Fine Mapping without Phenotyping: Identification of Selection Targets in Secondary Evolve and Resequence Experiments

Anna Maria Langmüller[1,2], Marlies Dolezal[3], and Christian Schlötterer[1,*]

[1]Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria

[2]Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria

[3]Plattform Bioinformatik und Biostatistik, Vetmeduni Vienna, Vienna, Austria

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

Accepted: 24 June 2021

Abstract

Evolve and Resequence (E&R) studies investigate the genomic selection response of populations in an Experimental Evolution setup. Despite the popularity of E&R, empirical studies in sexually reproducing organisms typically suffer from an excess of candidate loci due to linkage disequilibrium, and single gene or SNP resolution is the exception rather than the rule. Recently, so-called "secondary E&R" has been suggested as promising experimental follow-up procedure to confirm putatively selected regions from a primary E&R study. Secondary E&R provides also the opportunity to increase mapping resolution by allowing for additional recombination events, which separate the selection target from neutral hitchhikers. Here, we use computer simulations to assess the effect of different crossing schemes, population size, experimental duration, and number of replicates on the power and resolution of secondary E&R. We find that the crossing scheme and population size are crucial factors determining power and resolution of secondary E&R: A simple crossing scheme with few founder lines consistently outcompetes crossing schemes where evolved populations from a primary E&R experiment are mixed with a complex ancestral founder population. Regardless of the experimental design tested, a population size of at least 4,800 individuals, which is roughly five times larger than population sizes in typical E&R studies, is required to achieve a power of at least 75%. Our study provides an important step toward improved experimental designs aiming to characterize causative SNPs in Experimental Evolution studies.

Key words: experimental evolution, secondary evolve and resequence, experimental design, *Drosophila*, fine mapping.

Significance

Despite the popularity of Evolve and Resequence (E&R) to investigate genomic selection responses, most studies that use sexually reproducing organisms have broad selection signatures and an excess of candidate loci due to linkage disequilibrium. In this study, we use computer simulations and statistical modeling to evaluate the effects of different experimental and population genetic parameters on the success of potential follow-up experiments (=secondary E&R) aiming to validate and fine-map selection signatures of primary studies. We found that a large population size in combination with a simple crossing scheme is key to the success of secondary E&R in *Drosophila*.

Introduction

Deciphering the genetic architecture of adaptation is one of the longstanding goals in evolutionary biology. Experimental Evolution (EE) has become a popular approach to study adaptation in real time (Garland and Rose 2009; Kawecki et al. 2012). In contrast to natural populations, EE offers the key advantage of replicating experiments under controlled laboratory conditions (Schlötterer et al. 2015). Evolve and

Resequence (E&R) (Turner et al. 2011; Long et al. 2015; Schlötterer et al. 2015)—a combination of EE with Next Generation Sequencing—facilitates in-depth analysis of the genomic responses to selection, with the ultimate goal to identify and characterize individual adaptive loci.

E&R has already been successful in investigating genomic selection responses from standing genetic variation in adapting sexually reproducing organisms, such as chicken (Johansson et al. 2010), yeast (Burke et al. 2014), and *Drosophila* (Teotónio et al. 2009; Remolina et al. 2012; Martins et al. 2014; Barghi et al. 2019). Despite its popularity, E&R typically suffers from an excess of candidates caused by linkage disequilibrium between true causative SNPs and neutral hitchhikers (Nuzhdin and Turner 2013; Tobler et al. 2014; Franssen et al. 2015), which decreases the resolution of E&R studies and makes single gene resolution (Martins et al. 2014) the exception rather than the rule.

The problem of candidate excess in E&R studies has been approached from different angles. More refined statistical tests have been developed (Topa et al. 2015; Iranmehr et al. 2017; Kelly and Hughes 2019; Spitzer et al. 2020), and the combination of time-series data with replicate populations has been identified as particularly powerful (Lang et al. 2013; Burke et al. 2014; Barghi et al. 2020). Organisms with a higher recombination rate and a lack of large segregating inversions that suppress recombination events have been suggested to be better suited for E&R studies (Barghi et al. 2017). Computer simulations showed that the power of E&R studies can be significantly improved by increasing the number of replicate populations, the experimental duration, or by adjusting the applied selection regime (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014; Kessner and Novembre 2015; Vlachos and Kofler 2019).

Burny et al. (2020) recently suggested an experimental follow-up procedure (secondary E&R) to validate selection signals of primary E&R studies. The basic idea of secondary E&R is that putative selection targets determined in the primary E&R study should rise in frequency again when exposed to the same environmental conditions during an additional E&R conducted after the primary experiment (fig. 1*A*). This experimental validation of selection signals is especially attractive before starting the time-consuming functional characterization of putatively selected alleles (e.g., based on the CRISPR/Cas technology) (Gratz et al. 2013). A hitherto underexplored potential of secondary E&R is that the additional recombination events during the secondary E&R can be used to fine map selection signals of primary experiments. In addition to mixing evolved genotypes of a primary E&R with nonadapted ancestral founder genotypes (coined "dilution" by Burny et al. [2020]), we propose several different secondary E&R crossing schemes for validating and fine-mapping of putative selection targets.

We evaluate the power and resolution of different secondary E&R designs to identify causative SNPs via extensive computer simulations. We use logistic regression to assess which simulated experimental and population genetic parameters have a significant effect on the success of secondary E&R. Selection coefficient, dominance coefficient, and mean starting allele frequency of the selection target all have a significant effect on the success of secondary E&R. However, crossing scheme and population size emerge as the most influential parameters. We show that the population size of secondary E&R experiments needs to be at least five times larger than currently used population sizes in typical primary E&R studies with *Drosophila* to achieve a power above 75%. Furthermore, we show that the crossing scheme is a crucial experimental parameter shaping the power of secondary E&R—a simple crossing scheme with few founder lines results in higher power and resolution compared with more complex crossing schemes.

Materials and Methods

Outline of the Simulation Framework

The nonadapted ancestral founder genotypes used in our simulation study are a randomly chosen subset of 100 haplotypes from a panel of 189 sequenced *D. simulans* haplotypes originally collected in Tallahassee (Florida, USA) capturing the amount of standing genetic variation in a natural *Drosophila* population (Barghi et al. 2019; Howie et al. 2019). We use the term "founder line" for an inbred isofemale line homozygous for one of these ancestral haplotypes. In order to speed up the calculations, we only simulated chromosome-arm 2L. Simulations with linkage were conducted with MimicrEE2 (v206) (Vlachos and Kofler 2018) using the *D. simulans* recombination map (Howie et al. 2019). MimicrEE2 is a forward-simulation framework for E&R studies that can simulate evolving experimental populations based on their haplotype information and genome-wide recombination rates. We used the w-mode of MimicrEE2 which computes the fitness of individuals directly from the selection coefficients. If not stated otherwise, we simulated positive, additive selection for biallelic SNPs, with selection coefficients being uniformly sampled between 0.07 and 0.1 for each positively selected SNP and tracked the frequency of all SNPs over time (ranging from 479,507 to 970,466 SNPs, supplementary table S1, Supplementary Material online). We chose to simulate rather strong selection reasoning that alleles with a high selection coefficient are more likely to be experimentally tested. On the other hand, strongly selected alleles result in many neutral linked hitchhikers producing false positive signals that adversely impact the mapping resolution (Kofler and Schlötterer 2014)—which requires follow-up studies to identify the target of selection. If not stated otherwise, selected SNPs were randomly chosen with an equal probability to be

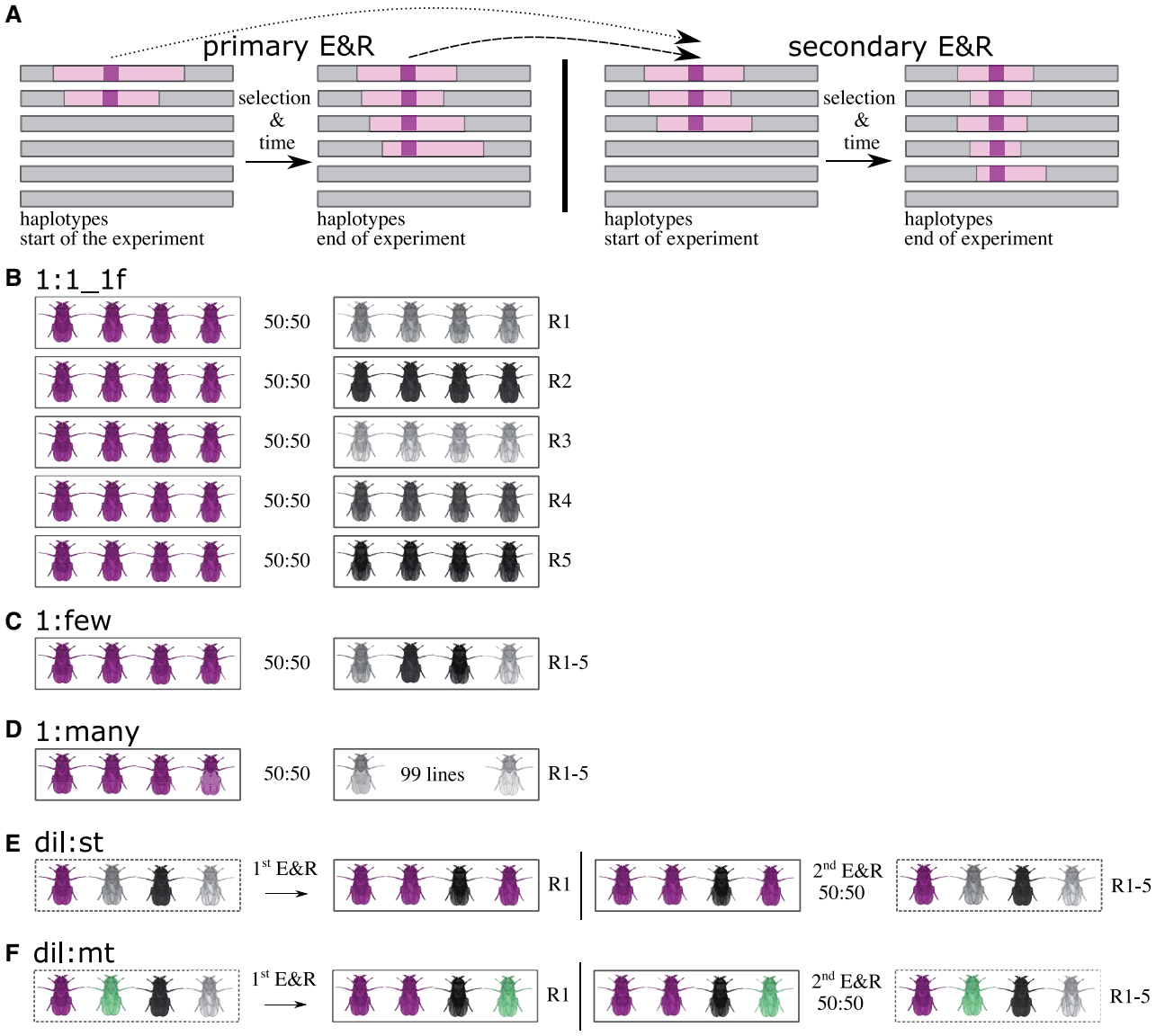

Fig. 1.—Basic idea of secondary Evolve and Resequence (E&R) and simulated crossing schemes. (*A*) The basic idea of secondary E&R is that a putative selection target (purple) determined in a primary E&R study (left) should rise in frequency again when exposed to the same environment (i.e., selection regime), during secondary E&R (right). Additional recombination events during the secondary E&R allow to fine map selected regions of primary E&R experiments (i.e., reduce the number hitchhikers indicated in light pink). The two arrows indicate that the secondary E&R is either started with evolved haplotypes (dashed arrow) or with specific founder lines of the primary E&R (dotted arrow). (*B*) 1:1_1f crossing scheme: Inbred flies with one target of selection (purple) are crossed to inbred flies without known beneficial variants. The starting frequency of each genotype is 50%. In each replicate the line with the beneficial allele (focal line, purple) is crossed to a different line lacking beneficial mutations (nonfocal lines, different shades of gray). (*C*) 1:few crossing scheme: The focal line is crossed to a pool of flies with five different genotypes without known selection targets. The starting frequency of the focal line is again 50%. (*D*) 1:many crossing scheme: The focal line is crossed to a pool of 99 lines without known selection targets. (*E*) dil:st crossing scheme: 50% of an evolved population originating from a primary E&R is replaced by ancestral genotypes of the primary E&R. The entire population has only one single target of selection (focal SNP, purple flies). (*F*) dil:mt crossing scheme: 50% of an evolved population originating from a primary E&R is replaced by ancestral genotypes of the primary E&R. The ancestral population carries 16 targets of selection (flies carrying different beneficial SNPs are shown in purple, and green).

either codominant (dominance coefficient $h = 0.5$), or dominant ($h = 1$). We did not consider recessive loci ($h = 0$), because we do not anticipate that fully recessive targets would result in sufficiently large allele frequency changes to be detected in primary E&R experiments (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014). Similar to Baldwin-Brown et al. (2014), we did not model allele frequency estimation errors caused for example by sequencing errors, limited read

Table 1

Simulation Overview. 1:1_1f, 1:few, 1:many, dil:st, dil:mt: For each selected SNP, selection coefficients were uniformly sampled between 0.07 and 0.1. Dominance coefficients were randomly chosen to be either 0.5 (co-dominant) or 1 (dominant). 1:1_2f, 1:1_1f1nf: For each simulation, the selection coefficient of the target of interest was uniformly sampled between 0.07 and 0.1. In contrast to 1:1_1f, we simulated an additional selection target, which was either located on the focal haplotype (1:1_2f) or one non-focal haplotype (1:1_1f1nf). The selection coefficient of the additional selection target was uniformly sampled between 0.07 and the selection coefficient of the target of interest. We simulated all possible combinations of dominance coefficients for the two selected SNPs (500 simulations per combination).

Crossing Scheme	Population Size	Generations	Replicates	Number of Simulations
1:1_1f	300; 1,200; 4,800; 19,200	60; 20	5/30	2,000/100
1:1_2f	300; 1,200; 4,800; 19,200	60	5	2,000
1:1_1f1nf	300; 1,200; 4,800; 19,200	60	5	2,000
1:few	300	60	5/30	2,000/100
1:many	300	60; 20	5/30	2,000/100
dil:st	300	60	5/30	2,000/100
dil:mt	300; 1,200; 4,800; 19,200	60; 20	5/30	2,000/100

depth, read depth heterogeneity across the chromosome, or the number of sequenced individuals. We used PoPoolation2 (Kofler et al. 2011) to rescale allele counts for each biallelic position to a uniform read depth of 80.

Experimental Parameters

The purpose of this study is to test the influence of different experimental parameters on the power and resolution of secondary E&R. For this, we systematically varied the crossing scheme, population size, experimental duration, and number of replicates to assess the effect of these experimental parameters on the power and resolution of secondary E&R (table 1).

We use the term "experimental design" to describe a distinct set of simulated experimental parameters (e.g., crossing scheme = 1:1_1f; population size = 1,200 individuals; experimental duration = 60 generations; five replicates).

Crossing Scheme

We simulated five different crossing schemes: 1:1 (versions 1f, 2f, and 1f1nf), 1:few, 1:many, dil:st, and dil:mt (fig. 1*B–F*). In the 1:1_1f crossing scheme (fig. 1*B*), two inbred founder lines are crossed at equal proportions. One inbred focal line carries a single target of selection—known from a primary E&R experiment—and is crossed with inbred nonfocal lines, not carrying known adaptive alleles. We consider this a best-case scenario. Replicates are created by crossing the same focal line to different nonfocal lines without known targets of selection. We used different nonfocal lines as crossing partners to account for the possibility that in real experiments these lines may contain unidentified selected loci. By using different lines, the influence of selection targets present in a single nonfocal line will be outweighed by the focal locus, which is present in every replicate (fig. 1*B*). To further explore the influence of unidentified selected loci, we simulated two more versions of the 1:1 crossing scheme with a more realistic genetic architecture (supplementary fig. S1, Supplementary Material online). In these versions, either the focal line itself, or one of the nonfocal lines carries one additional target of selection. We call these scenarios 2f, for a total of two selected loci in the focal line (supplementary fig. S1*B*, Supplementary Material online), and 1f1nf, for one selected locus in the focal line and one selected locus in one of the nonfocal lines (supplementary fig. S1*C*, Supplementary Material online). For each simulation, we sampled the selection coefficient of the additional selection target uniformly between 0.07 and the selection coefficient of the selected SNP we consider for our analysis. We simulated all possible combinations of dominance coefficients for the two selected SNPs (0.5–0.5; 0.5–1; 1–0.5; 1–1). We ran 500 simulations for each of the four combinations of dominance coefficients for 2f and 1f1nf, respectively.

In the 1:few crossing scheme (fig. 1*C*) the focal line is crossed with a pool of five nonfocal lines that do not carry known beneficial alleles. The starting frequency of the focal line is 50%, whereas each nonfocal line has a starting frequency of 10%. Each replicate consists of the same focal/nonfocal line mixture. In the 1:many crossing scheme (fig. 1*D*), the focal line is crossed with a pool of 99 nonfocal lines. Replicates consist of the same mixture of lines, and the starting frequency of the focal line is 50%.

It has been recently suggested, that "diluting" evolved populations of a primary E&R experiment with many nonadapted ancestral genotypes of the very same primary E&R and exposing the diluted populations to the same selection regime is a promising approach to validate selection candidates (Burny et al. 2020). However, we lack a systematic power assessment of such experiments with computer simulations. We thus included two "dilution" crossing schemes (dil:mt, dil:st) into our analysis. To evaluate the power of dilution crossing schemes, it is important to first simulate a primary E&R study. We chose to simulate a population consisting of 100 different founder lines, a population size of 300 individuals, 60 generations of adaptation and one replicate, which can then be diluted with nonadapted ancestral genotypes. SNPs in the primary E&R were tested for allele frequency change with the χ^2 test.

The population of the dil:st crossing scheme carries only one beneficial SNP (dil:st = dilution:single target, fig. 1*E*). The starting allele frequency of this beneficial, focal SNP is sampled from the empirical starting allele frequency distribution of putatively selected alleles from a previous E&R study in which *D. simulans* populations adapted to a new temperature regime (mean starting frequency = 0.1) (Barghi et al. 2019). After simulating 60 generations of adaptation (primary E&R), 50% of the evolved population is replaced by flies of the nonadapted ancestral founder population. After this dilution step, the secondary E&R was simulated under the exact same selection regime as in the primary E&R. A region with a strong selection signal from the primary E&R is chosen for validation in the dilution crossing schemes (Burny et al. 2020). Hence, we investigated a 1 Mb window, which is the previously reported median selected haplotype block length on chromosome-arm 2L (Barghi et al. 2019), around the SNP with the highest χ^2 test statistic in the primary E&R, in the secondary E&R.

The dil:mt (fig. 1*F*) crossing scheme has multiple selection targets (dil:mt = dilution:multiple targets; 16 selection targets on chromosome-arm 2L; Barghi et al. 2019). Again, we investigate a 1 Mb window around the SNP with the highest χ^2 test statistic in the primary E&R, in the secondary E&R. In case of multiple selection targets in the 1 Mb window, we consider the target that is closest to the SNP with the highest CMH test statistic in the 1 Mb window of the secondary E&R in our analysis.

Population Size

For two crossing schemes that are relatively easy to implement in empirical studies 1:1 (1f, 2f, and 1f1nf) and dil:mt, we performed simulations (experimental duration = 60 generations; five replicates) with independently sampled selection targets for population sizes of 300; 1,200; 4,800; and 19,200 individuals per replicate to test for the effect of the population size on the power and resolution of secondary E&R (table 1). All other crossing schemes were evaluated at a population size of 300 individuals.

Experimental Duration

All crossing schemes were evaluated after 60 generations. To explore the possibility that shorter experiments with less than 60 generations may already be sufficient to achieve satisfactory power, we analyzed the simulations for 1:1 (without 2f and 1f1nf) and dil:mt crossing schemes also already after 20 generations (table 1).

Number of Replicates

To assess the impact of the number of replicates on power of secondary E&R, we conducted for each crossing scheme (1:1 without 2f and 1f1nf) 100 additional simulations (population size = 300 individuals; experimental duration = 60 generations) with 30 replicates (table 1).

Statistical Analysis

All statistical analyses were performed using the R statistical computing environment (v3.5.3) (R Core Team 3.5.3 2019).

Power

For each simulation, we tested all SNPs on chromosome-arm 2L for allele frequency increase between the start and the end of the simulated secondary E&R using the Cochran–Mantel–Haenszel test (CMH test, implemented in the R package poolSeq [v0.3.2]; Taus et al. 2017). The CMH test allows to test for independence of matched data—for example, allele counts of replicated ancestral and evolved populations (Agresti and Kateri 2011). We ranked SNPs based on their CMH test statistic using the dense ranking method in the R package data.table (v.1.12-8) (Dowle and Srinivasan 2019). In dense ranking, SNPs with identical test statistics receive the same rank, and the following SNP is assigned the immediately following rank.

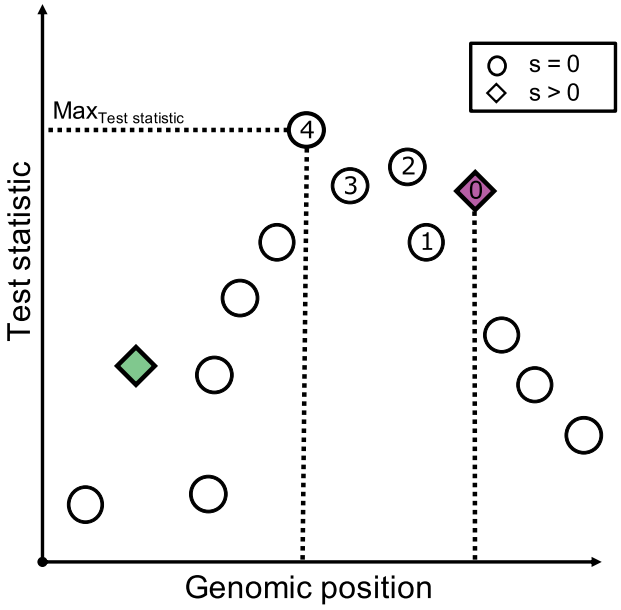

Fig. 2.—Schematic overview of the definition of success-B in a secondary Evolve and Resequence simulation. All SNPs (neutral=circle, beneficial=diamond) are tested for an allele frequency change with the Cochran–Mantel–Haenszel test, and are ranked based on their test statistic (*y* axis). If the focal selection target (purple diamond; one additional selected SNP is shown as green diamond—for crossing scheme 1:1_2f, 1:1_1fnf, and dil:mt) is less than 100 SNPs away from the SNP with the highest test statistic, the simulation run is deemed a success. In the example depicted, the distance in number of SNPs between the focal target of selection, and the SNP with the highest test statistic is 4, and the simulation is classified as success.

Based on this ranking, we used two different approaches to classify simulations being either successful, or unsuccessful. First, we only considered a simulation to be successful if the true target of selection (i.e., the focal SNP with a selection coefficient >0) was the SNP with the highest test statistic (success-A). In a second analysis step, we classified a simulation as success, if the focal target of selection was not more than 100 SNPs away from the SNP with the highest test statistic (success-B, fig. 2). We acknowledge that success-B depends on the maximum distance allowed between the true target of selection and the SNP with the highest CMH test statistic (fig. 2). However, varying the maximum distance threshold did not alter the relative performance of different experimental designs (data not shown). The power of an experimental design is defined as the proportion of simulations that were able to detect the true target of selection.

Resolution

For simulations where the selection target was detected (success-B), we determined the resolution of fine mapping of the selection target by counting the number of SNPs between the true selection target, and the SNP with the highest CMH test statistic.

Assessment of Experimental and Population Genetic Parameters

We used logistic regressions with a binomial error structure (ε) and a logit link function (Baayen 2008) to test how different experimental and population genetic parameters affect success and failure to identify targets of selection, that is, success (Y) is treated as a binary response encoded in 0 (failure) and 1 (success) of secondary E&R.

We fitted three different models with R function *glm* with μ being the overall mean per model: Model 1 includes only the two crossing schemes 1:1_1f and dil:mt for which we also varied population size. Model 2 includes the three versions of crossing scheme 1:1 (1f, 2f, and 1f1nf) at varying population sizes: 1f with only one positively selected SNP in the focal line; version 2f with two selected SNPs in the focal line; and version 1f1nf with one selected SNP in the focal line and one selected SNP in one nonfocal line. Model 3 includes all five simulated crossing schemes (1:1 without 2f and 1f1nf) at a constant population size of 300 individuals.

Prior to model fitting, the two covariates selection coefficient and mean starting allele frequency (not applicable to model 2 as starting allele frequency is always 50%) were multiplied by 100, and z-transformed to a mean of zero and standard deviation of one for easier interpretable estimates (Schielzeth 2010).

$$\begin{aligned} Y_{ijklmn} = {} & \mu + \text{cross}_i + h_j + s_k + af_l + N_m \\ & + (\text{cross} : h)_{ij} + (h : af)_{jl} + (\text{cross} : N)_{im} \\ & + (h : N)_{jm} + (\text{cross} : h : N)_{ijm} + \varepsilon_{ijklmn} \end{aligned} \quad (1)$$

$$\begin{aligned} Y_{ijkmn} = {} & \mu + \text{architecture}_i + h_j + s_k + N_m \\ & + (\text{architecture} : h)_{ij} + (\text{architecture} : N)_{im} \\ & + (h : N)_{jm} + (\text{architecture} : h : N)_{ijm} + \varepsilon_{ijkmn} \end{aligned} \quad (2)$$

$$\begin{aligned} Y_{ijklm} = {} & \mu + \text{cross}_i + h_j + s_k + af_l + (\text{cross} : h)_{ij} + (h : af)_{jl} \\ & + \varepsilon_{ijklm} \end{aligned} \quad (3)$$

Model 1 (eq. 1) contained: 1) five explanatory variables as main effects; crossing scheme (cross_i), a fixed categorical effect with two levels—1:1_1f and dil:mt; a fixed categorical effect of dominance coefficient (h_j) with levels 0.5 and 1; selection coefficient (s_k) and mean starting allele frequency (af_l) over replicated populations (both as continuous covariate); and population size (N_m), a fixed categorical effect with four levels—300; 1,200; 4,800; 19,200, 2) an interaction term between dominance coefficient and mean starting allele frequency ($(h : af)_{jl}$), 3) and a triple interaction between crossing scheme, dominance coefficient, and population size ($(\text{cross} : h : N)_{ijm}$), and all pairwise interaction terms of effects involved in the triple interaction term ($(\text{cross} : h)_{ij}$; $(\text{cross} : N)_{im}$; $(h : N)_{jm}$). We included interaction terms into the model that have population genetic interpretations.

Data analyzed with model 1 contained 16,000 observations, namely 2,000 independent simulation runs for each crossing scheme (1:1_1f, dil:mt), and each of the four different population sizes (300; 1,200; 4,800; 19,200) (table 1). To avoid potential bias introduced by specific haplotypes being sampled, we randomly chose four sets of focal/nonfocal founder lines for the 1:1_1f crossing scheme and performed 500 simulation runs per set. The chosen set of founder lines did not have a significant effect on the success of secondary E&R and is thus not included in the final model (likelihood ratio test [LRT] full-reduced model comparison; data not shown).

Model 2 (eq. 2) contained: 1) four explanatory variables as main effects; a fixed categorical effect "architecture" ($\textit{architecture}_i$) that describes the version of the 1:1 crossing scheme in combination with the dominance coefficient of the additional selected target (if present) resulting in five levels: 1f; 2f_h05; 2f_h1; 1f1nf_h05; 1f1nf_h. Model 2 further contained a fixed categorical effect of dominance coefficient for the focal SNP (h_j) with levels 0.5 and 1; selection coefficient (s_k) as continuous covariate; and population size (N_m), a fixed categorical effect with four levels: 300; 1,200; 4,800; 19,200, 2) a triple interaction between architecture, dominance coefficient, and population size ($(\text{architecture} : h : N)_{ijm}$), and all pairwise interaction terms of effects involved in the triple interaction term ($(\text{architecture} : h)_{ij}$; $(\text{architecture} : N)_{im}$; $(h : N)_{jm}$).

Data analyzed with model 2 contained 24,000 observations, namely 2,000 independent simulation runs for each version of the 1:1 crossing scheme (1f, 2f, 1f1nf), and four different population sizes (300; 1,200; 4,800; 19,200

individuals). Because model 1 showed that the chosen set of founder lines did not have a significant effect on the power of secondary E&R, we simulated only one set of focal/nonfocal founder lines.

Model 3 (eq. 3) contained: 1) four explanatory variables; crossing scheme ($cross_i$), a fixed categorical effect with five levels: 1:1_1f; 1:few; 1:many; dil:st; dil:mt, dominance coefficient (h_j), selection coefficient (s_k), and mean starting allele frequency (af_l) over replicated populations as main effects, as described for model 1, 2) an interaction term between crossing scheme and dominance coefficient ($(cross : h)_{ij}$), 3) and an interaction term between dominance coefficient and mean starting allele frequency ($(h : af)_{jl}$).

We analyzed 10,000 samples (2,000 independent simulation runs for each crossing scheme) (table 1). For crossing schemes using only few different founder lines (1:1_1f, 1:few), the simulation runs are based on four randomly chosen sets of founder lines each (500 simulation runs per set). As in model 1, the set of founder lines does not have a significant effect on secondary E&R success and is thus not included in the final model (LRT full-reduced model comparison; data not shown). We performed additional analysis with 500 observations (100 independent simulation runs for each crossing scheme) to determine the influence of the number of replicates on the power of secondary E&R (table 1).

We performed all diagnostic checks required for logistic regression. Absence of collinearity was confirmed by computing the generalized Variance Inflation Factors (Fox and Monette 1992) using function *vif* in R package *car* (v3.0-8; Fox and Weisberg 2019). Model stability was checked with the R function *dfbeta*. For visualization, linear predictors (LP) were back-transformed to success probabilities using the inverse logit transformation: $p_{sucess} = \frac{e^{LP}}{1+e^{LP}}$. 95% confidence intervals of the fitted values were investigated with the R function *predict.glm*. Significance of single explanatory variables was tested with a Type II ANOVA using function *Anova* in R package car (Fox and Weisberg 2019) and are provided in the Supplement. Significance of explanatory variables including all their modeled interactions was tested with a LRT comparing the full model with a nested reduced model with the same structure as the full model, but lacking the assessed explanatory variable (and its interactions). Significance is declared at an alpha cut-off of 5%. We used Nagelkerke's R^2 index (Nagelkerke 1991) to calculate the improvement of each model parameter upon the prediction of a reduced model.

We observed two cases where a combination of explanatory variables resulted in complete separation of data points (fig. 5*A* [dominance coefficient: 0.5, population size: 1,200, architecture: 2f_h05], supplementary fig. S8*A*, Supplementary Material online [dominance coefficient: 0.5, crossing scheme: dil:mt]). To obtain interpretable model estimates, we added one pseudo-observation with the missing response (success) to the data for each of these two cases.

Results

We used forward simulations to assess the influence of different experimental and population genetic parameters, more specifically crossing scheme, population size, dominance coefficient, selection strength, and mean starting allele frequency on the success to detect and fine map selection targets in secondary E&R experiments (fig. 1*A*). We used a CMH test to identify SNPs rising in allele frequency (number of SNPs see supplementary table S1, Supplementary Material online). The CMH test allows to test for independence of matched categorical data (Agresti and Kateri 2011), and compares favorably to other statistical methods in reliably identifying possible targets of selection in E&R setups (Vlachos et al. 2019). A simulation run was considered successful, if the true target of selection was the SNP with the highest CMH test statistic (success-A). In a second analysis step, we considered simulation runs as successful, if the true target of selection was not more than 100 SNPs away from the SNP with the highest CMH test statistic (success-B, fig. 2). We used logistic regression to assess which experimental and population genetic parameters have a significant effect on the success of secondary E&R.

First, we evaluated two crossing schemes that can be easily implemented in empirical studies, 1:1_1f (fig. 1*B*) and dil:mt (fig. 1*F*). The 1:1_1f crossing scheme (fig. 1*B*) is based on founder lines only (a founder line is an inbred isofemale line homozygous for one ancestral haplotype). Using information about the selected haplotype from a primary E&R study, it is possible to determine which founder lines carry a selection target (Barghi et al. 2019). Note, this requires focal founder lines to be sequenced. Crossing a focal founder line with the selected haplotype with a nonfocal line without known selection targets offers the advantage of reducing the number of selection targets dramatically. Using different nonfocal lines without known strong selection targets in each replicate reduces the potential of consistent confounding effects of unidentified selection targets outside the region of interest—a signal the CMH test is particularly sensitive to as it scans for consistent allele frequency changes across replicates.

Dil:mt (fig. 1*F*) represents an entirely different approach. Dil:mt has multiple targets of selection at different frequencies and is probably the simulated crossing scheme with the most straight forward empirical implementation (Barghi et al. 2019; Burny et al. 2020). It is based on a "dilution" approach, where evolved individuals from a primary E&R experiment are crossed to the nonadapted ancestral founder population from the same primary E&R (Burny et al. 2020). In contrast to 1:1_1f, a dil:mt crossing scheme requires both—ancestral and evolved—populations of the primary E&R, but relatively limited information about selection targets on individual founder lines.

To evaluate the 1:1_1f and dil:mt crossing scheme, we simulated secondary E&R consisting of five replicated

populations with constant population size (300; 1,200; 4,800, or 19,200 individuals per replicate) that evolve for 60 generations. We simulated positively selected SNPs (selection coefficient is uniformly sampled between 0.07 and 0.1) that were randomly chosen with equal probability to be either codominant ($h = 0.5$) or dominant ($h = 1$). We used logistic regression to assess the effects of model parameters on secondary E&R success (Model 1, Assessment of Experimental and Population Genetic Parameters in the Materials and Methods section). Although selection strength (LRT full-reduced null model comparison [Model 1]: $\chi^2 = 67.4$, df = 1, $P < 0.001$ [success-A]; $\chi^2 = 73.4$, df = 1, $P < 0.001$ [success-B]), dominance coefficient (LRT full-reduced model comparison [Model 1]: $\chi^2 = 733.4$, df = 9, $P < 0.001$ [success-A]; $\chi^2 = 714.5$, df = 9, $P < 0.001$ [success-B]), and mean starting allele frequency (LRT full-reduced model comparison [Model 1]: $\chi^2 = 22.4$, df = 2, $P < 0.001$ [success-A]; $\chi^2 = 17.4$, df = 2, $P < 0.001$ [success-B]) all have a significant effect on the power of secondary E&R (supplementary table S2, Supplementary Material online), our analysis reveals that crossing scheme and population size have by far the strongest influence on secondary E&R success. Both the crossing scheme (LRT full-reduced model comparison [Model 1]: $\chi^2 = 3,510.1$, df = 8, $P < 0.001$ [success-A]; $\chi^2 = 3,211.3$, df = 8, $P < 0.001$ [success-B]); and the population size (LRT full-reduced model comparison [Model 1]: $\chi^2 = 2,142.9$, df = 12, $P < 0.001$ [success-A]; $\chi^2 = 2,633.4$, df = 12, $P < 0.001$ [success-B]) have a significant effect on the success of secondary E&R, and are the only parameters with a Nagelkerke's R^2 index above 0.2 (supplementary table S3, Supplementary Material online).

Although a 1:1_1f crossing scheme results in higher power values than the dil:mt crossing scheme independently of the population size, the difference to the dil:mt crossing scheme is more pronounced in larger experimental populations (fig. 3). This does not hold only for the power, but also for the resolution (=the distance in SNPs between the true target of selection, and the SNP with the highest CMH test statistic, fig. 4). A potential disadvantage of the 1:1_1f crossing scheme is that the low number of different founder lines ($n = 6$, fig. 1*B*) causes more linkage disequilibrium, indicated by the number of neighboring SNPs with identical test statistics (=ties) and thus broadens signatures compared with dil:mt crossing scheme that has more founder lines ($n = 100$) (supplementary fig. S2, Supplementary Material online). However, this is outweighed by superior power of the 1:1_1f crossing scheme at every population size investigated (figs. 3 and 4).

Experimental Duration

Our analysis showed that after 60 generations of adaptation, only experimental designs with a population size of at least 1,200 individuals have more power than 50% (fig. 3). The 1:1_1f crossing scheme clearly outperforms dil:mt regardless of the population size, and reaches power above 75% only with a population size of at least 4,800 individuals. However, the maintenance of five replicates for 60 generations is labor intensive and can be quite time consuming for most sexual organisms. We explored the possibility that shorter experiments with less than 60 generations may already be sufficient to achieve satisfactory power values. We thus reanalyzed the power of the 1:1_1f, and dil:mt crossing scheme after 20 generations of adaptation (table 1).

Consistent with previous computer simulation studies (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014; Kessner and Novembre 2015) and empirical results (Langmüller and Schlötterer 2020), we observe reduced power for experimental designs with shorter experimental duration (supplementary fig. S3, Supplementary Material online). Consistent with the results after 60 generations, the crossing scheme (LRT full-reduced model comparison [Model 1]: $\chi^2 = 2,672.7$, df = 8, $P < 0.001$ [success-A]; $\chi^2 = 2,234.4$, df = 8, $P < 0.001$ [success-B]) and the population size (LRT full-reduced model comparison [Model 1]: $\chi^2 = 3,176.4$, df = 12, $P < 0.001$ [success-A]; $\chi^2 = 3,375.3$, df = 12, $P < 0.001$ [success-B]) have the biggest effects on the success of secondary E&R (supplementary tables S2 and S3, Supplementary Material online).

In contrast to our analysis after 60 generations of adaptation, the dil:mt crossing scheme results in higher power than the 1:1_1f crossing scheme for populations with the smallest simulated population size (300 individuals) after 20 generations of adaptation (supplementary fig. S4, Supplementary Material online). A possible explanation for this is that the dil:mt crossing scheme has multiple beneficial targets. Because we simulate additive selection, linked selection targets can act synergistically and increase the frequency of the focal SNP over shorter time scales. This phenomenon of pronounced allele frequency increase due to linked selection will be especially important if the population size is small (i.e., drift is not neglectable), and the experimental duration is short. This is reflected in a higher Nagelkerke's R^2 index for population size in secondary E&R with an experimental duration of 20 generations compared with 60 generations of adaptation (supplementary table S3, Supplementary Material online).

With increasing population size and thus, reduced genetic drift, the 1:1_1f crossing scheme outcompetes dil:mt, as already seen in our analysis after 60 generations (supplementary figs. S3 and S4, Supplementary Material online). Although shorter experimental duration reduces power, our analysis after 20 generations of adaptation highlights that if the maintenance of experimental populations for many generations is not feasible, shorter experiments can achieve similar power if they are maintained at larger population sizes (fig. 3 and supplementary fig. S3, Supplementary Material online).

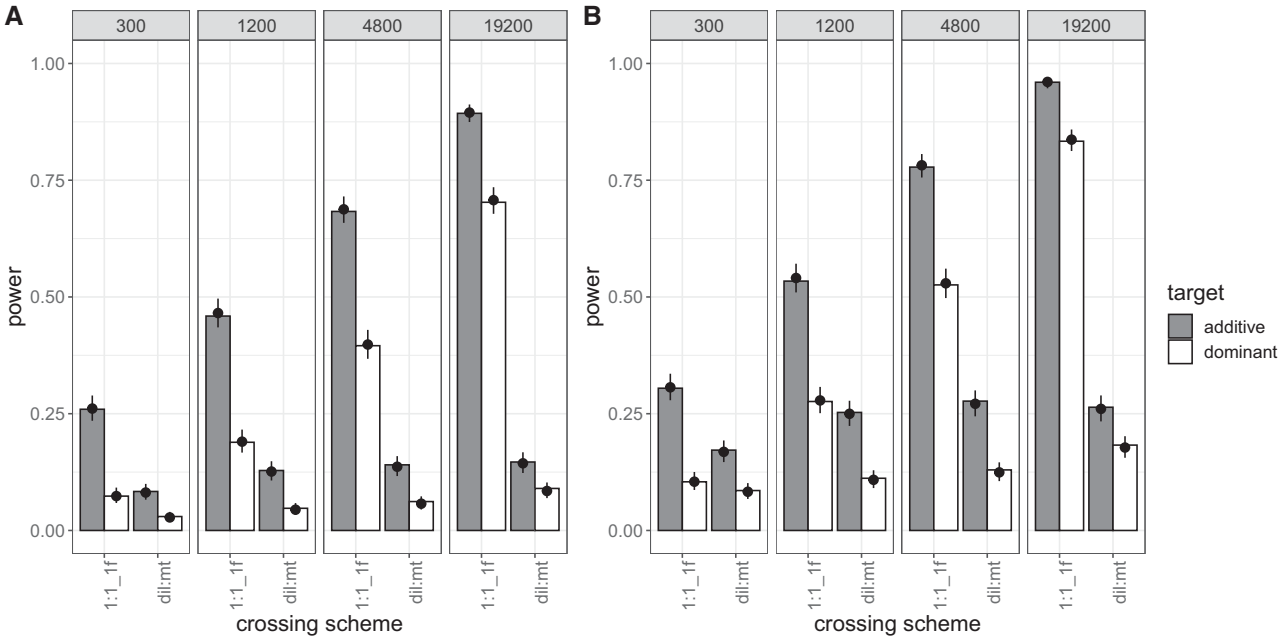

Fig. 3.—Power of the 1:1_1f and dil:mt crossing scheme at different population sizes (2,000 simulations/experimental design). Bars show the power (i.e., the proportion of successful simulations) separately for each combination of crossing scheme (1:1_1f, dil:mt), population size (300; 1,200; 4,800; 19,200 individuals), and dominance coefficient (additive in gray, dominant in white). The dots with error bars display the estimate from the fitted model (Model 1) and its 95% confidence interval. For the model fit, the selection coefficient was fixed to its global average, and combination-specific average starting allele frequencies were used. (*A*) shows the results for success-A (=selection target is the SNP with the highest Cochran–Mantel–Haenszel [CMH] test statistic), (*B*) shows the results for success-B (=selection target is not more than 100 SNPs away from the SNP with the highest CMH test statistic).

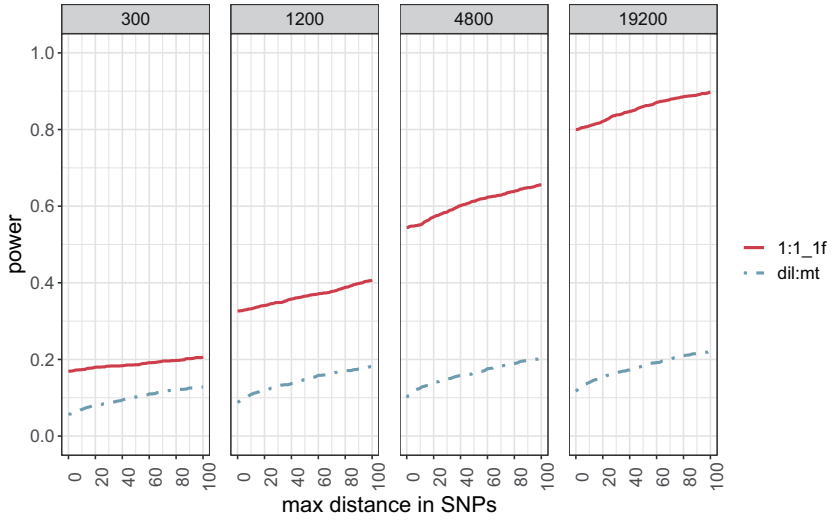

Fig. 4.—Resolution of the 1:1_f and dil:mt crossing scheme at different population sizes. Proportion of simulations (*y* axis) that do not exceed a maximum distance in SNPs (*x* axis) between the SNP with the highest Cochran–Mantel–Haenszel test statistic and the true target of selection. Each panel shows the result for one simulated population size.

Additional Target of Selection in the 1:1 Crossing Scheme

In contrast to dil:mt, the 1:1_1f crossing scheme harbors only one target of selection. We simulated two additional versions of the 1:1 crossing scheme to investigate how one additional target of selection influences the power of secondary E&R (supplementary fig. S1, Supplementary Material online; Model 2, Assessment of Experimental and Population Genetic Parameters in the Materials and Methods section). We observed that one additional beneficial SNP has a significant effect on the power of secondary E&R using the 1:1

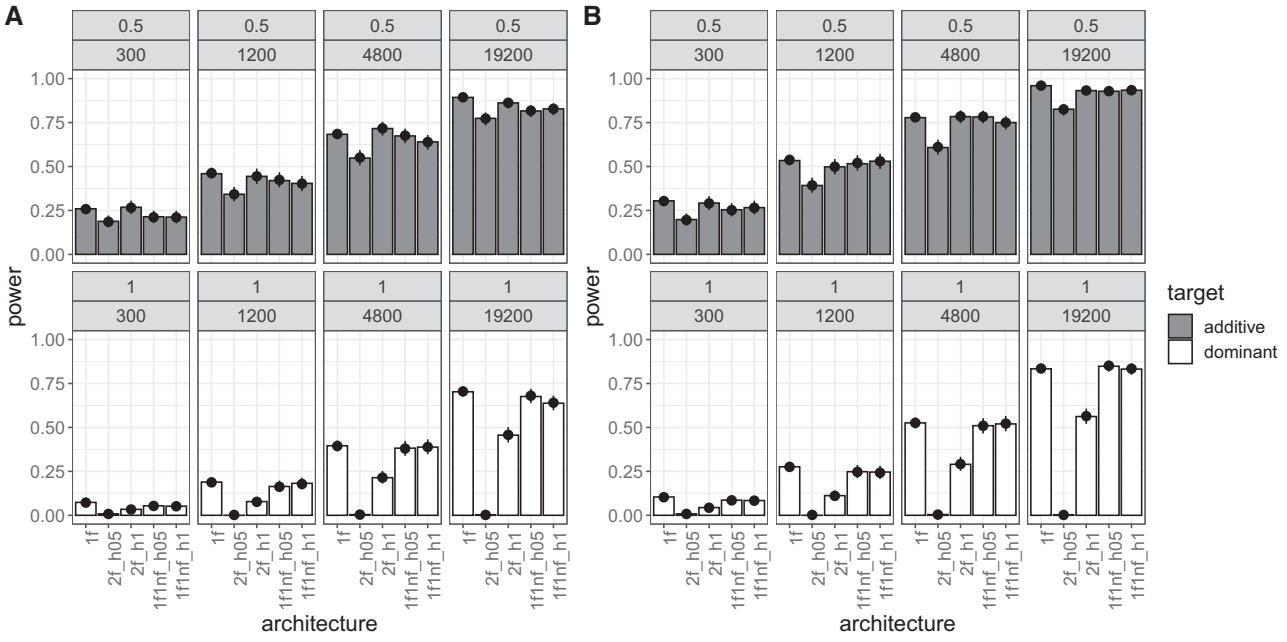

Fig. 5.—Power of the different 1:1 crossing schemes at different population sizes (Model 2) (2,000 simulations/experimental design) after 60 generations of adaptation. Bars show the power (i.e., proportion of successful simulations) separately for each combination of 1:1 crossing scheme version with the dominance coefficient of the additional target of selection: 1f, only one target of selection on the focal haplotype; 2f_h05/2f_h1, two targets of selection on the focal haplotype; 1f1nf_h05/1f1nf_h1, one target of selection on the focal haplotype and one on one nonfocal haplotype. Columns depict different population sizes (300; 1,200; 4,800; 19,200 individuals), and rows show different dominance coefficients of the target of interest (additive in top row, dominant in bottom row). The dots with error bars display the model fit (Model 2) and its 95% confidence interval. For the model fit, the selection coefficient was fixed to its global average. (*A*) shows the results for success-A (=selection target is the SNP with the highest Cochran–Mantel–Haenszel [CMH] test statistic). (*B*) shows the results for success-B (=selection target is not more than 100 SNPs away from the SNP with the highest CMH test statistic).

crossing scheme (LRT full-reduced model comparison [Model 2]: $\chi^2 = 1,727.7$, df = 32, $P < 0.001$ [success-A]; $\chi^2 = 2,588.42$, df = 32, $P < 0.001$ [success-B]; supplementary tables S4 and S5, Supplementary Material online). This significant effect is mainly driven by one scenario: when the focal line harbors two targets of selection with the target of interest being dominant and the additional target being codominant (fig. 5) the power to detect the target of interest is close to zero. The reason is that with a starting frequency of 50%, codominant alleles respond more to strong selection than dominant alleles, because nonfavored alleles are masked by high frequency dominant alleles (supplementary fig. S5, Supplementary Material online). This differential behavior is particularly pronounced for high allele frequencies. Note that our definitions of success do not use prior information on the location of the focal SNP of interest. We propose that a substantial fraction of the power can be recovered if the analysis is restricted to the approximate location of the focal target determined in the primary E&R study. On the other hand, for additive targets of interest one additional target of selection hardly reduces the power of the 1:1 crossing scheme, regardless of the dominance coefficient of the additional target and whether the additional target is positioned on the focal or one nonfocal haplotype (fig. 5). Overall, our results show that the 1:1 crossing scheme still outperforms dil:mt even in the presence of one additional target of selection (figs. 3 and 5). For the remaining analysis, we will focus on a 1:1_1f crossing scheme with only one target of selection.

Alternative Crossing Schemes

Since large population sizes can be challenging to maintain, we also evaluated additional crossing schemes which may perform better even for smaller population sizes (Model 3, Assessment of Experimental and Population Genetic Parameters in the Materials and Methods section). 1:few (fig. 1*C*) is a modification of the 1:1_1f crossing scheme, which combines all nonfocal lines in each replicate—this provides a consistent genetic composition in each replicate. The 1:many crossing scheme (fig. 1*D*) uses one focal, and 99 nonfocal lines without selection target. We reasoned that the larger number of segregating variants (970,466 SNPs compared with max 527,771 SNPs in 1:1 crossing scheme, supplementary table S1, Supplementary Material online) potentially provides a higher mapping resolution. Finally, we modified the dilution crossing scheme by simulating only one selection target in the founder population (dil:st, fig. 1*E*). Hence, no additional selection targets can create potentially

confounding selection signatures. For each experimental design, we simulated a population size of 300 individuals, 60 generations of adaptation, and five replicates (table1).

Regardless of the experimental design, the simulated focal selection targets experienced a pronounced allele frequency increase (supplementary fig. S6, Supplementary Material online). In "dilution" crossing schemes (dil:st, dil:mt; fig. 1*E* and *F*) the frequency trajectories of the focal SNPs were highly variable because of the heterogeneous starting frequency. The crossing schemes where a single focal line with a starting frequency of 50% carries the beneficial allele typically had a superior performance than dilution crossing schemes (fig. 6). Notably, the significant influence of crossing scheme on secondary E&R success (LRT full-reduced model comparison [Model 3]: $\chi^2 = 295.9$, df = 8, $P < 0.001$ [success-A]; $\chi^2 = 190.23$, df = 8, $P < 0.001$ [success-B]) (supplementary tables S6 and S7, Supplementary Material online) cannot be explained by a loss of the focal SNP in dilution crossing schemes, which occurred in less than 1% of the simulations.

We found that the dil:st crossing scheme is still inferior to the 1:1_1f crossing scheme for additive loci (fig. 6). Also, the modifications of the 1:1_1f crossing scheme (1:few, 1:many) do not provide a substantial improvement (fig. 6). Surprisingly, the 1:many crossing scheme performed very poorly. This may be at least partly attributed to the larger number of SNPs/kb, which will affect the SNP-based success rate (success-B, fig. 2). Given that the three additional crossing schemes do not provide a clear advantage (figs. 6 and 7; supplementary fig. S7, Supplementary Material online), and are not easier to execute experimentally, we did not evaluate them with larger population sizes.

Number of Replicates

Given the superior performance of experimental designs with larger population sizes (fig. 3), we were also interested whether more replicates with a smaller population size may provide further improvements. We performed additional simulations for all five different crossing schemes with a population size of 300 individuals per replicate, and 30 replicates (table 1). Consistent with other simulation studies (Kofler and Schlötterer 2014; Kessner and Novembre 2015), more replicates result in increased power (supplementary fig. S8, Supplementary Material online). The power of secondary E&R with 30 replicates (supplementary table S6, Supplementary Material online) is only significantly affected by the crossing scheme considering all modeled interactions (LRT full-reduced model comparison [Model 3]: $\chi^2 = 31.467$, df = 8, $P < 0.001$ [success-A]; $\chi^2 = 16.692$, df = 8, $P = 0.033$ [success-B]; supplementary table S7, Supplementary Material online). The 1:1_1f crossing scheme with 30 replicates also has the highest resolution (supplementary fig. S9*A*, Supplementary Material

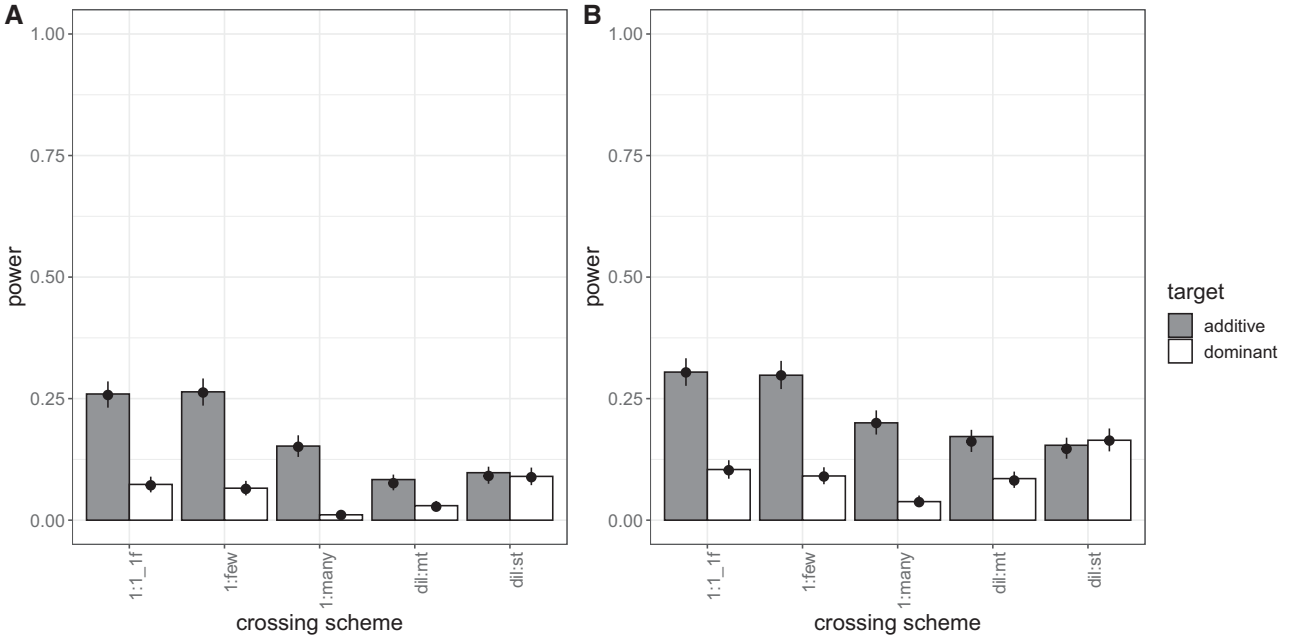

Fig. 6.—Power of five different crossing schemes (population size=300 individuals; 2,000 simulations/experimental design). Bars show the power (i.e., proportion of successful simulations) separately for each combination of crossing scheme and dominance coefficient (additive in gray; dominant in white). The dots with error bars display the estimate from the fitted model (Model 3) and its 95% confidence interval. For the model fit, the selection coefficient was fixed to its global average, and combination-specific average starting allele frequencies were used. (*A*) shows the results for success-A (=selection target is the SNP with the highest Cochran–Mantel–Haenszel (CMH) test statistic), (*B*) shows the results for success-B (=selection target is not more than 100 SNPs away from the SNP with the highest CMH test statistic).

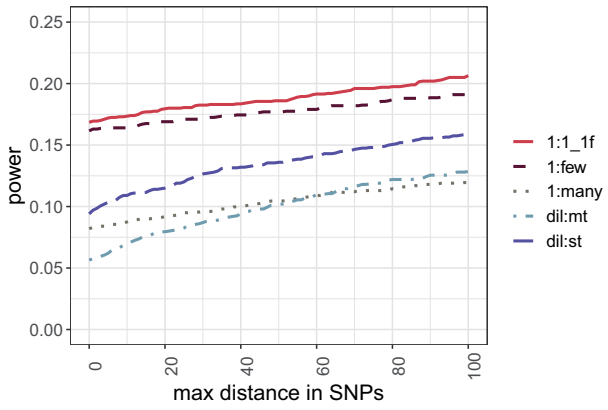

Fig. 7.—Resolution of five different crossing schemes. Proportion of simulations (*y* axis) that do not exceed a maximum distance in SNPs (*x* axis) between the SNP with the highest Cochran–Mantel–Haenszel test statistic and the true target of selection.

online). As expected, the impact of linkage disequilibrium, indicated by the number of ties, is dramatically reduced for all crossing schemes when more replicates are simulated (supplementary fig. S9*B*, Supplementary Material online, compared with supplementary fig. S7, Supplementary Material online). However, similar improvements as seen with 30 replicates can be achieved if five replicates of the 1:1_1f crossing scheme are used with a population size of 1,200 individuals (fig. 3). Thus, the same power can be achieved while maintaining about 33% fewer individuals.

Dominance Coefficient, Selection Coefficient, and Mean Starting Allele Frequency

As expected, dominance coefficient, selection coefficient, and average starting allele frequency have a significant effect on the success of secondary E&R in all simulation scenarios, except for data with 30 replicates. We observed that selection targets with a higher selection coefficient and/or a higher mean starting allele frequency are easier to fine map. The lack of significant effects of these parameters on secondary E&R success for the data set with 30 replicates can probably be explained by the smaller number of conducted simulations (100 for 30 replicates and 2,000 for five replicates, table 1). After crossing scheme and population size, the dominance coefficient is the parameter with the highest Nagelkerke's R^2 index in our analysis, with additive loci being easier to fine map. This is caused by the fact that heterozygotes and target homozygotes with a dominant beneficial SNP have the same fitness, resulting in less efficient selection that becomes especially apparent if the selected allele has already reached a high frequency in the population, and nonselected allele homozygotes become rare (indicated by a significant negative effect of the interaction term between dominance coefficient and mean starting allele frequency in most of our statistical models).

Discussion

This work was inspired by the difficulty of most E&R studies with sexually reproducing organisms to pinpoint selection targets, mostly due to numerous neutral hitchhikers resulting in large haplotype blocks. Secondary E&R—a follow-up EE validating putative selection targets of a primary E&R under an identical selection regime—has been recently suggested as experimental approach for selection target confirmation (Burny et al. 2020). We used extensive computer simulations to evaluate how experimental and population genetics parameter shape the power and resolution of secondary E&R. As expected, dominance coefficient, selection strength, and mean starting allele frequency of the selected target all have a significant effect on the success of E&R, where dominant selection targets at high frequency are particularly challenging to detect. However, population size and crossing scheme emerged as the most influential parameters in our analysis.

The Crossing Scheme Has a Pronounced Effect on Secondary E&R Success

We show that a simple crossing scheme, which only requires that (a subset of) the founder lines are sequenced and that founder lines with and without the selection target of interest can be distinguished, has the best power and resolution of the five crossing schemes tested. The 1:1 crossing scheme is particularly well-suited when many selection targets are detected in the evolved populations of the primary E&R, because it uses only a subset of the lines from the nonadapted ancestral founder population of the primary E&R study. This reduces potential confounding effects between the selection target of interest and other adaptive loci in two ways. First, nonadapted ancestral founder haplotypes will harbor on average less selection targets than evolved haplotypes that can acquire multiple selection targets through recombination events during the primary E&R (Otte and Schlötterer 2021). Because evolved haplotypes will often harbor multiple selection targets, beneficial alleles will not propagate independently from each other, which makes it challenging to fine-map single selection targets. Second, the total number of potential beneficial alleles in a 1:1 crossing scheme is deliberately reduced by picking only a subset of the founder lines of the primary E&R, which facilitates fine mapping of one particular selection target of interest with secondary E&R.

We would like to point out that in our study "crossing schemes" are defined such that they do not only differ in the actual crossing procedure, but also in the number and starting allele frequency distribution of simulated selection targets (e.g., 1:1_1f vs. dil:mt). Additional beneficial alleles can without doubt have an impact on the power of secondary E&R. Although the 1:1 crossing scheme still outperforms dil:mt in the presence of one additional selection target, the relative performance of crossing schemes might change given an even more complex underlying genetic architecture.

Furthermore, we focused in this study on the assessment of experimental and population genetic parameters given a directional selection regime. Future research on the potential of secondary E&R to fine map selection targets under a complex/polygenic genetic architecture is needed, such as for example a quantitative trait under stabilizing selection that experienced a recent shift in trait optimum.

Sufficient E&R Power Requires Large Population Sizes

Our analyses also indicate that rather large population sizes are crucial to identify the causative variant with sufficient confidence using a secondary E&R approach. This observation explains why an empirical secondary E&R study in *D. simulans* with a population size of 1,250 flies per replicated population failed to narrow down a pronounced candidate region from a primary E&R in a secondary E&R after 30 generations, and only confirmed the presence of selected alleles (Burny et al. 2020).

Our results suggest that large population sizes become especially crucial for short-term secondary E&R. The low power of short-term E&R with reduced experimental duration (Kofler and Schlötterer 2014) can be improved by a larger experimental population size. Our analyses show that to achieve satisfactory power after only 20 generations requires population sizes of tens of thousands of individuals—dimensions that are rather rare in current EE designs with strictly sexually reproducing organisms, where population sizes are mostly limited by the capacities to conduct such large experiments. Similar to previous simulation studies (Kofler and Schlötterer 2014; Kessner and Novembre 2015), we demonstrated that increasing the number of replicates results in more powerful secondary E&R. If large population sizes are not feasible, maintaining many replicates at small population size can help to boost secondary E&R performance. Additionally, the loss of one replicate in a highly replicated setup has a smaller impact than in an experimental setup with very few replicates. However, we show that additional replication cannot completely compensate the advantage of large population sizes in secondary E&R. Based on this observation, we conclude that independent of the crossing scheme, the power of secondary E&R will benefit from larger population sizes. Furthermore, the resolution will significantly increase with more replicates—a result that we anticipate can be generalized to different underlying genetic architectures as well as (model) organisms.

Allele Frequency Estimation Errors and Read Depth

The results of this study are based on true allele frequencies estimated without error—a rather optimistic assumption that will not hold for empirical data. In empirical E&R studies, numerous factors influence the accuracy of allele frequency estimates, such as the rate of sequencing errors, genome-wide read depth heterogeneity, and the average read depth (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014; Tilk et al. 2019).

Individual whole-genome sequencing for entire populations becomes quickly prohibitive with increasing sample size. Sequencing pools of individuals (Pool-Seq) (Schlötterer et al. 2014) provides a cost-effective approach that allows to robustly estimate allele frequency estimates and has become the method of choice for most E&R studies (Turner et al. 2011; Schlötterer et al. 2015). Since typical Pool-Seq studies combine all individuals from a given generation to reduce the sampling error (Schlötterer et al. 2014), the read depth of Pool-Seq studies will affect all experimental designs to the same extent as long as the coverage is considerably lower than the pool size.

Choice of Model Organism

Our simulations are parameterized for *D. simulans* because this species is better suited for E&R studies than *D. melanogaster* (Barghi et al. 2017). Although other species, such as *Saccharomyces cerevisiae* (Burke et al. 2014) or *Caenorhabditis remanei* (Castillo et al. 2015) have been used to study adaptive response from standing genetic variation in outcrossing species, *Drosophila* is currently the most popular organism. Nevertheless, we anticipate that the results of this study can be generalized to other species, but the availability of sequenced inbred founder lines is probably a more severe restriction, which could limit the widespread use of secondary E&R.

Selection Regime

Our simulations are targeted for selection regimes with a moderate number of selection targets, with rather strong effects, as seen in empirical E&R studies using *Drosophila* (Mallard et al. 2018; Barghi et al. 2019; Michalak et al. 2019). Since strongly selected haplotype blocks typically have low starting frequencies, we consider the 1:1 crossing scheme a very realistic case, as most of the selection targets will be found only in a few ancestral founder lines. As outlined above, evolved haplotypes in dil:mt on the other hand will most likely carry multiple beneficial alleles that have recombined over the course of the primary E&R, which makes it challenging to fine-map one distinct selection target of interest as our results suggest.

Furthermore, we assumed the rather simple selection scenario of directional selection. More complex scenarios that include pleiotropy and/or epistasis were not considered. Although both are very important factors that could affect the allele frequency changes in an E&R study, we would like to point out that secondary E&R is designed to study loci, which experienced a substantial allele frequency increase in a primary experiment, thus pleiotropic constraints are not expected to be a major confounding factor. Epistasis, on the other hand may be more pronounced in the 1:1 crossing

scheme. It may be possible that the focal allele is only selected in a subset of the replicates, but not in others-or the response may differ between replicates. Although this reduces the power of secondary E&R to fine-map the selection target, such a result may open the possibility to study the impact of epistatic effects. Since the founder lines will be available, it will be possible to repeat the experiment with replication for each genotype combination to assure that the heterogeneous response comes from epistatic effects and is not linked to stochastic changes.

Finally, it is important to keep in mind that the secondary E&R designs discussed here are not tailored to study highly polygenic traits, because in this study, we were assuming that recombination facilitates the identification of a selection target with a major effect. For a highly polygenic trait, selected haplotype blocks identified in the primary E&R experiment (with multiple selected loci) would be broken down during the secondary E&R, which reduces their selective advantage. To what extent this different behavior in secondary E&R experiments can be used to distinguish these different architectures requires further work.

Supplementary Material

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

Acknowledgments

We thank the members of the Institute of Population Genetics for fruitful discussion and support. Thanks to Roger Mundry who shared R scripts for the statistical analyses. Special thanks to Rupert Mazzucco, Robert Kofler, Kathrin Otte, and Christos Vlachos for their support regarding the computer simulations. Special thanks to Claire Burny and Robert Kofler for helpful comments on earlier versions of the manuscript. Last, but not least we are thankful to the anonymous reviewers for their constructive feedback. This work was supported by the Austrian Science Funds (FWF) (Grant Nos. W1225 and P-32935) and by the European Research Council (ERC) (Grant ArchAdapt).

Author Contributions

AML, MD, and CS designed the analysis. AML performed the forward simulations and the bioinformatic analysis. AML and MD performed the statistical analysis. AML, MD, and CS wrote the manuscript.

Data Availability

Information regarding the accessibility of raw sequence reads, phased haplotypes of the ancestral *Drosophila simulans* haplotypes as well as MimicrEE2 ready text files for the *D. simulans* recombination map (file used in this project: Dsim_recombination_map_LOESS_100kb_1.txt) can be found in Howie et al. (2019). MimicrEE2 ready input files of the different experimental designs, the simulated selection regimes, processed simulation results and all scripts that are necessary to reproduce the results are available at SourceForge (https://sourceforge.net/projects/secondary-e-r-sim/files/, last accessed July 9, 2021).

Literature Cited

Agresti A, Kateri M. 2011. Categorical data analysis. In: Lovric M, editor. International Encyclopedia of Statistical Science. Berlin Heidelberg: Springer. p. 206–208.

Baayen RH. 2008. Analyzing linguistic data: a practical introduction to statistics using R. Cambridge, England: Cambridge University Press.

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.

Barghi N, et al. 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*. PLoS Biol. 17(2):e3000128.

Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. Nat Rev Genet. 21(12):769–781.

Barghi N, Tobler R, Nolte V, Schlötterer C. 2017. *Drosophila simulans*: a species with improved resolution in evolve and resequence studies. G3 (Bethesda) 7(7):2337–2343.

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

Burny C, et al. 2020. Secondary evolve and resequencing: an experimental confirmation of putative selection targets without phenotyping. Genome Biol Evol. 12(3):151–159.

Castillo DM, Burger MK, Lively CM, Delph LF. 2015. Experimental evolution: assortative mating and sexual selection, independent of local adaptation, lead to reproductive isolation in the nematode *Caenorhabditis remanei*. Evolution 69(12):3141–3155.

Dowle M, Srinivasan A. 2019. data.table: extension of 'data.frame'. Available from: https://cran.r-project.org/package=data.table. Accessed May 1, 2020.

Fox J, Monette G. 1992. Generalized collinearity diagnostics. J Am Stat Assoc. 87(417):178–183.

Fox J, Weisberg S. 2019. An R companion to applied regression. Thousand Oaks (CA): Sage. Available from: https://socialsciences.mcmaster.ca/jfox/Books/Companion/. Accessed May 1, 2020.

Franssen SU, Nolte V, Tobler R, Schlötterer C. 2015. Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila melanogaster* populations. Mol Biol Evol. 32(2):495–509.

Garland T, Rose M. 2009. Experimental evolution: concepts, methods, and applications of selection experiments. Berkeley (CA): University of California Press.

Gratz SJ, et al. 2013. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. Genetics 194(4):1029–1035.

Howie JM, Mazzucco R, Taus T, Nolte V, Schlötterer C. 2019. DNA motifs are not general predictors of recombination in two *Drosophila* sister species. Genome Biol Evol. 11(4):1345–1357.

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

Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö. 2010. Genome-wide effects of long-term divergent selection. PLoS Genet. 6(11):e1001188.

Kawecki TJ, et al. 2012. Experimental evolution. Trends Ecol Evol. 27(10):547–560.

Kelly JK, Hughes KA. 2019. Pervasive linked selection and intermediate-frequency alleles are implicated in an evolve-and-resequencing experiment of *Drosophila simulans*. Genetics 211(3):943–961.
Kessner D, Novembre J. 2015. Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. Genetics 199(4):991–1005.
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, et al. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500(7464):571–574.
Langmüller AM, Schlötterer C. 2020. Low concordance of short-term and long-term selection responses in experimental *Drosophila* populations. Mol Ecol. 29(18):3466–3475.
Long A, Liti G, Luptak A, Tenaillon O. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat Rev Genet. 16(10):567–582.
Mallard F, Nolte V, Tobler R, Kapun M, Schlötterer C. 2018. A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in *Drosophila*. Genome Biol. 19(1):119.
Martins NE, et al. 2014. Host adaptation to viruses relies on few genes with different cross-resistance properties. Proc Natl Acad Sci U S A. 111(16):5938–5943.
Michalak P, Kang L, Schou MF, Garner HR, Loeschcke V. 2019. Genomic signatures of experimental adaptive radiation in *Drosophila*. Mol Ecol. 28(3):600–614.
Nagelkerke N. 1991. A note on a general definition of the coefficient of determination. Biometrika 78(3):691–692.
Nuzhdin SV, Turner TL. 2013. Promises and limitations of hitchhiking mapping. Curr Opin Genet Dev. 23(6):694–699.
Otte KA, Schlötterer C. 2021. Detecting selected haplotype blocks in evolve and resequence experiments. Mol Ecol Resour. 21(1):93–109.
R Core Team 3.5.3. 2019. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.r-project.org">https://www.R-project.org. http://www.r-project.org. Accessed February 15, 2019.
Remolina SC, Chang PL, Leips J, Nuzhdin SV, Hughes KA. 2012. Genomic basis of aging and life-history evolution in *Drosophila melanogaster*. Evolution 66(11):3390–3403.
Schielzeth H. 2010. Simple means to improve the interpretability of regression coefficients. Methods Ecol Evol. 1(2):103–113.
Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity (Edinb). 114(5):431–440.
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.
Spitzer K, Pelizzola M, Futschik A. 2020. Modifying the chi-square and the CMH test for population genetic inference: adapting to overdispersion. Ann Appl Stat. 14(1):202–220.
Taus T, Futschik A, Schlötterer C. 2017. Quantifying selection with pool-seq time series data. Mol Biol Evol. 34(11):3023–3034.
Teotónio H, Chelo IM, Bradić M, Rose MR, Long AD. 2009. Experimental evolution reveals natural selection on standing genetic variation. Nat Genet. 41(2):251–257.
Tilk S, et al. 2019. Accurate allele frequencies from ultra-low coverage pool-seq samples in evolve-and-resequence experiments. G3 (Bethesda) 9:4159–4168.
Tobler R, et al. 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, 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.
Vlachos C, et al. 2019. Benchmarking software tools for detecting and quantifying selection in evolve and resequencing studies. Genome Biol. 20(1):169.
Vlachos C, Kofler R. 2018. MimicrEE2: genome-wide forward simulations of Evolve and Resequencing studies. PLoS Comput Biol. 14(8):e1006413.
Vlachos C, Kofler R. 2019. Optimizing the power to identify the genetic basis of complex traits with evolve and resequence studies. Mol Biol Evol. 36(12):2890–2905.

Associate editor: Charles Baer