# CLEAR: Composition of Likelihoods for Evolve And Resequence Experiments

Arya Iranmehr[1], Ali Akbari[1], Christian Schlötterer[2], and Vineet Bafna[3]

[1]Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA.
[2]Institut für Populationsgenetik, Vetmeduni, Vienna, Austria.
[3]Computer Science and Engineering, University of California, San Diego, La Jolla, CA, USA.

**Abstract**

The advent of next generation sequencing technologies has made whole-genome and whole-population sampling possible, even for eukaryotic organisms. With this development, experimental evolution studies can be designed to observe molecular evolution "in-action" by Evolving-and-Resequencing (E&R) populations.Among other applications, E&R studies can be used to locate the genes and variants responsible for genetic adaptation. To analyze E&R datasets, majority of the existing literature on time-series data analysis, is mainly devoted to the settings with large population sizes, allele frequency as input data and wide time spans, these assumptions does not hold in many E&R studies.

In this article, we propose a method, Composition of Likelihoods for Evolve-And-Resequence experiments (CLEAR), to identify selection in short-term (as well as long-term), E&R of *small* sexual populations. CLEAR takes whole-genome sequence of pool of individuals (pool-seq) as input, and properly addresses heterogeneous ascertainment bias, due to uneven coverages. CLEAR also provides unbiased estimates of model parameters, including selection strength and over-dominance, and population size, while being computationally efficient. Extensive simulations show that CLEAR achieves higher power in detecting and localizing selection over a wide range of parameters. Moreover, we show CLEAR statistic is robust to variation of coverage. We applied CLEAR statistic to previously published datasets, including, data from a study of *D. melanogaster* adaptation to alternating temperaturesand study of outcrossing Yeast populations. Result XXXX

## 1 Introduction

Natural selection is a key force in evolution, and a mechanism by which populations can adapt to external 'selection' constraints. Examples of adaptation abound in the natural world [27], including for example, classic examples like lactose tolerance in Northern Europeans [12], human adaptation to high altitudes [72, 89], but also drug resistance in pests [19], HIV [30], cancer [35, 90], malarial parasite [5, 56], and other antibiotic resistance [73]. In these examples, understanding the genetic basis of adaptation can provide actionable information, underscoring the importance of the problem.

Experimental evolution refers to the study of the evolutionary processes of a model organism in a controlled [9, 14, 37, 47, 48, 59, 60] or natural [7, 11, 20, 21, 52, 65, 88] environment. Recent advances in whole genome sequencing have enabled us to sequence populations at a reasonable cost even for large genomes. Perhaps more important for experimental evolution studies, we can now evolve and resequence (E&R) multiple replicates of a population to obtain *longitudinal time-series data*, in order to investigate the dynamics of evolution at molecular level. Although constraints

such as small sizes, limited timescales, and oversimplified laboratory environments may limit the interpretation of E&R results, these studies are increasingly being used to test a wide range of hypotheses [44] and have been shown to be more predictive than static data analysis [16, 22, 69]. In particular, longitudinal E&R data is being used to estimate model parameters including population size [43, 62, 78, 83, 84, 86], strength of selection [15, 39, 40, 51, 54, 74, 78], allele age [51] recombination rate [78], mutation rate [8, 78], quantitative trait loci [6] and for tests of neutrality hypotheses [11, 17, 29, 78].

While objectives, designs and organisms of E&R studies can be entirely different [8, 70], here we restrict our attention to the adaptive evolution of multi-cellular sexual organisms. For simplicity, we assume fixed population size, and for the most part, positive single locus selection (only one favored mutation). This regime has been considered earlier, typically with *D. melanogaster* as the model organism of choice, to identify adaptive genes in longevity and aging [17, 66] (600 generations), courtship song [81] (100 generations), hypoxia tolerance [91] (200 generations), adaptation to new laboratory environments [32, 59] (59 generations), egg size [42] (40 generations), C virus resistance [53] (20 generations), and dark-fly [41] (49 generations).

The task of identifying genetic adaptation can be addressed at different levels of specificity. At the coarsest level, identification could simply refer to deciding whether some genomic region (or a gene) is under selection or not. In the following, we refer to this task as *detection*. In contrast, the task of *site-identification* corresponds to the process of finding the favored mutation/allele at nucleotide level. Finally, *estimation of model parameters*, such as strength of selection and overdominance at the site, can provide a comprehensive description of the selection process.

A wide range of computational methods [82] have been developed to detect regions under positive selection. A majority of the existing methods focus on static data analysis; analysis of a single sample of the population at a specific time, either during the sweep, or subsequent to fixation of the favored allele. Static analysis is focused on reduction in genetic diversity [28, 33, 67, 77] shift in allele-frequencies, prevalence of long haplotypes [68, 82], population differentiation [17, 36, 38] in multiple-population data and others. Many existing methods use the Site Frequency Spectrum (SFS, see S2 Fig) to identify departure from neutrality. Classical examples including Tajima's $D$ [77], Fay and Wu's $H$ [28], Composite Likelihood Ratio [58], were all shown to be weighted linear combinations of the SFS values [1]. While successful, these methods are prone to both, false negatives [55], and also false-discoveries due to confounding factors such as demography, including bottleneck and population expansions, and ascertainment bias [3, 55, 57, 63, 64]. Nevertheless, SFS based tests continue to be used successfully, often in combination with other tests [3, 82]. One of the contributions of this paper is the extension of SFS based methods to analyze time-series data, and the identification of selection regimes where these methods perform well.

Relative to the analysis of static samples, fewer tests-of-selection for dynamic time-series data have been proposed. Often, existing tests for static data are adopted for dynamic data with two time-points. Zhu *et al.* [91] used the ratio of the estimated population size of case and control populations to compute test statistic for each genomic region. Burke *et al.* [17] applied Fisher exact test to the last observation of data on case and control populations. Orozco-terWengel *et al.* [59] used the Cochran-Mantel-Haenszel (CMH) test [2] to detect SNPs whose read counts change consistently across all replicates of two time-point data. Turner *et al.* [81] proposed the diffStat statistic to test whether the change in allele frequencies of two populations deviate from the distribution of change in allele frequencies of two drifting populations. Bergland *et al.* [11] applied $F_{st}$ to populations throughout time to signify their differentiation from ancestral (two time-point data) as well as geographically different populations. Jha *et al.* [42] computed test statistic of generalized linear-mixed model directly from read counts.

Early *direct* methods for analyzing time-series data devoted to estimate population size in

neutral populations [4, 10, 13, 83, 86], using statistical (HMM) and population genetics (Coalescent) models. The first effort to properly address the problem of parameter estimation in time series selection data was done by Bollback *et al.* [15]. They provided a diffusion approximation to the continues Wright Fisher Markov process and estimated $s$ numerically for large population sizes. Arya Note: (this is the only additional reference I added.) Malaspinas *et al.* [51] extended Bollback et al.s method to estimate allele age in ancient-DNA (aDNA). Steinrücken and Song [74] proposed a general diploid selection model which takes into account of dominance of the favored allele and approximates likelihood analytically. Mathieson and McVean [54] adopted HMMs to structured populations and estimated parameters using an Expectation Maximization (EM) procedure, on discretized allele frequency. Feder *et al.* [29] modeled increments in allele frequency with a Brownian motion process, proposed the Frequency Increment Test (FIT). More recently, Topa *et al.* [80] proposed a Gaussian Process (GP) for modeling single-locus time-series pool-seq data. Terhorst *et al.* [78] extended GP to compute joint likelihood of multiple loci under null and alternative hypotheses. Recently, schraiber *et al.* [71] proposed a Bayesian framework to estimate parameters using Monte Carlo Markov chain sampling.

While existing methods have been successfully applied to their corresponding application, they make several key assumptions which does not hold in many of E&R studies. First, they assume that the underlying population size is large, if not infinity, (See Table 3 in [50] for illustration). As a result, they provide continues state models for dynamics of allele frequencies. In addition, most these methods, originally designed to process wide time spans, i.e. aDNA studies. Also, they assume input data is allele frequency or at least is *sampled* allele frequency, where ascertainment bias is uniform along genome.

This manuscript, explicitly posits a "small-population-size" assumption on Williamson *et al.* [86] and Bollback *et al.*'s model [15], and as a consequence, the resulting model become a discrete state (frequency) model. We show that for small population sizes discrete models can compute likelihood exactly, which make a difference in the statistical performance, especially in short term experiments. Additionally, we add another level of sampling-noise to the traditional HMM, which models heterogeneous ascertainment bias due to uneven coverages among variants. We show for a wide range of parameters that CLEAR provides higher power for detecting selection, is robust to ascertainment bias due to coverage heterogeneity, estimates model parameters consistently, and localizes favored allele more accurately compared to the state-of-the-art methods, while being computationally efficient.

## 2 Materials and Methods

To identify the genes and variants that are responding to the selection pressure, we consider the following statistical procedure:

(i) **Estimating population size.** The procedure starts by finding the maximum likelihood estimate of population size, $\widehat{N}$, over the whole genome.

(ii) **Estimating selection parameters.** Given $\widehat{N}$, maximizing the likelihood of the time series data w.r.t. selection and overdominance parameters $s, h$, for each polymorphism.

(iii) **Computing likelihood statistics.** For each variant, it calculates the log-odds ratio of the likelihood of selection model to the likelihood of neutral evolution/drift model. Likelihood ratios in a genomic region are combined to compute the CLEAR statistic.

(iv) **Hypothesis testing.** The null distribution of the CLEAR (or likelihood ratio) statistics are computed on a set of whole-genome (single locus, respectively) drift simulations with

population size of $\widehat{N}$, and variant starting frequency and coverage of the experimental data. Given the null distribution of statistics, $p$-values and corresponding False Discovery Rate (FDR) are calculated. The overlapping genes with the regions (or variants) that satisfy FDR criterion, will be reported for functional analysis or imported to the Gene Set Enrichment Analysis (GSEA).

In the subsequent of this section, we outline different steps of the statistical procedure.

## 2.1 Estimating Population Size

Estimating population sizes from temporal neutral evolution data has been previously studied [4, 15, 43, 78, 86]. Existing methods are well designed for when the allele frequencies are computed from a finite sample, that is the ascertainment bias is uniform over the genome. However, in the case of pool-seq data in addition to uniform ascertainment bias, each variant is sampled at different rate, due to finite sequencing coverage. In addition, bulk of existing models [15, 29, 78, 80] are designed for large populations, and model frequency as a continues quantity. However, we show that smooth approximations is inadequate for small populations, low starting frequencies and sparse sampling (in time) that are typical in experimental evolution (see Results, Fig 3A-C, and Fig 2). To this end, we model Wright-Fisher Markov process for generating pool-seq data (S1 Fig) via a *discrete* HMM, Fig 1-B. In order to find an estimate of population size, we first need to find likelihood of the population size given neutral pool-seq data.

**Likelihood for Neutral Model.** Consider a neutrally evolving diploid population with fixed size of $N$ individuals where $\nu_t$ denotes allele frequency of the derived allele at generation $t$. Experimental evolution for $R$ replicates is conducted so that at generations $\mathcal{T} = \{\tau_i : 0 \leq \tau_0 < \tau_1, \ldots < \tau_T\}$, $n$ individuals are chosen for sequencing.

At the highest level, the consecutive allele frequencies of the population in a fixed-size Wright-Fisher model evolves by Binomial sampling

$$\nu_0 \sim \pi, \qquad 2N\nu_t|\nu_{t-1} \sim \text{Binomial}(2N, \nu_{t-1}) \tag{1}$$

where $\pi$ is the global (marginal) distribution of allele frequencies in the base population. In general, $\pi$ depends on demographic history of the founder lines. Here we simply assume is $\pi$ is the site frequency spectrum of fixed sized neutral population S2 Fig.

To compute distributions after $\tau$ transitions, it is enough to specify the $2N \times 2N$ transition matrix $P^{(\tau)}$, where $P^{(\tau)}[i,j]$ denotes probability of change in allele frequency from $i/2N$ to $j/2N$ in $\tau$ generations:

$$P^{(1)}[i,j] = \Pr\left(\nu_{t+1} = \frac{j}{2N} \middle| \nu_t = \frac{i}{2N}\right) = \binom{2N}{j} \nu_t^j (1-\nu_t)^{2N-j}, \tag{2}$$

$$P^{(\tau)} = P^{(\tau-1)} P^{(1)} \tag{3}$$

As at each generation $n$ out of $N$ individuals are randomly selected for sequencing. The sampled allele frequencies, $\{y_t\}_{t\in\mathcal{T}}$, are also Binomially distributed

$$2ny_t \sim \text{Binomial}(2n, \nu_t) \tag{4}$$

We introduce the $2N \times 2n$ sampling matrix $Y$, where $Y[i,j]$ stores the probability that the sample allele frequency is $i/2n$ given that the true allele frequency is $i/2N$.

We denote the pool-seq data for that variant as $\{x_t = \langle c_t, d_t\rangle\}_{t\in\mathcal{T}}$ where $d_t, c_t$ represent the read depth, and the read count of the derived allele, respectively, at time $\tau_t$. Let $\{\lambda_t\}_{t\in\mathcal{T}}$ be the sequencing coverage at different generations, then, the observed data are sampled according to

$$d_t \sim \text{Poisson}(\lambda_t), \qquad\qquad c_t \sim \text{Binomial}(d_t, y_t) \tag{5}$$

where the the emission probabilities for a observed tuple $x_t = \langle d_t, c_t\rangle$ is

$$\mathbf{e}_i(x_t) = \binom{d_t}{c_t}\left(\frac{i}{2n}\right)^{c_t}\left(1-\frac{i}{2n}\right)^{d_t-c_t}. \tag{6}$$

For $1 \le t \le T$, let $\alpha_{t,i}$ denote the probability of emitting $x_1, x_2, \ldots, x_t$ and reaching state $i$ at $\tau_t$. Then, $\alpha_{t,i}$ can be computed using the forward-procedure [23]:

$$\begin{aligned}\alpha_{t,i} &= \left(\sum_{1\le j\le 2N} \alpha_{t-1,j}\; P^{(\delta_t)}[j,i]\right) Y\mathbf{e}_i(x_t)\\ \alpha_t &= \text{Diag}(\alpha_{t-1})P^{(\delta_t)}Y\mathbf{e}(x_t)\end{aligned} \tag{7}$$

where $\delta_t = \tau_t - \tau_{t-1}$. The joint likelihood of the observed data from $R$ independent observations is given by

$$\Pr(\{\boldsymbol{x}^{(r)}\}|N,n) = \mathcal{L}(N|\{\boldsymbol{x}^{(r)}\},n) = \prod_{r=1}^{R}\mathcal{L}(N|\boldsymbol{x}^{(r)},n) = \prod_{r=1}^{R}\sum_i \alpha_{T,i}^{(r)} \tag{8}$$

where $\boldsymbol{x} = \{x_t\}_{t\in\mathcal{T}}$. The graphical model and the generative process for which data is being generated is depicted in Fig 1-B and S1 Fig, respectively.

Finally, the last step is to find the $\widehat{N}$ in which maximizes the likelihood of the all the $M$ variants in whole genome:

$$\widehat{N} = \arg\max_N \prod_i^M \mathcal{L}(N|\{\boldsymbol{x}_i^{(r)}\} \tag{9}$$

## 2.2 Estimating Selection Parameters

**Likelihood for Selection Model.** Assume that the site is evolving under selection constraints $s \in \mathbb{R}$, $h \in \mathbb{R}_+$, where $s$ and $h$ denote selection strength and overdominance parameters , respectively. By definition, the relative fitness values of genotypes 0|0, 0|1 and 1|1 are given by $w_{00} = 1$, $w_{01} = 1 + hs$ and $w_{11} = 1 + s$. Recall that $\nu_t$ denotes the frequency of the site at time $\tau_t \in \mathcal{T}$. Then, $\nu_{t^+}$, the frequency at time $\tau_t + 1$ (one generation ahead), can be estimated using:

$$\begin{aligned}\hat{\nu}_{t^+} = \mathbb{E}[\nu_{t^+}|s,h,\nu_t] &= \frac{w_{11}\nu_t^2 + w_{01}\nu_t(1-\nu_t)}{w_{11}\nu_t^2 + 2w_{01}\nu_t(1-\nu_t) + w_{00}(1-\nu_t)^2}\\ &= \nu_t + \frac{s(h+(1-2h)\nu_t)\nu_t(1-\nu_t)}{1+s\nu_t(2h+(1-2h)\nu_t))}.\end{aligned} \tag{10}$$

The machinery for computing likelihood of the selection parameters is identical to that of population size, except for transition matrices. Hence, here we only describe the definition transition matrix $Q_{s,h}$ of the selection model. Let $Q_{s,h}^{(\tau)}[i,j]$ denote the probability of transition from $i/2N$ to $j/2N$ in $\tau$ generations, then (See [25], Pg. 24, Eqn. 1.58-1.59):

$$Q_{s,h}^{(1)}[i,j] = \Pr\left(\nu_{t^+} = \frac{j}{2N}\,\middle|\,\nu_t = \frac{i}{2N}; s,h,N\right) = \binom{2N}{j}\hat{\nu}_{t^+}^j(1-\hat{\nu}_{t^+})^{2N-j} \tag{11}$$

$$Q_{s,h}^{(\tau)} = Q_{s,h}^{(\tau-1)}Q_{s,h}^{(1)} \tag{12}$$

The maximum likelihood estimates are given by

$$\widehat{s}, \widehat{h} = \arg\max_{s,h} \prod_{i}^{M} \mathcal{L}(s, h | \{\boldsymbol{x}_i^{(r)}, \widehat{N}\} \tag{13}$$

The parameters in Eqs. 9, 13 are optimized using grid search. By broadcasting and vectorizing the grid search operations across all variants, the genome scan on millions of polymorphisms can be done in significantly smaller time than iterating a numerical optimization routine for each variant(see Results and Fig 4).

## 2.3 Empirical Likelihood Ratio Statistics

The likelihood ratio statistic for testing directional selection, to be computed for each variant, is given by

$$H = -2 \log \left( \frac{\mathcal{L}(\widehat{s}, 0.5 | \{\boldsymbol{x}^{(r)}\}, \widehat{N})}{\mathcal{L}(0, 0.5 | \{\boldsymbol{x}^{(r)}\}, \widehat{N})} \right) \tag{14}$$

Similarly we can define test statistic for testing if selection is over-dominant

$$D = -2 \log \left( \frac{\mathcal{L}(\widehat{s}, \widehat{h} | \{\boldsymbol{x}^{(r)}\}, \widehat{N})}{\mathcal{L}(\bar{s}, 0.5 | \{\boldsymbol{x}^{(r)}\}, \widehat{N})} \right), \text{ where } \quad \bar{s} = \arg\max_{s} \prod_{i}^{M} \mathcal{L}(s, 0.5 | \{\boldsymbol{x}_i^{(r)}, \widehat{N}\}. \tag{15}$$

While extending the single-locus WF model to a multiple linked-loci can improve the power of the model [78], it is computationally and statistically expensive to compute exact likelihood. haplotype resolved data, which pool-seq does not provide. Instead, similar to Nielse *et al* [58], we calculate Composite Likelihood Ratio score for a genomic region.

$$\mathcal{H} = \frac{1}{|L|} \sum_{\ell \in L} H_\ell. \tag{16}$$

where $L$ to be a collection of segregating sites and $H_\ell$ is the likelihood ratio score based for each variant $\ell$ in $L$. The optimal value of the hyper-parameter $L$ depends upon a number of factors, including, initial frequency of the favored allele, recombination rates, initial linkage of the favored allele to its surrounding variation, population size, coverage, and time since the onset of selection (duration of the experiment). However, we provide a heuristic choose size of $L$ for an experiment.

In general, as selection acts locally in the genome, size of $L$ have a direct effect on the power. For instance, when $L$ is chosen to be a large region (e.g. chromosome), power will be degraded since distribution of null and alternative $\mathcal{H}$ statistics converge together. Hence, we choose $L$ to be the largest, such that it provides enough discoveries that satisfies experiment's FDR.

## 2.4 Hypothesis Testing

**Single-Locus.** Under neutrality, Wilkss theorem [85] states that $H \sim \mathcal{X}_1^2$, asymptotically, and $p$-values can be computed directly. However, Feder *et al.* [29] showed that when the number of independent samples (replicates) is small, $\mathcal{X}^2$ is a crude approximation to the true null distribution and underestimates FDR. They suggested to compute $p$-values based on the empirical distribution of statistic on simulations with $\widehat{N}$ . Here, take the same approach and conduct single locus drift simulations starting from initial frequencies of the experimental data. Then we sample read counts

of the derived allele given coverage and allele frequency, see S1 Fig for step-by-step procedure. Test statistic is then computed for simulated drifting pool-seq data and $p$-value of an experimental samples is computed as fraction of null statistics that have a higher or equal value than testing statistic. Finally, we use Storey and Tibshirani's method [76], to control for False Discovery Rate in multiple testing.

**Regions.** To compute null distribution of $\mathcal{H}$ one should compute $\mathcal{H}$ on a set of neutral genome-wide neutral simulations. Also, as selection is expected to have local effect on the genome, we normalize $\mathcal{H}$ with respect to each chromosome both in simulated and experimental data in computing $p$-values:

$$\mathcal{H}_i^* = \frac{\mathcal{H}_i - \mu_{\mathcal{C}}}{\sigma_{\mathcal{C}}}, \qquad \forall i \in \mathcal{C} \tag{17}$$

where $\mu_{\mathcal{C}}$ and $\sigma_{\mathcal{C}}$ are the chromosome-wise mean and standard deviation. $p$-values and FDR can be computed in the same regime as single locus. After discovering intervals that satisfy FDR requirement, we further select those variants within selected intervals that their individual score $H$ is significantly large and find the intersected genes accordingly.

## 2.5 Simulations

We performed extensive simulations using parameters that have been used for *D. melanogaster* experimental evolution [46]. See also Fig 1-A for illustration. To implement real world pool-seq experimental evolution, we conducted simulations as follows:

I. **Creating initial founder line haplotypes.** Using `msms` [26], we created neutral populations for $F$ founding haplotypes with command `$./msms <F> 1 -t <2μLNe> -r <2rNeL> <L>`, where $F = 200$ is number of founder lines, $N_e = 10^6$ is effective population size, $r = 2 \times 10^{-8}$ is recombination rate, $\mu = 2 \times 10^{-9}$ is mutation rate and $L = 50K$ is the window size in base pairs which gives $\theta = 2\mu N_e L = 200$ and $\rho = 2N_e r L = 2000$.

II. **Creating initial diploid population.** To simulate experimental evolution of diploid organisms, initial haplotypes were first cloned to create $F$ diploid homozygotes. Next, each diploid individual was cloned $N/F$ times to yield diploid population of size $N$.

III. **Forward Simulation.** We used forward simulations for evolving populations under selection. We also consider selection regimes which the favored allele is chosen from standing variation (not *de novo* mutations). Given initial diploid population, position of the site under selection, selection strength $s$, number of replicates $R = 3$, recombination rate $r = 2 \times 10^{-8}$ and sampling times $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$, `simuPop` [61] was used to perform forward simulation and compute allele frequencies for all of the $R$ replicates. For hard sweep (respectively, soft sweep) simulations we randomly chose a site with initial frequency of $\nu_0 = 0.005$ (respectively, $\nu_0 = 0.1$) to be the favored allele.

IV. **Sequencing Simulation.** Give allele frequency trajectories we sampled depth of each site identically and independently from Poisson($\lambda$), where $\lambda \in \{30, 100, 300\}$ is the coverage for the experiment. Once depth $d$ is drawn for the site with frequency $\nu$, the number of reads $c$ carrying the derived allele are sampled according to Binomial$(d, \nu)$. For experiments with finite depth the tuple $\langle c, d \rangle$ is the input data for each site.

# 3 Results

**Modeling Allele Frequency Trajectories in Small Populations.** We first tested the goodness of fit of the discrete versus continuous models in modeling allele frequency trajectories, under general E&R parameters. For this purpose, we conducted 100K simulations with two time samples $\mathcal{T} = \{0, \tau\}$ where $\tau \in \{1, 10, 100\}$ is the parameter controlling the density of sampling in time. In addition, we repeated simulations for different values of starting frequency $\nu_0 \in \{0.005, 0.1\}$ (i.e., hard and soft sweep) and selection strength $s \in \{0, 0.1\}$ (i.e., neutral and selection). Then, given initial frequency $\nu_0$, we computed the expected distribution of the frequency of the next sample $\nu_\tau$ under two models and compared them with empirical distributions calculated from simulated data. Fig 2A-F shows that Brownian motion (continuous model) is inadequate when $\nu_0$ is far from 0.5, or when sampling times are sparse ($\tau > 1$). If the favored allele arises from standing variation in a neutral population, it is unlikely to have frequency close to 0.5, and the starting frequencies are usually much smaller (see S2 Fig). Moreover, in typical *D. melanogaster* experiments for example, sampling is sparse. Often, the experiment is designed so that $10 \leq \tau \leq 100$ [32, 46, 59, 91].

In contrast to the Brownian motion results, Markov chain can provide predictions when the the allele is under selection. In addition Fig 2A-M also shows that Markov chain predictions (Eq. 12) are highly consistent with empirical data for a wide range of simulation parameters.

**Detection Power.** We compared the performance of CLEAR against other methods for detecting selection. For each method we calculated detection power as the percentage of true-positives identified with false-positive rate $\leq 0.05$. For each configuration (specified with values for selection coefficient $s$, starting allele frequency $\nu_0$ and coverage $\lambda$), power of each method is evaluated over 2000 distinct simulations, half of which modeled neutral evolution and the rest modeled positive selection.

We compared the power of CLEAR with Gaussian process (GP) [78], FIT [29], and CMH [2] statistics. FIT and GP convert read counts to allele frequencies prior to computing the test statistic. CLEAR shows the highest power in all cases and the power stays relatively high even for low coverage (Fig 3 and S1 Table). In particular, the difference in performance of CLEAR with other methods is pronounced when starting frequency is low. The advantage of CLEAR stems from the fact that favored allele with low starting frequency might be missed by low coverage sequencing. In this case, incorporating the signal from linked sites becomes increasingly important. We note that methods using only two time points, such as CMH, do relatively well for high selection values and high coverage. However, the use of time-series data can increase detection power in low coverage experiments or when starting frequency is low. Moreover, time-series data provide means for estimating selection parameters $s, h$ (see below). Finally, as CLEAR is robust to change of coverage, our results (Fig 3B,C) suggest that taking many samples with lower coverage is preferable to sparse sampling with higher coverage.

**Site-identification.** In general, localizing the favored variant, using pool-seq data is a nontrivial task [79]. We used the simple approach of ranking each site in a region detected as being under selection. The variants were ranked according to the likelihood ratio scores (Eqn. 14). For each setting of $\nu_0$ and $s$, we conducted 1000 simulations and computed the rank of the favored mutation in each simulation. The cumulative distribution of the rank of the favored allele in 1000 simulation for each setting (Fig 5) shows that CLEAR outperforms other statistics.

An interesting observation is revisiting the contrast between site-identification and detection [49, 79]. When selection coefficient is high, detection is easier (Fig 3A-F), but site-identification is harder due to the high LD between hitchhiking sites and the favored allele (Fig 5A-F). Moreover, site-identification is harder in hard sweep scenarios relative to soft sweeps. For example, when coverage

$\lambda = 100$ and selection coefficient $s = 0.1$, the detection power is 75% for hard sweep, but 100% for soft sweep (Fig 3B-E). In contrast, the favored site was ranked as the top in 14% of hard sweep cases, compared to and 95% of soft sweep simulations.

**Estimating Parameters.** CLEAR computes the selection parameters $\hat{s}$ and $\hat{h}$ as a byproduct of the hypothesis testing. We computed bias of selection fitness $(s - \hat{s})$ and overdominance $(h - \hat{h})$ for of CLEAR and GP in each setting. The distribution of the error (bias) for $100\times$ coverage is presented in Fig 6 for different configurations. S5 Fig and S6 Fig provide the distribution of estimation errors for $30\times$, and infinite coverage, respectively. For hard sweep, CLEAR provides estimates of $s$ with lower variance of bias (Fig 6A). In soft sweep, GP and CLEAR both provide unbiased estimates with low variance (Fig 6B). Fig 6C-D shows that CLEAR provides unbiased estimates of $h$ as well.

**Running Time.** As CLEAR does not compute exact likelihood of a region (i.e., does not explicitly model linkage between sites), the complexity of scanning a genome is linear in number of polymorphisms. Calculating score of each variant requires and $\mathcal{O}(TRN^2)$ computation for $\mathcal{H}$. However, most of the operations are can be vectorized for all replicates to make the effective running time for each variant. We conducted 1000 simulations and measured running times for computing site statistics $H$, FIT, CMH and GP with different number of linked-loci. Our analysis reveals (Fig 4) that CLEAR is orders of magnitude faster than GP, and comparable to FIT. While slower than CMH on the time per variant, the actual running times are comparable after vectorization and broadcasting over variants (see below).

These times can have a practical consequence. For instance, to run GP in the single locus mode on the entire pool-seq data of the *D. melanogaster* genome from a small sample (≈1.6M variant sites), it would take 1444 CPU-hours (≈ 1 CPU-month). In contrast, after vectorizing and broadcasting operations for all variants operations using `numba` package, CLEAR took 75 minutes to perform an scan, including precomputation, while the fastest method, CMH, took 17 minutes.

## 3.1 Analysis of a *D. melanogaster* Adaptation to Cold and Hot Temperatures

We applied CLEAR to the data from a study of *D. melanogaster* adaptation to alternating temperatures [32, 59], where 3 replicate samples were chosen from a population of *D. melanogaster* for 59 generations under alternating 12-hour cycles of hot (28°C) and cold (18°C) temperatures and sequenced. In this dataset, sequencing coverage is different across replicates and generations (see S2 Fig of [78]) which makes variant depths highly heterogeneous (S4 Fig and S3 Fig).

We first filtered out heterochromatic, centromeric and telomeric regions [31], and those variant that have collective coverage of more that 1500 in all 13 populations: three replicates at the base population, two replicates at generation 15, one replicate at generation 23, one replicate at generation 27, three replicates at generation 37 and three replicates at generation 59. After filtering, we ended up with 1,605,714 variants.

First we estimated population size $\hat{N} = 250$ over the whole genome, Fig 10. The likelihood curves of CLEAR is sharper around the optimum compared to Bollback et. al [15] (see Supplementary Fig. 1 in [59]). While chromosomes 3L and 3R appear to have smaller population size Fig 10-D, which can taken into account of chromosome-wise significance cutoffs, we took the genome-wide population size of $\widehat{N}$ =250 to calculate $p$-values and FDR. Next, we found maximum likelihood estimates of $s$, and computed the test statistic $H$ for each variant, given $N = 250, h = 0.5$. To compute $p$-values of $H$ statistics, we calculated single locus Wright-Fisher simulations for $\widehat{N}$ =250, initial frequencies and variant depths of real data (see S1 Fig). We repeated forward simulations 50 times for whole genome, to collect ≈90M pool-seq trajectories with starting frequencies and coverage of real data. Then, $p$-value of each variant in the real data is calculated as the fraction

| CHROM | Start | End |
|---|---|---|
| 2L | 16853191 | 16964749 |
| 2R | 2740567 | 2796964 |
| 3L | 14342590 | 14500095 |
| X | 1569717 | 1655232 |
| X | 7129589 | 7231085 |

Table 1: Coordinates of the selected intervals found by CLEAR $\mathcal{H}^*$ statistic.

of null statistics that are grater than or equal of test statistic(see S9 Fig). Finally, to correct for multiple testing, we used Storey & Tibshirani [76] method to compute FDR. Fig 7 shows the distribution of 804 variants passing FDR$\leq$ 0.001. We also performed GO enrichment using `Gowinda` [45] and found 16 enriched GO terms including XXXX, XXXX S4 Table. Also, we tested if variants showing signal of overdominance, we computed $D$ statistic on simulated and experimental data, and computed $p$-values accordingly. After correcting for multiple testing, 96 variants discovered with FDR$\leq$ 0.01 Fig 7. However, it has been shown that [32, 79] large number of candidate loci is due to low frequency haplotype blocks as well as large inversions and most of candidate loci are rather hitchhikers to the favored allele.

We also applied composite statistic to the *D. melanogaster* data by computing $\mathcal{H}^*$ with sliding windows of size of 500 SNPs and step size of 100 variants over the genome. We computed null distribution of $\mathcal{H}^*$ by creating 100 chromosome simulations using experimental data parameters and length of 20Mbp. After correcting for multiple testing, only 16 intervals Fig 8 satisfy FDR$\leq$ 0.05, which form 5 contiguous interval Table 1 covering 3740 polymorphic sites. We then selected the 216 significant variants with FDR $\leq$ 0.05 within selected regions. These variants fall within 33 genes Table 2 and enrich 4 GO terms Table 3.

Here is UCSC track for this data.

### 3.2 Analysis of Outcrossing Yeast Populations

We also applied CLEAR to outcrossing Yeast populations [18], with 12 replicates where samples are taken at generations $\mathcal{T} = \{0, 180, 360, 540\}$. While this experiment is being conducted with larger set of replicates, population size, and number of generations, it appears that a number of replicates undergoing severe demographic events S10 Fig. Hence we chose seven replicates $r \in \{3, 7, 8, 9, 10, 11, 12\}$ that exhibit consistent genome-wide site-frequency spectrum over the whole experiment S11 Fig.

We estimated population size to be $\widehat{N} = 2000$ haplotypes, and computed $\widehat{s}$, $\widehat{h}$ and $H$ statistic accordingly. To compute $p$-values, we created 1M single-locus neutral simulations according to experimental data's initial frequency and coverage. By setting FDR cutoff to 0.05, only 18 and 16 variants show significant signal for directional and overdominant selection, respectively, see Fig 7.

Here is UCSC track for this data.

## 4 Discussion

We developed a computational tool, CLEAR, that can detect regions under selection experimental evolution experiments of sexual populations. Using extensive simulations, we show that CLEAR outperforms existing methods in detecting selection, locating the favored allele, and estimating selection parameters. Importantly, we make design choices that make CLEAR very fast in practice, facilitating genome-wide studies.

| index | FBgn | CHROM | start | end | name |
|---|---|---|---|---|---|
| 1 | FBgn0051810 | 2L | 16849018 | 16852368 | CG31810 |
| 2 | FBgn0051809 | 2L | 16853620 | 16858714 | CG31809 |
| 3 | FBgn0264440 | 2L | 16861497 | 16862431 | CR43858 |
| 4 | FBgn0051781 | 2L | 16869346 | 16878300 | CR31781 |
| 5 | FBgn0052832 | 2L | 16878326 | 16879290 | CG32832 |
| 6 | FBgn0032618 | 2L | 16879517 | 16886319 | CG31743 |
| 7 | FBgn0085342 | 2L | 16879517 | 16886319 | CG34313 |
| 8 | FBgn0040985 | 2L | 16887109 | 16887966 | CG6115 |
| 9 | FBgn0261671 | 2L | 16888490 | 16917052 | tweek |
| 10 | FBgn0026150 | 2L | 16908229 | 16910418 | ApepP |
| 11 | FBgn0262355 | 2L | 16944723 | 16945374 | CR43053 |
| 12 | FBgn0053179 | 2L | 16973091 | 16993984 | beat-IIIb |
| 13 | FBgn0040674 | 2R | 2725579 | 2726560 | CG9445 |
| 14 | FBgn0033109 | 2R | 2749506 | 2760223 | coro |
| 15 | FBgn0033110 | 2R | 2760501 | 2763324 | CG9447 |
| 16 | FBgn0033113 | 2R | 2768500 | 2770912 | Spn42Dc |
| 17 | FBgn0028988 | 2R | 2770785 | 2772378 | Spn42Dd |
| 18 | FBgn0033115 | 2R | 2773057 | 2775767 | Spn42De |
| 19 | FBgn0050158 | 2R | 2779265 | 2810118 | CG30158 |
| 20 | FBgn0001085 | 3L | 14267443 | 14361739 | fz |
| 21 | FBgn0036421 | 3L | 14362025 | 14362807 | CG13481 |
| 22 | FBgn0262580 | 3L | 14375013 | 14376399 | CG43120 |
| 23 | FBgn0036422 | 3L | 14393869 | 14395825 | CG3868 |
| 24 | FBgn0003459 | 3L | 14396030 | 14403039 | stwl |
| 25 | FBgn0087007 | 3L | 14405928 | 14529376 | bbg |
| 26 | FBgn0023531 | X | 1567143 | 1586801 | CG32809 |
| 27 | FBgn0023130 | X | 1587648 | 1589922 | a6 |
| 28 | FBgn0025378 | X | 1602839 | 1604215 | CG3795 |
| 29 | FBgn0025391 | X | 1629978 | 1648098 | Scgdelta |
| 30 | FBgn0026086 | X | 1667758 | 1682098 | Adar |
| 31 | FBgn0029939 | X | 7031756 | 7135965 | CG9650 |
| 32 | FBgn0029941 | X | 7175122 | 7299830 | CG1677 |
| 33 | FBgn0029944 | X | 7218247 | 7222839 | Dok |

Table 2: aa

| GO ID | -log($p$-value) | Hits | Num of Genes | Total Genes | GO Term |
|---|---|---|---|---|---|
| GO:0019233 | 1.6237e-06 | 4 | 435 | 552 | sensory perception of pain |
| GO:0005615 | 1.6237e-06 | 5 | 306 | 405 | extracellular space |
| GO:0004867 | 1.6237e-06 | 4 | 48 | 83 | serine-type endopeptidase inhibitor activity |
| GO:0045861 | 1.6237e-06 | 4 | 16 | 21 | negative regulation of proteolysis |

Table 3: aa

Many factors play a role in adaptation during experimental evolution studies. The statistics used by CLEAR perform well because they account for many of these aspects. CLEAR is not restricted to two-time points, but uses the complete time-series data. Because it uses an exact model, CLEAR achieves robust predictions for all values of the initial frequency. It adjusts for heterogeneous ascertainment bias in finite-depth pooled-seq data to avoid hard filtering variants. It exploits presence of high linkage within a region to compute composite likelihood ratio statistic. Finally, CLEAR uses $s, h$ as model parameters in its likelihood calculation, and provides optimized estimates of these parameters, which can provide extra information such as fixation time, and dominance (**??**).

In our simulations, we found that the power of detection can be severely affected by the sampling schedule as well as initial frequency of the favored allele. In general, while E&R studies are powerful, they also pose some challenges that are not adequately considered by other tools. One serious constraint is the sampling time span, the gap between the first and last sampled generations, which depends upon the generation time of the organism. It can be very small relative to the time of fixation of the favored allele. In *D. melanogaster* for example, 30-50 generations are typical [46], although there are some notable exceptions [91]. Therefore, unless the selection coefficient is very strong, the time series data will only capture a 'partial sweep'. This limitation is more pronounced in controlled experimental evolution, where the sampling often starts at the onset of selection. In particular, in a hard sweep scenario, the initial frequency of the favored allele is low, and may not reach detectable frequency in sequencing, given the sampling time span. Through exact (discrete-time, discrete-frequency) modeling, CLEAR performs better than competing tools even when initial frequency is low and sampling time span is limited.

However, even if it were possible to sample over a larger time-span, many methods, especially the ones that compute full likelihoods, would simply not scale to allow computation of evolutionary trajectories over a large time-span. In contrast, CLEAR precomputes the transition matrices, and scales linearly with number of samples, irrespective of the time-span in which they were acquired.

Sequence coverage is a practical consideration that is often ignored by other tools. Low sequencing coverage can lead to incorrect frequency estimates, even for the favored allele, especially when the initial frequency is low. CLEAR uses HMMs to explicitly model variation in sequence coverage. Moreover, it computes the composite likelihood from multiple linked sites, reducing the impact of coverage on any one site, and detects selection even when the favored site is not sampled due to low sequencing depth.

In controlled experimental evolution experiments, populations are evolved and inbred. As this scenario involves picking a small number of founders, the effective population size significantly drops from the large number of wild type (e.g., for *D. melanogaster* , $N_e \approx 10^6$) to a small number of founder lines ($F \approx 10^2$). This creates a severe population bottleneck. The bottleneck confounds SFS-based statistics and makes it difficult to fit a model or test a hypothesis (**??**). Hence, statistical testing based on SFS statistic provides poor performance in controlled experiments where the initial sampling time is close to the onset of selection. However, SFS-based methods perform very well when sampling is started long after the onset of selection (e.g., sampling from natural populations). The larger time gap from the onset of selection provides an opportunity for the site frequency spectrum to shift away from neutrality.

The comparison of hard and soft sweep scenarios lead to interesting observations. First, when LD is high in the selected region, as is often the case in a hard sweep, composition of scores significantly improves power of detection. When LD is low, as in soft sweep scenarios, composition of scores does not work as well. However, the favored allele is well established at the onset of selection, and will grow faster compared to the hard sweep scenario under identical selection regimes. This makes it possible to detect selection even in soft sweep scenarios. The situation is a little different

with respect to localizing the favored allele. In soft sweep scenarios, the favored allele is not in high LD with nearby variants, and its frequency change is independent of them. Therefore, we obtain better localization results in soft sweep scenarios.

There are many directions to improve the analyses presented here. In particular, we plan to focus our attention on other organisms with more complex life cycles, experiments with variable population size and longer sampling-time-spans. As evolve and resequencing experiments continue to grow, deeper insights into adaptation will go hand in hand with improved computational analysis.

**Software and Data Availability.** The source code and running scripts for CLEAR are publicly available at `https://github.com/bafnalab/clear`.

*D. melanogaster* data originally published [32, 59]. The dataset of the *D. melanogaster* study, until generation 37, is obtained from Dryad digital repository (`http://datadryad.org`)under accession DOI: 10.5061/dryad.60k68. Generation 59 of the *D. melanogaster* study is accessed from European Sequence Read Archive (http://www.ebi.ac.uk/ena/) under the project accession number: PRJEB6340. The dataset containing experimental evolution of Yeast populations [18] is downloaded from `http://wfitch.bio.uci.edu/~tdlong/PapersRawData/BurkeYeast.gz` (last accessed 01/24/2017). UCSC browser tracks for *D. melanogaster* and Yeast data analysis are found in Suppl. Data 1 and 2, respectively.

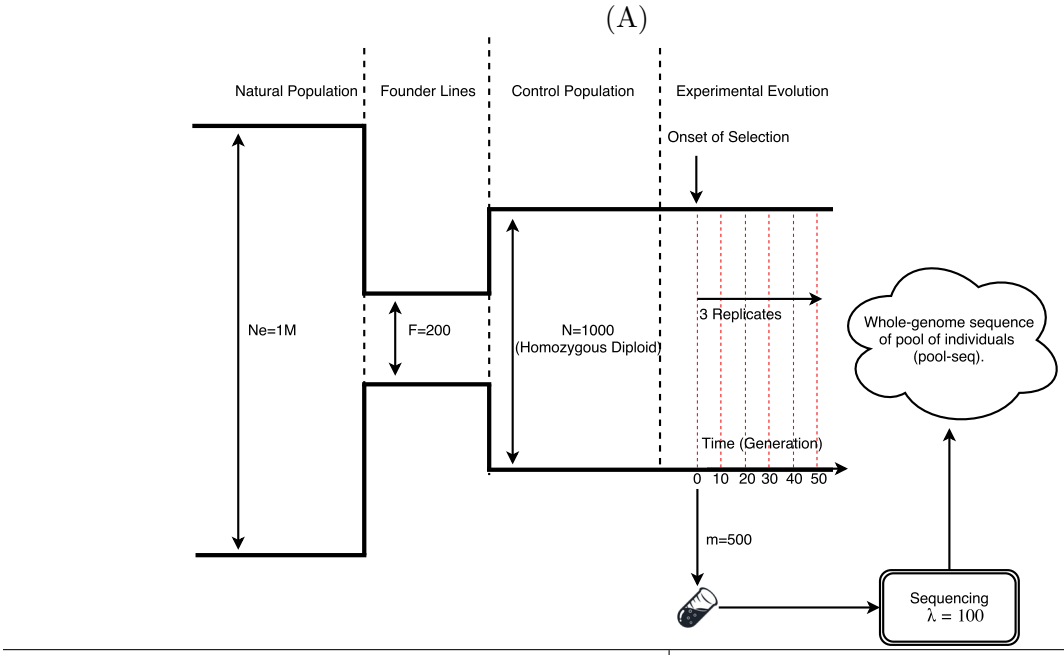
## Acknowledgments

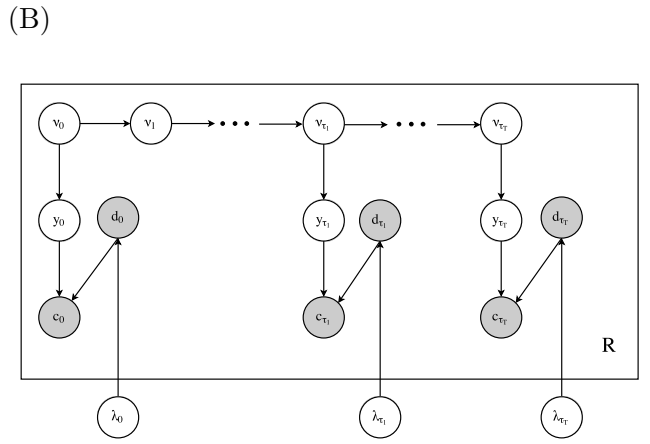
AI, AA, and VB were supported by grants from the NIH (1R01GM114362) and NSF (DBI-1458557 and IIS-1318386). CS is supported by the European Research Council grant ArchAdapt.

## Conflict of interest

VB is a co-founder, has an equity interest, and receives income from Digital Proteomics, LLC (DP). The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. DP was not involved in the research presented here.

# Figures



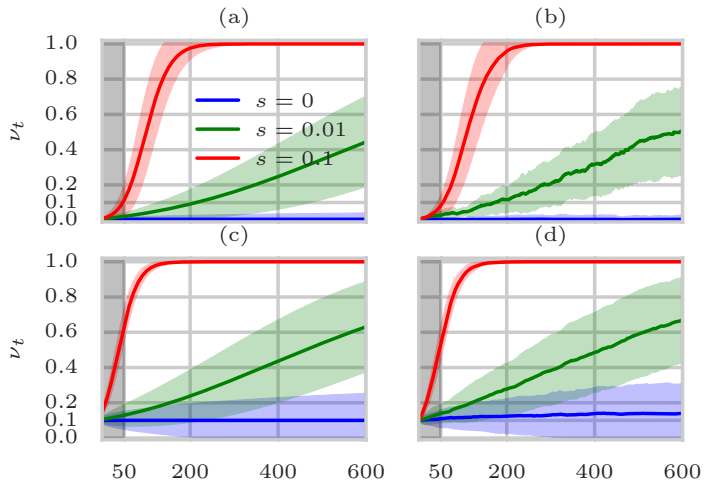




Fig 1: **Evolve and Resequence Selection Experiments on *D. melanogaster* .** (A) Typical configuration in which dynamic data is collected for *D. melanogaster* . (B) Graphical model showing dependence of the random variables in the single-locus model used to compute CLEAR statistics. Only shaded variables are observed. (C) Mean and the 95% confidence interval of the theoretical (a,b) and empirical (b,d) trajectories of the favored allele for hard (a,b) and soft (c,d) sweep scenarios and $N = 1000$. The first 50 generations are shaded in gray to represent the sampling span of sampling in short-term experiments, implying the difficulty in predicting selection at early stages of selective sweep.

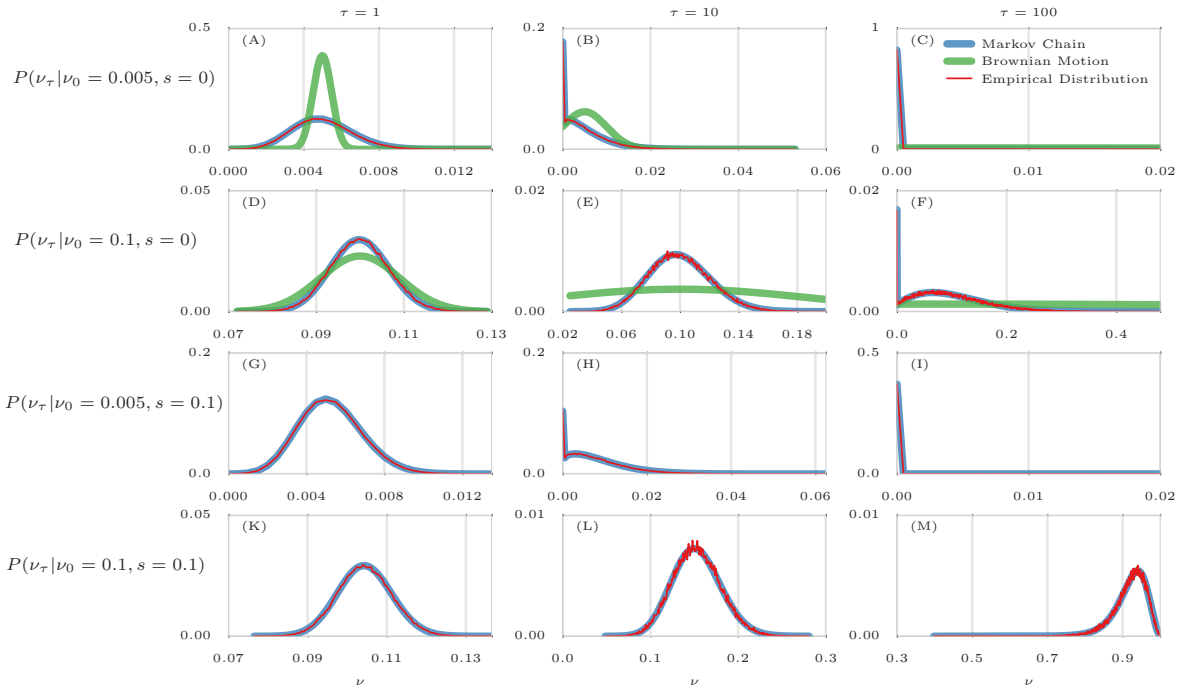



Fig 2: **Comparison of empirical distributions of allele frequencies (red) versus predictions from Brownian Motion (green), and Markov chain (blue).**
Comparison of empirical and theoretical distributions under neutral evolution (panels A-F) and selection (panels G-M) with different starting frequencies $\nu_0 \in \{0.005, 0.1\}$ and sampling times of $\mathcal{T} = \{0, \tau\}$, where $\tau \in \{1, 10, 100\}$. For each panel, the empirical distribution was computed over 100,000 simulations. Brownian motion (Gaussian approximation) provides poor approximations when initial frequency is far from 0.5 (A) or sampling is sparse (B,C,E,F). In addition, Brownian motion can only provide approximations under neutral evolution. In contrast, Markov chain consistently provide a good approximation in all cases.

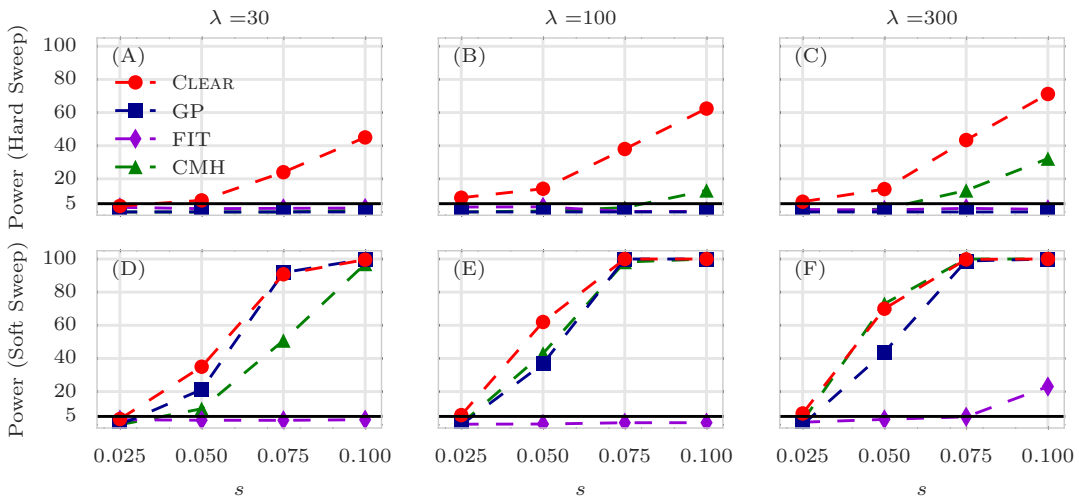



Fig 3: **Power calculations for detection of selection.**
Detection power for CLEAR($\mathcal{H}$), Frequency Increment Test (FIT), Gaussian Process (GP), and CMH under hard (A-C) and soft sweep (D-F) scenarios. $\lambda$, $s$ denote the mean coverage and selection coefficient, respectively. The $y$-axis measures power – sensitivity with false positive rate FPR $\leq 0.05$ – for $2,000$ simulations of 50Kbp regions. The horizontal line reflects the power of a random classifier. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

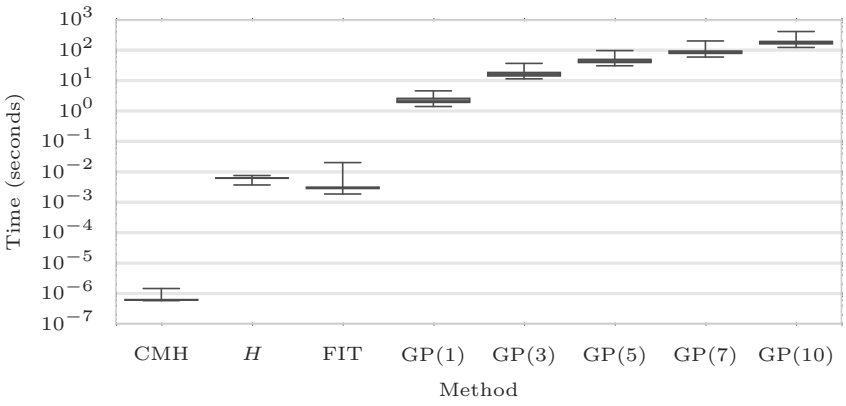



Fig 4: **Running time.**
Box plots of running time per variant (CPU-secs.) of CLEAR($\mathcal{H}$), CMH, FIT, and GP with single, 3, 5, 7, and 10 loci over 1000 simulations conducted on a workstation with Intel Core i7 processor. The average running time for each method is shown on the x-axis. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

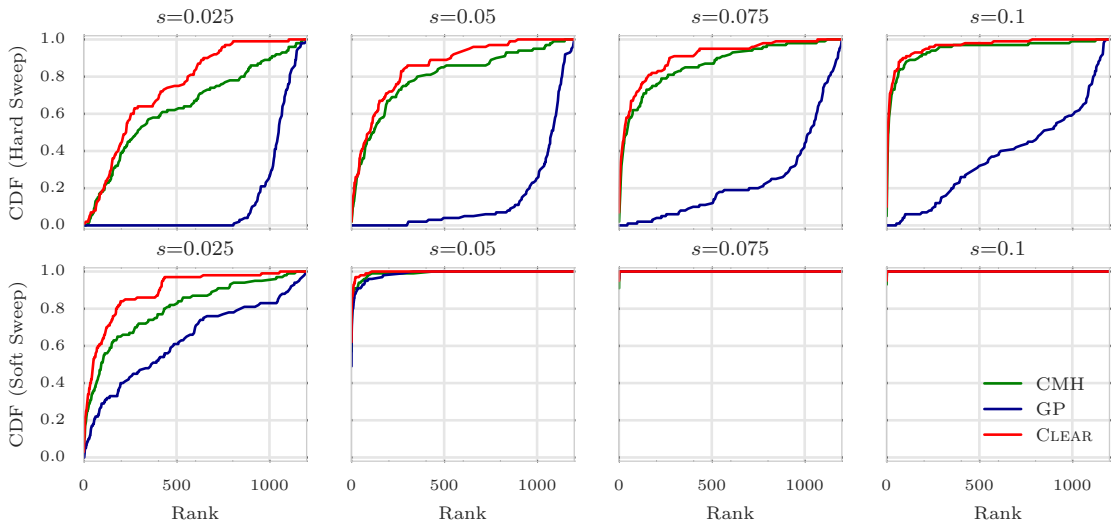



Fig 5: **Ranking performance for 100× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR ($H$), Gaussian Process (GP), CMH, and Frequency Increment Test (FIT), for different values of selection coefficient $s$ and initial carrier frequency. Note that the individual variant CLEAR score ($H$) is used to rank variants. The Area Under Curve (AUC) is computed as an overall quantitative measure to compare the performance of methods for each configuration. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

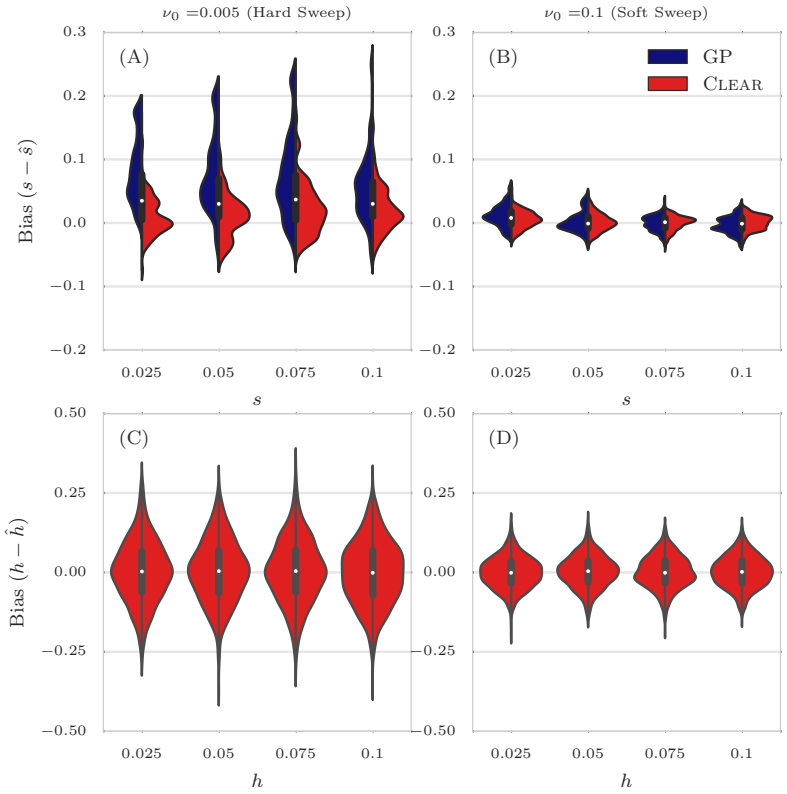



Fig 6: **Distribution of bias for 100× coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR $(H)$ is shown for a range of choices for the selection coefficient $s$ and starting carrier frequency $\nu_0$, when coverage $\lambda = 100$ (Panels A,B). GP and CLEAR have similar variance in estimates of $s$ for soft sweep, while CLEAR provides lower variance in hard sweep. Also see S3 Table. Panels C,D show the variance in the estimation of $h$. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

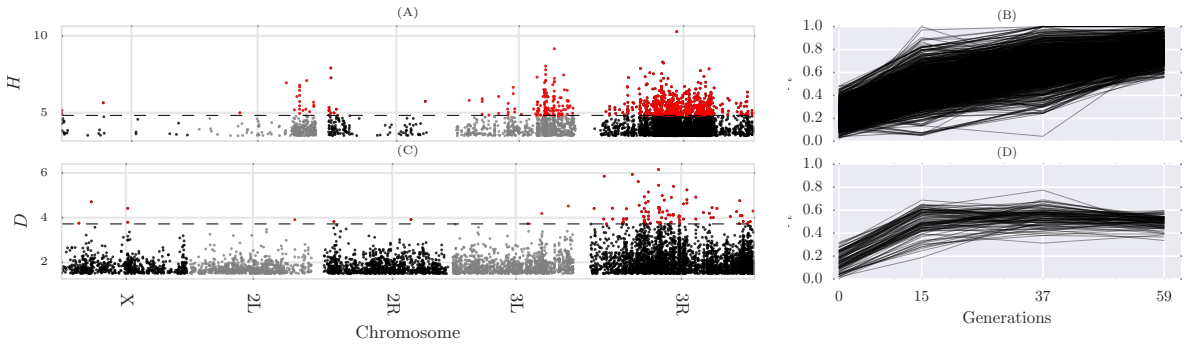



Fig 7: **Single locus analysis of the data from a study of *D. melanogaster* adaptation to alternating temperatures.**
Manhattan plot of scan for testing directional selection (A) and overdominant selection (C). Significant variants with FDR$\leq$ 0.001 and FDR $\leq$ 0.01 are denoted in red, and their trajectories are depicted in panels (B) and (D).

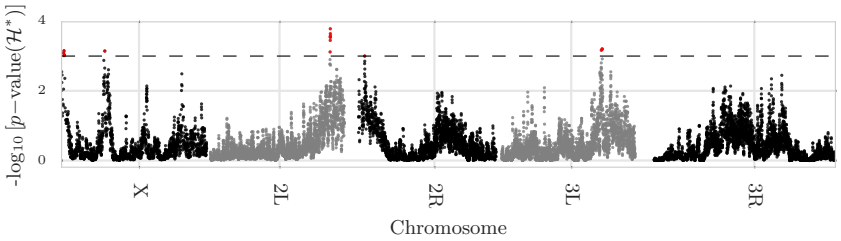



Fig 8: **Scan of composite statistic on data from a study of *D. melanogaster* adaptation to alternating temperatures.** Manhattan plot of scan for $\mathcal{H}^*$ statistic over the genome. The dashed line represents cutoff for genome-wide FDR$\leq$ 0.05, selecting 16 regions.

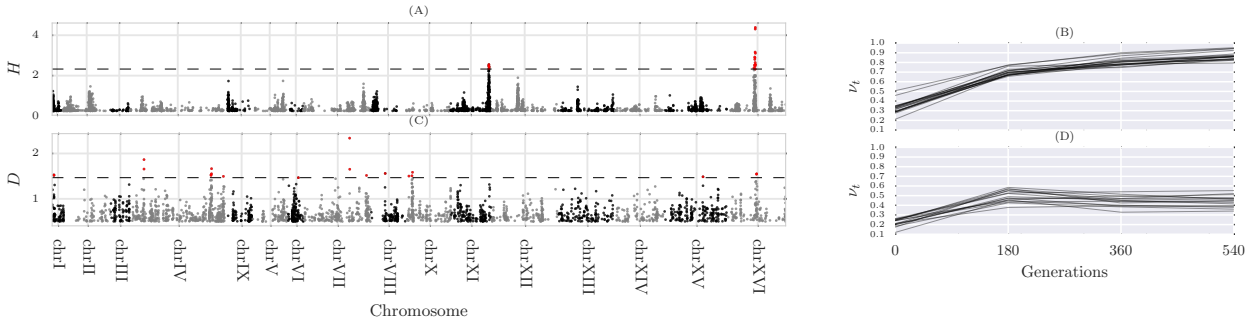



Fig 9: **Single locus analysis of the data from a study of *D. melanogaster* adaptation to alternating temperatures.**
Manhattan plot of scan for testing directional selection (A) and overdominant selection (C). dashed The dashed line represents cutoff for genome-wide FDR$\leq$ 0.05. Trajectories of the selected variants are depicted in panels (B) and (D).

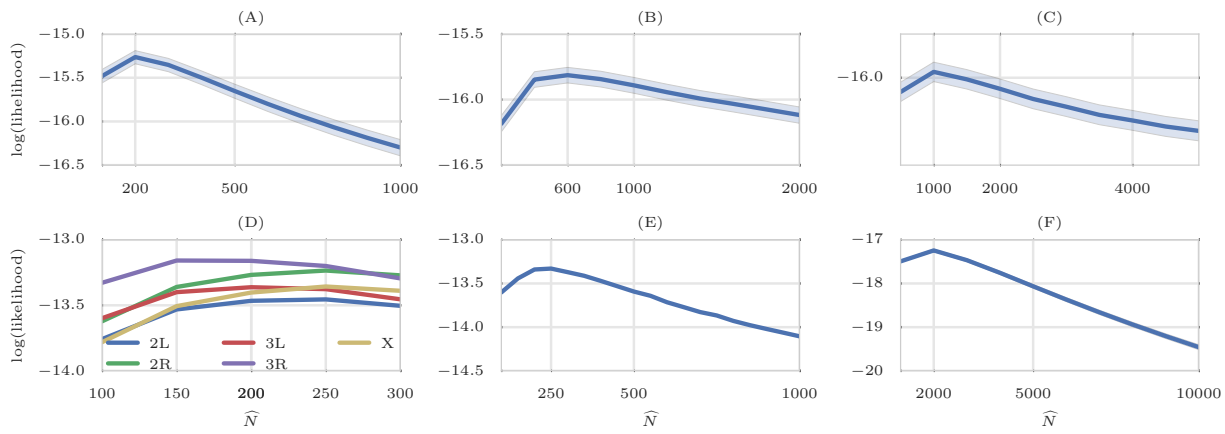



Fig 10: **Estimating N on simulated and real data.** Average and 95% confidence interval if likelihood of simulated data with $N = 200$ (A), $N = 600$(B), and $N = 1000$ individuals, over 100 simulations, shows that estimator is unbiased. Chromosome-wise (D) and genome-wide (E) estimation of population size for data from a study of *D. melanogaster* adaptation to alternating temperatures. Chromosome 3R fits population size of 150, while genome-wide population size is 250. (F) Despite large census population size ($10^6 - 10^7$ [18]), Yeast dataset exhibits much smaller ($\widehat{N} = 2000$) populations size.

**Generative Process 1:** The Generative Process for Dynamic Pool-seq Data.

**Input:** $N, n, R, \{\lambda_{\tau_0}, \ldots, \lambda_{\tau_T}\}, \mathcal{T} = \{\tau_0, \ldots \tau_T\}$

**Output:** Time-series pool-seq data for $R$ replicates of a single locus $\{\mathbf{c}^{(r)}\}$ and $\{\mathbf{d}^{(r)}\}$.

**for** $r \leftarrow 1$ **to** $R$ **do**

  **for** $t \leftarrow \tau_0$ **to** $\tau_T$ **do**

    $2N\nu_t \sim \text{Binomial}(2N, \nu_{t-1})$;

    **if** $t \in \mathcal{T}$ **then**

      $d_t^{(r)} \sim \text{Poisson}(\lambda_{\tau_i})$ ;

      $2ny_t \sim \text{Binomial}(2n, \nu_t)$;

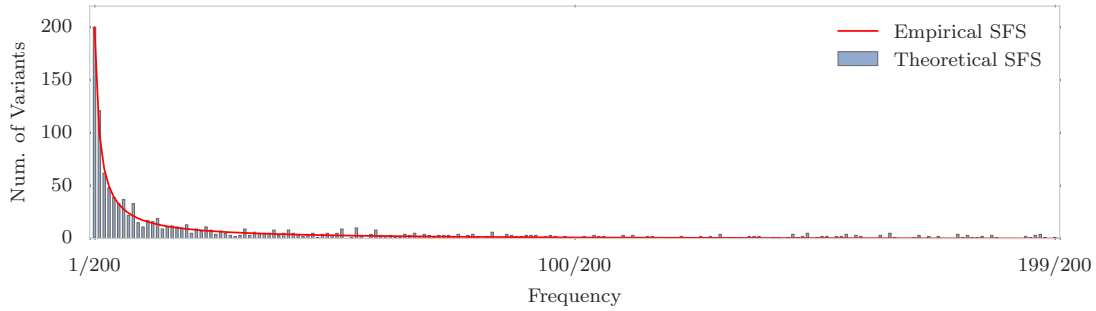
      $c_t^{(r)} \sim \text{Binomial}(d_t^{(r)}, y_t)$;

    **end**

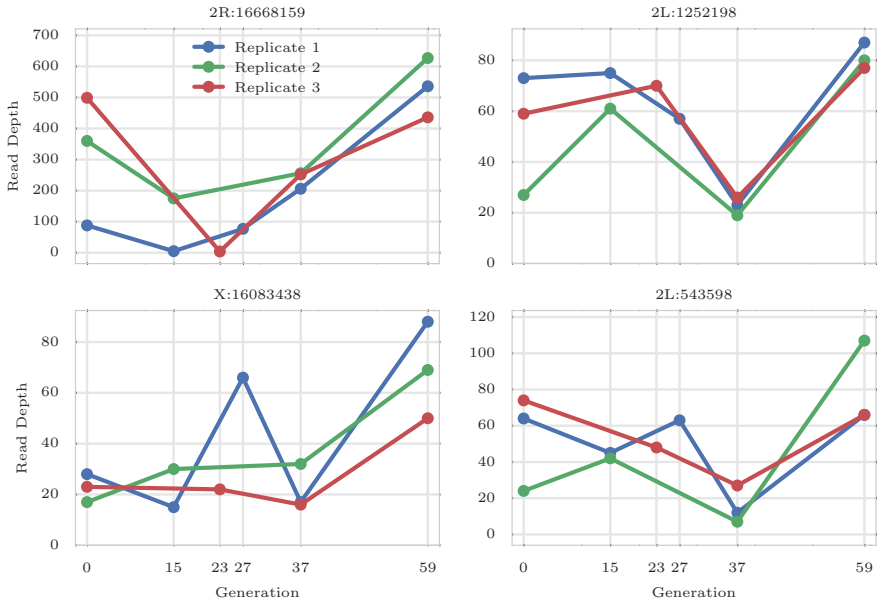
  **end**

**end**

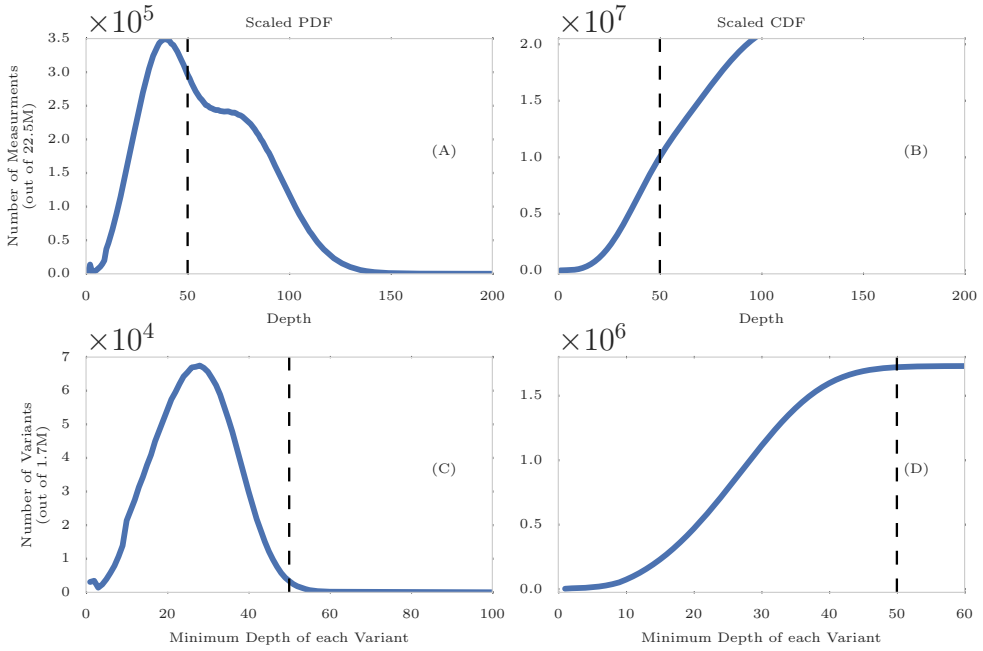
S1 Fig: aa



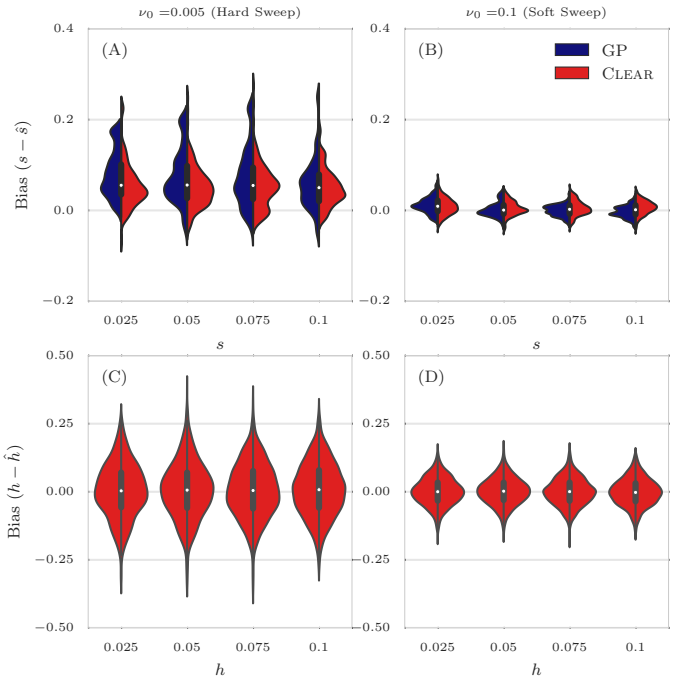

S2 Fig: **Site Frequency Spectrum.**
Theoretical and Empirical SFS in a 50Kbp region for a neutral population of 200 individuals when $N_e = 10^6$ and $\mu = 10^{-9}$. The $x$-axis corresponds to site frequency, and the $y$-axis to the number of variants with a specific frequency. In a neural population, majority of the variations stand in low frequency.



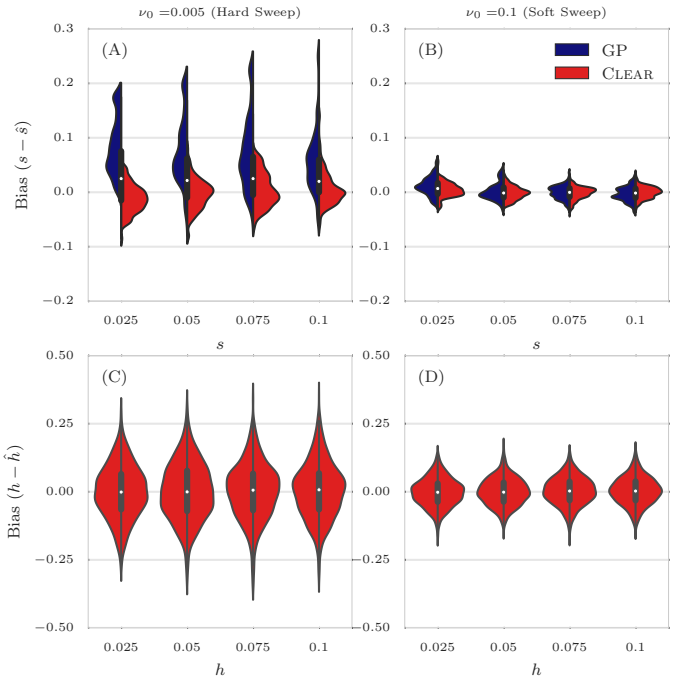

S3 Fig: **Coverage heterogeneity in time series data.**
Each panel shows the read depth for 3 replicates of the data from a study of *D. melanogaster* adaptation to alternating temperatures data (see section 3.1). Heterogeneity in depth of coverage is seen between replicates, and also at different time points, in all 4 variants. None of these sites pass the the hard filtering with minimum depth of 30.



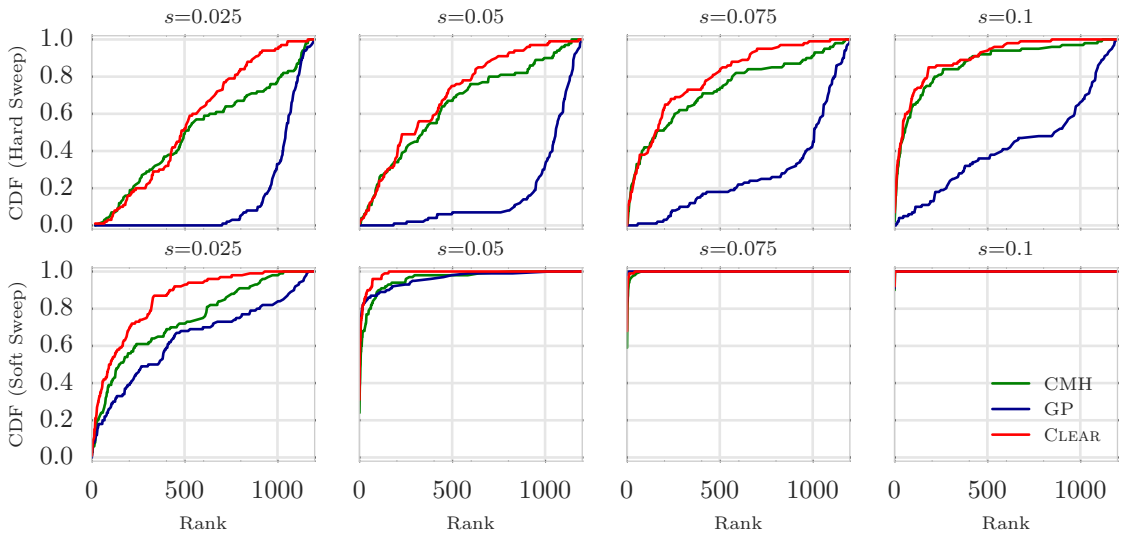

S4 Fig: **Distribution of depth in the real data.**
Scaled PDF (A) and CDF (B) of the read depths of all (≈22.1M) measurements, i.e., all replicates and time points of the all (≈1.7M) variants. Scaled PDF (C) and CDF (D) of the minimum depth of sites. While more than half most (≈12.5M) of the measurements have depth of 50 or greater (dashed line in (A),(B)), only a small fraction (≈11K) of variants (dashed line in (C),(D)) pass the filter of having minimum depth of 50.



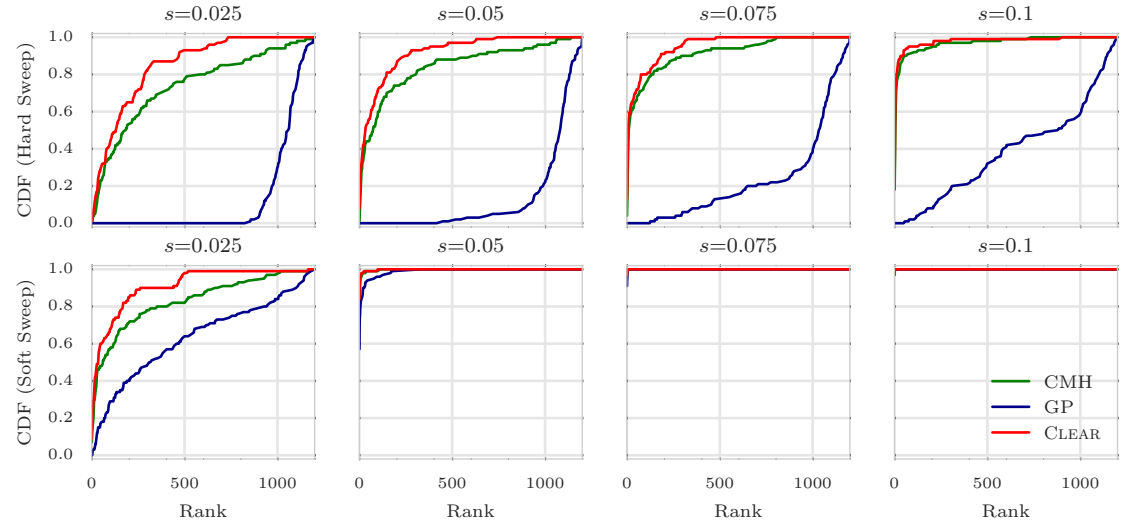

S5 Fig: **Distribution of bias for 30× coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR ($H$) is shown for a range of choices for the selection coefficient $s$ and starting carrier frequency $\nu_0$, when coverage $\lambda = 30$ (Panels A,B). GP and CLEAR have similar variance in estimates of $s$ for soft sweep, while CLEAR provides lower variance in hard sweep. Also see S3 Table. Panels C,D show the variance in the estimation of $h$.



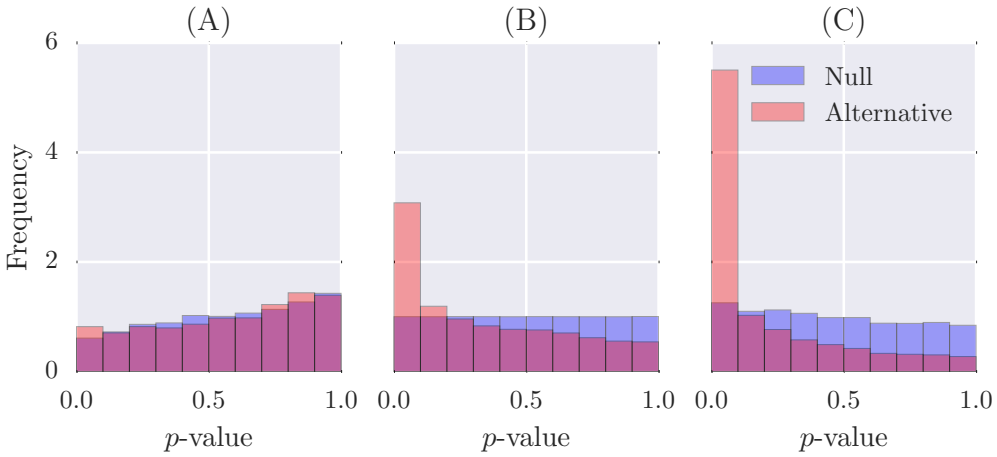

S6 Fig: **Distribution of bias for infinite coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR ($H$) is shown for a range of choices for the selection coefficient $s$ and starting carrier frequency $\nu_0$, when coverage $\lambda = \infty$ (Panels A,B). GP and CLEAR have similar variance in estimates of $s$ for soft sweep, while CLEAR provides lower variance in hard sweep. Also see S3 Table. Panels C,D show the variance in the estimation of $h$.



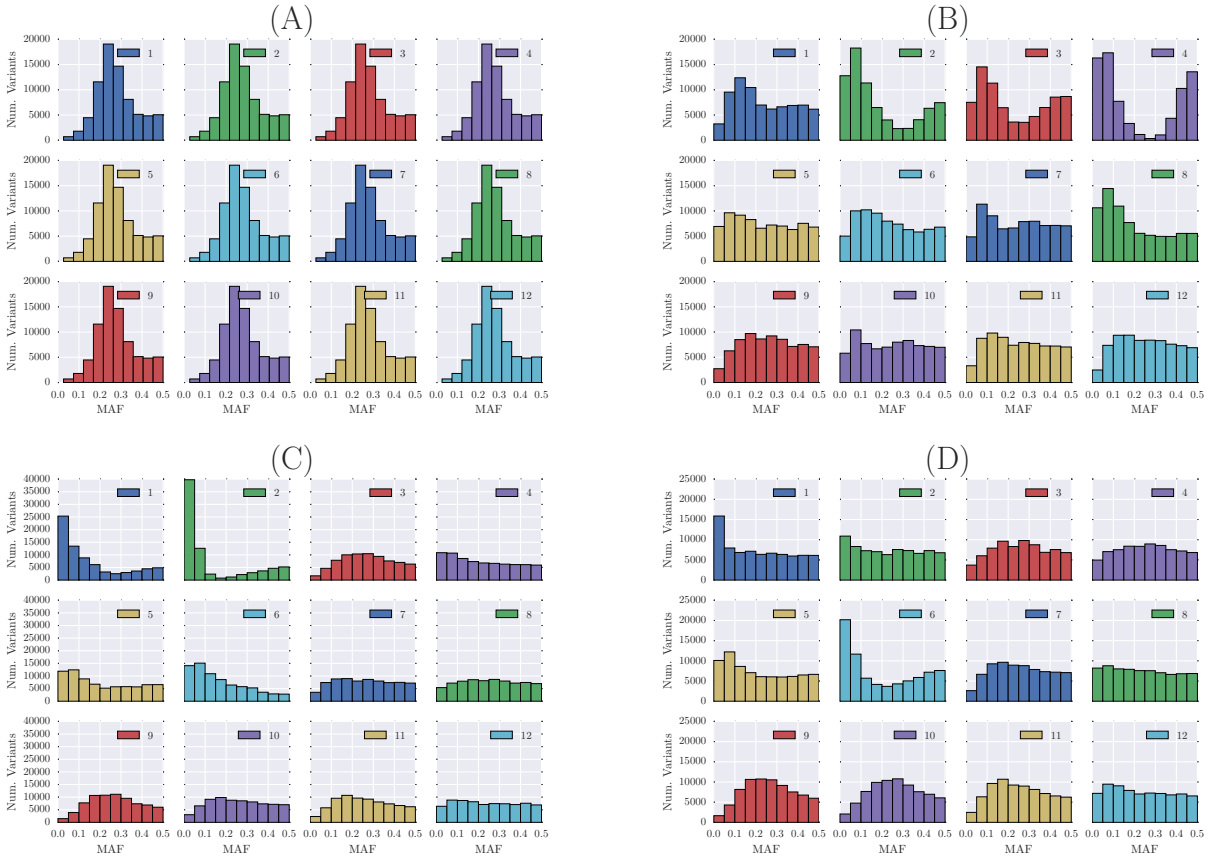

S7 Fig: **Ranking performance for 30× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR ($H$ score), Gaussian Process (GP), and Cochran Mantel Haenszel (CMH), for different values of selection coefficient $s$ and initial carrier frequency.



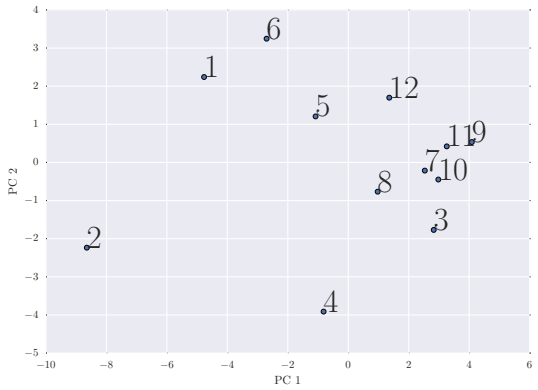

S8 Fig: **Ranking performance for 300× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR ($H$ score), Gaussian Process (GP), and Cochran Mantel Haenszel (CMH), for different values of selection coefficient $s$ and initial carrier frequency.




S9 Fig: **..**




S10 Fig: **Yeast SFS.**




S11 Fig: **Yeast SFS.**

S1 Table: **Average of power for detecting selection.**

Hard Sweep

| $\lambda$ | Method | Avg Power |
|---|---|---|
| 300 | CLEAR | 34 |
| 300 | CMH | 12 |
| 300 | FIT | 2 |
| 300 | GP | 0 |
| 100 | CLEAR | 31 |
| 100 | CMH | 4 |
| 100 | FIT | 2 |
| 100 | GP | 0 |
| 30 | CLEAR | 20 |
| 30 | FIT | 2 |
| 30 | CMH | 0 |
| 30 | GP | 0 |

Soft Sweep

| $\lambda$ | Method | Avg Power |
|---|---|---|
| 300 | CLEAR | 69 |
| 300 | CMH | 69 |
| 300 | GP | 61 |
| 300 | FIT | 8 |
| 100 | CLEAR | 67 |
| 100 | CMH | 60 |
| 100 | GP | 59 |
| 100 | FIT | 1 |
| 30 | CLEAR | 57 |
| 30 | GP | 53 |
| 30 | CMH | 39 |
| 30 | FIT | 3 |

Average power is computed for 8000 simulations with $s \in \{0.025, 0.05, 0.075, 0.1\}$. Frequency Increment Test (FIT), Gaussian Process (GP), CLEAR ($\mathcal{H}$ statistic) and Cochran Mantel Haenszel (CMH) are compared for different initial carrier frequency $\nu_0$. For all sequencing coverages, CLEAR outperform other methods. When coverage is not high ($\lambda \in \{30, 100\}$) and initial frequency is low (hard sweep), CLEAR significantly perform better than others.

S2 Table: **Average running time per variant in seconds for different methods.**

| Method | Avg. Time per Locus |
|---|---|
| CMH | 0.0 |
| $H$ | 0.003 |
| FIT | 0.006 |
| GP(1) | 2.551 |
| GP(3) | 19.177 |
| GP(5) | 50.291 |
| GP(7) | 95.602 |
| GP(10) | 202.017 |

S3 Table: **Mean and standard deviation of the distribution of bias $(s - \hat{s})$ of 8000 simulations with coverage $\lambda = 100\times$ and $s \in \{0.025, 0.05, 0.075, 0.1\}$.**

| Method | $\nu_0$ | Mean | STD |
|---|---|---|---|
| GP | 0.005 | 0.073 | 0.061 |
| CLEAR | 0.005 | 0.016 | 0.035 |
| GP | 0.1 | 0.002 | 0.016 |
| CLEAR | 0.1 | 0.002 | 0.013 |

S4 Table: **GO enrichment analysis of data from a study of *D. melanogaster* adaptation to alternating temperaturesusing `Gowinda`.**

| GO ID | -log($p$-value) | Hits | Num of Genes | Total Genes | GO Term |
|---|---|---|---|---|---|
| GO:0007501 | 9e-06 | 4 | 10 | 11 | mesodermal cell fate specification |
| GO:0010629 | 9e-06 | 4 | 43 | 56 | negative regulation of gene expression |
| GO:0048190 | 9e-06 | 6 | 33 | 45 | wing disc dorsal/ventral pattern formation |
| GO:0004222 | 2.92222e-05 | 10 | 66 | 81 | metalloendopeptidase activity |
| GO:0004252 | 4.40476e-05 | 23 | 243 | 283 | serine-type endopeptidase activity |
| GO:0005509 | 5.33318e-05 | 14 | 152 | 198 | calcium ion binding |
| GO:0006508 | 6.27391e-05 | 36 | 407 | 481 | proteolysis |
| GO:0004386 | 9.20577e-05 | 4 | 29 | 37 | helicase activity |
| GO:0008586 | 0.000172871 | 4 | 33 | 48 | imaginal disc-derived wing vein morphogenesis |
| GO:0007017 | 0.0002060152 | 5 | 37 | 48 | microtubule-based process |
| GO:0008513 | 0.0002800775 | 8 | 18 | 23 | secondary active organic cation transmembrane transporter activity |
| GO:0007517 | 0.0003799091 | 7 | 53 | 71 | muscle organ development |
| GO:0004046 | 0.0005088 | 4 | 6 | 7 | aminoacylase activity |
| GO:0007507 | 0.00056285 | 8 | 50 | 57 | heart development |
| GO:0006520 | 0.0007301093 | 5 | 15 | 18 | cellular amino acid metabolic process |
| GO:0030431 | 0.0007301093 | 8 | 94 | 119 | sleep |

S5 Table: **Distribution of candidate variants for data from a study of *D. melanogaster* adaptation to alternating temperatures.**

| Annotation | Count |
|---|---|
| intron variant | 356 |
| intergenic region | 254 |
| synonymous variant | 93 |
| 3 prime UTR variant | 29 |
| missense variant | 27 |
| splice region variant and intron variant | 12 |
| 5 prime UTR variant | 12 |
| non coding exon variant | 1 |
| Total | 784 |

## References


[1] Guillaume Achaz. Frequency spectrum neutrality tests: one for all and all for one. *Genetics*, 183(1):249–258, 2009.

[2] Alan Agresti and Maria Kateri. *Categorical data analysis*. Springer, 2011.

[3] Joshua M Akey. Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome research*, 19(5):711–722, 2009.

[4] Eric C Anderson, Ellen G Williamson, and Elizabeth A Thompson. Monte Carlo evaluation of the likelihood for Ne from temporally spaced samples. *Genetics*, 156(4):2109–2118, 2000.

[5] Frédéric Ariey, Benoit Witkowski, Chanaki Amaratunga, Johann Beghain, Anne-Claire Langlois, Nimol Khim, Saorin Kim, Valentine Duru, Christiane Bouchier, Laurence Ma, and Others. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature*, 505(7481):50–55, 2014.

[6] James G Baldwin-Brown, Anthony D Long, and Kevin R Thornton. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Molecular biology and evolution*, page msu048, 2014.

[7] Rowan D H Barrett, Sean M Rogers, and Dolph Schluter. Natural selection on a major armor gene in threespine stickleback. *Science*, 322(5899):255–257, 2008.

[8] Jeffrey E Barrick and Richard E Lenski. Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14(12):827–839, 2013.

[9] Jeffrey E Barrick, Dong Su Yu, Sung Ho Yoon, Haeyoung Jeong, Tae Kwang Oh, Dominique Schneider, Richard E Lenski, and Jihyun F Kim. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, 461(7268):1243–1247, 2009.

[10] Mark A Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003.

[11] Alan O Bergland, Emily L Behrman, Katherine R O'Brien, Paul S Schmidt, and Dmitri A Petrov. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genet*, 10(11):e1004775, 2014.

[12] Todd Bersaglieri, Pardis C Sabeti, Nick Patterson, Trisha Vanderploeg, Steve F Schaffner, Jared A Drake, Matthew Rhodes, David E Reich, and Joel N Hirschhorn. Genetic signatures of strong recent positive selection at the lactase gene. *The American Journal of Human Genetics*, 74(6):1111–1120, 2004.

[13] Pierre Berthier, Mark A Beaumont, Jean-Marie Cornuet, and Gordon Luikart. Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, 160(2):741–751, 2002.

[14] Jonathan P Bollback and John P Huelsenbeck. Clonal interference is alleviated by high mutation rates in large populations. *Molecular biology and evolution*, 24(6):1397–1406, 2007.

[15] Jonathan P Bollback, Thomas L York, and Rasmus Nielsen. Estimation of 2Nes from temporal allele frequency data. *Genetics*, 179(1):497–502, 2008.

[16] Adam R Boyko, Scott H Williamson, Amit R Indap, Jeremiah D Degenhardt, Ryan D Hernandez, Kirk E Lohmueller, Mark D Adams, Steffen Schmidt, John J Sninsky, Shamil R Sunyaev, and Others. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*, 4(5):e1000083, 2008.

[17] Molly K Burke, Joseph P Dunham, Parvin Shahrestani, Kevin R Thornton, Michael R Rose, and Anthony D Long. Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature*, 467(7315):587–590, 2010.

[18] Molly K Burke, Gianni Liti, and Anthony D Long. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of Saccharomyces cerevisiae. *Molecular biology and evolution*, page msu256, 2014.

[19] P Daborn, S Boundy, J Yen, B Pittendrigh, and Others. DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Molecular Genetics and Genomics*, 266(4):556–563, 2001.

[20] Rachel Daniels, Hsiao-Han Chang, Papa Diogoye Séne, Danny C Park, Daniel E Neafsey, Stephen F Schaffner, Elizabeth J Hamilton, Amanda K Lukens, Daria Van Tyne, Souleymane Mboup, and Others. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS One*, 8(4):e60780, 2013.

[21] Vincent J Denef and Jillian F Banfield. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science*, 336(6080):462–466, 2012.

[22] Michael M Desai and Joshua B Plotkin. The polymorphism frequency spectrum of finitely many sites under selection. *Genetics*, 180(4):2175–2191, 2008.

[23] Richard Durbin, Sean R Eddy, Anders Krogh, and Graeme Mitchison. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.

[24] Eyal Elyashiv, Shmuel Sattath, Tina T Hu, Alon Strutsovsky, Graham McVicker, Peter Andolfatto, Graham Coop, and Guy Sella. A Genomic Map of the Effects of Linked Selection in Drosophila. *PLoS Genet*, 12(8):1–24, 2016.

[25] Warren J Ewens. *Mathematical Population Genetics 1: Theoretical Introduction*, volume 27. Springer Science & Business Media, 2012.

[26] Gregory Ewing and Joachim Hermisson. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics*, 26(16):2064–2065, 2010.

[27] Shaohua Fan, Matthew E B Hansen, Yancy Lo, and Sarah A Tishkoff. Going global by adapting local: A review of recent human adaptation. *Science*, 354(6308):54–59, 2016.

[28] Justin C Fay and Chung-I Wu. Hitchhiking under positive Darwinian selection. *Genetics*, 155(3):1405–1413, 2000.

[29] Alison F Feder, Sergey Kryazhimskiy, and Joshua B Plotkin. Identifying signatures of selection in genetic time series. *Genetics*, 196(2):509–522, 2014.

[30] Alison F Feder, Soo-Yon Rhee, Susan P Holmes, Robert W Shafer, Dmitri A Petrov, and Pleuni S Pennings. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *eLife*, 5, jan 2016.

[31] Anna-Sophie Fiston-Lavier, Nadia D Singh, Mikhail Lipatov, and Dmitri A Petrov. Drosophila melanogaster recombination rate calculator. *Gene*, 463(1):18–20, 2010.

[32] Susanne U Franssen, Viola Nolte, Ray Tobler, and Christian Schlötterer. Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental Drosophila melanogaster populations. *Molecular biology and evolution*, 32(2):495–509, 2015.

[33] Nandita R Garud, Philipp W Messer, Erkan O Buzbas, and Dmitri A Petrov. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet*, 11(2):e1005004, 2015.

[34] John H Gillespie. *Population genetics: a concise guide*. JHU Press, 2010.

[35] Michael M Gottesman. Mechanisms of cancer drug resistance. *Annual review of medicine*, 53(1):615–627, 2002.

[36] Torsten Günther and Graham Coop. Robust identification of local adaptation from allele frequencies. *Genetics*, 195(1):205–220, 2013.

[37] Matthew Hegreness, Noam Shoresh, Daniel Hartl, and Roy Kishony. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science*, 311(5767):1615–1617, 2006.

[38] Kent E Holsinger and Bruce S Weir. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, 10(9):639–650, 2009.

[39] Christopher J R Illingworth and Ville Mustonen. Distinguishing driver and passenger mutations in an evolutionary history categorized by interference. *Genetics*, 189(3):989–1000, 2011.

[40] Christopher J R Illingworth, Leopold Parts, Stephan Schiffels, Gianni Liti, and Ville Mustonen. Quantifying selection acting on a complex trait using allele frequency time series data. *Molecular biology and evolution*, 29(4):1187–1197, 2012.

[41] Minako Izutsu, Atsushi Toyoda, Asao Fujiyama, Kiyokazu Agata, and Naoyuki Fuse. Dynamics of Dark-Fly Genome Under Environmental Selections. *G3: Genes— Genomes— Genetics*, pages g3—-115, 2015.

[42] Aashish R Jha, Cecelia M Miles, Nodia R Lippert, Christopher D Brown, Kevin P White, and Martin Kreitman. Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in Drosophila melanogaster. *Molecular biology and evolution*, 32(10):2616–2632, 2015.

[43] Ágnes Jónás, Thomas Taus, Carolin Kosiol, Christian Schlötterer, and Andreas Futschik. Estimating the Effective Population Size from Temporal Allele Frequency Changes in Experimental Evolution. *Genetics*, aug 2016.

[44] Tadeusz J Kawecki, Richard E Lenski, Dieter Ebert, Brian Hollis, Isabelle Olivieri, and Michael C Whitlock. Experimental evolution. *Trends in ecology & evolution*, 27(10):547–560, 2012.

[45] Robert Kofler and Christian Schlötterer. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, 28(15):2084–2085, 2012.

[46] Robert Kofler and Christian Schlötterer. A guide for the design of evolve and resequencing studies. *Molecular biology and evolution*, page mst221, 2013.

[47] Gregory I Lang, David Botstein, and Michael M Desai. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics*, 188(3):647–661, 2011.

[48] Gregory I Lang, Daniel P Rice, Mark J Hickman, Erica Sodergren, George M Weinstock, David Botstein, and Michael M Desai. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500(7464):571–574, 2013.

[49] Quan Long, Fernando A Rabanal, Dazhe Meng, Christian D Huber, Ashley Farlow, Alexander Platzer, Qingrun Zhang, Bjarni J Vilhjálmsson, Arthur Korte, Viktoria Nizhynska, and Others. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature genetics*, 45(8):884–890, 2013.

[50] Anna-Sapfo Malaspinas. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. *Molecular ecology*, 25(1):24–41, 2016.

[51] Anna-Sapfo Malaspinas, Orestis Malaspinas, Steven N Evans, and Montgomery Slatkin. Estimating allele age and selection coefficient from time-serial data. *Genetics*, 192(2):599–607, 2012.

[52] Frank Maldarelli, Mary Kearney, Sarah Palmer, Robert Stephens, JoAnn Mican, Michael A Polis, Richard T Davey, Joseph Kovacs, Wei Shao, Diane Rock-Kress, and Others. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *Journal of virology*, 87(18):10313–10323, 2013.

[53] Nelson E Martins, Vítor G Faria, Viola Nolte, Christian Schlötterer, Luis Teixeira, Élio Sucena, and Sara Magalhães. Host adaptation to viruses relies on few genes with different cross-resistance properties. *Proceedings of the National Academy of Sciences*, 111(16):5938–5943, 2014.

[54] Iain Mathieson and Gil McVean. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, 193(3):973–984, 2013.

[55] Philipp W Messer and Dmitri A Petrov. Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*, 28(11):659–669, 2013.

[56] Shalini Nair, Denae Nash, Daniel Sudimack, Anchalee Jaidee, Marion Barends, Anne-Catrin Uhlemann, Sanjeev Krishna, François Nosten, and Tim J C Anderson. Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Molecular Biology and Evolution*, 24(2):562–573, 2007.

[57] Rasmus Nielsen and James Signorovitch. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical population biology*, 63(3):245–255, 2003.

[58] Rasmus Nielsen, Scott Williamson, Yuseob Kim, Melissa J Hubisz, Andrew G Clark, and Carlos Bustamante. Genomic scans for selective sweeps using SNP data. *Genome research*, 15(11):1566–1575, 2005.

[59] Pablo Orozco-ter Wengel, Martin Kapun, Viola Nolte, Robert Kofler, Thomas Flatt, and Christian Schlötterer. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular ecology*, 21(20):4931–4941, 2012.

[60] Tugce Oz, Aysegul Guvenek, Sadik Yildiz, Enes Karaboga, Yusuf Talha Tamer, Nirva Mumcuyan, Vedat Burak Ozan, Gizem Hazal Senturk, Murat Cokol, Pamela Yeh, and Others. Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. *Molecular biology and evolution*, page msu191, 2014.

[61] Bo Peng and Marek Kimmel. simuPOP: a forward-time population genetics simulation environment. *Bioinformatics*, 21(18):3686–3687, 2005.

[62] Edward Pollak. A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104(3):531–548, 1983.

[63] Susan E Ptak and Molly Przeworski. Evidence for population growth in humans is confounded by fine-scale population structure. *Trends in Genetics*, 18(11):559–563, 2002.

[64] Sebastian E Ramos-Onsins and Julio Rozas. Statistical properties of new neutrality tests against population growth. *Molecular biology and evolution*, 19(12):2092–2100, 2002.

[65] Brian J Reid, Rumen Kostadinov, and Carlo C Maley. New strategies in Barrett's esophagus: integrating clonal evolutionary theory with clinical management. *Clinical Cancer Research*, 17(11):3512–3519, 2011.

[66] Silvia C Remolina, Peter L Chang, Jeff Leips, Sergey V Nuzhdin, and Kimberly A Hughes. Genomic basis of aging and life-history evolution in Drosophila melanogaster. *Evolution*, 66(11):3390–3403, 2012.

[67] Roy Ronen, Nitin Udpa, Eran Halperin, and Vineet Bafna. Learning natural selection from the site frequency spectrum. *Genetics*, 195(1):181–193, 2013.

[68] P C Sabeti, S F Schaffner, B Fry, J Lohmueller, P Varilly, O Shamovsky, A Palma, T S Mikkelsen, D Altshuler, and E S Lander. Positive natural selection in the human lineage. *science*, 312(5780):1614–1620, 2006.

[69] Stanley A Sawyer and Daniel L Hartl. Population genetics of polymorphism and divergence. *Genetics*, 132(4):1161–1176, 1992.

[70] Christian Schlötterer, R Kofler, E Versace, R Tobler, and S U Franssen. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, 114(5):431–440, 2015.

[71] Joshua G Schraiber, Steven N Evans, and Montgomery Slatkin. Bayesian inference of natural selection from allele frequency time series. *Genetics*, 203(1):493–511, 2016.

[72] Tatum S Simonson, Yingzhong Yang, Chad D Huff, Haixia Yun, Ga Qin, David J Witherspoon, Zhenzhong Bai, Felipe R Lorenzo, Jinchuan Xing, Lynn B Jorde, and Others. Genetic evidence for high-altitude adaptation in Tibet. *Science*, 329(5987):72–75, 2010.

[73] Brad Spellberg, Robert Guidos, David Gilbert, John Bradley, Helen W Boucher, W Michael Scheld, John G Bartlett, John Edwards, Infectious Diseases Society of America, and Others. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clinical Infectious Diseases*, 46(2):155–164, 2008.

[74] Matthias Steinrücken, Anand Bhaskar, and Yun S Song. A novel spectral method for inferring general diploid selection from time series genetic data. *The annals of applied statistics*, 8(4):2203, 2014.

[75] Wolfgang Stephan, Yun S Song, and Charles H Langley. The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics*, 172(4):2647–2663, 2006.

[76] John D Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.

[77] Fumio Tajima. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595, 1989.

[78] Jonathan Terhorst, Christian Schlötterer, and Yun S Song. Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution. *PLoS Genet*, 11(4):e1005069, 2015.

[79] Ray Tobler, Susanne U Franssen, Robert Kofler, Pablo Orozco-terWengel, Viola Nolte, Joachim Hermisson, and Christian Schlötterer. Massive habitat-specific genomic response in D. melanogaster populations during experimental evolution in hot and cold environments. *Molecular biology and evolution*, 31(2):364–375, 2014.

[80] Hande Topa, Ágnes Jónás, Robert Kofler, Carolin Kosiol, and Antti Honkela. Gaussian process test for high-throughput sequencing time series: application to experimental evolution. *Bioinformatics*, page btv014, 2015.

[81] Thomas L Turner, Andrew D Stewart, Andrew T Fields, William R Rice, and Aaron M Tarone. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. *PLoS Genet*, 7(3):e1001336, 2011.

[82] Joseph J Vitti, Sharon R Grossman, and Pardis C Sabeti. Detecting natural selection in genomic data. *Annual review of genetics*, 47:97–120, 2013.

[83] Jinliang Wang. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetical research*, 78(03):243–257, 2001.

[84] Robin S Waples. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 121(2):379–391, 1989.

[85] David Williams and David Williams. *Weighing the odds: a course in probability and statistics*, volume 548. Springer, 2001.

[86] Ellen G Williamson and Montgomery Slatkin. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, 152(2):755–761, 1999.

[87] Scott H Williamson, Melissa J Hubisz, Andrew G Clark, Bret A Payseur, Carlos D Bustamante, and Rasmus Nielsen. Localizing recent adaptive evolution in the human genome. *PLoS Genet*, 3(6):e90, 2007.

[88] Mark A Winters, Robert M Lloyd Jr, Robert W Shafer, Michael J Kozal, Michael D Miller, and Mark Holodniy. Development of elvitegravir resistance and linkage of integrase inhibitor mutations with protease and reverse transcriptase resistance mutations. *PloS one*, 7(7):e40514, 2012.

[89] Xin Yi, Yu Liang, Emilia Huerta-Sanchez, Xin Jin, Zha Xi Ping Cuo, John E Pool, Xun Xu, Hui Jiang, Nicolas Vinckenbosch, Thorfinn Sand Korneliussen, and Others. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329(5987):75–78, 2010.

[90] Hiba Zahreddine and K L Borden. Mechanisms and insights into drug resistance in cancer. *Front Pharmacol*, 4(28.10):3389, 2013.

[91] Dan Zhou, Nitin Udpa, Merril Gersten, DeeAnn W Visk, Ali Bashir, Jin Xue, Kelly A Frazer, James W Posakony, Shankar Subramaniam, Vineet Bafna, and Gabriel G. Haddad. Experimental selection of hypoxia-tolerant Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 108(6):2349–2354, 2011.