

August 28, 2019

1 Introduction

2 Results

We measure two types of covariances between allele frequency caused by selection: temporal autocovariances, and across-replicate covariances. First, positive temporal autocovariance in a neutral allele frequency’s trajectory occurs when the allele becomes associated with a high or low fitness chromosomal background, and this association persists through the generations due to linkage disequilibrium. As long as the direction of selection remains constant, the fitness background is predictive of the direction in the neutral allele’s frequency changes, creating positive covariance. Even though these magnitude of frequency changes at each site may be subtle, as they would be under polygenic selection, cumulatively these perturb neighboring sites in a predictable manner and build up temporal autocovariance which acts as a genome-wide signal of linked selection. Second, if evolution occurs in replicate populations undergoing convergent selection pressure, neutral sites linked to fitness backgrounds shared across replicates are expected to change in the same direction. This creates across-replicate covariance, which is a measure of the extent to which convergent selection pressures across replicate populations cause similar allele frequency changes. Finally, it is important to note that under the null model where the fitness differences between individuals are entirely random and non-heritable (e.g. when selection is not acting), both forms of covariances are expected to be zero.

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

We first analyzed Barghi et al. (2019), an evolve-and-resequence study with ten replicate populations exposed to a high temperature environment and evolved for 60 generations, and sequenced every ten generations. Using the seven timepoints and ten replicate populations, we estimated a bias-corrected 60×60 temporal-replicate variance-covariance matrix (see XXX for details). Since each replicate population was sequenced every ten generations, the timepoints $t_0 = 0$ generations, $t_1 = 10$ generations, $t_2 = 20$ generations, etc., lead to observed allele frequency changes across ten generation blocks, $\Delta p_{t_0}, \Delta p_{t_1}, \dots, \Delta p_{t_6}$. Consequently, the ten temporal covariance matrices for each of the ten replicate populations have off-diagonal elements of the form $\text{Cov}(\Delta p_{t_0}, \Delta p_{t_1}) = \text{Cov}(p_{t_1} - p_{t_0}, p_{t_2} - p_{t_1}) = \sum_{i=0}^{10} \sum_{j=10}^{20} \text{Cov}(\Delta p_i, \Delta p_j)$. Each diagonal element

has the form $\text{Var}(\Delta p_{t_0}) = \sum_{i=0}^{t_0} \text{Var}(\Delta p_i) + \sum_{i \neq j}^{t_0} \text{Cov}(\Delta p_i, \Delta p_j)$, and is thus a combination of the effects of drift and selection, as both the variance in allele frequency changes and cumulative temporal autocovariances terms increase the variance in allele frequency. With sampling each generation, one could more accurately partition the total variance in allele frequency change (Buffalo and Coop 2019); while we cannot directly estimate the contribution of linked selection to the variance in allele frequency change here, the presence of a positive observed covariance between allele frequency change can only be caused linked selection. Overall, we expect observed temporal covariances between timepoints to be weak, some fraction of the additive genetic variance and linkage disequilibrium that creates temporal autocovariance decays in the ten generations between sequencing.

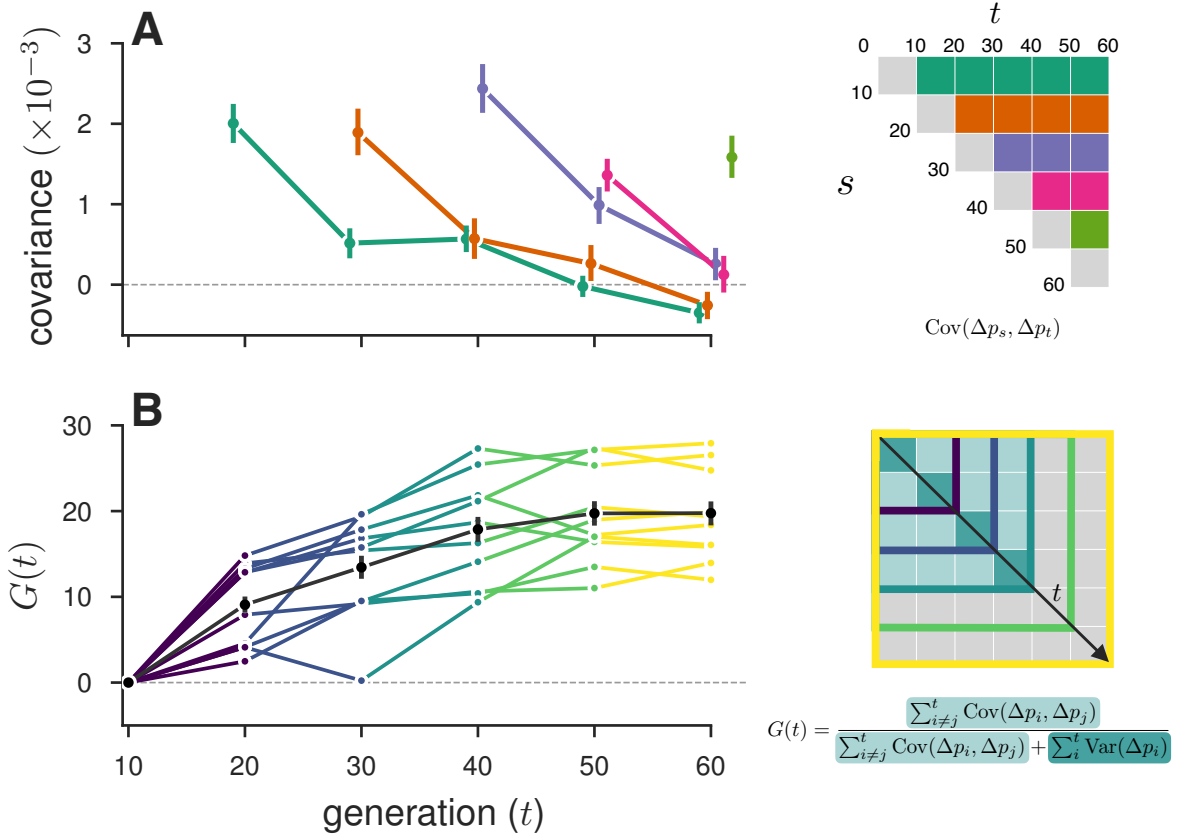


Figure 1: A: Temporal covariance, averaged across all ten replicate populations, through time from the Barghi et al. (2019) study. Each line depicts the temporal covariance $\text{Cov}(\Delta p_s, \Delta p_t)$ from some reference generation s to a later time t which varies along the x-axis; each line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

Averaging across replicate populations, we find positive temporal covariances across time that

are statistically significant ($p < XXX$) consistent with linked selection acting to affect allele frequency changes over very short time periods. We visualize these covariances in Figure (A), which depicts the temporal covariances through time, for each of the five rows covariance matrix. Each row represents the temporal covariance $\text{Cov}(\Delta p_s, \Delta p_t)$, between some initial reference generation s (the row of the matrix), and some later timepoint t (the column of the matrix). For each row, the covariances at first are positive, and then decay towards zero as expected when directional selection affects linked variants' frequency trajectories until ultimately linkage disequilibrium and additive genetic variance for decay (Buffalo and Coop 2019). Note that per replicate, the signal is a bit noisier; see Supplementary Figures XXX.

While the presence of positive temporal covariances is consistent with linked selection affecting allele frequencies over time, this measure not easily interpretable. Additionally, we can quantify the impact of linked selection on allele frequency change as the ratio of total covariance in allele frequency change to the total variance in allele frequency change. Since the total variation in allele frequency change can be decomposed into variance and covariance components, $\text{Var}(p_t - p_0) = \sum_{i \neq j} \text{Cov}(\Delta p_i, \Delta p_j) / \text{Var}(p_t - p_0)$, and the covariances are zero when drift acts alone, this is a lower bound on how much of the variance in allele frequency change is caused by linked selection (Buffalo and Coop 2019). We call this measure $G(t)$, which is the total effect of linked selection between the initial generation 0 and some later generation t , which can be varied to see how this quantity grows through time. As with the temporal covariances, the study design of **Barghi2019** leads our measure $G(t)$ to be even more conservative, since the temporal covariances with each ten-generation block between sequenced timepoints are not directly observable, and are not included in the numerator of $G(t)$. Still, we find a remarkably strong signal that greater than 20% of total variation in allele frequency change over 60 generations is directly the result of linked selection.

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

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

where A and B here are two replicate labels.

We've calculated the convergent correlation for all rows of the replicate covariance matrices, which unlike temporal covariance matrices have diagonal elements that are also covariances (e.g. $\text{Cov}(\Delta p_{t,A}, \Delta p_{t,B})$). Like temporal covariances, we visualize these through time (Figure 2 A), with each line representing the convergent correlation from a particular reference generation s as it varies with t (shown on the x-axis). In other words, each of the colored lines corresponds to the like colored row of the convergence correlation matrix (upper left in Figure 2 A). We find these decay very quickly, from an initial convergence correlation coefficient of about 0.1 (block bootstrap confidence intervals XXX; see the note in Supplementary Material XXX why these are so narrow), to around 0.01 within 20 generations.

A benefit of between-replicate covariances is unlike temporal covariances, these can be calculated with only two sequenced timepoints and a replicated study design. This allowed us to assess the impact of linked selection in driving convergent patterns of allele frequency change across replicate populations in two other studies. First, we reanalyzed the selection experiment of Kelly and Hughes (2019), which evolved three replicate populations of *Drosophila simulans* for 14 generations in a novel laboratory environment. Since each replicate was exposed to the same selection pressure and share linkage disequilibria common to the original natural founding population, we expected each of the three replicate populations to have positive between-replicate covariances. We find all three pairwise between-replicate covariances are positive and statistically significant ($p < XXX$, estimated via block bootstrap). We estimate the convergent correlation coefficient across these replicates as 0.36 (95% block-bootstrap confidence interval [0.31, 0.40]).

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

Paragraph about large effect loci.

Finally, we observed that in the longest run evolve-and-resequence study we analyzed (Barghi et al. 2019), some temporal covariances appeared to become negative at future timepoints (see the first two rows in Figure 1 A). This could be caused by advantageous fitness backgrounds later becoming disadvantageous, either due to frequency-dependent selection, recombination breaking up coupled beneficial and deleterious alleles, or an environmental change in the direction of selection. However, the signal of a change in the direction of selection localized to a particular region of the genome may be washed out when we calculate genome-wide temporal covariances. Averaging over such large numbers of loci is necessary to detect a signal of covariance in allele frequency changes when both drift (inflated in evolve-and-sequence studies which use small population sizes) and sampling variance can create spurious covariances. To address this limitation, we calculated temporal covariances over XXX Mb genomic windows, with on average XXX loci per window. While the covariance of each tile is noisy, we can analyze the genome-wide distribution of tile covariances for changes through time. We would like to compare the distribution of these tile covariances to a neutral expectation, but a theoretically-determined neutral null distribution would be inappropriate given complex dependencies across the genome due to LD (which can occur over long physical distances in ER and selection studies, Baldwin-Brown et al. 2014; Nuzhdin and Turner 2013).

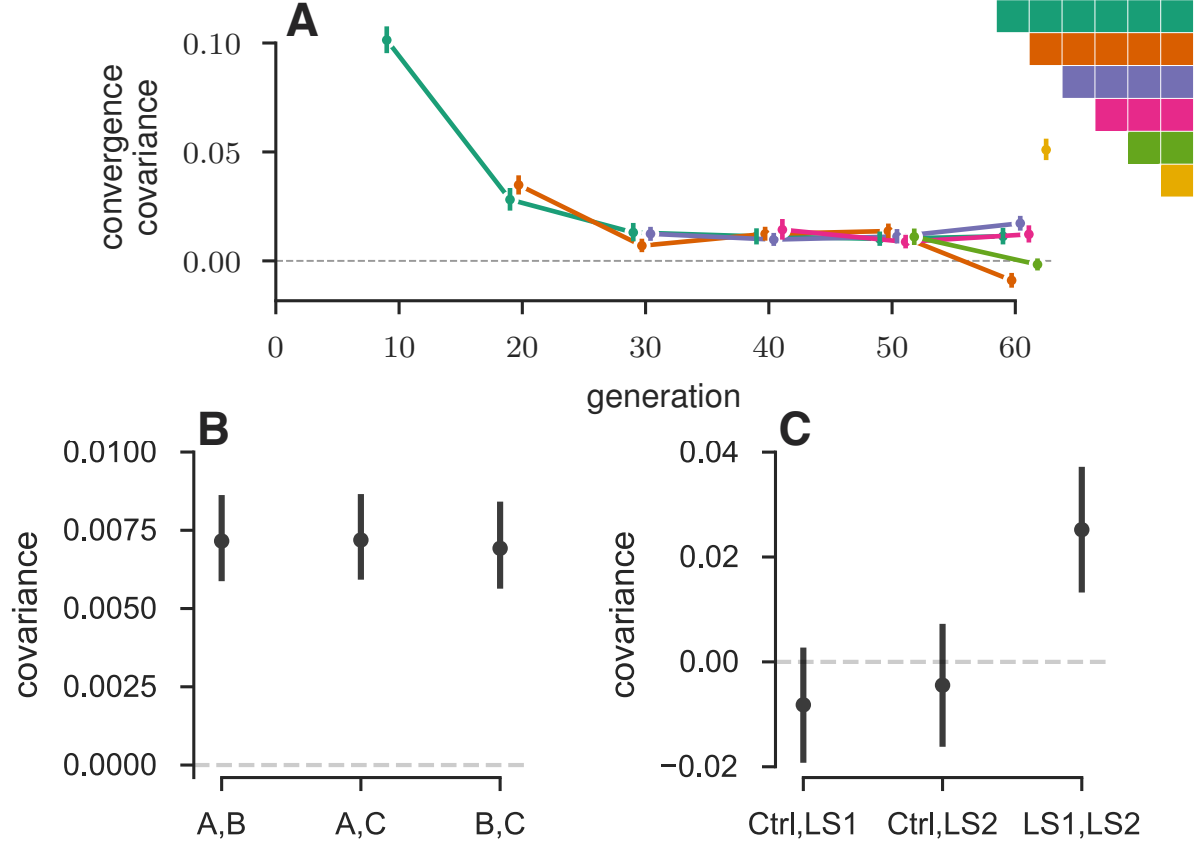


Figure 2: A: The convergence correlation, averaged across replicate pairs, through time. Each line represents the convergence correlation $\text{cor}(\Delta p_s, \Delta p_t)$ from a starting reference generation s to a later time t , which varies along the x-axis; each line corresponds to a row of the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

3 Appendix

4 Sampling Bias Corrections

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

$$\text{Var}(\tilde{p}_t) = \mathbb{E}(\text{Var}(\tilde{p}_t|p_t)) + \text{Var}(\mathbb{E}(\tilde{p}_t|p_t)) \quad (2)$$

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

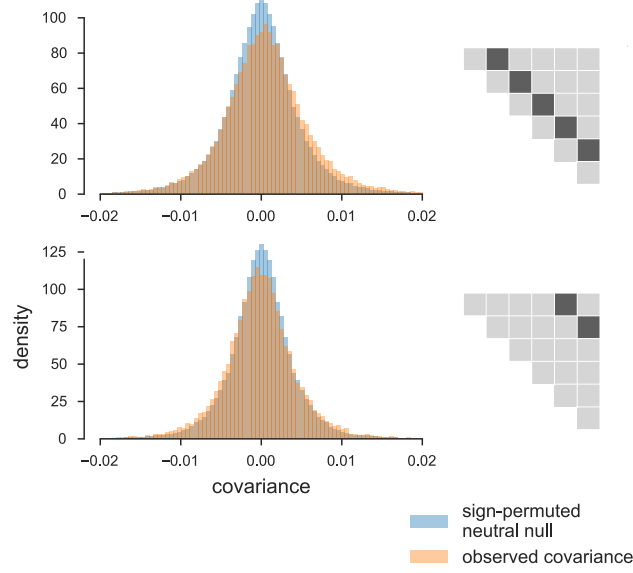


Figure 3

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 (Buffalo and Coop 2019). We can write

$$V(p_t) = V(p_0 + (p_1 - p_0) + (p_2 - p_1) + \dots + (p_t - p_{t-1})) \quad (4)$$

$$= V(p_0 + \Delta p_0 + \Delta p_1 + \dots + \Delta p_{t-1}) \quad (5)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{d_0-1}{d_0} \frac{\text{Var}(\tilde{p}_1 - \tilde{p}_0)}{\tilde{p}_0(1-\tilde{p}_0)} - \frac{(d_0-1)}{d_0(d_1-1)} \frac{\tilde{p}_1(1-\tilde{p}_1)}{\tilde{p}_0(1-\tilde{p}_0)} - \frac{1}{d_0} + \frac{C}{N} = \frac{\text{Var}(\Delta_H p_0)}{p_0(1-p_0)} + \frac{1}{2N} \quad (17)$$

4.1 Individual and depth sampling process

$X_t \sim \text{Binom}(n_t, p_t)$ where X_t is the count of alleles and n_t is the number of diploids sampled at time t . Then, these individuals are sequenced at a depth of d_t , and $Y_t \sim \text{Binom}(d_t, X_t/n_t)$ reads have

the tracked allele. We let $\tilde{p}_t = Y_t/d_t$ be the observed sample allele frequency. Then, the sampling noise is

$$\text{Var}(\tilde{p}_t|p_t) = \mathbb{E}(\text{Var}(\tilde{p}_t|X_t)) + \text{Var}(\mathbb{E}(\tilde{p}_t|X_t)) \quad (18)$$

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

(see also Jónás et al. 2016).

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

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

$$+ \sum_{0 \leq i < j}^{t-1} \text{Cov}(\Delta p_i, \Delta p_j) \quad (22)$$

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

$$\frac{n_t d_t}{(n_t - 1)(d_t - 1)} \tilde{p}_t(1 - \tilde{p}_t) \quad (23)$$

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

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

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

$$= \tilde{p}_t(1 - \tilde{p}_t) \frac{d_t + n_t - 1}{(n_t - 1)(d_t - 1)} + \tilde{p}_0(1 - \tilde{p}_0) \frac{d_0 + n_0 - 1}{(n_0 - 1)(d_0 - 1)} + \quad (25)$$

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

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

4.2 Covariance Correction

We also need to apply a bias correction to the temporal covariances (and possibly the replicate covariances if the initial sample frequencies are all shared).

The basic issue is that $\text{Cov}(\Delta\tilde{p}_t, \Delta\tilde{p}_{t+1}) = \text{Cov}(\tilde{p}_{t+1} - \tilde{p}_t, \tilde{p}_{t+2} - \tilde{p}_{t+1})$, and thus shares the sampling noise of timepoint $t + 1$. Thus acts to bias the covariance by subtracting off the noise variance term of $\text{Var}(\tilde{p}_{t+1})$, so we add that back in.

4.3 Variance-Covariance Matrix Correction

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

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

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

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

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

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

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

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

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

where division here is elementwise. Thus, \mathbf{B} is a $R \times (T + 1)$ matrix. As explained in XXX, each temporal covariance submatrix $\mathbf{Q}_{r,r}$ requires two corrections.

$$Q_{i,j} = \begin{cases} Q'_{t,s} - b_t - b_s, & \text{if } t = s \\ Q'_{t,s} + b, & \text{if } |i - j| = 1 \end{cases}$$

$$Q_{i,j} = \text{Cov}(\Delta p_i, \Delta p_j)$$

Additionally, in some study designs, the frequencies from the first timepoint

$$\mathbf{Q} = \mathbf{Q}' - \text{diag}(\text{flat}(\mathbf{B}_{:,2:(T+1)})) - \text{diag}(\text{flat}(\mathbf{B}_{:,1:T})) + \text{offdiag}_1(\text{flat}(\mathbf{B}_{:,2:T})) + \text{offdiag}_{-1}(\text{flat}(\mathbf{B}_{:,2:T})) \quad (31)$$

where $\text{diag}(\mathbf{x})$ is an operation that takes the vector \mathbf{x} and places them along the diagonal of a matrix. Similarly, $\text{offdiag}_k(\mathbf{x})$ places vector \mathbf{x} along the offdiagonal where for row i , and column j , $j - i = k$. We represent subset of columns s through t of a matrix \mathbf{B} as $\mathbf{B}_{:,s:t}$.

4.4 Block Bootstrap Procedure

4.5 Correction

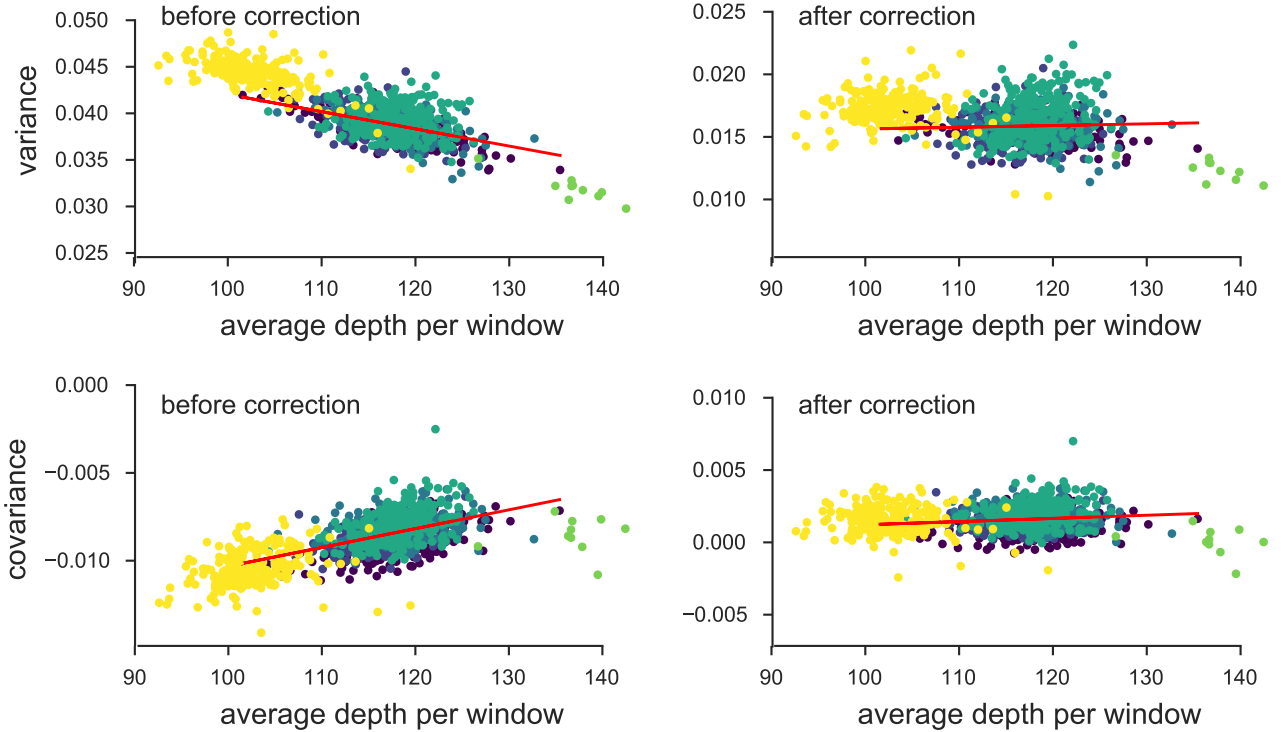


Figure 4