A Hidden Markov Model for Analyzing Adaptive Experimental Evolutions with Pooled Sequencing Data

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

[1]Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA.
[2]Institut fr Populationsgenetik, Vetmeduni, Vienna, Austria.
[3]Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Experimental evolution (EE) studies are powerful tools for observing molecular evolution "in-action" in wild and controlled environments. This paradigm of experiment was infeasible until recently when the whole-genome and whole-population was made possible by next-generation sequencing technologies. However, one of the primary constraints of the EE studies is the limited time for the experiment, which primarily depend on the organism's generation time. This constraint impedes adaptation and optimization (evolvability) studies, where the population can only evolved and re-sequenced in a small number of generations, relative to the number of generations required for fixation of adaptive allele. Although a powerful library of tests-of-selection has already been developed, they are mainly designed for static data to to identify adaptation when the sample is taken close enough (before/after) to the fixation of adaptive allele. In this article, we study the problem of identifying selective sweep in short-term experimental evolution of sexual organisms and propose Composite Of MArkovian Likelihoods for Experimental evolution (COMALE) statistic which computes its score by averaging likelihood ratios of polymorphisms for a genomic region. The likelihood of null (neutral) and alternative (selection) hypotheses calculated using the Wright-Fisher Markov chain model for each variant. Extensive simulation study shows that COMALE achieves higher detection power methods on both soft and hard sweep simulations for various selection strengths. Finally, we apply the COMALE statistic to the controlled experimental evolution of D. melanogaster to detect adaptive genes/alleles under alternating cold and hot temperatures.

1 Introduction

Experimental Evolution. Recent advances in whole genome sequencing have enabled us to sequence populations at a reasonable cost to perform *longitudinal studies* and study different forces of evolution in real-time. Modern experimental evolution refers to the study of the evolutionary processes of a model organism at genomic level in a controlled [6, 8, 23, 31, 32, 41, 42] or natural [4, 7, 13, 14, 34, 46, 66] environment. Although constraints such as small population sizes, limited timescales and oversimplified laboratory environments limits interpreting experimental evolution results, they can be used to test different hypotheses [30] regarding mutation rate, inbreeding, environmental variability, sexual selection & conflict, kin selection and cooperation life history and sex allocation, sexual reproduction and mating systems, behavior and cognition, hostparasite interactions, speciation repeatability of evolution and make more accurate inferences than static data analysis [10, 16, 51]. In addition, dynamic data has been used to estimate model parameters including population size [43, 58, 62–64] strength of selection [9, 25, 26, 33, 36, 55, 58], allele age [33] recombination rate [58], mutation rate [5, 58] and test neutrality hypotheses [7, 11, 20, 58].

Among different types of evolution experiments [5, 52] in this paper we only focus on adaptive evolution of multicellular sexual organisms with continuous culture , fixed population size, single locus selection (only one causal mutation). For this setting, D. melanogaster is usually the model organism and it has been used to identify adaptive genes in longevity and aging [11, 47] (600 and generations respectively), courtship song [60] (100 generations), hypoxia tolerance [68] (200 generations), adaptation to new temperatures[41, 59] (59 generations), egg size [28] (40 generations), C virus resistance [35] (20 generations), and dark-fly [27] (49 generations) experiments.

Natural selection. Natural selection is one of the main forces of the evolutionary process and identifying it at the genomic level is one of the critical problems facing humanity. For example, drug resistance in HIV [21], cancer [22, 67], malaria [3, 38], pests [12] or antibiotic resistance [54] are instances of genetic adaptations that yet to be understood at genomic level, which could potentially return us to the pre-antibiotic age. Although, a wide range of computational methods [61] enable us to identify different regimes of genetic adaptations, Messer and Petrov [37] argued that "many, if not most, cases of adaptation are yet to be discovered", false-negatives. In addition, current methods prone to "pathelogical false-positives" due to other confounding factors such as demography.

Selective sweep. Selective sweep [29, 53] is the model for describing directional single-locus selection, which takes into account of associations of the beneficial mutation with its surrounding loci. The extent of genetic loci that are in association with the adaptive allele depends on the amount of accumulated recombination events between adaptive allele and the rest of genome.

In the asexual populations, where no crossover occurs, the whole chromosome is perfectly linked to the adaptive allele the whole chromosome "hitchhikes" with the beneficial mutation in the sweep process. Also, when (beneficial) mutation rate is high, it is possible that more than one beneficial allele exist in the population at the same time and "clonal interference" [15, 32] best describes the adaptation process.

On the other hand, in sexual populations, the favored mutation is only in linkage-disequilibrium (LD) with its nearby polymorphisms. Hence, methods for identifying selective sweep in sexual populations often analyze polymorphism data of a population of in a genomic region, rather than a single site.

Identifying Natural Selection. Adaptation leaves a variety of signatures in different kinds of genomic data, and methods for identifying natural selection are essentially *data-driven*. For instance, reduction in genetic diversity[19, 49, 57] in allele-frequency data, prevalence of long haplotypes [50, 61] in haplotype (phased) data, population differentiation [11, 24] in multiple-population data and rapid increase in allele frequencies [7] in the dynamic data are different signatures of selective sweep in the polymorphism data. In this paper, we restrict our attention to the experimental evolution experiments with pooled-sequencing, where dynamic allele frequency of the population is available.

The identification of an selection event can be done in different levels of detail. At the coarsest level, identification can be done by determining whether a region (e.g. a small region with no or low recombination) on genome is under selection. In the rest, we consider this task to be the task of *detection*. Then, finding the causal mutation/allele would be a more elaborate identification of selection, henceforth, *locating* selection. Finally, estimating model parameters such as strength of selection and overdominance at the site fully describes the selective sweep.

Static Data. Traditionally, given static allele frequency data, Site Frequency Spectrum (SFS) is computed to perform neutrality tests including Tajima's D [57], Fay and Wu's H [19], Composite

Likelihood Ratio [40], SFSelect [49], in a genomic region[1]. Despite their simplicity and clarity, it has been shown that SFS-based tests often fail to distinguish demographic changes from adaptation. They are also prone to pathological false-positive/negatives due to low linkage of the adaptive allele to its surrounding variation and ascertainment bias [2, 37, 39, 44, 45].

SFS (Figure S1) shows the distribution of allele frequencies in a genomic region. Historically, SFS of static data has been extensively used to "detect" genetic adaption and demographic changes in a population by measuring the diversity in a genomic region. In general, reduction in diversity is a signal of selection, and detecting selection based on the reduction in genomic diversity is a subtle task, because

(i) in soft sweeps the genomic diversity does not necessarily reduce.

(ii) even in hard sweeps with no recombination, the reduction diversity is significant only when the SFS sample is taken close to fixation (not too far prior or after fixation)

(iii) SFS change in the same way for both selection an demography changes.

Although conditions (hard sweep with no recombination, not far from fixation, and random-mating and constant size population) for detecting selection based on SFS are very restrictive, SFS-based tests are simple and inexpensive to use and often used in combination with other tests [2, 61].

In practice, (positive and sum-to-one) weighted linear combinations scaled SFS distribution [1] is used as different estimators of θ and pairwise discrepancy between them is used as a test statistic for detecting selection. Under neutrality, the discrepancies should be distributed around zero, and a simple t test can provide p-value for rejecting neutrality. For example, test statistics for Tajima's D[57], Fay Wu's H [19] and SFSelect[49] can be obtained by a dot product of the scaled SFS vector with their corresponding weight vector.

Dynamic Data (ad-hoc). On the other hand, tests-of-selection for dynamic data is less studied, and often existing tests for static data are adopted for dynamic data in an ad-hoc manner. For example, Zhu et al. [68] used the ratio of the estimated population size of case and control populations to compute test statistic for each window. Burke et al. [11] applied Fisher exact test to the last observation of data on case and control populations. Bergland et al. [7] applied F_{st} to populations throughout time to signify their differentiation from ancestral as well as geographically different populations. Jha et al. [28] computed test statistic of generalized linear-mixed model (GLMM) directly from read counts.

Dynamic Data (general). To perform statistical test on time-series data directly, Bollback et al. [9] provided diffusion approximation to the continues Wright Fisher Markov process and estimated s numerically. Then, they tested likelihood ration on $\mathcal{X}^2$ distribution.

Feder et al. [20] proposed Frequency Increment Test(FIT) for dynamic frequency data and Empirical Likelihood Ratio Test (ELRT) instead of $\mathcal{X}^2$ test which performs poorly when the number of independent samples (replicates) are small. Specifically, FIT is a neutrality test which uses (continuous-time continuous-state) Brownian motion process for modeling variation of allele frequencies under genetic drift. Basically, given current allele frequency ν_t at a site, Brownian motion approximation assumes future-generations allele frequencies are drawn from the Gaussian distribution

$$\nu_{t+\tau} \sim \mathcal{N}\left(\nu_t, \frac{2\nu_t(1-\nu_t)}{N_e}\tau\right) \tag{1}$$

[1]The extent of genomic region is mainly depend on the amount background linkage in the genome.

In other words, under neutrality, increments of the consecutive observations drawn from a Gaussian with zero mean and variance of (1), and p-value of and be readily computed via a Student's t-test. More recently, Song et al. [58] computed LRT statistic by fitting parameters to a Gaussian process model to the time series frequency data.

Notation. Let $\mathcal{V} = (\nu_{ijk}) \in [0,1]^{T\times M\times R}$ denote the population frequency where T is the number of samples in time, M is the number of segregating sites, and R is the number of replicates. Samples in time are taken such that $\tau_1 < \tau_2, \cdots < \tau_T$.

Hard and Soft Sweep. Selective sweeps are classified by the amount of variation exist in the individuals carrying adaptive allele. By definition, hard sweep is the case when all the carriers coalesce after onset of selection and thus diversity between carriers is at its minimum. In general, excess of genetic variation in carriers of adaptive allele, makes it difficult to detect selection, i.e., soft sweep is difficult to detect. In the following, we conduct and evaluate our simulations for soft and hard sweep separately. Since here we do not take into account of de novo mutations, soft sweep can only happen in standing variation where the site under selection is at frequency is larger than $1/F$. In contrast, hard sweep experiments are those that their adaptive allele is at its minim frequency, $1/F$, at the onset of selection.

Challenges. The main constraint in the adaptive experimental evolution is the sampling-time-span (STS), the number of generations between the first and last sampled generations. Given a fixed amount of time for an study, the longer generation times of the organism, the smaller number of generations can be evolved and re-sequenced. This implies that, only quite strong selection pressures, STS will be of the same order of fixation time. As a result, the time series data usually only captures a "partial sweep".

Moreover, in controlled experimental evolution experiments, populations are evolved and inbred. This scenario in which population size significantly drops from the large number of wild type (e.g.D. melanogaster $N_e \approx 10^6$) to a small number (typically F is between 100-1000) of founder lines for EE, resembles a severe population bottleneck. Such a intense reduction in effective population size increases the variance of bionomial sampling in the Wright-Fisher model, and consequently makes genetic drift quite strong. Also, signal of selection in SFS is absorbed by such a strong bottleneck and makes it harder to identify selection.

2 Materials and Methods

2.1 The COMALE statistic

Consider a locus with starting derived allele frequency ν_0. Frequencies are sampled at T distinct generations specified by $\mathcal{T} = \{\tau_i : 1 \leq \tau_1 < \tau_2, \ldots \leq \tau_T\}$, and denoted by $\boldsymbol{\nu} = \{\nu_1, \ldots, \nu_T\}$. Moreover, R replicate measurements are made, and we denote the r-th replicate frequency data as $\boldsymbol{\nu}^{(r)}$.

To identify if the locus is evolving under positive selection, we follow previous approaches to focus on a parametrized likelihood based model that (a) maximizes the likelihood of the time series data under selection; and, (b) computes the log odds of the likelihood of selection compared to the likelihood of neutral evolution/drift.

To model neutral evolution, it is natural to model the change in frequency ν_t over time via Brownian motion or Gaussian process (e.g., FIT []). 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 population sizes and low starting frequencies that are typical in experimental evolution (see Results, and Figure **??**).

Instead, we use a discrete-time discrete-state-space Wright-Fisher Markov Chain with transition matrix P for population of N diploid individuals [18], 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:

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

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

$$\tag{4}$$

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

$$\hat{\nu}_{t^+} = \mathbb{E}[\nu_{t^+}(s,h,\nu_t)] = \frac{w_{11}\nu_t^2 + w_{01}\nu_t(1-\nu_t)}{w_{11}\nu_t^2 + 2w_{01}\nu_t(1-\nu_t) + w_{00}(1-\nu_t)^2} = \nu_t + \frac{s(h+(1-2h)\nu_t)\nu_t(1-\nu_t)}{1+s\nu_t(2h+(1-2h)\nu_t))}. \tag{5}$$

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 [18], eq 1.58-59, page 24):

$$Q_{s,h}^{(1)}[i,j] = \Pr\left(\nu_{t^+} = \frac{j}{2N} \middle| \nu_t = \frac{i}{2N}; s, h\right) = \binom{2N}{j} \hat{\nu}_{t^+}^j (1-\hat{\nu}_{t^+})^{2N-j} \tag{6}$$

$$Q_{s,h}^{(\tau)} = Q_{s,h}^{(\tau-1)} Q_{s,h}^{(1)} \tag{7}$$

The likelihood of observing the trajectory $\boldsymbol{\nu}$ is computed using:

$$\mathcal{L}_M(s,h|\boldsymbol{\nu}) = \Pr(\boldsymbol{\nu}; \nu_0, s, h) = \prod_{t=1}^{T} \Pr(\nu_t|\nu_{t-1}; \nu_0, s, h) = \prod_{t=1}^{T} Q_{s,h}^{(\delta_t)}[\hat{i}, \hat{j}], \tag{8}$$

where, $(\hat{i}, \hat{j}) = (\lfloor 2N\nu_{t-1} \rfloor, \lfloor 2N\nu_t \rfloor)$, and $\delta_t = \tau_t - \tau_{t-1}$. Moreover, for R independent replicates,

$$\mathcal{L}_M(s,h|\{\boldsymbol{\nu}^{(r)}\}) = \prod_r \mathcal{L}_M(s,h|\boldsymbol{\nu}^{(r)}) \tag{9}$$

In its simplest form, the COMALE test statistic is given by

$$\mathcal{M} = \operatorname{sgn}(s^*) \log\left(\frac{\mathcal{L}_M(s^*,h^*|\{\boldsymbol{\nu}^{(r)}\})}{\mathcal{L}_M(0,0|\{\boldsymbol{\nu}^{(r)}\})}\right). \tag{10}$$

In the following, we extend COMALE to handle cases where mapped reads are used to estimate frequencies, and second, to compute composite likelihood scores in a genomic window.

COMALE statistic for sequence data. When sequencing coverage is low, and particularly for extreme frequencies, or when the read depths vary across different replicates and generations of a locus, computing allele frequencies reliably is challenging. Moreover, even in a high coverage experiment ascertainment bias in computing SNP frequencies can have nontrivial impact [**?**]. To incorporate uncertainties in allele frequency estimates, we extend the Markov chain in Eq. 9,10 to a Hidden Markov Model.

Instead of sampling frequencies, we sample a sequence of tuples $\mathbf{x} = x_1, x_2, \ldots, x_T$, where x_t is sampled at time $\tau_t \in \mathcal{T}$. $x_t = \langle c_t, d_t \rangle$ where d_t, c_t represent the read depth, and the count of the derived allele, respectively, at time τ_t. Consider an HMM H, with the state space defined by the tuples $\langle t, i \rangle$, where $1 \leq t \leq T$, $1 \leq i \leq 2N$. State $\langle t, i \rangle$ emits x with probability independent of τ_t. Specifically,

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

For $1 \leq t \leq T$, let $\alpha_{t,i}$ denote the probability of emitting $x_1, x_2, \ldots, x_t$ and ending in state $\langle t, i \rangle$. $\boldsymbol{\alpha}$ can be computed using dynamic programming [**?**].

$$\alpha_{t,i} = \left(\sum_{1 \leq j \leq 2N} \alpha_{t-1,j}\, Q_{s,h}^{(\delta_t)}[j, i]\right) \mathbf{e}_i(x_t) \ . \tag{11}$$

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

$$\mathcal{L}_H(s, h|\{\boldsymbol{x}^{(r)}\}) = \prod_{r=1}^{R} \mathcal{L}_H(s, h|\mathbf{x}^{(r)}) = \prod_{r=1}^{R} \sum_i \alpha_{T,i}^{(r)} \ . \tag{12}$$

The HMM based COMALE test statistic $\mathcal{H}$ is computed similar to Eq. 10.

$$\mathcal{H} \quad = \quad \mathrm{sgn}(s^*) \log\left(\frac{\mathcal{L}_{\mathcal{H}}(s^*, h^*|\{\boldsymbol{\nu}^{(r)}\})}{\mathcal{L}_{\mathcal{H}}(0, 0|\{\boldsymbol{\nu}^{(r)}\})}\right). \tag{13}$$

Composite Likelihood. Consider a small genomic window defined by a collection of segregating sites L, with little or no recombination between sites. The frequency of a site $\ell \in L$ is governed by the selection, drift, *and* linkage with other sites. However, modeling the linkage is computationally expensive. Instead, we combine the likelihood ratio scores of all sites $\ell \in L$ to get a Composite Likelihood Ratio (CLR) for the region [40, 61, 65]. Specifically, we compute the test statistic $\mathcal{M}(\ell)$ (or, $\mathcal{H}(\ell)$) for all sites ℓ. For percentile cut-off π, let $L_\pi \subseteq L$ denote the set of sites whose likelihood scores had percentile π or better. For all π, the CLR statistic is computed using:

$$\mathcal{M}_\pi = \frac{1}{|L_\pi|} \sum_{\ell \in L_\pi} \mathcal{M}(\ell) \ , \tag{14}$$

or (for the HMM),

$$\mathcal{H}_\pi = \frac{1}{|L_\pi|} \sum_{\ell \in L_\pi} \mathcal{H}(\ell) \ . \tag{15}$$

Note that in this notation, $\mathcal{M}_{100}$ (respectively, $\mathcal{H}_{100}$) denote the statistic for the maximum scoring site.

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

$$s^*, h^* \quad = \quad \arg\max_{s,h} \sum_{r}^{R} \log\left(\mathcal{L}_M(s, h|\boldsymbol{\nu}^{(r)}\right), \quad \text{or,} \tag{16}$$

$$s^*, h^* \quad = \quad \arg\max_{s,h} \sum_{r}^{R} \log\left(\mathcal{L}_H(s, h|\boldsymbol{x}^{(r)}\right). \tag{17}$$

where likelihoods are defined in Eq. 9 and Eq. 12, respectively. As shown in Appendix **??**, for a given h, Eq. **??** and **??** quasiconvex (unimodal) functions in s and admit efficient algorithms, (e.g., bisection) to compute the optimal in $\mathcal{O}(\log_2(1/\epsilon))$ steps for each site, where ϵ is the upper bound on the error in the solution.

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. Following Fu, 1995 [**?**], 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 test statistics test neutrality by computing differences of two estimates to check, or performing cross-population tests comparing the statistic in two different populations [1, 49**?**]. Finally, SFS can be computed with frequency information, and does not need full haplotypes.

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. However, these tests are not independent. Evans et al. [17] developed diffusion equations for evolution of SFS in time series, but they are difficult to solve. Here, we explicitly derive many of the common test statistics as functions of frequency of the favored site in the form $S_t = f_S(\nu_t, \nu_0)$, where ν_t itself can be written as a function of s, t (Eq. S5). This allows us to compute $\mathcal{L}_S(s, h; \{S_t\})$ for many SFS based statistics. Then, a likelihood ratio, similar to Eq. **??**, provides[2] a predictor for detecting selection in each window.

Tajima's D. Let D_t denote the value of Tajima's D in a hard sweep at time t. We show (Appendix 6.4), that

$$D_t = D_0 - \log(1-\nu_t)\frac{W_0}{\log(2N)} - \nu_t^2 \Pi_0 \tag{18}$$

where W_0 and Π_0 are Watterson and Tajima estimates of θ at the the initial generation.

Fay Wu's H. We show (Appendix 6.3), that the dynamics of the H statistic are directly related to average of Haplotype Allele Frequency (HAF) score [48], and can be written as a function of $\vec{u}_t$ as follows:

$$nH_t = \theta\nu_t\left(\frac{\nu_t+1}{2} - \frac{1}{(1-\nu_t)n+1}\right) + \theta(1-\nu_t)\left(\frac{n+1}{2n} - \frac{1}{(1-\nu_t)n+1}\right) \tag{19}$$

SFSelect. The SFSelect statistic was proposed by Ronen et al. [49] to predict selection by empirically learning the so-called *optimal* weights using Support Vector Machines.

Dominance. The value of the overdominance parameter can provide an insight into the kind of adaptation *using population frequency data*[3]. In fact, for $s > 0$ we have [**?**]

condition	comment
$h < 0$	underdominance
$h = 0$	recessive adaptive allele
$h = 0.5$	directional selection
$h = 1$	dominant adaptive allele
$h > 1$	overdominance

[2]The likelihood of the data to the model is the least-squares loss between model $D_t((\hat{s}))$ and the observed D.
[3]It is trivial for genotyped data.

Software. Pre-computation of 1313 transition matrices for $s \in \{-0.5, -0.49, \ldots, 0.5\}$ and $h \in \{-1, -0.75, \ldots, 2\}$ took in second on a desktop computer with a Core i7 CPU and 16GB of RAM.

P value According to the Wilks theorem [**?**] log odds ratios are asymptotically distributed according to $\mathcal{X}^2$, yet [20] shown that in the empirical distribution is better when the number of independent samples (replicates) is small. Therefore we create 10^6 neutral unlinked loci simulations with the same sampling rate and the same number of replicates.

2.3 Simulations

For each experiment a diploid population is created and evolved as follows.

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

$$\mathbb{E}[M] = \theta \sum_{i=1}^{F-1} \frac{1}{i} = 1175$$

II. **Creating initial diploid population** To implement similar setting for experimental evolution of diploid organisms, initial haplotypes first cloned to create F diploid homozygotes. Then each diploid individual is cloned N/F times to yield diploid population of size N.

III. **Forward Simulation** 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 $\mathcal{T} = \{10, 20, 30, 40, 50\}$, `simuPop` is used to perform forward simulation and compute allele frequencies for all of the R replicates. Also, to avoid spurious simulation samples, simulation results are constrained to those that the beneficial allele escapes stochastic loss of genetic drift and *establishes* in all the replicates.

3 Results

Modeling neutral trajectories in finite populations. We tested the closeness of fit for the Markov Likelihood as a model for neutral trajectories, compared to Brownian motion. We performed 150K simulations for different values of ν_0 ($\nu_0 \in \{0.005, 0.1\}$) and time τ generations $\tau \in \{1, 10, 100\}$. Figure 1 shows that Brownian motion is inadequate when ν_0 is far from 0.5, and when sampling is done after many generations $\tau > 1$. (sampling times are sparse). In most experimental evolution scenarios, a site is unlikely to have frequency close to 0.5, and the starting frequencies are usually much smaller. Moreover, sampling times are sparse, with sampling done between 10 and 100 generations in *Drosophila* experiments [41, 68].

In contrast, Figure 1A-F also shows that Markov Likelihood predictions (Eq. 7) are highly consistent with empirical data for a wide range of simulation parameters. We also tested the predictions under selection regime by conducting 100K simulations with selection strength $s = 0.1$ on a site with initial frequency $\nu_0 = 0.005$ and sampling after τ ($\tau \in \{1, 10, 100\}$) generations. The empirical and theoretical distributions tracked closely (Figure 1G-I)VB note: Give P-value of match.

Power. We compared the detection power of COMALE against Gaussian process (GP) [58], FIT [20] statistics. For each experiment, (specified with values for selection coefficient s, starting allele frequency $\vec{u}_0$, sampling time schedule $\mathcal{T}$, and number of replicates R), we conducted 1000 simulations. Half of these modeled neutral evolution and the rest were under selection. Define power of a statistic as the average true-positive rate when false-positive rate is less than 0.05 over 1000 simulations. In other words, a cutoff was set so that at most 25 of 500 neutral simulations were predicted as being under selection, in order to compute the fraction of true positives.

Before computing against other methods, we first tested the use of percentile cut-off π in computing compute $\mathcal{H}_\pi$ VB note: Put results here., and the advantage of using HMM based statistic $\mathcal{H}$ versus Markov Chain statistics ($\mathcal{M}$) for read data.

Under a range of simulation parameters, COMALE ($\mathcal{H}_{99}$) provides higher detection power compaed to FIT or GP(Figure 2).

Running Time. A COMALE does not compute full likelihoods, or explicitly model linkage between sites, the complexity of computing likelihoods is $\mathcal{O}(TR)$, and can be efficiently vectorized for multiple replicates and loci. Therefore, it is expected to be faster than other approaches like Gaussian Process (GP) [58].

We conducted 1000 simulations and measured running time for COMALE and GP. COMALE is $\sim 1000\times$ faster than single locus GP (Figure 9), while maintaining high power (Figure 2). I'm going to measure time for multiple locus GP as well, currently figure shows single locus GP which linkage isn't involved.

SFS for Detection in Natural Experimental Evolution. See figure 4 We did not show the SFS based statistics in Figure 2 as they do not perform well. In the specific mode of experimental evolution, we sample a restricted set of F founder lines. Here, $F << N_e$ (Fig. **??**) creating a severe bottleneck, and this bottleneck confounds SFS. Figure S4 demonstrates the effect of experimental evolution on different SFS statistics under neutral evolution for 1000 simulations. The mean of neutral simulations can be used to empirically filter out the effect of bottleneck in dynamic data. VB note: Natural experiment results here.

Locating the Adaptive Mutation. The secondary task in identifying selection is to locate the position of the adaptive allele. We simply consider the site with highest score in the window as the locus of the beneficial allele. For each setting of ν_0 and s, we conducted 500 simulations and computed the rank of the beneficial mutation in each simulation. We plotted the cumulative distribution of the rank in Figure 5.

COMALE works the best in hard sweep regime. For example, when $\nu_0 = 0.005$ and $s = 0.01$ (weakest selection), the beneficial allele was ranked first in more than 60 experiments, and was ranked ≤ 5 in over 90% of experiments of hard sweep. The accuracy decreased of locating the adaptive allele diminished in soft-sweep scenarios (larger values of ν_0). Yet, in the worst case the beneficial allele was ranked among top 50 SNPs.

Strength of Selection. As the likelihood calculation is model based, we can also output the model parameters $\hat{s}$ that maximized likelihood (**??**). We computed bias, $s-\hat{s}$ for each experiment of COMALE , and GP. The distribution of the bias is presented in Figure 6 for different configurations. In general, both GP and COMALE have biased results for weak selections, where genetic drift dominates. However, for stronger sweeps, e.g.$s = 0.1$, COMALE provides estimates with smaller bias and variance.

SFS based statistics in time series. We also considered SFS based tests including Tajima's D, Fay & Wu's H and SFSelect for detecting selection. As these tests work only on static data,

we extend them for time series data for both null and alternative hypotheses. More precisely, we explicitly defined functionals D_t and H_t as function of initial carrier frequency ν_0 and strength of selection s (see section 2.2 for details). To show that the proposed models of D_t and H_t are valid models, we simulated 1000 populations and computed SFS based statistics every 10 generation and compared them with the proposed model. As shown in the Figures S1 proposed models are more consistent with data.

SFS based tests will fail when carrier frequency ν_t is low. Importantly, Figures S1 shows the increase in variance of the trajectories associated with SFS, compared to carrier frequency. Moreover, it is difficult to distinguish SFS trajectories of selection and neutral populations, when carrier frequency is small, e.g. first 50 generations of Figures S1. This, observation can be verified by examining the terms in the functional form of D_t. As shown in the Figure S3 right, in early generations of hard sweep where carrier frequency is low, D_t is either positive or close zero. In other words, the reduction in diversity become significant when carrier frequency is high enough. It can be shown that this argument holds for H_t in early generations of hard sweep, where carrier frequency is not high.

3.1 Analysis of Real Data

We finally apply COMALE method to the controlled experimental evolution experiment of [41], which evolves 5 replicates of a population of Drosophila melanogaster for 37 generations under alternating 12-hour cycles of hot (28°C) and cold (18°C) temperatures. Three replicates are sampled at the firs generation, 2 replicates at generation 15, one replicate at generation 23, one replicate at generation 27 and three replicates at generation 59.

Heterogenocity of read depths. COMALE statistic is basically computed using site allele frequencies at different generations for all replicates. For real Pool-Seq data, however, allele frequencies is unknown and only read counts at each site measured. As shown in Figure **??** read depths are highly heterogeneousa and filtering low coverage sites from data strictly reduces the number of SNPs. For example, by setting minimum read depth at each site (for all replicates and generations), the number of SNPs shrinks from 1,544,374 to 10,387.

Uncertainity in allele frequency. In addition to coverage heterogenocity, allele frequencies in principle are *hidden* variables for *observed* data, alleles read counts. We take into account of allele frequency uncertainty by computing the likelihood of data using Hidden Markov Model (HMM), see section **??** for details.

SNP-based analysis with ELRT statistic Manhattan plots in Figure 7 depicts the distribution of the top 2000 SNPs and corresponding 1961 genes. The list of genes is available in the same directory under name `genes_SNP.txt`.

The following genes are enriched with P-value of 0.0006725 using Fisher exact test and genes of 6.1.

Window-based analysis with COMALE statistic Manhattan plots in Figure 7 depicts the distribution of the top 100 50K regions and corresponding 421 genes. The list of genes is available in the same directory under name `genes_SNP.txt`.

4 Discussion

5 Figures

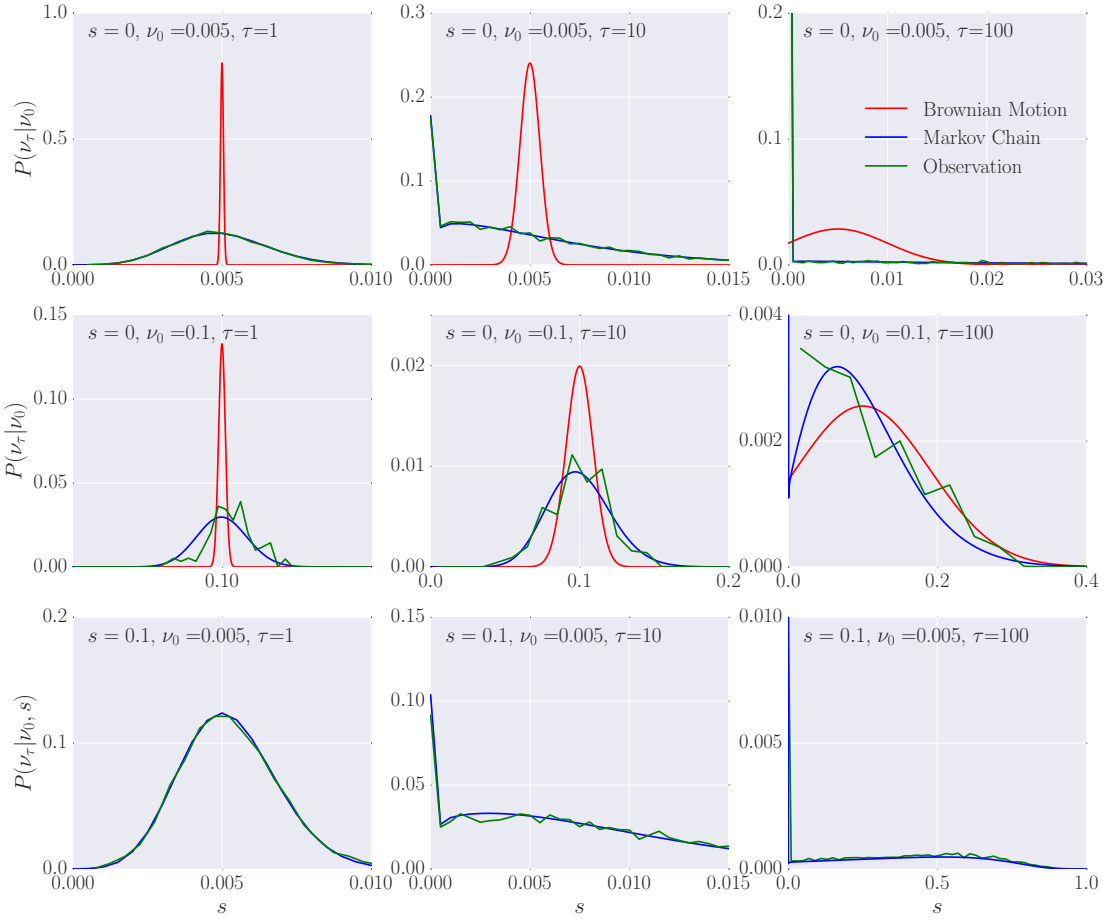

Figure 1: Comparison of empirical distributions of allele frequencies (green) versus predictions from Brownian Motion (red), and Markov Likelihoods. Panels A-F: Experiments were conducted under neutral evolution with different starting frequencies $\nu_0 \in \{0.005, 0.1\}$ and sampling times $\tau \in \{1, 10, 100\}$ generations. The empirical distribution was computed by sampling 143,900 sites with $\nu_0 = 0.005$ and 47,500 with variants $\nu_0 = 0.1$. computed from neutrally evolving simulations is depicted in green lines. Panels G,H,I: Comparisons of Empirical and Markov chain based predicted allele frequency value distributions under a selection regime with $s = 0.1$. Initial frequency was chosen as $\nu_0 = 0.005$ and sampling performed after $t = \{1, 10, 100\}$ generations.

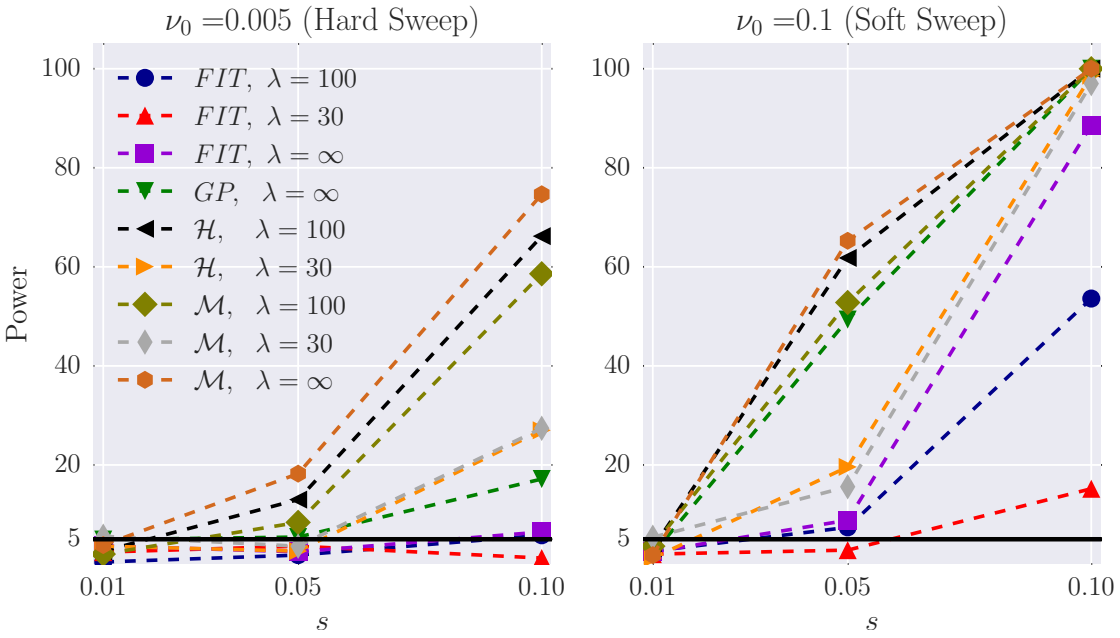

Figure 2: Predictive performance of different method is evaluated on 1000 simulations for different values of selection strength s and initial carrier frequency ν_0.

Method	Depth Rate	Hard Sweep	Soft Sweep
$\mathcal{M}$	∞	32.2	55.7
GP	∞	9.2	50.4
FIT	∞	4.2	33.3
$\mathcal{M}$	30	12.2	39.3
$\mathcal{H}$	30	11.1	39.8
FIT	30	2.4	6.7
$\mathcal{H}$	100	27.3	55.0
$\mathcal{M}$	100	23.0	52.1
FIT	100	2.7	21.2

Table 1: Power of methods

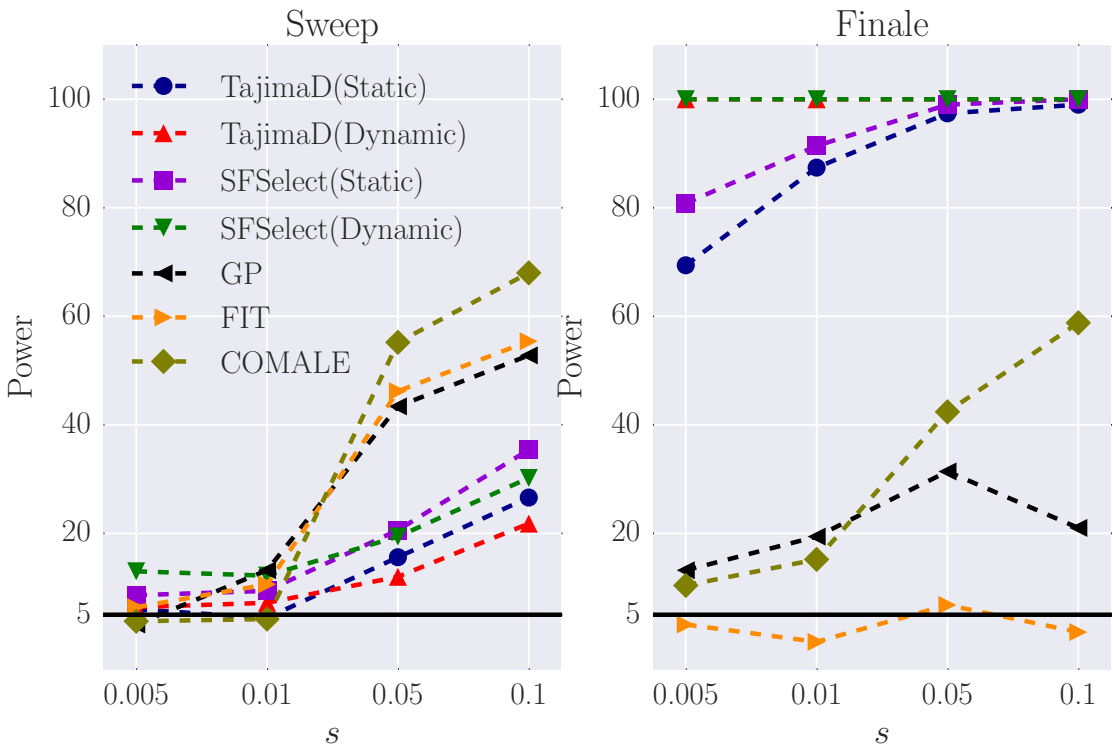

Figure 3: Predictive performance of different method is evaluated on 400 simulations for different values of selection strength s and initial carrier frequency ν_0.

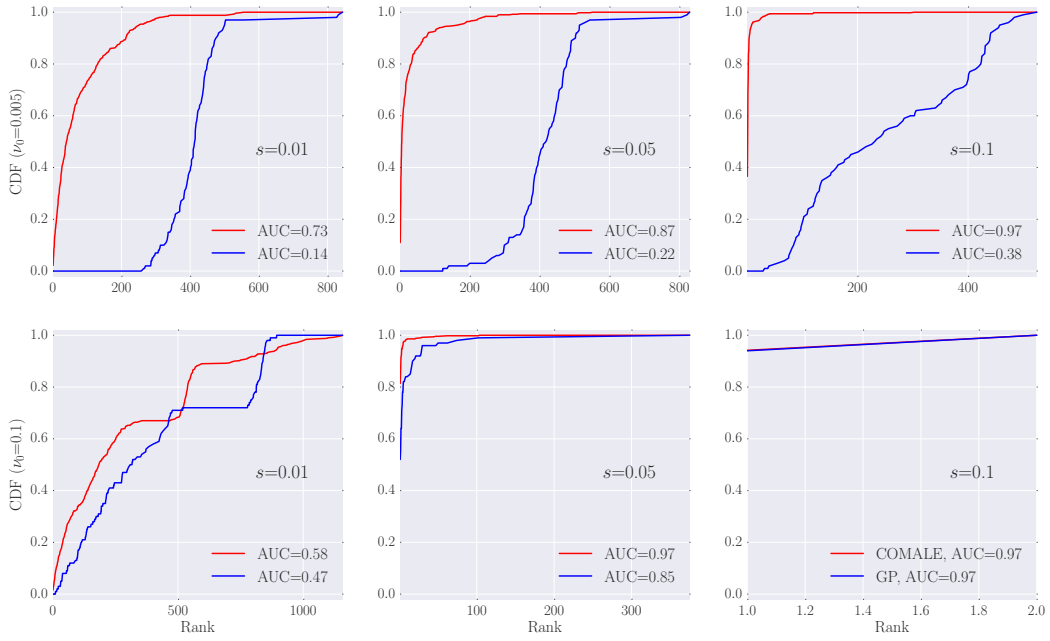

Figure 4: CDF of the rank of the adaptive allele in 100 simulations.

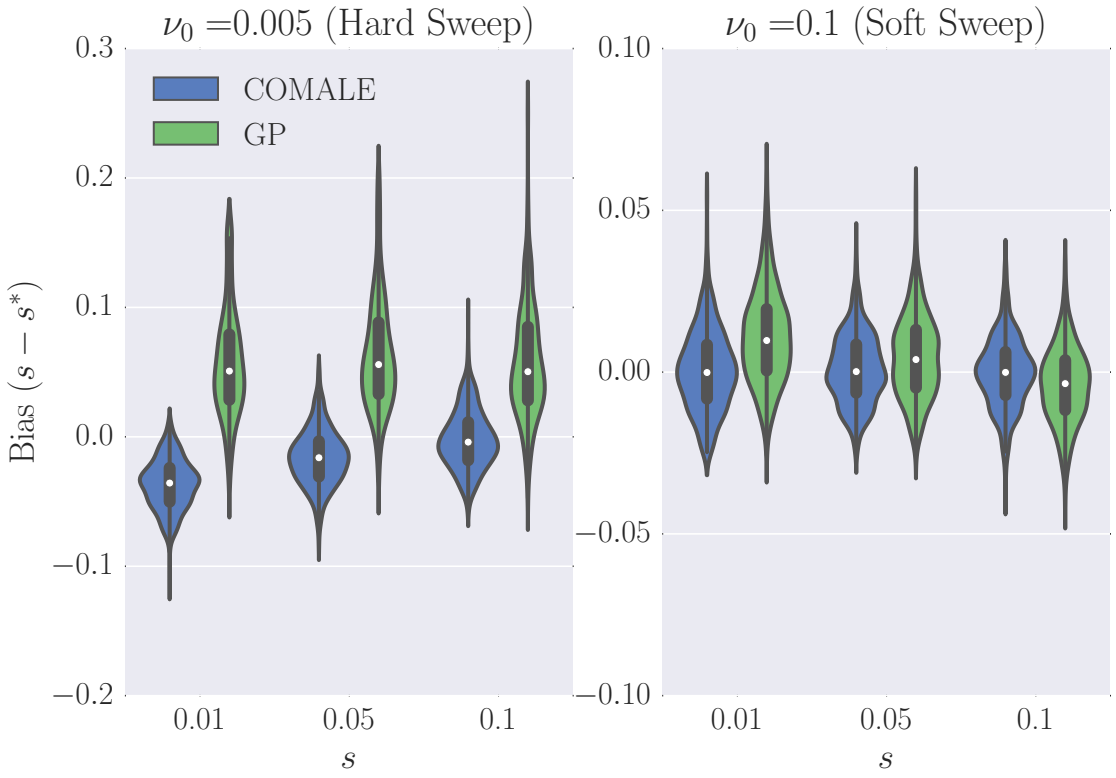

Figure 5

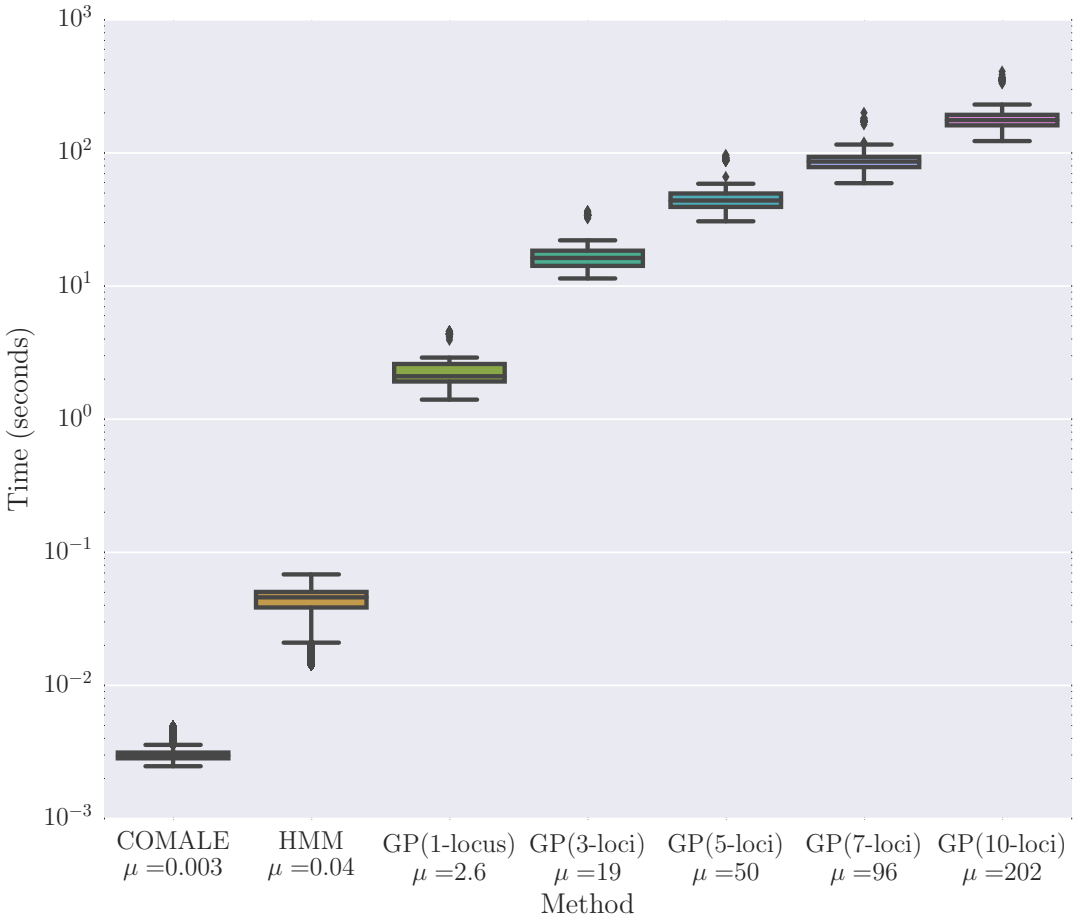

Figure 6: Average running time of COMALE , HMM, GP with single,3,5,7 and 10 loci over 1000 simulations.

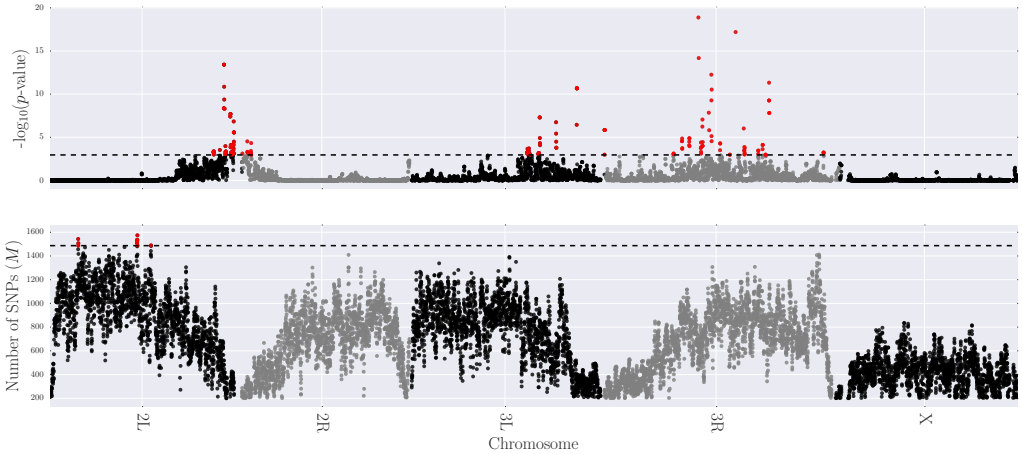

Figure 7: To be added.

6 Appendix

6.1 Allele frequencies under selective sweep

(S1)

where $s \in \mathbb{R}$ is the selection coefficient and $o \in [0, 1]$ is the overdominance parameter which for $u = 0.5$ we have

$$x_{t+1} = x_t + \frac{sx_t(1 - x_t)}{2 + 2sx_t} \, . \tag{S2}$$

we also have

$$\frac{\mathbf{d}x_t}{\mathbf{d}t} = \frac{sx_t(1 - x_t)}{2 + 2sx_t} \tag{S3}$$

which is a differential equation that is difficult to solve. However if take the approximation $2+2sx_t \approx 2$, it becomes an ordinary differential equation that can be readily solved

$$\nu_t = \frac{1}{1 + \frac{1-x_0}{x_0} e^{-st/2}} = \sigma(st/2 + \eta(x_0)) \tag{S4}$$

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

6.2 Logistic Model for Selection.

As maximum likelihood for estimating s using Markov chain is computationally expensive, here we use and logistic function for modeling allele frequencies undergoing selective sweep. In a pure selection process with no drift, i.e. infinite population size, dynamic of allele frequencies can be well approximated by the logistic function (see Appendix 6.1 for derivation)

$$\nu_t = \sigma(st + \eta(\nu_0)) \tag{S5}$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the logistic function, and $\eta(x) = \log(x)/\log(1 - x)$ is inverse of the logistic, aka logit, function. Figure S2 depicts the behavior of the logistic model for site allele frequencies and the default sampling time span(STS). Without genetic drift, soft sweep is easier to detect, because the logistic function happens to have steeper slope in th STS than those of hard sweeps, due to standing variation frequency. In addition, even under infinite population size regime, it is difficult differentiate between weak selections and genetic drift (a horizontal line), in short (e.g. 50 generations) experimental evolutions.

As shown in Figures **??** and **??** (first row), the approximate logistic model is consistent with simulated data and we use it to estimate the strength of selection for each site by solving a linear system of equations(see Section 2.1 for details).

6.3 Fay Wu's H

In any finite population size of n with m segregating sites, allele frequencies take discrete values, i.e., $x_j \in \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\}$, $\forall j \in 1, \cdot, m$ and we can write

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{m} x_j^2 = \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^2 \xi_i = \frac{(n-1)}{2n} H \tag{S6}$$

where ξ_i is the number of sites with frequency i/n and H is the Fay & Wu's estimate of θ.

Recently, Ronen et al. [48] devised the 1-HAF statistic for identifying selection on static data, which has the expected value related to the $\|\mathbf{x}\|^2$:

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

where

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

where easily follows that

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

6.4 Tajima's D

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 if individuals, then $\Pi = \frac{1}{n^2}\sum P_{ij}$. 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:

$$\Pi_t = (1-\nu_t^2)\Pi_0 \tag{S10}$$

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

$$\frac{W_t}{W_0} = \frac{\frac{m_t}{S}}{\frac{m_0}{S}} \quad \Rightarrow W_t = \frac{m_t}{m_0}W0 \tag{S11}$$

where m_t to be interpreted as the expected number of segregating sites at time t, under neutral evolution. At time t, the number of individuals that undergone neutral evolution is $(1-\nu_t)n+1$, which leads to

$$\frac{m_t}{m_0} = \frac{\log\left((1-\nu_t)n+1\right)\theta}{\log(n)\theta} \approx \frac{\log\left((1-\nu_t)n\right)}{\log(n)} = \frac{\log(1-\nu_t)+\log(n)}{\log(n)} = 1 + \frac{\log(1-\nu_t)}{\log(n)} \tag{S12}$$

putting all together

$$D_t = (1-\nu_t^2)\Pi_0 - (1+\frac{\log(1-\nu_t)}{\log(n)})W_0 = D_0 - \log(1-\nu_t)\frac{W_0}{\log(n)} - \nu_t^2\Pi_0 \tag{S13}$$

6.5 Linkage Disequilibrium

Nonrandom associations between polymorphisms are established in the substitution process according to the phylogeny, broken by recombination events and reinforced by selection. Although in EE the experiments with pooled sequencing, LD can not be measured throughout evolution, it is still worthwhile to examine the behavior of LD as a result of the interaction between recombination and natural selection, to take into account of some of EE implicit constraints.

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

$$\rho_t = \alpha_t\beta_t\rho_0 = e^{-rtl}\left(\frac{H_t}{H0}\right)\rho_0 \tag{S14}$$

where $H_T = 2\nu_0(1-\nu_0)$ is the heterozygousity at the selected site, r is the recombination rate/bp/gen. The decay factor $\alpha_t = e^{-rtl}$ is the product of recombination and growth factor β_t (eq. 30-31 in [56])is the outcome of selection. For $s = 0.01$, $l = 100K$bp, the log of decay, growth and product of both is depicted in Figure S6. It is evident that, for these parameters LD does not start to decay until generation 1000, which would be problematic when ρ_0. For example, in the case of hard sweep, the selection is imposed on the site with minimum AF, which is at perfect linkage ($|D'| = 1$) with all the other loci.[4] We this phenomenon is shown in the Figures S7, S8 where at generation zero the site at position 500K is at perfect linkage with all the other sites, and linkage of the middle site with all the genome is depicted for both genetic drift and natural selection, in different generations. Also, a window of 50Kbp around the selected site is shaded in Figure S7 to demonstrate the value of LD in the window under drift and hard sweep. This implies that the precision of locating the selection on the genome is tightly dependent on a set of parameters including, recombination rate, selection strength, initial carrier frequency, and the initial linkage.

6.6 Likelihood Functions

Actually, it is enough to show (S2) is linear fractional in s, which is nby definition: linear numerator and denominator, and strictly positive denominator.

[4] This is because, between the selected site and all the other sites frequency of one gamete is zero.

Supplemental Figures

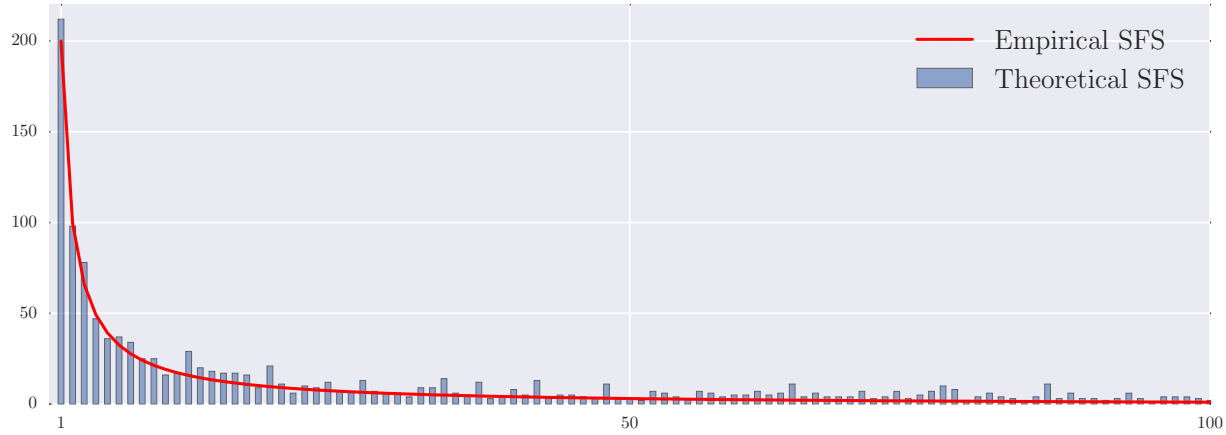

Figure S1: Theoretical and Empirical SFS for a neutral population of 200 individuals.

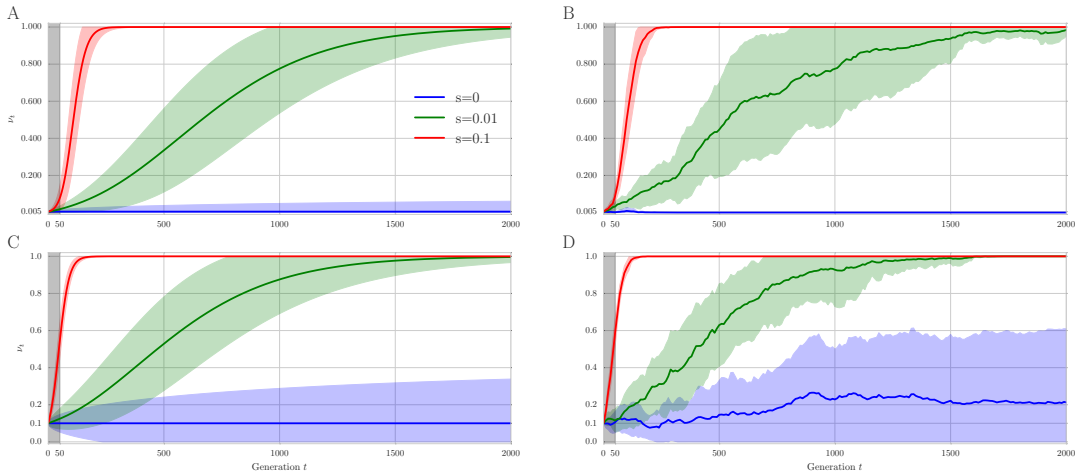

Figure S2: Logistic model for different selection strengths for soft (left) and hard (right) sweep as a function of time in generations. The first 50 generations, which observations are sampled is shaded.

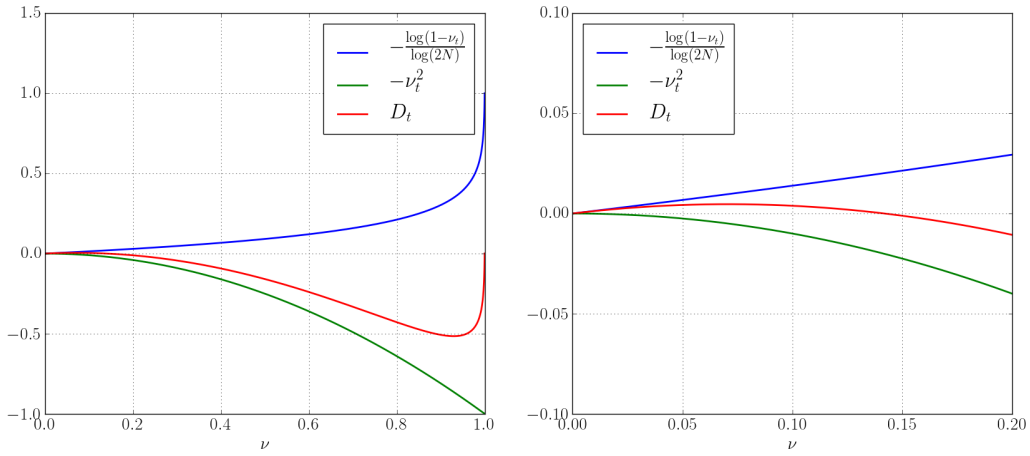

Figure S3: Interactions of two terms in D. W.l.o.g when $D_0 = 0$ and $\Pi_0 = W_0 = 1$, D_t is sum of the logarithmic $\frac{-\log(1-\nu_t)}{\log(2N)}$ and the squared term ν_t^2.

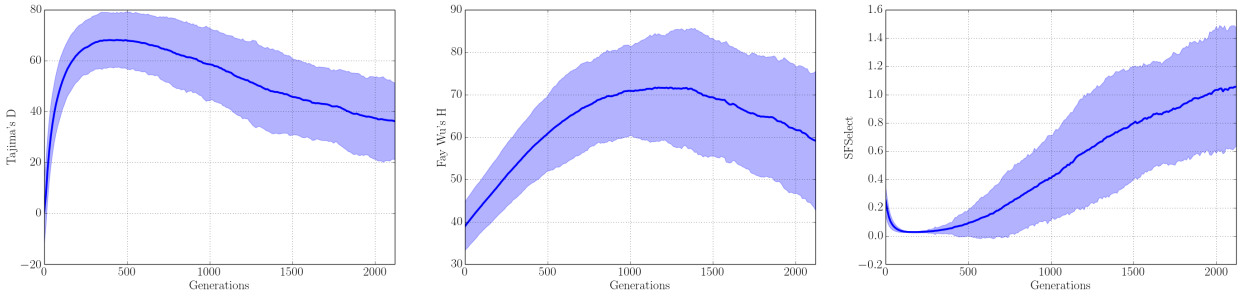

Figure S4: Effect of bottle neck in a typical experimental evoloution experiment where a restricted number of founder lines (here $F = 200$) is selected out of a larger population size ($N_e = 10^{-6}$). Tajima's D (left), Fay Wu's H (middle) and SFSelect is computed for 1000 neutral simulations and mean and 95% confidence interval is plotted.

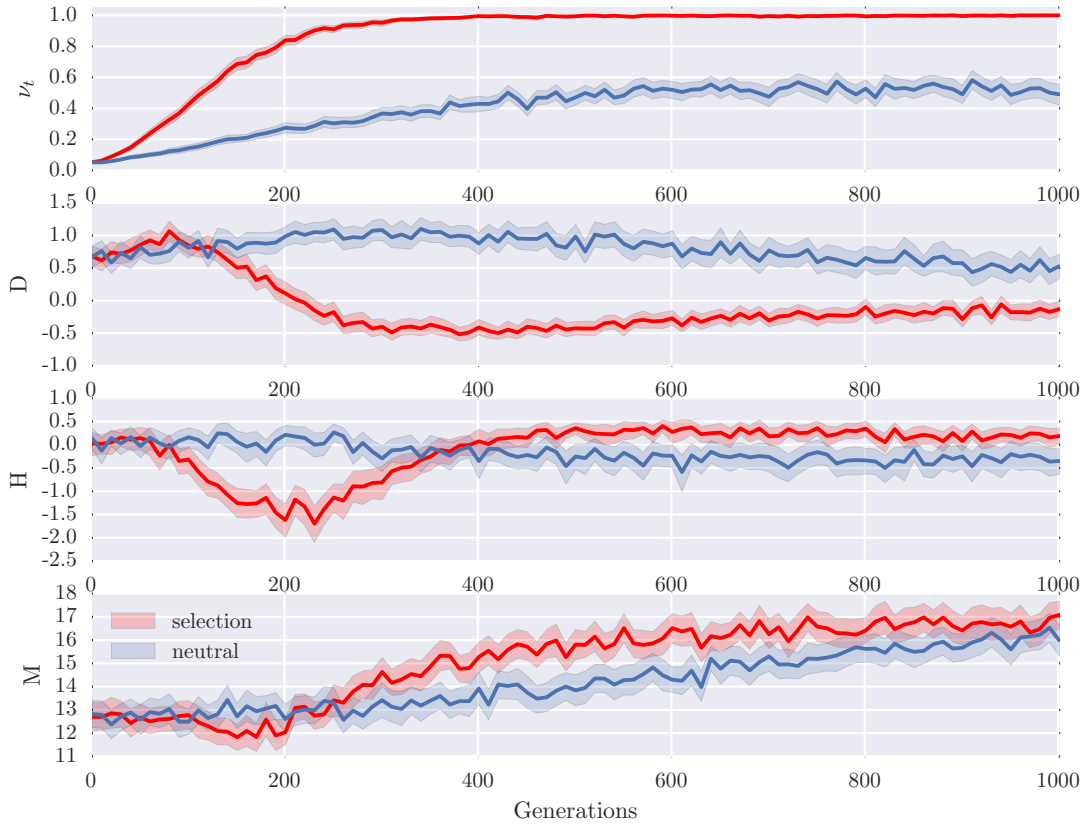

Figure S5: Mean and 95% CI of 1000 simulations for neutral (blue trajectories) selection with $s = 0.1$ (red trajectories).

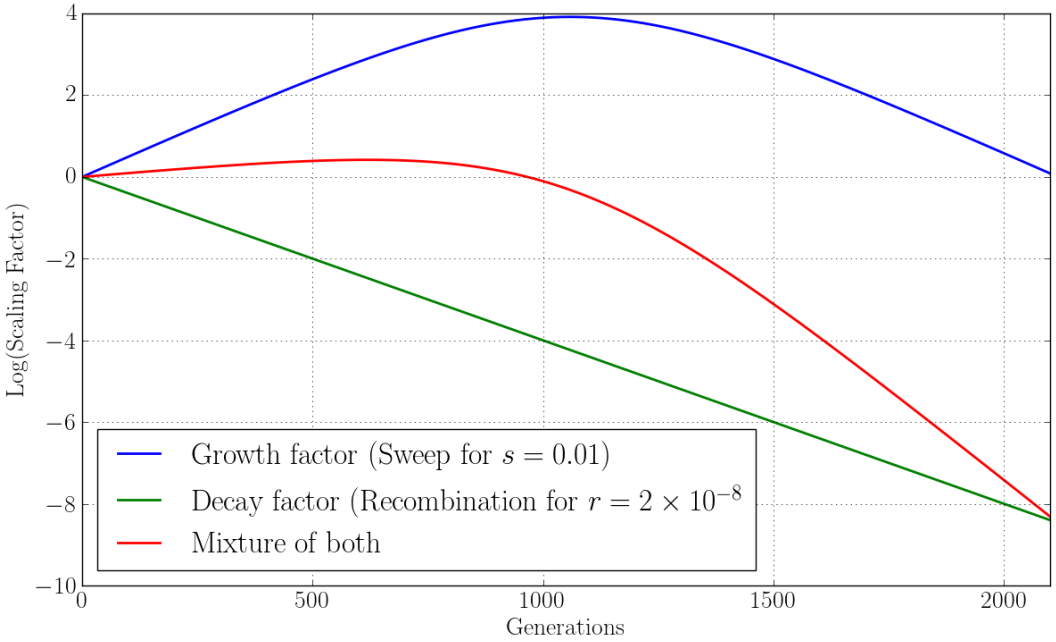

Figure S6: Interaction between productive factors of LD under natural selection for weak selection (s=0.01) and a distance of 100Kb between sites. In this setting, after about 1000 generations LD start to decay (red curve).

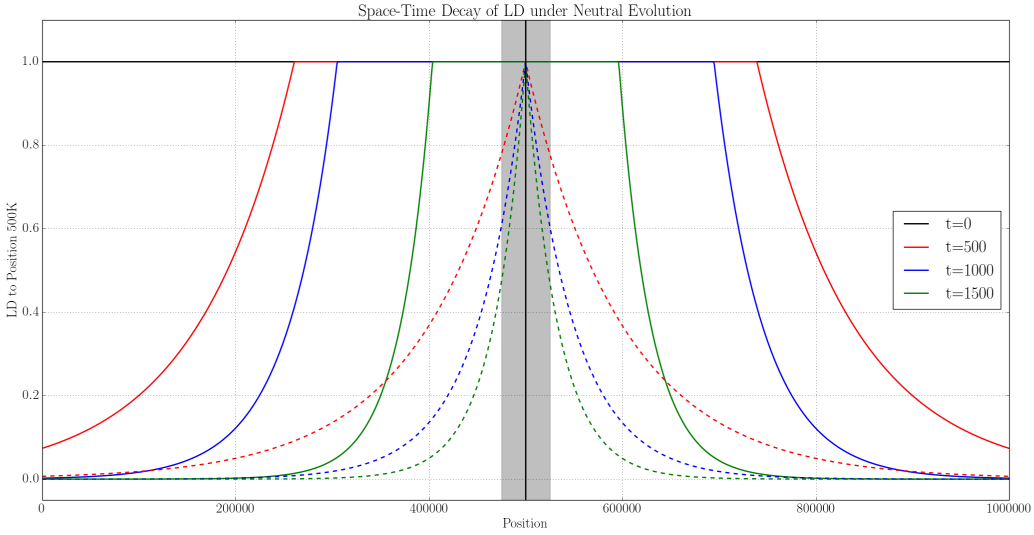

Figure S7: Decay of LD ($|D'|$ measure) of the minimum AF site at position 500K with the rest of genome when $s = 0.01$ and $r = 2 \times 10^{-8}$. A window of 50Kb is shaded at the center of genome to illustrate high values of linkage in both selection and drift.

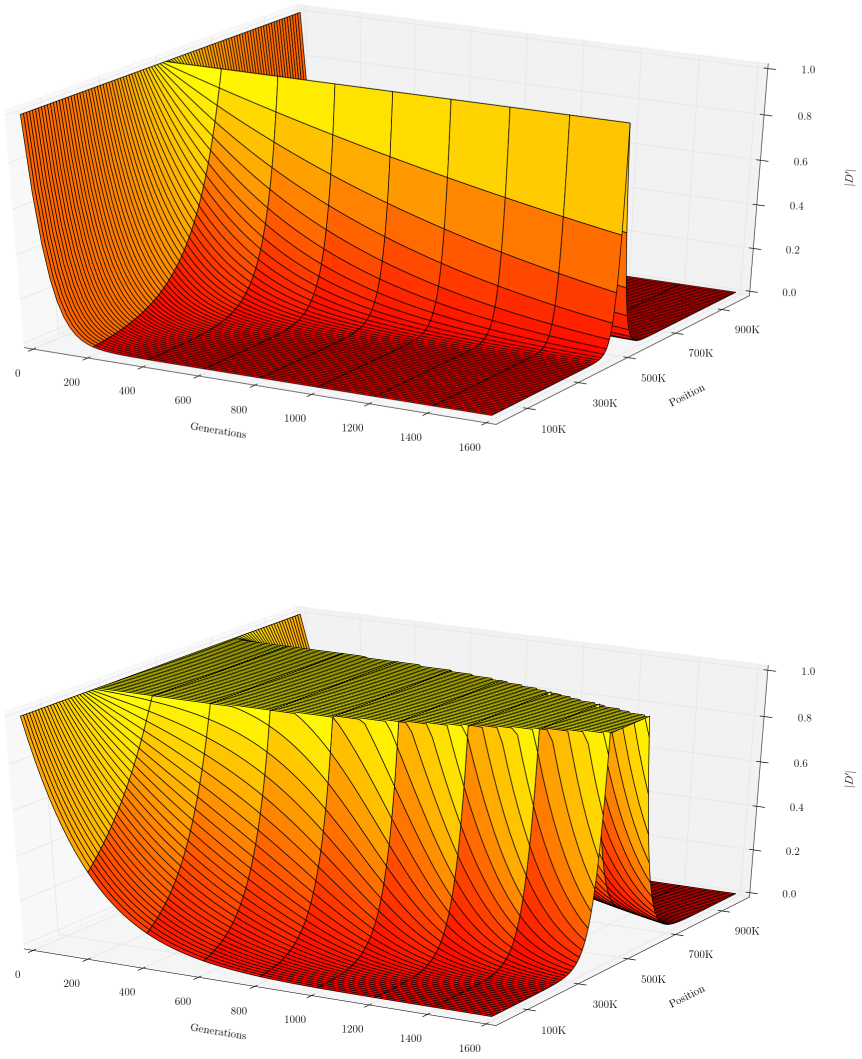

Figure S8: ld

Figure S9: Decay of LD ($|D'|$ measure) of the minimum AF site at position 500K with the rest of genome in genetic drift with $r = 2 \times 10^{-8}$ (top) and hard sweep with $s = 0.01$ (bottom).

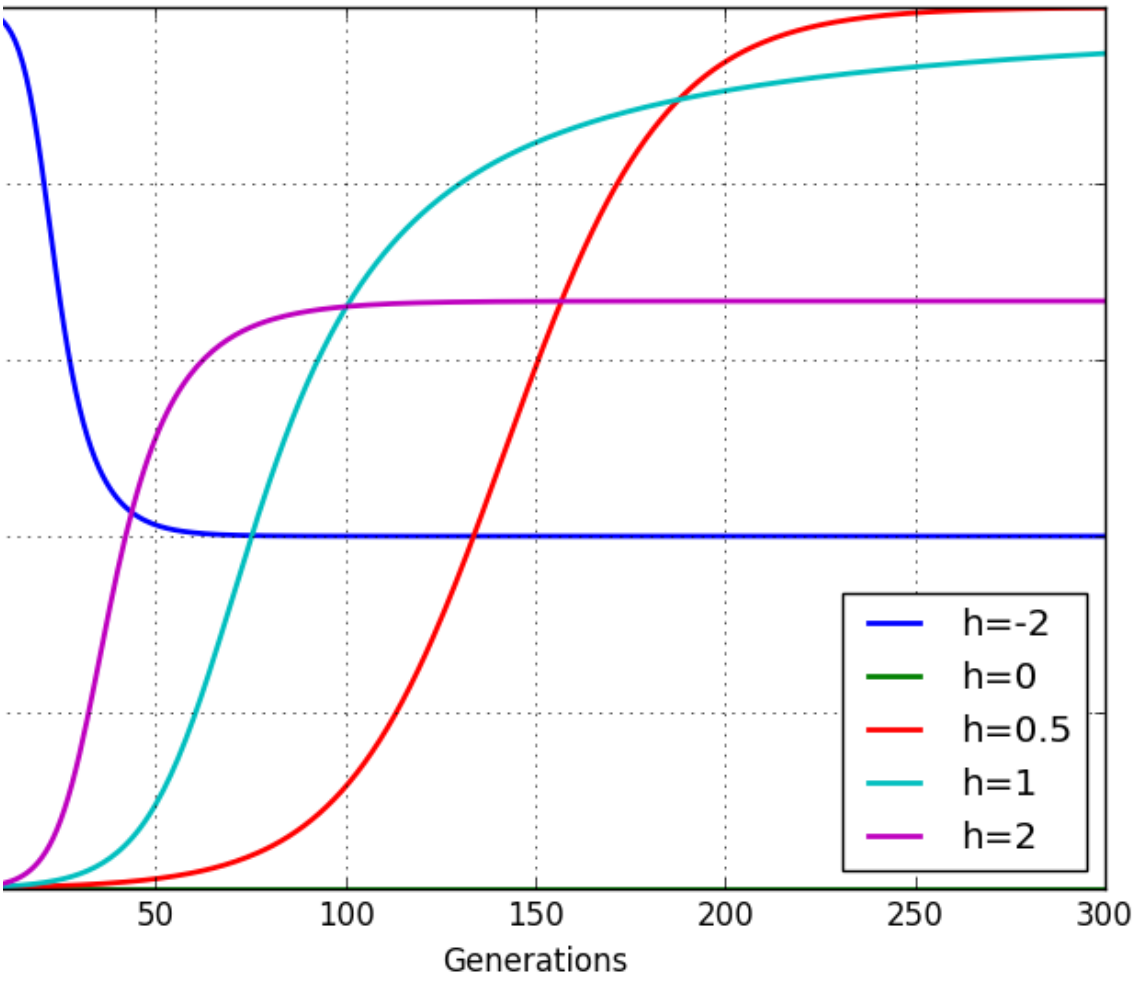

Figure S10: Dominance for $s = 0.1$.

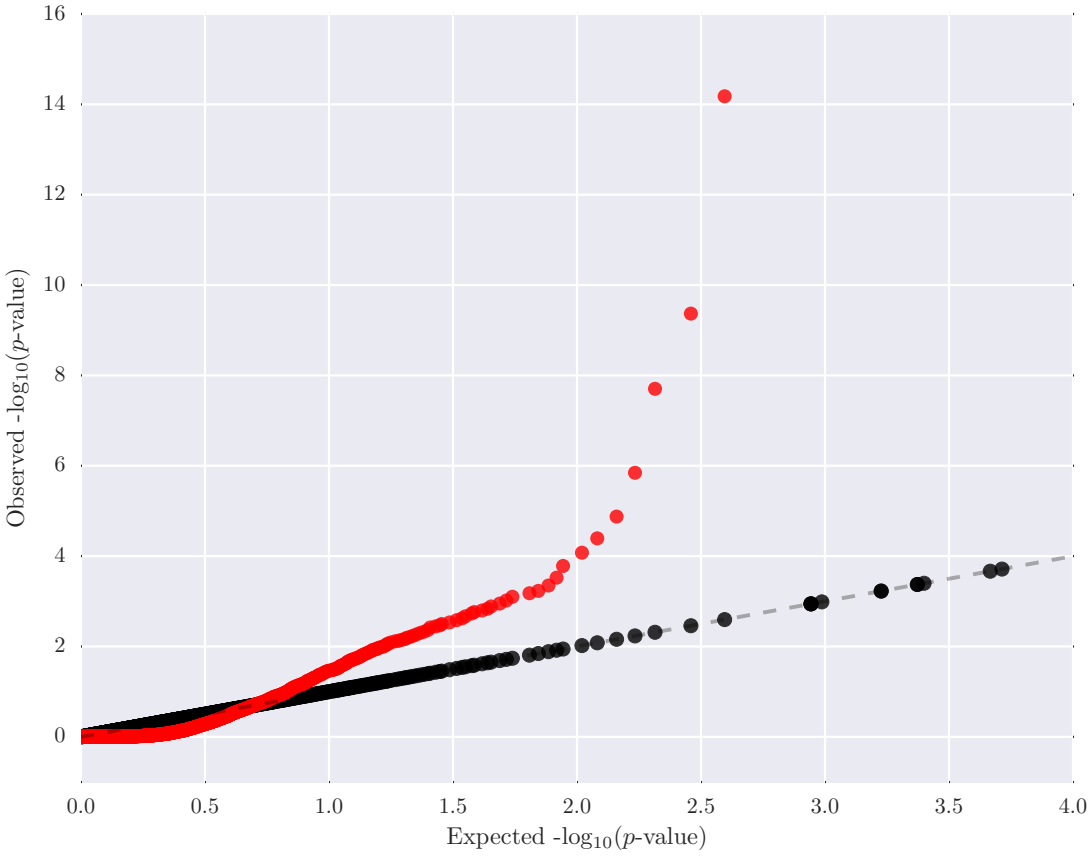

Figure S11: QQ plot for a simulation.

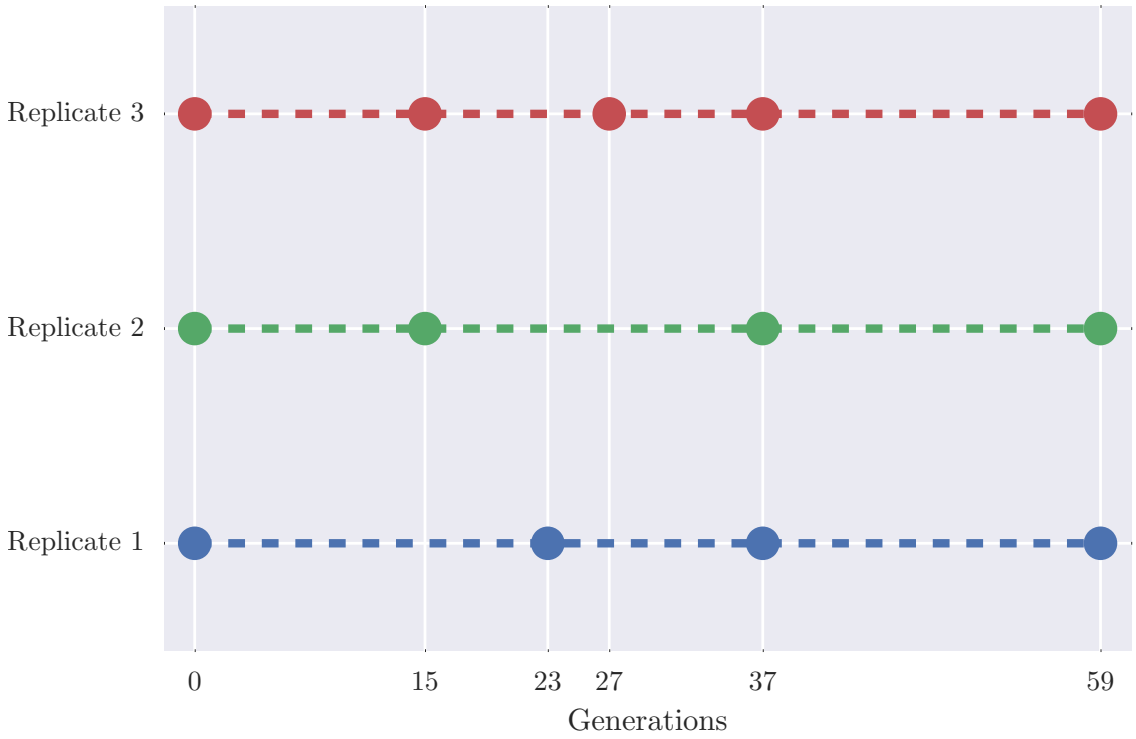

Figure S12: sampling times.

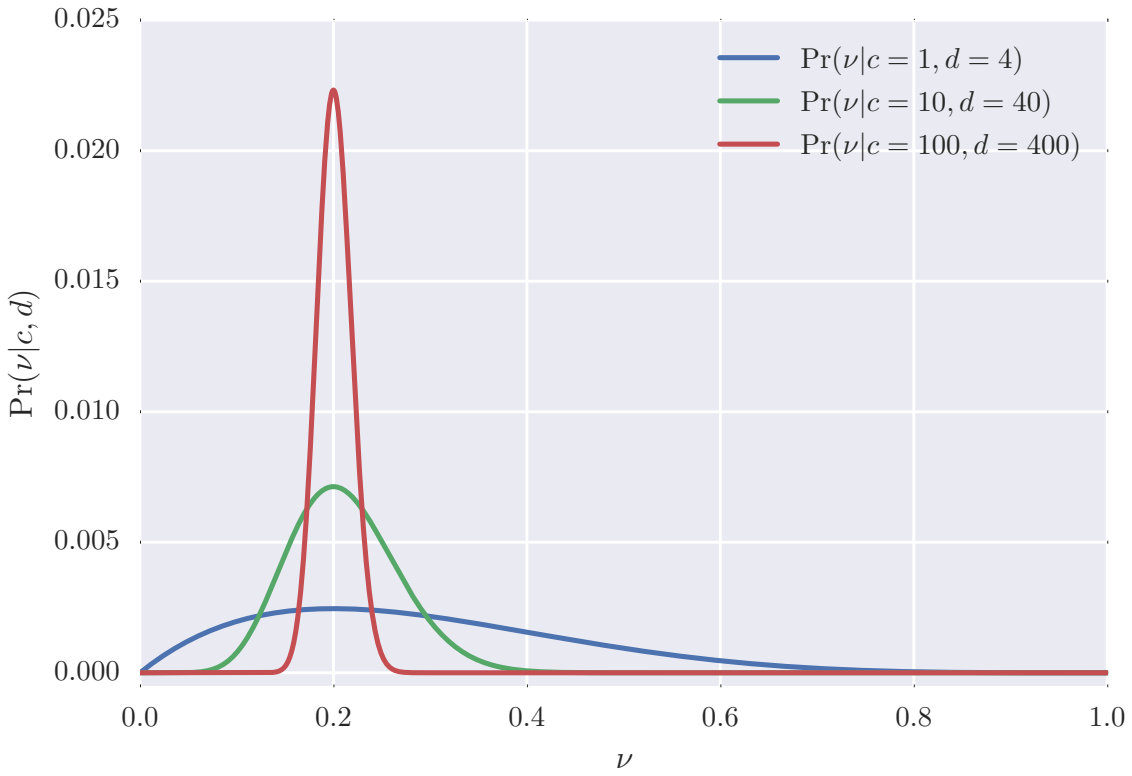

Figure S13: observation state transition.

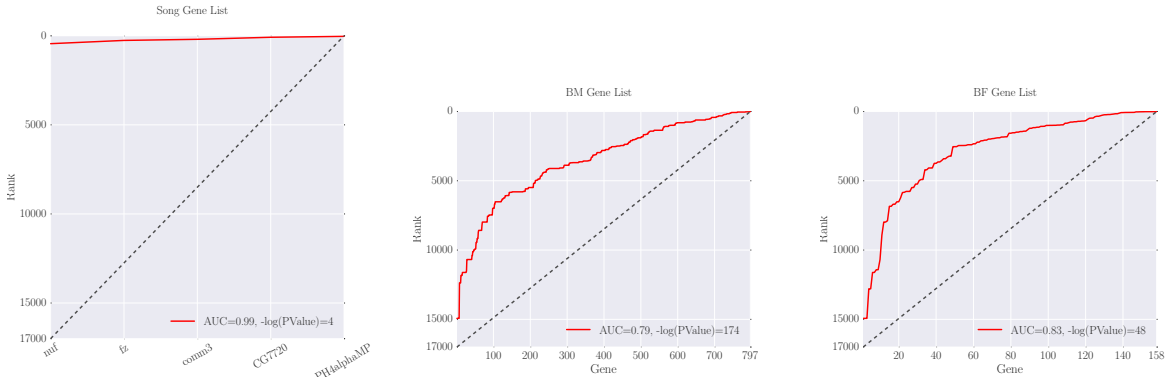

Figure S14: Song, HSF, HSP.

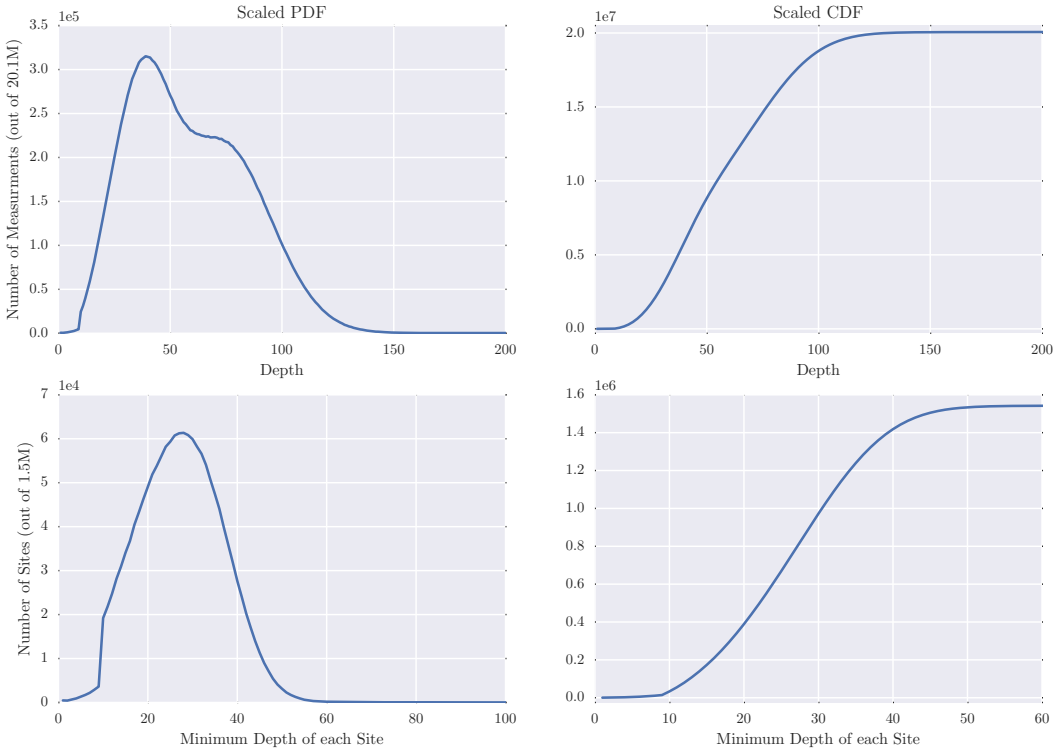

Figure S15: Scaled PDF (left) and CDF (right) of the overall read depth distribution (top) and minimum depth of sites (bottom). Top row depicts the overall distribution of rad of all reads.

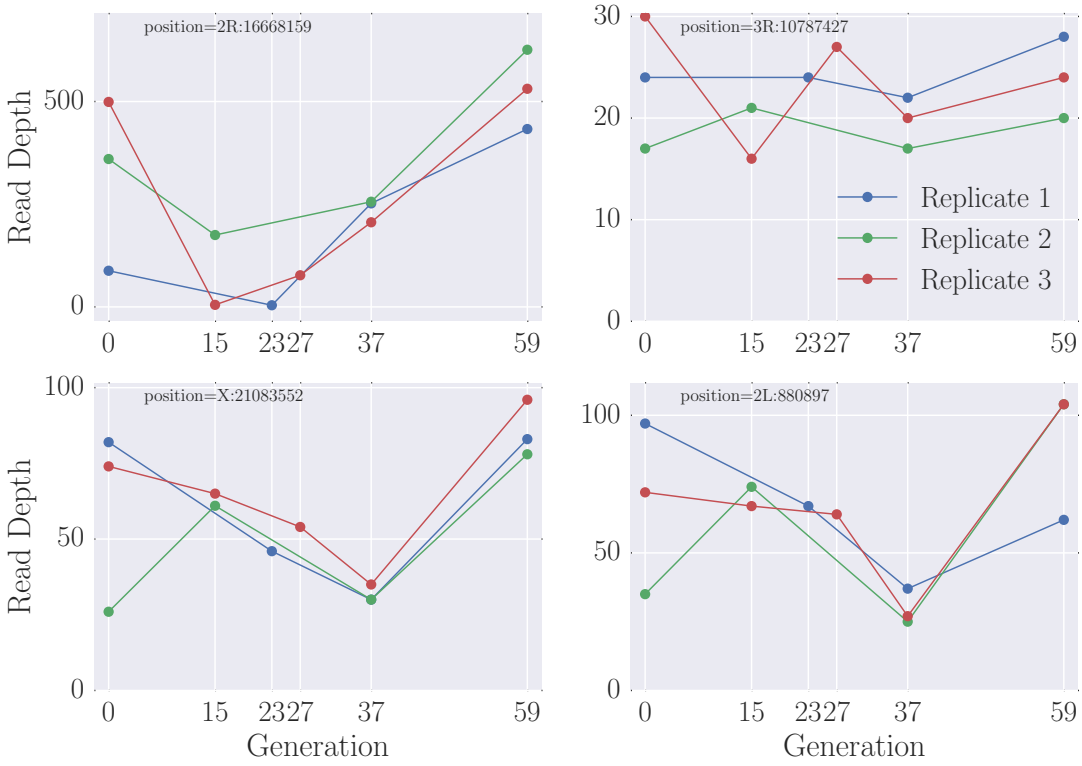

Figure S16: Read depth at four different sites. (which would be filtered if the min depth is set to 30X)

References

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

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

[3] Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., and Others. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature 505*, 7481 (2014), 50–55.

[4] Barrett, R. D. H., Rogers, S. M., and Schluter, D. Natural selection on a major armor gene in threespine stickleback. *Science 322*, 5899 (2008), 255–257.

[5] Barrick, J. E., and Lenski, R. E. Genome dynamics during experimental evolution. *Nat Rev Genet 14*, 12 (dec 2013), 827–839.

[6] Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., and Kim, J. F. Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature 461*, 7268 (2009), 1243–1247.

[7] Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., and Petrov, D. A. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. *PLoS Genet 10*, 11 (2014), e1004775.

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

[9] Bollback, J. P., York, T. L., and Nielsen, R. Estimation of 2Nes from temporal allele frequency data. *Genetics 179*, 1 (2008), 497–502.

[10] Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., Hernandez, R. D., Lohmueller, K. E., Adams, M. D., Schmidt, S., Sninsky, J. J., Sunyaev, S. R., and Others. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet 4*, 5 (2008), e1000083.

[11] Burke, M. K., Dunham, J. P., Shahrestani, P., Thornton, K. R., Rose, M. R., and Long, A. D. Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature 467*, 7315 (2010), 587–590.

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

[13] Daniels, R., Chang, H.-H., Séne, P. D., Park, D. C., Neafsey, D. E., Schaffner, S. F., Hamilton, E. J., Lukens, A. K., Van Tyne, D., Mboup, S., and Others. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS One 8*, 4 (2013), e60780.

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

[15] Desai, M. M., and Fisher, D. S. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics 176*, 3 (2007), 1759–1798.

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

[17] Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.

[18] Evans, S. N., Shvets, Y., and Slatkin, M. Non-equilibrium theory of the allele frequency spectrum. *Theoretical population biology 71*, 1 (2007), 109–119.

[19] Ewens, W. J. *Mathematical Population Genetics 1: Theoretical Introduction*, vol. 27. Springer Science & Business Media, 2012.

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

[21] Feder, A. F., Kryazhimskiy, S., and Plotkin, J. B. Identifying signatures of selection in genetic time series. *Genetics 196*, 2 (2014), 509–522.

[22] Feder, A. F., Rhee, S.-Y., Holmes, S. P., Shafer, R. W., Petrov, D. A., and Pennings, P. S. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *eLife 5* (jan 2016).

[23] Fu, Y.-X. Statistical properties of segregating sites. *Theoretical population biology 48*, 2 (1995), 172–197.

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

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

[26] Hegreness, M., Shoresh, N., Hartl, D., and Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science 311*, 5767 (2006), 1615–1617.

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

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

[29] Illingworth, C. J. R., Parts, L., Schiffels, S., Liti, G., and Mustonen, V. Quantifying selection acting on a complex trait using allele frequency time series data. *Molecular biology and evolution 29*, 4 (2012), 1187–1197.

[30] Izutsu, M., Toyoda, A., Fujiyama, A., Agata, K., and Fuse, N. Dynamics of Dark-Fly Genome Under Environmental Selections. *G3: Genes— Genomes— Genetics* (2015), g3—-115.

[31] Jha, A. R., Miles, C. M., Lippert, N. R., Brown, C. D., White, K. P., and Kreitman, M. Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in Drosophila melanogaster. *Molecular biology and evolution 32*, 10 (2015), 2616–2632.

[32] Kaplan, N. L., Hudson, R. R., and Langley, C. H. The" hitchhiking effect" revisited. *Genetics 123*, 4 (1989), 887–899.

[33] Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M. C. Experimental evolution. *Trends in ecology & evolution 27*, 10 (2012), 547–560.

[34] Lang, G. I., Botstein, D., and Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics 188*, 3 (2011), 647–661.

[35] Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D., and Desai, M. M. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature 500*, 7464 (2013), 571–574.

[36] Malaspinas, A.-S., Malaspinas, O., Evans, S. N., and Slatkin, M. Estimating allele age and selection coefficient from time-serial data. *Genetics 192*, 2 (2012), 599–607.

[37] Maldarelli, F., Kearney, M., Palmer, S., Stephens, R., Mican, J., Polis, M. A., Davey, R. T., Kovacs, J., Shao, W., Rock-Kress, D., and Others. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. *Journal of virology 87*, 18 (2013), 10313–10323.

[38] Martins, N. E., Faria, V. G., Nolte, V., Schlötterer, C., Teixeira, L., Sucena, É., and Magalhães, S. Host adaptation to viruses relies on few genes with different cross-resistance properties. *Proceedings of the National Academy of Sciences 111*, 16 (2014), 5938–5943.

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

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

[41] Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A.-C., Krishna, S., Nosten, F., and Anderson, T. J. C. Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Molecular Biology and Evolution 24*, 2 (2007), 562–573.

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

[43] Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. Genomic scans for selective sweeps using SNP data. *Genome research 15*, 11 (2005), 1566–1575.

[44] OROZCO-terWENGEL, P., Kapun, M., Nolte, V., Kofler, R., Flatt, T., and Schlötterer, C. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Molecular ecology 21*, 20 (2012), 4931–4941.

[45] Oz, T., Guvenek, A., Yildiz, S., Karaboga, E., Tamer, Y. T., Mumcuyan, N., Ozan, V. B., Senturk, G. H., Cokol, M., Yeh, P., and Others. Strength of selection pressure

is an important parameter contributing to the complexity of antibiotic resistance evolution. *Molecular biology and evolution* (2014), msu191.

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

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

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

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

[50] Remolina, S. C., Chang, P. L., Leips, J., Nuzhdin, S. V., and Hughes, K. A. Genomic basis of aging and life-history evolution in Drosophila melanogaster. *Evolution 66*, 11 (2012), 3390–3403.

[51] Ronen, R., Tesler, G., Akbari, A., Zakov, S., Rosenberg, N. A., and Bafna, V. Predicting Carriers of Ongoing Selective Sweeps Without Knowledge of the Favored Allele. *PLoS Genet 11*, 9 (2015), e1005527.

[52] Ronen, R., Udpa, N., Halperin, E., and Bafna, V. Learning natural selection from the site frequency spectrum. *Genetics 195*, 1 (2013), 181–193.

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

[54] Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., and Others. Genome-wide detection and characterization of positive selection in human populations. *Nature 449*, 7164 (2007), 913–918.

[55] Sawyer, S. A., and Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics 132*, 4 (1992), 1161–1176.

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

[57] Smith, J. M., and Haigh, J. The hitch-hiking effect of a favourable gene. *Genetical research 23*, 01 (1974), 23–35.

[58] Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., Edwards, J., of America, I. D. S., and Others. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clinical Infectious Diseases 46*, 2 (2008), 155–164.

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

[60] Stephan, W., Song, Y. S., and Langley, C. H. The Hitchhiking Effect on Linkage Disequilibrium Between Linked Neutral Loci. *Genetics 172*, 4 (apr 2006), 2647–2663.

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

[62] Terhorst, J., Schlötterer, C., and Song, Y. S. Multi-locus Analysis of Genomic Time Series Data from Experimental Evolution. *PLoS Genet 11*, 4 (2015), e1005069.

[63] Tobler, R., Franssen, S. U., Kofler, R., Orozco-terWengel, P., Nolte, V., Hermisson, J., and Schlötterer, C. Massive habitat-specific genomic response in D. melanogaster populations during experimental evolution in hot and cold environments. *Molecular biology and evolution 31*, 2 (2014), 364–375.

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

[65] Vitti, J. J., Grossman, S. R., and Sabeti, P. C. Detecting natural selection in genomic data. *Annual review of genetics 47* (2013), 97–120.

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

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

[68] Williams, D., and Williams, D. *Weighing the odds: a course in probability and statistics*, vol. 548. Springer, 2001.

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

[70] Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., and Nielsen, R. Localizing recent adaptive evolution in the human genome. *PLoS Genet 3*, 6 (2007), e90.

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

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

[73] Zhou, D., Udpa, N., Gersten, M., Visk, D. W., Bashir, A., Xue, J., Frazer, K. A., Posakony, J. W., Subramaniam, S., Bafna, V., and Others. Experimental selection of hypoxia-tolerant Drosophila melanogaster. *Proceedings of the National Academy of Sciences 108*, 6 (2011), 2349–2354.