the 20% tail probability expected under the neutral null. Similar figures for different values of k are in Supplementary Figures **??**.

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.

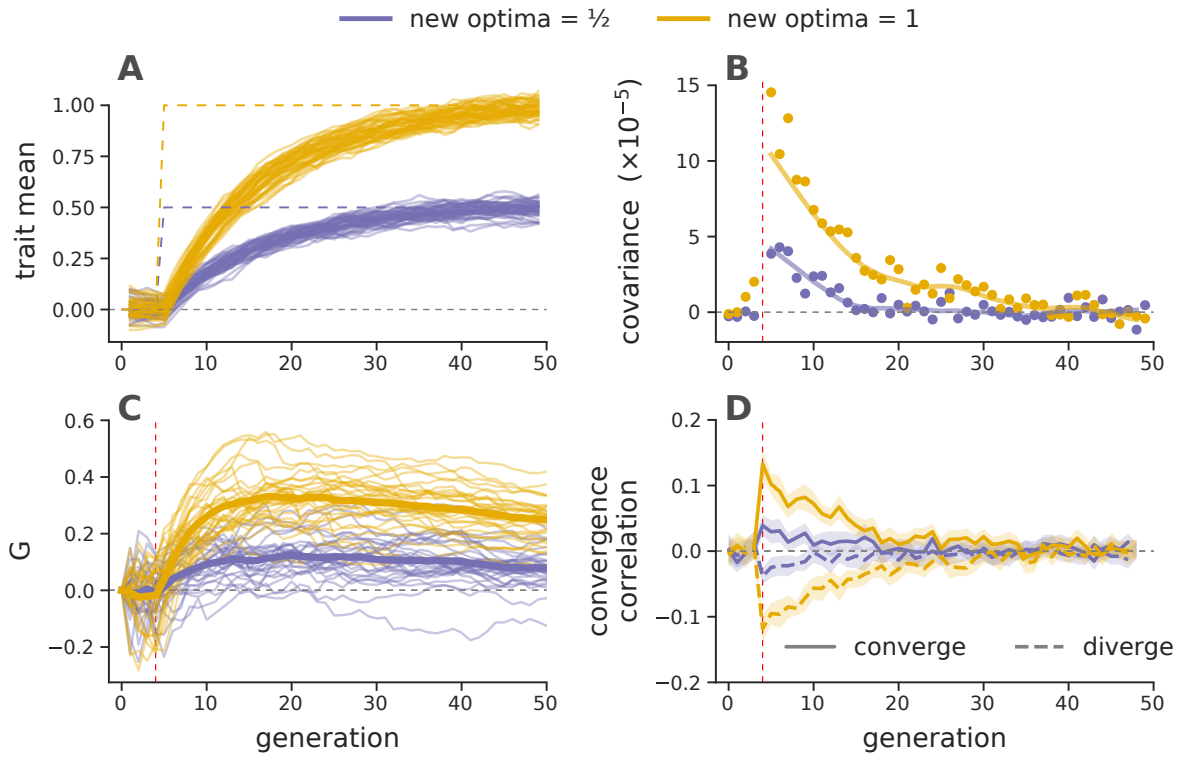

Figure 4: Forward-in-time simulations of demonstrate how temporal covariance, $G(t)$ trajectories, and convergence correlations arise during optima shifts of two different magnitudes, under Gaussian stabilizing selection. (A) Trait means across 30 replicate before and after optima shifts (solid lines), for two different magnitudes (indicated by color). The optimal trait values are indicated by the purple and yellow dashed lines. (B) Mean temporal covariance $\mathrm{Cov}(\Delta p_5, \Delta p_t)$ across 30 simulation replicates, where t varies along the x-axis (points), with a loess-smoothed average (solid line). (C) $G(t)$ trajectories through time, for 30 replicate simulations across two optima shifts. The solid line is a loess-smoothed average. (D) The convergence correlations between two populations (each 1000 diploids) split from a common population, that underwent either an optima shift in the same direction (converge) and opposite directions (diverge) at generation five. In subfigures (B), (C), and (D), directional selection begins at generation five, when the optima shifts; this is indicated by the vertical dashed red line.

Finally we used forward-in-time simulations to explore the conditions under which temporal and convergent correlations arise. We show a subset of our results for a model of *stabilizing selection on a phenotype where directional selection is induced by a sudden shift in the optimum phenotype of varying magnitudes (Figure 4A)*. We find that positive temporal covariances are produced by such selection (Figure 4B), and that these positive temporal covariances can compound together to generate a large proportion of allele frequency change being due to selection (i.e. large G) over the relatively short time periods similar to our analyzed selection datasets span (Figure 4C). The magnitude of these covariances increases with the strength of selection, i.e. the variance in fitness, such that stronger selection generates larger proportions of allele frequency change. We find a similar picture of stronger convergent selection pressures generating larger convergence correlations

(Figure 4D; see also Supplementary Materials Figure **??** *for how other factors impact convergence correlations*).

These *simulation* results are often relatively insensitive to the number of loci underlying the trait, suggesting that they may be capturing highly polygenic signals. Indeed if only a small number of loci influence the trait, the $G(t)$ trajectories are typically much more stochastic, such that replicates do not reliably generate temporal covariances (compare Figure 1B to Supplementary Material Figure **??**). This again suggests that the genome-wide linked selection response we see in *the* **Barghi2019-qy** *data* is highly polygenic. Using our simulations we find that sampling only every 10 generations does indeed mean that our estimates of $G(t)$ are an underestimate of the contribution of linked selection as they cannot include the covariance between closely spaced generations *(see Supplementary Material Figure* **??***).*

Additionally, we explored other selection schemes. We find that the long term dynamics of the covariances under directional truncation selection, which generates substantial epistasis, are richer than we see under GSS and multiplicative selection (Supplementary Material Figure **??**). We also conducted simulations of purifying selection alone (i.e. background selection) and find that this can also generate positive temporal covariances (Supplementary Material Figure **??**) *and under some circumstances, can even generate convergence correlations (Supplementary Material Figure* **??***). Thus it is unlikely that the signatures of linked selection we see are entirely* the result of the novel selection pressure the populations are exposed to, and some of this selection may be ongoing purifying selection. Only in the case of the Longshanks experiment, does the control line *allow us to conclude that* selection that is due to the novel selection pressure.

While none of our experiments have selected the populations in divergent directions, in our simulations we find that such selection can generate negative convergent correlations (Figure 4D). This suggests that selection experiments combining multiple replicates, control lines, as well as divergent selection pressures might be quite informative for unpacking the contribution of particular selection pressures to genome-wide allele frequency *changes*.

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 underused 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**; **Wray1990-zf**; **Woolliams1993-qo**; 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 of linked selection may not be differentiable from genetic drift using only single contemporary population samples or looking at at temporal allele frequency change at each locus in isolation. In this way, averaging summaries of temporal data allows us to sidestep the key problem of detecting selection from standing variation: that the genomic footprint leaves too soft of a signature to differentiate from a background of genetic drift. In fact we find that the temporal covariance signal is detectable even in the most extremely difficult to detect soft

sweep case: polygenic selection on highly polygenic traits (**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 reasonable proportion of the response is shared across short timescales. These likely represent substantial under-estimates of the contribution of linked selection because the studies we have reanalyzed do not sequence the population each generation, preventing us from including the effects of stronger correlations between adjacent generations. Furthermore, our estimation methods are intentionally conservative, for example they exclude the contribution of selection that does not persist across generations and selection that reverses sign; thus they can be seen as a strong lower bound of the effects of selection, *which we have confirmed through forward-in-time simulations*.

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. While there are LD data for a natural population of *D. simulans* (**Signor2018-wg**; **Howie2018-ay**), 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 parametric models 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 that in natural populations (**Hendry1999-zu**; **Hairston2005-ga**). Conversely, these lab populations were maintained at very small effective population sizes, estimated at 300, 450, and 45 for the **Barghi2019-qy**, **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. Adapting these methods to natural populations will require either populations that are reasonably closed to migration, or for the effect of migration to be accounted for possibly either by knowledge of allele frequencies in source populations or the identification of migrant individuals.

While it challenging to apply temporal methods to natural populations there is a lot of promise for these approaches (**Bergland2014-ij**; **Machado2018-cs**). Efforts to quantify the impact of

linked selection have found obligately sexual organisms have up to an 89% reduction in genome-wide diversity over long time periods (**McVicker2009-ax**; **Elyashiv2016-vt**; **Corbett-Detig2015-gt**; **Coop2016-gx**; **Comeron2014-nh**) Thus linked selection makes a sizeable contribution to long-term allele frequency change in some species, and there is reason to be hopeful that we could detect this from temporal data, which would help to resolve the timescales that linked selection act over. In our reanalysis of the **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 to have a much richer picture of the impact of selection as temporal sequencing becomes more common, allowing us to observe the effects of ecological dynamics in genomic data (**Hairston2005-ga**).

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

4 Materials and Methods

4.1 Datasets Analyzed

We used available genomic data data from four studies: pooled population resequencing (pool-Seq) data from **Barghi2019-qy**, **Kelly2019-dc**, **Bergland2014-ij**, and **Castro2019-uk**. In all cases, we used the variants kept after the filtering criteria of the original studies.

4.2 Variance and Covariance Estimates

To remove systematic covariances in allele frequency change caused by tracking the reference or minor allele, we randomly choose which allele's frequency to track for each locus. Then, we calculate the variance-covariance matrix of allele frequency changes using a Python software package we have written, available at `http://github.com/vsbuffalo/cvtk`. This simultaneously calculates temporal variances and covariances, and replicate covariances and uses the sampling depth and number of diploid individuals to correct for bias in the variance estimates and a bias that occurs in covariance estimates between adjacent timepoints due to shared sampling noise (see Supplementary Material Sections **??**, **??**, and **??** for mathematical details of these estimators). We assess that our bias correction procedure is working adequately through a series of diagnostic plots that ensure that the procedure removes the relationship between sampling depth and uncorrected variance and covariances (Supplementary Figure **??**). *Through our simulations we find that our estimates can differ based on how fixations and losses are handled in long time-series (Supplementary Material Section **??**) but none of our findings in the main text are altered by this decision (Supplementary Material Figures **??** and **??**).*

4.3 Estimating Uncertainty with a Block Bootstrap

To infer the uncertainty of covariance, convergence correlation, and $G(t)$ estimates, we used a block bootstrap procedure. This bootstrap procedure resamples blocks of data points, rather than individual data points, to infer the uncertainty of an statistic in the presence of unknown correlation

between loci. As most estimators in this paper are ratios (e.g. covariance standardize by sample heterozygosity, $G(t)$, and the convergence correlation) which we estimate with a ratio of averages, we exploit the linearity of expectation for efficient computation of bootstrap samples (see **??** for details).

4.4 Partitioning Unique and Shared Selection Effects in the Longshanks Study

The unselected control line in the Longshanks experiment allows us to additionally partition the total variance in allele frequency change into drift, shared effects of selection, and unshared effects of selection between selected replicates. We begin by decomposing the allele frequency change in Longshanks line 1 (LS1) as $\Delta p_{t,\mathrm{LS1}} = \Delta_D p_{t,\mathrm{LS1}} + \Delta_U p_{t,\mathrm{LS1}} + \Delta_S p_{t,\mathrm{LS}}$ where these terms are the drift in Longshanks replicate 1 ($\Delta_D p_{t,\mathrm{LS1}}$), selection unique to the LS1 replicate ($\Delta_U p_{t,\mathrm{LS1}}$), and selection response shared between the two Longshanks replicates ($\Delta_S p_{t,\mathrm{LS}}$) respectively (and similarly for the Longshanks line 2, LS2). By construction, this decomposition assumes that each of these terms are uncorrelated within replicates, so the contribution of each term to the total variance in allele frequency change, $\mathrm{Var}(\Delta p_{t,\mathrm{LS1}})$, is the variance in of that term's allele frequency change.

We estimate the effects of selection by first calculating the fraction of the total variance explained by drift. We assume the variance in allele frequency change observed in the unselected control line ($\mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}})$) is driven entirely by neutral genetic drift, and since an identical breeding scheme was used across all three replicates (except breeders for the control line were chosen at random), we can use this as an estimate of the contribution of neutral genetic drift in the selected lines, $\mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}}) = \mathrm{Var}(\Delta_D p_{t,\mathrm{LS1}}) = \mathrm{Var}(\Delta_D p_{t,\mathrm{LS2}})$. Then, we can estimate the increase in variance in allele frequency change due to selection as $(\mathrm{Var}(\Delta p_{t,\mathrm{LS1}}) + \mathrm{Var}(\Delta p_{t,\mathrm{LS2}}))/2 - \mathrm{Var}(\Delta p_{t,\mathrm{Ctrl}})$ and the shared effect of selection across selected lines as $\mathrm{Cov}(\Delta p_{t,\mathrm{LS1}}, \Delta p_{t,\mathrm{LS2}})$. Finally, the covariance in allele change between replicates is used to estimate the shared effects of selection between lines, $\mathrm{Cov}(\Delta p_{t,\mathrm{LS1}}, \Delta p_{t,\mathrm{LS2}}) = \mathrm{Var}(\Delta_S p_{t,\mathrm{LS}})$.

4.5 Windowed Covariance and the Empirical Neutral Null

Throughout the paper, we use genomic windows for the block-bootstrap procedure. For the *D. simulans* and *D. melanogaster* data from the **Barghi2019-qy**, **Kelly2019-dc**, and **Bergland2014-ij** studies, we used large megabase windows for the block bootstrap procedure, while we used a ten megabase window for the large mouse genome data from the **Castro2019-uk** study.

Given evidence of a reversal in the direction of selection at later timepoints in the **Barghi2019-qy** study, we calculated windowed temporal covariances on 10 kilobase windows and looked at the distribution of these covariances through time. We compare these distributions of windowed covariances to an empirical neutral null created by randomly permuting the sign of allele frequency change at the block level (to preserve the correlation structure between loci due to LD). This destroys the systematic covariances in allele frequency change created by linked selection, which emulates a frequency trajectory under drift. This approach is conservative, since heritable fitness variation also inflates the magnitude of allele frequency change more than expected under drift, but we do not change these magnitudes. Using this empirical neutral null distribution of windowed covariances, we calculate how much of the observed windowed covariance distribution falls outside of empirical null distribution for different tail probabilities. While the comparison between the distribution of 10 kilobase windowed covariances to the empirical neutral null created from sign-permuting 10 kilobase windows is most natural, we wanted to ensure that our finding that the

shift from mostly positive to mostly negative windowed covariances through time (Figure 3) was robust to LD extending beyond the range of these 10 kilobase windows. We took the conservative approach of also sign-permuting at the chromosome-level, and found the same qualitative shift (Supplementary Figure **??**).

4.6 Forward-in-time Simulations

*To explore how aspects of genetic architecture, model of selection, and experimental design impact temporal covariance, the $G(t)$ trajectories, and convergence correlations, we ran extensive forward-in-time using SLiM (***Haller2019-vu***); here we discuss the Gaussian Stabilizing Selection simulations in Figured 4, but Supplementary Materials Section* **??** *describes these simulation routines in much more detail as well as additional simulations we have conducted.*

We simulated directional selection on a trait by first burning in a population of $N = 1000$ diploids under GSS for $10N$ generations with the stabilizing selection trait variance $V_s = 1$ and an optima set at zero. We targeted a polygenic architecture by setting the trait mutation rate to 10^{-8} per basepair, per generation, in addition to having a separate neutral mutation of 10^{-8} which created neutral mutations which we used to calculate the temporal covariances. Our simulated region was 50 megabases in length (about one quarter of a Drosophila *chromosome), and trait alleles were randomly selected to have a ± 0.01 effect size. By tracking the trait mean through the burnin, we found it converged as expected. After a $10N$ burnin, the population was split into two different replicate population sizes, to capture the effect of bottlenecks in selection experiments (these population sizes were, 50, 500, and 1000 diploids; the later representing no bottleneck). Each population then underwent an optima shift of either 0.1, 0.5, or 1 on generation five, with the first four generations serving as a control. These optima shifts were either in the same direction (converging), different directions (diverging), or only one optima shifted (as a control). By tracking the trait mean, we saw that it converged as expected during burnin, and had the appropriate response to selection (Supplementary Material Figure* **??***). Using the neutral population frequency data from these simulations, we calculated the temporal covariances, $G(t)$ trajectories, and convergence correlations.*

5 Acknowledgments

We would like to thank the authors of the original studies we've analyzed, including Neda Barghi, Christian Schlötterer, John Kelly, Kimberly Hughes, Frank Chan, Campbell Rolian, Nick Barton, Alan Bergland, and Dmitri Petrov. We would also like to thank Doc Edge for helpful statistical advice, and Matt Osmond, Erin Calfee, Andy Kern, Sivan Yair, *Chuck Langley, Dave Begun, and Michael Turelli* for helpful discussions. *Additionally, we thank Guy Sella and one additional anonymous reviewer for greatly improving the manuscript.* This research was supported by an NSF Graduate Research Fellowship grant awarded to VB (1650042), and NIH (R01-GM108779) and NSF (1353380) awarded to GC.

Supplementary Material

1.1 Estimator Bias Correction

1.1.1 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,

$$\begin{aligned}\text{Var}(\widetilde{p_t}) &= \mathbb{E}(\text{Var}(\widetilde{p_t}|p_t)) + \text{Var}(\mathbb{E}(\widetilde{p_t}|p_t)) &(3)\\ &= \underbrace{\frac{p_t(1-p_t)}{d_t}}_{\text{generation } t \text{ sampling noise}} + \underbrace{\text{Var}(p_t)}_{\text{variance due to evolutionary process}}. &(4)\end{aligned}$$

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

$$\begin{aligned}\text{Var}(p_t) &= \text{Var}\left(p_0 + (p_1 - p_0) + (p_2 - p_1) + \ldots + (p_t - p_{t-1})\right) &(5)\\ &= \text{Var}\left(p_0 + \Delta p_0 + \Delta p_1 + \ldots + \Delta p_{t-1}\right) &(6)\\ &= \text{Var}(p_0) + \sum_{i=0}^{t-1}\text{Cov}(p_0, \Delta p_i) + \sum_{i=0}^{t-1}\text{Var}(\Delta p_i) + \sum_{0\le i<j}^{t-1}\text{Cov}(\Delta p_i, \Delta p_j). &(7)\end{aligned}$$

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

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

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**)

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

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

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

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

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

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

which is the usual variance in allele frequency change due to drift. Then, the total allele frequency change from generations 0 to t is $\mathrm{Var}(\widetilde{p}_t - \widetilde{p}_0) = \mathrm{Var}(\widetilde{p}_t) + \mathrm{Var}(\widetilde{p}_0) - 2\,\mathrm{Cov}(\widetilde{p}_t, \widetilde{p}_0)$, where the covariance depends on the nature of the sampling plan (see **Nei1981-oy**; **Waples1989-sj**). 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 (**Waples1989-sj**), we write,

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

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

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

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

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

where $C = 1$ if Plan I is used, and $C = 0$ if Plan II is used (see **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. $\frac{d_t}{d_t - 1}\widetilde{p}_t(1-\widetilde{p}_t)$,

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

1.1.2 Correcting variance bias with individual and depth sampling processes

Here, we extend the sampling bias correction described above to handle two binomial sampling processes: one as individuals are binomially sampled from the population, and another as reads are

binomially sampled during sequencing. (see also **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 $\widetilde{p}_t = {}^{Y_t}/_{d_t}$ be the observed sample allele frequency. Then, the sampling noise is

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

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

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

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

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

Through the law of total expectation (see **Kolaczkowski2011-ee** Supplementary File 1 for a sample proof), one can find that an unbiased estimator of the half the heterozygosity is

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

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

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

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

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

$$\tag{26}$$

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

1.1.3 Covariance Correction

We also need to apply a bias correction to the temporal covariances (and possibly the replicate covariances if the initial sample frequencies are all shared). The basic issue is that $\text{Cov}(\Delta\widetilde{p}_t, \Delta\widetilde{p}_{t+1}) =$

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

1.1.4 Temporal-Replicate Covariance Matrix Correction

In practice, we simultaneously estimate the temporal and replicate covariance matrices for each replicate, which we call the temporal-replicate covariance matrix. This needs a bias correction; we extend the bias corrections for single locus variance and covariance described in Supplementary Material Sections **??**, **??**, 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 $\mathbf{\Delta F}$ which contains the allele frequency changes between adjacent generations, and has dimension $R \times T \times L$. The operation $\mathrm{flat}(\mathbf{\Delta F})$ 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

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

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

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

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

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

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

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

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

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

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

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

1.1.5 Barghi2019-qy Temporal Covariances

Since each replicate population was sequenced every ten generations, the timepoints $t_0 = 0$ generations, $t_1 = 10$ generations, $t_2 = 20$ generations, etc., lead to observed allele frequency changes across ten generation blocks, $\Delta p_{t_0}, \Delta p_{t_1}, \ldots, \Delta p_{t_6}$. Consequently, the ten temporal covariance matrices for each of the ten replicate populations have off-diagonal elements of the form $\mathrm{Cov}(\Delta p_{t_0}, \Delta p_{t_1}) = \mathrm{Cov}(p_{t_1} - p_{t_0}, p_{t_2} - p_{t_1}) = \sum_{i=0}^{10} \sum_{j=10}^{20} \mathrm{Cov}(\Delta p_i, \Delta p_j)$. Each diagonal element has the form $\mathrm{Var}(\Delta p_{t_0}) = \sum_{i=0}^{t_0} \mathrm{Var}(\Delta p_i) + \sum_{i \neq j}^{t_0} \mathrm{Cov}(\Delta p_i, \Delta p_j)$, and is thus a combination of the effects of drift and selection, as both the variance in allele frequency changes and cumulative temporal autocovariances terms increase the variance in allele frequency. With sampling each generation, one could more accurately partition the total variance in allele frequency change (**Buffalo2019-io**); while we cannot directly estimate the contribution of linked selection to the variance in allele frequency change here, the presence of a positive observed covariance between allele frequency change can only be caused linked selection.

1.2 Block Bootstrap Procedure

The estimators used in this paper are predominantly ratios, e.g. temporal-replicate covariance standardized by half the heterozygosity, $G(t)$ which is the ratio of covariance to total variance, and the convergence correlation (equation (2)). In these cases, we can exploit the linearity of the expectation to make the bootstrap procedure more computationally efficient, by pre-calculating the statistics of the ratio's numerator and denominator, $N(\mathbf{x}_i)$ and $D(\mathbf{x}_i)$, on the data $\mathbf{x}_i$ for all blocks $i \in \{1, 2, \ldots, W\}$ in the genome. Then we draw W bootstrap samples with replacement, and compute the estimate for bootstrap sample b with an average weighted by the fraction w_i of total loci contained in each block,

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

Note that computing the ratio of averages rather than the average of a ratio is a practice common for population genetic statistics like F_{ST} (**Bhatia2013-zy**). With these B bootstrap estimates, we calculate the $\alpha/2$ and $1-\alpha/2$ quantiles, which we use to estimate the $1-\alpha = 95\%$ pivot confidence intervals (p. 33 **Wasserman2006-jl**, p. 194 **Davison2013-oy**) throughout the paper,

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

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

1.3 Replicate G and Partitioning the Variance in Allele Frequency

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

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

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

Extending our theoretic work in **Buffalo2019-io**, we can partition the allele frequency change in two replicates into drift, and shared selection and replicate-specific selection components of allele frequency change. For two replicates, A and B,

$$\Delta p_{t,A} = \Delta_{D} p_{t,A} + \Delta_{U} p_{t,A} + \Delta_{S} p_t \tag{35}$$
$$\Delta p_{t,B} = \Delta_{D} p_{t,B} + \Delta_{U} p_{t,B} + \Delta_{S} p_t \tag{36}$$

where $\Delta_{D} p_{t,A}$ is allele frequency change due to drift (this is specific to a replicate, and equal to $\Delta_{N} p_{t,A} + \Delta_{M} p_{t,A}$ in the notation of **Buffalo2019-io**), $\Delta_{U} p_{t,A}$ is the allele frequency change from indirect selection specific to replicate A (and likewise with $\Delta_{U} p_{t,A}$ for replicate B), and $\Delta_{S} p_t$ is the allele frequency change from indirect selection shared across the replicates A and B (this term lacks a replicate subscript since by construction it is identical between replicates). By construction, each of these terms is uncorrelated, so the variance and be written as:

$$\mathrm{Var}(\Delta p_{t,A}) = \mathrm{Var}(\Delta_{D} p_{t,A}) + \mathrm{Var}(\Delta_{U} p_{t,A}) + \mathrm{Var}(\Delta_{S} p_t) \tag{37}$$
$$\tag{38}$$

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

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

In artificial selection studies with a control (non-selected) line, such as the **Castro2019-uk** study, this allows us to estimate the contribution of the effects of shared and unique indirect selection. In the case of this study, we can estimate the drift, unique selection effect, and shared selection effect terms using the fact that,

$$\Delta p_{t,\mathrm{LS1}} = \Delta_D p_{t,\mathrm{LS1}} + \Delta_U p_{t,\mathrm{LS1}} + \Delta_{\mathrm{LS}} p_t \tag{40}$$
$$\Delta p_{t,\mathrm{LS2}} = \Delta_D p_{t,\mathrm{LS2}} + \Delta_U p_{t,\mathrm{LS2}} + \Delta_{\mathrm{LS}} p_t \tag{41}$$
$$\Delta p_{t,\mathrm{Ctrl}} = \Delta_D p_{t,\mathrm{Ctrl}}. \tag{42}$$

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

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

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

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

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

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

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

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

1.4 The Empirical Neutral Null Windowed Covariance Distribution

To detect an excess of genomic regions with unusually high or low covariances, we need to compare the distribution of observed windowed covariances to a null distribution of windowed covariances that we would expect under no selection. While we could construct a theoretic sampling distribution of the spurious covariances created by neutral genetic drift at particular site, the unknown linkage disequilibrium between sites would mean that this is not an adequate null model for the distribution of windowed covariances in our data.

To address this limitation, we construct a neutral null model by sign-permuting the observed allele frequency changes. This destroys the covariances built up by selection, mimicking a neutral allele's frequency trajectory. This approach is conservative, since selection also acts to increase the magnitude of allele frequency changes (see equation 1 of **Buffalo2019-io**), but this magnitude is not affected by the sign-permutation procedure. Consequently, the resulting empirical null distribution has higher variance than would be expected under neutrality alone.

Still, we wanted to ensure that LD between sign-permuted blocks, which will affect the variance of the empirical null distribution, does not impact our primary finding that the distribution of temporal covariances becomes increasingly negative in the **Barghi2019-qy** dataset through time. To address this, we also sign-permuted at the whole chromosome level finding we recapitulate the same pattern (Supplementary Figure **??**).

1.5 Bergland2014-ij Re-Analysis

We also applied our temporal covariance approach to **Bergland2014-ij**, which found evidence of genome-wide fluctuating selection between Spring and Fall seasons across three years *Drosophila melanogaster*. As described in **Buffalo2019-io**, if fluctuating selection pressure among time-periods are the dominant genome-wide pattern, we might expect positive covariances between like seasons changes (e.g. Spring 2010 to Fall 2010 and Spring 2011 to Fall 2011), and negative covariances between dislike seasonal changes (e.g. Fall 2009 to Spring 2010 and Fall 2010 to Spring 2011). However, while we find temporal covariances that are non-zero, we find only weak support for a seasonal fluctuating model driving these covariances. In Supplementary Figure **??**, we show the temporal covariances from varying reference generations, across seasonal transitions that are alike (e.g. the covariance between the allele frequency changes between Fall 2009 and Spring 2009, and frequency changes between Fall 2010 and Spring 2010), and dislike (e.g. the covariance between the allele frequency change between Fall 2009 and Spring 2009, and the frequency changes between Spring 2010 and Fall 2009). The first row of temporal covariance matrix is consistent with fluctuating selection operating for two timepoints, as the first covariance is negative, and the second is positive, and later covariances are not statistically differentiable from zero (which could occur if LD and additive genetic variance decay). However, all other temporal covariances do not fit the pattern we would expect under genome-wide fluctuating selection.

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 a genome-wide excess of loci showing evidence

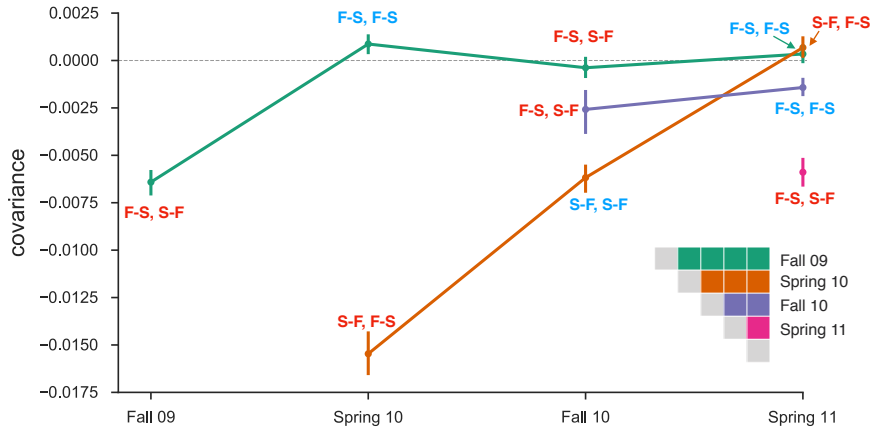

Figure S1: Temporal covariances from the **Bergland2014-ij** study, from varying reference generations (e.g. rows along the temporal covariance matrix). Each covariance is labeled indicating whether the covariance is between two like seasonal transitions (e.g. the covariance between allele frequency 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.

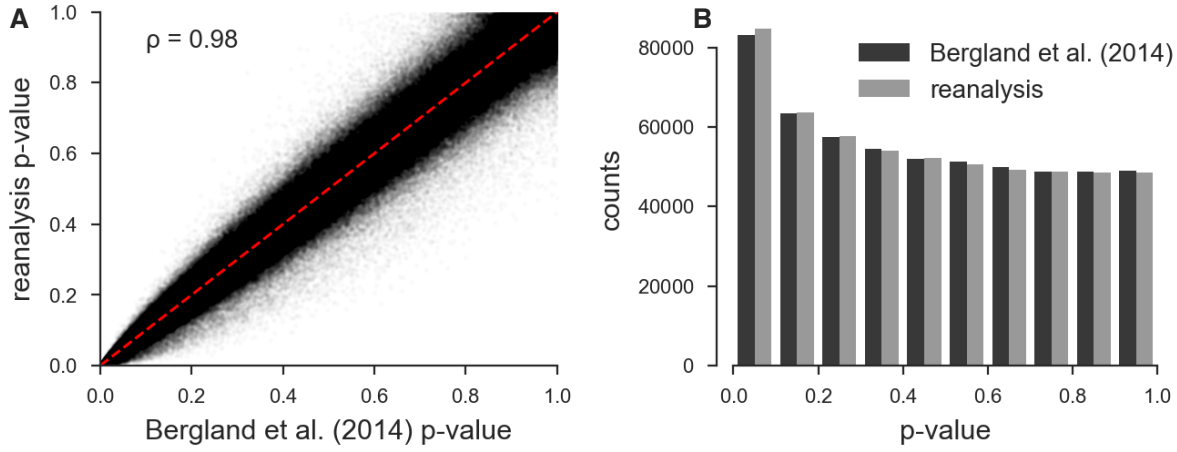

Figure S2: **A**: Scatterplot of the original unadjusted p-values from **Bergland2014-ij** and the p-values from our reanalysis of the same data using the same statistical methods; the minor discrepancy is likely due to software version differences. **B**: The histograms of the p-values of our reanalysis 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.

of fluctuating selection we reanalyzed the data of (**Bergland2014-ij**) using the same seasonal fluctuating model as the original paper. This model is a Binomial logit-linked GLM fit per-locus, where the frequencies are regressed on the Spring/Fall seasons are encoded as a dummy variable. 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.

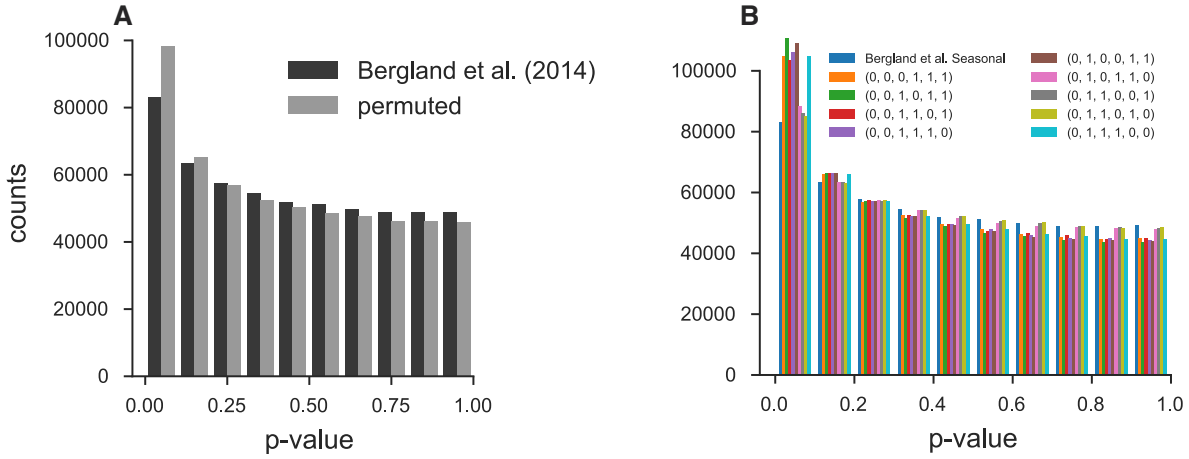

Figure S3: **A**: Histogram of original **Bergland2014-ij** seasonal p-values and p-values creating by randomly permuting the seasons at each locus. **B**: Histogram of original **Bergland2014-ij** p-values alongside all unique permutations (ignoring symmetries that lead to identical p-values).

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 **??**), where we see strong evidence of selection (non-zero temporal covariances), yet the pattern

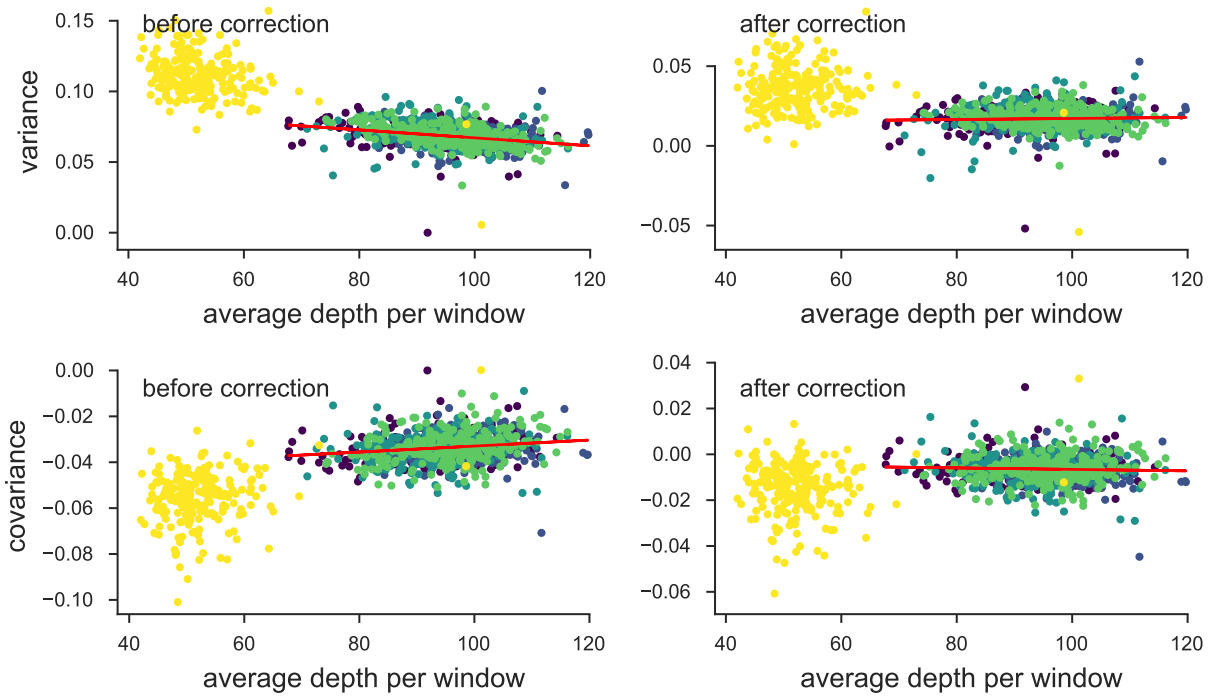

Figure S4: The variance and covariances from the **Bergland2014-ij** study, calculated in 100kb genomic windows plotted against average depth in a window before and after bias correction. Each panel has a least-squares estimate between the variance and covariance, and the average depth. The bias correction 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.

does not follow that of seasonal fluctuating selection.

1.6 Simulation Results

We conducted extensive simulations to understand how temporal covariance, $G(t)$, and convergence correlations behave under (1) different quantitative genetic fitness models, (2) different trait architectures (e.g. varying levels of V_A for fitness and the number of sites affecting fitness), (3) background selection, and (4) different sampling periods. Furthermore, we use two *replicate* population simulations to investigate how convergence correlations depend on (1) the population sizes of each selection line sampled from the main population, and (2) the direction the trait is selected on in each line (i.e. in the same direction, differing directions, or only one lines elected).

Due to the high computational burden of forward simulations over this wide breadth of parameters, we modeled a single 50 megabase region in a population of $N = 1000$ diploid individuals with a neutral variant mutation rate of 10^{-8} and a recombination rate of 10^{-8} per basepair. This is roughly analogous to a quarter of an autosome of *Drosophila melanogaster*; however with this small population size and mutation rate, the population mutation rate θ for the entire region leads to far fewer neutral sites to calculate covariances and other statistics on than expected for a region this

length in *D. melanogaster*. Since our main goal is to understand the dynamics of statistics used in the paper and how they are affected by different quantitative genetic fitness models, background selection, and trait architecture, we use population frequencies rather than sampling frequencies.

All forward simulations were conducted using SLiM (**Haller2019-vu**) and run and processed using Snakemake (**Koster2012-iv**); all simulation routines are available in the Github repository `https://github.com/vsbuffalo/cvtk/`.

1.6.1 The Effects of the Genetic Architecture under Exponential Directional Selection

We first investigated the effects of the selected trait's genetic architecture on temporal covariances and $G(t)$ by neutrally burning in a population for $10N$ generations, and selecting on the trait with an exponential fitness function. The exponential fitness function corresponds to multiplicative selection across sites and so serves as the simplest directional selection model of a trait to understand the effects of genetic architecture on the statistics we have used in the paper.

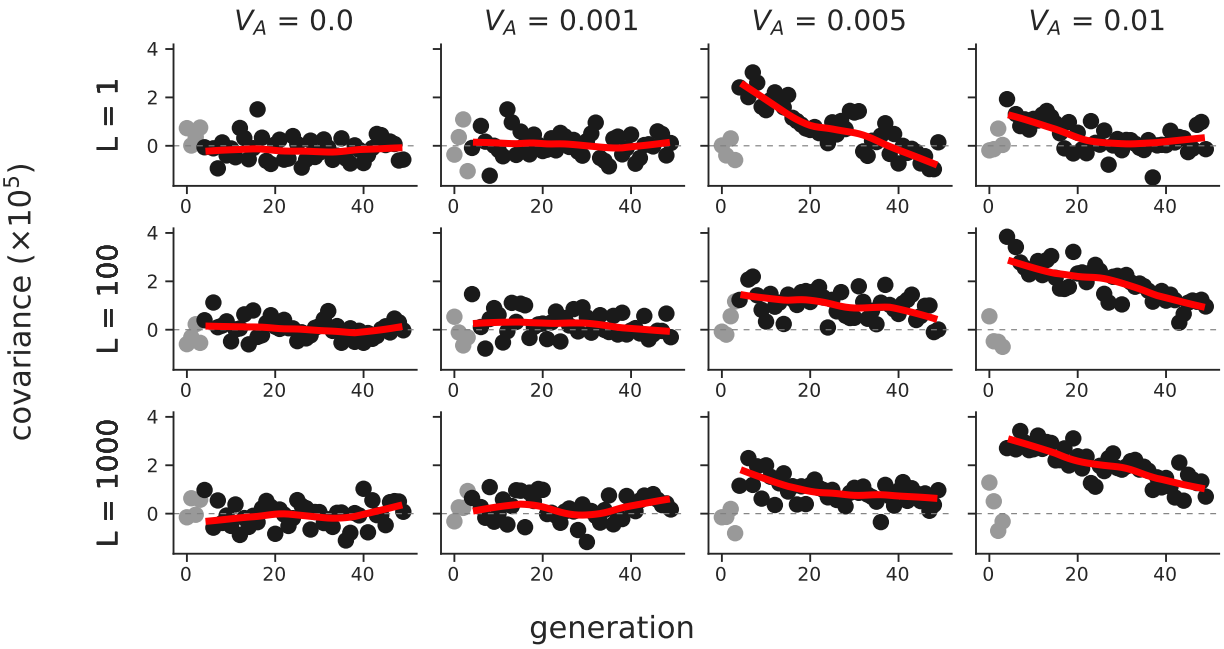

Figure S5: The temporal covariances $\mathrm{Cov}(\Delta p_5, \Delta p_t)$ from the onset of selection (generation 5) to a later time point t, which varies along the x-axis, across a variety of different trait additive genetic variances (V_A, columns) and number of sites contributing to the trait (L, rows). Each point is the temporal covariance averaged over 50 replicate simulations; dark gray points are temporal covariances after the onset of selection, and light gray points are before. The red line is a loess-smoothed curve through the covariances after the onset of selection. Selection on the trait was imposed through an exponential fitness function.

During this burnin, sites were either marked as neutral (with mutation rate $\mu_{\text{neutral}} = 10^{-8}$ per gamete per generation) or contributed to the trait's value (with mutation rate μ_{trait}), but were not selected until generation $10N + 5$ (the five generations after burnin serve as a neutral control). The trait mutation rate, μ_{trait} was set by targeting a particular architecture, the number of selected sites, L. Each site contributing to the trait's value was randomly chosen to have effect size $\pm\alpha$ with

equal probability, where α was set to target a particular additive genetic variance for the trait, V_A, for the target number of selected sites L.

Overall, we confirm a finding in **Buffalo2019-io** that the initial expected temporal covariance conditioned on V_A, is invariant to the number of loci determining the trait's value, L (Supplementary Figure **??**). We do find some evidence that the decay in temporal covariance is faster when the trait has a monogenic basis (see the third column of Supplementary Figure **??**); this is expected the selection coefficients are larger for these monogenic simulations, leading to faster allele frequency changes and a rapid change in additive genetic variance.

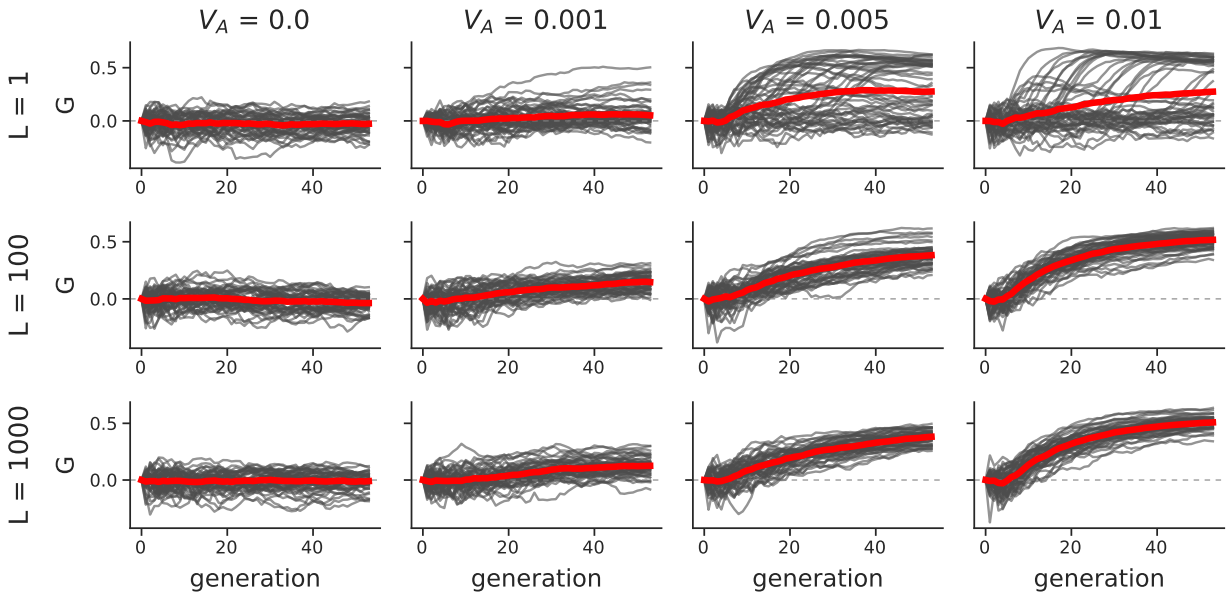

Figure S6: The $G(t)$ trajectories of 50 replicate simulations, across different trait architectures (L is the target number of sites affecting the trait's value, and V_A is the target trait additive genetic variance). The red line is the mean trajectory across all replicate simulations. Like Supplementary Figure **??**, the onset of selection is five generations after the $10N$ generation burnin; this is evident by the initial flat period of the $G(t)$ trajectory.

In our previous work, we did not investigate the affect of trait architecture on our measure $G(t)$. Using the exponential fitness function simulations, we also calculated $G(t)$ for each of the replicate simulations. We find that the $G(t)$ trajectories can vary considerably across replicates depending on the number of sites (L) determining the trait's value (Supplementary Figure **??**). When a trait is reasonably monogenic ($L \approx 1$), $G(t)$ trajectories vary considerably across replicate lines, as certain lines may stochastically lose the few copies of the selected alleles (top row of Supplementary Figure **??**). However, with a polygenic trait, ($L \geq 100$), the $G(t)$ trajectories across replicates are similar as each replicate contains an abundance of trait alleles (bottom rows of Supplementary Figure **??**). Comparing the simulated $G(t)$ replicate trajectories of Supplementary Figure **??** with the **Barghi2019-qy** $G(t)$ trajectories in Figure 1B, we again confirm a finding of **Barghi2019-qy**: that there is considerable genetic redundancy among beneficial alleles, *meaning because of the polygenic architecture, there are multiple routes to adaptation*. We should note that our simplified simulation routines are slightly different from the **Barghi2019-qy** study in that the burnin populations are all independent; however we expect the same qualitative result.

1.6.2 Temporal Covariances and $G(t)$ under Gaussian Stabilizing Selection

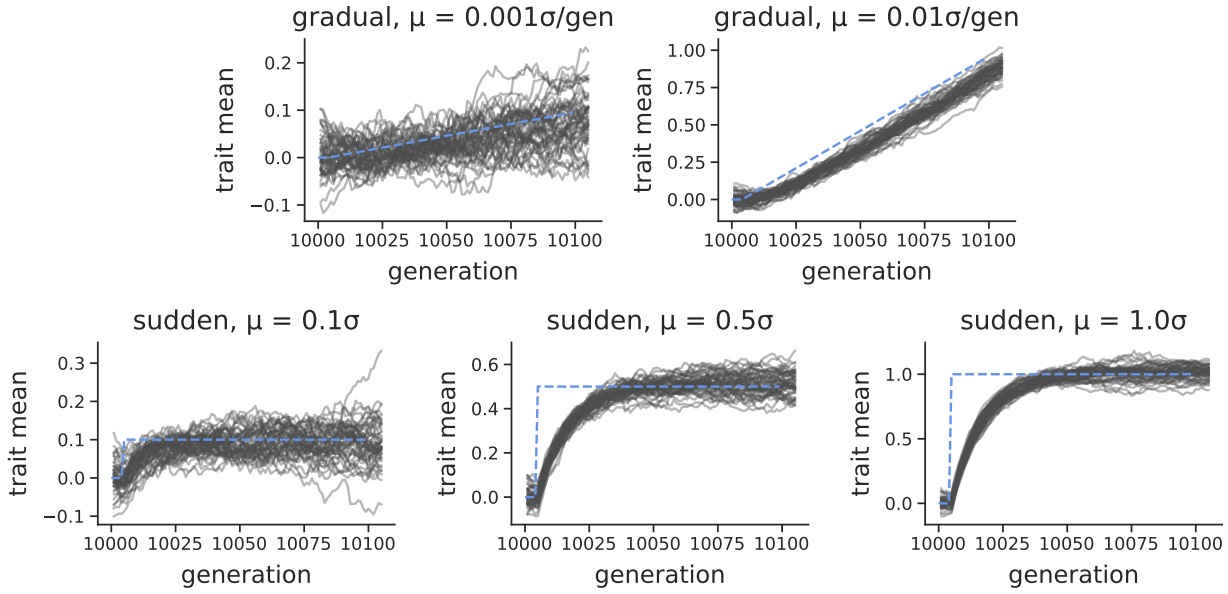

Figure S7: The population mean trait value under the Gaussian stabilizing selection simulations (gray lines) and the trait optima (dashed blue lines). The first row shows the selection response during a graduate shift in optima per generation, while the second row shows the selection response during a sudden optima shift.

Additionally, we wanted to ensure that our temporal covariances and $G(t)$ trajectories were robust to more complicated, but realistic fitness models. To this end, we also simulated Gaussian stabilizing selection (GSS) on a trait during burnin, followed by one of two two optima shift routines: (1) sudden optima shifts of $\mu_{\text{sudden}} = \{0.1\sigma, 0.5\sigma, 1\sigma\}$, and (2) very graduate optima shifts of $\mu_{\text{gradual}} = \{0.001\sigma, 0.01\sigma\}$ per generation using the same two population simulation scheme described above. We used a polygenic architecture for these simulations, *with trait alleles assigned a* ± 0.01 *effect size with equal probability, trait mutation rate* 10^{-8}, and the optima shift began at five generations after a $10N$ generation burnin. Across our GSS simulations, we see the expected selection response (Supplementary Material Figure **??**).

Overall, we see the same qualitative results under Gaussian stabilizing selection with optima shifts as under exponential directional selection. Stronger directional selection, here determined by larger sudden optima shifts or larger gradual shifts per generation, lead to stronger temporal covariances (Supplementary Materials Figure **??**). Furthermore, we see a stronger effect of linked selection, as measured by $G(t)$, under stronger directional selection (Supplementary Material Figure **??**).

Additionally, we looked at the effect the size of each replicate population has on a single population's $G(t)$ *trajectories. These simulations had the same* $10N$ *generation burnin, followed by a change in population size emulating the bottlenecks associated with creating selection lines. Overall, we find that smaller population sizes lead to a reduced* $G(t)$ *(Supplementary Material Figure* **??***). This is expected, as the denominator of* $G(t)$ *is* $\text{Var}(p_t - p_0)$*, which has an inverse relationship with* N_e*; as replicate population size is reduced, the proportion of allele frequency change driven by*

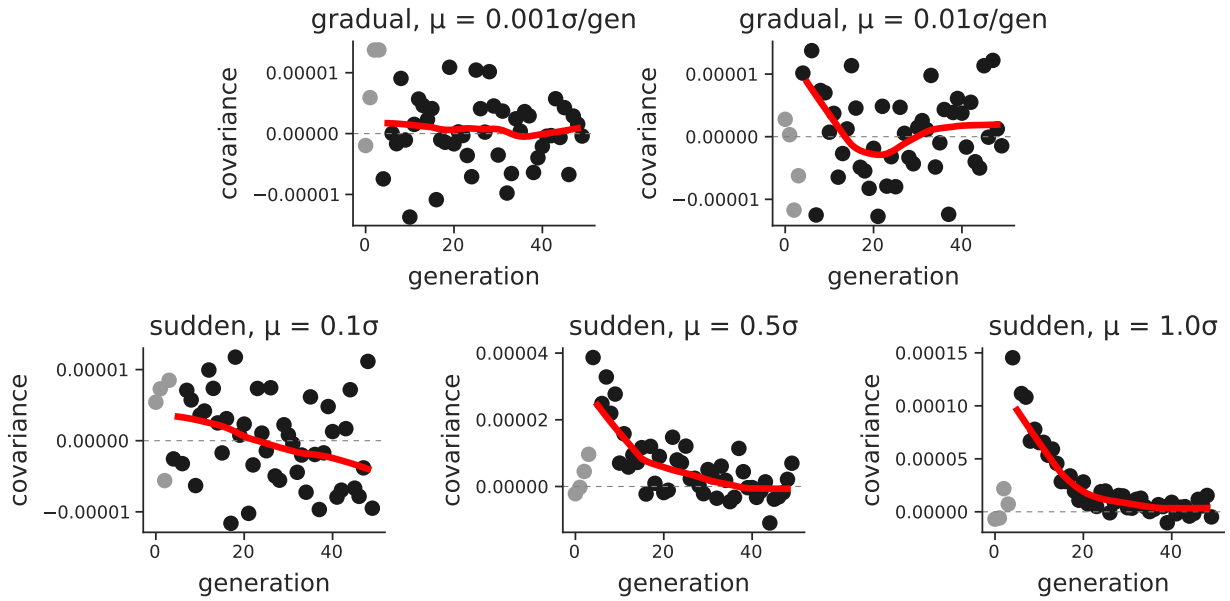

Figure S8: Mean temporal covariance (Cov $\Delta p_5, p_t$, with t varying across the x-axis) across 30 replicate simulations (light gray points are before the onset of selection; dark gray points are after selection begins), under different Gaussian stabilizing selection with optima shift regimes. The solid red line is a loess-smoothed average of these points.

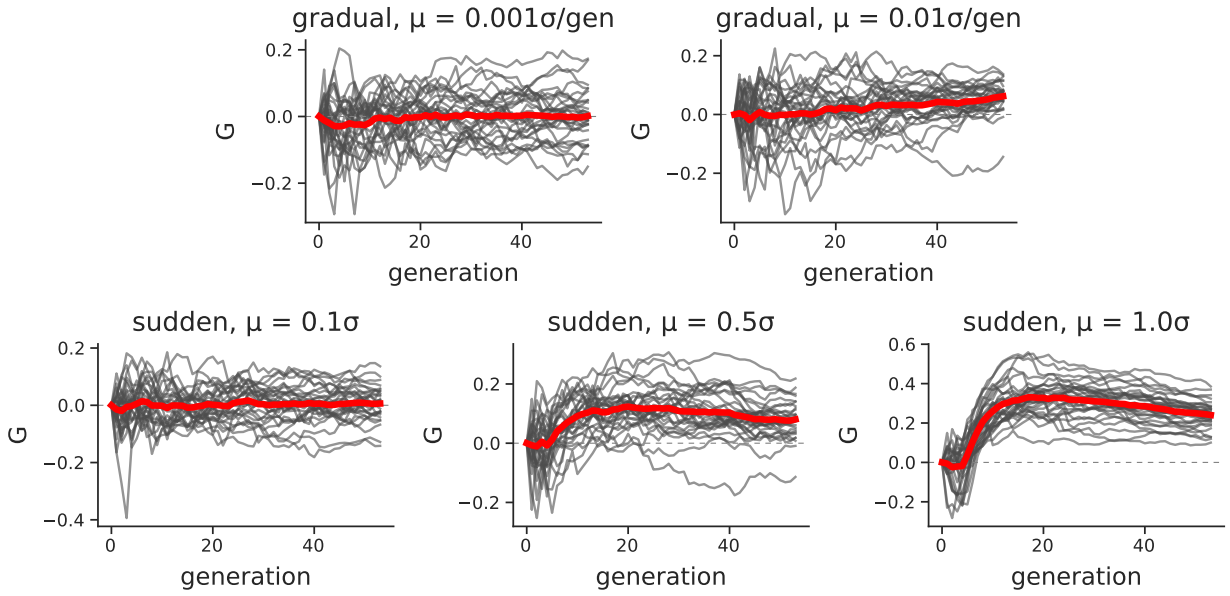

Figure S9: $G(t)$ trajectories across 30 replicate Gaussian stabilizing selection with optima shift regimes. The solid red line is a loess-smoothed average across replicates.

linked selection is lower, since the rate of drift is increased.

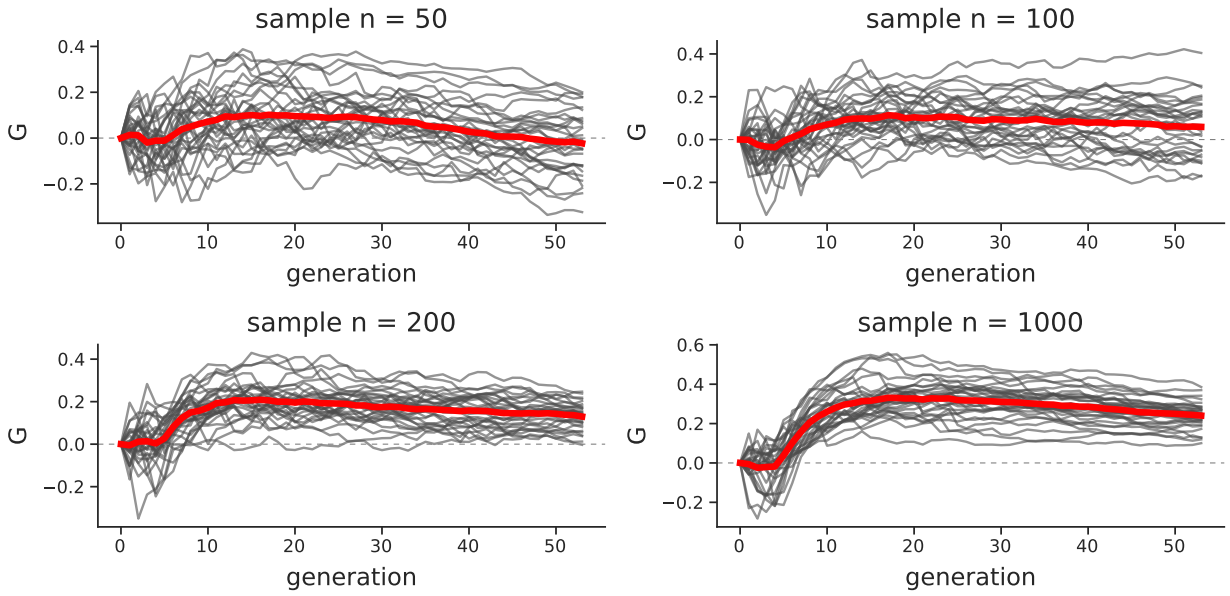

Figure S10: $G(t)$ trajectories for a sudden optima shift of 1, for varying replicate population sizes.

1.6.3 Convergence Correlations

Using the same exponential fitness function simulations described above, we also investigated how the convergence correlation is impacted by (1) genetic architecture, (2) the design of the selection experiment, e.g. how many individuals are selected for each line from the founding population, and (3) the direction of selection across the two populations "lines". After burning in $N = 1000$ diploid populations for $10N$ generations, we simulated two equally-sized lines of sizes $n = \{50, 500, 1000\}$ diploids, and imposed three selection schemes across different simulation runs. First, we imposed a convergent selection scheme, where the populations undergo exponential directional selection in the same direction. We expect that the convergent correlation under this convergent scheme should be positive, as the two lines should share some haplotypes carrying beneficial alleles, and these are selected in the same direction across the two lines. Second, we imposed divergent selection, where the two lines again undergo exponential directional selection, except in different directions. Here, we expect the convergence correlation to be negative, as haplotypes that increase the selected trait in one population are beneficial in the upward selected line, but deleterious in the downward selected line. Third, we have a control selection scheme, where one line is selected and the other is not; this is akin to the control line in the **Castro2019-uk** study (see Figure 2C). In this case, we expect to see no convergence correlation, as only one line is being selected. Finally, across these two-line simulation studies, we expect that smaller selection line sizes should show weaker convergent correlations, as the probability that the same haplotypes are selected between the two lines decreases with size.

Overall, our simulations confirm our hypotheses; see Supplementary Material Figure **??**. We also find that in simulations where we target a monogenic genetic architecture (i.e. the target number of trait-affecting loci is $L = 1$), the convergence correlations are generally much weaker than those under a polygenic architecture. However, this effect is mediated by the line population size; the difference in convergence correlation between $L = 1$ and $L = 1000$ are more dissimilar when

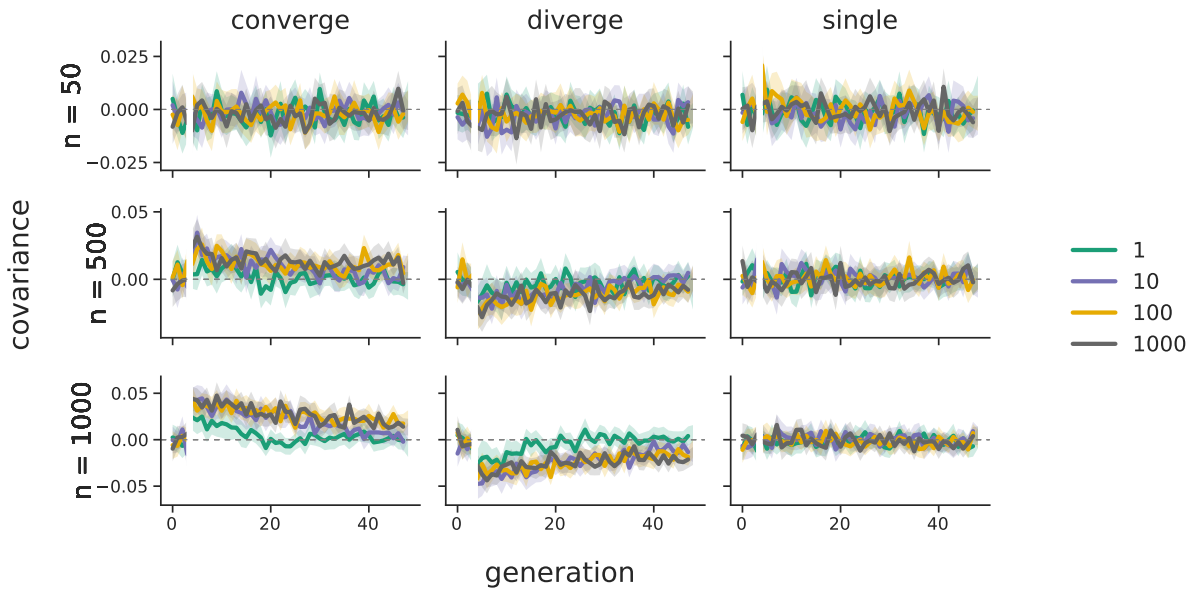

Figure S11: The convergence correlations across the two population line exponential directional selection simulations; panel rows are for differing line population sizes, and panel columns are the modes of selection across the lines (convergent, divergent, and only a single selected line control). Line color indicates the target genetic architecture, in number of loci affecting the trait's value. 95% confidence intervals are also shown. Note that selection begins at generation five, which is the reference generation; this is indicated by the split in the lines.

the line population sizes are larger (compare the first column, last two rows). Like the convergence correlations calculated on the **Barghi2019-qy** data, we find in simulations convergence correlations decay through time. Additionally, populations selected in opposite directions lead to negative convergence correlations, as expected. Overall, we find that the convergence correlation is affected by both genetic architecture and the size of the selected population lines.

We also wanted to test whether we see similar convergence correlations under Gaussian stabilizing selection. In these simulations, rather than targeting a particular V_A, we fix the trait mutation rate at 10^{-8} (thus region-wide $\theta = 2000$). Like the exponential directional selection simulations, we impose directional selection in the same direction across the two populations (converge), different directions (diverge), and only in one population (single). We also vary the type (gradual versus sudden) and magnitude of optima shifts in the two populations. Overall, simulations show convergence correlations for the sudden optima shifts in Supplementary Figure **??**. Importantly, optima shifts in the opposite direction cause negative convergence correlations. We found that for slow moving optima shifts, the convergence correlations are generally too weak to be distinguished from zero reliably, (top row of Supplementary Material Figure **??**).

1.6.4 Sampling in Temporal Blocks

In our analysis of the **Barghi2019-qy** data, we describe our statistic $G(t)$ as a lower bound for two reasons: (1) the population is sequenced every ten generations, meaning the temporal covariances between adjacent generations cannot contribute to the numerator of $G(t)$ but contribute

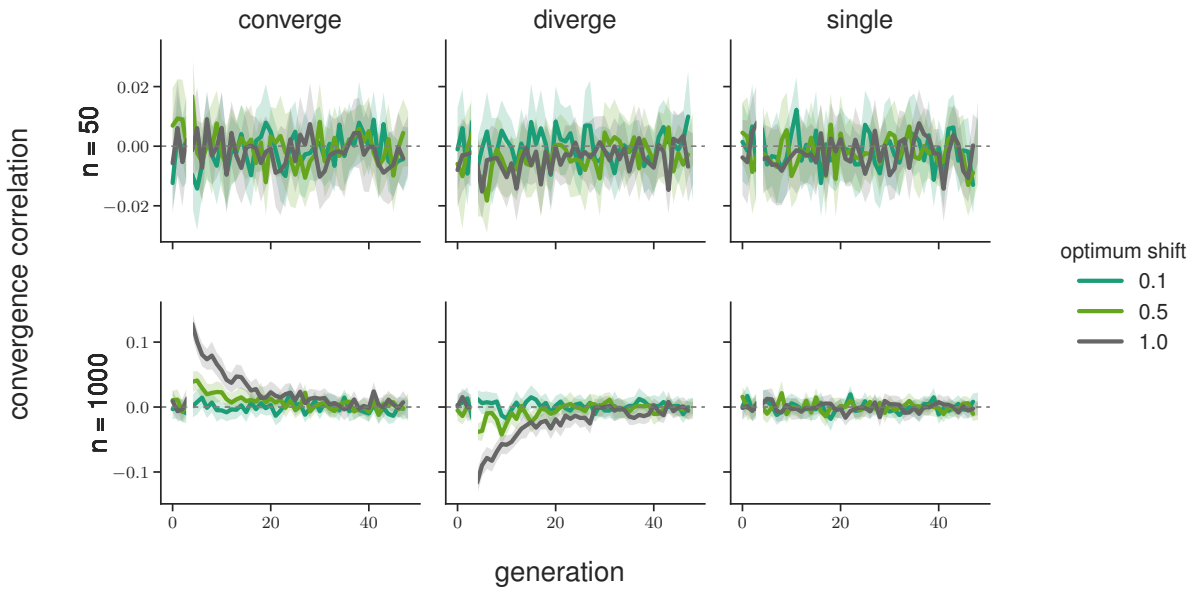

Figure S12: The convergence correlations across the two population line Gaussian stabilizing selection sudden optima shift simulations; selection line population sizes vary across rows, and panel columns are the modes of selection across the lines (convergent, divergent, and only a single selected line control). All simulations have a target number of loci affecting the trait of $L = 1000$; line color indicates the size of the sudden optima shift in standard deviations of V_S 95% confidence intervals are also shown. Note that selection begins at generation five, which is the reference generation; this is indicated by the split in the lines.

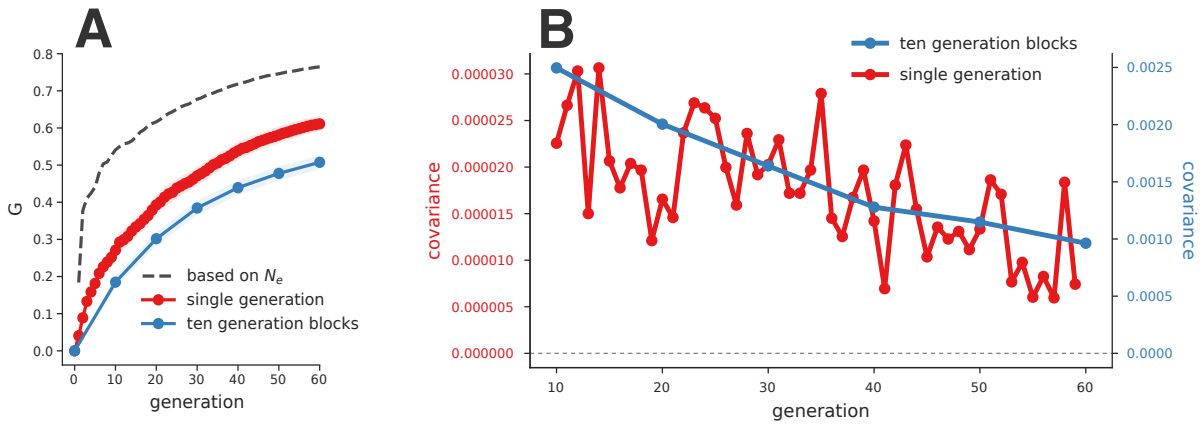

Figure S13: A: The $G(t)$ averaged over 50 replicate simulations with $V_A = 0.01$ and $L = 1000$. The blue line shows $G(t)$ calculated over ten generation blocks, similar to the calculation of temporal covariances of the **Barghi2019-qy** study. The red line shows the average $G(t)$ estimates when the population is sampled every generation and all covariances can contribute to the numerator of $G(t)$. The dashed gray line indicates the $G(t)'$ estimate, which uses the known drift effective population size of the simulations. B: The temporal covariances calculated each generation (red line) and on ten generation blocks (blue line) using the same simulation data.

to the denominator, and (2) the estimate of $G(t)$ ignores linked selection's contribution to the *per-generation* variance in allele frequency change (this is the difference between *our* $G(t)$ and $G'(t)$ *estimators*, the latter of which includes these variance terms (**Buffalo2019-io**)). To verify that $G(t)$ estimated every ten generations is indeed a lower bound, we used a simulation procedure similar to the exponential fitness function simulations (described in Supplementary Material Section **??**), and calculated the temporal covariances and $G(t)$ both each generation, and every ten generations. Unlike the simulations described in **??**, we began selection at $10N$ generations, and used trait $V_A = 0.01$ and targeted $L = 1000$ sites affecting the trait.

First, comparing $G(t)$ when sampling population frequencies every generation versus every ten generations, we confirm that the ten-generation block $G(t)$ is a lower bound of the $G(t)$ trajectory when sampling is every generation (red and blue lines in Supplementary Figure **??**A). Furthermore, since we control the population size in our simulations at $N = 1000$ diploids, we know the drift effective population size in the absence of selection. This allows us to estimate $G(t)'$, which is a measure of $G(t)$ that accounts for the linked selection's inflation of the variance in allele frequency change between two generations (equation 26, **Buffalo2019-io**). Plugging in the drift effective population size $N_e = 1000$ into the expression for $G'(t)$ and using the $\text{Var}(p_t - p_0)$ calculated for different t's, we see that the every generation $G(t)$ that does not account for linked selection's inflation of $\text{Var}(\Delta p_t)$ does underestimate the true impact of linked selection as expected (dashed gray line in Supplementary Figure **??**A).

To further understand the effects of calculating temporal covariances every ten generations rather than every generation, we also compared their magnitudes and decay rates using the simulations described above. We find that ten generation block temporal covariances are orders of magnitude larger but decay at similar rates (see Supplementary Figure **??**B; note the two y-axis scales are different). The larger magnitude is expected, as each ten generation block temporal covariance is the sum of 45 temporal covariances between adjacent generations (e.g. $\binom{10}{2}$).

1.6.5 Background Selection

In our previous work, **Buffalo2019-io**, we did not investigate whether background selection can lead to temporal autocovariance. Here, using Forward-in-time simulations, we find that background selection can indeed generate temporal autocovariance and lead to convergence correlations when deleterious haplotypes are shared between populations and both removed by selection.

We simulated background selection in a 50 megabase region, where deleterious alleles are randomly introduced by mutation. Following background selection literature (**Charlesworth1993-gb**; **Nordborg1996-nq**; **Hudson1994-oh**; **Hudson1995-xc**), we parameterize the mutation rate as the total number of deleterious mutations introduced per diploid genome, per generation, and simulate values $U = \{0.5, 1.0, 1.5\}$. We also vary the strength of selection against the deleterious mutations, $s = \{0.0, 0.01, 0.05, 0.1\}$ (where $s = 0$ is a neutral control), as well as different recombination rates ($r_{bp} = \{10^{-7}, 10^{-8}\}$). Like other simulations, we burnin the population for $10N$ generations under backgrounds selection. Overall, we find background selection does create temporal covariance (Supplementary Material Figure **??**), which are stronger under (1) higher deleterious mutation rates and (2) larger selection coefficients. This latter point initially seems at odds with background selection theory, as the level of pairwise diversity in a region under *stroong* background selection is invariant with respect to the selection coefficient. However, looking at the background selection $G(t)$ trajectories, we find that over time, the impact of linked selection created by background selection appears to trend towards an equilibrium in the $r_{BP} = 10^{-7}$ subfigure, and

reaches an equilibrium in the $r_{BP} = 10^{-8}$ subfigure *that seems reasonably invariant to the choice of s* (Supplementary Material Figure **??**). *We believe that these observations can be reconciled by* $Cov(\Delta ap_t,\ \Delta p_{t'})$ *being larger for larger s when* $|t' - t|$ *is small, but also decaying more rapidly with* $|t' - t|$*, such that the overall contribution of selection to allele frequency variance is invariant to s (for strong background selection).* However, further future work is needed to fully explore and understand the temporal covariance dynamics of background selection.

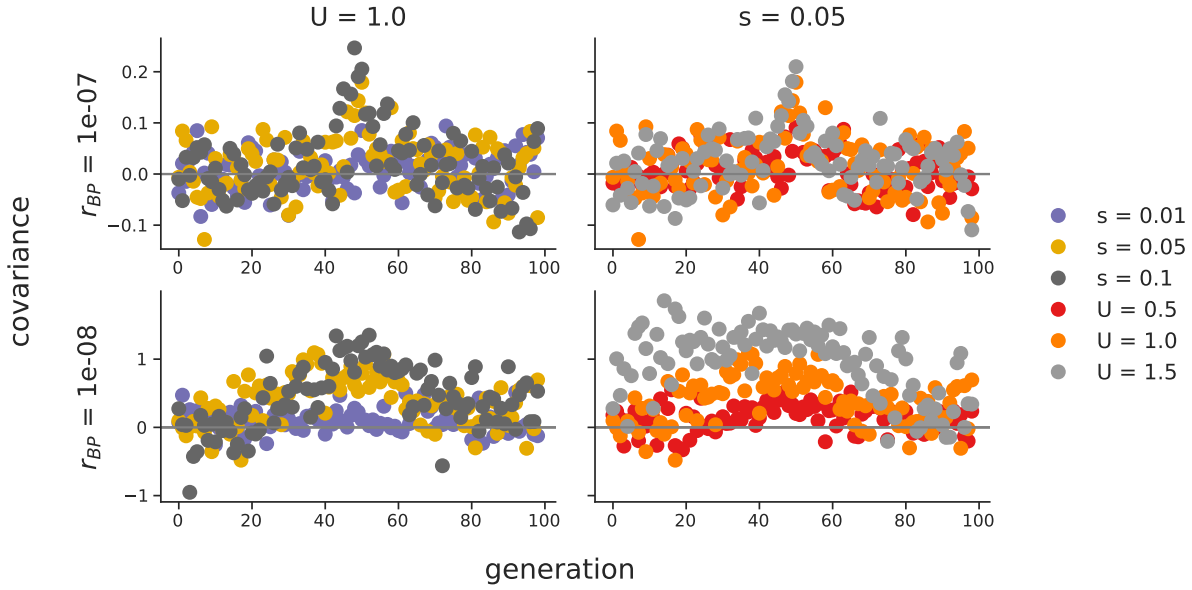

Figure S14: The temporal covariances, $\mathrm{Cov}(\Delta p_{50}, \Delta p_t)$ (where t varies along the x-axis) created by background selection, under different recombination rates (r_{BP}, rows), selection coefficients (s), and deleterious mutation rates (U). Unlike directional selection figures, where we choose the reference generation to be the first generation after the onset of selection, here we choose an arbitrary reference generation (generation 50). The symmetry of temporal covariance around the reference generation, is expected, since unlike directional selection the level of additive genetic variance for fitness has hit mutation-selection-drift balance. Note that the first column sets constant $U = 1.0$, and s varies, while the second column sets $s = 0.05$ constant, and varies U.

Additionally, we investigated whether background selection can create convergence correlations between two replicate populations. Much like the exponential directional selection and Gaussian stabilizing selection simulations, we burned in a population for $10N$ generations with background selection, which continued after the population was split into two replicate populations. These simulations fixed $U = 1.0$, $r_{BP} = 10^{-8}$, and varied the replicate population size $n = \{200, 1000\}$. We find that background selection can create convergence correlations (Supplementary Material Figure **??**). We find the convergence correlation is weaker in smaller replicate population sizes, as there are fewer shared haplotypes carrying *the same* deleterious alleles between the two populations.

While investigating whether different selection coefficients converged to the same levels of $G(t)$, we extended the number of generations we calculated our statistics on (e.g. in earlier exponential and Gaussian stabilizing selection fitness function, these were calculated to fifty generations; here we've calculated them to one hundred generations). This lead us to discover a subtle bias temporal covariance: whether sites that fix or are lost through time are included or excluded in the temporal