CLEAR: Composition of Likelihoods for Evolve and Resequence Experiments

Arya Iranmehr,*,[1] Ali Akbari,* Christian Schlötterer,[†] and Vineet Bafna[‡]

*Department of Electrical and Computer Engineering, and [‡]Department of Computer Science and Engineering, University of California, San Diego, California 92093, and [†]Institut für Populationsgenetik, Veterinärmedizinische Universität Wien, 1210 Vienna, Austria

ORCID IDs: 0000-0002-8404-9797 (A.I.); 0000-0002-5000-5876 (A.A.); 0000-0003-4710-6526 (C.S.)

ABSTRACT The advent of next generation sequencing technologies has made whole-genome and whole-population sampling possible, even for eukaryotes with large genomes. With this development, experimental evolution studies can be designed to observe molecular evolution "in action" via evolve-and-resequence (E&R) experiments. Among other applications, E&R studies can be used to locate the genes and variants responsible for genetic adaptation. Most existing literature on time-series data analysis often assumes large population size, accurate allele frequency estimates, or wide time spans. These assumptions do not hold in many E&R studies. In this article, we propose a method—composition of likelihoods for evolve-and-resequence experiments (CLEAR)—to identify signatures of selection in small population E&R experiments. CLEAR takes whole-genome sequences of pools of individuals as input, and properly addresses heterogeneous ascertainment bias resulting from uneven coverage. CLEAR also provides unbiased estimates of model parameters, including population size, selection strength, and dominance, while being computationally efficient. Extensive simulations show that CLEAR achieves higher power in detecting and localizing selection over a wide range of parameters, and is robust to variation of coverage. We applied the CLEAR statistic to multiple E&R experiments, including data from a study of adaptation of *Drosophila melanogaster* to alternating temperatures and a study of outcrossing yeast populations, and identified multiple regions under selection with genome-wide significance.

KEYWORDS experimental evolution; selection; genetic drift; time-series data; hidden Markov model; Wright–Fisher process

NATURAL selection is a key force in evolution, and a mechanism by which populations can adapt to external "selection" pressure. Examples of adaptation abound in the natural world (Fan *et al.* 2016), including classic examples like lactose tolerance in Northern Europeans (Bersaglieri *et al.* 2004), human adaptation to high altitudes (Simonson *et al.* 2010; Yi *et al.* 2010), but also drug resistance in pests (Daborn *et al.* 2001), HIV (Feder *et al.* 2016), cancer (Gottesman 2002; Zahreddine and Borden 2013), malarial parasite (Nair *et al.* 2007; Ariey *et al.* 2014), and others

(Spellberg *et al.* 2008). In these examples, understanding the genetic basis of adaptation can provide valuable information, underscoring the importance of the problem.

Experimental evolution refers to the study of the evolutionary processes of a model organism in a controlled (Hegreness *et al.* 2006; Bollback and Huelsenbeck 2007; Barrick *et al.* 2009; Lang *et al.* 2011; Orozco-ter Wengel *et al.* 2012; Lang *et al.* 2013; Oz *et al.* 2014) or natural (Barrett *et al.* 2008; Reid *et al.* 2011; Denef and Banfield 2012; Winters *et al.* 2012; Daniels *et al.* 2013; Maldarelli *et al.* 2013; Bergland *et al.* 2014) environment. Recent advances in whole-genome sequencing have enabled us to sequence populations at a reasonable cost, even for large genomes. Perhaps more important for experimental evolution studies, we can now evolve and resequence (E&R) multiple replicates of a population to obtain *longitudinal time-series data*, to investigate the dynamics of evolution at the molecular level. Although constraints such as small sizes, limited timescales, and oversimplified laboratory environments may limit the interpretation of E&R results,

these studies are increasingly being used to test a wide range of hypotheses (Kawecki *et al.* 2012) and have been shown to be more predictive than static data analysis (Sawyer and Hartl 1992; Boyko *et al.* 2008; Desai and Plotkin 2008). In particular, longitudinal E&R data are being used to estimate model parameters including population size (Pollak 1983; Waples 1989; Williamson and Slatkin 1999; Wang 2001; Terhorst *et al.* 2015; Jónás *et al.* 2016), strength of selection (Bollback *et al.* 2008; Illingworth and Mustonen 2011; Illingworth *et al.* 2012; Malaspinas *et al.* 2012; Mathieson and McVean 2013; Steinrücken *et al.* 2014; Terhorst *et al.* 2015), allele age (Malaspinas *et al.* 2012), recombination rate (Terhorst *et al.* 2015), mutation rate (Barrick and Lenski 2013; Terhorst *et al.* 2015), quantitative trait loci (Baldwin-Brown *et al.* 2014), and for tests of neutrality hypotheses (Burke *et al.* 2010; Bergland *et al.* 2014; Feder *et al.* 2014; Terhorst *et al.* 2015).

While many E&R study designs are being used (Barrick and Lenski 2013; Schlötterer *et al.* 2015), we restrict our attention to the adaptive evolution due to standing variation in fixed-size populations. This regime has been considered earlier, typically with *Drosophila melanogaster* as the model organism of choice, to identify adaptive genes in longevity and aging (600 generations) (Burke *et al.* 2010; Remolina *et al.* 2012), courtship song (100 generations) (Turner *et al.* 2011), hypoxia tolerance (200 generations) (Zhou *et al.* 2011), adaptation to new laboratory environments (59 generations) (Orozco-ter Wengel *et al.* 2012; Franssen *et al.* 2015), egg size (40 generations) (Jha *et al.* 2015), C-virus resistance (20 generations) (Martins *et al.* 2014), and dark-fly (49 generations) (Izutsu *et al.* 2015).

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

In the effort to analyze E&R selection experiments, many authors chose to adapt existing tests that were originally used for static data, pairwise comparisons (two time points), and single replicates to perform a null scan. For instance, Zhou *et al.* (2011) used the ratio of the estimated population size of case and control populations to compute a test statistic for each genomic region. Burke *et al.* (2010) applied the Fisher exact test to the last observation of data on case and control populations. Orozco-ter Wengel *et al.* (2012) used the Cochran–Mantel–Haenszel (CMH) test (Agresti and Kateri 2011) to detect SNPs whose read counts change consistently across all replicates of two time-point data. Turner *et al.* (2011) proposed the diffstat statistic to test whether the change in allele frequencies of two populations deviate from the distribution of change in allele frequencies of two drifting

populations. Bergland *et al.* (2014) calculated F_{st} to populations throughout time to signify their differentiation from ancestral (two time-point data) as well as geographically different populations. Jha *et al.* (2015) computed a test statistic of generalized linear-mixed model directly from read counts.

Alternatively, *direct* methods have been developed to analyze time-series data by taking a likelihood approach, and estimating population genetics parameters. Bollback *et al.* (2008) proposed a hidden Markov model (HMM) to estimate the selection coefficient *s* and population size by using a diffusion approximation to the Wright–Fisher process. Steinrücken *et al.* (2014) proposed a general diploid selection model which takes into account dominance of the favored allele and approximates likelihood analytically. Recently, Schraiber *et al.* (2016) proposed a Bayesian framework to estimate parameters using Markov chain Monte Carlo sampling. Mathieson and McVean (2013) adopted HMMs to structured populations and estimated parameters using an expectation maximization procedure on a discretized allele frequency. Feder *et al.* (2014) modeled increments in allele frequency with a Brownian motion process, proposed as the frequency increment test (FIT). More recently, Topa *et al.* (2015) proposed a Gaussian process (GP) for modeling single-locus, time-series, whole-genome sequencing of pools of individuals (pool-seq) data. Terhorst *et al.* (2015) extended GP to compute joint likelihood of multiple loci under null and alternative hypotheses. Finally, Levy *et al.* (2015) proposed a Bayesian model to handle sequencing, amplification, and growth noise in a large population of barcoded lineages.

Among the methods specifically designed for time-series data, many make assumptions which may not hold in E&R studies. One common assumption is that the underlying population size is large, so it is reasonable to model dynamics of allele frequencies using continuous-state models (Bollback *et al.* 2008; Feder *et al.* 2014; Terhorst *et al.* 2015). Second, many existing methods were originally designed to process the wider time spans seen in ancient DNA studies, an assumption that does not hold for E&R experiments (Steinrücken *et al.* 2014; Schraiber *et al.* 2016). Finally, many E&R analysis tools assume that allele frequencies in the input data are unbiased (*e.g.*, Bollback *et al.* 2008), which may not be valid for shotgun sequencing experiments.

Here, we consider an HMM, similar to Williamson and Slatkin (1999) and Bollback *et al.* (2008) but under a "small-population-size" regime. Specifically, we use a discrete state (frequency) model. We show that for small population sizes, discrete models can compute likelihood exactly, which improves statistical performance, especially for short time-span experiments. Additionally, we add another level of sampling noise to the traditional HMM model, allowing for heterogeneous ascertainment bias due to uneven coverage among variants. We show that for a wide range of parameters, CLEAR provides higher power for detecting selection, estimates model parameters consistently, and localizes the favored allele more

accurately compared to the state-of-the-art methods, while being computationally efficient.

Materials and Methods

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

1. Estimating population size: The procedure starts by estimating the effective population size, $\hat{N}$, under the assumption that much of the genome is evolving neutrally.
2. Estimating selection parameters: For each polymorphic site, selection and dominance parameters s, h are estimated so as to maximize the likelihood of the time-series data, given $\hat{N}$.
3. Computing likelihood statistics: For each variant, a log-odds ratio of the likelihood of selection model ($s > 0$) to the likelihood of neutral evolution/drift model is computed. Likelihood ratios in a genomic region are combined to compute the CLEAR statistic for the region.
4. Hypothesis testing: An empirical null distribution of the CLEAR statistic is calculated using genome-wide drift simulations, and used to compute P-values and thresholds for a specified false discovery rate (FDR). We perform single-locus hypothesis testing within selected regions to identify significant variants and report genes that intersect with the selected variants.

These steps are described in detail below.

Estimating population size

Methods for estimating population sizes from temporal neutral evolution data have been developed (Williamson and Slatkin 1999; Anderson *et al.* 2000; Bollback *et al.* 2008; Terhorst *et al.* 2015; Jónás *et al.* 2016). Here, we aim to extend these models to explicitly model the sampling noise that arise in pool-seq data. Specifically, we model the variation in sequence coverage over different locations, and the noise due to sequencing only a subset of the individuals in the population. In addition, many existing methods (Bollback *et al.* 2008; Feder *et al.* 2014; Terhorst *et al.* 2015; Topa *et al.* 2015) are designed for large populations, and model frequency as a continuous quantity. We observed that using Brownian motion to model frequency drift may be inadequate for small populations, low starting frequencies, and sparse sampling (in time); factors that are common in experimental evolution (see *Results*, Figure 3, A–C, and Figure 2). To this end, we model the Wright–Fisher Markov process for generating pool-seq data (Supplemental Material, Figure S1 in File S3) via a discrete HMM (Figure 1B). We start by computing a likelihood function for the population size given neutral pool-seq data.

Likelihood for the neutral model: We model the allele frequency counts $2N\nu_t$ as being sampled from a binomial distribution. Specifically,

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

where π is the global distribution of allele frequencies in the base population. Note that π depends on the demographic history of the founder lines and can be estimated from the site-frequency spectrum (see Figure S2 in File S3) of the initial population. For notational convenience, henceforth we omit the dependence of likelihoods to the parameter π.

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

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

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

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

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

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

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

$$d_t \sim \text{Poisson}(\lambda_t), \quad c_t \sim \text{binomial}(d_t, y_t). \tag{4}$$

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

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

For $1 \leq t \leq T$, $1 \leq j \leq 2N$, let $\alpha_{t,j}$ denote the probability of emitting $x_1, x_2, \ldots, x_t$ and reaching state j at τ_t. Then, α_t can be computed using the forward procedure (Durbin *et al.* 1998):

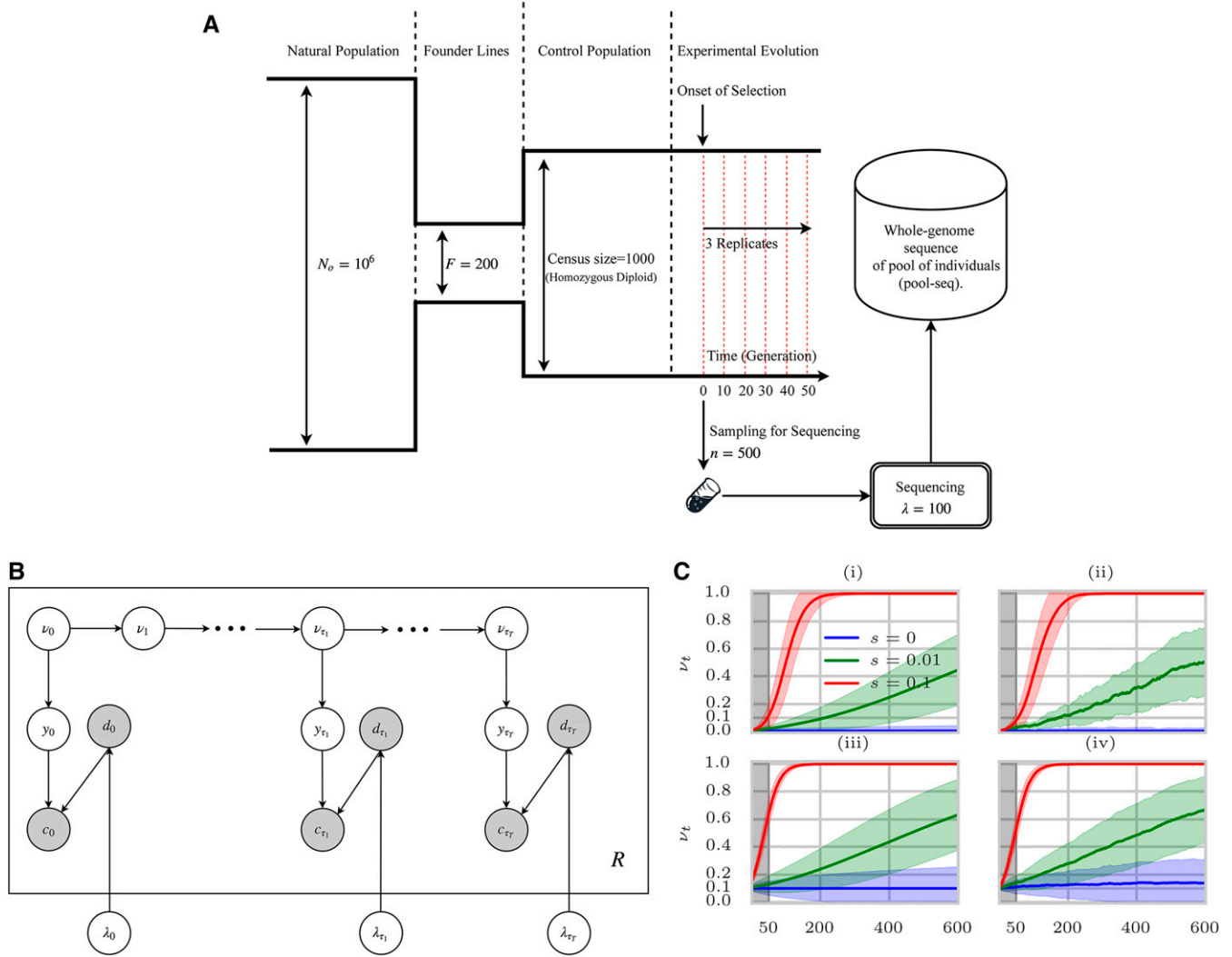$$\alpha_t^T = \alpha_{t-1}^T P^{(\delta_t)} \text{diag}[Y\mathbf{e}(x_t)], \tag{6}$$

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

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

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

where $\boldsymbol{x} = \{x_t\}_{t \in \mathcal{T}}$. The graphical model and the generative process for which data are being generated is depicted in Figure 1B and Figure S1 in File S3, respectively.

Finally, the last step is to compute an estimate $\hat{N}$ that maximizes the likelihood of all M variants in the whole ge-

nome. Let $\boldsymbol{x}_i^{(r)}$ denote the time-series data of the ith variant in replicate r. Then,

$$\hat{N} = \arg \max_N \prod_{i=1}^{M} \prod_{r=1}^{R} \mathcal{L}\left[N|\boldsymbol{x}_i^{(r)}\right]. \tag{8}$$

Estimating selection parameters

Likelihood for the selection model: Assume that the site is evolving under selection constraints $s \in \mathbb{R}$ and $h \in \mathbb{R}_+$, where s and h denote selection strength and dominance parameters, respectively. By definition, the relative fitness values of genotypes $0|0$, $0|1$, and $1|1$ are given by $w_{00} = 1$, $w_{01} = 1 + hs$,
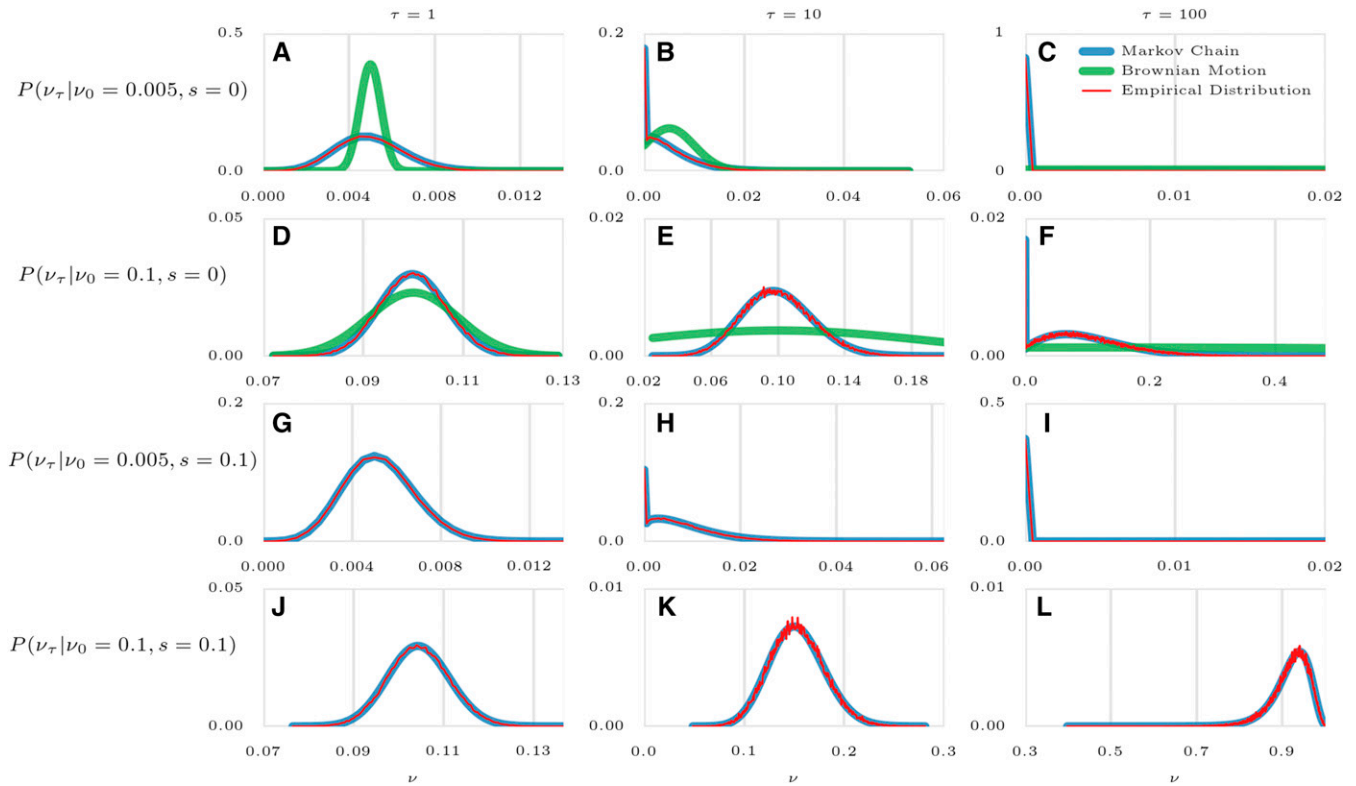
Figure 2 Comparison of empirical distributions of allele frequencies (red) *vs.* predictions from Brownian motion (green), and Markov chain (blue). Comparison of empirical and theoretical distributions under neutral evolution (A–F) and selection (G–M) with different starting frequencies $\nu_0 \in \{0.005, 0.1\}$ and sampling times of $\mathcal{T} = \{0, \tau\}$, where $\tau \in \{1, 10, 100\}$ and $N = 1000$. For each panel, the empirical distribution was computed over 100,000 simulations. Brownian motion (Gaussian approximation) provides poor approximations when initial frequency is far from 0.5 (A) or sampling is sparse (B, C, E , and F). In addition, Brownian motion can only provide approximations under neutral evolution. In contrast, Markov chain consistently provides a good approximation in all cases.

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

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

$$= \nu_t + \frac{s(h + (1-2h)\nu_t)\nu_t(1-\nu_t)}{1 + s\nu_t[2h + (1-2h)\nu_t]}.$$

(9)

The machinery for computing likelihood of the selection parameters is identical to that of population size, except for transition matrices. Hence, here we only describe the definition transition matrix $Q_{s,h}$ of the selection model. Let $Q_{s,h}^{(\tau)}[i,j]$ denote the probability of transition from $i/2N$ to $j/2N$ in τ generations, then (see Ewens 2004, p. 24, equations 1.58–1.59):

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

$$= \binom{2N}{j}\hat{\nu}_{t^+}^j(1-\hat{\nu}_{t^+})^{2N-j}$$

(10)

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

(11)

The maximum likelihood estimates are given by

$$\hat{s}, \hat{h} = \arg\max_{s,h} \prod_{r=1}^{R} \mathcal{L}\left(s, h|\mathbf{x}^{(r)}, \hat{N}\right).$$

(12)

Using grid search, we first estimate N (Equation 8), and subsequently, we estimate parameters s and h (Equation 12, Figure S3 in File S3). By broadcasting and vectorizing the grid search operations across all variants, the genome scan on millions of polymorphisms can be done in a significantly smaller time than iterating a numerical optimization routine for each variant (see *Results* and Figure 6).

Empirical likelihood-ratio statistics

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

$$H = -2\log\left[\frac{\mathcal{L}(\bar{s}, 0.5|\{\mathbf{x}\}_R, \hat{N})}{\mathcal{L}(0, 0.5|\{\mathbf{x}\}_R, \hat{N})}\right],$$

(13)

where $\bar{s} = \arg\max_{s} \prod_{r=1}^{R} \mathcal{L}(s, 0.5|\mathbf{x}^{(r)}, \hat{N})$. Similarly, we can define a test statistic for testing if selection is dominant by
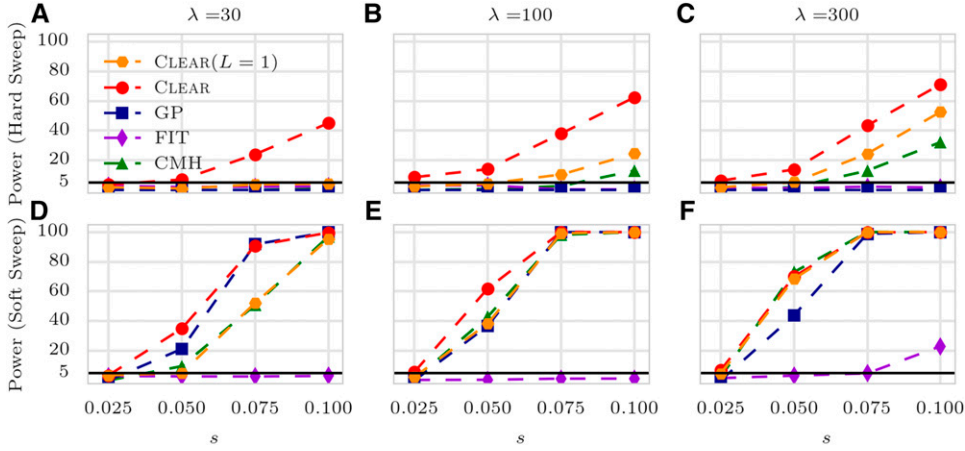
Figure 3 Power calculations for detection of selection. Detection power for CLEAR ($\mathcal{H}$), FIT, GP, and CMH under (A–C) hard- and (D–F) soft-sweep scenarios. λ and s denote the mean coverage and selection coefficient, respectively. Orange hexagons represent the performance of CLEAR when the maximum of the single-locus statistic is used to make a decision for the genomic region, while the red ● corresponds to the performance of CLEAR when single-locus statistics are averaged over the region. The y-axis measures power—sensitivity with false-positive rate (FPR) ≤ 0.05—for 2000 simulations with $N = 1000$ and $L = 50$ kbp. The horizontal line reflects the power of a random classifier. In all simulations, three replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

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

While extending the single-locus Wright–Fisher model to multiple linked loci can improve the power of the model (Terhorst *et al.* 2015), it is computationally and statistically expensive to compute exact likelihood. In addition, computing linked-loci joint likelihood requires haplotype resolved data, which pool-seq does not provide. Here, similar to Nielsen *et al.* (2005), we calculate the *composite-likelihood-ratio* score for a genomic region,

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

where L is a collection of segregating sites and H_ℓ is the likelihood-ratio score based for each variant ℓ in L. The optimal value of the hyper-parameter L depends upon a number of factors, including initial frequency of the favored allele, recombination rates, linkage of the favored allele to neighboring variants, population size, coverage, and time since the onset of selection (duration of the experiment). In File S3, we provide a heuristic to compute a reasonable value of L, based on experimental data.

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

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

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

Hypothesis testing

Single-locus tests: Under neutrality, log-likelihood ratios can be approximated by χ^2 distribution (Williams and Williams 2001), and P-values can be computed directly. However,

Feder *et al.* (2014) showed that when the number of independent samples (replicates) is small, χ^2 is a crude approximation to the true null distribution and results in more false positives. Following their suggestion, we first compute the empirical null distribution using simulations with the estimated population size (see Figure S1 in File S3). The empirical null distribution of statistic H is used to compute P-values as the fraction of null values that exceed the test score. Finally, we use the method of Storey and Tibshirani (2003) to control for FDR in multiple testing.

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

Simulations

We use the same simulation procedure for two purposes. First, we use the simulations to test the power of CLEAR against other methods in small genomic windows. Second, we use them to generate the distribution of null values for the statistic to compute empirical P-values. We mainly chose parameters that are relevant to *D. melanogaster* experimental evolution (Kofler and Schlötterer 2013). See also Figure 1A for illustration.

1. Creating initial founder line haplotypes: Using msms (Ewing and Hermisson 2010), we created neutral populations for F founding haplotypes with command $./msms $\langle F \rangle\ 1\ -t\ \langle 2\mu W N_o \rangle\ -r\ \langle 2r W N_o \rangle\ \langle W \rangle$, where $F = 200$ is the number of founder lines, $N_o = 10^6$ is the effective founder population size, $r = 2 \times 10^{-8}$ is the recombination rate, and $\mu = 2 \times 10^{-9}$ is the mutation rate. The window size W is used to compute $\theta = 2\mu N_o W$ and $\rho = 2N_o r W$. We chose $W = 50$ kbp for simulating individual windows for performance evaluations, and $W = 20$ Mbp for simulating *D. melanogaster* chromosomes for P-value computations.

2. Creating initial diploid population: An initial set of $F = 200$ haplotypes was created from step 1, and duplicated to
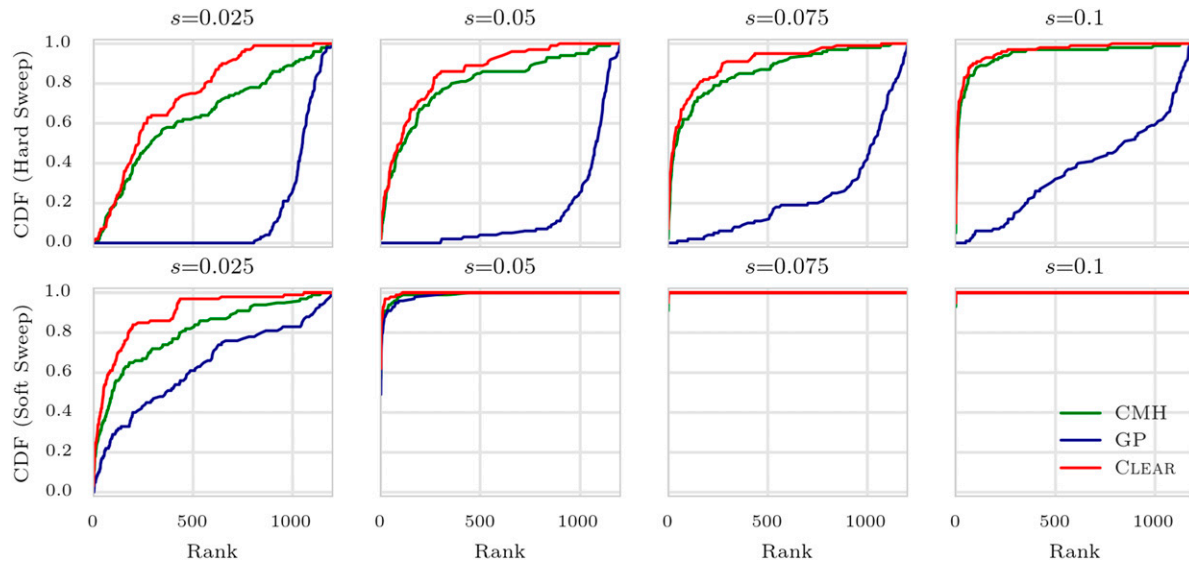
Figure 4 Ranking performance for 100× coverage. Cumulative distribution function (CDF) of the distribution of the rank of the favored allele in 1000 simulations for CLEAR (*H*), GP, CMH, and FIT, for different values of selection coefficient *s* and initial carrier frequency. Note that the individual variant CLEAR score (*H*) is used to rank variants. The area under the curve is computed as an overall quantitative measure to compare the performance of methods for each configuration. In all simulations, three replicates with $N = 1000$ are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

create *F* homozygous diploid individuals to simulate generation of inbred lines. *N* diploid individuals were generated by sampling with replacement from the *F* individuals.

3. Forward simulation: We used forward simulations for evolving populations under selection. We also consider selection regimes in which the favored allele is chosen from standing variation (not *de novo* mutations). Given initial diploid population, position of the site under selection, selection strength *s*, number of replicates $R = 3$, recombination rate $r = 2 \times 10^{-8}$, and sampling times $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$; simuPop (Peng and Kimmel 2005) was used to perform the forward simulation and compute allele frequencies for all of the *R* replicates. For hard sweep (respectively, soft sweep) simulations we randomly chose a site with an initial frequency of $\nu_0 = 0.005$ (respectively, $\nu_0 = 0.1$) to be the favored allele. For generating the null distribution with drift for *P*-value computations, we used this procedure with $s = 0$.

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

Data availability

The source code and running scripts for CLEAR are publicly available at https://github.com/airanmehr/clear.

D. melanogaster data was originally published (Orozco-ter Wengel *et al.* 2012; Franssen *et al.* 2015). The data set of the

D. melanogaster study, until generation 37, is obtained from the Dryad digital repository (http://datadryad.org) under accession DOI: 10.5061/dryad.60k68. Generation 59 of the *D. melanogaster* study is accessed from the European Sequence Read Archive (http://www.ebi.ac.uk/ena/) under the project accession number PRJEB6340. The data set containing the experimental evolution of yeast populations (Burke *et al.* 2014) is downloaded from http://wfitch.bio.uci.edu/~tdlong/PapersRawData/BurkeYeast.gz (last accessed January 24, 2017). University of California, Santa Cruz browser tracks for *D. melanogaster* and yeast data analysis are found in File 1 and File 2, respectively.

Results

Modeling allele frequency trajectories in small populations

We first tested the goodness of fit of the discrete *vs.* Brownian motion (a continuous-state model) in modeling allele frequency trajectories, under general E&R parameters. For this purpose, we conducted 100 K simulations with two time samples $\mathcal{T} = \{0, \tau\}$ where $\tau \in \{1, 10, 100\}$ is the parameter controlling the density of sampling in time. In addition, we repeated simulations for different values of starting frequency $\nu_0 \in \{0.005, 0.1\}$ (*i.e.*, hard and soft sweep) and selection strength $s \in \{0, 0.1\}$ (*i.e.*, neutral and selection). Then, given initial frequency ν_0, we computed the expected distribution of the frequency of the next sample ν_τ under two models to make a comparison. Figure 2, A–F, shows that Brownian motion (continuous model) is inadequate when ν_0 is far from 0.5, or when sampling times are sparse $(\tau > 1)$. If the favored allele arises from standing variation in a neutral population, it is unlikely to have a frequency close to 0.5, and
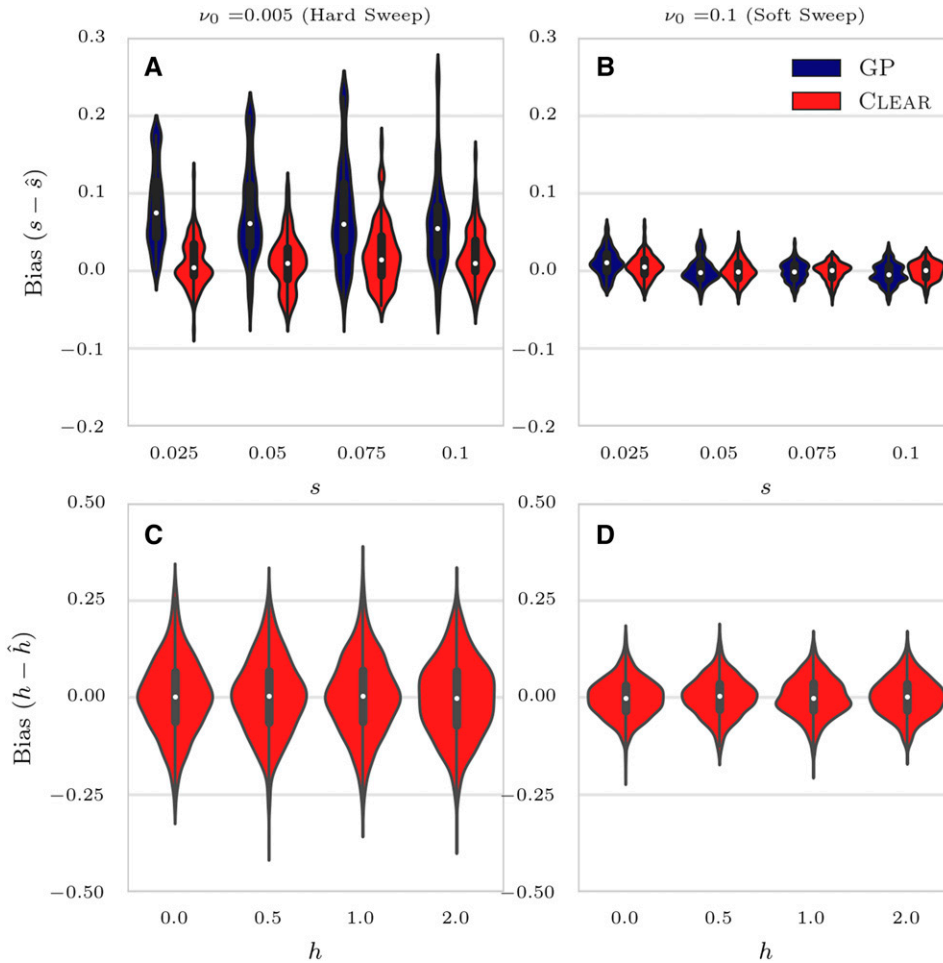
Figure 5 Distribution of bias for 100× coverage. The distribution of bias ($s - \hat{s}$) in estimating selection coefficient over 1000 simulations using GP and CLEAR (H) is shown for a range of choices for the selection coefficient s and starting carrier frequency ν_0, when coverage $\lambda = 100$ (A and B). GP and CLEAR have similar variance in estimates of s for soft sweep, while CLEAR provides lower variance in hard sweep. Also see Table S2 in File S3 (C and D). The variance in the estimation of h is shown. In all simulations, three replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.

the starting frequencies are usually much smaller (see Figure S2 in File S3). Moreover, in typical *D. melanogaster* experiments, for example, sampling is sparse. Often, the experiment is designed so that $10 \leq \tau \leq 100$ (Zhou *et al.* 2011; Orozco-ter Wengel *et al.* 2012; Kofler and Schlötterer 2013; Franssen *et al.* 2015).

In contrast to the Brownian motion approximation, discrete Markov chain predictions (Equation 11) are highly consistent with empirical data for a wide range of simulation parameters (Figure 2, A–M). Moreover, the discrete Markov chain can be modified to model the case when the the allele is under selection.

Detection power

We compared the performance of CLEAR against other methods for detecting selection. For each method we calculated detection power as the percentage of true positives identified with a false-positive rate ≤ 0.05. For each configuration (specified with values for selection coefficient s, starting allele frequency ν_0, and coverage λ), the power of each method is evaluated over 2000 distinct simulations, half of which modeled neutral evolution and the rest modeled positive selection.

We compared the power of CLEAR with GP (Terhorst *et al.* 2015), FIT (Feder *et al.* 2014), and CMH (Agresti and Kateri

2011) statistics. FIT and GP convert read counts to allele frequencies prior to computing the test statistic. CLEAR shows the highest power in all cases and the power stays relatively high even for low coverage (Figure 3 and Table S1 in File S3). In particular, the difference in performance of CLEAR with other methods is pronounced when starting frequency is low. The advantage of CLEAR stems from the fact that the
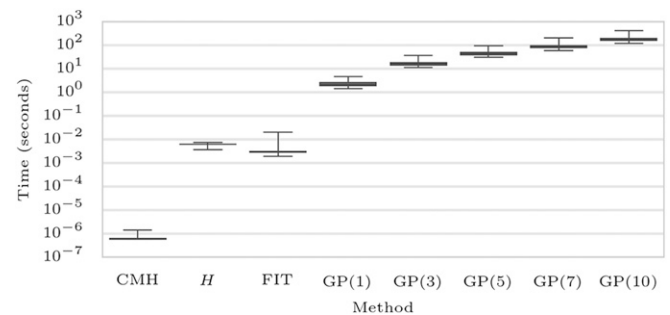

Figure 6 Running time. Box plots of running time per variant (CPU seconds) of CLEAR ($\mathcal{H}$), CMH, FIT, and GP with 1, 3, 5, 7, and 10 loci over 1000 simulations conducted on a workstation with Intel Core i7 processor. The average running time for each method is shown on the *x*-axis. In all simulations, three replicates are evolved and sampled at generations $\mathcal{T} = \{0, 10, 20, 30, 40, 50\}$.
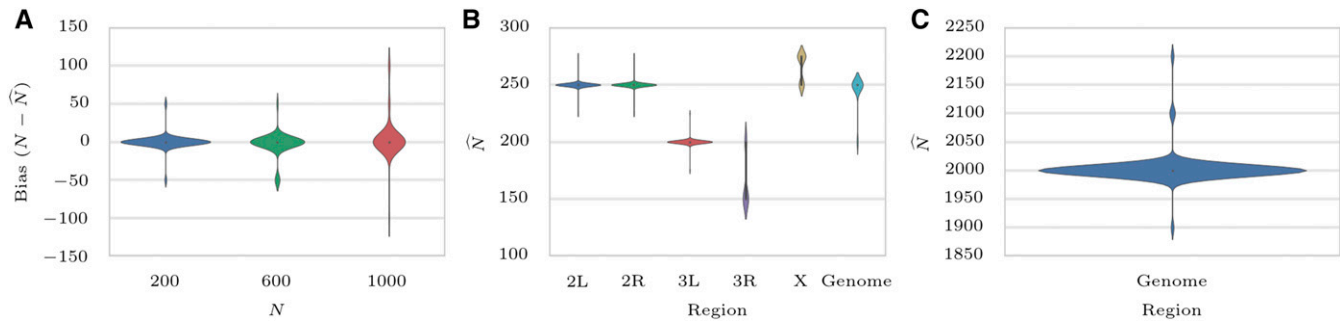
Figure 7 Estimating population size. (A) Distribution of bias in estimating N, computed on 1000 neutral simulations for each $N \in \{200, 600, 1000\}$ when $W = 10$ Mbp and $r = 2 \times 10^{-8}$. (B) Estimates of population size for data from a study of adaptation of *D. melanogaster* to alternating temperatures. For each case, the distribution of estimator is computed by 100 bootstrap computations using 1000 variants each. The multiple modes are an artifact of grid search used to speed up computation. (C) Distribution of the population size estimates on the yeast data set. Despite large census population size ($10^6 - 10^7$; Burke *et al.* 2014), this data set exhibits a much smaller effective population size ($\hat{N} = 2000$).

favored allele with a low starting frequency might be missed by low coverage sequencing. In this case, incorporating the signal from linked sites becomes increasingly important. We note that methods using only two time points, such as CMH, do relatively well for high selection values and high coverage. However, the use of time-series data can increase detection power in low coverage experiments or when the starting frequency is low. Moreover, time-series data provide means for estimating selection parameters s and h (see below). Finally, as CLEAR is robust to change of coverage, our results (Figure 3, B and C) suggest that taking many samples with lower coverage is preferable to sparse sampling with higher coverage. For comparison purposes, we also tested CLEAR using the single-locus statistic ($L = 1$). For the most part, CLEAR showed an improvement over other methods even with $L = 1$, or showed similar performance. The performance improved with higher L.

Site identification

In general, localizing the favored variant using pool-seq data is a nontrivial task due to extensive linkage disequilibrium (LD) (Tobler *et al.* 2014). To measure performance, we sorted variants by their H scores and computed rank of the favored allele for each method. For each setting of ν_0 and s, we conducted 1000 simulations and computed the rank of the favored mutation in each simulation. The cumulative distribution of the rank of the favored allele in 1000 simulations for each setting (Figure 4) shows that CLEAR outperforms other statistics.

An interesting observation is revisiting the contrast between site identification and detection (Long *et al.* 2013; Tobler *et al.* 2014). When selection strength is high, detection is easier (Figure 3, A–F), but site identification is harder, due to the high LD between flanking variants and the favored allele (Figure 4, A–F). Moreover, site identification becomes more difficult whenever the initial frequency of the favored allele is low; *i.e.*, at the onset of selection, LD between a favored allele and its nearby variants is high. 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 (Figure 3, B–E). In contrast, the favored site was ranked as the top in 14% of hard sweep cases, compared to 95% of soft sweep simulations.

Estimating parameters

CLEAR estimates effective population size $\hat{N}$ and selection parameters, $\hat{s}$ and $\hat{h}$, as a byproduct of the hypothesis testing. We computed bias of selection fitness ($s - \hat{s}$) and dominance ($h - \hat{h}$) for CLEAR and GP for 1000 simulations in each setting. The distribution of the error (bias) for $100\times$ coverage is presented in Figure 5 for different configurations. Figures S4 and S5 in File S3 provide the distribution of estimation errors for $30\times$ and $300\times$ coverage, respectively. For hard sweep, CLEAR provides estimates of s with lower variance of bias (Figure 5A and Figure S6 in File S3). In soft sweep, GP and CLEAR both provide unbiased estimates of s with low variance (Figure 5B). Figure 5, C and D, shows that CLEAR provides unbiased estimates of h as well when $h \in \{0, 0.5, 1, 2\}$ and $s = 0.1$. We also tested if CLEAR provides unbiased estimates of N by estimating population size on 1000 simulations when $N \in \{200, 600, 1000\}$. As shown in Figure 7A and Figure S9, A–C, in File S3 maximum likelihood is attained at the true value of the parameter.

Running time

As CLEAR does not compute exact likelihood of a region (*i.e.*, does not explicitly model linkage between sites), the complexity of scanning a genome is linear in the number of polymorphisms. Calculating the score of each variant requires an $\mathcal{O}(TRN^3)$ computation for $\mathcal{H}$. However, most of the operations are can be vectorized for all replicates to make the effective running time for each variant faster. We conducted 1000 simulations and measured running times for computing site statistics H, FIT, CMH, and GP with different numbers of linked loci. Our analysis reveals (Figure 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).
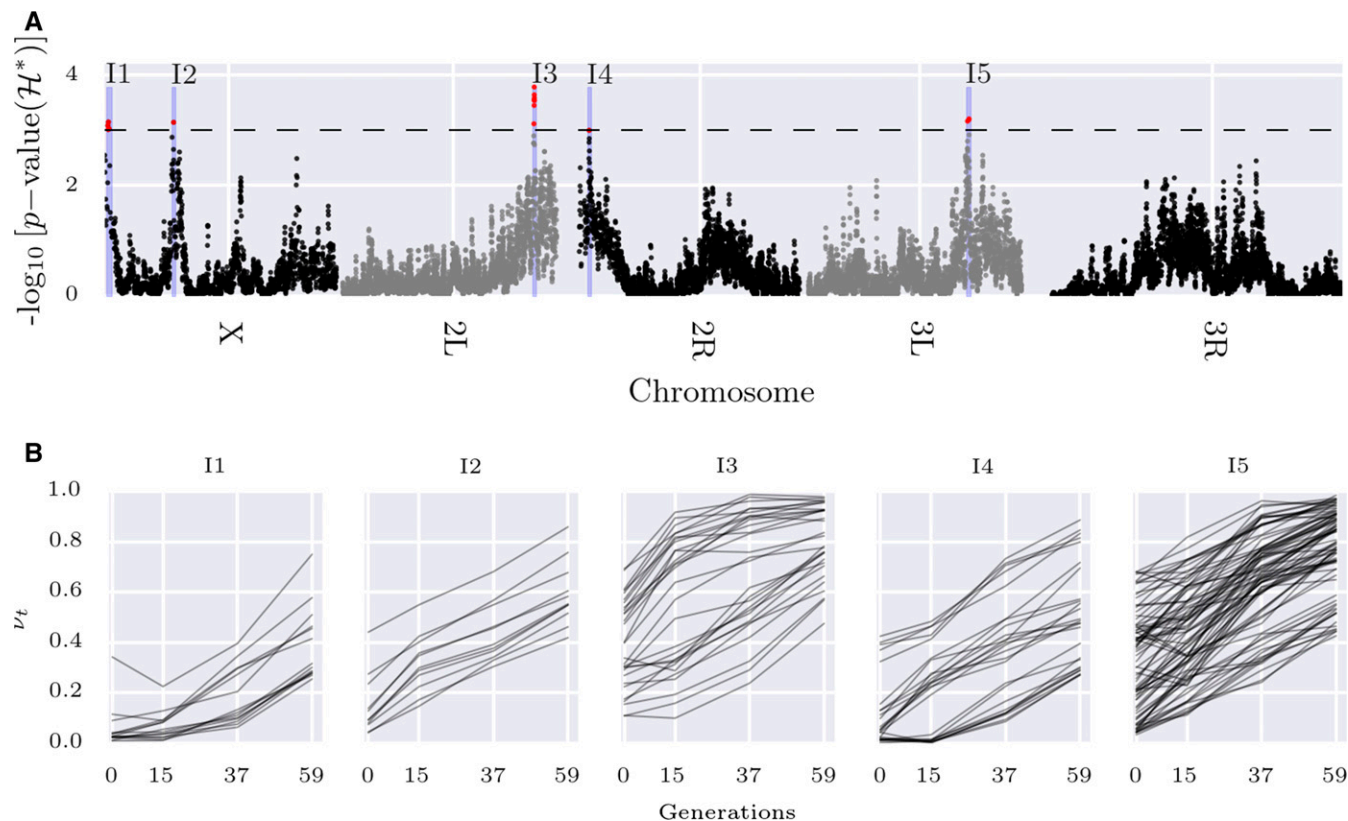
Figure 8 Scan of Clear statistic on data from a study of adaptation of *D. melanogaster* to alternating temperatures. (A) Manhattan plot of scan for $\mathcal{H}^*$ statistic using sliding window of size $L = 3000$ over the genome. The dashed line represents cutoff for genome-wide FDR ≤ 0.05, and identifies five contiguous intervals, I1–I5, which are shaded in blue. (B) Trajectories of the selected variants within intervals I1–I5.

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 ($\approx$1.6-M variant sites), it would take 1444 CPU hours ($\approx$1 CPU month). In contrast, after vectorizing and broadcasting operations for all variant operations using the numba package, Clear took 75 min to perform a scan, including precomputation; while the fastest method, CMH, took 17 min.

Analysis of Real Data

Analysis of a D. melanogaster adaptation to alternating temperatures: We applied Clear to the data from a study of the adaptation of *D. melanogaster* to alternating temperatures (Orozco-ter Wengel *et al.* 2012; Franssen *et al.* 2015), where three replicate samples were chosen from a population of *D. melanogaster* for 59 generations under alternating 12-hr cycles of hot-stressful (28°) and nonstressful (18°) temperatures, and sequenced. In this data set, sequencing coverage is different across replicates and generations (see figure S2 of Terhorst *et al.* 2015) which makes variant depths highly heterogeneous (Figure S10 in File S3).

We first filtered out heterochromatic, centromeric, and telomeric regions (Fiston-Lavier *et al.* 2010), and those variants that have a collective coverage of >1500 in all 13 populations: three replicates at the base population, two

replicates at generation 15, one replicate at generation 23, one replicate at generation 27, three replicates at generation 37, and three replicates at generation 59. After filtering, we ended up with 1,605,714 variants.

Next, we estimated genome-wide population size $\hat{N} = 250$ (Figure 7B and Figure S9E in File S3), which is consistent with previous studies (Orozco-ter Wengel *et al.* 2012; Jónás *et al.* 2016). The likelihood curves of Clear are sharper around the optimum compared to that of the method of Bollback *et al.* (2008) (see figure S1 in Orozco-ter Wengel *et al.* 2012). Also, chromosomes 3L and 3R appear to have a smaller population size, $\hat{N} = 200$, and 150, respectively. Others have made similar observations on this data. In particular, Jónás *et al.* (2016) showed that the chromosome-wise population size varies even more when it is computed for each replicate separately (see table 1 in Jónás *et al.* 2016). For instance, $\hat{N}$ is 131 for chromosome 3R replicate 1, while it is 328 for chromosome X replicate 2.

While it would be ideal to compute the Clear statistic for each replicate and chromosome separately, computing empirical *P*-values and significant regions become computationally intensive as the empirical null distribution of each replicate and each chromosome needs to be computed. Hence, we use a single genome-wide estimate $\hat{N} = 250$ in all analyses, but we normalize statistic $\mathcal{H}^*$ separately for each chromosome.
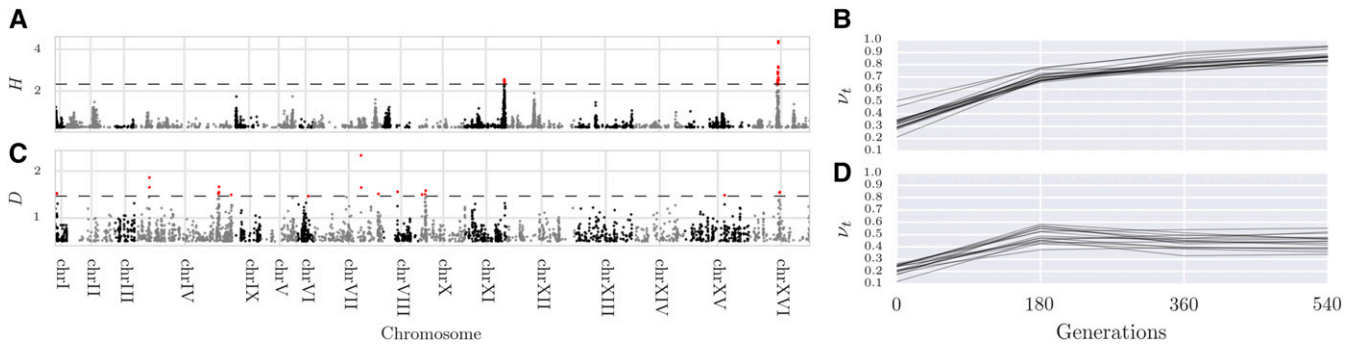
Figure 9 Single-locus analysis of the yeast outcrossing populations. Manhattan plot of scan single-locus CLEAR statistic ($L = 1$) for testing (A) directional selection and (C) dominant selection. The dashed line represents cutoff for genome-wide FDR ≤ 0.05. Trajectories of the selected variants are depicted in panels (B) and (D).

We used a heuristic calculation (see File S3) to choose the sliding window size L as the distance where the LD between the favored mutation and a site $L/2$-bp away remains strong. For *D. melanogaster* parameters, we obtained $L = 30$ kbp. We computed the normalized test statistic $\mathcal{H}^*$ on sliding windows of size of 30 kbp and step size of 5 kbp over the genome (see Figure 8A).

The empirical null distribution of $\mathcal{H}^*$ was estimated by creating 100 whole-genome simulations (400 K statistic values), as described in *Material and Methods*. Then, the P-value of the test statistic in each region in the experimental data was calculated as the fraction of the null statistic values that are greater than or equal to the test statistic(see Figure S11 in File S3). After correcting for multiple testing, we identified five contiguous intervals (Figure 8) satisfying FDR ≤ 0.05, and covering 2829 polymorphic sites. We further performed single-locus hypothesis testing on the 2829 sites to identify 174 individual variants with an FDR ≤ 0.01 (Figure 8B).

The final set of 174 variants fall within 32 genes (Table S3 in File S3), including many serine inhibitory proteases (serpins) and other genes involved in endocytosis. Recycling of synaptic vesicles is seen to be blocked at high temperature in temperature-sensitive *Drosophila* mutants (Kosaka and Ikeda 1983). This is also supported by gene ontology (GO)-enrichment analysis, where a single GO term "inhibition of proteolysis" is found to be enriched (corrected P-value = 0.0041). To test for dominant selection, we computed the D statistic on simulated neutral and experimental data, and computed P-values accordingly. After correcting for multiple testing, 96 variants were discovered with an FDR ≤ 0.01 (Figure S12 in File S3).

Analysis of outcrossing yeast populations

We also applied CLEAR to 12 replicate samples of outcrossing yeast populations (Burke *et al.* 2014), where samples were taken at generations $\mathcal{T} = \{0, 180, 360, 540\}$ We observed a significant variation in the genome-wide, site-frequency spectrum of certain populations over different time points for some replicates (Figure S13 in File S3). The variation does not have an easily identifiable cause. Therefore, we focused analysis on seven replicates $r \in \{3, 7, 8, 9, 10, 11, 12\}$ with a

genome-wide, site-frequency spectrum over the time range (Figure S14 in File S3).

We estimated population size to be $\hat{N} = 2000$ haplotypes (Figure 7C and Figure S9F in File S3), and computed the $\hat{s}$, $\hat{h}$, and H statistics accordingly. To compute P-values, we created 1-M single-locus neutral simulations according to the experimental data's initial frequency and coverage. By setting the FDR cutoff to 0.05, only 18 and 16 variants show significant signal for directional and dominant selection, respectively (Figure 9 S12 in File S3). Selected variants for directional selection are clustered in two regions, which match two of the five regions (regions C and E in figure 2A in Burke *et al.* 2014) identified by Burke *et al.* (2014) in their preliminary analysis. UCSC browser tracks for analysis of D. *melanogaster* and Yeast datasets are available as File S1 and File S2, respectively.

Discussion

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

Many factors; such as small population size, finite coverage, LD, finite sampling for sequencing, duration of the experiment, and the small number of replicates; can limit the power of tools for analyzing E&R. Here, by discrete modeling, CLEAR estimates population size and provides unbiased estimates of s and h. It adjusts for heterogeneous coverage of pool-seq data, and exploits the presence of linkage within a region to compute a composite likelihood ratio statistic.

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

The comparison of hard- and soft-sweep scenarios showed that the initial frequency of the favored allele can have a

nontrivial effect on the statistical power for identifying selection. Interestingly, while it is easier to detect a region undergoing strong selection, it is harder to locate the favored allele in that region.

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

Acknowledgments

Literature Cited

Agresti, A. Categorical Data Analysis, Ed. 2, in *Wiley Series in Probability and Statistics*. Wiley-Interscience, July 2002.

Anderson, E. C., E. G. Williamson, and E. A. Thompson, 2000   Monte Carlo evaluation of the likelihood for Ne from temporally spaced samples. Genetics 156: 2109–2118.

Ariey, F., B. Witkowski, C. Amaratunga, J. Beghain, A.-C. Langlois *et al.*, 2014   A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature 505: 50–55.

Baldwin-Brown, J. G., A. D. Long, and K. R. Thornton, 2014   The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Mol. Biol. Evol.* 31: 1040–1055.

Barrett, R. D. H., S. M. Rogers, and D. Schluter, 2008   Natural selection on a major armor gene in threespine stickleback. Science 322: 255–257.

Barrick, J. E., and R. E. Lenski, 2013   Genome dynamics during experimental evolution. Nat. Rev. Genet. 14: 827–839.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh *et al.*, 2009   Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461: 1243–1247.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov, 2014   Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS Genet. 10: e1004775.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner *et al.*, 2004   Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 74: 1111–1120.

Bollback, J. P., and J. P. Huelsenbeck, 2007   Clonal interference is alleviated by high mutation rates in large populations. Mol. Biol. Evol. 24: 1397–1406.

Bollback, J. P., T. L. York, and R. Nielsen, 2008   Estimation of 2Nes from temporal allele frequency data. Genetics 179: 497–502.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008   Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4: e1000083.

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010   Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature 467: 587–590.

Burke, M. K., G. Liti, and A. D. Long, 2014   Standing genetic variation drives repeatable experimental evolution in outcrossing populations of Saccharomyces cerevisiae. *Mol. Biol. Evol.* 31: 3228–3239.

Daborn, P., S. Boundy, J. Yen, and B. Pittendrigh, and R. ffrench-Constant, 2001   DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. Mol. Genet. Genomics 266: 556–563.

Daniels, R., H.-H. Chang, P. D. Séne, D. C. Park, D. E. Neafsey *et al.*, 2013   Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. PLoS One 8: e60780.

Denef, V. J., and J. F. Banfield, 2012   In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. Science 336: 462–466.

Desai, M. M., and J. B. Plotkin, 2008   The polymorphism frequency spectrum of finitely many sites under selection. Genetics 180: 2175–2191.

Durbin, R., S. R. Eddy, A. Krogh, and G. Mitchison, 1998   *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, United Kingdom.

Ewens, W. J., 2004   *Mathematical Population Genetics 1: Theoretical Introduction* (Interdisciplinary Applied Mathematics, Vol. 27). Springer Science & Business Media, New York.

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

Fan, S., M. E. B. Hansen, Y. Lo, and S. A. Tishkoff, 2016   Going global by adapting local: a review of recent human adaptation. Science 354: 54–59.

Feder, A. F., S. Kryazhimskiy, and J. B. Plotkin, 2014   Identifying signatures of selection in genetic time series. Genetics 196: 509–522.

Feder, A. F., S.-Y. Rhee, S. P. Holmes, R. W. Shafer, D. A. Petrov *et al.*, 2016   More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. Elife 5: e10670 [corrigenda: Elife 6: e24879 (2017)].

Fiston-Lavier, A.-S., N. D. Singh, M. Lipatov, and D. A. Petrov, 2010   Drosophila melanogaster recombination rate calculator. Gene 463: 18–20.

Franssen, S. U., V. Nolte, R. Tobler, and C. Schlötterer, 2015   Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental Drosophila melanogaster populations. Mol. Biol. Evol. 32: 495–509.

Gottesman, M. M., 2002   Mechanisms of cancer drug resistance. Annu. Rev. Med. 53: 615–627.

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

Illingworth, C. J. R., and V. Mustonen, 2011   Distinguishing driver and passenger mutations in an evolutionary history categorized by interference. Genetics 189: 989–1000.

Illingworth, C. J. R., L. Parts, S. Schiffels, G. Liti, and V. Mustonen, 2012   Quantifying selection acting on a complex trait using allele frequency time series data. Mol. Biol. Evol. 29: 1187–1197.

Izutsu, M., A. Toyoda, A. Fujiyama, K. Agata, and N. Fuse, 2015   Dynamics of dark-fly genome under environmental selections. G3 (Bethesda) 6: 365–376.

Jha, A. R., C. M. Miles, N. R. Lippert, C. D. Brown, K. P. White *et al.*, 2015   Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in Drosophila melanogaster. Mol. Biol. Evol. 32: 2616–2632.

Jónás, Á., T. Taus, C. Kosiol, C. Schlötterer, and A. Futschik, 2016 Estimating the effective population size from temporal allele frequency changes in experimental evolution. Genetics 204: 723–735.

Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri *et al.*, 2012 Experimental evolution. Trends Ecol. Evol. 27: 547–560.

Kofler, R., and C. Schlötterer, 2013 A guide for the design of evolve and resequencing studies. *Mol. Biol. Evol.* 31: 474–483.

Kosaka, T., and K. Ikeda, 1983 Reversible blockage of membrane retrieval and endocytosis in the Garland cell of the temperature-sensitive. J. Cell Biol. 97: 499–507.

Lang, G. I., D. Botstein, and M. M. Desai, 2011 Genetic variation and the fate of beneficial mutations in asexual populations. Genetics 188: 647–661.

Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock *et al.*, 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500: 571–574.

Levy, S. F., J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher *et al.*, 2015 Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature 519: 181–186.

Long, Q., F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow *et al.*, 2013 Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat. Genet. 45: 884–890.

Malaspinas, A.-S., O. Malaspinas, S. N. Evans, and M. Slatkin, 2012 Estimating allele age and selection coefficient from time-serial data. Genetics 192: 599–607.

Maldarelli, F., M. Kearney, S. Palmer, R. Stephens, J. Mican *et al.*, 2013 HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. J. Virol. 87: 10313–10323.

Martins, N. E., V. G. Faria, V. Nolte, C. Schlötterer, L. Teixeira *et al.*, 2014 Host adaptation to viruses relies on few genes with different cross-resistance properties. Proc. Natl. Acad. Sci. USA 111: 5938–5943.

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

Nair, S., D. Nash, D. Sudimack, A. Jaidee, M. Barends *et al.*, 2007 Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. Mol. Biol. Evol. 24: 562–573.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. Genome Res. 15: 1566–1575.

Orozco-ter Wengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt *et al.*, 2012 Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol. Ecol. 21: 4931–4941.

Oz, T., A. Guvenek, S. Yildiz, E. Karaboga, Y. T. Tamer *et al.*, 2014 Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution. *Mol. Biol. Evol.* 31: 2387–2401.

Peng, B., and M. Kimmel, 2005 simuPOP: a forward-time population genetics simulation environment. Bioinformatics 21: 3686–3687.

Pollak, E., 1983 A new method for estimating the effective population size from allele frequency changes. Genetics 104: 531–548.

Reid, B. J., R. Kostadinov, and C. C. Maley, 2011 New strategies in Barrett's esophagus: integrating clonal evolutionary theory with clinical management. Clin. Cancer Res. 17: 3512–3519.

Remolina, S. C., P. L. Chang, J. Leips, S. V. Nuzhdin, and K. A. Hughes, 2012 Genomic basis of aging and life-history evolution in Drosophila melanogaster. Evolution 66: 3390–3403.

Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics 132: 1161–1176.

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

Schraiber, J. G., S. N. Evans, and M. Slatkin, 2016 Bayesian inference of natural selection from allele frequency time series. Genetics 203: 493–511.

Simonson, T. S., Y. Yang, C. D. Huff, H. Yun, G. Qin *et al.*, 2010 Genetic evidence for high-altitude adaptation in Tibet. Science 329: 72–75.

Spellberg, B., R. Guidos, D. Gilbert, J. Bradley, H. W. Boucher *et al.*, 2008 The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. Clin. Infect. Dis. 46: 155–164.

Steinrücken, M., A. Bhaskar, and Y. S. Song, 2014 A novel spectral method for inferring general diploid selection from time series genetic data. Ann. Appl. Stat. 8: 2203–2222.

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100: 9440–9445.

Terhorst, J., C. Schlötterer, and Y. S. Song, 2015 Multi-locus analysis of genomic time series data from experimental evolution. PLoS Genet. 11: e1005069.

Tobler, R., S. U. Franssen, R. Kofler, P. Orozco-terWengel, V. Nolte *et al.*, 2014 Massive habitat-specific genomic response in D. melanogaster populations during experimental evolution in hot and cold environments. Mol. Biol. Evol. 31: 364–375.

Topa, H., Á. Jónás, R. Kofler, C. Kosiol, and A. Honkela, 2015 Gaussian process test for high-throughput sequencing time series: application to experimental evolution. Bioinformatics 31: 1762–1770.

Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. PLoS Genet. 7: e1001336.

Wang, J., 2001 A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Genet. Res. 78: 243–257.

Waples, R. S., 1989 A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121: 379–391.

Williams, D., 2001 *Weighing the Odds: A Course in Probability and Statistics*. Cambridge University Press, Cambridge, United Kingdom.

Williamson, E. G., and M. Slatkin, 1999 Using maximum likelihood to estimate population size from temporal changes in allele frequencies. Genetics 152: 755–761.

Winters, M. A., R. M. Lloyd, Jr., R. W. Shafer, M. J. Kozal, M. D. Miller *et al.*, 2012 Development of elvitegravir resistance and linkage of integrase inhibitor mutations with protease and reverse transcriptase resistance mutations. PLoS One 7: e40514.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo *et al.*, 2010 Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329: 75–78.

Zahreddine, H., and K. L. Borden, 2013 Mechanisms and insights into drug resistance in cancer. Front. Pharmacol. 4: 28.

Zhou, D., N. Udpa, M. Gersten, D. W. Visk, A. Bashir *et al.*, 2011 Experimental selection of hypoxia-tolerant Drosophila melanogaster. Proc. Natl. Acad. Sci. USA 108: 2349–2354.

Communicating editor: Y. S. Song