# Quantifying Selection with Pool-Seq Time Series Data

Thomas Taus,[1,2] Andreas Futschik,[3] and Christian Schlötterer*,[1]

[1]Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria
[2]Vienna Graduate School of Population Genetics, Vienna, Austria
[3]Department of Applied Statistics, Johannes Kepler Universität Linz, Linz, Austria

*Corresponding author: E-mail: christian.schloetterer@vetmeduni.ac.at.
Associate editor: Nadia Singh

## Abstract

Allele frequency time series data constitute a powerful resource for unraveling mechanisms of adaptation, because the temporal dimension captures important information about evolutionary forces. In particular, Evolve and Resequence (E&R), the whole-genome sequencing of replicated experimentally evolving populations, is becoming increasingly popular. Based on computer simulations several studies proposed experimental parameters to optimize the identification of the selection targets. No such recommendations are available for the underlying parameters selection strength and dominance. Here, we introduce a highly accurate method to estimate selection parameters from replicated time series data, which is fast enough to be applied on a genome scale. Using this new method, we evaluate how experimental parameters can be optimized to obtain the most reliable estimates for selection parameters. We show that the effective population size ($N_e$) and the number of replicates have the largest impact. Because the number of time points and sequencing coverage had only a minor effect, we suggest that time series analysis is feasible without major increase in sequencing costs. We anticipate that time series analysis will become routine in E&R studies.

*Key words:* evolve and resequence, time series, selection, software.

## Introduction

The underlying molecular mechanisms of adaption in natural populations to novel environments have been of long-standing interest in evolutionary genetics. Nevertheless, it is also becoming increasingly clear that information about adaptive alleles provides an enormous potential for a broad range of disciplines including vaccine and drug development (e.g., Andries et al. 2005; Sequist et al. 2011), animal and plant breeding (e.g., Hufford et al. 2012; Daetwyler et al. 2014), pest management, biomedicine and many more. A widely used approach for the identification of selection relies on pattern of sequence variation, which are not compatible with the null hypothesis of neutral drift, preferentially accounting for past demographic processes. One of the most simple selection signatures is generated by a beneficial allele sweeping through the population; linked neutral alleles are hitchhiking with the target of selection (Smith and Haigh 1974; Barton 2000) generating patterns of reduced variability, increased linkage disequilibrium and skewed site frequency spectra around the targeted genomic region. These signals are generated by sequence variants flanking the selected site and provide inference power for popular tests for selection (e.g., Tajima 1983; Fay and Wu 2000; Kim and Stephan 2002; Sabeti et al. 2002; Kim and Nielsen 2004; Jensen et al. 2005; Nielsen et al. 2005; Voight et al. 2006; Foll and Gaggiotti 2008; Pavlidis et al. 2013; Ferrer-Admetlla et al. 2014). While these tests utilize polymorphism data from a single time point, the analysis of samples collected during multiple time points may be powerful enough to infer selection for a single site taking advantage of the allele frequency changes across the different time intervals. Although the advantage of time series for population genetic analyses has been appreciated for a long time (Fisher and Ford 1947; Wright 1948), due to the lack of adequate data this approach had not been receiving much attention. With the arrival of second-generation sequencing the situation has changed and with decreasing sequencing costs, the collection of time series data has become feasible and is being recognized as a powerful approach to study adaptive processes.

One recent example for the successful use of time series data is the sequencing of ancient human remains to characterize trajectories of selected alleles and to identify candidate loci in European human populations (Mathieson et al. 2015). Another study of seasonal population samples collected over 3 years demonstrated that natural *Drosophila melanogaster* populations rapidly respond to selection driven by seasonal changes in the environment (Bergland et al. 2014). Nevertheless, one particular challenge for the inference of adaptive processes from natural populations arises from demographic processes, which may result in biased estimates if not accounted for (e.g., Thornton et al. 2007; Crisci et al. 2012; Haasl and Payseur 2016).

An alternative source of high resolution time series data comes from experimental evolution studies combined with sequencing pools of individuals (Pool-Seq) (Futschik and Schlötterer 2010; Schlötterer et al. 2014), which provide replicated samples at multiple time points. Because the environmental conditions are well controlled and evolutionary trajectories can be monitored in replicates, such evolve and resequence (E&R) studies (Turner et al. 2011) are particularly

Article

informative to link selection signatures and adaptive changes to a given environment. Over the past years adaptation to novel environments was studied within an E&R framework in a wide range of organisms, including bacteria (e.g., Barrick et al. 2009; Woods et al. 2011; Tenaillon et al. 2012, 2016), viruses (e.g., Foll et al. 2014), yeast (e.g., Kao and Sherlock 2008; Lang et al. 2013; Burke et al. 2014; McDonald et al. 2016), *Drosophila* (e.g., Burke et al. 2010; Turner et al. 2011; Zhou et al. 2011; Orozco-terWengel et al. 2012; Turner and Miller 2012; Tobler et al. 2014; Griffin et al. 2017), and mice (e.g., Chan et al. 2012).

Despite the conceptual appeal and an increasing number of suitable data sets, the inference of selection parameters from time series data remains a significant challenge. Unlike in cases where random drift can be ignored (Illingworth et al. 2012), calculating the exact probability of allele frequency trajectories for finite population sizes comes at a huge computational cost (Jewett et al. 2016). Hence, transition probabilities are approximated (Terhorst et al. 2015; Topa et al. 2015; Khatri 2016) and parameters are estimated in a maximum likelihood (ML) (Bollback et al. 2008; Malaspinas et al. 2012; Mathieson and McVean 2013; Steinrücken et al. 2014; Iranmehr et al. 2017) or a Bayesian framework (Ferrer-Admetlla et al. 2016; Schraiber et al. 2016). Because these methods are still computationally rather demanding, estimates were also obtained by forward simulations in combination with approximate Bayesian computation (Foll et al. 2015). While all these methods account for random allele frequency fluctuations owing to finite population sizes, some are capable of estimating both $N_e$ and $s$ (Bollback et al. 2008; Malaspinas et al. 2012; Steinrücken et al. 2014; Foll et al. 2015; Ferrer-Admetlla et al. 2016; Iranmehr et al. 2017), whereas others rely on independent $N_e$ estimates (Mathieson and McVean 2013). Recently, Jewett et al. (2016) suggested that purely deterministic models, despite ignoring the effects of random drift, could provide accurate estimates and reduce the computation time by several orders of magnitude. This implies that the characterization of selection dynamics for millions of loci genome-wide is possible in a reasonable time frame. However, it remains unclear if accuracy is sufficiently high, when such deterministic models are applied to E&R studies in sexual organisms, such as *Drosophila*, where $N_e$ is in the hundreds (e.g., Tobler et al. 2014). Additionally, available methods are mostly limited to the analysis of single trajectories, while state-of-the-art E&R studies generate replicated allele frequency trajectories. Thus, the inference potential provided by modern experimental designs is often not fully exploited.

Several computer simulation studies provided guidelines about the experimental design of E&R studies to optimize the detection of selected loci (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014; Kessner and Novembre 2015). In contrast, not much is known about the influence of the experimental design on the inference of selection parameters. We introduce a fast and highly accurate approach to estimate the selection coefficient $s$ and dominance $h$ and evaluate different designs for E&R. We show that the number of replicates and the effective population size are the primary factors

determining the accuracy of estimates. Because a large number of time points and a high sequencing depth are not needed in terms of accuracy, time series analyses are affordable and will most likely be the standard for future E&R analyses.

## Results and Discussion

We introduce a new approach to estimate $s$ for haploid and diploid populations that combines information of replicated evolutionary trajectories. In its basic form, it employs linear least squares regression (LLS) to fit allele frequency data to a purely deterministic selection model.

### Robustness under a Wide Range of Scenarios
We tested the robustness of the LLS method by applying it to a set of 1.2 million simulated trajectories varying $s$ and $p_0$. These simulations covered a wide range of experimental setups differing in the number of replicates, $N_e$, sequencing coverage and the total number of generations. Selection coefficients were estimated with LLS and contrasted with the actual parameter values (fig. 1). $N_e$ and the number of replicates had the largest effect on the precision of selection estimates, whereas sequencing coverage and number of generations only played a minor role. Although the precision varied across experimental parameters, LLS-based estimates were unbiased under all scenarios examined. For low sequencing coverage ($20\times$) only, our selection estimators showed some bias with values that were systematically too small. This bias appears to be negligible compared with the low precision that results from the sampling noise. Interestingly, despite this drop in precision for low sequence coverage, our method was still powerful to identify selected loci (87%, $\alpha = 0.01$).

We related the performance of our approach to WFABC (Foll et al. 2015) and CLEAR (Iranmehr et al. 2017). The former was previously favorably compared to other methods estimating $s$ from time series data—in particular for accuracy under strong selection and computational efficiency (Foll et al. 2015). CLEAR is a method to detect selection in E&R data (Iranmehr et al. 2017), which also estimates $s$. We simulated allele frequency trajectories for 100,000 loci and 60 generations with $p_0 \in [0, 1]$ and $s \in [0, 0.3]$ assuming $N_e = 300$. Allele frequencies were sampled every ten generations mimicking an average Pool-Seq sequencing depth of $80\times$. The strength of selection was estimated for single trajectories only assuming $h = 0.5$, with LLS, WFABC, and CLEAR. Accuracy was measured by the relative root–mean–square error (rRMSE, see Materials and Methods for details), since this measure captures both bias and variance of an estimator.

All methods provided the most reliable estimates for large selection coefficients (see fig. 2A) and intermediate (between 10% and 30%) starting allele frequencies (see fig. 2B). With decreasing selective advantage, estimates became less accurate, but LLS-based estimates provided lower rRMSE values than those of CLEAR and WFABC. The accuracy dropped (and rRMSE increased) also with increasing starting allele frequency for all methods, but the effect was less pronounced
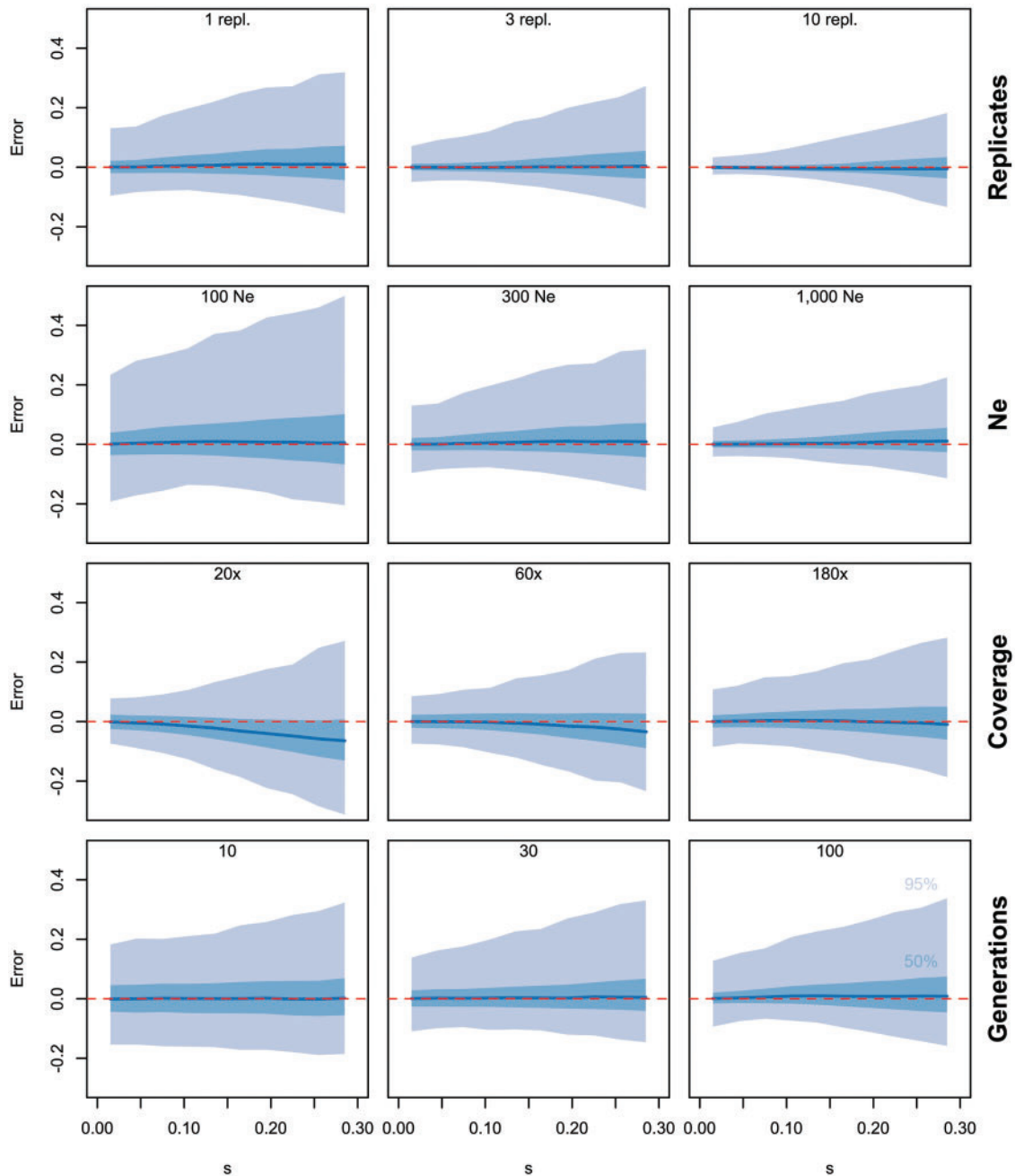
**FIG. 1.** Robustness of LLS under a wide range of scenarios. Allele frequency trajectories were simulated for 100,000 unlinked loci. Unless noted otherwise the default parameters were 60 geneations, $N_e = 300$, $p_0 \in [0, 1]$ and $s \in [0, 0.3]$ and 1 replicate. We tested the influence of the number of replicates, $N_e$, sequencing coverage or the total number of generations (values on top of each panel) on the robustness of LLS. A bias correction was applied to all estimates (see Materials and Methods). The absolute error between estimated and true $s$ is shown: median (dark blue), 25–75% quantile (light blue) and 2.5–97.5% quantile (gray) of values.

for LLS and CLEAR. Under weak selection, as well as high starting allele frequency of the beneficial allele, it is challenging to disentangle directional selection from random drift. Our results suggest that the good performance of LLS becomes particularly evident for those cases where it is challenging to distinguish selection from drift.

Applying any selection inference procedure on a genomic scale requires a computationally effective method. It has recently been suggested that runtimes of ML-based estimators can be reduced by several orders of magnitude by ignoring

demographic histories, while maintaining high accuracy if $N_e$ and $s$ are large enough (Jewett et al. 2016). WFABC is considerably faster than several ML-based approaches (Foll et al. 2015), but we found that LLS required on average 0.0015 s per locus, while CLEAR and WFABC needed about 0.1 and 1.4 s, respectively. For LLS this translates into a processing time of 25 min for one million loci. Notably, the processing time of CLEAR increased dramatically with $N_e$, while it remained fairly constant with LLS (see supplementary table S1, Supplementary Material online). Most importantly, the
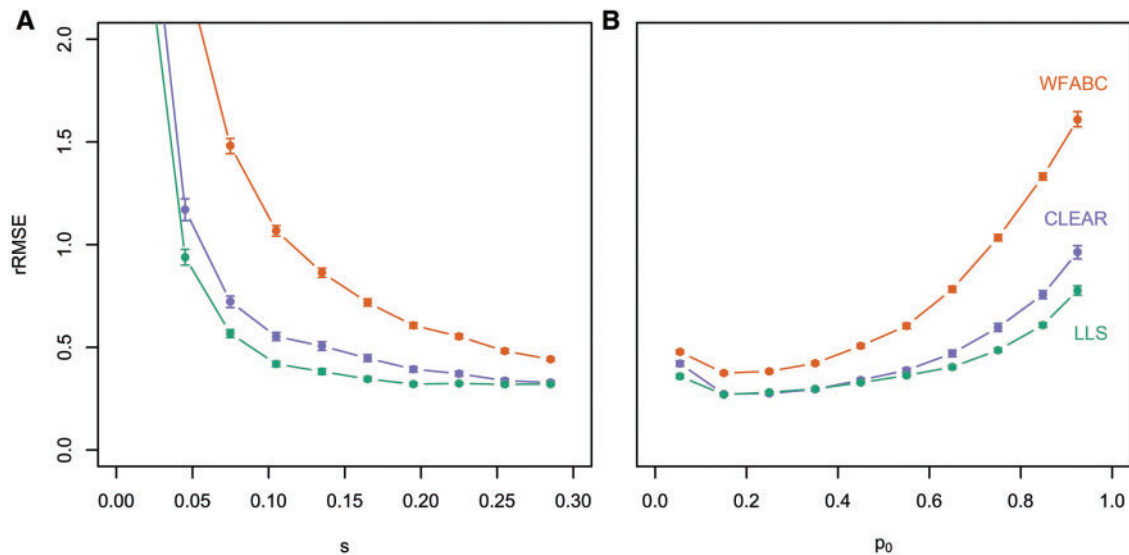
**FIG. 2.** Comparison of LLS, WFABC, and CLEAR. Allele frequency tajectories were simulated for 100,000 unlinked loci over 60 generations assuming $N_e = 300$ with $p_0 \in [0, 1]$ and $s \in [0, 0.3]$. Selection coefficients were estimated with LLS (green), as well as WFABC (orange), and CLEAR (purple). The relative deviation between estimated and true selection coefficient is shown as a function of $s$ (A) and $p_0$ (B).

higher computational efficiency of LLS does not come at the cost of reduced accuracy; as for challenging scenarios (small $s$ or large $p_0$) it is more accurate than WFABC, while providing a level of accuracy similar to CLEAR.

So far, we assumed codominance ($h = 0.5$), but since dominance has a pronounced effect on the frequency trajectory of selected alleles, we tested the robustness of LLS with different degrees of dominance. We simulated replicated allele frequency trajectories with $h \in \{0, 0.5, 0.75, 1\}$ and estimated selection coefficients with LLS. While $s$ is upward biased for recessive alleles, we notice a downward bias in the case of dominance (see fig. 3).

These results emphasize that ignoring dominance could result in highly biased results with the LLS method. Using the nonlinear least squares regression (NLS) approach to coestimate $s$ and $h$ from time series data, we obtain unbiased selection coefficient estimates for all values of $h$ (see fig. 3). While the width of the 50% band of errors is comparable between LLS and NLS, we obtain more outliers with NLS, which is reflected in a rather broad 95% band of errors. Analyzing the 1% most extreme outliers suggests that that the numerical procedure to optimize the sum of squared errors is robust with respect to the starting value and therefore multiple local optima appear to be only rarely responsible for the reduced precision (see supplementary fig. S1, Supplementary Material online). Rather, it seems that outliers occur frequently, when $p_0$ of the beneficial allele is high and the respective trajectories are not informative enough to coestimate $s$ and $h$ (see supplementary figs. S2 and S3, Supplementary Material online).

To combine the precision of LLS with the robustness of NLS, we developed a procedure, which relies on NLS estimates only when the data suggest a deviation from codominance. We inferred the deviation from codominance by adding a quadratic term to the linear model. If the $P$ value of the quadratic term is sufficiently small, NLS is used instead of LLS. A too stringent threshold implies that LLS will be preferred, which will provide potentially biased estimates, if the assumption of $h = 0.5$ is violated. On the other hand, too liberal $P$ value thresholds favor NLS, which is in general less precise but always unbiased. Using a $P$ value threshold of 0.1, our results suggested that our automated switching method constitutes a good compromise between precision and bias (see fig. 3). The specificity of the switching procedure appeared to be best for $p_0 < 1/3$ (see supplementary fig. S4, Supplementary Material online). In the case of higher starting allele frequencies, trajectories may not be informative enough to detect deviations from codominance and estimates may be biased.

## Benefit of Replicated Allele Frequency Trajectories

Genetic drift and sampling noise during allele frequency measurements (e.g., Pool-Seq) are responsible for deviations between true and estimated parameter values. The former can be reduced either by increasing $N_e$ or by inferring parameters using the information from multiple evolutionary trajectories simultaneously, as available in experimental evolution with replicate populations. The latter can be decreased by increasing pool size and sequencing depth. This raises the question whether multiple replicates or increased sequencing coverage are more effective to improve our estimates. To address this, we simulated time series data under two scenarios with the same sequencing effort: 1) single trajectories with high sequencing coverage ($480\times$) and 2) six replicates with $80\times$ each. An effective population size of 300 was chosen, because it reflects a realistic value for E&R studies in sexual organisms, such as *Drosophila*. Selection coefficients were estimated with LLS.

With the same sequencing effort, we found for the entire parameter space that more replicates provided more accurate results than a single replicate with higher sequencing
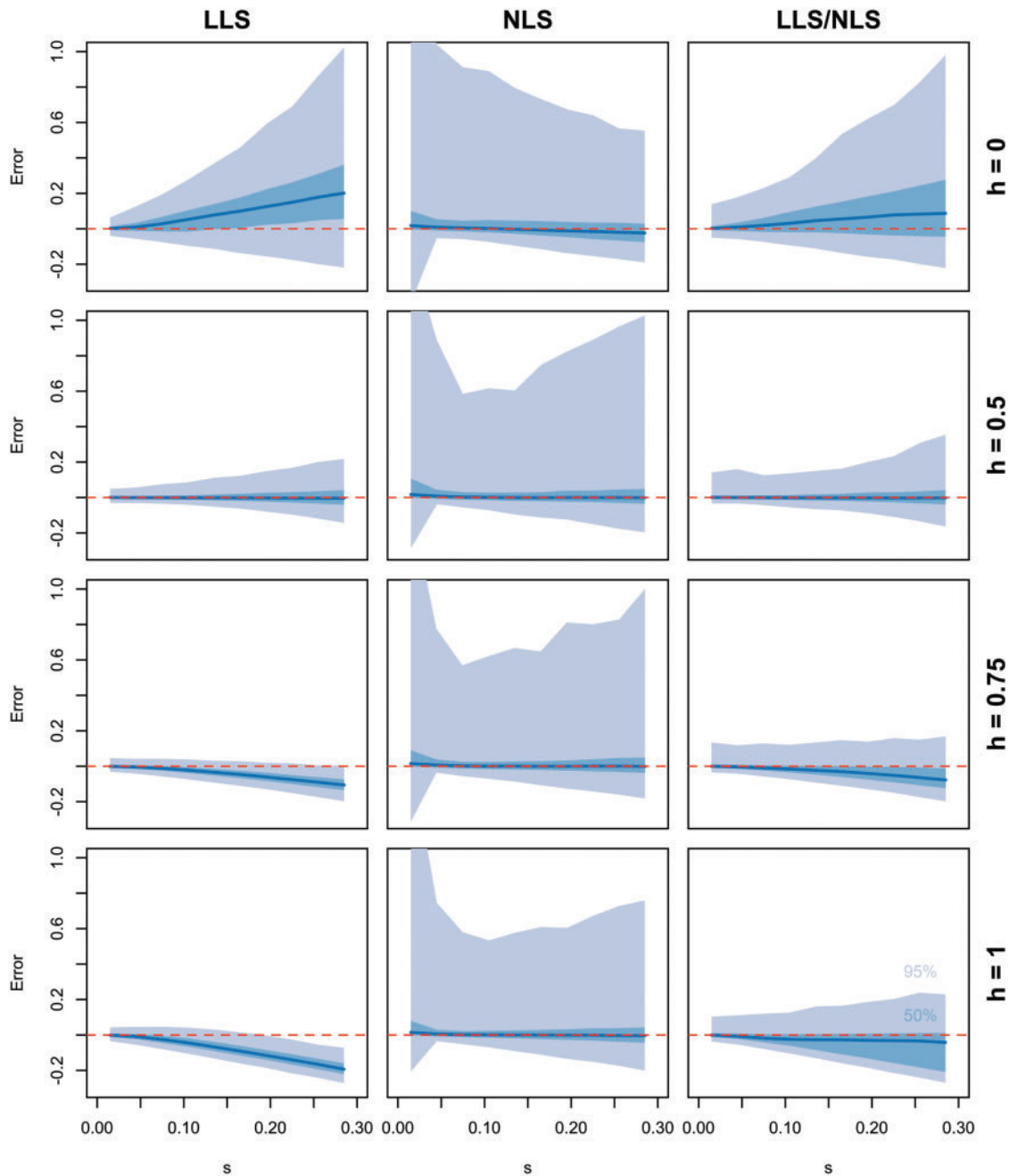
**Fig. 3.** Impact of the dominance parameter on the accuracy of selection estimates. Allele frequency trajectories were simulated in six replicates for 100,000 unlinked loci over 60 geneations assuming $N_e = 300$ with $p_0 \in [0, 1]$ and $s \in [0, 0.3]$. The dominance parameter ($h$) was assumed to be 0, 0.5, 0.75, or 1. Selection coefficients were estimated with both LLS and NLS. A bias correction was applied to all estimates (see Materials and Methods). Additionally, an automatic switching approach was employed (LLS/NLS) that uses NLS-based estimates instead of those of LLS, if there is sufficient statistical evidence for the need of a more complex selection model. Illustrated is the median (dark blue), 25–75% quantile (light blue) and 2.5–97.5% quantile (gray) of absolute errors between true and estimated selection coefficient as a function of $s$.

coverage (see fig. 4). This result makes intuitive sense, because by combining allele frequency estimates across independently evolved replicates the effect of random drift is reduced, while increasing the read depth reduces solely the loss in accuracy due to the sequencing noise. This emphasizes the importance of sequencing multiple replicates, when studying populations with small $N_e$. It is important to note that the reduction in rRMSE by increasing the number of replicates will be less pronounced for large $N_e$, for example, in E&R studies with

microbes, because single trajectories will be close to deterministic ones.

## Performance under a Typical E&R Scenario

We further characterized the accuracy of $s$ estimates, by simulating allele frequency trajectories for unlinked loci under experimental parameters that resemble those of recent E&R studies with *Drosophila* (e.g., Turner et al. 2011; Tobler et al. 2014; Griffin et al. 2017). For a total of ten million loci, six
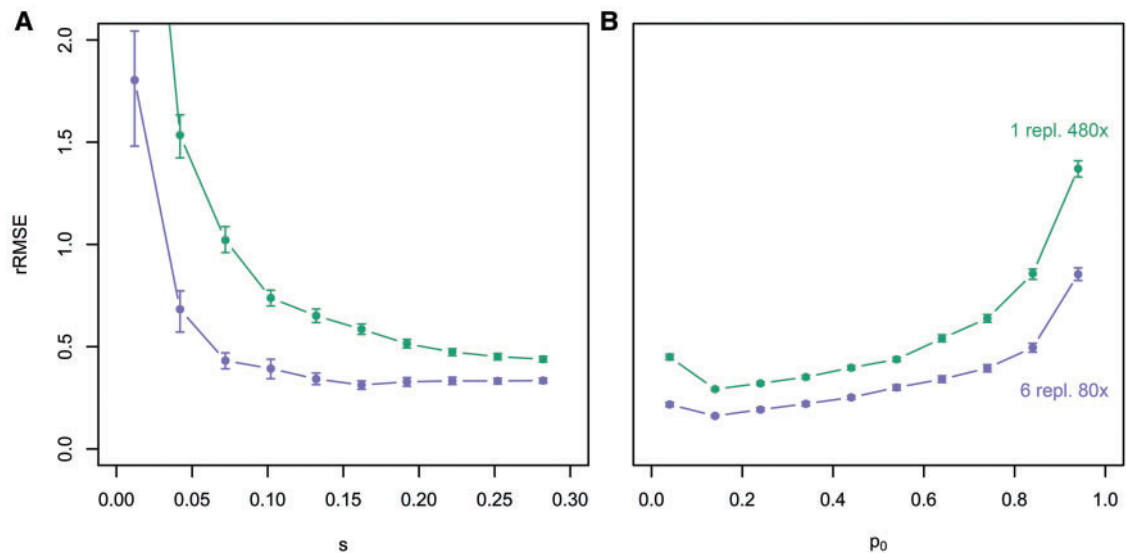
**FIG. 4.** The advantage of replication over high sequencing coverage. Allele frequency tajectories were simulated for 100,000 unlinked loci over 60 generations assuming $N_e = 300$ with $p_0 \in [0, 1]$ and $s \in [0, 0.3]$. Selection coefficients were estimated with LLS using either a single replicate per locus (green), or by combining the information of multiple replicates (purple). The overall sequencing effort was assumed to be constant. The relative deviation between estimated and true selection coefficient is shown as a function of $s$ (A) and $p_0$ (B).

independent trajectories were simulated over 60 generations each with population allele frequency estimates at every 10th generation. Sampling noise was added to mimic Pool-Seq with an average sequencing coverage of 80×. The entire parameter space was split into 2,500 equally sized bins based on both $p_0$ and $s$, and accuracy was assessed based on the rRMSE.

For 60% of the parameter space, estimates of $s$ were highly accurate with rRMSE below 0.3 (see fig. 5). Notably, this included selection coefficients as low as 0.05, if the beneficial allele started at a frequency between 0.1 and 0.7. For low starting frequencies ($p_0 < 0.1$) and weak selection ($s < 0.05$) the estimates were less accurate. In general, rRMSE exceeds 0.9 if either $p_0 > 95\%$ or $s < 0.01$. High starting frequencies result in short and uninformative trajectories, which translates in the higher uncertainty. If the fitness benefit is too small, it is challenging to distinguish random drift from selection and the relative error of estimates increases.

Analyzing the same data set with the automatic switching method provided comparable results for variants under strong selection ($s > 0.1$) with a starting allele frequency below 60% (see supplementary fig. S5, Supplementary Material online). For rare alleles under weak selection, or variants at high frequency ($p_0 > 0.75$) estimates were imprecise and the rRMSE exceeded 0.9.

Genetic drift results in random allele frequency changes, thus time series data could distinguish between random (genetic drift) and directional (selection) allele frequency changes. Nevertheless, by chance genetic drift can also generate time series data that resemble those of selected alleles. We used computer simulations to distinguish selection from neutral expectations under drift. Specifically, we determined the probability of a given selection coefficient under the null hypothesis of random genetic drift by parametric bootstrapping (see Materials and Methods). Applying this neutrality test to the aforementioned simulations we determined the
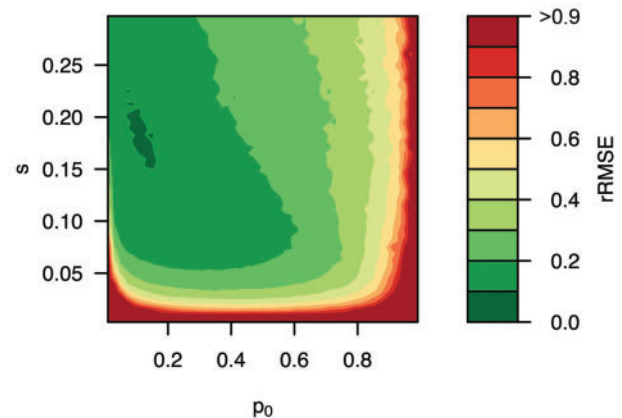


**FIG. 5.** Accuracy of selection coefficient estimates under a state-of-the-art E&R experiment. Allele frequency trajectories were simulated in six replicates for ten million unlinked loci over 60 geneations assuming $N_e = 300$ with $p_0 \in [0, 1]$ and $s \in [0, 0.3]$. Selection coefficients were estimated with LLS. Estimates were divided into 2,500 bins based on $p_0$ and $s$, before computing the rRMSE for each bin. Illustrated are the color-coded rRMSE values for each region of the parameter space, with green and red corresponding to high and low accuracy, respectively.

fraction of significant loci ($\alpha = 0.01$)—that is, the estimated power of our test—for different values of $s$ and $p_0$ (see fig. 6).

LLS-based selection tests turned out to be very powerful. More than 80% of the selected loci with $s > 0.03$, $s > 0.06$, or $s > 0.09$ were detected for starting allele frequencies of $p_0 = 0.25$, $p_0 = 0.1$, or $p_0 = 0.05$, respectively. Fitness advantages as low as 2.5% can be detected in $>40\%$ of the cases if $p_0$ is between 0.3 and 0.7. Variants with 5% selective advantage can be detected in $>80\%$ of the cases if they start between 0.1 and 0.8. Thus, even under challenging conditions, selected loci can be detected with a state-of-the-art E&R design at reasonable sensitivity using our method. However, linkage between loci
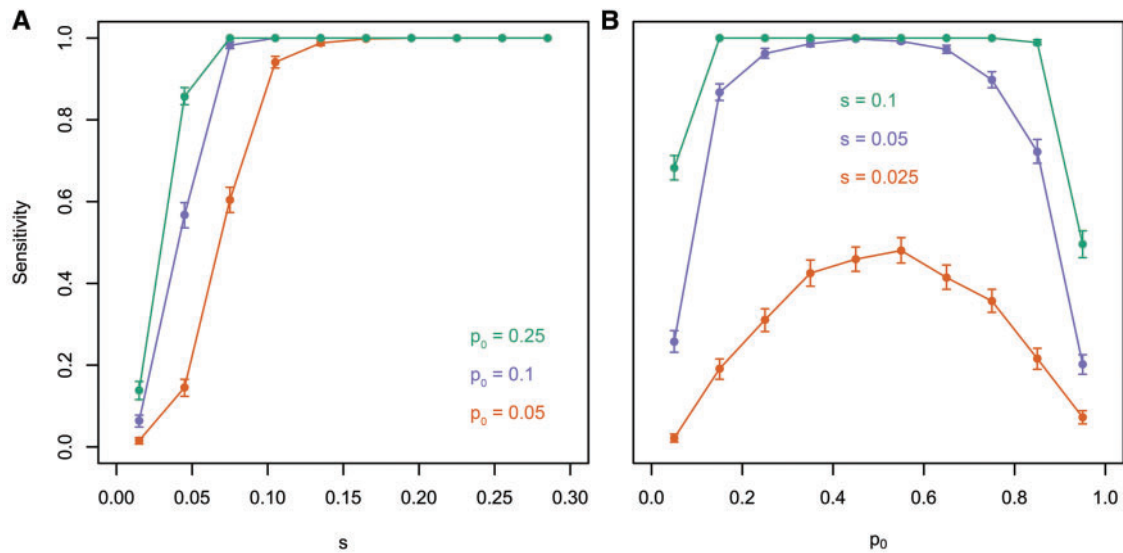
**FIG. 6.** Sensitivity to identify selected loci. Allele frequency trajectories were simulated in six replicates for 10,000 unlinked loci over 60 geneations assuming $N_e = 300$ varying both $p_0$ and $s$. Selection coefficients were estimated and the null hypothesis of random drift was tested with LLS. Illustrated is the sensitivity in identifying selected loci ($\alpha = 0.01$) as a function of $s$ and $p_0$. (A) Beneficial variants started at 5% (orange), 10% (purple) or 25% (green) with $s \in [0, 0.3]$. (B) The selection advantage was equal to 0.025 (orange), 0.05 (purple), or 0.1 (green) with $p_0 \in [0, 1]$.

and competition between beneficial loci in close proximity (Hill and Robertson 1966) will complicate the identification of selection targets.

These results also emphasize that the detection limit of weakly selected alleles can be improved by avoiding small starting allele frequencies. Although this may seem trivial it is tricky to achieve. On one hand, starting the experiment from a diverse set of haplotypes is beneficial because levels of linkage disequilibrium (LD) are low (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014; Kessner and Novembre 2015). On the other hand, under such an experimental setup, most alleles will be at low frequency, thus increasing the number of founder chromosomes will also increase the probability of low frequency alleles under selection.

Recently, a highly outcrossed diploid *Saccharomyces cerevisiae* population originating from four founder genotypes was exposed to competitive growth in liquid media for 540 generations (Burke et al. 2014). Small starting allele frequencies were avoided, because the founding population was derived from four haplotypes only. Strain specific alleles occurred on average at a frequency around 25%. Still, levels of LD were kept low owing to the intense outcrossing scheme between the four founder haplotypes. Such starting frequencies greatly facilitate the identification of targets of selection, even for weakly selected sites (see also fig. 6). Using LLS with an FDR of 0.01, we identified all selected genomic regions of the original analysis and inferred selection coefficients as low as 0.002 (see fig. 7). Compared with the previous study, we found more candidate loci, which we attribute to differences in statistical power between the methods, particularly when selection is acting on alleles with extreme $p_0$ (see supplementary fig. S6, Supplementary Material online). Notably, a rather complex life cycle was imposed in this E&R study and it is unclear if assumptions of a constant selective pressure underlying our analytic model are met. With a considerable

number of generations of asexual growth, clonal interference could further complicate the analysis. Thus, we caution that deviations from the model assumptions—if present—may have influenced the inferred selection coefficients and the exact number and location of true targets of selection require further investigation.

## Impact of Experimental Parameters

Previous studies evaluating the experimental design of E&R studies (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014; Kessner and Novembre 2015) focused on the power to identify selection targets, but did not evaluate the influence of the experimental design on the inference of selection parameters. Thus, we evaluated the impact of several experimental parameters on the accuracy of estimated selection coefficients. Allele frequency trajectories were simulated for weak ($s = 0.025$) and strong ($s = 0.1$) selection on rare ($p_0 = 0.05$) or common ($p_0 = 0.25$) variants under the standard E&R scenario (300 $N_e$, 6 replicates, and 60 generations, measuring allele frequencies every ten generations at 80× coverage). The selection coefficient $s$ was estimated with LLS. Furthermore, we studied the influence of dominance (for $p_0 = 0.25$ and $s = 0.1$) using the switching method. The performance of the estimation procedure was measured again using the relative root mean squared error rRMSE, since it captures both precision and bias (see fig. 8).

The measurement interval determines the temporal resolution of allele frequency trajectories. Keeping the total number of generations constant, shorter intervals result in higher resolution and more measurements. Altering the time interval between 5, 10, and 20 generations, which correspond to 13, 7, and 4 allele frequency measurements, respectively, did not change the accuracy of $s$ estimates. Already four time points (F0, F20, F40, and F60) resulted in accurate selection coefficient estimates. Unlike codominant targets of selection,
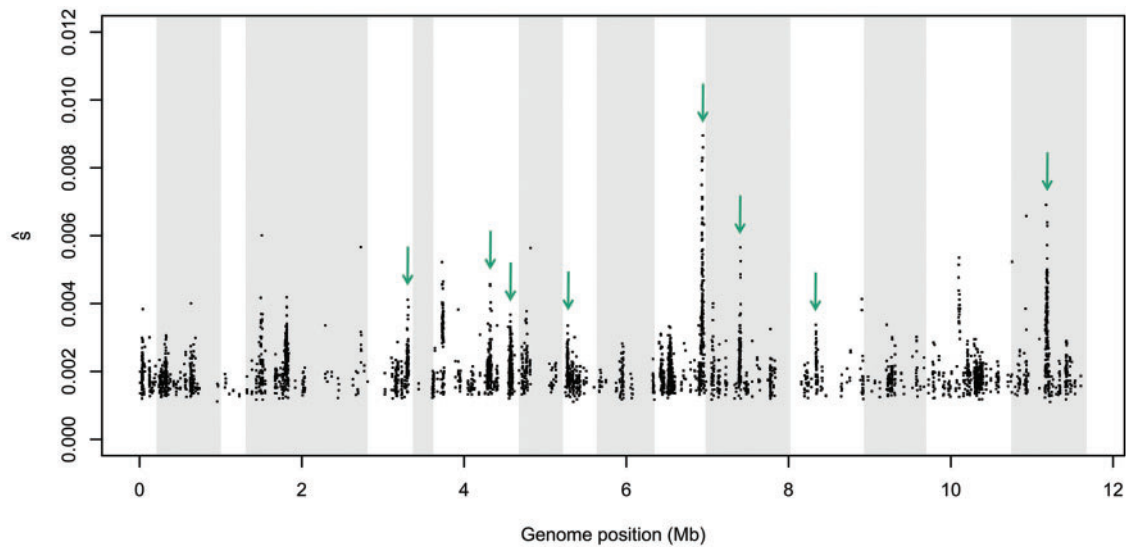
**Fig. 7.** Genome-wide quantification of selection coefficients. The time series data of a recent E&R study in yeast (Burke et al. 2014) were analyzed with LLS. Selection coefficients are plotted for significant SNPs (FDR 0.01) with the X-axis specifying their genomic positions in Mb. Different shadings of gray illustrate individual chromosomes. Green arrows mark genomic regions that were identified as targets of selection in the original analysis.
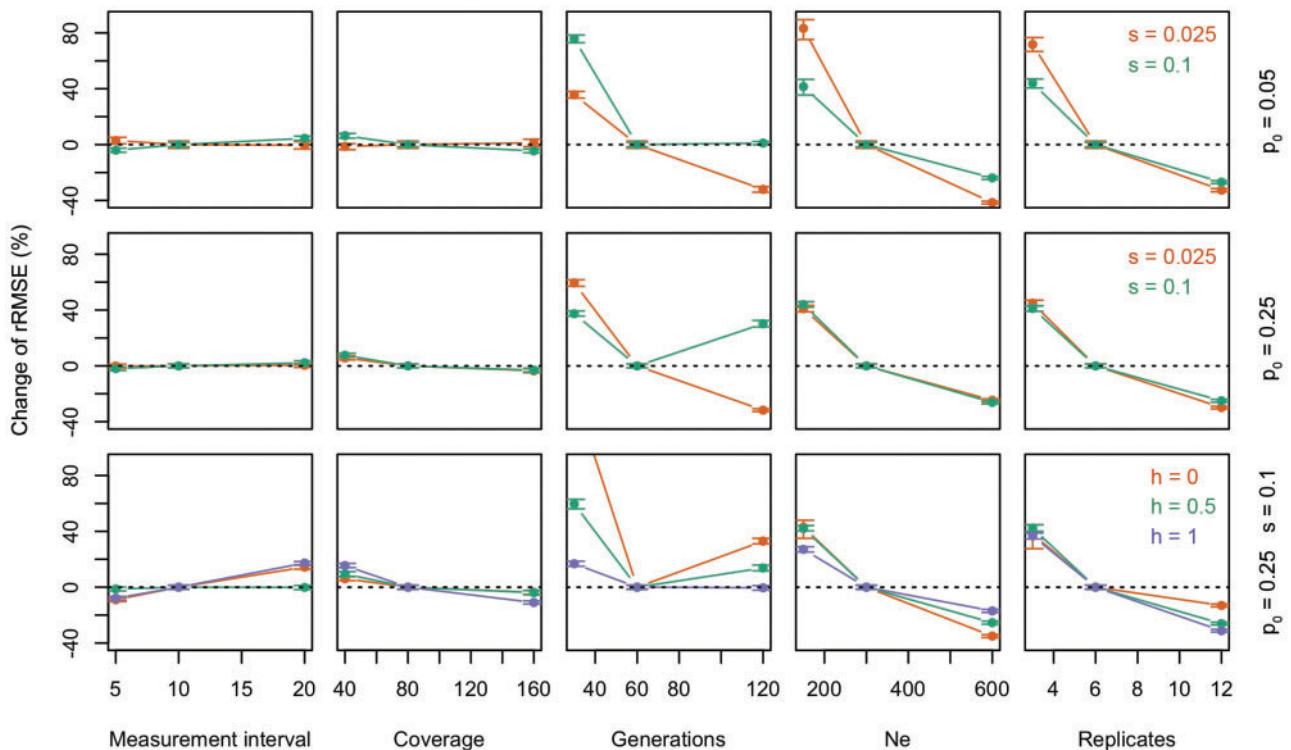


**Fig. 8.** Impact of experimental parameters on the accuracy of s estimates. Allele frequency trajectories were simulated in six replicates for 10,000 unlinked loci over 60 geneations assuming $N_e = 300$ for rare ($p_0 = 0.05$) and common ($p_0 = 0.25$) alleles under weak ($s = 0.025$) or strong ($s = 0.1$) selection. Additionally, we also varied the dominance parameter between 0, 0.5, and 1. We tested the influence of the measurement interval, sequencing coverage, total number of generations, $N_e$, and the number of replicates on the accuracy of s estimates. Illustrated is the change of rRMSE in percent, when modifying one of the experimental parameters.

the analysis of dominant and recessive alleles benefits from shorter measurement intervals. Similarly, sequencing coverage played a minor role regarding the accuracy of selection estimates and already a coverage of 40× produced reasonable estimates. This is in contrast to analyses performed for two

time points, where sequencing coverage was identified as a key factor for reliable identification of selection targets (Kofler and Schlötterer 2014). We conclude that the availability of multiple time points compensates for the increased sampling variance at lower coverage.

The duration of the experiment had a pronounced effect on the accuracy. While for weakly selected alleles rRMSE decreased continuously with the duration of the experiment, for strongly selected alleles the optimal rRMSE was seen for an intermediate number of generations. For strongly selected alleles ($s = 0.1$) with a starting frequency of 0.25, the rRMSE increased even when trajectories covered 120 generations. Deterministic trajectories show that under strong selection ($s = 0.1$) on a rare variant ($p_0 = 0.05$), allele frequencies start to plateau at generation 90, while they keep increasing under weak selection ($s = 0.025$, see supplementary fig. S7, Supplementary Material online). This is even more extreme for beneficial variants that are common ($p_0 = 0.25$, see supplementary fig. S7, Supplementary Material online). Hence, monitoring alleles under strong selection for a large number of generations may not provide additional information and the accuracy of selection estimates does not improve (see supplementary fig. S8, Supplementary Material online). It is important to note that for both LLS and the automatic switching method, reducing the number of generations from 60 to 30 severely deteriorated the accuracy of $s$-estimates for all parameter tested.

Independent of $p_0$, $s$, and $h$, $N_e$ and the number of replicates had the most pronounced influence on accuracy. At low starting frequencies, weakly selected variants benefited more than strongly selected ones from more replicates or larger $N_e$. Under codominance both $N_e$ and the number of replicates had similar effects, but for recessive alleles ($h = 0$) the influence of $N_e$ was more pronounced and for dominance ($h = 1$) increasing the number of replicates had a larger effect.

While $N_e$ and number of replicates were the parameters with the largest potential to improve E&R studies, we also caution that the costs associated with either increased population size or more replicates may be a limiting factor—in particular for studies with a substantial phenotyping component.

## Analyzing Time Series Data with the poolSeq R-Package

We implemented the methods described here in the user-friendly R-package poolSeq (available at https://github.com/ThomasTaus/poolSeq). The major functions of poolSeq are the simulation Pool-Seq data for unlinked loci, $N_e$ estimation (Jónás et al. 2016) and the inference of selection parameters. Forward in time simulations are performed for unlinked loci of haploid or diploid individuals under a Wright–Fisher model. Both the selection coefficient and dominance parameter can be specified. Sampling noise can be added to allele frequencies to mimic the Pool-Seq process. Alternatively to simulating data, allele counts can be loaded into R from the commonly used "sync" file format (Kofler et al. 2011). The package also provides fast implementations of the chi-squared and Cochran–Mantel–Haenszel test.

## Materials and Methods

### Estimating s and h from Time Series Data

We consider unlinked biallelic loci with alleles $A$ and $a$ that evolve at a constant population size $N_e$ for $t$ nonoverlapping generations. Relative fitness is given by $w_A = 1 + s$ and $w_a = 1$ for the two alleles in haploids. In the diploid case, the relative fitness of the three possible genotypes $AA$, $Aa$, and $aa$ is defined as $w_{AA} = 1 + s$, $w_{Aa} = 1 + hs$ and $w_{aa} = 1$. Allele frequencies are denoted $p$ and $1$-$p$ for $A$ and $a$, respectively. In populations of infinite size the allele frequency after one generation of selection ($p'$) can be computed for haploid individuals as:

$$p' = \frac{w_A p}{w_A p + w_a(1 - p)} \qquad (1)$$

and for diploid individuals as:

$$p' = \frac{w_{AA}p^2 + w_{Aa}p(1 - p)}{w_{AA}p^2 + w_{Aa}2p(1 - p) + w_{aa}(1 - p)^2}. \qquad (2)$$

Using a continuous time approximation with overlapping generations and assuming weak selection, the trajectory of the selected allele for $h = 0.5$ (Crow and Kimura 1970) in infinite haploid populations is given by:

$$p(t) = \frac{1}{1 + \left(\frac{1 - p_0}{p_0}\right)e^{-st}}. \qquad (3)$$

The corresponding formula for diploids is:

$$p(t) = \frac{1}{1 + \left(\frac{1 - p_0}{p_0}\right)e^{\frac{-st}{2}}}. \qquad (4)$$

From this a linear relationship between logit-transformed allele frequencies and the number of generations is obtained, both in the haploid and in the diploid case:

$$\ln\left(\frac{p(t)}{1 - p(t)}\right) = \ln\left(\frac{p_0}{1 - p_0}\right) + st, \qquad (5)$$

$$\ln\left(\frac{p(t)}{1 - p(t)}\right) = \ln\left(\frac{p_0}{1 - p_0}\right) + \frac{s}{2}t. \qquad (6)$$

Using equations (5) and (6), we can infer both $s$ and $p_0$ by fitting a linear model with least squares regression (LLS) to logit-transformed allele frequency trajectories. The slope of the linear model provides the selection coefficient, while the starting allele frequency is given by the intercept. However, with increasing strength of selection, the discrepancy between the discrete and continuous model grows (see supplementary fig. S9, Supplementary Material online), ultimately biasing the selection coefficient estimates for populations evolving in discrete generations. We therefore propose a bias correction, in cases where the estimated selection coefficient is large. This is done by adding the difference between the selection coefficient used to compute a deterministic allele frequency trajectory and the value obtained by reestimating the parameter with LLS (see equation [5] or [6]).

It turns out that the logit transform stabilizes the drift variance (see supplementary fig. S10, Supplementary Material online). If sequencing coverage varies across the time points, the fit to the regression model could be further improved in principle by taking nonhomogeneous variances into account. This would imply to fit a weighted linear model, with the inverse coverage as weights. As shown by our simulations however (see fig. 1), the sequencing coverage does not affect the accuracy of estimation much, unless it is very small.

Relaxing the assumption of codominance ($h = 0.5$) requires a more general function describing the allele frequency depending on $p_0$, $s$, $h$, and $t$. Although a general solution to the underlying differential equation (Crow and Kimura 1970) is available, numerical methods are required to estimate both $s$ and $h$ for a given allele frequency trajectory. Therefore, we coded a function to perform discrete forward in time computations based on equations (1) and (2) to obtain deterministic allele frequency trajectories. This function was used within the nonlinear least squares routine *nls* in R, to obtain parameter estimates for both $s$ and $h$. While this approach worked for a broad range of parameter combinations, the algorithm did not converge in a few cases (see supplementary fig. S11, Supplementary Material online).

To rule out that an estimated selection coefficient could be obtained by random genetic drift, we simulate neutral ($s = 0$) allele frequency trajectories and estimate $s$ with either LLS or NLS. The $P$ value for the null hypothesis of neutrality is the fraction of simulations resulting in estimates of $s$, which are at least as large as the selection coefficient inferred from the empirical data.

Assuming that $s$ and $h$ do not change, replicated trajectories improve the inference when the information of all replicates is combined. Assuming that all replicates start from the same allele frequencies, we averaged the allele frequencies at each time point across replicates. In the case of small $N_e$, selected alleles with a low starting frequency have a high probability of being lost which would bias the inferred selection parameters. We therefore excluded observed allele frequencies in a replicate from the time an allele gets lost onwards. In cases where the starting allele frequency is low compared to $N_e$, this conditioning on fixation needs to be accounted for. We therefore propose for a further bias correction in such cases, that is based on neutral simulations for each locus individually, leading to corrected empirical consensus trajectories. The corrected consensus trajectories can then be used in such cases for LLS- and NLS-based inference of $s$ (and $h$).

## Precision and Bias of Estimates

To assess the precision and bias of estimated selection coefficients $s$, we used the relative root–mean–squared error (rRMSE), a measure of the relative deviation between estimated and true selection coefficient. It was computed for bins of $n$ loci:

$$rRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n} (\hat{s}_i - s_i)^2}{n}}}{\bar{s}}, \quad (7)$$

where $\hat{s}_i$ is the estimated and $s_i$ is the true selection coefficient of locus $i$, and $\bar{s}$ is the mean selection coefficient of all $n$ loci. In the comparison between LLS, CLEAR, and WFABC, estimates were binned according to either $s$ or $p_0$ and then the rRMSE was computed for each bin. Similarly, when assessing the accuracy of LLS under a state-of-the-art E&R design, ten million estimates were divided into 2,500 bins depending on $s$ and $p_0$. This way, the dependence of the rRMSE on $s$ and $p_0$ has been illustrated. Confidence intervals of rRMSE were estimated by bootstrapping loci in each bin.

## Benchmark between LLS, CLEAR, and WFABC

We compared LLS-based estimates of $s$ to those obtained with CLEAR (downloaded from https://github.com/airan mehr/clear on May 5, 2017) and WFABC (version 1.1) using the same allele frequency trajectories. To match the default parameter setting of WFABC, only trajectories with a minor allele frequency larger than 0.01 at any of the time points were used. We provided the true effective population size to WFABC (fixed_N 600). A flat prior between $-0.2$ (min_s) and 1.0 (max_s) was used. For CLEAR, $N_e$ was set to 300 (–N 300) and the maximum selection coefficient was set to 1, scanning in steps of 0.01 (maxS = 1 and stepS = 0.01). We further specified that all 300 individuals were sampled using Pool-Seq ($n = 300$).

The maximal deviation of ABC-based estimates from the true value depends on the range of the prior distribution, while LLS estimates could theoretically range from negative and positive infinity. To compare the methods, we projected LLS and CLEAR estimates to the range of the priors of WFABC.

Allele frequency trajectories were simulated under the Wright–Fisher model for diploid individuals assuming linkage equilibrium between loci. Starting allele frequencies were drawn from a uniform distribution between 0 and 1 and selection coefficients were between 0 and 0.3. To mimic sampling noise of Pool-Seq, first sequence coverage for each locus was drawn from a Poisson distribution and then binomial sampling was performed with sample size matching coverage.

## Analysis of E&R Study in Yeast

Allele frequency trajectories of all SNPs were downloaded from http://wfitch.bio.uci.edu (last accessed April 10, 2017). Without further filtering, selection coefficients were estimated and $P$ values under the null hypothesis of random genetic drift were computed with LLS for all 75,410 loci. We assumed a population of haploid individuals, since the life-cycle imposed in this E&R study involved haploids for most of the time. We used the Benjamini–Hochberg correction (Benjamini and Hochberg 1995) to correct the $P$ values for multiple hypothesis testing. We only considered outliers, which were supported by at least one additional outlier SNP within 1 kb.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## References

Andries K, Verhasselt P, Guillemont J, Göhlmann HWH, Neefs J-M, Winkler H, Van Gestel J, Timmerman P, Zhu M, Lee E. 2005. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 307(5707):223–227.

Baldwin-Brown JG, Long AD, Thornton KR. 2014. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Mol Biol Evol*. 31(4):1040–1055.

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

Barton NH. 2000. Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci*. 355(1403):1553–1562.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57: 289–300.

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

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

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

Burke MK, Liti G, Long a. D. 2014. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Mol Biol Evol*. 31(12):3228–3239.

Chan YF, Jones FC, McConnell E, Bryk J, Bünger L, Tautz D. 2012. Parallel selection mapping using artificially selected mice reveals body weight control loci. *Curr Biol*. 22(9):794–800.

Crisci JL, Poh YP, Bean A, Simkin A, Jensen JD. 2012. Recent progress in polymorphism-based population genetic inference. *J Hered*. 103(2):287–296.

Crow JF, Kimura M. 1970. An introduction to population genetics theory. New York: Harper and Row

Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, Liao X, Djari A, Rodriguez SC, Grohs C, et al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 46(8):858–865.

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

Ferrer-Admetlla A, Leuenberger C, Jensen JD, Wegmann D. 2016. An approximate markov model for the Wright–Fisher diffusion and its application to time series data. *Genetics* 203(2):831–846.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol*. 31(5):1275–1291.

Fisher RA, Ford EB. 1947. The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* 1(2): 143–174.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180(2):977–993.

Foll M, Poh YP, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas AS, Ewing G, Liu P, Wegmann D, et al. 2014. Influenza virus drug resistance: a time-sampled population genetics perspective. *PLoS Genet*. 10(2):e1004185.

Foll M, Shim H, Jensen JD. 2015. WFABC: a Wright–Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol Ecol Resour*. 15(1):87–98.

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186(1):207–218.

Griffin PC, Hangartner SB, Fournier-Level A, Hoffmann AA. 2017. Genomic trajectories to desiccation resistance: convergence and divergence among replicate selected drosophila lines. *Genetics* 205(2):871–890.

Haasl RJ, Payseur BA. 2016. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Mol Ecol*. 25(1):5–23.

Hill WG, Robertson a. 1966. The effect of linkage on limits to artificial selection. *Genet Res*. 8(3):269–294.

Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat Genet*. 44(7):808–811.

Illingworth CJR, Parts L, Schiffels S, Liti G, Mustonen V. 2012. Quantifying selection acting on a complex trait using allele frequency time series data. *Mol Biol Evol*. 29(4):1187–1197.

Iranmehr A, Akbari A, Schlötterer C, Bafna V. 2017. Clear: composition of likelihoods for evolve and resequence experiments. *Genetics* 206(2):1011–1023.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170(3): 1401–1410.

Jewett EM, Steinrücken M, Song YS. 2016. The effects of population size histories on estimates of selection coefficients from time-series genetic data. *Mol Biol Evol*. 33:048355.

Jónás A, Taus T, Kosiol C, Schlötterer C, Futschik A. 2016. Estimating the effective population size from temporal allele frequency changes in experimental evolution. *Genetics* 206:723–735.

Kao KC, Sherlock G. 2008. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nat Genet*. 40(12):1499–1504.

Kessner D, Novembre J. 2015. Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. *Genetics* 199(4):991–1005.

Khatri BS. 2016. Quantifying evolutionary dynamics from variant-frequency time series. *Sci Rep*. 6:32497.

Kim Y, Nielsen. R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167(3):1513–1524.

Kim Y, Stephan W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160(2):765–777.

Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24):3435–3436.

Kofler R, Schlötterer C. 2014. A guide for the design of evolve and resequencing studies. *Mol Biol Evol*. 31(2):474–483.

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

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

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583): 499–503.

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

McDonald MJ, Rice DP, Desai MM. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* 531(7593):233–236.

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

Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol Ecol.* 21(20):4931–4941.

Pavlidis P, Živković D, Stamatakis A, Alachiotis N. 2013. SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol.* 30(9): 2224–2234.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419(6909):832–837.

Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals: mining genome-wide polymorphism data without big funding. *Nat Rev Genet.* 15(11):749–763.

Schraiber JG, Evans SN, Slatkin M. 2016. Bayesian inference of natural selection from allele frequency time series. *Genetics* 203(1):493–511.

Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, Bergethon K, Shaw AT, Gettinger S, Cosper AK, et al. 2011. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med.* 3(75):75ra26.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.* 23(1):23–35.

Steinrücken M, Bhaskar A, Song YS. 2014. A novel spectral method for inferring general diploid selection from time series genetic data. *Ann Appl Stat.* 8(4):2203–2222.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460.

Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C, et al. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536(7615):165–170.

Tenaillon O, Rodriguez-Verdugo a, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* 335(6067):457–461.

Terhorst J, Schlötterer C, Song YS. 2015. Multi-locus analysis of genomic time series data from experimental evolution. *PLoS Genet.* 11(4):e1005069.

Thornton KR, Jensen JD, Becquet C, Andolfatto P. 2007. Progress and prospects in mapping recent selection in the genome. *Heredity* 98(6):340–348.

Tobler R, Franssen SU, Kofler R, Orozco-Terwengel P, Nolte V, Hermisson J, Schlötterer C. 2014. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol Biol Evol.* 31(2):364–375.

Topa H, Jónás Á, Kofler R, Kosiol C, Honkela A. 2015. Gaussian process test for high-throughput sequencing time series: application to experimental evolution. *Bioinformatics* 31(11):1762–1770.

Turner TL, Miller PM. 2012. Investigating natural variation in drosophila courtship song by the evolve and resequence approach. *Genetics* 191(2):633–642.

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

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4(3):e72.

Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE. 2011. Second-order selection for evolvability in a large *Escherichia coli* population. *Science* 331(6023):1433–1436.

Wright S. 1948. On the roles of directed and random changes in gene frequency in the genetics of populations. *Evol Int J Org Evol.* 2(4):279–294.

Zhou D, Udpa N, Gersten M, Visk DW, Bashir A, Xue J, Frazer K. a, Posakony JW, Subramaniam S, Bafna V, et al. 2011. Experimental selection of hypoxia-tolerant *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 108(6):2349–2354.