CLEAR: Composition of Likelihoods for Evolve And Resequence Experiments

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

[1]Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA
[2]Institut für Populationsgenetik, Vetmeduni, Vienna, Austria
[3]Department of 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 eukaryotes with large genomes. With this development, experimental evolution studies can be designed to observe molecular evolution "in-action" via Evolve-and-Resequence (E&R) experiments. Among other applications, E&R studies can be used to locate the genes and variants responsible for genetic adaptation. Most of existing literature on time-series data analysis often assumes large population size, accurate allele frequency estimates, or wide time spans. These assumptions do 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 signatures of selection in small population E&R experiments. CLEAR takes whole-genome sequence of pool of individuals (pool-seq) as input, and properly addresses heterogeneous ascertainment bias resulting from uneven coverage. CLEAR also provides unbiased estimates of model parameters, including population size, selection strength and dominance, while being computationally efficient. Extensive simulations show that CLEAR achieves higher power in detecting and localizing selection over a wide range of parameters, and is robust to variation of coverage. We applied CLEAR statistic to multiple E&R experiments, including, data from a study of *D. melanogaster* adaptation to alternating temperatures and a study of outcrossing yeast populations, and identified multiple regions under selection with genome-wide significance.

1 Introduction

Natural selection is a key force in evolution, and a mechanism by which populations can adapt to external 'selection' pressure. Examples of adaptation abound in the natural world [22], including for example, classic examples like lactose tolerance in Northern Europeans [9], human adaptation to high altitudes [56, 70], but also drug resistance in pests [15], HIV [24], cancer [27, 71], malarial parasite [3, 45], and others [57]. In these examples, understanding the genetic basis of adaptation can provide valuable information, underscoring the importance of the problem.

Experimental evolution refers to the study of the evolutionary processes of a model organism in a controlled [7, 10, 28, 37, 38, 47, 48] or natural [5, 8, 16, 17, 42, 51, 69] 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 [34] and have been shown to be more predictive than static data analysis [12, 18, 53]. In particular, longitudinal E&R data is being used to estimate model parameters including population size [33, 50, 61, 65, 66, 68], strength of selection [11, 29, 30, 41, 44, 58, 61], allele age [41] recombination rate [61], mutation rate [6, 61], quantitative trait loci [4] and for tests of neutrality hypotheses [8, 13, 23, 61].

While many E&R study designs are being used [6, 54], we restrict our attention to the adaptive evolution due to standing variation in fixed size populations. This regime has been considered earlier, typically with *D. melanogaster* as the model organism of choice, to identify adaptive genes in longevity and aging [13, 52] (600 generations), courtship song [64] (100 generations), hypoxia tolerance [72] (200 generations), adaptation to new laboratory environments [26, 47] (59 generations), egg size [32] (40 generations), C virus resistance [43] (20 generations), and dark-fly [31] (49 generations).

The task of identifying selection signatures 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 dominance at the site, can provide a comprehensive description of the selection process.

In the effort to analyze E&R selection experiments, many authors chose to adapt existing tests that were originally used for static data, pairwise comparisons (two time-points) and single replicates to perform a null scan. For instance, Zhu *et al.* [72] used the ratio of the estimated population size of case and control populations to compute test statistic for each genomic region. Burke *et al.* [13] applied Fisher exact test to the last observation of data on case and control populations. Orozco-terWengel *et al.* [47] used the Cochran-Mantel-Haenszel (CMH) test [1] to detect SNPs whose read counts change consistently across all replicates of two time-point data. Turner *et al.* [64] 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.* [8] calculated 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.* [32] computed test statistic of generalized linear-mixed model directly from read counts.

Alternatively, *direct* methods have been developed to analyze time-series data by taking a likelihood approach, and estimating population genetics parameters. Bollback *et al.* [11] proposed a Hidden Markov Model (HMM) to estimate the selection coefficient s and population size by using a diffusion approximation to the Wright Fisher process. Steinrücken *et al.* [58] proposed a general diploid selection model which takes into account of dominance of the favored allele and approximates likelihood analytically. Recently, Schraiber *et al.* [55] proposed a Bayesian framework to estimate parameters using Monte Carlo Markov chain sampling. Mathieson and McVean [44] adopted HMMs to structured populations and estimated parameters using an Expectation Maximization (EM) procedure on discretized allele frequency. Feder *et al.* [23] modeled increments in allele frequency with a Brownian motion process, proposed the Frequency Increment Test (FIT). More recently, Topa *et al.* [63] proposed a Gaussian Process (GP) for modeling single-locus time-series pool-seq data. Terhorst *et al.* [61] extended GP to compute joint likelihood of multiple loci under null and alternative hypotheses. Finally, Levy *et al.* [39] proposed a Bayesian model to handle sequencing, amplification and growth noise in a large population of barcoded lineages.

Among the methods specifically designed for time-series data, many make assumptions which may not hold in E&R studies. One common assumption is that the underlying population size is large, so it is reasonable to model dynamics of allele frequencies using continuous state models [11,

23, 61]. Second, many existing methods were originally designed to process wider time spans seen in ancient DNA studies, an assumption that does not hold for E&R experiments [55, 58]. Finally, many E&R analysis tools assume that allele frequencies in the input data are unbiased (e.g. [11]), which may not be valid for shotgun sequencing experiments.

Here, we consider a Hidden Markov Model (HMM), similar to Williamson *et al.* [68] and Bollback *et al.*'s [11] but under a "small-population-size" regime. Specifically, we use a discrete state (frequency) model. We show that for small population sizes, discrete models can compute likelihood exactly, which improves statistical performance, especially for short time-span experiments. Additionally, we add another level of sampling-noise to the traditional HMM model, allowing for heterogeneous ascertainment bias due to uneven coverage among variants. We show that for a wide range of parameters, CLEAR provides higher power for detecting selection, 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

Consider a panmictic diploid population with fixed size of N individuals. Let $\boldsymbol{\nu} = \{\nu_t\}_{t\in\mathcal{T}}$ be frequencies of the derived allele at generations $t \in \mathcal{T}$ for a given variant, where at generations $\mathcal{T} = \{\tau_i : 0 \leq \tau_0 < \tau_1 \ldots < \tau_T\}$ samples of n individuals are chosen for pooled sequencing. The experiment is replicated R times. We denote allele frequencies of the R replicates by the set $\{\boldsymbol{\nu}\}_R$. To identify the genes and variants that are responding to selection pressure, we use the following procedure:

1. **Estimating population size.** The procedure starts by estimating the effective population size, $\widehat{N}$, under the assumption that much of the genome is evolving neutrally.

2. **Estimating selection parameters.** For each polymorphic site, selection and dominance parameters s, h are estimated so as to maximize the likelihood of the time series data, given $\widehat{N}$.

3. **Computing likelihood statistics.** For each variant, a log-odds ratio of the likelihood of selection model ($s > 0$) to the likelihood of neutral evolution/drift model is computed. Likelihood ratios in a genomic region are combined to compute the CLEAR statistic for the region.

4. **Hypothesis testing.** An empirical null distribution of the CLEAR statistic is calculated using genome-wide drift simulations, and used to compute p-values and thresholds for a specified FDR. We perform single locus hypothesis testing within selected regions to identify significant variants and report genes that intersect with the selected variants.

These steps are described in detail below.

2.1 Estimating Population Size

Methods for estimating population sizes from temporal neutral evolution data have been developed [2, 11, 33, 61, 68]. Here, we aim to extend these models to explicitly model the sampling noise that arise in pool-seq data. Specifically, we model the variation in sequence coverage over different locations, and the noise due to sequencing only a subset of the individuals in the population. In addition, many existing methods [11, 23, 61, 63] are designed for large populations, and model frequency as a continuous quantity. We observed that using Brownian motion to model frequency

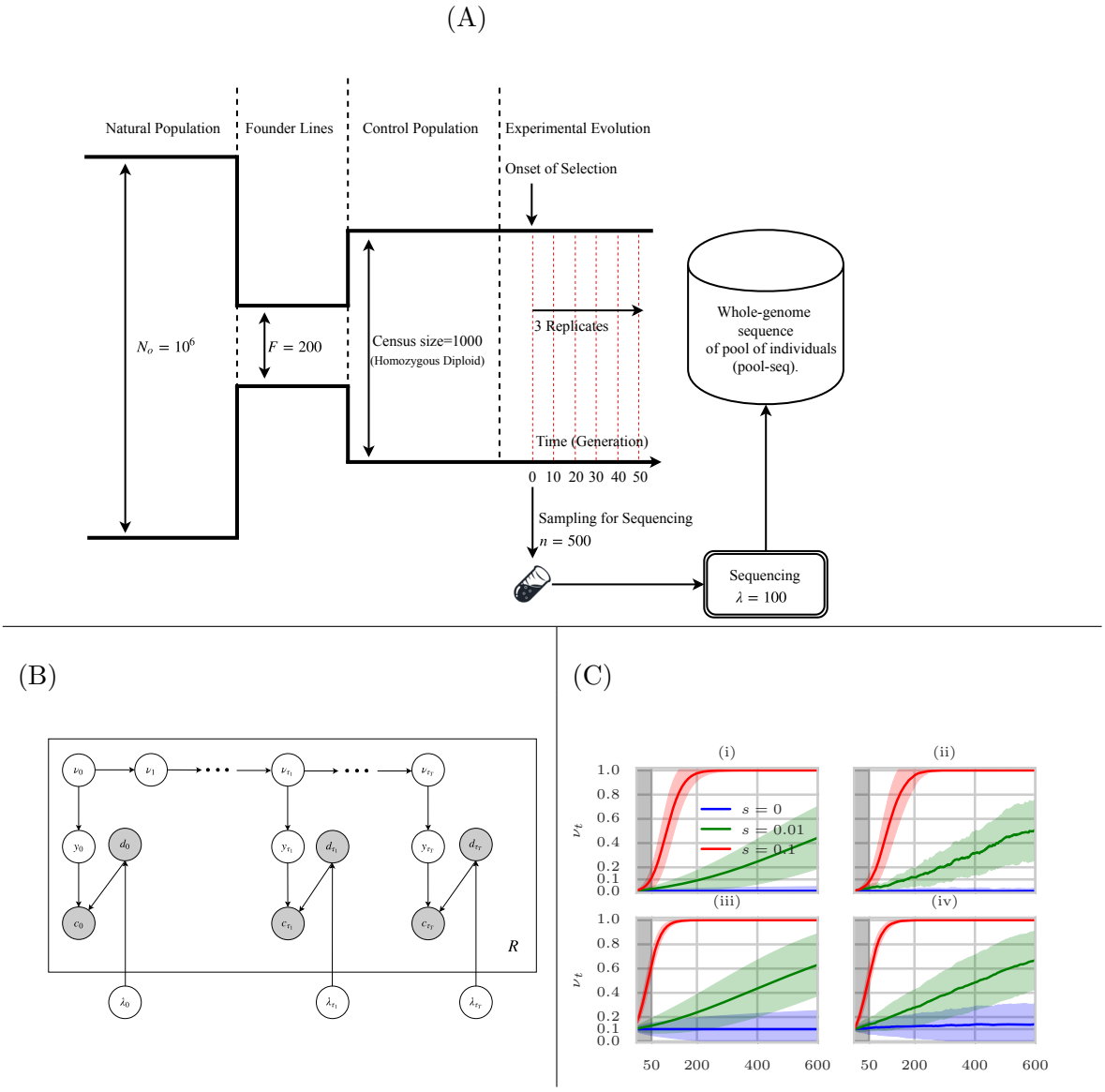

Fig 1: **Evolve and Resequence Selection Experiments on *D. melanogaster* .** (A) Typical configuration in which time-series data is collected for *D. melanogaster* . A small set of founder lines ($F = 200$) is selected from a large population ($N_o = 10^6$), and used to create a sub-population of isofemale lines. Multiple replicates of the population are evolved and resequenced to collect time-series genomic data. For sequencing, n individuals are randomly sampled and sequenced with coverage λ. (B) Graphical model showing dependence of the random variables in the single-locus model used to compute CLEAR statistics. Observed variables, c (derived allele read count) and d (total read count) are shaded. The variables ν, y, λ denote allele frequency, sampled allele frequency, and mean sequencing coverage, respectively. (C) Mean and 95% confidence interval of the theoretical (i,iii) and empirical (ii,iv) trajectories of the favored allele for hard (i,ii) and soft (iii,iv) sweep scenarios and $N = 1000$. The first 50 generations are shaded in gray to represent the sampling span of sampling in short-term experiments, illustrating the difficulty in predicting selection at early stages of selective sweep.

drift may be inadequate for small populations, low starting frequencies and sparse sampling (in time), factors that are common in experimental evolution (see Results, Fig 12A-C, and Fig 11). To this end, we model the Wright-Fisher Markov process for generating pool-seq data (Fig S1) via a discrete HMM (Fig 10-B). We start by computing a likelihood function for the population size given neutral pool-seq data.

Likelihood for Neutral Model. We model the allele frequency counts $2N\nu_t$ as being sampled from a Binomial distribution. Specifically,

$$\begin{aligned} \nu_0 &\sim \pi, \\ 2N\nu_t|\nu_{t-1} &\sim \text{Binomial}(2N, \nu_{t-1}) \end{aligned}$$

where π is the global distribution of allele frequencies in the base population. Note that π depends on the demographic history of the founder lines and can be estimated from site frequency spectrum(see Fig S2) of the initial population. For notational convenience, henceforth we omit the dependence of likelihoods to the parameter π.

To estimate frequency after τ 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 τ generations:

$$\begin{aligned} P^{(1)}[i,j] &= \Pr\left(\nu_{t+1} = \frac{j}{2N} \mid \nu_t = \frac{i}{2N}\right) \\ &= \binom{2N}{j} s\nu_t^j (1-\nu_t)^{2N-j}, \end{aligned} \tag{1}$$

$$P^{(\tau)} = P^{(\tau-1)}P^{(1)} \tag{2}$$

Furthermore, in an E&R experiment, $n \leq N$ individuals are randomly selected for sequencing. The sampled allele frequencies, $\{y_t\}_{t\in\mathcal{T}}$, are also Binomially distributed

$$2ny_t \sim \text{Binomial}(2n, \nu_t) \tag{3}$$

We introduce the $2N \times 2n$ sampling matrix Y, where $Y[i,j]$ stores the probability that the sample allele frequency is $j/2n$ 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 coverage, and the read count of the derived allele, respectively. Let $\{\lambda_t\}_{t\in\mathcal{T}}$ be the sequencing coverage at different generations. Then, the observed data are sampled according to

$$d_t \sim \text{Poisson}(\lambda_t), \qquad c_t \sim \text{Binomial}(d_t, y_t) \tag{4}$$

The emission probability for a observed tuple $x_t = \langle d_t, c_t\rangle$ is

$$\mathbf{e}_i(x_t) = \binom{d_t}{c_t}\left(\frac{i}{2n}\right)^{c_t}\left(1-\frac{i}{2n}\right)^{d_t-c_t}. \tag{5}$$

For $1 \leq t \leq T, 1 \leq j \leq 2N$, let $\alpha_{t,j}$ denote the probability of emitting $x_1, x_2, \ldots, x_t$ and reaching state j at τ_t. Then, α_t can be computed using the forward-procedure [19]:

$$\alpha_t^T = \alpha_{t-1}^T P^{(\delta_t)} \text{diag}(Y\mathbf{e}(x_t)) \tag{6}$$

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

$$\begin{aligned}\mathcal{L}(N|\{\boldsymbol{x}\}_R, n) &= \prod_{r=1}^{R} \mathcal{L}(N|\boldsymbol{x}^{(r)}, n) = \Pr(\{\boldsymbol{x}\}_R|N, n) \\ &= \prod_{r=1}^{R} \sum_i \alpha_{T,i}^{(r)}\end{aligned} \tag{7}$$

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 10-B and Fig S1, respectively.

Finally, the last step is to compute an estimate $\widehat{N}$ that maximizes the likelihood of all M variants in whole genome. Let $\boldsymbol{x}_i^{(r)}$ denote the time-series data of the i-th variant in replicate r. Then,

$$\widehat{N} = \arg\max_N \prod_{i=1}^{M}\prod_{r=1}^{R} \mathcal{L}(N|\boldsymbol{x}_i^{(r)}) \tag{8}$$

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 dominance 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 $w_{11} = 1 + s$. Then, ν_{t^+}, the frequency at time $\tau_t + 1$ (one generation ahead), can be estimated using:

$$\begin{aligned}\hat{\nu}_{t^+} = \mathbb{E}[\nu_{t^+}|s, h, \nu_t] &= \frac{w_{11}\nu_t^2 + w_{01}\nu_t(1-\nu_t)}{w_{11}\nu_t^2 + 2w_{01}\nu_t(1-\nu_t) + w_{00}(1-\nu_t)^2} \\ &= \nu_t + \frac{s(h + (1-2h)\nu_t)\nu_t(1-\nu_t)}{1 + s\nu_t(2h + (1-2h)\nu_t))}.\end{aligned} \tag{9}$$

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 τ generations, then (See [20], Pg. 24, Eqn. 1.58-1.59):

$$\begin{aligned}Q_{s,h}^{(1)}[i,j] &= \Pr\left(\nu_{t^+} = \frac{j}{2N} \,\middle|\, \nu_t = \frac{i}{2N}; s, h, N\right) \\ &= \binom{2N}{j} \hat{\nu}_{t^+}^{j} (1-\hat{\nu}_{t^+})^{2N-j}\end{aligned} \tag{10}$$

$$Q_{s,h}^{(\tau)} = Q_{s,h}^{(\tau-1)} Q_{s,h}^{(1)} \tag{11}$$

The maximum likelihood estimates are given by

$$\widehat{s}, \widehat{h} = \arg\max_{s,h} \prod_{r=1}^{R} \mathcal{L}(s, h|\boldsymbol{x}^{(r)}, \widehat{N}) \tag{12}$$

Using grid search, we first estimate N (Eq. 8), and subsequently, we estimate parameters s, h (Eq. 12, Fig S3). 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 13).

2.3 Empirical Likelihood Ratio Statistics

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

$$H = -2\log\left(\frac{\mathcal{L}(\bar{s}, 0.5|\{\boldsymbol{x}\}_R, \widehat{N})}{\mathcal{L}(0, 0.5|\{\boldsymbol{x}\}_R, \widehat{N})}\right), \tag{13}$$

where $\bar{s} = \arg\max_s \prod_{r=1}^{R} \mathcal{L}(s, 0.5|\boldsymbol{x}^{(r)}, \widehat{N})$. Similarly we can define a test statistic for testing if selection is dominant by

$$D = -2\log\left(\frac{\mathcal{L}(\widehat{s}, \widehat{h}|\{\boldsymbol{x}\}_R, \widehat{N})}{\mathcal{L}(\bar{s}, 0.5|\{\boldsymbol{x}\}_R, \widehat{N})}\right). \tag{14}$$

While extending the single-locus WF model to a multiple linked-loci can improve the power of the model [61], it is computationally and statistically expensive to compute exact likelihood. In addition, computing linked-loci joint likelihood requires haplotype resolved data, which pool-seq does not provide. Here, similar to Nielsen *et al* [46], we calculate *composite likelihood ratio* score for a genomic region.

$$\mathcal{H} = \frac{1}{|L|}\sum_{\ell \in L} H_\ell. \tag{15}$$

where L is a collection of segregating sites and H_ℓ is the likelihood ratio score based for each variant ℓ 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, linkage of the favored allele to neighboring variants, population size, coverage, and time since the onset of selection (duration of the experiment). In S1 Text, we provide a heuristic to compute a reasonable value of L, based on experimental data.

We work with a normalized value of $\mathcal{H}$, given by

$$\mathcal{H}_i^* = \frac{\mathcal{H}_i - \mu_{\mathcal{C}}}{\sigma_{\mathcal{C}}}, \qquad \forall i \in \mathcal{C}, \tag{16}$$

where $\mu_{\mathcal{C}}$ and $\sigma_{\mathcal{C}}$ are the mean and standard deviation of $\mathcal{H}$ values in a large region $\mathcal{C}$. We found different chromosomes to have different distribution of $\mathcal{H}_i$ values, and therefore decided to use single chromosomes as $\mathcal{C}$.

2.4 Hypothesis Testing

Single-Locus tests. Under neutrality, Log-likelihood ratios can be approximated by $\mathcal{X}^2$ distribution [67], and p-values can be computed directly. However, Feder *et al.* [23] showed that when the number of independent samples (replicates) is small, $\mathcal{X}^2$ is a crude approximation to the true null distribution and results in more false positive. Following their suggestion, we first compute the empirical null distribution using simulations with the estimated population size (See Fig S1). The empirical null distribution of statistic H is used to compute p-values as the fraction of null values that exceed the test score. Finally, we use Storey and Tibshirani's method [60] to control for False Discovery Rate in multiple testing.

Composite likelihood tests. Similar to single-locus tests, we compute the null distribution of the $\mathcal{H}^*$ statistic using whole-genome simulations with the estimated population size, and subsequently compute FDR. The simulations for generating the null distribution of $\mathcal{H}^*$ are described next.

2.5 Simulations

We use the same simulation procedure for two purposes. First, we use them to test the power of CLEAR against other methods in small genomic windows. Second, we use the simulations to generate the distribution of null values for the statistic to compute empirical p-values. We mainly chose parameters that are relevant to *D. melanogaster* experimental evolution [35]. See also Fig 10-A for illustration.

1. **Creating initial founder line haplotypes.** Using `msms` [21], we created neutral populations for F founding haplotypes with command `$./msms <F> 1 -t <2`$\mu W N_o$`> -r <2`rWN_o`> <W>`, where $F = 200$ is number of founder lines, $N_o = 10^6$ is effective founder population size, $r = 2 \times 10^{-8}$ is recombination rate, $\mu = 2 \times 10^{-9}$ is mutation rate. The window size W is used to compute $\theta = 2\mu N_o W$ and $\rho = 2N_o rW$. We chose $W = 50$Kbp for simulating individual windows for performance evaluations, and $W = 20$Mbp for simulating *D. melanogaster* chromosomes for p-value computations.

2. **Creating initial diploid population.** An initial set of $F = 200$ haplotypes was created from step I, and duplicated to create F homozygous diploid individuals to simulate generation of inbred lines. N diploid individuals were generated by sampling with replacement from the F individuals.

3. **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 $r = 2\times10^{-8}$ and sampling times $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$, `simuPop` [49] 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. For generating the null distribution with drift for p-value computations, we used this procedure with $s = 0$.

4. **Sequencing Simulation.** Given allele frequency trajectories we sampled depth of each site in each replicate identically and independently from Poisson(λ), where $\lambda \in \{30, 100, 300\}$ is the coverage for the experiment. Once depth d is drawn for the site with frequency ν, the number of reads c carrying the derived allele are sampled according to Binomial(d, ν). 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 Brownian motion (a continuous-state model) 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 ν_0, we computed the expected distribution of the frequency of the next sample ν_τ under two models to make a comparison. Fig 11A-F shows that Brownian motion (continuous model) is inadequate when ν_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

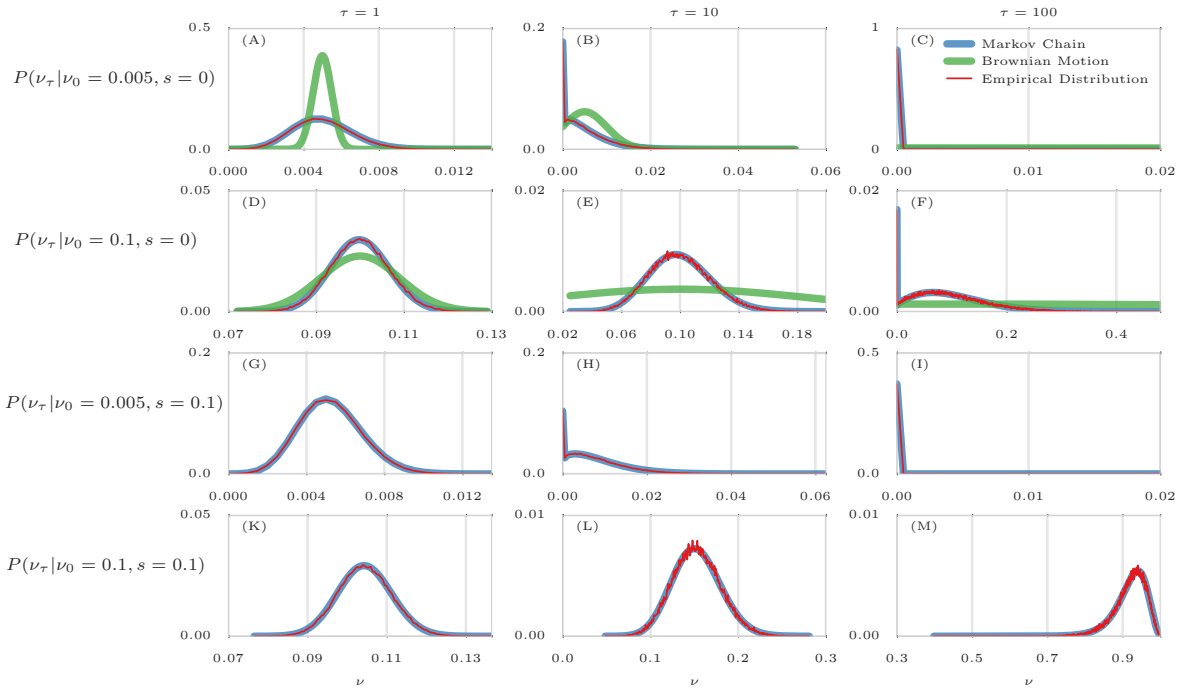

Fig 2: **Comparison of empirical distributions of allele frequencies (red) versus predictions from Brownian Motion (green), and Markov chain (blue).**
Comparison of empirical and theoretical distributions under neutral evolution (panels A-F) and selection (panels G-M) with different starting frequencies $\nu_0 \in \{0.005, 0.1\}$ and sampling times of $\mathcal{T} = \{0, \tau\}$, where $\tau \in \{1, 10, 100\}$ and $N = 1000$. 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 provides a good approximation in all cases.

Fig S2). Moreover, in typical *D. melanogaster* experiments for example, sampling is sparse. Often, the experiment is designed so that $10 \leq \tau \leq 100$ [26, 35, 47, 72].

In contrast to the Brownian motion approximation, discrete Markov chain predictions (Eq. 11) are highly consistent with empirical data for a wide range of simulation parameters (Fig 11A-M). Moreover, the discrete markov chain can be modified to model the case when the the allele is under selection.

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 ≤ 0.05. For each configuration (specified with values for selection coefficient s, starting allele frequency ν_0 and coverage λ), 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) [61], FIT [23], and CMH [1] 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 12 and Table S1). 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

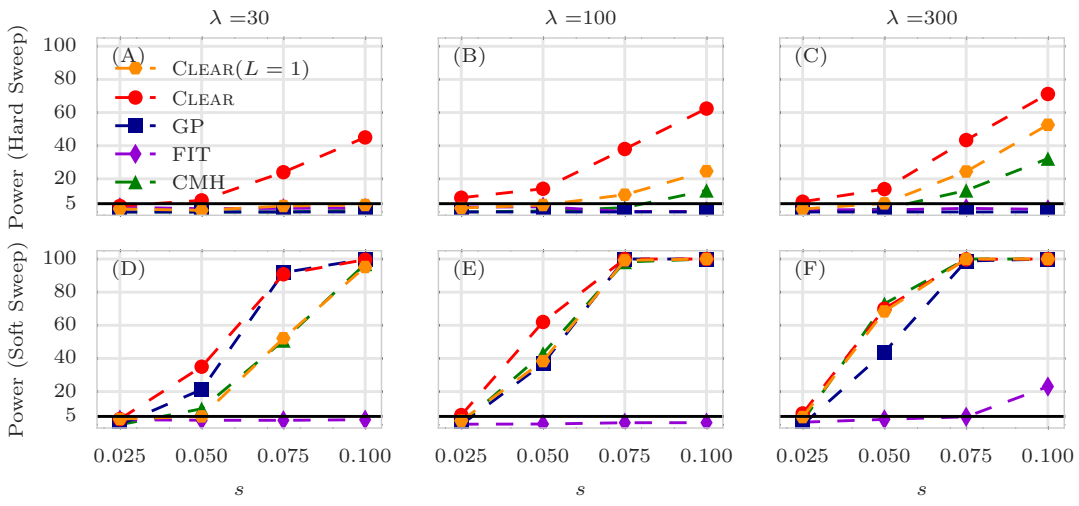

Fig 3: **Power calculations for detection of selection.**
Detection power for CLEAR($\mathcal{H}$), Frequency Increment Test (FIT), Gaussian Process (GP), and CMH under hard (A-C) and soft sweep (D-F) scenarios. λ, s denote the mean coverage and selection coefficient, respectively. Orange hexagons represent the performance of CLEAR when the maximum of the single-locus statistic is used to make a decision for the genomic region, while the red circle corresponds to the performance of CLEAR when single locus statistics are averaged over the region. The y-axis measures power – sensitivity with false positive rate FPR ≤ 0.05 – for $2,000$ simulations with $N = 1,000$, $L = 50$Kbp. The horizontal line reflects the power of a random classifier. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

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 12B,C) suggest that taking many samples with lower coverage is preferable to sparse sampling with higher coverage. For comparison purposes, we also tested CLEAR using the single locus statistic ($L = 1$). For the most part, CLEAR showed an improvement over other methods even with $L = 1$, or showed similar performance. The performance improved with higher L.

Site-identification. In general, localizing the favored variant, using pool-seq data is a nontrivial task due to extensive linkage disequilibrium [62]. To measure performance, we sorted variants by their H scores and computed rank of the favored allele for each method. For each setting of ν_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 14) shows that CLEAR outperforms other statistics.

An interesting observation is revisiting the contrast between site-identification and detection [40, 62]. When selection strength is high, detection is easier (Fig 12A-F), but site-identification is harder, due to the high LD between flanking variants and the favored allele (Fig 14A-F). Moreover, site-identification becomes more difficult whenever the initial frequency of the favored allele is low, i.e., at the onset of selection, LD between favored allele and its nearby variants is high. For example,

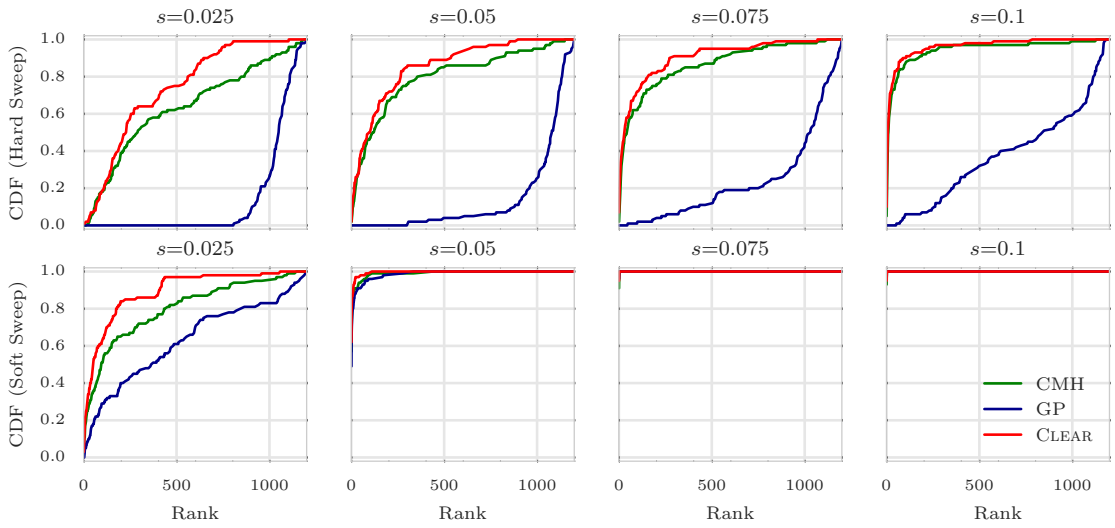

Fig 4: **Ranking performance for 100× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H), Gaussian Process (GP), CMH, and Frequency Increment Test (FIT), for different values of selection coefficient s and initial 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 with $N = 1000$ are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

when coverage $\lambda = 100$ and selection coefficient $s = 0.1$, the detection power is 75% for hard sweep, but 100% for soft sweep (Fig 12B-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 estimates effective population size $\widehat{N}$ and selection parameters, $\hat{s}$ and $\hat{h}$, as a byproduct of the hypothesis testing. We computed bias of selection fitness $(s - \hat{s})$ and dominance $(h - \hat{h})$ for of CLEAR and GP for 1000 simulations in each setting. The distribution of the error (bias) for 100× coverage is presented in Fig 15 for different configurations. Fig S4 and Fig S5 provide the distribution of estimation errors for 30×, and 300× coverage, respectively. For hard sweep, CLEAR provides estimates of s with lower variance of bias (Fig 15A and Fig S6). In soft sweep, GP and CLEAR both provide unbiased estimates of s with low variance (Fig 15B). Fig 15 C-D shows that CLEAR provides unbiased estimates of h as well when $h \in \{0, 0.5, 1, 2\}$ and $s = 0.1$. We also tested if CLEAR provide unbiased estimates of N, by estimating population size on 1000 simulations when $N \in \{200, 600, 1000\}$. As shown in Fig 16-A and Fig S9A-C, maximum likelihood is attained at true value of the parameter.

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 $\mathcal{O}(TRN^3)$ 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 13) that CLEAR is orders of magnitude faster than GP, and comparable to FIT. While slower than

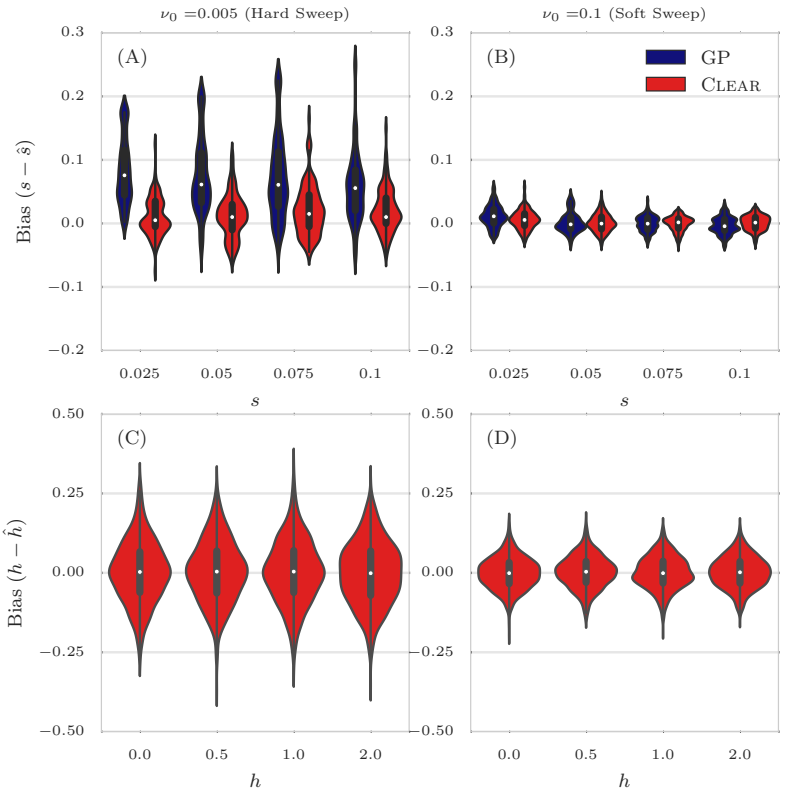

Fig 5: **Distribution of bias for 100× coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = 100$ (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 Table S2. Panels C,D show the variance in the estimation of h. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

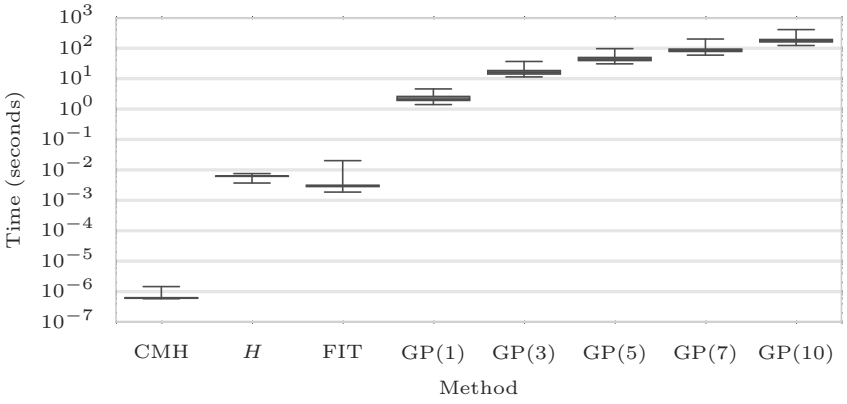

Fig 6: **Running time.**
Box plots of running time per variant (CPU-secs.) of CLEAR($\mathcal{H}$), CMH, FIT, and GP with single, 3, 5, 7, and 10 loci over 1000 simulations conducted on a workstation with Intel Core i7 processor. The average running time for each method is shown on the x-axis. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

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 Alternating Temperatures

We applied CLEAR to the data from a study of *D. melanogaster* adaptation to alternating temperatures [26, 47], where 3 replicate samples were chosen from a population of *D. melanogaster* for 59 generations under alternating 12-hour cycles of hot stressful (28°C) and non-stressful (18°C) temperatures and sequenced. In this dataset, sequencing coverage is different across replicates and generations (see S2 Fig of [61]) which makes variant depths highly heterogeneous (Fig S10).

We first filtered out heterochromatic, centromeric and telomeric regions [25], and those variants 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.

Next, we estimated genome-wide population size $\widehat{N} = 250$ (Fig 16-B and Fig S9-E) which is consistent with previous studies [33, 47]. The likelihood curves of CLEAR are sharper around the optimum compared to that of Bollback et. al [11]'s method (see Supplementary Fig. 1 in [47]). Also, chromosomes 3L and 3R appear to have smaller population size, $\widehat{N} = 200, 150$, respectively. Others have made similar observations on this data. In particular, Jónás *et al.* [33] shown that the chromosome-wise population size varies even more when it is computed for each replicate separately (see Table 1 in [33]). For instance, $\widehat{N}$ is 131 for chromosome 3R replicate 1, while it is 328 for chromosome X replicate 2.

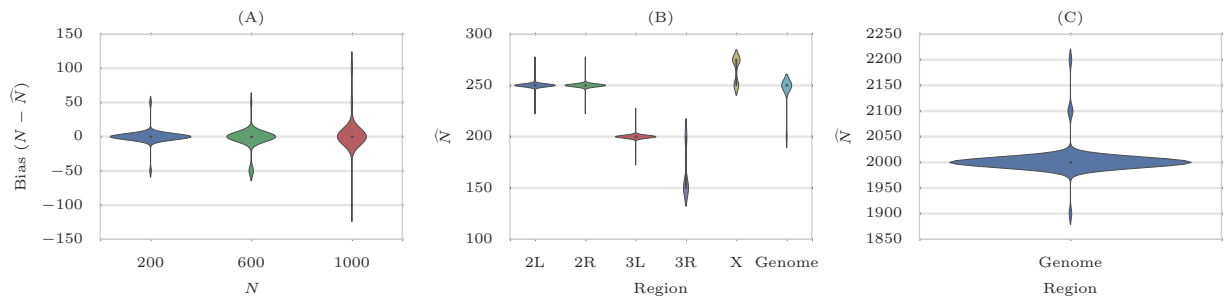

Fig 7: **Estimating population size.** (A) Distribution of bias in estimating N, computed on 1000 neutral simulations for each $N \in \{200, 600, 1000\}$ when $W = 10\text{Mbp}$ and $r = 2 \times 10^{-8}$. (B) Estimates of population size for data from a study of *D. melanogaster* adaptation to alternating temperatures. For each case, the distribution of estimator is computed by 100 bootstrap computations using 1000 variants each. The multiple modes are an artifact of grid search used to speed up computation. (C) Distribution of the population size estimates on the yeast dataset. Despite large census population size ($10^6 - 10^7$ [14]), this dataset exhibits much smaller effective population size ($\widehat{N} = 2000$).

While it would be ideal to compute CLEAR statistic for each replicate and chromosome separately, computing empirical p-values and significant regions become computationally intensive as empirical null distribution of each replicate and each chromosome needs to be computed. Hence, we use a single genome-wide estimate $\widehat{N} = 250$ in all analyses, but we normalize statistic $\mathcal{H}^*$ separately for each chromosome.

We use a heuristic calculation (See S1 Text) to choose the sliding window size L as the distance where the LD between the favored mutation and a site $L/2$bp away remains strong. For *D. melanogaster* parameters, we obtained $L = 30\text{Kbp}$. We computed the normalized test statistic $\mathcal{H}^*$ on sliding windows of size of 30Kbp and step size of 5Kbp over the genome (See Fig 17-A).

Empirical null distribution of $\mathcal{H}^*$ was estimated by creating 100 whole genome simulations (400K statistic values) as described in Section 2.5. Then, p-value of the test statistic in each region in the experimental data was calculated as the fraction of the null statistic values that are greater than or equal to the test statistic(see Fig S11). After correcting for multiple testing, we identified 5 contiguous intervals (Fig 17) satisfying FDR≤ 0.05, and covering $2,829$ polymorphic sites. We further performed single-locus hypothesis testing on the $2,829$ sites to identify 174 individual variants with FDR ≤ 0.01 (Fig 17-B).

The final set of 174 variants fall within 32 genes(Table S3) including many Serine inhibitory proteases (serpins), and other genes involved in endocytosis. Recycling of synaptic vesicles is seen to be blocked at high temperature in temperature sensitive Drosophila mutants [36]. This is also supported by GO enrichment analysis, where a single GO term 'inhibition of proteolysis' is found to enriched (corrected p-value:0.0041). To test for dominant selection, we computed D statistic on simulated neutral and experimental data, and computed p-values accordingly. After correcting for multiple testing, 96 variants were discovered with FDR≤ 0.01 (Fig S12).

3.2 Analysis of Outcrossing Yeast Populations

We also applied CLEAR to 12 replicate samples of outcrossing yeast populations [14], where samples are taken at generations $\mathcal{T} = \{0, 180, 360, 540\}$. We observed a significant variation in the genome-wide site frequency spectrum of certain populations over different time points for some

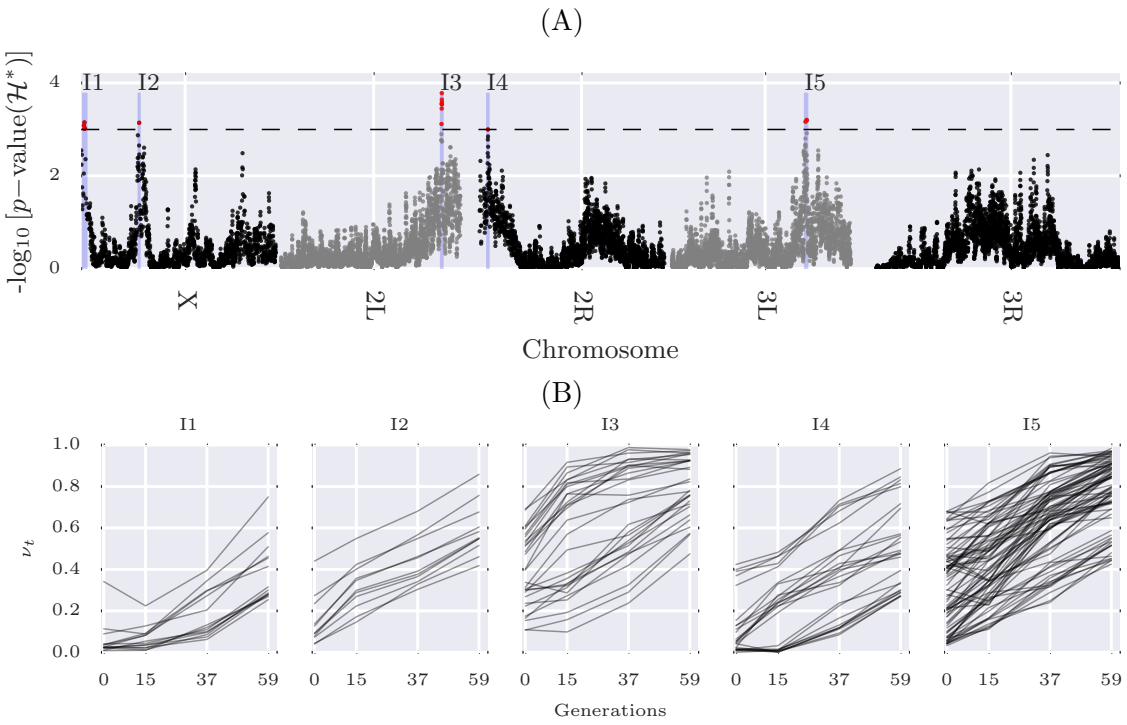

Fig 8: **Scan of CLEAR statistic on data from a study of *D. melanogaster* adaptation to alternating temperatures.** (A) Manhattan plot of scan for $\mathcal{H}^*$ statistic using sliding window of size $L = 3000$ over the genome. The dashed line represents cutoff for genome-wide FDR≤ 0.05, and identifies 5 contiguous intervals, I1-I5, which are shaded in blue. (B) Trajectories of the selected variants within intervals I1-I5.

replicates (Fig S13). The variation does not have an easily identifiable cause. Therefore, we focused analysis on seven replicates $r \in \{3, 7, 8, 9, 10, 11, 12\}$ with genome-wide site-frequency spectrum over the time range (Fig S14).

We estimated population size to be $\widehat{N} = 2000$ haplotypes (Fig 16-C and Fig S9-F), 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 dominant selection, respectively (Fig S12). Selected variants for directional selection are clustered in two regions, which match 2 of the 5 regions (regions C and E in Fig. 2-a in [14]) identified by Burke *et al.* in their preliminary analysis.

4 Discussion

We developed a computational tool, CLEAR, that can detect regions and variants under selection E&R experiments. Using extensive simulations, we show that CLEAR outperforms existing methods in detecting selection, locating the favored allele, and estimating model parameters. Also, while being computationally efficient, CLEAR provide means for estimating populations size and hypothesis testing.

Many factors such as small population size, finite coverage, linkage disequilibrium, finite sampling for sequencing, duration of the experiment and the small number of replicates can limit the

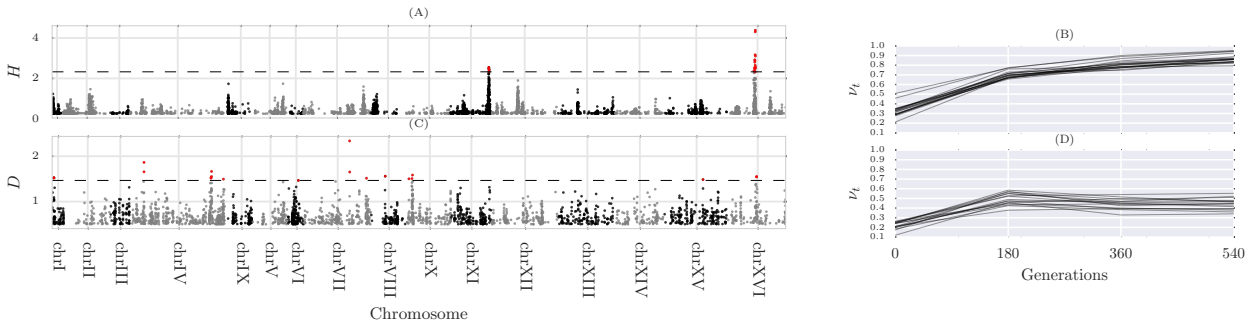

Fig 9: **Single locus analysis of the yeast outcrossing populations.**
Manhattan plot of scan single locus CLEAR statistic ($L = 1$) for testing directional selection (A) and dominant selection (C). The dashed line represents cutoff for genome-wide FDR≤ 0.05. Trajectories of the selected variants are depicted in panels (B) and (D).

power of tools for analyzing E&R. Here, by an discrete modeling, CLEAR estimates population size, and provides unbiased estimates of s, h. It adjusts for heterogeneous coverage of pool-seq data, and exploits presence of linkage within a region to compute composite likelihood ratio statistic.

It should be noted that, even though we described CLEAR for small fixed-size populations, the statistic can be adjusted for other scenarios, including changing population sizes when the demography is known. For large populations, transitions can be computed on sparse data structures, as for large N the transition matrices become increasingly sparse. Alternatively, frequencies can be binned to reduce dimensionality.

The comparison of hard and soft sweep scenarios showed that initial frequency of the favored allele can have an nontrivial effect on the statistical power for identifying selection. Interestingly, while it is easier to detect a region undergoing strong selection, it is harder to locate the favored allele in that region.

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/airanmehr/clear`.

D. melanogaster data originally published [26, 47]. 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 [14] 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. The

authors also thank the anonymous reviewers whose comments substantially improved quality and clarity of this manuscript.

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

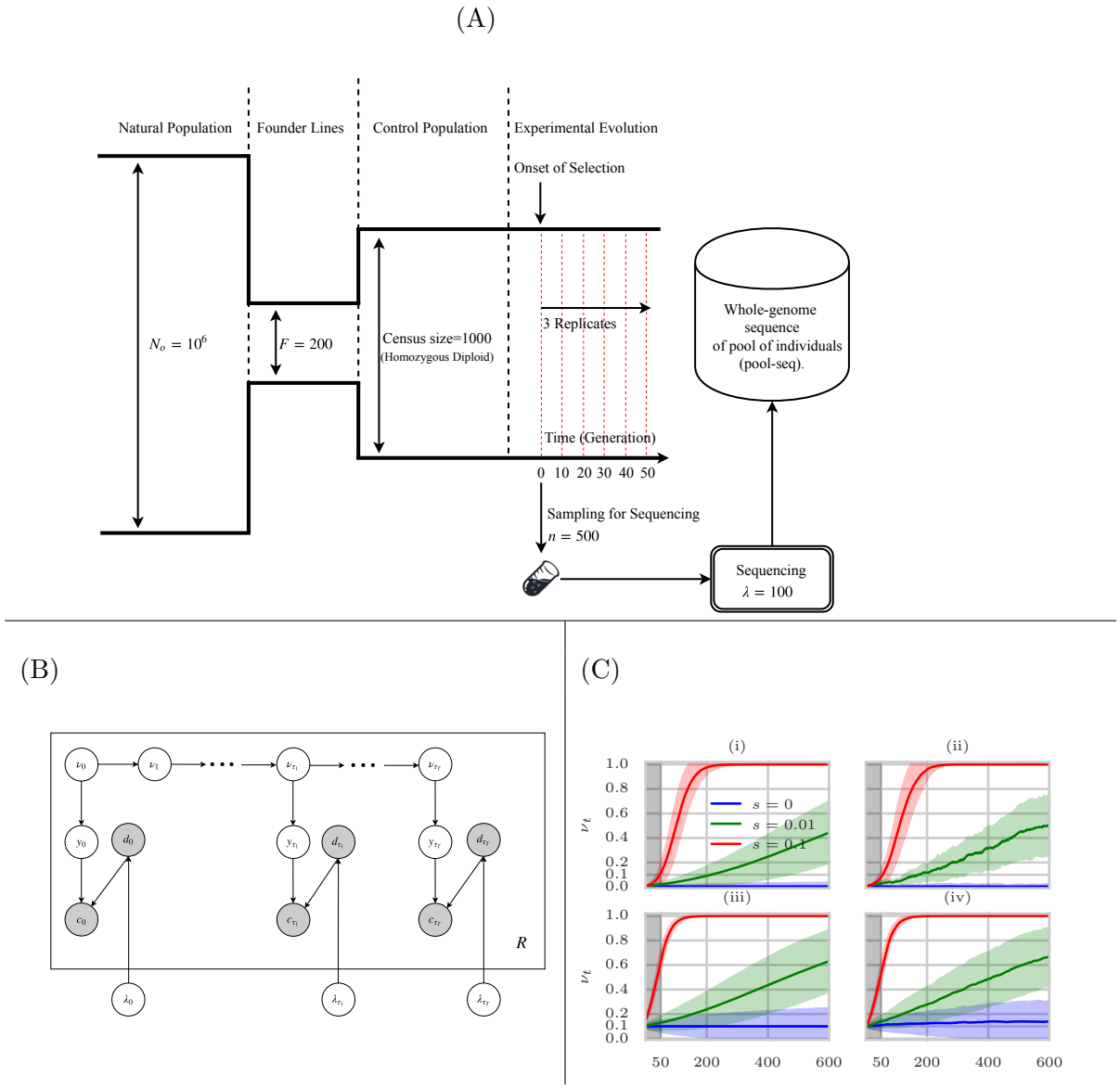

Fig 10: **Evolve and Resequence Selection Experiments on *D. melanogaster* .** (A) Typical configuration in which time-series data is collected for *D. melanogaster* . A small set of founder lines ($F = 200$) is selected from a large population ($N_o = 10^6$), and used to create a sub-population of isofemale lines. Multiple replicates of the population are evolved and resequenced to collect time-series genomic data. For sequencing, n individuals are randomly sampled and sequenced with coverage λ. (B) Graphical model showing dependence of the random variables in the single-locus model used to compute CLEAR statistics. Observed variables, c (derived allele read count) and d (total read count) are shaded. The variables ν, y, λ denote allele frequency, sampled allele frequency, and mean sequencing coverage, respectively. (C) Mean and 95% confidence interval of the theoretical (i,iii) and empirical (ii,iv) trajectories of the favored allele for hard (i,ii) and soft (iii,iv) sweep scenarios and $N = 1000$. The first 50 generations are shaded in gray to represent the sampling span of sampling in short-term experiments, illustrating the difficulty in predicting selection at early stages of selective sweep.

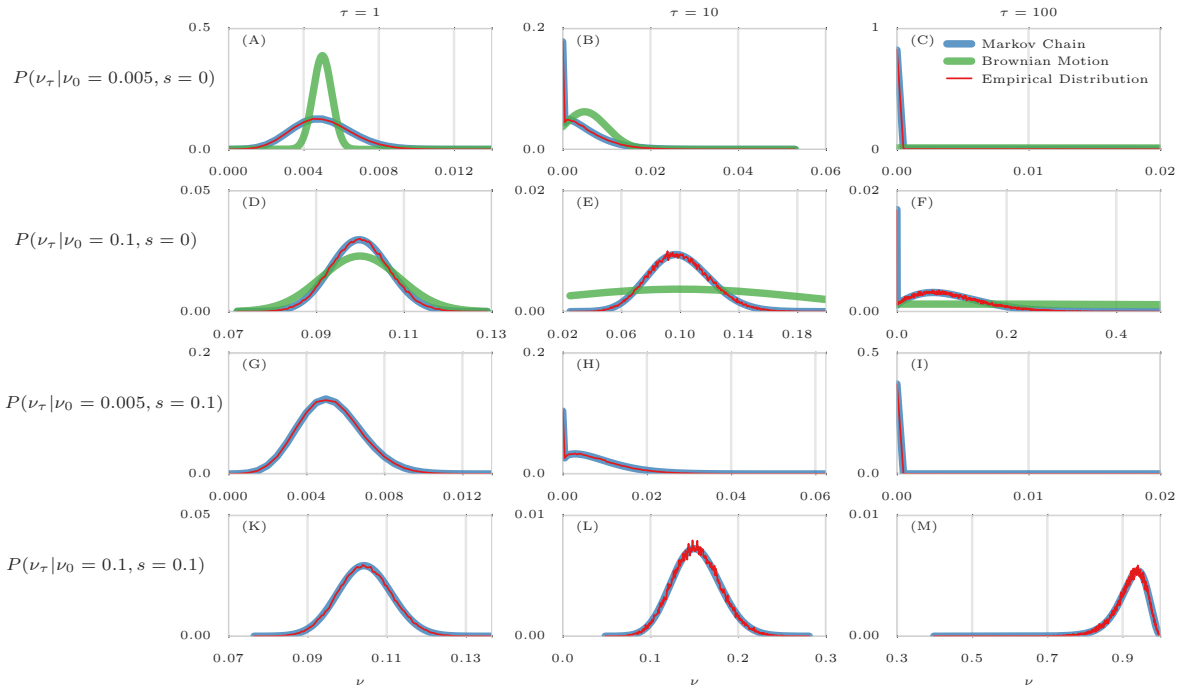

Fig 11: **Comparison of empirical distributions of allele frequencies (red) versus predictions from Brownian Motion (green), and Markov chain (blue).**
Comparison of empirical and theoretical distributions under neutral evolution (panels A-F) and selection (panels G-M) with different starting frequencies $\nu_0 \in \{0.005, 0.1\}$ and sampling times of $\mathcal{T} = \{0, \tau\}$, where $\tau \in \{1, 10, 100\}$ and $N = 1000$. 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 provides a good approximation in all cases.

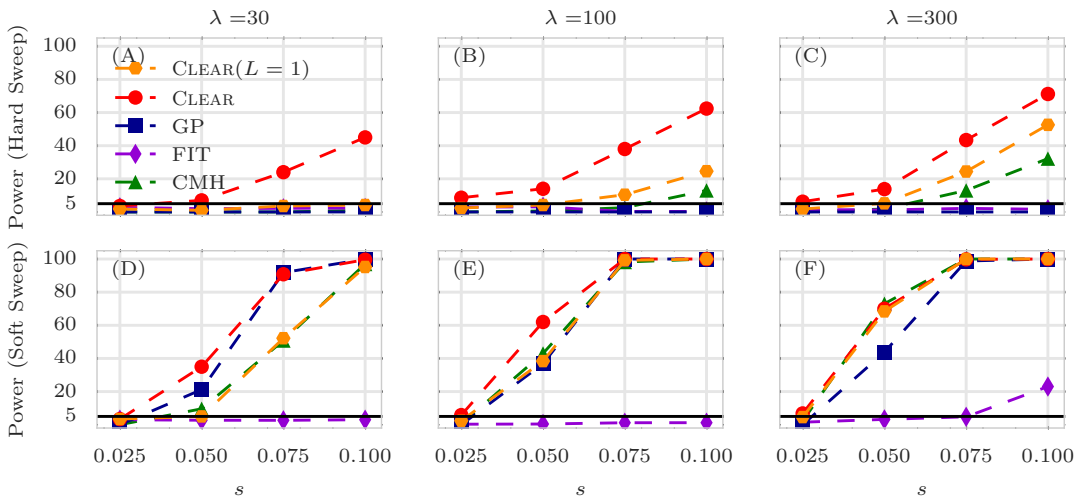

Fig 12: **Power calculations for detection of selection.**
Detection power for CLEAR($\mathcal{H}$), Frequency Increment Test (FIT), Gaussian Process (GP), and CMH under hard (A-C) and soft sweep (D-F) scenarios. λ, s denote the mean coverage and selection coefficient, respectively. Orange hexagons represent the performance of CLEAR when the maximum of the single-locus statistic is used to make a decision for the genomic region, while the red circle corresponds to the performance of CLEAR when single locus statistics are averaged over the region. The y-axis measures power – sensitivity with false positive rate FPR ≤ 0.05 – for $2,000$ simulations with $N = 1,000$, $L = 50$Kbp. The horizontal line reflects the power of a random classifier. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

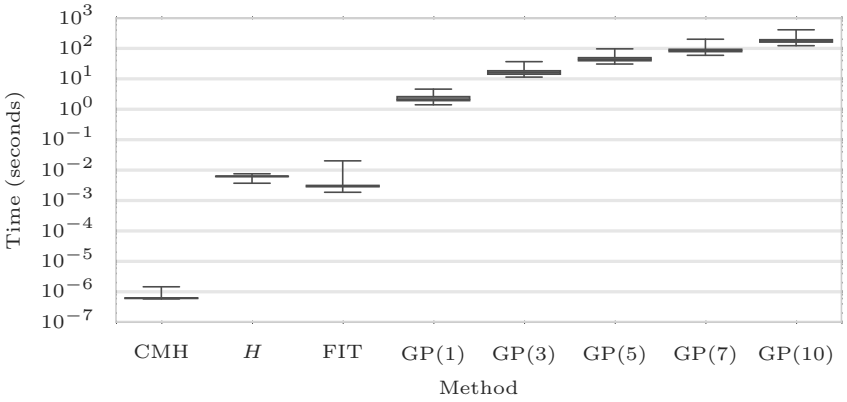

Fig 13: **Running time.**
Box plots of running time per variant (CPU-secs.) of CLEAR($\mathcal{H}$), CMH, FIT, and GP with single, 3, 5, 7, and 10 loci over 1000 simulations conducted on a workstation with Intel Core i7 processor. The average running time for each method is shown on the x-axis. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

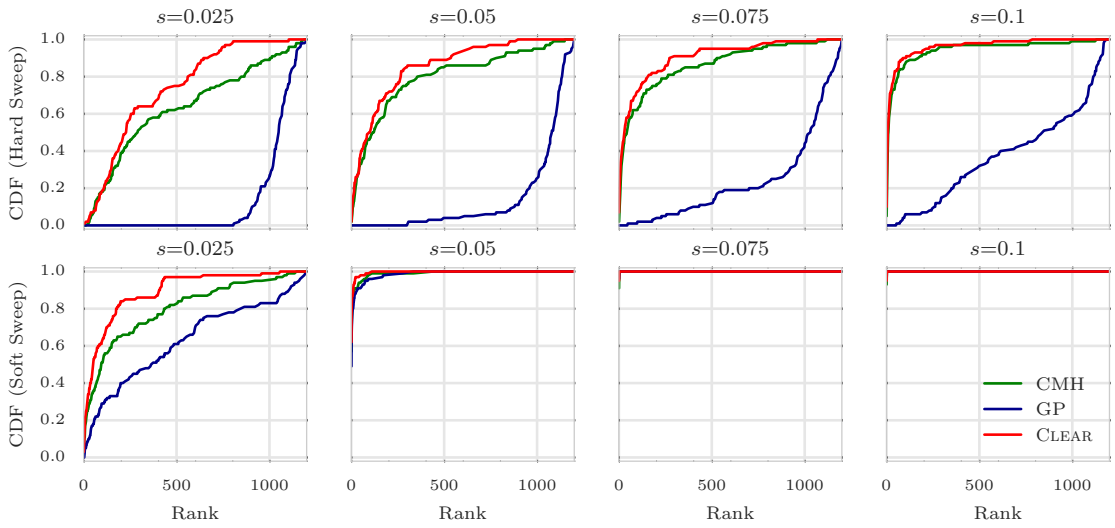

Fig 14: **Ranking performance for 100× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H), Gaussian Process (GP), CMH, and Frequency Increment Test (FIT), for different values of selection coefficient s and initial 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 with $N = 1000$ are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

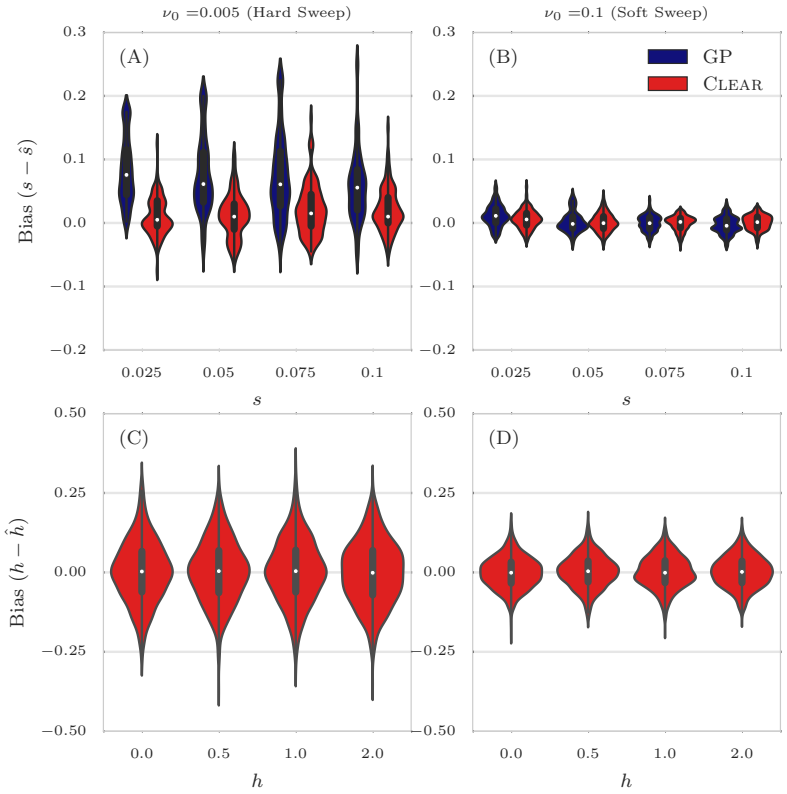

Fig 15: **Distribution of bias for 100× coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = 100$ (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 Table S2. Panels C,D show the variance in the estimation of h. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

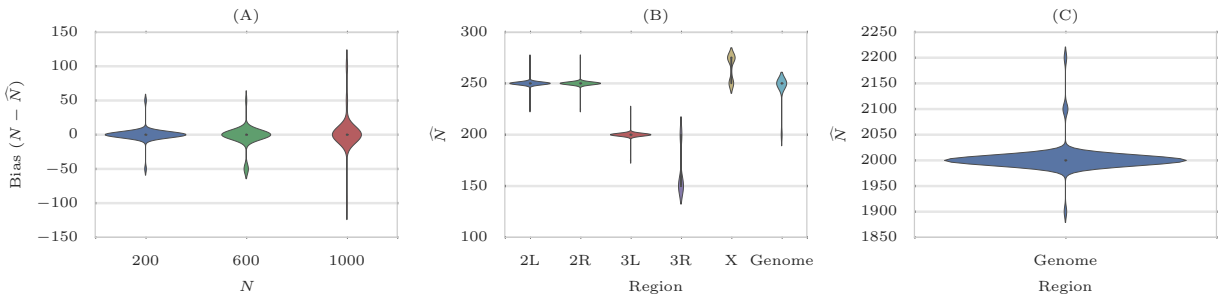

Fig 16: **Estimating population size.** (A) Distribution of bias in estimating N, computed on 1000 neutral simulations for each $N \in \{200, 600, 1000\}$ when $W = 10$Mbp and $r = 2 \times 10^{-8}$. (B) Estimates of population size for data from a study of *D. melanogaster* adaptation to alternating temperatures. For each case, the distribution of estimator is computed by 100 bootstrap computations using 1000 variants each. The multiple modes are an artifact of grid search used to speed up computation. (C) Distribution of the population size estimates on the yeast dataset. Despite large census population size ($10^6 - 10^7$ [14]), this dataset exhibits much smaller effective population size ($\widehat{N} = 2000$).

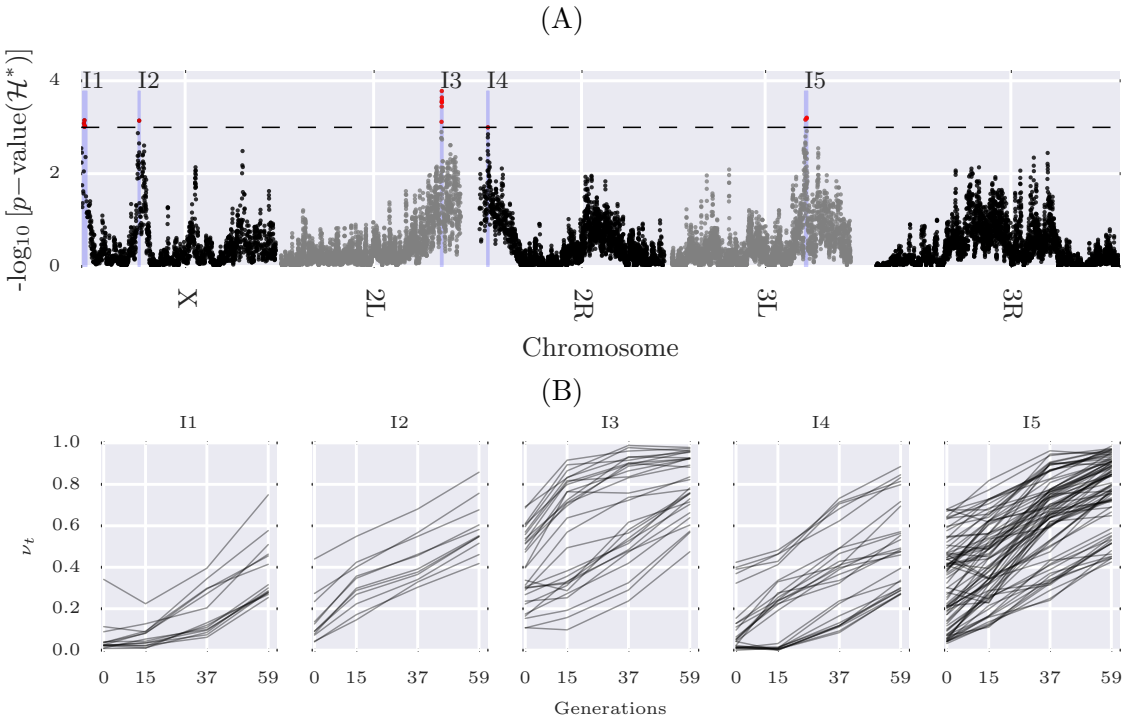

Fig 17: **Scan of CLEAR statistic on data from a study of *D. melanogaster* adaptation to alternating temperatures.** (A) Manhattan plot of scan for $\mathcal{H}^*$ statistic using sliding window of size $L = 3000$ over the genome. The dashed line represents cutoff for genome-wide FDR≤ 0.05, and identifies 5 contiguous intervals, I1-I5, which are shaded in blue. (B) Trajectories of the selected variants within intervals I1-I5.

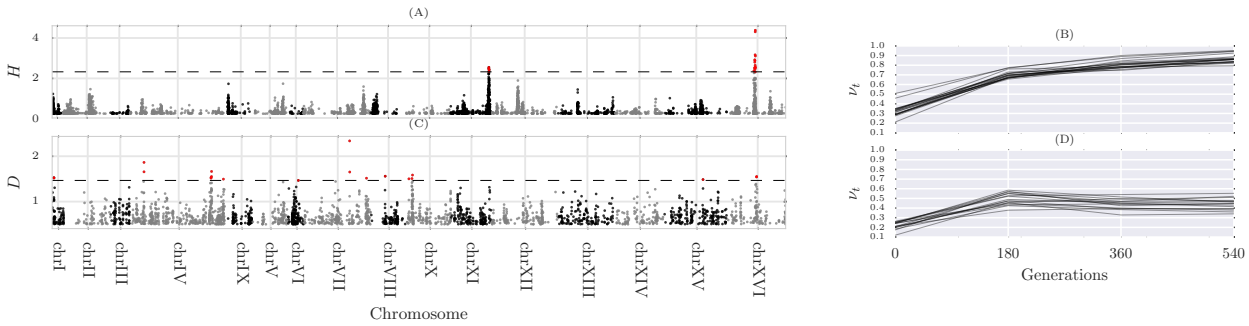

Fig 18: **Single locus analysis of the yeast outcrossing populations.**
Manhattan plot of scan single locus CLEAR statistic ($L = 1$) for testing directional selection (A) and dominant selection (C). The dashed line represents cutoff for genome-wide FDR≤ 0.05. Trajectories of the selected variants are depicted in panels (B) and (D).

S1 Text Choosing Window Size

In genome-wide scans for detecting selection, we apply the CLEAR statistic on sliding windows of length Lbp. The single locus statistic values within the window are averaged to get the composite statistic. While the statistic is robust to variation in window-size, choosing a very large window where LD has decayed will weaken the composite signal, and choosing a small window will decrease the power of composite likelihoods. Here, we use a systematic calculation to choose L as the distance where the LD between the favored mutation and a site $L/2$bp away remains strong.

Consider a segregating site l bp away from the favored allele in a selective sweep. Let ρ_τ be the LD between the favored allele and the site, τ generations after the onset of selection. Then, we have (see Eqs. 30-31 in [59]):

$$\rho_\tau = \alpha_\tau \beta_\tau \rho_0 = e^{-r\tau l} \left(\frac{K^{(\tau)}}{K^{(0)}} \right) \rho_0, \tag{S1}$$

where $K^{(\tau)} = 2\nu_\tau(1-\nu_\tau)$ is the heterozygosity at the selected site, r is the recombination rate (crossovers/bp/gen). The 'decay factor', $\alpha_\tau = e^{-r\tau l}$, and 'growth factor', β_τ, are due to recombination and selection, respectively. Under regular parameter settings, linkage to the favored allele is expected to increase after onset of selection and then decreases due to crossover events (See Fig S15-A). While ρ_0 is unknown in pool-seq E&R experiments, we compute the value of l so that

$$\alpha_\tau \beta_\tau = 1. \tag{S2}$$

In E&R scenarios, we let τ be the time of the last sampling. For given s, we aim to compute the smallest window size L over all possible starting frequencies. Specifically,

$$L = 2 \min_{\nu_0} \left\{ \frac{1}{r\tau} \log \left(\frac{\hat{\nu}_\tau(1-\hat{\nu}_\tau)}{\nu_0(1-\nu_0)} \right) \right\}, \tag{S3}$$

where the term $\hat{\nu}_\tau$ depends on initial frequency ν_0 and selection strength s (Eq. 9).

We used *D. melanogaster* dataset parameters, $N = 250, r = 2 \times 10^{-8}$ and $\tau = 59$ to compute the optimal window size for different values of Ns, ranging from weak selection to strong selection: $Ns \in \{20, 100, 200, 500\}$, or $s \in \{0.08, 0.4, 0.8, 2\}$. We set $L = 30$Kbp (See Fig S15-B) to provide good resolution for detecting weak selection.

Generative Process 1: The Generative Process for Neutral Wright-Fisher Time-series Pool-seq Data.

Input: $N, n, R, \{\lambda_t\}_{t\in\mathcal{T}}, \mathcal{T} = \{\tau_0, \dots \tau_T\}$
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** τ_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

Fig S1: **The Generative Process for Neutral Wright-Fisher Time-series Pool-seq Data.**

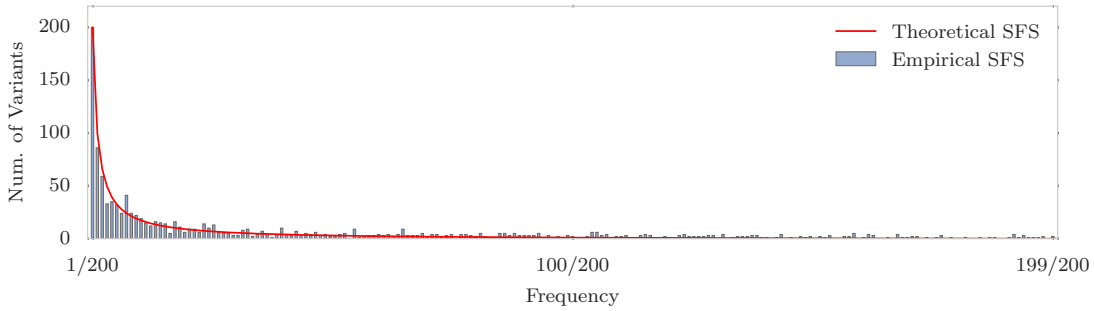

Fig S2: **Site Frequency Spectrum.**
Theoretical and Empirical SFS in a 50Kbp region for a neutral population of 200 individuals when $N_e = 10^6$ and $\mu = 10^{-9}$. The x-axis corresponds to site frequency, and the y-axis to the number of variants with a specific frequency. In a neural population,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

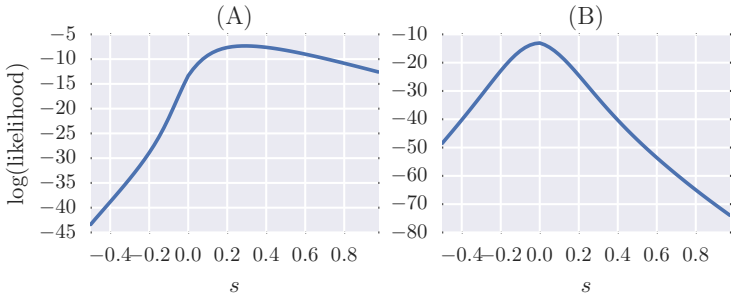

Fig S3: **Likelihoods of the parameter s.**
Likelihood of the parameter s in *D. melanogaster* data for a variant with $\widehat{s} = 0.2$ (A) and $\widehat{s} = 0$ (B).

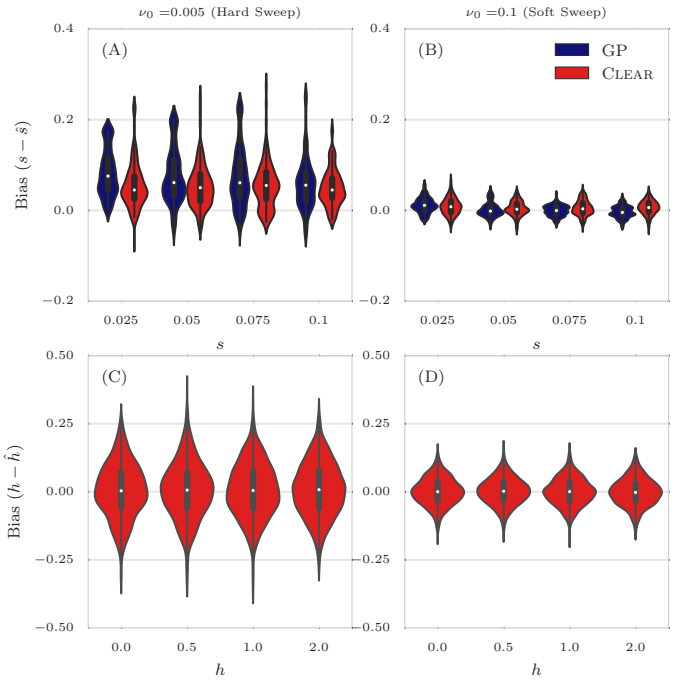

Fig S4: **Distribution of bias for 30× coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = 30$ (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 Table S2. Panels C,D show the variance in the estimation of h.

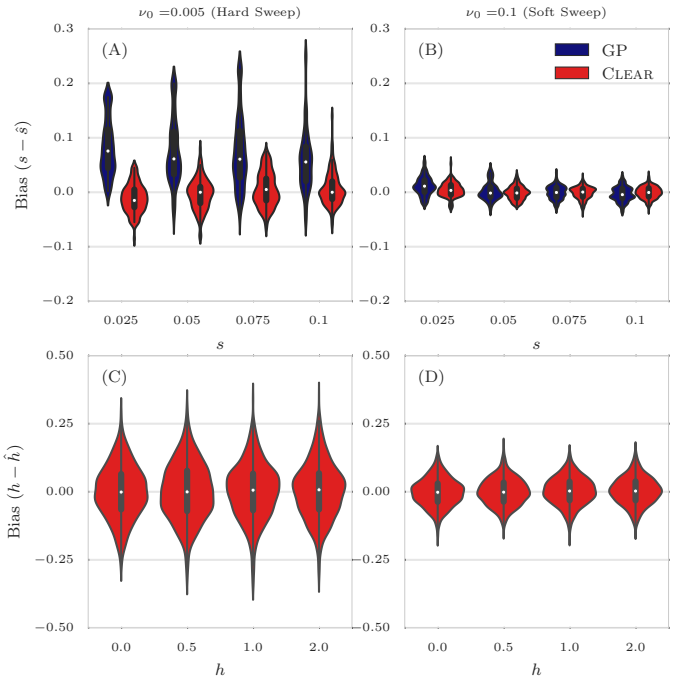

Fig S5: **Distribution of bias for 300× coverage.**
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = \infty$ (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 Table S2. Panels C,D show the variance in the estimation of h.

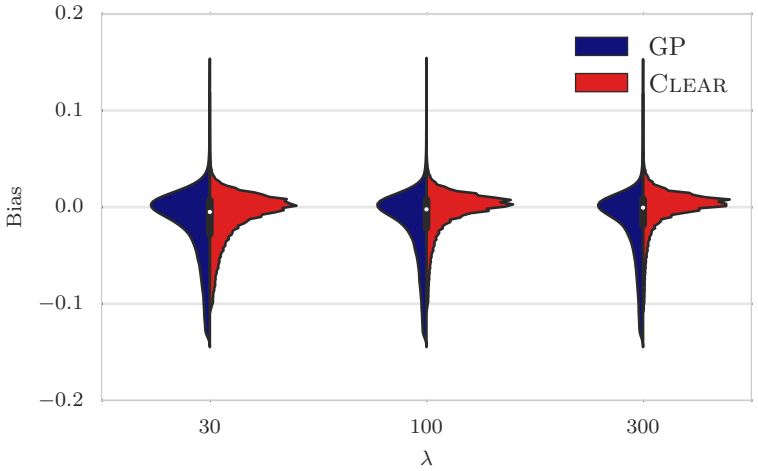

Fig S6: Distribution of bias for null simulations with coverage $\lambda \in \{30, 100, 300\}$.

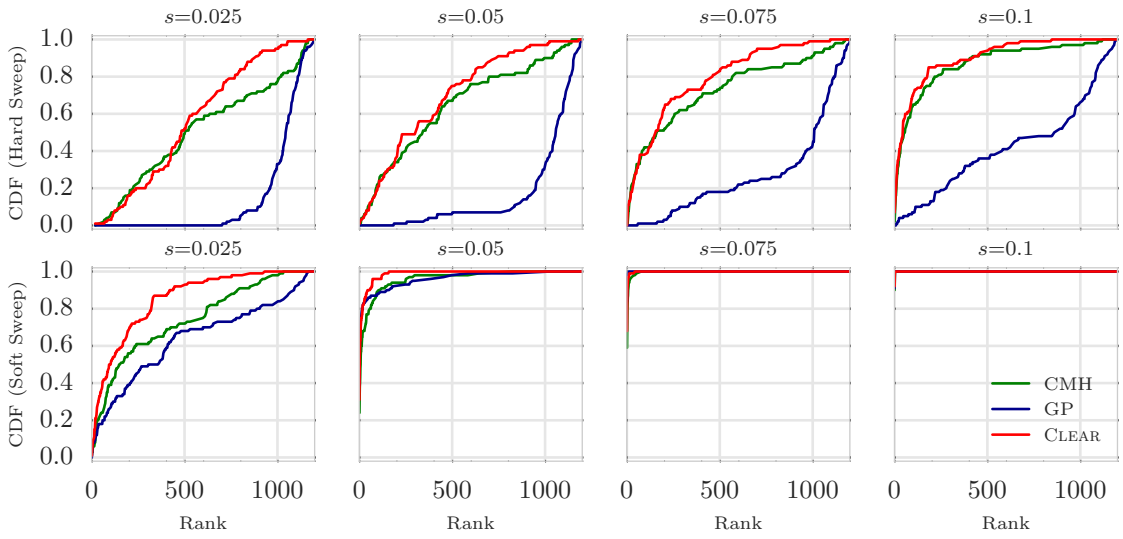

Fig S7: **Ranking performance for 30× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H score), Gaussian Process (GP), and Cochran Mantel Haenszel (CMH), for different values of selection coefficient s and initial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

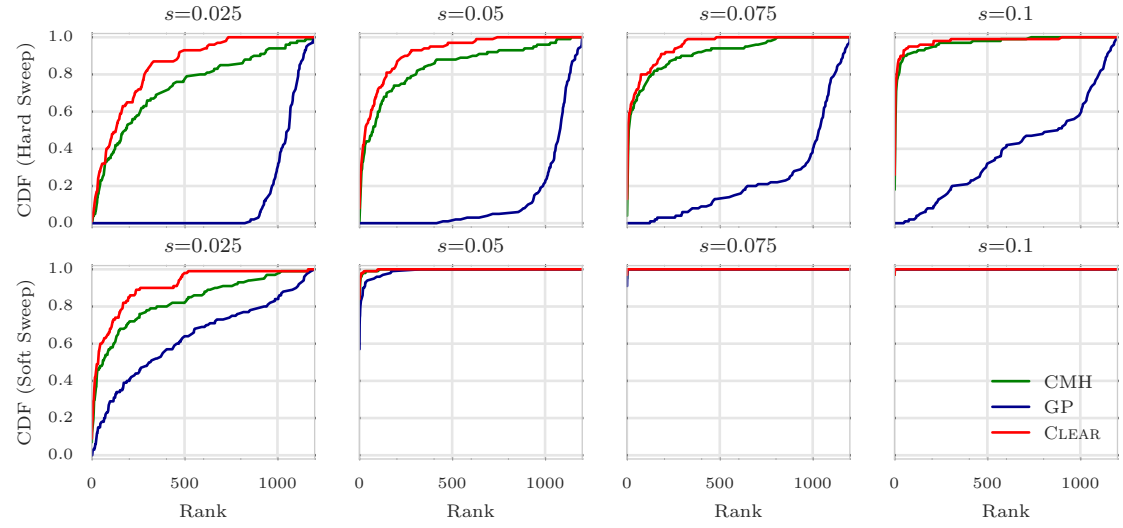

Fig S8: **Ranking performance for 300× coverage.**
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H score), Gaussian Process (GP), and Cochran Mantel Haenszel (CMH), for different values of selection coefficient s and initial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

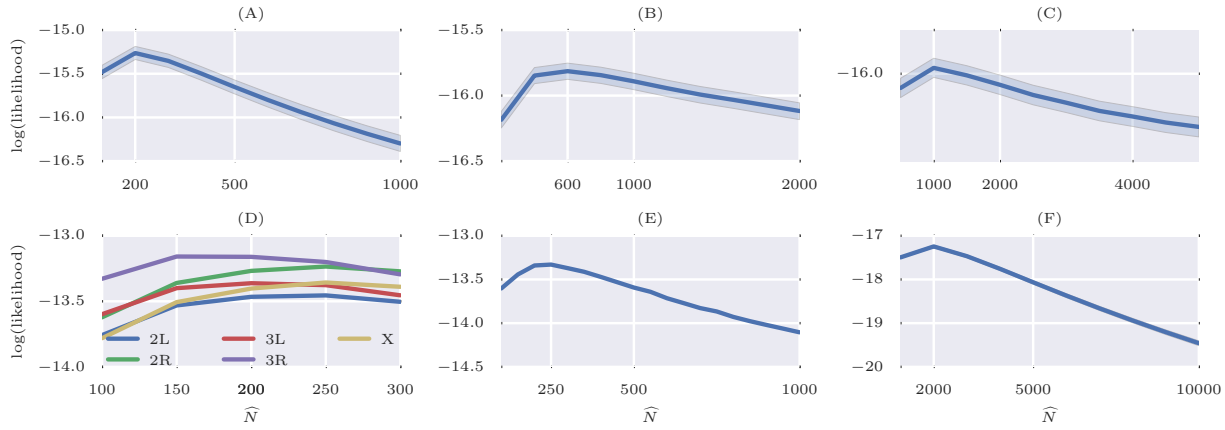

Fig S9: **Maximum likelihood Estimates of N.** Mean and 95% confidence interval of likelihoods of N on simulated data with $N = 200$ (A), $N = 600$(B), and $N = 1000$ individuals, over 1000 simulations. Chromosome-wise (D) and genome-wide (E) likelihood of population size for data from a study of *D. melanogaster* adaptation to alternating temperatures. Likelihood of the Chromosome 3R is attained at 150, while genome-wide maximum likelihood estimate for population size is 250. (F) Likelihood of the population size with respect to all the variants in the yeast dataset. Despite large census population size ($10^6 - 10^7$ [14]), this dataset exhibits much smaller effective population size ($\widehat{N} = 2000$).

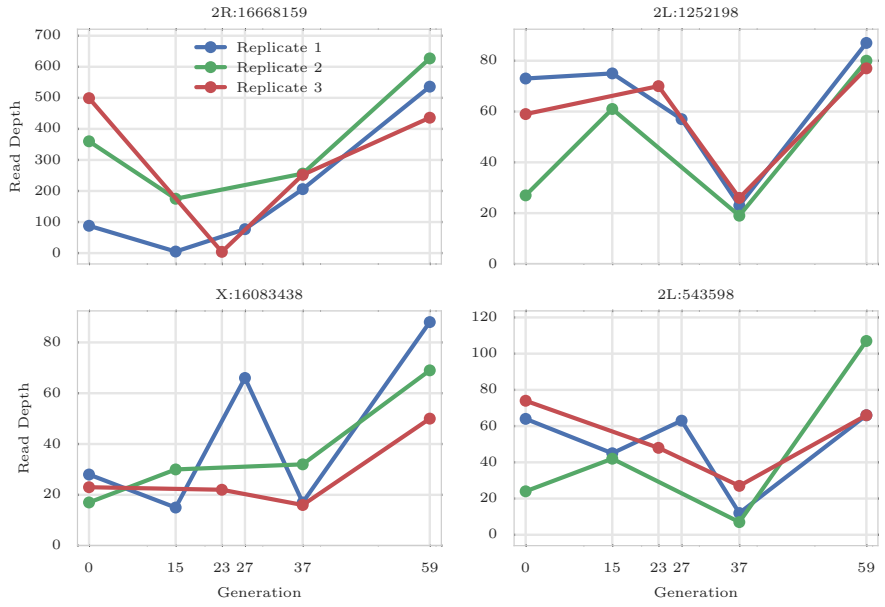

Fig S10: **Coverage heterogeneity in time series data.**
Each panel shows the read depth for 3 replicates of the data from a study of *D. melanogaster* adaptation to alternating temperatures data (see section 3.1). Heterogeneity in depth of coverage is seen between replicates, and also at different time points, in all 4 variants. None of these sites pass the the hard filtering with minimum depth of 30.

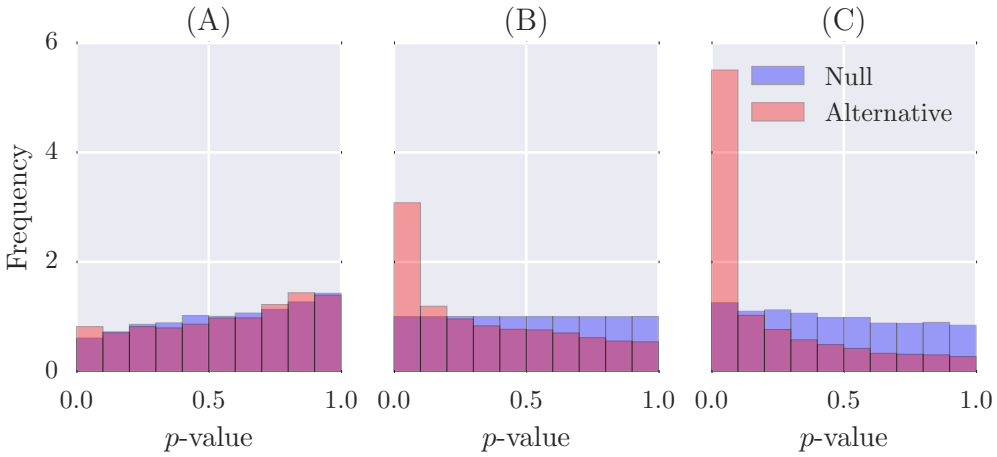

Fig S11: **Distribution of p-values.** Distribution of p-values of CLEAR in null simulations and experimental data when $N = 250$. Panel (A),(C) shows the effect of under estimations ($\widehat{N} = 100$) and over-estimation ($\widehat{N} = 500$) of population size in computing p-values, and panel (B) shows the distribution of $p-$values when unbiased estimate is used to create simulations. .

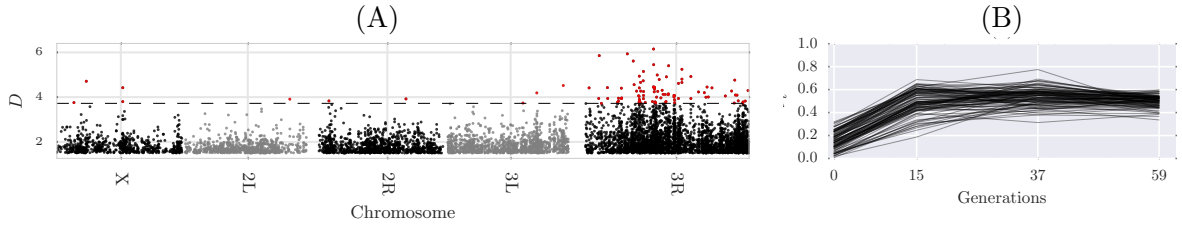

Fig S12: **Single locus analysis of the data from a study of *D. melanogaster* adaptation to alternating temperatures.**
Manhattan plot of scan for testing dominant selection (A). Significant variants with FDR ≤ 0.01 are denoted in red, and their trajectories are depicted in panel (B).

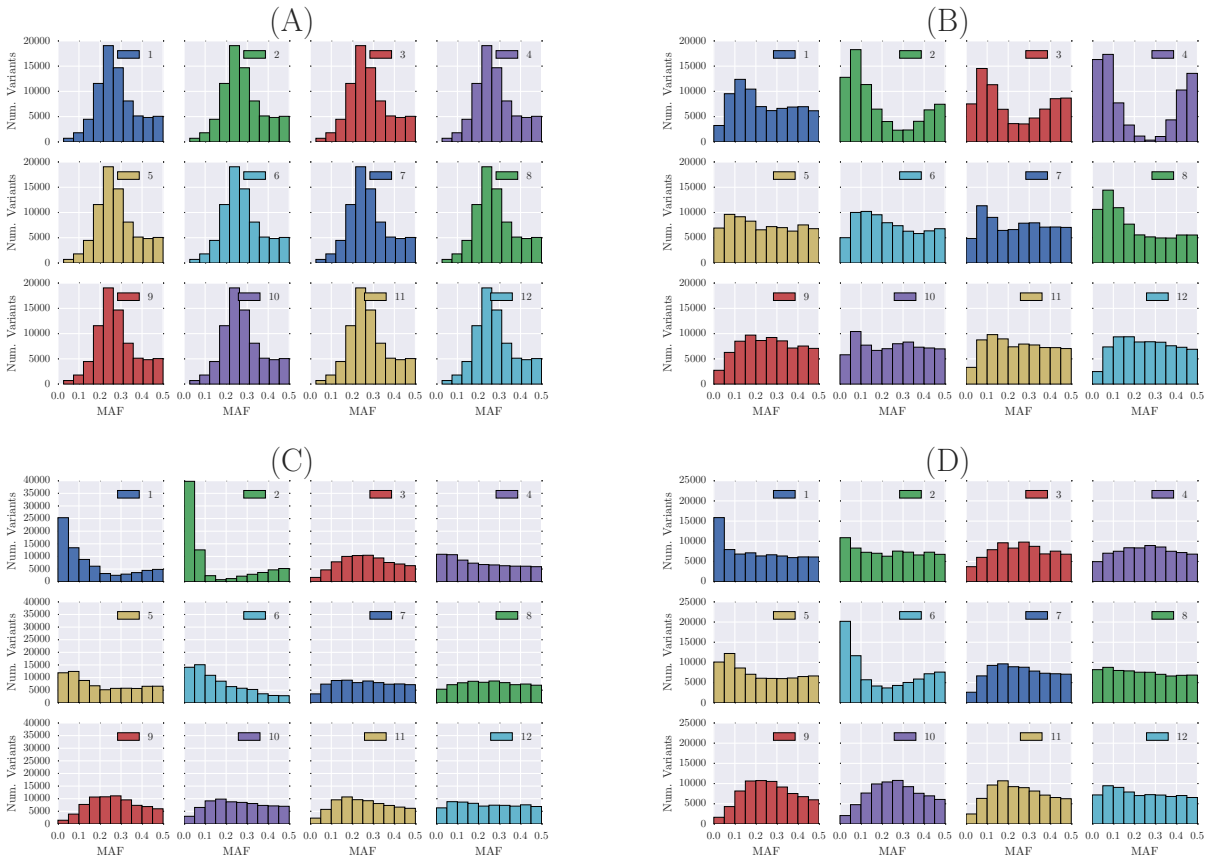

Fig S13: **Site frequency spectrum of the Yeast dataset.** Whole-genome site frequency spectrum of the Yeast dataset at generations 0 (A), 180 (B), 360 (C) and 540 (D). Some replicates, e.g. replicate 2, undergoing severe demographic events.

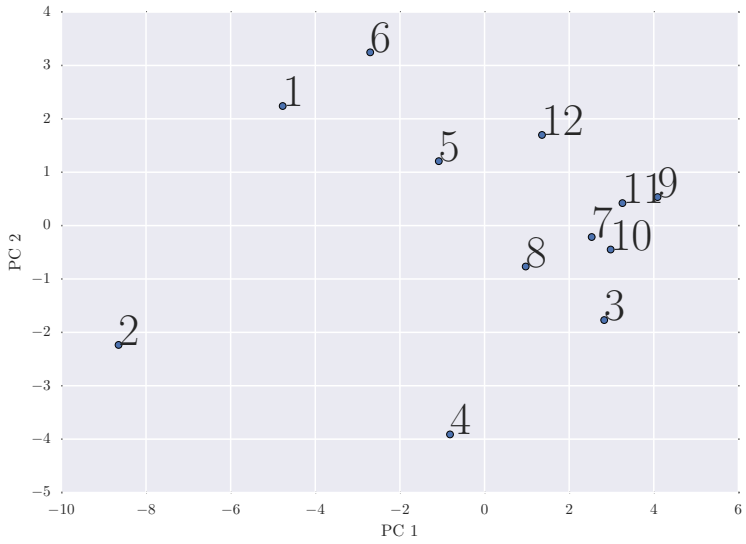

Fig S14: **Population similarity.** Principle component analysis of the 12 replicates throughout the experiment, showing that some populations exhibiting distinct frequency spectra.

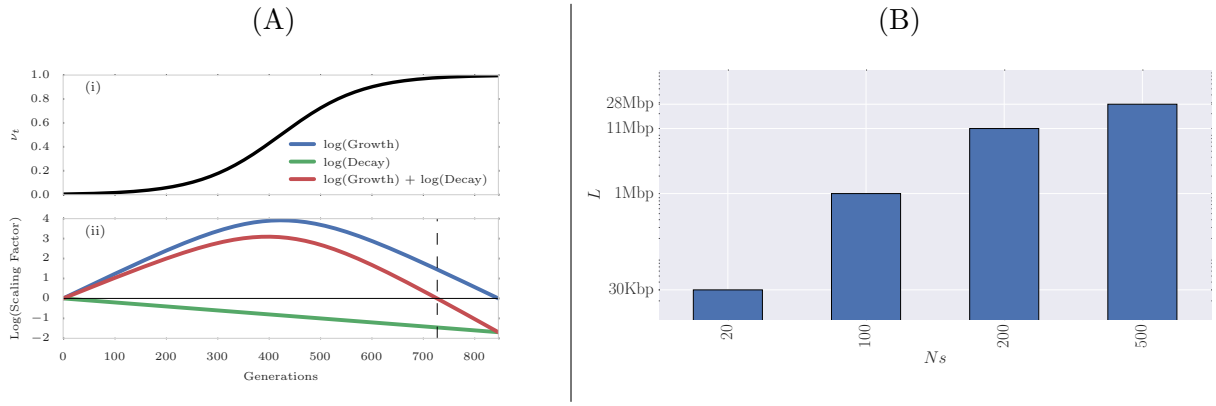

Fig S15: **Choosing window size for** CLEAR **statistic.** (A) Expected dynamics of LD between favored allele ($s = 0.025$) and a variant 50Kbp away, with initial frequency $\nu_0 = 0.01$. (A-i) depicts the dynamic of the favored allele during the selective sweep. (A-ii) illustrates interaction of the growth and decay factors introduced in Eq. S1, with the red line describing overall effect of selection and recombination on LD. The vertical dashed line points to the time when the LD value starts to decrease below original LD. (B) Alternatively, we can fix time, and find the window-size at which LD decays below the original LD (Eq. S3). The plot shows the window size as a function of Ns (20,100,200,500), after fixing other model parameters to match *D. melanogaster* E&R experiments ($N = 250$, $r = 2 \times 10^8, \tau = 59$).

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

Hard Sweep

λ	Method	Avg Power
300	CLEAR	34
300	CLEAR($L=1$)	21
300	CMH	12
300	FIT	2
300	GP	0
100	CLEAR	31
100	CLEAR($L=1$)	10
100	CMH	4
100	FIT	2
100	GP	0
30	CLEAR	20
30	CLEAR($L=1$)	3
30	FIT	2
30	CMH	0
30	GP	0

Soft Sweep

λ	Method	Avg Power
300	CLEAR	69
300	CMH	69
300	CLEAR($L=1$)	68
300	GP	61
300	FIT	8
100	CLEAR	67
100	CMH	60
100	CLEAR($L=1$)	60
100	GP	59
100	FIT	1
30	CLEAR	57
30	GP	53
30	CMH	39
30	CLEAR($L=1$)	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 ν_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.

Table S2: **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	ν_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

Table S3: **Overlapping genes with the 174 candidate variants.**

Interval	Position	FBgn	Gene Name	GO Function
I1	X:1.567-1.824M	FBgn0023531	CG32809	NA
		FBgn0023130	a6	embryonic development via the syncytial blastoderm
		FBgn0025378	CG3795	serine-type endopeptidase activity
		FBgn0025391	Scgdelta	heart contraction, mesoderm development
		FBgn0261548	CG42666	NA
		FBgn0026086	Adar	RNA editing
		FBgn0026090	CG14812	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process
		FBgn0023522	CG11596	NA
I2	X:7.175-7.241M	FBgn0029941	CG1677	NA
		FBgn0029944	Dok	stress activated protein kinase signaling
		FBgn0029946	CG15034	NA
I3	2L:16.878-16.993M	FBgn0052832	CG32832	mitochondrial pyruvate transport
		FBgn0032618	CG31743	sulfotransferase activity
		FBgn0085342	CG34313	NA
		FBgn0040985	CG6115	NA
		FBgn0261671	tweek	synaptic vesicle endocytosis
		FBgn0026150	ApepP	metalloaminopeptidase activity
		FBgn0262355	CR43053	NA
		FBgn0053179	beat-IIIb	NA
I4	2R:2.725-2.810M	FBgn0040674	CG9445	NA
		FBgn0265935	coro	adult somatic muscle development
		FBgn0033110	CG9447	NA
		FBgn0033113	Spn42Dc	Inhibitory Serpins
		FBgn0028988	Spn42Dd	Inhibitory Serpins
		FBgn0033115	Spn42De	Inhibitory Serpins
		FBgn0050158	CG30158	small GTPase mediated signal transduction
I5	3L:14.362-14.514M	FBgn0036421	CG13481	ubiquitin-protein transferase activity
		FBgn0262580	CG43120	NA
		FBgn0036422	CG3868	NA
		FBgn0087007	bbg	PDZ domain
		FBgn0036426	CG9592	NA
		FBgn0036427	CG4613	serine-type endopeptidase activity

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

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

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

[21] 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.

[22] 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.

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

[24] 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.

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

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

[36] Toshio Kosaka and Kazuo Ikeda. Reversible Blockage of Membrane Retrieval and Endocytosis in the Garland Cell of the Temperature-sensitive. *The Journal of cell biology*, 97, 1983.

[37] 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.

[38] 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.

[39] Sasha F Levy, Jamie R Blundell, Sandeep Venkataram, Dmitri A Petrov, Daniel S Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, 2015.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

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

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

[51] 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.

[52] 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.

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

[54] 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.

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

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

[61] 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.

[62] 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.

[63] 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.

[64] 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.

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

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

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

[68] 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.

[69] 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.

[70] 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.

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

[72] 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.