

A Guide for the Design of Evolve and Resequencing Studies

Robert Kofler¹ and Christian Schlötterer*¹

¹Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria

*Corresponding author: E-mail: christian.schloetterer@vetmeduni.ac.at.

Associate editor: John Parsch

Abstract

Standing genetic variation provides a rich reservoir of potentially useful mutations facilitating the adaptation to novel environments. Experimental evolution studies have demonstrated that rapid and strong phenotypic responses to selection can also be obtained in the laboratory. When combined with the next-generation sequencing technology, these experiments promise to identify the individual loci contributing to adaptation. Nevertheless, until now, very little is known about the design of such evolve & resequencing (E&R) studies. Here, we use forward simulations of entire genomes to evaluate different experimental designs that aim to maximize the power to detect selected variants. We show that low linkage disequilibrium in the starting population, population size, duration of the experiment, and the number of replicates are the key factors in determining the power and accuracy of E&R studies. Furthermore, replication of E&R is more important for detecting the targets of selection than increasing the population size. Using an optimized design, beneficial loci with a selective advantage as low as $s = 0.005$ can be identified at the nucleotide level. Even when a large number of loci are selected simultaneously, up to 56% can be reliably detected without incurring large numbers of false positives. Our computer simulations suggest that, with an adequate experimental design, E&R studies are a powerful tool to identify adaptive mutations from standing genetic variation and thereby provide an excellent means to analyze the trajectories of selected alleles in evolving populations.

Key words: experimental evolution, evolve & resequencing, next-generation sequencing, bioinformatics, adaptation.

Introduction

The importance of standing genetic variation to the adaptation of natural populations is well recognized (Hermisson and Pennings 2005; Przeworski et al. 2005; Garud et al. 2013). Experimental evolution studies have successfully demonstrated that standing genetic variation enables a rapid phenotypic response in laboratory populations (Burke et al. 2010; Turner et al. 2011; Orozco-terWengel et al. 2012). The advent of next-generation sequencing (NGS) has generated renewed interest in experimental evolution as it now has become possible to identify the loci contributing to adaptation (Burke et al. 2010; Parts et al. 2011; Orozco-terWengel et al. 2012; Remolina et al. 2012). This approach, which has been termed uniformly evolve & resequencing (E&R; Turner et al. 2011), promises to unify two branches of genetics that have been separated for most of the 20th century: molecular genetics and population genetics (Rose et al. 2011). E&R studies have the potential to identify the loci contributing to adaptation (molecular genetics) as well as to analyze the trajectories of these loci during adaptation (population genetics). At present little is known on how the design of E&R studies can be optimized to identify the maximal number of selectively favored loci (Burke and Long 2012). In fact, it is not yet known whether a few causal loci can be distinguished from millions of neutrally evolving ones, raising the important question whether E&R studies can work at all. Here, we address these questions by forward simulations of entire genomes and evaluate the power of different experimental designs to

identify beneficial loci. We base the simulations on *Drosophila melanogaster*, because it is widely used in experimental evolution studies (Rose 1984; Burke et al. 2010; Burke 2012; Kawecki et al. 2012; Orozco-terWengel et al. 2012). We demonstrate that the experimental design has a pronounced influence on the power to detect selected loci. Although the major challenge for weakly selected loci is the distinction from unlinked neutrally evolving ones, strongly selected loci frequently result in a large number of linked false positives. Increasing the number of replicates, population size, and the duration of the experiment increases the power of E&R studies. Using such an optimized design, up to 56% of the selected loci can be detected at the resolution of a single nucleotide. Even weakly selected loci with a selective advantage as low as $s = 0.005$ can be identified in the laboratory. Our computer simulations suggest that, provided an adequate experimental design, E&R studies are a powerful tool to identify adaptive mutations from standing genetic variation and thereby provide an excellent means for analyzing the trajectories of selected alleles in evolving populations.

Results

To minimize the parameter space for the forward simulations, we picked commonly used default conditions and varied only the parameter of interest within these defaults. By default, we used a base population consisting of 1,000 homozygous genomes (i.e., inbred isofemale lines) that capture the pattern of natural variation found in a population of *D. melanogaster*

© The Author 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

from Vienna (Bastide et al. 2013) (for details on the base population, see [supplementary results 1.1, Supplementary Material](#) online), the recombination rate of *D. melanogaster* (Fiston-Lavier et al. 2010), 3 biological replicates, and 60 generations of selection (see Materials and Methods). For every experimental design, ten independent simulations were performed. We separately simulated 150 codominant ($h = 0.5$) loci with two different selection coefficients of $s = 0.1$ and $s = 0.025$ (see Materials and Methods), which roughly cover the range of selection coefficients detected in experimental evolution studies (Imhof and Schlötterer 2001; Kassen and Bataillon 2006). Beneficial mutations were randomly picked from the segregating sites in the base population with no distinction being made between the ancestral or derived allele. Simulations were performed for the major autosomes and chromosomal regions with low recombination rates being excluded because they inflate the false-positive rate ([supplementary results 1.3 and fig. S4, Supplementary Material](#) online). Selected loci were identified by contrasting allele counts between the base and the evolved populations. Although several different test statistics have been used in E&R studies (Cochran-Mantel-Haensel [CMH] test [Orozco-Wengel et al. 2012], diffStat [Turner et al. 2011], association statistic [Turner and Miller 2012], and the F_{ST} [Remolina et al. 2012]), we used the CMH test because it showed the best performance ([supplementary results 1.2 and fig. S3, Supplementary Material](#) online).

The performance of different experimental designs was compared using receiver-operating characteristic (ROC) plots (Hastie et al. 2009) that contrasts the true-positive rate (TPR) with the false-positive rate (FPR). The TPR can be calculated as $TPR = TP / (TP + FN)$, where TP refers to the true positives and FN to the false negatives. The FPR instead can be calculated as $FPR = FP / (FP + TN)$, where FP refers to the false positives and TN to the true negatives. The

significance of each ROC curve in turn was summarized with the partial area under the curve (pAUC) statistic which is a subregion of the area under the ROC curve (McClish 1989; Thompson and Zucchini 1989). As we are here mostly interested in the power to identify beneficial loci at low FPR, we use the pAUC with a FPR range from 0 to 0.01 [$pAUC = \int_0^{0.01} ROC(f)df$].

Influence of the Number of Generations

The duration of an experimental evolution study is a key parameter, because it determines the feasibility of an experiment. We varied the number of generations from 10 to 300 which roughly corresponds to experimental lengths from 4 months to 10 years, respectively. As expected, for both selection coefficients, the number of true positives increased with the duration of the selection experiment (Kruskal–Wallis rank-sum test with pAUC; $s = 0.1$: $\chi^2 = 46.07$, $P = 5.8e-07$; $s = 0.025$: $\chi^2 = 94.72$, $P < 2.2e-16$). Overall, the effect was more pronounced for weakly selected loci than for strongly selected ones ([fig. 1 and supplementary fig. S15, Supplementary Material](#) online). With fewer generations, the differences between strong and weak selection became more pronounced. Based on an $FPR < 0.01$, up to 60% of the selected loci can be identified, independent of the selection coefficient. Notably, only a moderate number of generations was necessary to identify a large number of selected sites. For example, 60 generations of selection resulted already in the identification of 48.4% ($s = 0.1$) or 36.2% ($s = 0.025$) of the beneficial loci.

Influence of the Number of Replicates

One of the most powerful features of experimental evolution is the possibility of replication. We evaluated different numbers of replicates ranging from 1 to 20. As expected, increasing the number of replicates had a significant effect for both

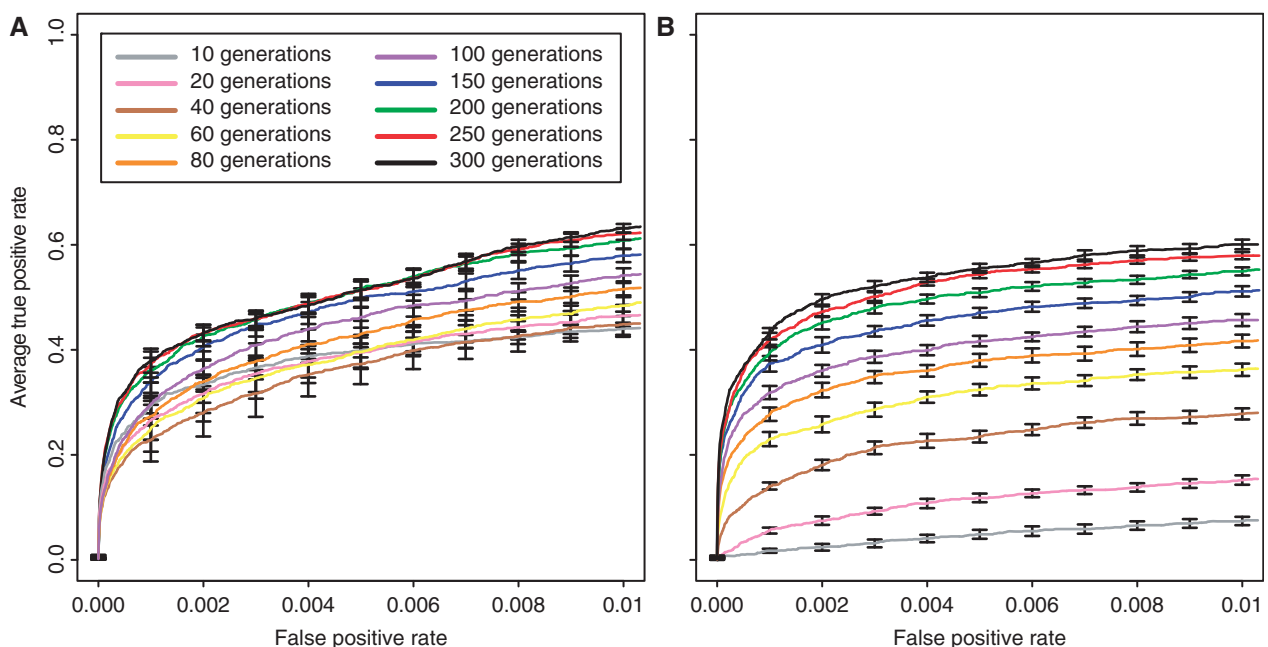


FIG. 1. Influence of the number of generations on identification of beneficial loci with selection coefficients of $s = 0.1$ (A) and $s = 0.025$ (B).

selection coefficients (Kruskal–Wallis rank-sum test with pAUC; $s = 0.1$: $\chi^2 = 25.45$, $P = 0.00011$; $s = 0.025$: $\chi^2 = 54.22$, $P = 1.88e-10$). Similar to the number of generations, also the number of replicates had a more pronounced effect for weakly selected sites (fig. 2 and supplementary fig. S16, Supplementary Material online). Interestingly, given a sufficiently large number of replicates, weakly selected loci can be more readily identified than strongly selected ones (Wilcoxon rank-sum test with pAUC [$0 \leq \text{FPR} \leq 0.001$]; $W = 100$; $P = 1.1e-05$; fig. 2). We reason that this is because neutral variants linked to weakly selected loci have more time to recombine onto neutral haplotypes as opposed to variants on strongly selected haplotypes, which will rapidly become fixed (supplementary results 1.5, Supplementary Material online). Although five replicates seem to be sufficient for the identification of strongly selected loci, reliable detection of weakly selected ones requires more replicates (fig. 2).

Influence of the Population Size

We also modified the number of individuals evolving in each replicate, ranging from 250 to 8,000. For both selection coefficients, we noticed that a larger population size improves the power to identify beneficial loci (Kruskal–Wallis rank-sum test with pAUC; $s = 0.1$: $\chi^2 = 49.80$, $P = 1.5e-09$; $s = 0.025$: $\chi^2 = 55.61$, $P = 9.8e-11$). Similar to the results for the number of replicates and generations, weakly selected sites benefitted more from an increase in population size (fig. 3 and supplementary fig. S17, Supplementary Material online). One complication arising from the comparison of different population sizes is that larger populations contained more starting variation (ranging from 2,011,991 to 3,129,057 single-nucleotide polymorphism [SNPs] in our study). Thus,

for a given FPR cutoff, the number of significant loci may depend on the population size. To account for this, we repeated the analysis using absolute numbers of significant loci but obtained similar results (supplementary fig. S18, Supplementary Material online).

Number of Haploid Genomes in the Base Population

In E&R studies, the base population is typically established from isofemale lines, where the number of lines is frequently smaller than the population size. In such cases, multiple females from each isofemale line are used to generate the base population. We tested the influence of the number of haploid genomes in the base population by varying the number of haploid genomes from 50 to 2,000. The respective base populations contained between 1,482,521 and 2,685,539 segregating loci. The number of haploid genomes had a significant influence on the power to identify beneficial loci for both selection coefficients (Kruskal–Wallis rank-sum test with pAUC; $s = 0.1$: $\chi^2 = 45.54$, $P = 1.1e-08$; $s = 0.025$: $\chi^2 = 39.92$, $P = 1.55e-07$). Contrary to our previous results, we found that strongly selected loci benefit more from an increase in the number of haploid genomes than weakly selected ones (fig. 4 and supplementary fig. S21, Supplementary Material online). The same trend was seen if the absolute number of significant loci is used instead of the FPR (supplementary fig. S20, Supplementary Material online). As neutral variants linked to strongly selected loci have less opportunity to recombine onto neutral haplotypes before becoming fixed (or at least reaching high frequencies) than variants linked to weakly selected loci, the identification of strongly selected loci benefits more from lower initial levels of linkage disequilibrium in the base population (supplementary results 1.6, Supplementary Material online).

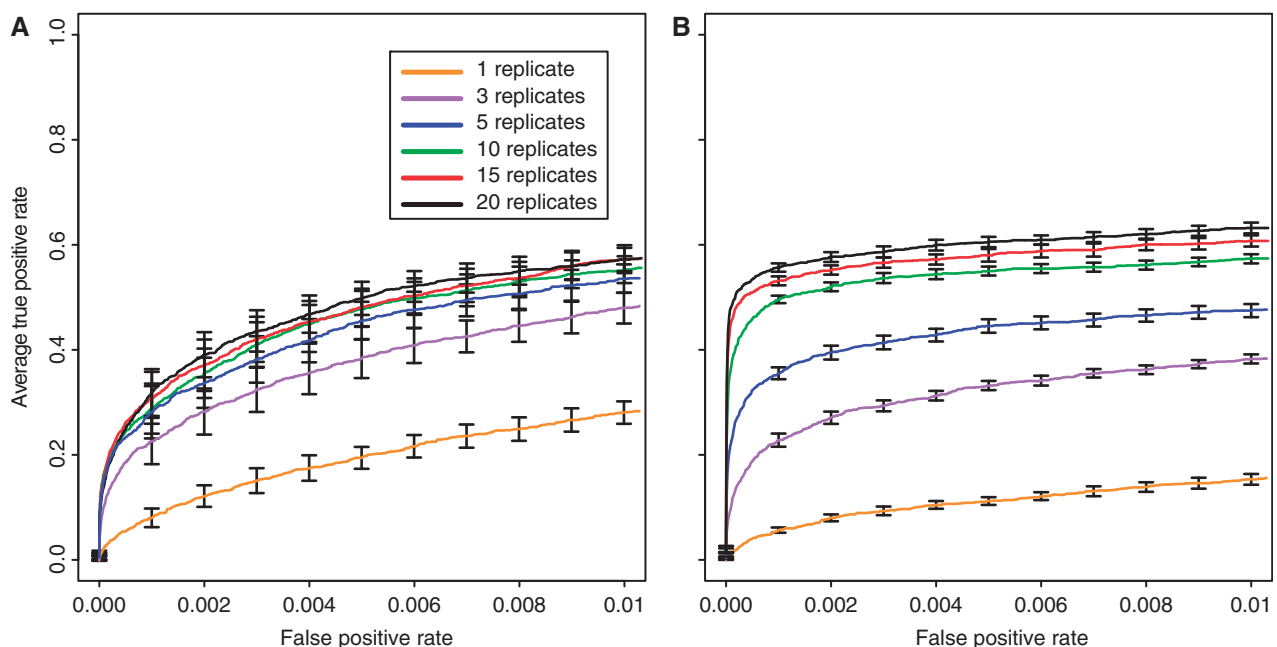


Fig. 2. Influence of the number of biological replicates on identification of beneficial loci with selection coefficients of $s = 0.1$ (A) and $s = 0.025$ (B).

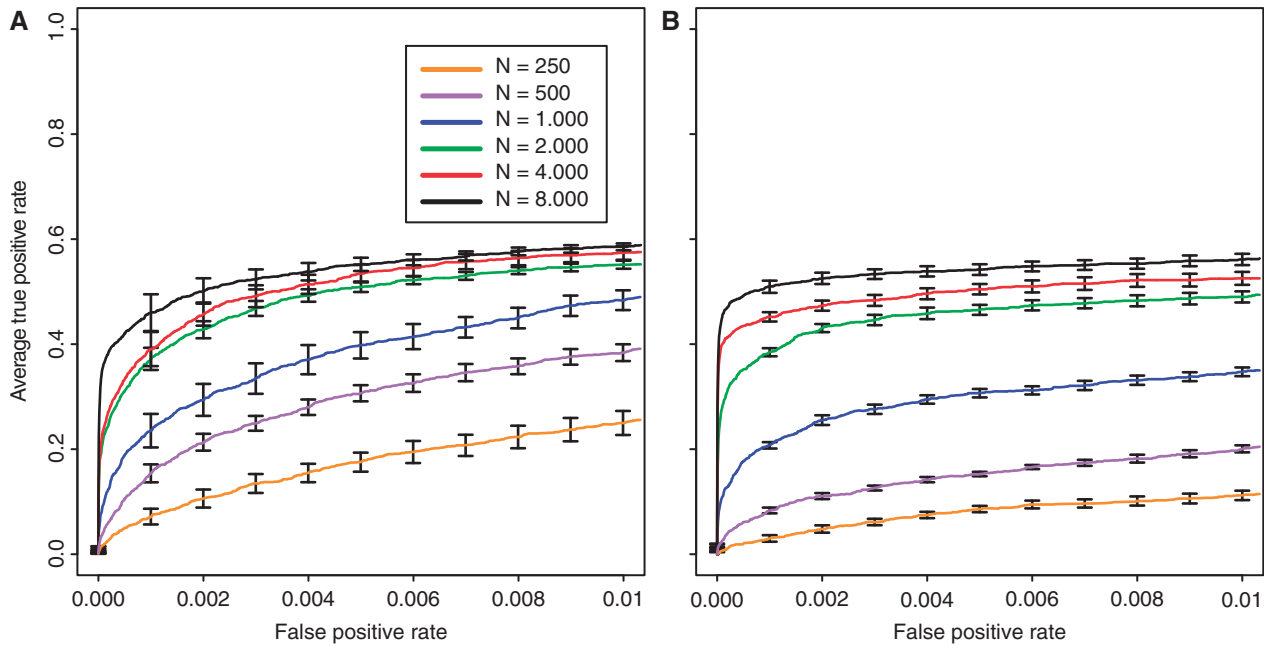


FIG. 3. Influence of the population size on identification of beneficial loci with selection coefficients of $s = 0.1$ (A) and $s = 0.025$ (B).

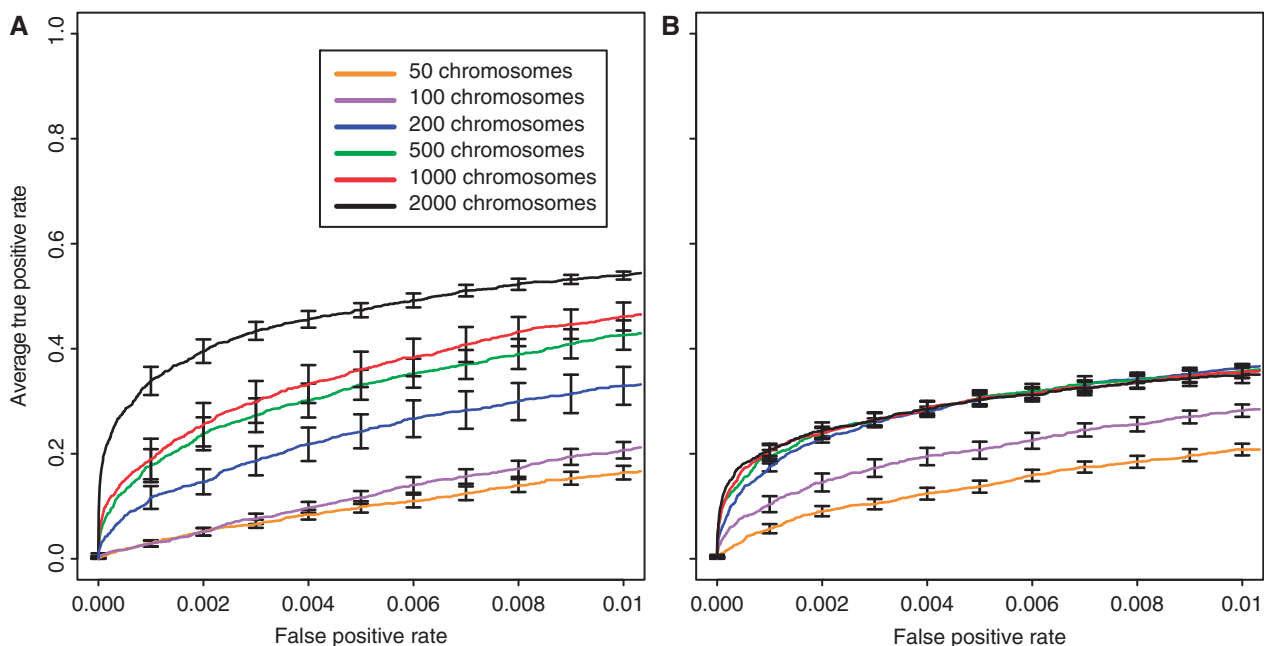


FIG. 4. Influence of the number of haploid genomes in a population with size of 1,000 on the identification of beneficial loci with selection coefficients of $s = 0.1$ (A) and $s = 0.025$ (B). Only homozygous individuals were used except in the population with 2,000 haploid genomes.

Generating an Outbred Population prior to the Experiment

A frequently used strategy is to generate an outbred population from a limited number of isofemale lines prior to the selection experiment. By propagating the flies for several generations at a large population size, linkage will be broken up, resembling a natural outbred population (Turner et al. 2011; Huang et al. 2012). We evaluated the efficiency of this procedure by using a population of 1,000 individuals founded by

50 inbred isofemale lines. The populations evolved under neutrality for up to 100 generations. During this neutral evolution phase, variation was lost by genetic drift, resulting in a loss of up to 10.7% of the beneficial alleles. We found that neutral evolution prior to E&R has a significant influence on the ability to identify beneficial loci of both selection coefficients (Kruskal–Wallis rank-sum test with pAUC; $s = 0.1$: $\chi^2 = 17.02$, $P = 0.009$; $s = 0.025$: $\chi^2 = 22.43$, $P = 0.001$). For strongly selected loci, the power to identify beneficial loci increased with the duration of the preceding neutral

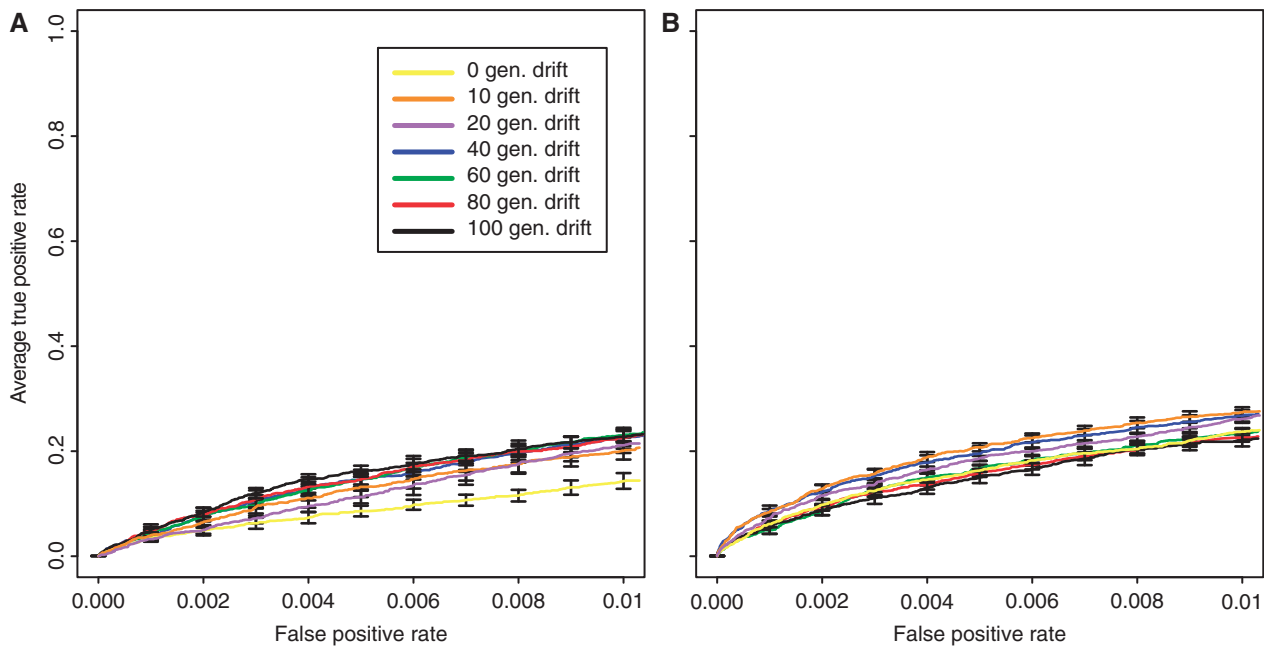


Fig. 5. Influence of drift prior to experimental evolution on the identification of beneficial loci with selection coefficients of $s = 0.1$ (A) and $s = 0.025$ (B). A number of haploid genomes of 50 and a population size of 1,000 were used. In the base population, every haploid genome was thus present in 20 homozygous individuals.

evolution (fig. 5 and supplementary fig. S21, Supplementary Material online). For weakly selected loci, we noticed a trade-off and only an intermediate duration of the neutral evolution phase improved the power of the experiment. Nevertheless, in all cases, the improvement in power was modest relative to the other factors evaluated in this study.

Tradeoff between Population Size and Number of Replicates

Because the power to identify beneficial loci was positively correlated with both the population size and the number of replicates, we were interested to determine the optimal strategy given finite resources. For simplicity, we assumed that costs scale linearly with the total number of individuals maintained during the experiment. We analyzed a target of 8,000 flies per experiment, ranging from a single replicate with 8,000 flies to 16 replicates with 500 flies. We found that increasing the number of replicates improved the power to identify strongly selected loci, but no effect was observed for weakly selected sites (Kruskal–Wallis rank-sum test with pAUC; $s = 0.1$: $\chi^2 = 27.52$, $P = 1.6e-05$; $s = 0.025$: $\chi^2 = 0.3$, $P = 1.0$; fig. 6 and supplementary fig. S22, Supplementary Material online; for $s = 0.005$, see supplementary fig. S23, Supplementary Material online). At small population sizes, genetic drift has a strong influence on the trajectories of selected loci. This provides an advantage for the identification of strongly selected loci as it may delay rapid rises in frequency of the selected loci and thus provide more opportunities for linked variants to recombine onto neutral haplotypes (supplementary results 1.7, Supplementary Material online).

Detection Limit of Beneficial Loci with Different Experimental Designs

One essential question is how the experimental design affects the minimum detectable effect size of beneficial loci. We addressed this question by evaluating two extreme experimental designs, one low-budget and one high-budget design. The low-budget design encompasses three replicates, each with a population size of 500, maintained for 60 generations. The high-budget design is based on 10 replicates of 2,000 individuals propagated for 120 generations. For both experimental designs, we evaluated the power to identify beneficial loci for selection coefficients ranging from 0.5 to 0.001. As expected, the selection coefficient had a significant influence on the ability to identify beneficial loci for both experimental designs (fig. 7 and supplementary fig. S24, Supplementary Material online; Kruskal–Wallis rank-sum test with pAUC; low budget: $\chi^2 = 53.15$, $P = 3.14e-10$; high budget: $\chi^2 = 55.58$, $P = 9.9e-11$). Very weakly selected loci ($s = 0.001$) could not be detected in either experimental design. For all other selection coefficients, the high-budget design drastically outperformed the low-budget design (Wilcoxon rank-sum test with pAUC; $s = 0.5-0.005$: $W = 0$, $P = 1.1e-05$; $s = 0.001$: $W = 25$, $P = 0.06$). Interestingly, for both experimental designs, intermediate selection coefficients ($s = 0.05$) performed best (fig. 7 and supplementary fig. S24, Supplementary Material online). We also tested the performance with beneficial loci having different selection coefficients and found that fewer targets of selection were correctly identified (supplementary results 1.12, Supplementary Material online). The most likely explanation for this observation is that neutral variants hitchhiking with strongly

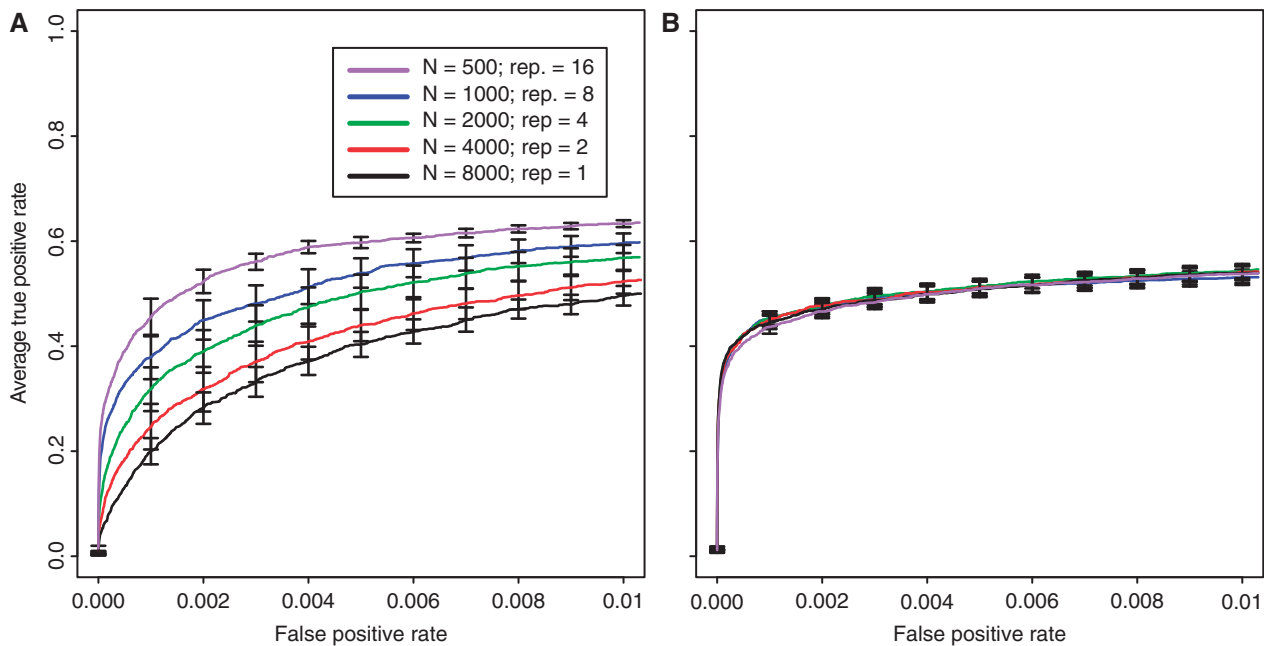


FIG. 6. Influence of the tradeoff between population size and number of replicates on the identification of beneficial loci with selection coefficients of $s = 0.1$ (A) and $s = 0.025$ (B).

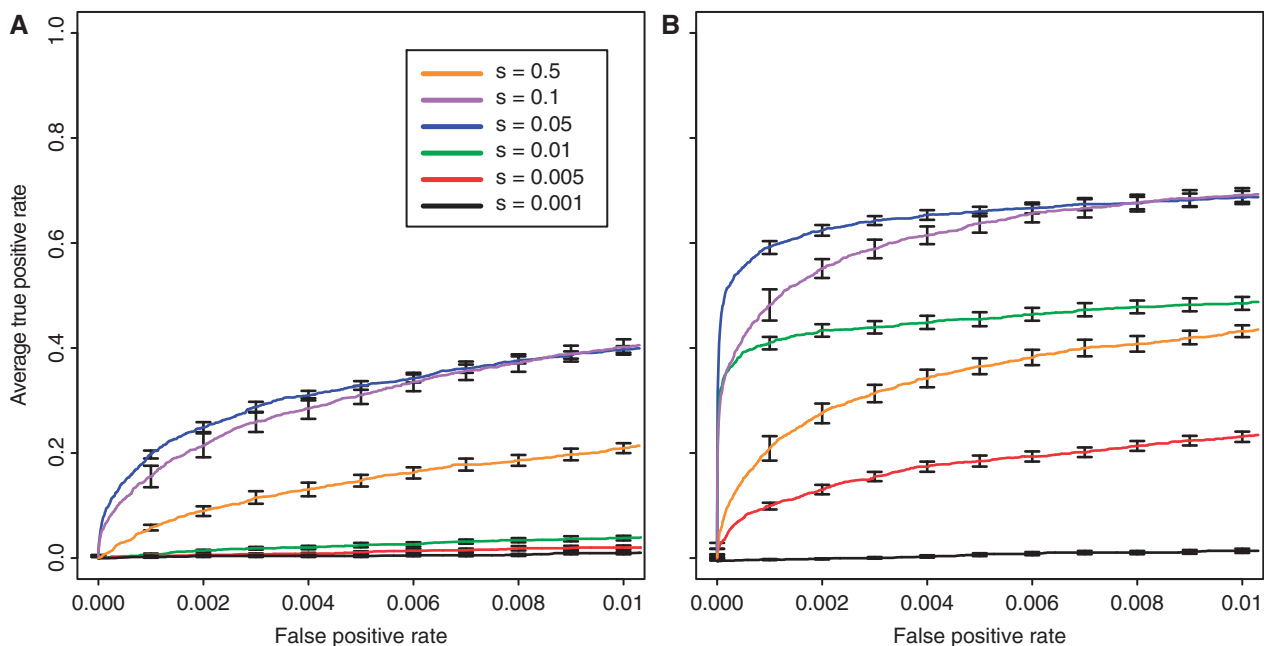


FIG. 7. Power to identify beneficial loci of different effect sizes with a low-budget (A) and a high-budget (B) study design. See text for definition of low-budget and high-budget study design.

selected loci will have more pronounced allele frequency changes than weakly selected loci, so that strongly selected loci are basically concealing the response to selection of weakly selected loci.

Another way to evaluate the power of different designs is the absolute number of selected loci found among the most significant SNPs. The results for the high-budget design were particularly encouraging since among the ten most significant loci, up to seven true positives were detected (table 1). Additionally, of the 100 most significant loci,

about 50 were true targets of selection. On the other hand, the low-budget design only resulted in moderately encouraging results. When the ten most significant loci were considered, at most two true positives were found. Considering a larger number of most significant SNPs only improved the result marginally (table 1). To allow for a more intuitive comparison of the two experimental designs with actual experimental evolution studies, we have visualized the results using Manhattan plots (low-budget design: supplementary figs. S25–S30, Supplementary Material online; high-budget

Table 1. Power to Identify Beneficial Loci of Different Selection Coefficients (s) with a Low-Budget (lb) and a High-Budget (hb) Study Design.

s		10	20	50	100	200	500	1,000	2,000	5,000	10,000
0.5	lb	0.3	0.5	0.8	1.7	2.8	5.4	9	13.5	22.5	29.8
	hb	1.5	2.6	4.7	7	9.5	15.5	21.6	29.6	42.9	53
0.1	lb	2.4	3.4	4.8	6.5	8.4	12.4	16.9	23.3	35.9	46
	hb	6.6	13.1	27.8	37.8	44.6	53.3	59.6	67.8	82.2	91.9
0.05	lb	2.4	3.5	5.5	8.2	11.7	16.6	21.7	29.4	39.7	49.1
	hb	7.7	15.4	35.3	52.4	65	76.4	80.9	86.4	93.2	97.7
0.01	lb	0	0	0.1	0.1	0.1	0.3	0.5	1	2.5	3.5
	hb	7.4	14.6	30.6	40.5	48	53.6	56.9	60.9	65.2	67.6
0.005	lb	0	0	0	0	0	0.1	0.4	0.5	0.9	1.5
	hb	1.7	2.2	3.2	4	5.2	7.7	11.1	14.4	20.1	26.8
0.001	lb	0	0	0	0	0	0	0.2	0.3	0.6	0.6
	hb	0	0	0	0	0	0.1	0.2	0.3	0.6	1.3

NOTE.—Reported values are average numbers of true positive identified with the given number of most significant SNPs. See text for definition of low- and high-budget study design.

design: [supplementary figs. S31–S36, Supplementary Material online](#)).

Discussion

Our simulations revealed an interesting difference between strongly and weakly selected sites. Experimental designs with low power typically favor strongly selected sites. This advantage is, however, frequently lost with more powerful designs. The power of an experimental design ultimately depends on the signal-to-noise ratio (the signal is the average log-transformed P value of selected sites and the noise the average log-transformed P value of nonselected sites). Strongly selected loci were characterized by a strong signal that was easily recognized even with less powerful experimental designs. In contrast, the signal of weakly selected loci needs to be detected with powerful experimental designs and distinguished from the background noise caused by neutrally evolving loci. Interestingly, once weakly selected loci are detected, the noise caused by hitchhiking variants is lower than for strongly selected loci. This different behavior of strongly and weakly selected loci is further illustrated in the online supplement by analyzing signal and noise separately ([supplementary results 1.5, Supplementary Material online](#)).

We showed that the following key factors could increase the sensitivity of an experimental design: 1) larger population size, 2) more replicates, 3) increasing number of generations, and 4) chromosome diversity at the beginning of the experiment. The benefits of higher chromosome diversity largely results from uncoupling linkage between selected and neutral loci. A longer duration of the selection experiment and a higher number of replicates, however, have a 2-fold influence. On one hand, they increase the sensitivity, particularly for weakly selected loci, by making consistent allele frequency change across replicates more apparent. On the other hand, they will increase the number of recombination events that occur during the experiment which will also reduce the noise by uncoupling selected loci from neutral linked ones. A larger population size has a similar effect: by reducing genetic drift, even weaker signals of selection can be

detected. Furthermore, the number of recombinant haplotypes is increased, uncoupling true and false positives.

Previously, we simulated genetic drift for individual loci to obtain a null expectation of P values from the CMH test that is based on neutrality (Orozco-terWengel et al. 2012). In contrast to Orozco-terWengel et al. (2012), we explicitly model linkage of neutral loci to the target of selection by computer simulations. Although the changes in allele frequency of individual loci (beneficial, neutral, or deleterious) can be easily worked out analytically (Haldane 1927; Kimura 1955, 1962), an exact analytical treatment of E&R will be challenging as usually 1) the allele frequency differences of several beneficial loci concurrently segregating in a population need to be calculated, 2) heterogeneity among biological replicates needs to be considered and 3) the allele frequency differences of millions of neutrally evolving loci (either hitchhiking or drifting) need to be calculated. However, even with extensive computer simulations, as presented here, it will be difficult to explore the full spectrum of possible evolutionary scenarios. Therefore, many interesting scenarios for EE studies remain to be explored by future works. For example, several E&R studies noted that selected SNPs do not reach fixation (Burke et al. 2010; Parts et al. 2011; Burke 2012; Orozco-terWengel et al. 2012). Although the biological forces underlying these observations are not yet understood, more complex selection schemes, such as frequency-dependent selection or a quantitative trait model (Chevin and Hospital 2008), could be studied. Other interesting scenarios that were not evaluated include epistasis, sexually antagonistic selection, and antagonistic pleiotropy.

Selected loci were detected with the CMH test, which is based on contrasting allele frequency differences between the base population and the evolved populations. Although this approach favors alleles starting at intermediate frequencies ([supplementary results 1.11, Supplementary Material online](#)), our results are robust compared with similar test statistics ([supplementary results 1.8, Supplementary Material online](#)). However, test statistics based on time series data may lead to different conclusions. Time series data can be easily obtained for E&R studies by sequencing the evolving populations at

certain intervals, which could improve the power to identify beneficial loci (e.g., Illingworth et al. 2012). Currently, suitable test statistics for time series data from E&R studies are still lacking, but once such tests are available it will be important to reevaluate whether the recommendations derived here are still valid. Furthermore, these data sets will require evaluation of additional factors such as the optimal number of sampled time points and the optimal interval between.

In this study, we have identified beneficial loci based on the actual allele frequencies in the populations. In an E&R study, exact allele frequencies require sequencing of all individuals in a population separately. Although technically feasible, budget constraints preclude this approach. Rather, E&R studies have relied on Pool-Seq as a more cost-effective approach (Futschik and Schlötterer 2010) to estimate allele frequencies (Burke et al. 2010; Parts et al. 2011; Turner et al. 2011; Orozco-terWengel et al. 2012; Remolina et al. 2012; Turner and Miller 2012; Tobler et al. 2013). As coverage is the key parameter determining the ability to detect allele frequency differences, the well-described phenomenon of coverage heterogeneity along chromosomes (Dohm et al. 2008; Minoche et al. 2011) provides an additional challenge for the analysis of E&R studies: selected sites in genomic regions with a lower coverage will be less likely to be detected than loci with a higher coverage. We modeled the sampling properties of Pool-Seq and found that a coverage of about 50 is sufficient to identify strongly selected loci, whereas for weakly selected loci, coverages of at least 200 will be required for a reliable identification of selected loci (supplementary results 1.13, Supplementary Material online).

To reduce the parameter space, we performed all simulations with 150 beneficial loci, as we found this to be the highest number of loci that could be simultaneously selected without disproportionately reducing the efficacy of selection (supplementary results 1.4 and fig. S5, Supplementary Material online). Fewer beneficial loci will moderately increase the efficacy of selection of strongly selected loci, but there seems to be little effect for weakly selected loci (supplementary results 1.9, Supplementary Material online). The power to identify beneficial loci in an E&R study with few strongly selected loci may therefore be moderately higher than suggested by our results. We only simulated codominant loci ($h = 0.5$) in this work. The power to identify beneficial loci is slightly increased for dominant loci ($1 \geq h > 0.5$) whereas it drops slightly for recessive loci ($0.5 > h \geq 0$) (supplementary results 1.10, Supplementary Material online).

E&R is a powerful tool to pick up even weakly selected loci, but the challenge of E&R studies lies in the distinction between true targets of selection and false positives. Our simulation results based on a small number of replicates and moderate population sizes mirror those from previously published experimental studies, which also identified a large number of candidate loci (Burke et al. 2010; Turner et al. 2011; Orozco-terWengel et al. 2012; Remolina et al. 2012). Depending on the strength of selection and the experimental design, neutral loci linked to the target of selection could create high levels of false positives (supplementary fig. S31, Supplementary Material online). On the other hand, it is also

possible that false positives are not flanking targets of selection (supplementary fig. S35, Supplementary Material online). Hence, it is essential for E&R studies to use an experimental setup which assures the maximum enrichment of truly selected sites among the top candidates. As the true number of selected loci is not known in experimental studies, our simulation results are an important guideline for future E&R studies. Based on our computer simulations, we recommend to maximize the number of replicates, the population size, the duration of the experiment, and the number of haploid genomes segregating in the base population. We advise to prioritize replication over population size for species such as *Drosophila*, where maintenance of large population sizes is resource intensive. Furthermore, we note that replication reduces the damage if one population is experiencing a severe bottleneck, contaminated, or otherwise compromised during the experiment.

In summary, we showed that selected loci with a selective advantage as low as $s = 0.005$ can be reliably identified at the nucleotide level when using an optimal experimental design. Even when a large number of loci are selected simultaneously, up to 56% can be detected in the laboratory with a low rate of false positives. Our computer simulations suggest that, provided an adequate experimental design, E&R studies are a powerful tool to identify adaptive mutations from standing genetic variation and thereby will be a major contributor in elucidating the dynamics of adaptation from standing genetic variation.

Materials and Methods

Simulation Software

We developed MimicrEE (*Mimicing Experimental Evolution*), a forward simulation tool designed to model experimental evolution for fully sequenced genomes in the presence of multiple selected loci. This software comprises several Java applications that utilize the same core library (mimcore). MimicrEE performs forward simulations for a population of diploid individuals. These diploid genomes are provided as haplotypes with two haplotypes constituting a diploid genome (further details are given in the manual: <https://code.google.com/p/mimicree/wiki/Manual>, last accessed November 20, 2013). MimicrEE can handle large populations with several million segregating sites. So far, it has successfully been tested with 8,000 diploid individuals each having 4 million polymorphic loci. Memory requirements scale linearly with the population size and number of loci. During the forward simulations, the population size is kept constant. A list of selected loci may be provided and in case that no selected locus is specified, neutral simulations will be performed. For each selected locus, the selection coefficient (s), the dominance coefficient (h), and the nucleotide of the nonselected allele needs to be provided (w_{11}). The fitness of the heterozygous and homozygous individuals is given by $w_{11} = 1$, $w_{12} = 1 + hs$, and $w_{22} = 1 + s$ (see also Gillespie 2010). We assume multiplicative fitness when several selected loci are specified. No de novo mutations are considered as we are specifically interested in adaptation from standing genetic

variation. The forward simulations are performed with non-overlapping generations of hermaphrodites with selfing being excluded. At each generation, matings are performed, where mating success (i.e., number of offspring) scales linearly with fitness, until the total number of offspring in the population equals the targeted population size (fecundity selection). Each parent contributes a single gamete to the offspring wherein crossing over events are introduced according to the specified recombination rate. The recombination rate can be provided for arbitrarily sized windows. MimicrEE provides either haplotypes (MimicrEEHaplotype) or a summary of the allele frequencies (MimicrEESummary) at generations specified by the user as output.

Forward Simulations

We simulated 8,000 haploid genomes with fastsimcoal v1.1.8 (Excoffier and Foll 2011), which capture the pattern of natural variation of a *D. melanogaster* population. The diversity (θ_T) of a *D. melanogaster* population, captured in Vienna in fall 2010 (Bastide et al. 2013), was estimated with PoPoolation v1.2.1 (Kofler, Orozco-terWengel et al. 2011). The recombination rate was obtained from the *D. melanogaster* recombination rate calculator v2.2 (Fiston-Lavier et al. 2010). Recombination rate and diversity were specified in 100 kb windows, where the diversity was provided as mutation rate parameter adjusted to the pairwise distance in a window for a population size of 750,000 ($\mu = \theta/4N_e$). We excluded the X-chromosome and low recombining regions (< 1 cM/Mb), including the entire 4th chromosome, from the analysis (supplementary results 1.3 and fig. S4, Supplementary Material online) Thus, we performed our simulations only with the high recombining regions of the chromosomes 2L, 2R, 3L, and 3R. To build a population for the start of the experimental evolution simulation (base population), we randomly sampled haploid genomes (default = 1,000) from the 8,000 simulated haploid genomes. If the number of haploid genomes sampled for the base population was smaller than the targeted population size, homozygous individuals were formed to mimic an inbred isofemale line. We found that up to 150 beneficial loci could be included in a simulation run without disproportionately reducing the efficacy of selection (supplementary results 1.4 and fig. S5, Supplementary Material online). We therefore randomly picked 150 beneficial loci in each experiment using an algorithm involving three steps. First, loci were randomly chosen from the base population irrespective of the starting allele frequency; second, either the ancestral or the derived allele was randomly assigned as the beneficial one, and third if the starting allele frequency exceeded 80%, the respective loci were discarded and novel ones were randomly chosen instead. This threshold of 80% was used to increase the probability of detecting the selected allele. All selected loci were codominant ($h = 0.5$) and, if not mentioned otherwise, had a selection coefficient of $s = 0.1$ or $s = 0.025$. The recombination rate for the forward simulations was also obtained from the *D. melanogaster* recombination rate calculator v2.2 (Fiston-Lavier et al. 2010). As males in *Drosophila* are

not recombining and we simulated hermaphrodites, we divided the female recombination rate by 2. Forward simulations were performed with MimicrEESummary for the given numbers of generations (default = 60). Biological replicates for every experiment were acquired by repeating forward simulations several times (default = 3). Confidence intervals for all experimental designs are based on ten independent repetitions using the same parameters but different selected loci. Note that each repetition included biological replicates.

Statistical Analysis

We used the CMH test (Landis et al. 1978), implemented in PoPoolation2 r185 (Kofler, Pandey et al. 2011), to identify selected loci as this test showed the best performance (supplementary results 1.2 and fig. S3, Supplementary Material online). The CMH test is based on a meta-analysis of a $2 \times 2 \times k$ contingency table, where k represents the number of replicates. The contingency table contains for each replicate (k) the counts of the major and the minor allele (2), for the base and the evolved population (2). The null hypothesis is no differentiation between base and evolved populations ($P = 1.0$). We wrote custom scripts in Python for calculating the diffStat (Turner et al. 2011), the association statistic (Turner and Miller 2012), and the average pairwise F_{ST} . For statistical analysis, we used the R programming language (R Core Team 2012) and the library ROCR (Sing et al. 2005).

Availability

MimicrEE has been released under the Mozilla Public License 1.1 and is available from <https://code.google.com/p/mimicree/> (last accessed November 20, 2013). A manual can be found at <http://code.google.com/p/mimicree/wiki/Manual> (last accessed November 20, 2013) and test data including a tutorial are available from <http://code.google.com/p/mimicree/wiki/Walkthrough> (last accessed November 20, 2013). Furthermore, we made the entire protocol of the analysis used in this manuscript available at <http://code.google.com/p/mimicree/wiki/PublicationsMaterial> (last accessed November 20, 2013).

Supplementary Material

Supplementary files S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

C.S. and R.K. conceived the study. R.K. wrote the software and conducted the analysis. R.K. and C.S. wrote the paper. The authors thank all members of the Institute of Population Genetics for feedback and support. They especially thank Liang Leng, Andrea Betancourt, Susanne Franssen, and Raymond Tobler for discussions and critical comments. This work was supported by the ERC grant “Archadapt” and Austrian Science Funds (FWF) grant P22725.

References

- Bastide H, Betancourt A, Nolte V, Tobler R, Stöbe P, Futschik A, Schlötterer C. 2013. A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genet.* 9: e1003534.
- Burke MK. 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc Biol Sci.* 279: 5029–5038.
- 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* 467:587–590.
- Burke MK, Long AD. 2012. What paths do advantageous alleles take during short-term evolutionary change? *Mol Ecol.* 21:4913–4916.
- Chevin LM, Hospital F. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* 180: 1645–1660.
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* 36:e105.
- Excoffier L, Foll M. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1332–1334.
- Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA. 2010. *Drosophila melanogaster* recombination rate calculator. *Gene* 463:18–20.
- Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186:207–218.
- Garud NR, Messer PW, Buzbas EO, Petrov DA. 2013. Soft selective sweeps are the primary mode of recent adaptation in *Drosophila melanogaster*. [cited 2013 Nov 20]. Available from: <http://arxiv.org/abs/1303.0906>
- Gillespie J. 2010. Population genetics: a concise guide. Baltimore (NC): Johns Hopkins University Press.
- Haldane JBS. 1927. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proceedings of the Cambridge Philosophical Society*, Vol. 23. p. 838–844.
- Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics. New York: Springer-Verlag.
- Hermisson J, Pennings PS. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.
- Huang W, Richards S, Carbone MA, et al. (25 co-authors). 2012. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A.* 109:15553–15559.
- 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.* 29:1187–1197.
- Imhof M, Schlötterer C. 2001. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc Natl Acad Sci U S A.* 98: 1113–1117.
- Kassen R, Bataillon T. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat Genet.* 38:484–8.
- Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012. Experimental evolution. *Trends Ecol Evol.* 27:547–560.
- Kimura M. 1955. Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci U S A.* 41: 144–150.
- Kimura M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- Kofler R, Orozco-terWengel P, DeMaio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011. Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* 6:e15925.
- Kofler R, Pandey RV, Schlötterer C. 2011. Popoolation2: identifying differentiation between populations using sequencing of pooled dna samples (pool-seq). *Bioinformatics* 27:3435–6.
- Landis JR, Heyman ER, Koch GG. 1978. Average partial association in three-way contingency tables: a review and discussion of alternative tests. *Int Stat Rev.* 46:237–254.
- McClish DK. 1989. Analyzing a portion of the ROC curve. *Med Decis Making.* 9:190–195.
- Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* 12:R112.
- 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.* 21:4931–4941.
- Parts L, Cubillos FA, Warringer J, et al. (14 co-authors). 2011. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21:1131–8.
- Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* 59:2312–2323.
- R Core Team. 2012. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Remolina SC, Chang PL, Leips J, Nuzhdin SV, Hughes KA. 2012. Genomic basis of aging and life-history evolution in *Drosophila melanogaster*. *Evolution* 66:3390–3403.
- Rose MR. 1984. Artificial selection on a fitness-component in *Drosophila melanogaster*. *Evolution* 38:516–526.
- Rose MR, Mueller LD, Burke MK. 2011. New experiments for an undivided genetics. *Genetics* 188:1–10.
- Sing T, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* 21:3940–3941.
- Thompson ML, Zucchini W. 1989. On the statistical analysis of ROC curves. *Stat Med.* 8:1277–1290.
- Tobler R, Franssen SU, Kofler R, Orozco-terWengel P, Nolte V, Hermisson J, Schlötterer C. Forthcoming 2013. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol Biol Evol.* Advance Access published October 22, 2013, doi: 10.1093/molbev/mst205.
- Turner TL, Miller PM. 2012. Investigating natural variation in *Drosophila* courtship song by the evolve & resequence approach. *Genetics* 191: 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:e1001336.