CLEAR: Composition of Likelihoods for Evolve And Resequence Experiments

Arya Iranmehr[1*], Ali Akbari[1], Christian Schlötterer[2], 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

* airanmehr@gmail.com (AI); vbafna@ucsd.edu (VB)

Abstract

Experimental evolution (EE) studies are powerful tools for observing molecular evolution "in-action" from populations sampled in controlled and natural environments. The advent of next generation sequencing technologies has made whole-genome and whole-population sampling possible, even for eukaryotic organisms with large genomes, and allowed us to locate the genes and variants responsible for genetic adaptation. While many computational tests have been developed for detecting regions under selection, they are mainly designed for static (single time) data, and work best when the favored allele is close to fixation. Conversely, EE studies provide samples over multiple time points, often at early stages of selective sweep.

While being more predictive than static data analysis, majority of the EE studies are constrained by the limited time span since onset of selection, depending upon the generation time of the organism. This constraint curbs the power adaptation studies, as the population can only be evolve-and-resequenced for a small number of generations relative to the fixation-time of the favored allele. Moreover, coverage in pool-sequenced experiments varies across replicates and time points for every variant. As a result, ascertainment bias in measuring population frequency is heterogeneous across different measurements of a single allele.

In this article, we directly address these issues while developing tools for identifying selective sweep in pool-sequenced EE of sexual organisms and propose Composition of Likelihoods for Evolve-And-Resequence experiments (CLEAR). Extensive simulations show that CLEAR achieves higher power in detecting and localizing selection over a wide range of parameters. In contrast to existing methods, the CLEAR statistic is robust to variation of coverage. CLEAR also provides robust estimates of model parameters, including selection strength and overdominance, as byproduct of the statistical testing, while being orders of magnitude faster. Finally, we applied the CLEAR statistic to data from a study of *D. melanogaster* adaptation to alternating temperatures, and discovered enrichment of genes related to "response to heat", "cold acclimation" and "defense response to bacterium" ontologies, many of which express Heat Shock Proteins (HSPs).

Author Summary

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 [1], including for example, classic examples like lactose tolerance in Northern Europeans [2], human adaptation to high altitudes [3, 4], but also drug resistance in pests [5], HIV [6], cancer [7, 8], malarial parasite [9, 10], and other antibiotic resistance [11]. 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 [12–18] or natural [19–25] 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 re-sequence 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 EE results, these studies are increasingly being used to test a wide range of hypotheses [26] and have been shown to be more predictive than static data analysis [27–29]. In particular, longitudinal EE data is being used to estimate model parameters including population size [30–35], strength of selection [34, 36–41], allele age [40] recombination rate [34], mutation rate [34, 42], quantitative trait loci [43] and for tests of neutrality hypotheses [25, 34, 44, 45].

While objectives, designs and organisms of EE studies can be entirely different [42, 46], 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 [45, 47] (600 generations), courtship song [48] (100 generations), hypoxia tolerance [49] (200 generations), adaptation to new laboratory environments [14, 50] (59 generations), egg size [51] (40 generations), C virus resistance [52] (20 generations), and dark-fly [53] (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 [54] 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 [55–58] shift in allele-frequencies, prevalence of long haplotypes [54, 59], population differentiation [45, 60] in multiple-population data and others. Many existing methods use the Site Frequency Spectrum (SFS, see Suppl. Fig. S1) to identify departure from neutrality. Classical examples including Tajima's D [55], Fay and Wu's H [56], Composite Likelihood Ratio [61], were all shown to be weighted linear combination of the SFS values [62]. While successful, these methods are prone to both, false negatives [63], as also false-discoveries due to confounding factors such as demography, including bottleneck and population expansions, and ascertainment bias [63–67]. Nevertheless, SFS based tests continue to be used successfully, often in combination with other tests [54, 66]. 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.* [49] used the ratio of the estimated population size of case and control populations to compute test statistic for each genomic region. Burke *et al.* [45] applied Fisher exact test to the last observation of data on case and control populations. Orozco-terWengel *et al.* [14] used the Cochran-Mantel-Haenszel (CMH) test [68] to detect SNPs whose read counts change consistently across all replicates of two time-point data. Turner *et al.* [48] proposed the diffStat statistic to test whether the change in allele frequencies of two population deviates from the distribution of change in allele frequencies of two drifting populations. Bergland *et al.* [25] 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.* [51] computed test statistic of generalized linear-mixed model directly from read counts. Bollback *et al.* [38] provided diffusion approximation to the continues Wright Fisher Markov process and estimated s numerically and performed standard likelihood ratio test under $\mathcal{X}^2$ distribution.

It is only recently that direct tests for analyzing time-series data have been developed, and much of it is based on whole-genome sequencing of pools of individuals (pool-seq) at specific times. Using continuous-time continuous-state Brownian motion process, Feder *et al.* [44] proposed the Frequency Increment Test (FIT). More recently, Topa *et al.* [69] proposed a Gaussian Process (GP) for modeling single-locus time-series pool-seq data. Terhorst *et al.* [34] extended GP to compute joint likelihood of multiple loci under null and alternative hypotheses.

A key contribution of our paper is the development of a direct, and significantly faster method, CLEAR, for identifying selection in short-term experimental evolution with pool-seq data. 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 orders of magnitude faster.

2 Materials and Methods

Notation. Consider a locus with starting derived allele frequency ν_0. Frequencies are sampled at $T+1$ distinct generations specified by $\mathcal{T} = \{\tau_i : 0 \leq \tau_0 < \tau_1, \ldots < \tau_T\}$, and denoted by $\boldsymbol{\nu} = \{\nu_0, \ldots, \nu_T\}$. Moreover, R replicate measurements are made, and we denote the r-th replicate frequency data as $\boldsymbol{\nu}^{(r)}$.

2.1 The CLEAR statistic

To test if a genomic region is evolving under natural selection, we consider a likelihood-based approach [34, 54, 61, 69] that (a) maximizes the likelihood of the time series data w.r.t. selection and overdominance parameters s, h, for each polymorphism in the region; and, (b) computes the log-odds ratio of the likelihood of selection model to the likelihood of neutral evolution/drift model, for every polymorphism in the region. Subsequently, (c) site likelihood ratios in a genomic region are combined to compute the CLEAR statistic, which is a composite likelihood score for the region being under selection. In addition to detecting selection, the CLEAR statistic can be used to rank variants for site-identification, and it provides maximum likelihood estimates of selection parameters.

Likelihood for Neutral Model. To model neutral evolution, it is natural to model the change in frequency ν_t over time via Brownian motion [44] or Gaussian process [34, 69]. Significant deviations from this Null could be indicative of non-neutrality. However, in our experiments, we found that the Brownian motion

approximation is inadequate for small populations, low starting frequencies and sparse sampling (in time) that are typical in experimental evolution (see Results, and Fig. 2). In fact, other "continuous" models such as Gaussian process for dynamic allele frequencies, are also susceptible to this issue (see Results and Fig. 5A-C).

Instead, by computing likelihood of data using a discrete-time discrete-state-space Wright-Fisher Markov chain, we turn the problem of small-population size into an advantage. Consider a neutrally evolving diploid population with N individuals. Define a $2N \times 2N$ transition matrix P, where $P^{(\tau)}[i,j]$ denotes probability of change in allele frequency from $\frac{i}{2N}$ to $\frac{j}{2N}$ in τ generations, solely due to genetic drift. P is defined as follows [70]:

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

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

Note that $P^{(\tau)}$ needs to be computed only once and can be reused for all the variants in the genome. Also, precomputing and storing $P^{(\tau)}$ is tractable and numerically stable for controlled experimental evolution experiments with $N \leq 5000$.

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 ν_t denotes the frequency of the site at time $\tau_t \in \mathcal{T}$. Then, ν_{t^+}, the frequency at time $\tau_t + 1$ (one generation ahead), can be estimated using:

$$\begin{aligned} \hat{\nu}_{t^+} = \mathbb{E}[\nu_{t^+}|s,h,\nu_t] &= \frac{w_{11}\nu_t^2 + w_{01}\nu_t(1-\nu_t)}{w_{11}\nu_t^2 + 2w_{01}\nu_t(1-\nu_t) + w_{00}(1-\nu_t)^2} \\ &= \nu_t + \frac{s(h+(1-2h)\nu_t)\nu_t(1-\nu_t)}{1+s\nu_t(2h+(1-2h)\nu_t))}. \end{aligned} \quad (3)$$

For finite populations, let $Q_{s,h}^{(\tau)}[i,j]$ denote the probability of transition from $\frac{i}{2N}$ to $\frac{j}{2N}$ in τ generations. We model Q as follows (See [70], Pg. 24, Eqn. 1.58-1.59):

$$Q_{s,h}^{(1)}[i,j] = \Pr\left(\nu_{t^+} = \frac{j}{2N} \middle| \nu_t = \frac{i}{2N}; s, h\right) = \binom{2N}{j} \hat{\nu}_{t^+}^j (1-\hat{\nu}_{t^+})^{2N-j} \quad (4)$$

$$Q_{s,h}^{(\tau)} = Q_{s,h}^{(\tau-1)} Q_{s,h}^{(1)} \quad (5)$$

For $s = 0$, Eq. 4 and 5 are identical to Eq. 1 and 2, respectively. The likelihood of observing the trajectory $\boldsymbol{\nu}$ is computed using:

$$\mathcal{L}_M(s,h|\boldsymbol{\nu}) = \Pr(\boldsymbol{\nu}; s, h) = \prod_{t=1}^{T} \Pr(\nu_t|\nu_{t-1}; s, h) = \prod_{t=1}^{T} Q_{s,h}^{(\delta_t)}[\hat{i},\hat{j}], \quad (6)$$

where, $(\hat{i},\hat{j}) = (2N\nu_{t-1}, 2N\nu_t)$, and $\delta_t = \tau_t - \tau_{t-1}$. Combining the likelihood over independent replicate samples $\boldsymbol{\nu}^{(r)}$, we get:

$$\mathcal{L}_M(s,h|\{\boldsymbol{\nu}^{(r)}\}) = \prod_r \mathcal{L}_M(s,h|\boldsymbol{\nu}^{(r)}). \quad (7)$$

Let $\hat{s}, \hat{h}$ denote the parameters that maximize the likelihood. The simplest form of the modified likelihood ratio test statistic for each variant is given by

$$M = \text{sgn}(\hat{s}).\log\left(\frac{\mathcal{L}_M(\hat{s},\hat{h}|\{\boldsymbol{\nu}^{(r)}\})}{\mathcal{L}_M(0,0|\{\boldsymbol{\nu}^{(r)}\})}\right). \quad (8)$$

Accounting for Heterogeneous Ascertainment Bias. In the discussion so far, we assumed that the exact allele frequencies are supplied. However, in most cases, allele frequencies are estimated from genotype data with small sample size or pool-seq data [71] (Suppl. Fig. S3,S2). In pool-seq datasets, the sequencing depth at a site varies for different replicates, and different time samples (Suppl. Fig. S4) and filtering low coverage variants can potentially discard useful information. For instance, in analyzing data from a study of *D. melanogaster* adaptation to alternating temperatures [14], by setting minimum read depth at a site to be 50, allowing the site to be retained only if the depth for all time points, and all replicates exceeded 50, the number of sites to be analyzed would drop from from 1,733,121 to 11,232 (Suppl. Fig. S5). To account for this heterogeneity, we extend the Markov chain likelihood in Eq. 7 to a Hidden Markov Model (HMM) for pool-seq data.

Consider a variant position being sampled at time point $\tau_t \in \mathcal{T}$. We denote the pool-seq data for that variant as $x_t = \langle c_t, d_t \rangle$ where d_t, c_t represent the read depth, and the read count of the derived allele, respectively, at time τ_t. The time-series data is represented by the sequence $\boldsymbol{x} = \{x_0, x_1, x_2, \ldots, x_T\}$. We model the dynamic pool-seq data using a an HMM with $2N+1$ states which state i $(0 \leq i \leq 2N)$ corresponds to allele frequency $i/2N$. Also, we note that HMM is stationary model in that transition and emission distributions do not change over time and defining these distributions completely specifies the HMM model. Eqs. 1, 2 define transition probabilities. The probability that state i emits $x = \langle d, c \rangle$ is given by

$$\mathbf{e}_i(x) = \binom{d}{c} \left(\frac{i}{2N}\right)^c \left(1 - \frac{i}{2N}\right)^{d-c} .$$

For $1 \leq t \leq T$, let $\alpha_{t,i}$ denote the probability of emitting $x_1, x_2, \ldots, x_t$ and reaching state i at τ_t. Then, $\alpha_{t,i}$ can be computed using the forward-procedure [72]:

$$\alpha_{t,i} = \left(\sum_{1 \leq j \leq 2N} \alpha_{t-1,j} \; Q_{s,h}^{(\delta_t)}[j,i] \right) \mathbf{e}_i(x_t) \quad . \tag{9}$$

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

$$\mathcal{L}_H(s,h|\{\boldsymbol{x}^{(r)}\}) = \prod_{r=1}^{R} \mathcal{L}_H(s,h|\boldsymbol{x}^{(r)}) = \prod_{r=1}^{R} \sum_i \alpha_{T,i}^{(r)} \quad . \tag{10}$$

Similar to Eq. 8, let $\hat{s}, \hat{h}$ denote the parameters that maximize likelihood. The modified likelihood ratio statistic for each variant of pool-seq data is given by

$$H \quad = \quad \mathrm{sgn}(\hat{s}).\log\left(\frac{\mathcal{L}_H(\hat{s},\hat{h}|\{\boldsymbol{x}^{(r)}\})}{\mathcal{L}_H(0,0|\{\boldsymbol{x}^{(r)}\})}\right). \tag{11}$$

Composite Likelihood Ratio. In general, the favored allele can be in linkage disequilibrium with some of its surrounding variation. The linked-loci hitchhike and share similar dynamics with the favored allele. Some models such as multi-locus Gaussian process [34] incorporate these associations by modeling linkage and recombination explicitly. However, these approaches are computationally expensive. Moreover, linkage computations are infeasible without haplotype resolved data, which pool-seq does not provide. Instead, we work with a simpler Composite Likelihood Ratio (CLR) [61, 73] computation to combine the individual scores of all variants into a composite score.

Consider a genomic region L to be a collection of segregating sites with little or no recombination between sites and the favored allele. This scenario holds when the starting frequency of the favored allele is not high and the region is small. Let M_ℓ (respectively, H_ℓ) denote the likelihood ratio score based on Markov chain (respectively, HMM) for each site ℓ in L. The classical CLR is computed by averaging scores of all the variants withing the testing region. However, the levels of LD between favored allele and its surrounding variation depend on initial frequency ν_0, strength of selection s and time since the onset of selection (see Appendix 6.5 for more details). Hence, we parameterize CLR to discard those polymorphisms that are in low LD with the favored allele, from computation of CLR. In particular, we can choose to only include sites whose likelihood ratio score is above a certain threshold. For percentile cut-off π, let $L_\pi \subseteq L$ denote the set of sites whose likelihood ratio scores had percentile π or better. For all π, the modified CLR statistic for Markov chain and HMM is computed using:

$$\mathcal{M}_\pi = \frac{1}{|L_\pi|}\sum_{\ell \in L_\pi} M_\ell, \qquad \mathcal{H}_\pi = \frac{1}{|L_\pi|}\sum_{\ell \in L_\pi} H_\ell. \tag{12}$$

In combining the sites, recall that sites that are not on the same lineage as the favored site, will reduce in frequency and have negative modified likelihood ratio values, but are still informative about selection. Therefore, we also define:

$$\mathcal{M}^+_\pi = \frac{1}{|L_\pi|}\sum_{\ell \in L_\pi} |M_\ell|, \qquad \mathcal{H}^+_\pi = \frac{1}{|L_\pi|}\sum_{\ell \in L_\pi} |H_\ell|. \tag{13}$$

We also note here that unlike $\mathcal{H}$, $\mathcal{H}^+_\pi$ can be computed without knowledge of the ancestral allele.

Final Word on Notation. In the following, we use M, Eq. 8 (respectively, H, Eq. 11) to denote the CLEAR scores for individual variants. Similarly, we use $\mathcal{M}_\pi, \mathcal{M}^+_\pi$ (respectively, $\mathcal{H}_\pi, \mathcal{H}^+_\pi$) to denote the composite CLEAR scores for a genomic region. In this notation, $\mathcal{H}_{100} = \mathcal{H}^+_{100}$ corresponds to the CLR based on choosing the best site in a region, while $\mathcal{H}_0, \mathcal{H}^+_0$ both correspond to the CLR based on all sites in the region. For ease of notation, we sometimes use $\mathcal{H}$, or $\mathcal{H}^+$ to denote CLEAR scores in results, when the choice of π is apparent.

Site-identification. Once a genomic region is claasified to be under selection by $\mathcal{H}$ or $\mathcal{M}$ statistic, individual variant scores (M or H) in a region are ranked to predict the favored site. In general, identifying the favored site in pool-seq data is difficult [50], due to extensive span of hitchhikers in an ongoing sweep (see Appendix 6.5 for more detail). In our analysis of the *D. melanogaster* EE data, we identify a set of "candidate" variants whose scores exceed a False Discovery Rate threshold based on the distribution of CLEAR scores on negative controls.

Estimating Parameters. Depending on data (read count or allele frequency) the optimal value of the parameters can be found by

$$\hat{s}, \hat{h} = \arg\max_{s,h} \sum_r^R \log\left(\mathcal{L}_{\mathcal{M}}(s, h|\boldsymbol{\nu}^{(r)})\right), \quad \text{or,} \tag{14}$$

$$\hat{s}, \hat{h} = \arg\max_{s,h} \sum_r^R \log\left(\mathcal{L}_{\mathcal{H}}(s, h|\boldsymbol{x}^{(r)})\right). \tag{15}$$

where likelihoods are defined in Eq. 7 and Eq. 10, respectively. The parameters in Eqs. 14, 15 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. 6). The value of the overdominance parameter can provide extra information regarding dominance of the favored allele and the kind of selection [74] (See Suppl. Fig. S6,S7):

value	condition
$h = 0$	recessive adaptive allele
$h = 0.5$	directional selection
$h = 1$	dominant adaptive allele
$h > 1$	overdominance

Precomputing Transition Matrices. CLEAR requires a one-time computation of matrices $Q^{(\tau)}_{s,h}$ for the entire range of s, h values. Precomputation of 909 transition matrices for $s \in \{-0.5, -0.49, \ldots, 0.5\}$ and $h \in \{0, 0.25, \ldots, 2\}$ took less that 15 minutes ($\approx$ 1 second per matrix) on a desktop computer with a Core i7 CPU and 16GB of RAM.

2.2 Extending Site Frequency Spectrum based tests for time series data

The site frequency spectrum (SFS) is a mainstay of tests of neutrality and selection, and can be computed using pool-seq data (does not need haplotypes). Following Fu, 1995 [75], any linear combination of the site frequencies is an estimate of θ. However, under non-neutral conditions, different linear combinations behave differently. Therefore, many popular tests of neutrality either compute differences of two estimates of θ, or perform cross-population tests comparing the θ estimates in two different populations [57, 62, 76].

We asked if SFS-based tests could be adapted for time-series data. A simple approach is to use cross-population SFS tests on the populations at time 0 (before onset of selection), and at time sample τ_t, for each t. Evans *et al.* [77] developed diffusion equations for evolution of SFS in time series, but they are difficult to solve. Instead, we derive a formula for computing D_t, the dynamic of Tajima's D at generation t. Specifically, for initial value D_0, initial carrier frequency ν_0 and selection coefficient s:

$$D_t = D_0 - \log(1-\nu_t)\frac{W_0}{\log(2N)} - \nu_t^2\Pi_0\ , \tag{16}$$

where W_0 and Π_0 are Watterson's and Tajima's estimates of θ in the initial generation (Appendix 6.2). See Suppl. Fig. S8 for comparison to empirical values from simulations. Similarly, we show (Appendix 6.3), that the dynamics of Fay and Wu's H statistic [56] are directly related to expected value of the of Haplotype Allele Frequency (HAF) score [78], and can be written as a function of ν_t as follows:

$$NH_t = \theta\nu_t\left(\frac{\nu_t+1}{2} - \frac{1}{(1-\nu_t)N+1}\right) + \theta(1-\nu_t)\left(\frac{N+1}{2N} - \frac{1}{(1-\nu_t)N+1}\right) \tag{17}$$

In both cases, ν_t itself can be written as a function of s, t (Suppl. Eq. S2). This allows us to compute likelihood functions $\mathcal{L}_S(s;\{D_t\})$ or $\mathcal{L}_S(s;\{H_t\})$. Then, a likelihood ratio, similar to Eqns. 8, 11 provides a statistic for detecting selection in each window.

However, as ν_0 and D_0 are often unknown in sampling from natural populations, we do not directly use Eqns. 16, 17. Instead, we heuristically aggregate statistics throughout time to compute a time-series score (Appendix 6.4).

2.3 Simulations

We performed extensive simulations using parameters that have been used for *D. melanogaster* experimental evolution [79]. See also Fig. 1 for illustration. To implement real world pool-seq experimental evolution, we conducted simulations as follows:

I. **Creating initial founder line haplotypes.** Using `msms` [80], 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 be chosen merely 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` [81] 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 chosen 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(λ), where $\lambda \in \{30, 100, 300, \infty\}$ 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. Infinite depth experiments refer to the case, where the true allele frequency is provided and Markov and HMM likelihood computations give identical results.

We also conducted simulations to evaluate performance for cases involving evolution and re-sequencing of natural populations. Sampling from natural populations differs from (controlled) experimental evolution in some important ways. First, the time of onset of selection may not be known. Second, as the start of sampling can be any generation during selective sweep, mutations that arose after the onset of selection appear in data can have a nontrivial effect on SFS, specifically if sampling is started long after onset of selection. While SFS based techniques expected to perform poorly in controlled EEs, their power is highest near fixation. To simulate these scenarios, `msms` was used to forward-simulate a population with $N_e = 10^4$, $\nu_0 = 10^{-4}$, and to record SFS of a 50Kbp region (Fig. 1A). The remaining parameters were identical to controlled experimental evolution simulations.

False Discovery Rate (FDR). We applied CLEAR to the data from a study of *D. melanogaster* adaptation to alternating temperatures [14], by computing the $\mathcal{H}^+$ statistic for sliding windows of 30Kbp with steps of 10Kbp over the entire genome. Due to the great variation in the density of polymorphic sites, we observed a large variation in scores. Moreover, high LD between proximal sites resulted in many neighboring windows with similar scores. Therefore, we computed a local false discovery rate for any

candidate window by choosing an encompassing genomic region of 2Mbp to preserve the genomic background. For each candidate window, we sampled a subset of variants from the encompassing region $10,000$ times, and computed the $\mathcal{H}^+$ score. The candidate window was selected if its score was among the top 1% of scores in the permuted tests.

3 Results

Modeling Allele Frequency Trajectories in Finite Populations. We first tested the goodness of fit of the discrete-time Markov chain versus continuous-time Brownian motion (Gaussian approximation) in modeling allele frequency trajectories in finite populations, under different sampling schemes and starting frequencies. 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 expected distribution of the frequency of the next sample ν_τ under two models and compared them with empirical distributions calculated from simulated data. Fig. 2A-F shows that Brownian motion 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 Suppl. Fig. S1). Moreover, in typical *D. melanogaster* experiments for example, sampling is sparse. Often, the experiment is designed so that $10 \leq \tau \leq 100$ [14, 49, 79].

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. 5) 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 ≤ 0.05 (Suppl. Fig. S9). 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.

Before comparing against other methods, we first evaluated the use of Clear with different percentile-cutoffs π (Eq. 12) in computing composite statistics of a region. For each configuration, we computed average Power for $s \in \{0.025, 0.05, 0.075, 0.1\}$, using $\mathcal{H}_\pi, \mathcal{H}_\pi^+$. We computed the optimal value of π using a line-search. Fig. 3 reveals several important trade-off between π, initial frequency, and coverage.

- $\mathcal{H}_\pi^+$ consistently achieves a high power for $\pi = 0$, and in the absence of knowledge of the selection regime or the ancestral allele, $\mathcal{H}_0^+$ is a powerful statistic to use.
- In every scenario tested, $\max_\pi\{\text{Power}(\mathcal{H}_\pi)\} \geq \max_\pi\{\text{Power}(\mathcal{H}_\pi^+)\}$, suggesting that it is beneficial to make predictions based on $\mathcal{H}_\pi$, if the selection regime is well-understood and ancestral allele is known.
- In soft sweep, relative to hard sweep, it helps to choose a higher value of the cut-off π. This is consistent with the fact that LD between the favored site and other sites is generally lower for soft sweep. For instance, in soft sweep with infinite coverage (Fig. 3F), optimum is gained at $\pi = 100$, equivalent to considering the score of the highest scoring site as $\mathcal{H}$ statistic of the region.
- When coverage is low (Fig. 3A,D), it helps to accumulate evidence from multiple sites, and the best results are achieved for lower values of π.

In the following tests, we fix the value of $\pi = 70, 97, 99$ for $\lambda = 30, 100, 300$, respectively, and simplify notation by denoting the optimum score as $\mathcal{H}$. For example, when $\lambda = 30$, $\mathcal{H}$ corresponds to $\mathcal{H}_{70}$. In addition, we use $\mathcal{H}^+$ to denote $\mathcal{H}_0^+$. We also compared $\mathcal{M}$ (Markov likelihoods) versus $\mathcal{H}$. As shown in Fig. 4, $\mathcal{H}$ has better power for low coverage ($\lambda = 30$) compared to $\mathcal{M}$ decays.

Finally, we compared the power of CLEAR with Gaussian process (GP) [34], FIT [44], and CMH [68] statistics. As CMH only takes read count data, here we used $\lambda = 300$ to implement infinite coverage scenario. All methods other than CLEAR and CMH 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. 5 and Table S1). In particular, the difference in performance of CLEAR with other methods is pronounced when starting frequency is low. Starting frequency is at its minimum in selection on an allele from *de novo* mutations and is likely to be low if selection is on an allele from standing variation (Suppl. Fig. S1). 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, taking into account of 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. 5B,C) suggest that taking many samples with lower coverage is preferable to sparse sampling with higher coverage.

SFS for Detection in Natural Samples. We did not show the SFS based statistics in Fig. 5 as they did not perform better than random. In majority of controlled experimental evolution studies, the population is restricted set of F founder lines, where $F << N_e$ (Fig. 1B) and inbred during the experiment. This creates a severe bottleneck, confounding SFS. Suppl. Fig. S11 demonstrates the effect of experimental evolution on different SFS statistics under neutral evolution for 1000 simulations. A second problem with using SFS for experimental evolution is that the sampling starts right after the onset of experimentally induced selection, and the favored allele may not reach high enough frequency to modify the site frequency spectrum (Suppl. Fig. S10).

However, in experiments involving naturally occurring populations, even if the span of the time-series is small, the onset of selection might occur many generations prior to sampling. To test performance of SFS-based statistics in natural evolution, we conducted 200 (100 neutral and 100 sweep) forward simulations for different values of s, λ using $N_e = 10K$ and accumulating new mutations. The start of sampling was done at a randomly picked time subsequent to the onset of selection in two distinct scenarios. Let $t_{\nu=x}(s, N_e)$ denote the expected time (in generations) required to reach carrier frequency x in a hard sweep and $U[a, b]$ denote discrete uniform distribution in the interval $[a, b]$. First we considered the case when start of sampling is chosen throughout the whole sweep. i.e., $\tau_1 \sim U\left[1, t_{\nu=1}(s, N_e)\right]$ (Fig. 7A). Next, we considered sampling start time chosen nearer to fixation of the favored allele, i.e., $\tau_1 \sim U\left[t_{\nu=0.9}(s, N_e), t_{\nu=1}(s, N_e)\right]$ (Fig. 7B). In both scenarios, sampling was done over 5 time points within 50 generations of τ_1. We compared $\mathcal{H}$, GP, FIT with both static and dynamic SFS based statistics of SFSelect and Tajima's D. Fig. 7A shows that SFS based statistics are outperformed by single locus and CLR methods. However, when sampling is performed close to fixation, i.e., when the favored allele has frequency of 0.9 or higher, SFS based statistics perform significantly better than GP, FIT and $\mathcal{H}$ (Fig. 7b). Moreover, dynamic SFS statistics outperform static SFS statistics, demonstrating that in these regimes, SFS based statistics could be used to detect selection.

Site-identification. Localizing the favored site is a nontrivial task. We used the simple approach of ranking each site in a region detected as being under selection. The sites were ranked according to the likelihood ratio scores (Eqns. 8, 11). 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. 8) shows that CLEAR outperforms other statistics. We also compared each method to see how often it ranked the favored site in as the top ranked site (Table 1A-B), among the top 10 ranked sites (Table 1C-D), and among the top 50 (Table 1E-F) ranked sites. In the 1150 variants tested, CLEAR performed consistently better than other methods in all of these measures.

An interesting observation is the contrast between site-identification and detection. When selection coefficient is high, detection is easier (Fig. 5A-F), but site-identification is harder due to the high LD between hitchhiking sites and the favored allele (Table 1A-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. 5B-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 (Table 1A-B). Our results are consistent with previous studies [50, 82]. See Appendix 6.5 for a detailed explanation.

Hard Sweep

(A)

s	CMH	FIT	GP	CLEAR
0.025	3	0	0	2
0.05	5	0	0	5
0.075	11	0	0	10
0.1	15	0	0	14

(C)

s	CMH	FIT	GP	CLEAR
0.025	21	3	0	15
0.05	39	2	0	28
0.075	57	4	0	49
0.1	78	4	0	71

(E)

s	CMH	FIT	GP	CLEAR
0.025	51	19	0	43
0.05	70	18	0	66
0.075	85	20	2	81
0.1	94	26	6	93

Soft Sweep

(B)

s	CMH	FIT	GP	CLEAR
0.025	9	1	4	11
0.05	53	10	51	63
0.075	87	31	92	91
0.1	93	54	95	95

(D)

s	CMH	FIT	GP	CLEAR
0.025	34	8	27	44
0.05	86	41	86	92
0.075	100	81	100	100
0.1	100	97	100	100

(F)

s	CMH	FIT	GP	CLEAR
0.025	60	22	62	75
0.05	96	69	97	99
0.075	100	97	100	100
0.1	100	100	100	100

Table 1. Percentage of simulations which appears in top of ranking. Percentage of simulations in which the favored allele is ranked first (A-B); appears in top 10 (C-D); or, appears in top 50 (E-F). In soft sweep simulations (B,D,F), the ranks are consistently better than hard sweep simulations (A,C,E). This can be attributed to lower LD between the hitchhikers (false positives) and favored allele in soft sweep scenarios.

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 over dominance $(h - \hat{h})$ for of CLEAR and GP in each setting. The distribution of the error (bias) for 100× coverage is presented in Fig. 9 for different configurations. Suppl. Fig. S12, S13 provide the distribution of estimation errors for 30×, and infinite coverage, respectively. For hard sweep, CLEAR provides estimates of s with lower

variance of bias (Fig.9A). In soft sweep, GP and CLEAR both provide unbiased estimates with low variance (Fig. 9B). Fig. 9C-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 $\mathcal{O}(TR)$ and $\mathcal{O}(TRN^2)$ computation for $\mathcal{M}$, and $\mathcal{H}$, respectively. However, most of the operations are can be vectorized for all replicates to make the effective running time for each variant, $\mathcal{O}(T)$ and $\mathcal{O}(TN)$, respectively. We conducted 1000 simulations and measured running times for computing site statistics M, H, FIT, CMH and GP with different number of linked-loci. Our analysis reveals (Fig. 6) 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 [14], 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 Fig. S2 of [34]) which makes variant depths highly heterogeneous (Suppl. Figs. S5, S4). We computed the $\mathcal{H}^+$ statistic for sliding windows of 30Kbp with steps of 10Kbp over the whole genome. After filtering for heterochromatin[83, 84], and applying a local false discovery rate ≤ 0.01 (Methods), we identified 89 intervals (Fig. 10) containing 968 genes (Suppl. Table S6).

We found 11 GO Biological Process terms to be enriched with Fisher exact P-value $\leq 10^{-3}$ (Table 2,S5), including many Heat Shock Proteins (HSPs) involved in 'cold acclimation' (Hsp22, 23, 26), and 'response to heat' (e.g., Adar, Syn). The effect of rapid changes in temperature is profound, and it is not surprising that the enriched GO terms include many generic stress and stress response genes. As longer genes contain more variants, the probability of a false variant being selected could increase with the length of the gene. Although the CLEAR statistic for genes does not favor longer genes, we also performed a single variant based GO enrichment using `Gowinda` [85]. The analysis identified 34 enriched GO terms (Suppl. Table S4) associated with Biological Process. 5 of 11 GO terms in the gene level analysis overlapped with the 34 `Gowinda` terms (Fisher exact p-val: 10^{-8}) suggesting consistency of single variant based and gene-based analysis.

Rank	GO ID	GO Term	-log(p-value)	Hits	Num of Genes
1	GO:0042742	defense response to bacterium	5.0	15	62
2	GO:0009408	response to heat	4.8	16	71
3	GO:0006719	juvenile hormone catabolic process	4.5	3	4
4	GO:0008363	larval chitin-based cuticle development	4.5	6	14
5	GO:0045664	regulation of neuron differentiation	4.5	3	4
6	GO:0051291	protein heterooligomerization	4.5	3	4
7	GO:0061077	chaperone-mediated protein folding	4.4	4	7
8	GO:0009631	cold acclimation	4.0	4	8
9	GO:0030837	negative regulation of actin filament polymerization	3.4	3	6
10	GO:0042026	protein refolding	3.1	4	11
11	GO:0007552	metamorphosis	3.0	8	36

Table 2. GO (Biological Process) enrichment.
A Fisher exact test was performed for GO enrichment in genes located in selected regions. All GO terms that contained at least 3 selected genes, and had Fisher exact p-value $\leq 10^{-3}$, are listed above.

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.

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 EE 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 [79], although there are some notable exceptions [49]. 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 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 ($F \approx 10^2$) for Experimental Evolution, and the evolution includes a severe population bottleneck. This bottleneck confounds SFS-based statistics and makes it difficult to fit a model or test a hypothesis (Suppl. Fig. S11). 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 and experiments with 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 Availability. The source code and running scripts for CLEAR are publicly available at

https://github.com/bafnalab/clear.

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.

5 Figures

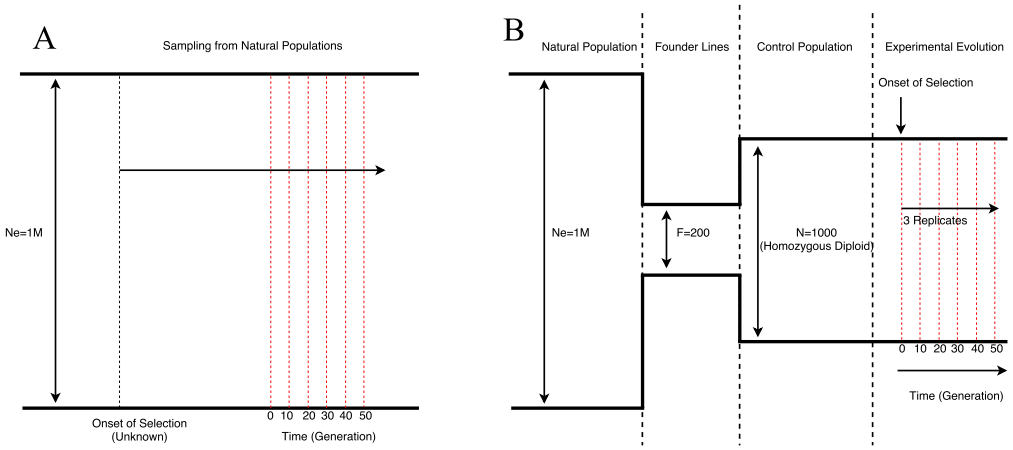

Fig 1. Two settings for collecting genomic time series data.
Different settings in which dynamic data is collected are depicted with typical parameters for *D. melanogaster*. In both settings, 6 samples (vertical red dashed lines) are taken every 10 generation. When sampling from naturally evolving populations (A), the time of onset of selection is unknown, and population size is larger. For (controlled) experimental evolution (B), founder lines are first sampled from a natural population to create a homogeneous population. Then, multiple replicates of this population are evolved and sampled over time.

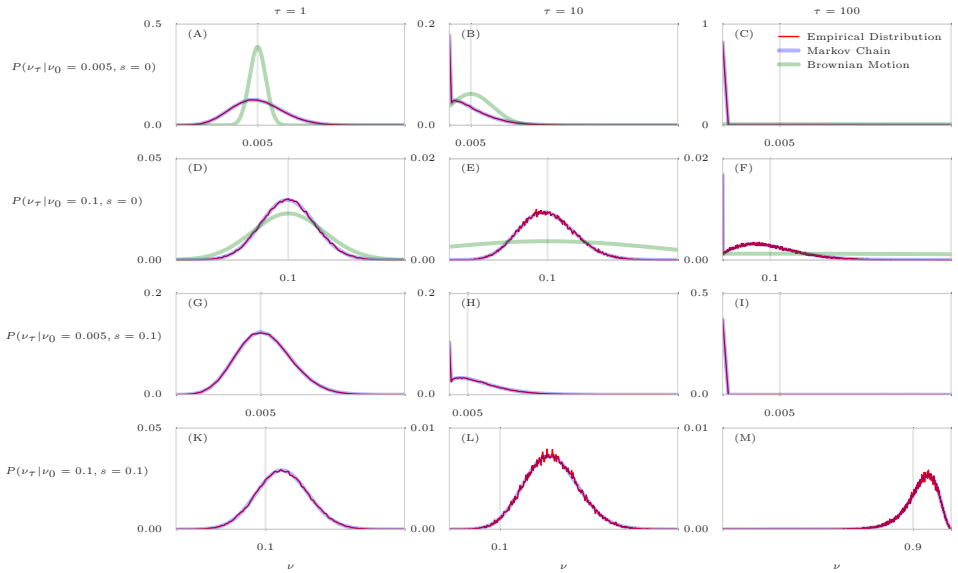

Fig 2. Comparison of empirical distributions of allele frequencies (red) versus predictions from Brownian Motion (green), and Markov chain (blue).
Comparison of empirical and theoretical distributions under neutral evolution (panels A-F) and selection (panels G-M) with different starting frequencies $\nu_0 \in \{0.005, 0.1\}$ 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.

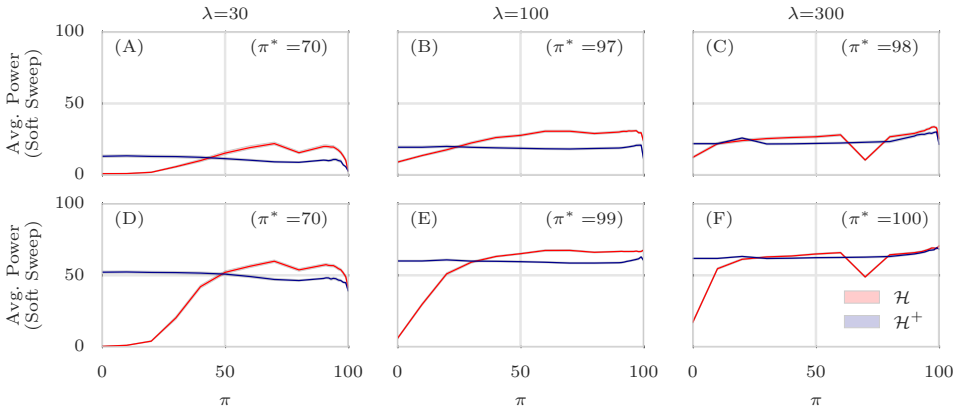

Fig 3. Average power of CLEAR statistics as function of percentile cutoffs.
Detection power is averaged for $s \in \{0.025, 0.05, 0.075, 0.1\}$ for CLEAR statistics, $\mathcal{H}_\pi$ and $\mathcal{H}^+_\pi$. We denote $\pi^* = \arg\max_\pi\{\text{Power}(\mathcal{H}), \text{Power}(\mathcal{H}^+)\}$. Average power was computed using 2000 simulations for each choice of s. The appropriate choice of π can be used to improve performance for different coverage values. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

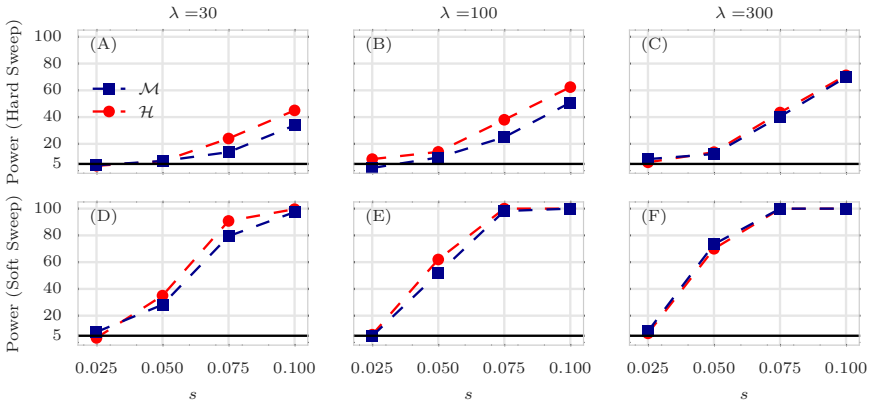

Fig 4. Comparison of power of $\mathcal{M}$ and $\mathcal{H}$.
Detection power for $\mathcal{M}$ and $\mathcal{H}$ under hard (top) and soft sweep (bottom) scenarios, for different settings of mean coverage λ and selection strength s. The y-axis measures power – sensitivity with false positive rate FPR ≤ 0.05 – for 2000 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 $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

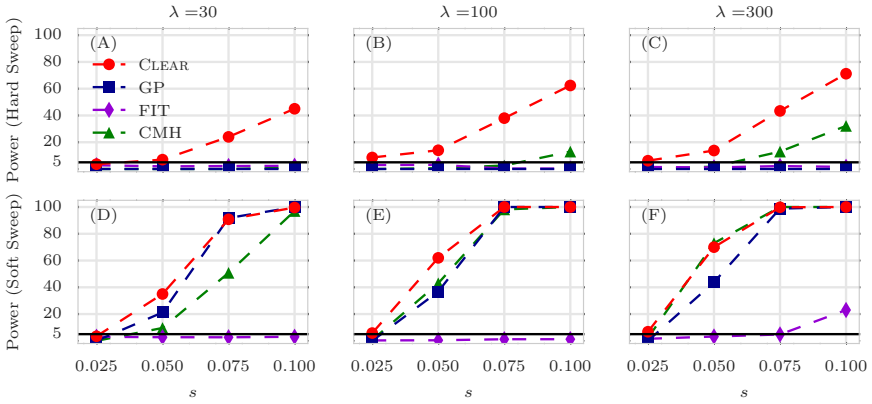

Fig 5. Power calculations for detection of selection.
Detection power for CLEAR($\mathcal{H}$), Frequency Increment Test (FIT), Gaussian Process (GP), and CMH under hard (top) and soft sweep (bottom) scenarios. λ, s denote the mean coverage and selection coefficient, respectively. The y-axis measures power – sensitivity with false positive rate FPR ≤ 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 $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

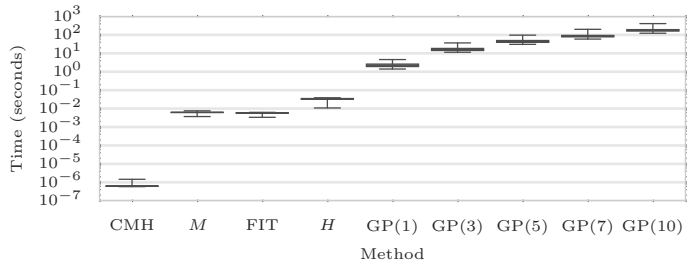

Fig 6. Running time.
Box plots of running time per variant (CPU-secs.) of CLEAR($\mathcal{M}$, $\mathcal{H}$), CMH, FIT, and GP with single, 3, 5, 7, and 10 loci over 1000 simulations conducted on a workstation with 4th Generation Intel Core i7 processor. The average running time for each method is shown on the x-axis. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

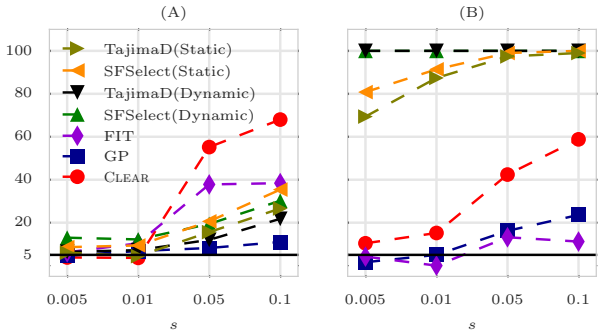

Fig 7. Power of SFS based statistics.
Power of detecting selection for Frequency Increment Test (FIT), Gaussian Process (GP), CLEAR($\mathcal{H}$) on hard sweep natural experimental evolution with $N_e = 10^4$ and depth $\lambda = \infty$. The measurements are conducted for a range of selection coefficients, s. Each point represents the mean of 200 simulations. For each simulation, sampling starts at a randomly chosen time, and subsequently 5 replicate samples are acquired every 10 generations. (A) Start of sampling is chosen randomly throughout the sweep $\tau_1 \sim U\left[1, t_{\nu=1}(s, N_e)\right]$, where $t_{\nu=x}(s, N_e)$ denotes is the expected time to reach carrier frequency x in a hard sweep and $U[a, b]$ is discrete uniform distribution. (B) The start of sampling is chosen near fixation of the favored allele, i.e. $\tau_1 \sim U\left[t_{\nu=0.9}(s, N_e), t_{\nu=1}(s, N_e)\right]$.

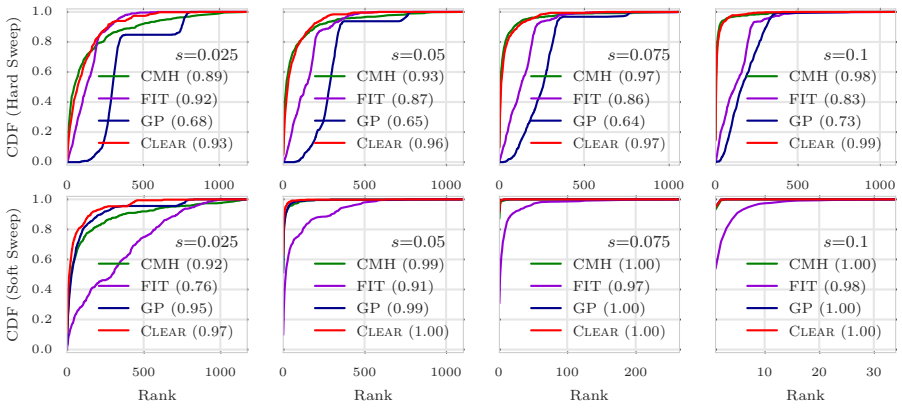

Fig 8. Ranking performance for 100× coverage.
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H), Gaussian Process (GP), CMH, and Frequency Increment Test (FIT), for different values of selection coefficient s and initial carrier frequency. Note that the individual variant CLEAR score (H) is used to rank variants. The Area Under Curve (AUC) is computed as a quantitative measure to compare the performance of methods for each configuration. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

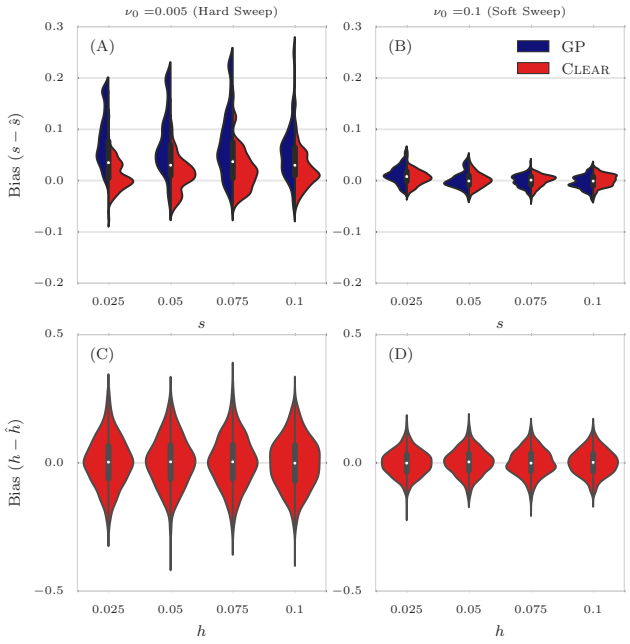

Fig 9. Distribution of bias for 100× coverage.
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = 100$ (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 S3. Panels C,D show the variance in the estimation of h. In all simulations, 3 replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

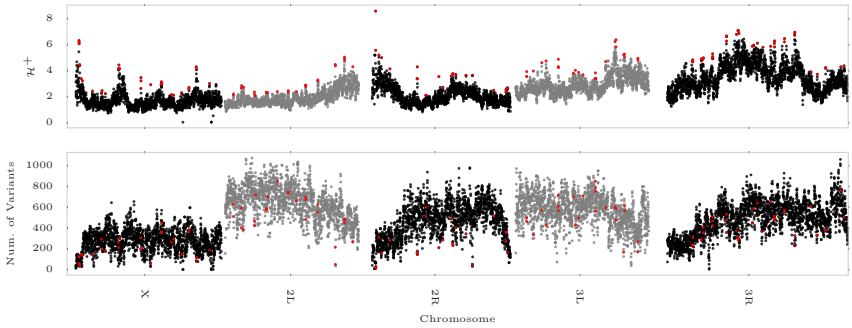

Fig 10. CLEAR **scan of the data from a study of *D. melanogaster* adaptation to alternating temperatures.**
Manhattan plot of the CLEAR ($\mathcal{H}^+$) statistic (A) and the number of SNPs (B) in 30Kbp sliding windows with steps of 10Kbp, excluding windows close to centromere and telomere. The dashed line corresponds to the top 1%-ile threshold. Regions that exceed the threshold are shown in red dots. For comparison to past results, only Chromosomes 2, 3, and X are shown.

References

1. Fan S, Hansen MEB, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. Science. 2016;354(6308):54–59. doi:10.1126/science.aaf5098.

2. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. The American Journal of Human Genetics. 2004;74(6):1111–1120.

3. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329(5987):75–78.

4. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, et al. Genetic evidence for high-altitude adaptation in Tibet. Science. 2010;329(5987):72–75.

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

6. Feder AF, Rhee SY, Holmes SP, Shafer RW, Petrov DA, Pennings PS. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. eLife. 2016;5. doi:10.7554/eLife.10670.

7. Gottesman MM. Mechanisms of cancer drug resistance. Annual review of medicine. 2002;53(1):615–627.

8. Zahreddine H, Borden KL. Mechanisms and insights into drug resistance in cancer. Front Pharmacol. 2013;4(28.10):3389.

9. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature. 2014;505(7481):50–55.

10. Nair S, Nash D, Sudimack D, Jaidee A, Barends M, Uhlemann AC, et al. Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. Molecular Biology and Evolution. 2007;24(2):562–573.

11. Spellberg B, Guidos R, Gilbert D, Bradley J, Boucher HW, Scheld WM, et al. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clinical Infectious Diseases. 2008;46(2):155–164.

12. Hegreness M, Shoresh N, Hartl D, Kishony R. An equivalence principle for the incorporation of favorable mutations in asexual populations. Science. 2006;311(5767):1615–1617.

13. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 2013;500(7464):571–574.

14. Orozco-ter Wengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Molecular ecology. 2012;21(20):4931–4941.

15. Lang GI, Botstein D, Desai MM. Genetic variation and the fate of beneficial mutations in asexual populations. Genetics. 2011;188(3):647–661.

16. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 2009;461(7268):1243–1247.

17. Bollback JP, Huelsenbeck JP. Clonal interference is alleviated by high mutation rates in large populations. Molecular biology and evolution. 2007;24(6):1397–1406.

18. Oz T, Guvenek A, Yildiz S, Karaboga E, Tamer YT, Mumcuyan N, et al. Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. Molecular biology and evolution. 2014; p. msu191.

19. Maldarelli F, Kearney M, Palmer S, Stephens R, Mican J, Polis MA, et al. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. Journal of virology. 2013;87(18):10313–10323.

20. Reid BJ, Kostadinov R, Maley CC. New strategies in Barrett's esophagus: integrating clonal evolutionary theory with clinical management. Clinical Cancer Research. 2011;17(11):3512–3519.

21. Denef VJ, Banfield JF. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. Science. 2012;336(6080):462–466.

22. Winters MA, Lloyd Jr RM, Shafer RW, Kozal MJ, Miller MD, Holodniy M. Development of elvitegravir resistance and linkage of integrase inhibitor mutations with protease and reverse transcriptase resistance mutations. PloS one. 2012;7(7):e40514.

23. Daniels R, Chang HH, Séne PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. PLoS One. 2013;8(4):e60780.

24. Barrett RDH, Rogers SM, Schluter D. Natural selection on a major armor gene in threespine stickleback. Science. 2008;322(5899):255–257.

25. Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS Genet. 2014;10(11):e1004775.

26. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. Experimental evolution. Trends in ecology & evolution. 2012;27(10):547–560.

27. Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, et al. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 2008;4(5):e1000083.

28. Desai MM, Plotkin JB. The polymorphism frequency spectrum of finitely many sites under selection. Genetics. 2008;180(4):2175–2191.

29. Sawyer SA, Hartl DL. Population genetics of polymorphism and divergence. Genetics. 1992;132(4):1161–1176.

30. Williamson EG, Slatkin M. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics. 1999;152(2):755–761.

31. Wang J. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Genetical research. 2001;78(03):243–257.

32. Pollak E. A new method for estimating the effective population size from allele frequency changes. Genetics. 1983;104(3):531–548.

33. Waples RS. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics. 1989;121(2):379–391.

34. Terhorst J, Schlötterer C, Song YS. Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution. PLoS Genet. 2015;11(4):e1005069.

35. Jónás A, Taus T, Kosiol C, Schlötterer C, Futschik A. Estimating the Effective Population Size from Temporal Allele Frequency Changes in Experimental Evolution. Genetics. 2016;.

36. Mathieson I, McVean G. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. Genetics. 2013;193(3):973–984.

37. Illingworth CJR, Mustonen V. Distinguishing driver and passenger mutations in an evolutionary history categorized by interference. Genetics. 2011;189(3):989–1000.

38. Bollback JP, York TL, Nielsen R. Estimation of 2Nes from temporal allele frequency data. Genetics. 2008;179(1):497–502.

39. Illingworth CJR, Parts L, Schiffels S, Liti G, Mustonen V. Quantifying selection acting on a complex trait using allele frequency time series data. Molecular biology and evolution. 2012;29(4):1187–1197.

40. Malaspinas AS, Malaspinas O, Evans SN, Slatkin M. Estimating allele age and selection coefficient from time-serial data. Genetics. 2012;192(2):599–607.

41. Steinrücken M, Bhaskar A, Song YS. A novel spectral method for inferring general diploid selection from time series genetic data. The annals of applied statistics. 2014;8(4):2203.

42. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nature Reviews Genetics. 2013;14(12):827–839.

43. Baldwin-Brown JG, Long AD, Thornton KR. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. Molecular biology and evolution. 2014; p. msu048.

44. Feder AF, Kryazhimskiy S, Plotkin JB. Identifying signatures of selection in genetic time series. Genetics. 2014;196(2):509–522.

45. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature. 2010;467(7315):587–590.

46. Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity. 2015;114(5):431–440.

47. Remolina SC, Chang PL, Leips J, Nuzhdin SV, Hughes KA. Genomic basis of aging and life-history evolution in Drosophila melanogaster. Evolution. 2012;66(11):3390–3403.

48. Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. PLoS Genet. 2011;7(3):e1001336.

49. Zhou D, Udpa N, Gersten M, Visk DW, Bashir A, Xue J, et al. Experimental selection of hypoxia-tolerant Drosophila melanogaster. Proceedings of the National Academy of Sciences. 2011;108(6):2349–2354.

50. Tobler R, Franssen SU, Kofler R, Orozco-terWengel P, Nolte V, Hermisson J, et al. Massive habitat-specific genomic response in D. melanogaster populations during experimental evolution in hot and cold environments. Molecular biology and evolution. 2014;31(2):364–375.

51. Jha AR, Miles CM, Lippert NR, Brown CD, White KP, Kreitman M. Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in Drosophila melanogaster. Molecular biology and evolution. 2015;32(10):2616–2632.

52. Martins NE, Faria VG, Nolte V, Schlötterer C, Teixeira L, Sucena É, et al. Host adaptation to viruses relies on few genes with different cross-resistance properties. Proceedings of the National Academy of Sciences. 2014;111(16):5938–5943.

53. Izutsu M, Toyoda A, Fujiyama A, Agata K, Fuse N. Dynamics of Dark-Fly Genome Under Environmental Selections. G3: Genes— Genomes— Genetics. 2015; p. g3—-115.

54. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annual review of genetics. 2013;47:97–120.

55. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–595.

56. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. Genetics. 2000;155(3):1405–1413.

57. Ronen R, Udpa N, Halperin E, Bafna V. Learning natural selection from the site frequency spectrum. Genetics. 2013;195(1):181–193.

58. Garud NR, Messer PW, Buzbas EO, Petrov DA. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS Genet. 2015;11(2):e1005004.

59. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. science. 2006;312(5780):1614–1620.

60. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nature Reviews Genetics. 2009;10(9):639–650.

61. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome research. 2005;15(11):1566–1575.

62. Achaz G. Frequency spectrum neutrality tests: one for all and all for one. Genetics. 2009;183(1):249–258.

63. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. Trends in ecology & evolution. 2013;28(11):659–669.

64. Ptak SE, Przeworski M. Evidence for population growth in humans is confounded by fine-scale population structure. Trends in Genetics. 2002;18(11):559–563.

65. Ramos-Onsins SE, Rozas J. Statistical properties of new neutrality tests against population growth. Molecular biology and evolution. 2002;19(12):2092–2100.

66. Akey JM. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome research. 2009;19(5):711–722.

67. Nielsen R, Signorovitch J. Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. Theoretical population biology. 2003;63(3):245–255.

68. Agresti A, Kateri M. Categorical data analysis. Springer; 2011.

69. Topa H, Jónás Á, Kofler R, Kosiol C, Honkela A. Gaussian process test for high-throughput sequencing time series: application to experimental evolution. Bioinformatics. 2015; p. btv014.

70. Ewens WJ. Mathematical Population Genetics 1: Theoretical Introduction. vol. 27. Springer Science & Business Media; 2012.

71. Lynch M, Bost D, Wilson S, Maruki T, Harrison S. Population-genetic inference from pooled-sequencing data. Genome biology and evolution. 2014;6(5):1210–1218.

72. Durbin R, Eddy SR, Krogh A, Mitchison G. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge university press; 1998.

73. Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. Localizing recent adaptive evolution in the human genome. PLoS Genet. 2007;3(6):e90.

74. Gillespie JH. Population genetics: a concise guide. JHU Press; 2010.

75. Fu YX. Statistical properties of segregating sites. Theoretical population biology. 1995;48(2):172–197.

76. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449(7164):913–918.

77. Evans SN, Shvets Y, Slatkin M. Non-equilibrium theory of the allele frequency spectrum. Theoretical population biology. 2007;71(1):109–119.

78. Ronen R, Tesler G, Akbari A, Zakov S, Rosenberg NA, Bafna V. Predicting Carriers of Ongoing Selective Sweeps Without Knowledge of the Favored Allele. PLoS Genet. 2015;11(9):e1005527.

79. Kofler R, Schlötterer C. A guide for the design of evolve and resequencing studies. Molecular biology and evolution. 2013; p. mst221.

80. Ewing G, Hermisson J. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics. 2010;26(16):2064–2065.

81. Peng B, Kimmel M. simuPOP: a forward-time population genetics simulation environment. Bioinformatics. 2005;21(18):3686–3687.

82. Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, et al. Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nature genetics. 2013;45(8):884–890.

83. Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA. Drosophila melanogaster recombination rate calculator. Gene. 2010;463(1):18–20.

84. Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in Drosophila melanogaster. PLoS Genet. 2012;8(10):e1002905.

85. Kofler R, Schlötterer C. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics. 2012;28(15):2084–2085.

86. Stephan W, Song YS, Langley CH. The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics. 2006;172(4):2647–2663.

6 Appendix

6.1 An approximate logistic function for allele frequency dynamics

Assume that a site is evolving under selection constraints $s, h \in \mathbb{R}$, where s and h denote selection strength and overdominance, respectively. Let ν_t denote the frequency of the site at time $\tau_t \in \mathcal{T}$. Then, ν_{t^+}, the frequency at time $\tau_t + 1$ can be estimated (See Eq. 3) using:

$$\hat{\nu}_{t^+} = \nu_t + \frac{s(h + (1-2h)\nu_t)\nu_t(1-\nu_t)}{1 + s\nu_t(2h + (1-2h)\nu_t))}.$$

We can show that the dynamic of the favored allele can be modeled via a logistic function, in the case of directional selection ($h = 0.5$). Taking derivatives of Eq. 6.1, we have

$$\frac{\mathbf{d}\nu_t}{\mathbf{d}t} = \frac{s\nu_t(1-\nu_t)}{2 + 2s\nu_t} \tag{S1}$$

To, solve the differential equation, note that for small s, $2 + 2s\nu_t \approx 2$. Substituting,

$$\nu_t = \frac{1}{1 + \frac{1-\nu_0}{\nu_0}e^{-st/2}} = \sigma(st/2 + \eta(\nu_0)) \tag{S2}$$

where $\sigma(.)$ is the logistic function and $\eta(.)$ is logit function (inverse of the logistic function).

6.2 Dynamic of Tajima's D

In this part we derive dynamic of Tajima's D statistic in *hard sweep* as function of its value at the onset of selection, D_0, selection strength and the frequency of the favored allele at the onset of selection. Let D_0, Π_0, W_0, be Tajima's D, Tajima's estimate of θ, and Watterson's estimate of θ at time zero and $D_0 = \Pi_0 - W_0$. In order to compute, $D_t = \Pi_t - W_t$ we compute Π_t and W_t separately as follows. Let P be the $n \times n$ matrix of pairwise heterozygosity of individuals, then $\Pi = 1/n^2 \sum P_{ij}$. So, if the population consist of νn identical carrier haplotype (due to lack of recombination), their pairwise hamming distance is zero and should be subtracted from the total Π_t:

$$\Pi_t = (1 - \nu_t^2)\Pi_0 \tag{S3}$$

To compute W_t, first remember that $W_t = \frac{m_t}{S_n}$ where m_t is the number of segregating sites at time t and $S_n = \sum_i^n 1/i \approx \log(n)$. Also we have

$$\frac{W_t}{W_0} = \frac{\frac{m_t}{S}}{\frac{m_0}{S}} \quad \Rightarrow W_t = \frac{m_t}{m_0} W0 \tag{S4}$$

Because of hard sweep and lack of recombination assumption, the population at time t consist of $(1-\nu_t)n$ non-carrier haplotypes and $\nu_t n$ identical carrier haplotypes. While not strictly correct, we assume that the $(1-\nu_t)n + 1$ individuals are evolving neutrally. Using this assumption, we have

$$\frac{m_t}{m_0} = \frac{\log\left((1-\nu_t)n + 1\right)\theta}{\log(n)\theta} \approx \frac{\log\left((1-\nu_t)n\right)}{\log(n)} = \frac{\log(1-\nu_t) + \log(n)}{\log(n)} = 1 + \frac{\log(1-\nu_t)}{\log(n)}. \tag{S5}$$

Finally, by putting Eqs. S3, S4, S5 together, we can explicitly write the dynmics of D statistic as

$$\begin{aligned} D_t &= (1-\nu_t^2)\Pi_0 - (1+\frac{\log(1-\nu_t)}{\log(n)})W_0 \\ &= D_0 - \log(1-\nu_t)\frac{W_0}{\log(n)} - \nu_t^2\Pi_0 \\ &\approx D_0 - \log(1-\sigma(st/2+\eta(\nu_0)))\frac{W_0}{\log(n)} - \sigma(st/2+\eta(\nu_0))^2\Pi_0. \end{aligned} \tag{S6}$$

where σ and η are logistic and logit functions.

6.3 Dynamics of Fay and Wu's H

In any finite population size of n with m segregating sites, allele frequencies take discrete values, i.e., $x_j \in \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\},\ \forall j \in 1, \ldots, m$. We have the following:

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{m} x_j^2 = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^2 \xi_i = \frac{(n-1)}{2n} H,$$

where ξ_i is the number of sites with frequency i/n and H is the Fay & Wu's estimate of θ and $\mathbf{x} \in (0,1)^m$ is the vector of allele frequency of a region with m segregating sites. Recently, Ronen *et al.* [78] devised the 1-HAF statistic for identifying selection on static data, and showed that the expected value of 1-HAF statistic is given by:

$$\mathbb{E}[\text{1-HAF}(t)] = n\|\mathbf{x}_t\|^2 \approx n g(\nu_t) \tag{S7}$$

where

$$g(\nu_t) = \theta\nu_t\left(\frac{\nu_t+1}{2} - \frac{1}{(1-\nu_t)n+1}\right) + \theta(1-\nu_t)\left(\frac{n+1}{2n} - \frac{1}{(1-\nu_t)n+1}\right) \tag{S8}$$

The dynamics of Fay & Wu's estimate are given by

$$H_t = \frac{n-1}{2} g(\nu_t) \tag{S9}$$

6.4 Greedy computation of time-series SFS-based statistics

As discussed in Section 2.2, modeling dynamic of Tajima's D (and Fay&Wu's H) requires knowledge of initial carrier frequency ν_0 and the value of D (and H) statistic at the onset of selection, which are often unknown. As these statistics are monotonically decreasing (or increasing for SFSelect) under no demographic changes, we chose to greedily aggregate statistics throughout time. For example, for Tajima's D, we have

$$\mathcal{D} = \sum_{t \in \mathcal{T}} D_t \tag{S10}$$

where the same procedure applies to Fay&Wu's H and SFSelect.

6.5 Linkage Disequilibrium

Nonrandom associations, Linkage Disequilibrium (LD), between polymorphisms are established in the substitution process, broken by recombination events and reinforced by selection. Although LD can not be measured in pooled sequencing data (phased haplotype data is required), it is still worthwhile to examine the behavior of LD as a

result of the interaction between recombination and natural selection. In this part we theoretically overview expected LD in short EEs.

Let ρ_0 be the LD at time zero between the favored allele and a segregating site l base-pairs away, then under natural selection we have

$$\rho_t = \alpha_t \beta_t \rho_0 = e^{-rtl} \left(\frac{K_t}{K_0} \right) \rho_0 \tag{S11}$$

where $K_t = 2\nu_t(1-\nu_t)$ is the heterozygousity at the selected site, r is the recombination rate/bp/gen. The *decay factor*, $\alpha_t = e^{-rtl}$, and *growth factor*, β_t (see Eq. 30-31 in [86]), are result of recombination and selection, respectively. Fig. S14 presents the expected theoretical value of LD when $\rho_0 = 0.5$ between favored allele (site at position 500K) and the rest of genome, and $\nu_0 = 0.1$. For neutral evolution (top), LD decays exponentially through space and time, while in natural selection (bottom), LD increases and then decreases. Interestingly, LD increases to its maximum value, 1, for the nearby region (the plateau in the Fig. S14 bottom) of the favored allele.

In principle, LD increases after the onset of selection, until $\log(\alpha_t) + \log(\beta_t) > 0$, see Eq. S11. Specifically, log of decay term is linear and, using Eq. S2, we write growth factor in term of initial frequency ν_0 and selection strength s. Fig. S15 depicts interaction of decay and growth factors for weak and strong selection and soft and hard sweeps. In all the case, LD of the favored allele with a segregating site 50Kbp away, increases in the first 50 generations, which give rise to increasing number of *hitchhikers*.

Increase of LD in a large (100Kbp) region is particularly advantageous to the task of identifying the region under selection, if the composite statistics is used. As a result, $\mathcal{H}$ statistic outperforms existing (single-loci) tools in identifying selection. In contrast, augmentation of LD, increases the number of candidates for the favored allele, which makes is difficult to localize the favored allele.

Hard Sweep			Soft Sweep		
λ	Method	Avg Power	λ	Method	Avg Power
300	Clear	34	300	Clear	69
300	CMH	12	300	CMH	69
300	FIT	2	300	GP	61
300	GP	0	300	FIT	8
100	Clear	31	100	Clear	67
100	CMH	4	100	CMH	60
100	FIT	2	100	GP	59
100	GP	0	100	FIT	1
30	Clear	20	30	Clear	57
30	FIT	2	30	GP	53
30	CMH	0	30	CMH	39
30	GP	0	30	FIT	3

Table S1. Average of power for detecting selection.
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.

Method	Avg. Time per Locus
CMH	0.0
M	0.006
FIT	0.006
H	0.033
GP(1)	2.551
GP(3)	19.177
GP(5)	50.291
GP(7)	95.602
GP(10)	202.017

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

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

GO ID	GO Term	-log(p-value)
GO:0001558	regulation of cell growth	4.1
GO:0001700	embryonic development via the syncytial blastoderm	4.1
GO:0003341	cilium movement	4.1
GO:0006030	chitin metabolic process	3.8
GO:0006355	regulation of transcription, DNA-templated	4.1
GO:0006367	transcription initiation from RNA polymerase II promoter	4.1
GO:0006508	proteolysis	4.1
GO:0006719	juvenile hormone catabolic process	4.1
GO:0006839	mitochondrial transport	4.1
GO:0007018	microtubule-based movement	4.1
GO:0007269	neurotransmitter secretion	3.6
GO:0007291	sperm individualization	4.1
GO:0007298	border follicle cell migration	4.1
GO:0007475	apposition of dorsal and ventral imaginal disc-derived wing surfaces	4.1
GO:0007552	metamorphosis	3.8
GO:0007602	phototransduction	4.1
GO:0008104	protein localization	3.1
GO:0008340	determination of adult lifespan	4.1
GO:0008362	chitin-based embryonic cuticle biosynthetic process	4.1
GO:0009312	oligosaccharide biosynthetic process	3.0
GO:0009408	response to heat	4.1
GO:0015991	ATP hydrolysis coupled proton transport	4.1
GO:0016079	synaptic vesicle exocytosis	4.1
GO:0016485	protein processing	4.1
GO:0031146	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	4.1
GO:0035556	intracellular signal transduction	4.1
GO:0042742	defense response to bacterium	3.8
GO:0043066	negative regulation of apoptotic process	3.1
GO:0045494	photoreceptor cell maintenance	4.1
GO:0045664	regulation of neuron differentiation	4.1
GO:0045861	negative regulation of proteolysis	4.1
GO:0048675	axon extension	4.1
GO:0055114	oxidation-reduction process	3.1
GO:0061024	membrane organization	4.1

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

FlyBase ID	GO Term	Gene Name
FBgn0001224	cold acclimation	Hsp23
FBgn0001225	cold acclimation	Hsp26
FBgn0001233	cold acclimation	Hsp83
FBgn0034758	cold acclimation	CG13510
FBgn0001224	response to heat	Hsp23
FBgn0001225	response to heat	Hsp26
FBgn0001233	response to heat	Hsp83
FBgn0001223	response to heat	Hsp22
FBgn0001226	response to heat	Hsp27
FBgn0001227	response to heat	Hsp67Ba
FBgn0001228	response to heat	Hsp67Bb
FBgn0001229	response to heat	Hsp67Bc
FBgn0003301	response to heat	rut
FBgn0004575	response to heat	Syn
FBgn0010303	response to heat	hep
FBgn0019949	response to heat	Cdk9
FBgn0023517	response to heat	Pgam5
FBgn0025455	response to heat	CycT
FBgn0026086	response to heat	Adar
FBgn0035982	response to heat	CG4461

Table S5. Enriched genes of analysis of data from a study of *D. melanogaster* adaptation to alternating temperaturesassociated with GO terms of "cold acclimation" and "response to heat".

Statistic	Value
Num. of Vatiants	1,608,032
Num. of Candidate Intervals	89
Total Num. of Genes	17,293
Num. of Variant Genes	12,834
Num. of Genes within Candidate Intervals	968
Total Num. of GO	6,983
Num. of GO with 3 or More Genes	3,447
Num. of Candidate Variants for Gowinda	2,886

Table S6. General statistic of analysis of data from a study of *D. melanogaster* adaptation to alternating temperatures.

7 Supplemental Figures and Tables

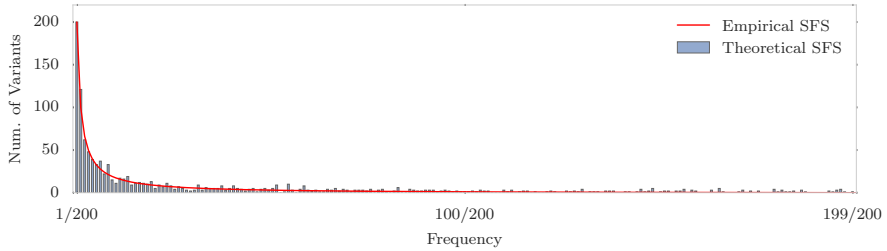

Fig S1. Site Frequency Spectrum.
Theoretical and Empirical SFS in a 50Kbp region for a neutral population of 200 individuals when $N_e = 10^6$ and $\mu = 10^{-9}$. The x-axis corresponds to site frequency, and the y-axis to the number of variants with a specific frequency. In a neural population, majority of the variations stand in low frequency.

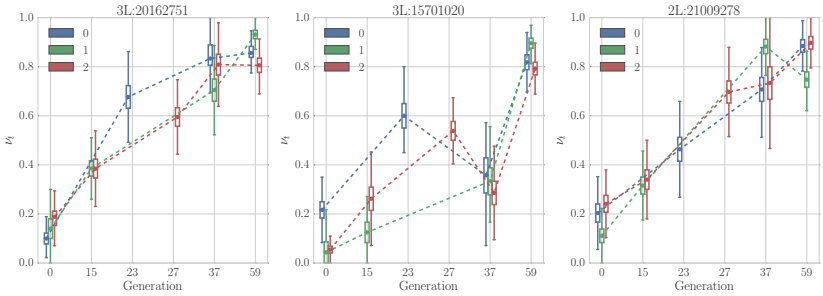

Fig S2. Trajectory of pool-sequenced variants.
Trajectory of three different variants that are increasing in frequency over time. Note that for read count data, the true allele frequency is not known. Here we draw the posterior distribution of the allele frequency at each time point using box plot. The median of each distribution is denoted by dots. The variance of each box is seen to be inversely related to the depth of the measurement. For instance, generation 59 is sequenced with higher coverage than generation 37. As a result, variance of observations in generation 59 is considerably smaller.

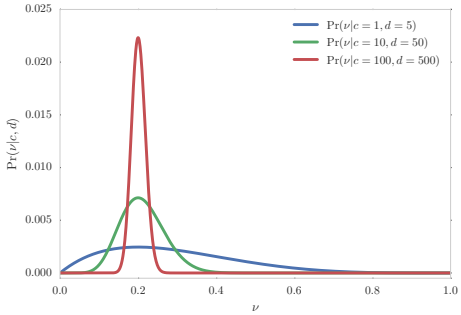

Fig S3. Posterior distribution of allele frequency.
Distribution of hidden allele frequency for different values of depth $d = \{5, 50, 500\}$. In all cases, the true frequency is 0.2. The estimated frequency values are binomially distributed, with different variances, around the true value in all cases with.

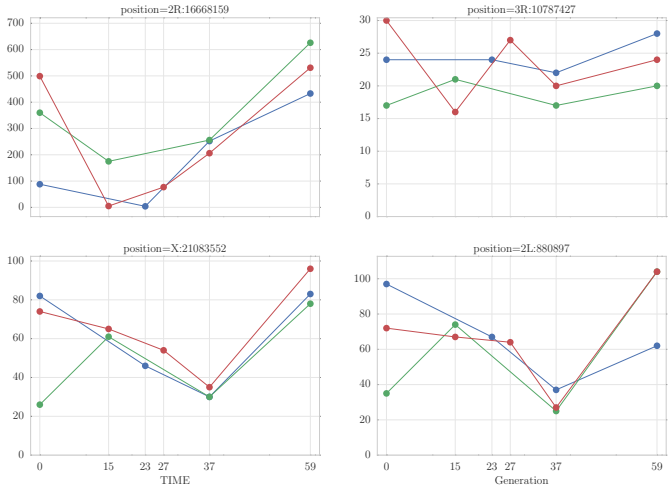

Fig S4. Coverage heterogeneity in time series data.
Each panel shows the read depth for 3 replicates of the data from a study of *D. melanogaster* adaptation to alternating temperatures data (see section 3.1). Heterogeneity in depth of coverage is seen between replicates, and also at different time points, in all 4 variants. None of these sites pass the the hard filtering with minimum depth of 30.

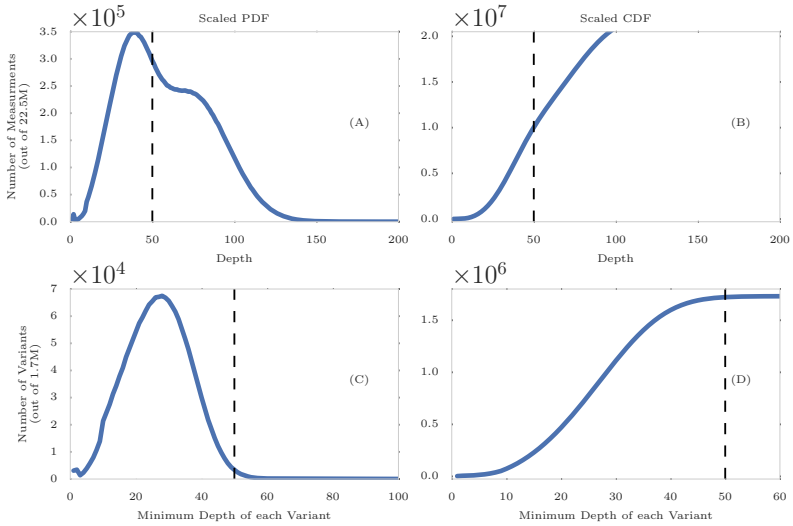

Fig S5. Distribution of depth in the real data.
Scaled PDF (A) and CDF (B) of the read depths of all (≈22.1M) measurements, i.e., all replicates and time points of the all (≈1.7M) variants. Scaled PDF (C) and CDF (D) of the minimum depth of sites. While more than half most (≈12.5M) of the measurements 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.

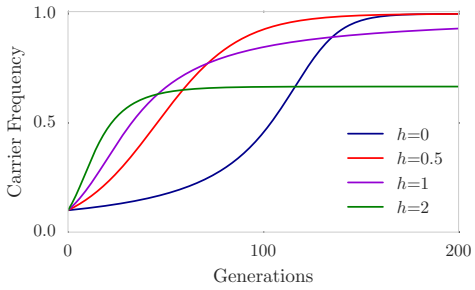

Fig S6. Natural selection in infinite population.
Trajectory of the favored allele in an infinite population with $s = 0.1$ for $h = 0$ (recessive favored allele), $h = 0.5$ (directional selection), $h = 1$ (dominant favored allele) and $h = 2$ (overdominant favored allele).

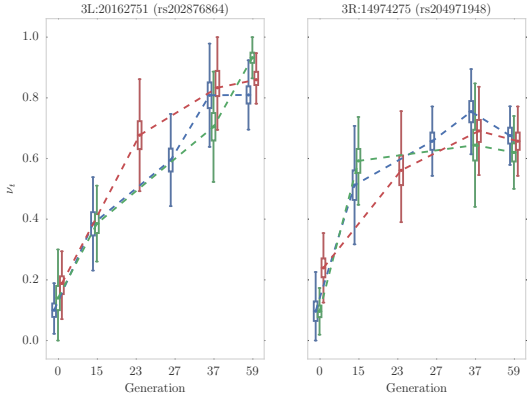

Fig S7. Variant showing strong signal of directional (Left) and over-dominant, aka balancing,(Right). selection. .

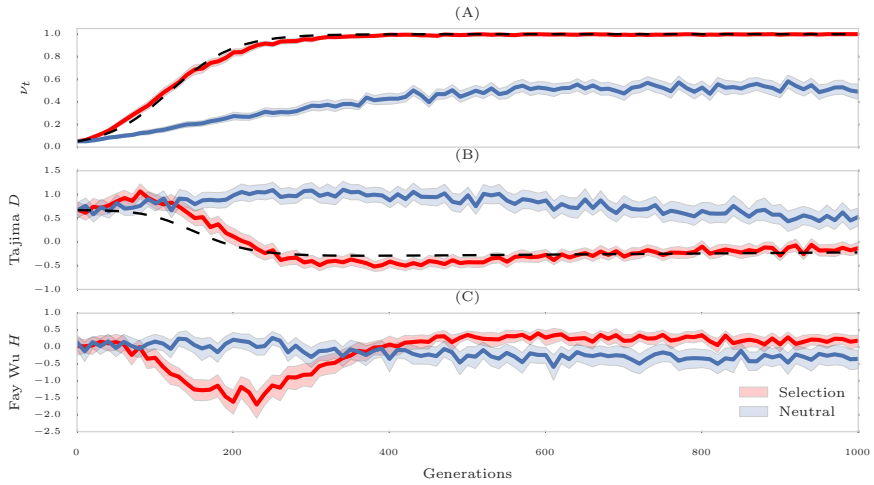

Fig S8. Dynamic SFS-based statistics.
Mean and 95% CI of 100 simulations for neutral (blue trajectories) selection with $s = 0.1$ (red trajectories). In all case, statistic computed for a 50Kbp window and $N_e = 10^4$, $\mu = 10^{-9}$. The dashed line shows the parametric models derived in Appendix 6.2.

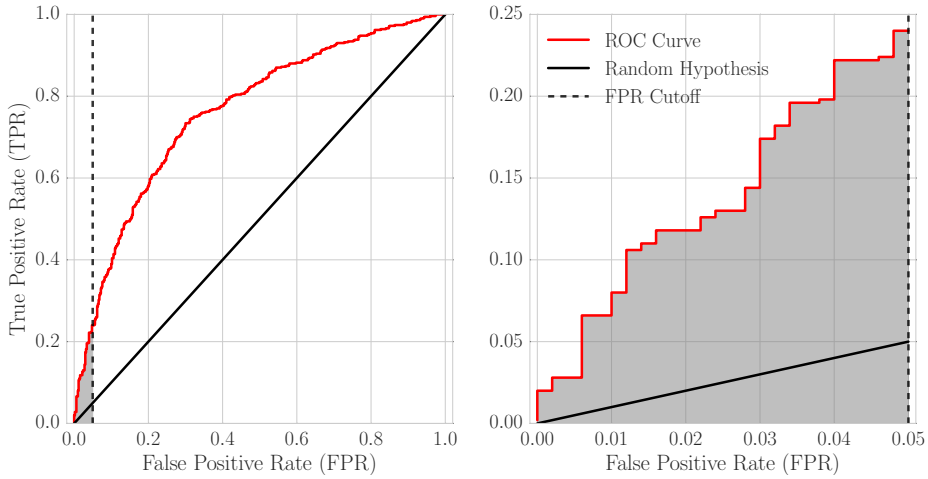

Fig S9. Schematic for computing power of detecting selection.
Receiver Operating Characteristics (ROC) curve (left) for classification of 2000 simulations (1000 selection and 1000 neutral). The Area Under the Curve (AUC) represents overall performance. The diagonal black line represents performance of a random hypothesis which achieves Area under the curve (AUC) of 0.5. To avoid computing AUC for the regions where FPR is unacceptably high, we restrict ROC curve to the region where FPR$\leq$ 0.05 (right). In this case, we define power to be the (scaled) AUC of the restricted region.

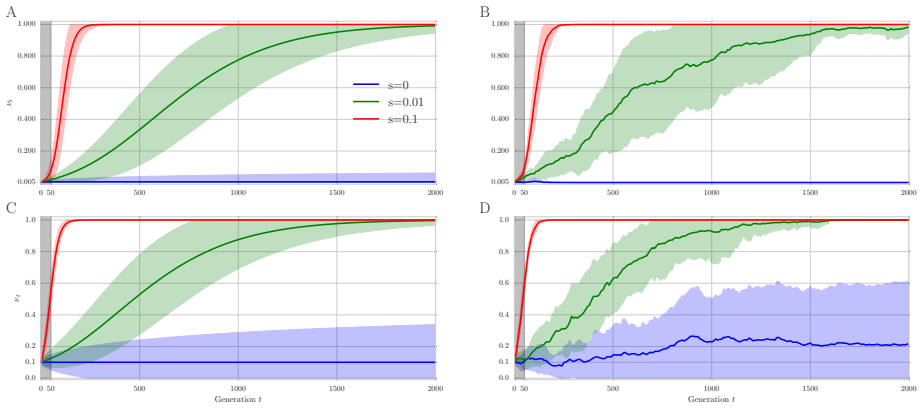

Fig S10. Theoretical (Markov chain) and Empirical trajectories of favored allele for hard and soft sweep scenarios.
Theoretical (A,C) and empirical (B,D) trajectories of the frequency of the favored allele are computed for 1000 simulations of populations with 1000 diploid individuals. Each curve shows the mean and the 95% confidence interval. Panels A and C represent theoretical Markov chain based calculations of the favored allele frequency under hard (ν_0 is small), and soft sweep due to standing variation (higher ν_0), for a range of values of s. Note that $s = 0$ corresponds to neutral evolution. Similarly, panels B and D show the empirical forward simulations of populations under the same selection regimes, and hard/soft-sweep scenarios. The first 50 generations are shaded in gray to represent the typical sampling span of EE experiments. The plot illustrates the difficulty of EE experiments in having to predict selection at a very early stage of the sweep. The signal is slightly stronger under standing variation scenario. The theoretical and empirical simulations are in close correspondence.

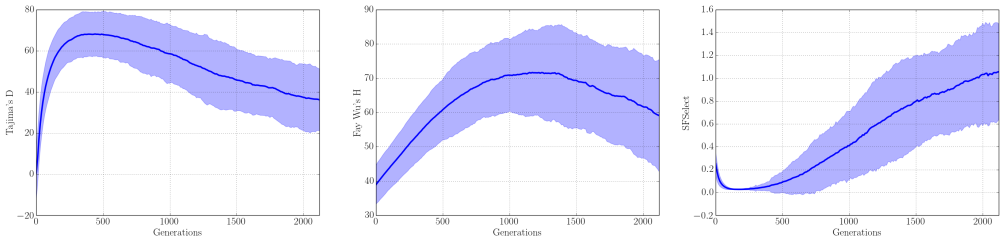

Fig S11. Effect of bottleneck in a typical experimental evolution experiment with restricted number of founder lines.
For the experiment, $F = 200$ founders were selected from a larger population size ($N_e = 10^{-6}$), and evolved under neutral scenario ($s = 0$). The statistics for Tajima's D (left), Fay Wu's H (middle) and SFSelect were computed for 1000 neutral simulations and the mean and 95% confidence interval plotted. Under neutral evolution, all the statistics are expected to vary around a fixed mean through time. However, under selective constraint, D and H take negative values, while SFSelect take positive values. In experimental evolution, bottleneck effect will suppress the signal of selection, especially in early generations.

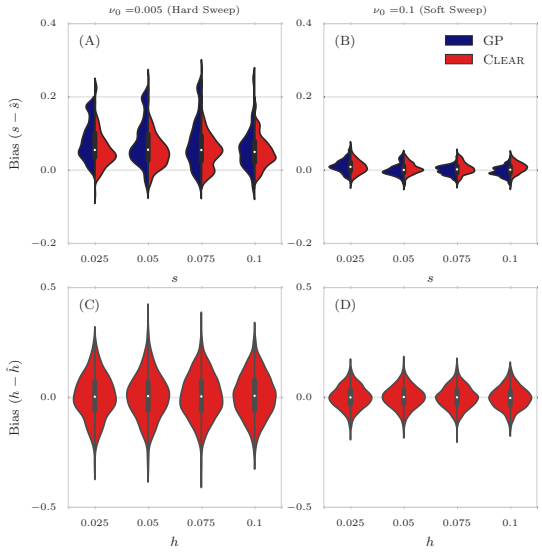

Fig S12. Distribution of bias for 30× coverage.
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = 30$ (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 S3. Panels C,D show the variance in the estimation of h.

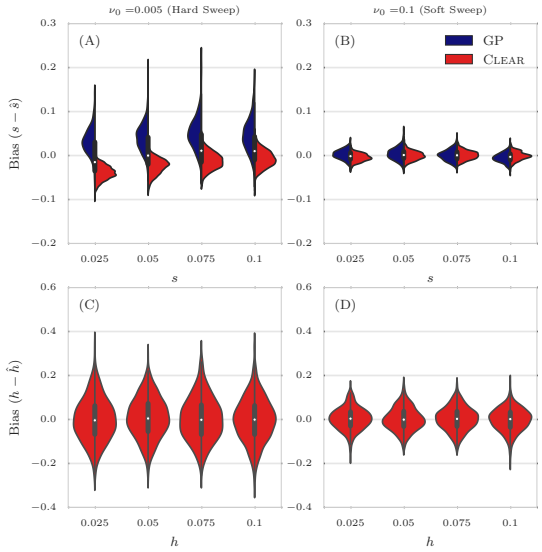

Fig S13. Distribution of bias for infinite coverage.
The distribution of bias $(s - \hat{s})$ in estimating selection coefficient over 1000 simulations using Gaussian Process (GP) and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = \infty$ (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 S3. Panels C,D show the variance in the estimation of h.

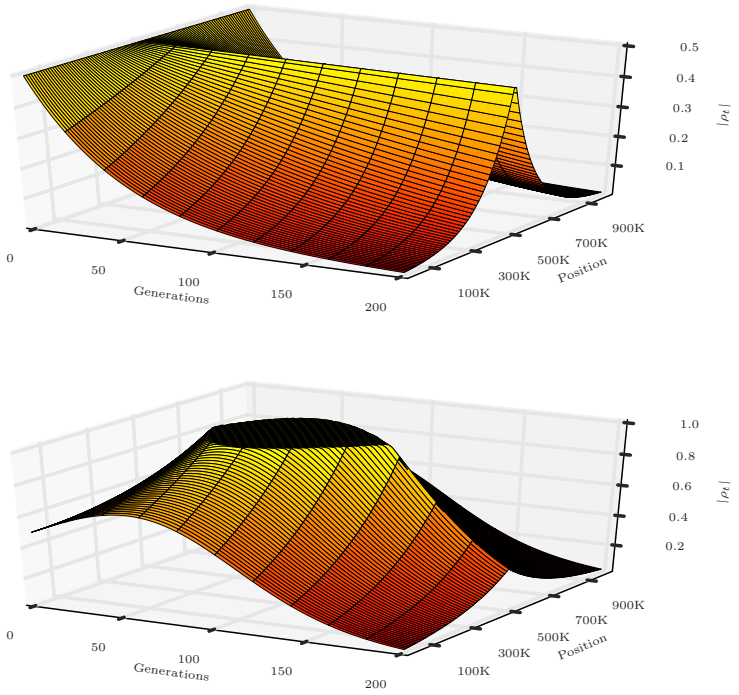

Fig S14. Expected dynamic of LD under selection and neutral evolution. Dynamic of LD (ρ_t) of a 1Mbp genome to the favored allele (at position 500K) is drawn as function of position and time for neutral (top) and selection(bottom) regimes. For sake of illustration, we assumed that at generations 0, LD of all variants with the favored allele is 0.5, initial frequency of the favored allele is 0.1, recombination rate is $r = 2 \times 10^{-8}$ (top). The selection strength is 0 and 0.05 for neutral and selection regimes, respectively. As expected LD decay exponentially through space and time. However, selection causes LD to increase then decrease.

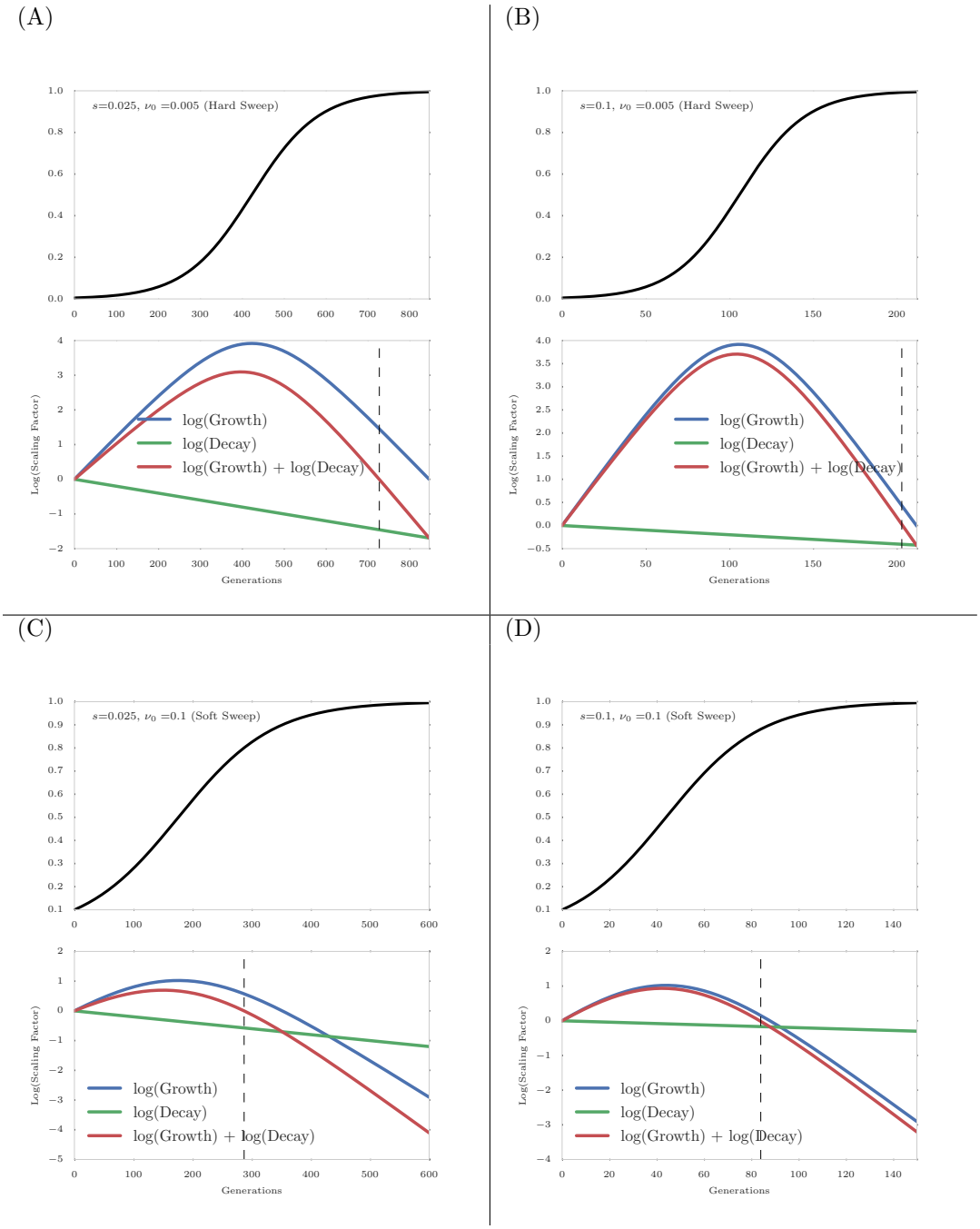

Fig S15. Interaction between growth and decay factors of LD. Expected evolution of LD under natural selection for weak selection (s=0.01) and a distance of 100Kb between sites for hard (A,B) and soft (C,D) sweeps. In addition to recombination, initial frequency of the favored allele and selection strength determine dynamic of LD. The vertical dashed line denotes the time in which LD start to decrease. In all cases, LD increase in first 50 generations, which implies that localizing adaptive allele in short term experimental evolution is a difficult task.

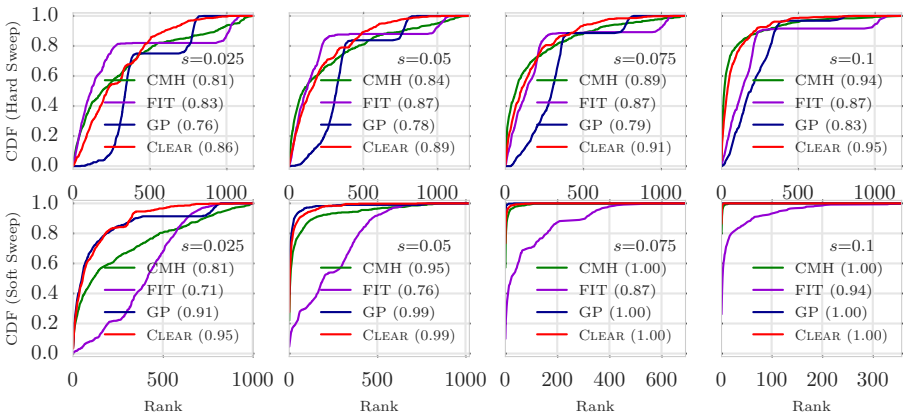

Fig S16. Ranking performance for 30× coverage.
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H), Gaussian Process (GP), CMH, and Frequency Increment Test (FIT), for different values of selection coefficient s and initial carrier frequency. Note that the individual variant CLEAR score (H) is used to rank variants. The Area Under Curve (AUC) is computed as a quantitative measure to compare the performance of methods for each configuration.

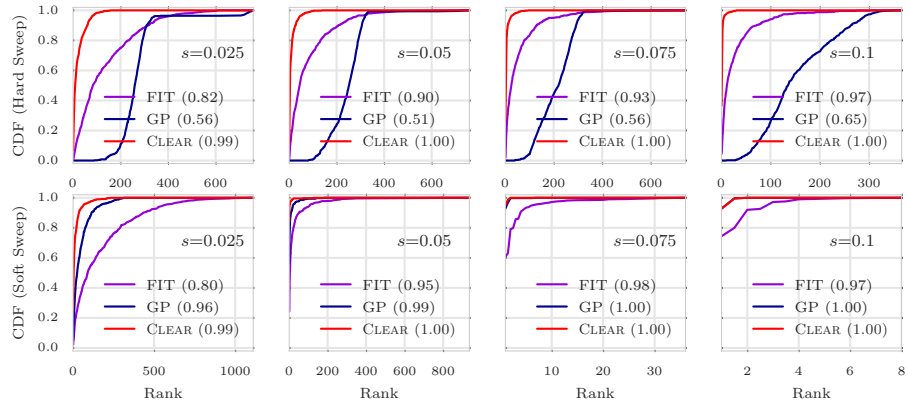

Fig S17. Ranking performance for infinite coverage.
Cumulative Distribution Function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (H), Gaussian Process (GP), CMH, and Frequency Increment Test (FIT), for different values of selection coefficient s and initial carrier frequency. Note that the individual variant CLEAR score (H) is used to rank variants. The Area Under Curve (AUC) is computed as a quantitative measure to compare the performance of methods for each configuration.