Fine mapping without phenotyping: Identification of selection targets in secondary Evolve and Resequence experiments

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

[1] Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

[2] Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

[3] Plattform Bioinformatik und Biostatistik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

*Author of Correspondence:

Christian Schlötterer, Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

+43 1 25077-4301, +43 25077-4390

christian.schloetterer@vetmeduni.ac.at

Abstract

Evolve and Resequence (E&R) studies investigate the genomic selection response of populations in an Experimental Evolution (EE) setup. Despite the popularity of E&R, empirical studies typically suffer from an excess of candidate loci due to linkage, and single gene or SNP resolution is the exception rather than the rule. Recently, a secondary E&R design – where unevolved founder genotypes are added to a primary E&R study - has been suggested as promising experimental procedure to confirm putative selection targets. Furthermore, secondary E&R provides 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 crossing scheme, population size, experimental duration, and number of replicates on the power, and resolution of secondary E&R. We find that the crossing scheme, population size, and the experimental duration are crucial factors for the 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 5 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 towards improved experimental designs aiming to characterize causative SNPs in EE studies.

Keywords

Experimental Evolution, secondary Evolve and Resequence, experimental design, *Drosophila*, fine mapping

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 (1, 2). In contrast to natural populations EE offers the key advantage of replicating experiments under controlled laboratory conditions (3). Evolve and Resequence (E&R) (3–5) – 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 (6), yeast (7), *C. elegans* (8) and *Drosophila* (9–12). Despite its popularity, E&R typically suffers from an excess of candidates caused by linkage disequilibrium between true causative SNPs and neutral hitchhikers (13–15), which decreases the resolution of E&R studies and makes single gene resolution (9) 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 (16–19), and the combination of time-series data with replicate populations has been identified as particularly powerful (7, 20, 21). 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 (22). 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 (23–26).

Secondary E&R (27) has been recently introduced as a promising experimental follow-up approach to validate selection signals of primary E&R. 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. 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 (28)). A hitherto under-explored 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.

Here we suggest additional secondary E&R experimental designs and evaluate their power and resolution 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 5 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 to more complex crossing schemes.

91 **Methods**

92 Outline of the simulation framework

93 The founder lines used in our simulation study are a randomly chosen subset of 100
94 haplotypes from a panel of 189 sequenced *D. simulans* haplotypes originally collected in
95 Tallahassee, Florida, USA capturing the amount of standing genetic variation in a natural
96 *Drosophila* population (12, 29). In order to speed up the calculations, we only simulated
97 chromosome-arm 2L. Simulations with linkage were conducted with MimicrEE2 (v206) (30)
98 using the *D. simulans* recombination map (29). MimicrEE2 is a forward-simulation framework
99 for E&R studies that can simulate evolving experimental populations based on their haplotype
100 information and genome-wide recombination rates. We used the w-mode of MimicrEE2
101 which computes the fitness of individuals directly from the selection coefficients. We
102 simulated positive, additive selection for biallelic SNPs, with selection coefficients being
103 uniformly sampled between 0.07 and 0.1 for each positively selected SNP and tracked the
104 frequency of all SNPs over time (ranging from 479,507 to 970,466 SNPs, Table S1). We chose
105 to simulate rather strong selection reasoning that alleles with a high selection coefficient are
106 more likely to be experimentally tested. On the other side, strongly selected alleles result in
107 many neutral linked hitchhikers producing false positive signals that adversely impact the
108 mapping resolution (25). Selected SNPs were randomly chosen with an equal probability to
109 be either co-dominant (dominance coefficient h=0.5), or dominant (h=1). We did not consider
110 recessive loci (h=0), because we do not anticipate that fully recessive targets would result in
111 sufficiently large allele frequency changes to be detected in primary E&R experiments (24,
112 25). We rescaled allele counts for each biallelic position to a uniform read depth of 80.

113

114

115 Experimental parameters

116 The purpose of this study is to test the influence of different experimental parameters on the
117 power and resolution of secondary E&R. For this, we systematically varied the crossing
118 scheme, population size, experimental duration, and number of replicates to assess the effect
119 of these experimental parameters on the power and resolution of secondary E&R (Table 1).
120 We use the term "experimental design" to describe a distinct set of simulated experimental
121 parameters (*e.g.,* crossing scheme=1:1; population size=1,200 individuals; experimental
122 duration=60 generations; 5 replicates).

123

124 *Crossing scheme*

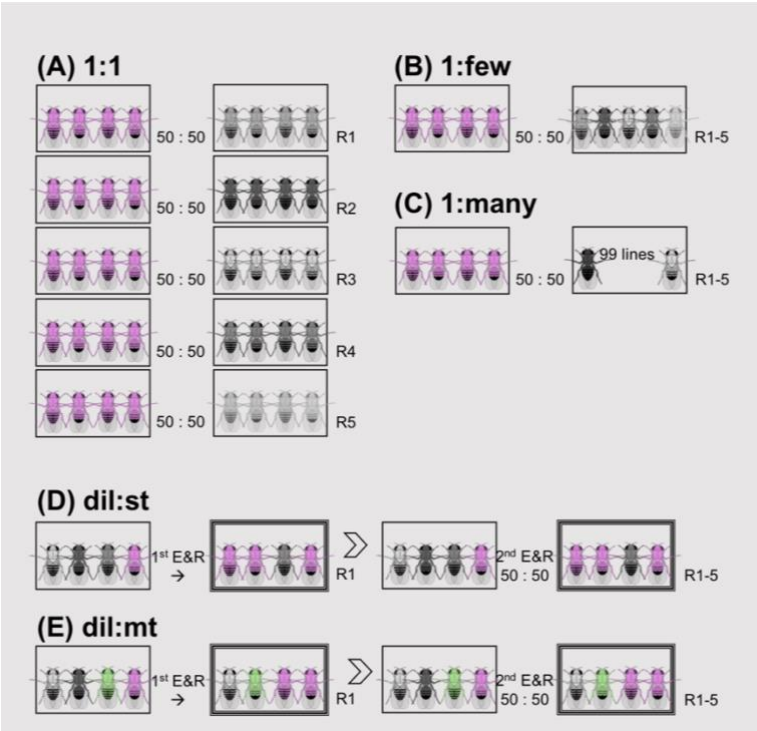
125 We simulated five different crossing schemes: (A) 1:1, (B) 1:few, (C) 1:many, (D) dil:st, and (E)
126 dil:mt (Figure 1). In the **1:1** crossing scheme (Figure 1A), two inbred lines are crossed at equal
127 proportions. One inbred focal line carries a single target of selection - known from a primary
128 E&R experiment – and is crossed with inbred non-focal lines, not carrying known adaptive
129 alleles. Replicates are created by crossing the same focal line to different non-focal lines
130 without known targets of selection. We used different non-focal lines as crossing partners to
131 account for the possibility that in real experiments these lines may contain unidentified
132 selected loci. By using different lines, the influence of selection targets present in a single line
133 will be outweighed by the focal locus, which is present in every replicate (Figure 1A).

134

135 In the **1:few** crossing scheme (Figure 1B) the focal line is crossed with a pool of five non-focal
136 lines that do not carry known beneficial alleles. The starting frequency of the focal line is 50
137 %, whereas each non-focal line has a starting frequency of 10 %. Each replicate consists of the
138 same focal/non-focal line mixture. In the **1:many** crossing scheme (Figure 1C), the focal line

139 is crossed with a pool of 99 non-focal lines. Replicates consist of the same mixture of lines,
140 and the starting frequency of the focal line is 50 %.
141
142 It has been recently suggested, that "diluting" evolved populations of a primary E&R
143 experiment with many ancestral genotypes of the very same primary E&R and exposing the
144 diluted populations to the same selection regime is a promising approach to validate selection
145 candidates (27). However, we lack a systematic power assessment of such experiments with
146 computer simulations. We thus included two "dilution" crossing schemes (**dil:mt, dil:st**) into
147 our analysis. To evaluate the power of dilution crossing schemes, it is important to first
148 simulate a primary E&R study. We chose to simulate a population consisting of 100 different
149 founder lines, a population size of 300 individuals, 60 generations of adaptation and one
150 replicate, which can then be diluted with genotypes of the ancestral population. SNPs in the
151 primary E&R were tested for allele frequency change with the χ^2 test.
152 The population of the **dil:st** crossing scheme carries only one beneficial SNP (**dil**:**st** = **dil**ution:
153 **s**ingle **t**arget, Figure 1D). The starting allele frequency of this beneficial, focal SNP is sampled
154 from the empirical starting allele frequency distribution of putatively selected alleles from a
155 previous E&R study in which *D. simulans* populations adapted to a new temperature regime
156 (mean starting frequency = 0.1) (12). After simulating 60 generations of adaptation (primary
157 E&R), 50 % of the evolved population is replaced by flies of the ancestral founder population.
158 After this dilution step, the secondary E&R was simulated under the exact same selection
159 regime as in the primary E&R. A region with a strong selection signal from the primary E&R is
160 chosen for validation in the dilution crossing schemes. Hence, we investigated a 1 Mb
161 window, which is the previously reported median selected haplotype block length on

162 chromosome-arm 2L (12), around the SNP with the highest χ^2 test-statistic in the primary
163 E&R, in the secondary E&R.
164
165 The **dil:mt** (Figure 1E) crossing scheme has multiple selection targets (**dil**:**mt** = **dil**ution:
166 **m**ultiple **t**argets; 16 selection targets on chromosome-arm 2L (12)). Again, we investigate a 1
167 Mb window around the SNP with the highest χ^2 test statistic in the primary E&R, in the
168 secondary E&R. In case there are multiple selection targets in the 1 Mb window, we consider
169 the target that is closest to the SNP with the highest CMH test statistic in the 1Mb window of
170 the secondary E&R in our analysis.



171
172 Figure 1 Simulated crossing schemes. (A) **1:1** crossing scheme: Inbred flies with one target of selection (purple)
173 are crossed to inbred flies without known beneficial variants. The starting frequency of each genotype is 50 %.
174 In each replicate the line with the beneficial allele (focal line, purple) is crossed to a different line lacking
175 beneficial mutations (non-focal lines, different shades of grey). (B) **1:few** crossing scheme: The focal line is
176 crossed to a pool of flies with five different genotypes without known selection targets. The starting frequency
177 of the focal line is again 50 %. (C) **1:many** crossing scheme: The focal line is crossed to a pool of 99 lines without
178 known selection targets. (D) **dil:st** crossing scheme: 50 % of an evolved population originating from a primary
179 E&R is replaced by ancestral genotypes of the primary E&R. The entire population has only one single target of
180 selection (focal SNP, purple flies) (E) **dil:mt** crossing scheme: 50 % of an evolved population originating from a
181 primary E&R is replaced by ancestral genotypes of the primary E&R. The ancestral population carries 16 targets
182 of selection (flies carrying different beneficial SNPs are shown in purple, and green).
183

184 *Population size*

185 For two crossing schemes that are relatively easy to implement in empirical studies (1:1,
186 dil:mt) we performed simulations (experimental duration = 60 generations; 5 replicates) with
187 independently sampled selection targets for population sizes of 300; 1,200; 4,800; and 19,200
188 individuals per replicate to test for the effect of the population size on the power and
189 resolution of secondary E&R (Table 1). All other crossing schemes were evaluated at a
190 population size of 300 individuals.

191

192 *Experimental duration*

193 All crossing schemes were evaluated after 60 generations. To explore the possibility that
194 shorter experiments with less than 60 generations may already be sufficient to achieve
195 satisfactory power, we analyzed the simulations for 1:1 and dil:mt crossing schemes also
196 already after 20 generations (Table 1).

197

198 *Number of replicates*

199 To assess the impact of the number of replicates on power of secondary E&R, we conducted
200 for each crossing scheme 100 additional simulations (population size= 300 individuals;
201 experimental duration = 60 generations) with 30 replicates (Table 1).

202

203 Table 1 Simulation overview.

Crossing scheme	Population size	Generations	Replicates	Number of simulations
1:1	300; 1,200; 4,800; 19,200	60; 20	5/30	2,000/100
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

204 Statistical analysis

205 All statistical analyses were performed using the R statistical computing environment (v3.5.3)
206 (31).

207

208 *Power*

209 For each simulation, we tested all SNPs on chromosome-arm 2L for allele frequency increase
210 between the start and the end of the simulated secondary E&R using the Cochran-Mantel-
211 Haenszel-test (CMH-test, implemented in the R package poolSeq (v0.3.2) (32)). The CMH-test
212 allows to test for independence of matched data – e.g. allele counts of replicated ancestral
213 and evolved populations (33). We ranked SNPs based on their CMH test statistic using the
214 dense ranking method in the R package data.table (v.1.12-8 (34)). In dense ranking, SNPs with
215 identical test statistics receive the same rank, and the following SNP is assigned the
216 immediately following rank.

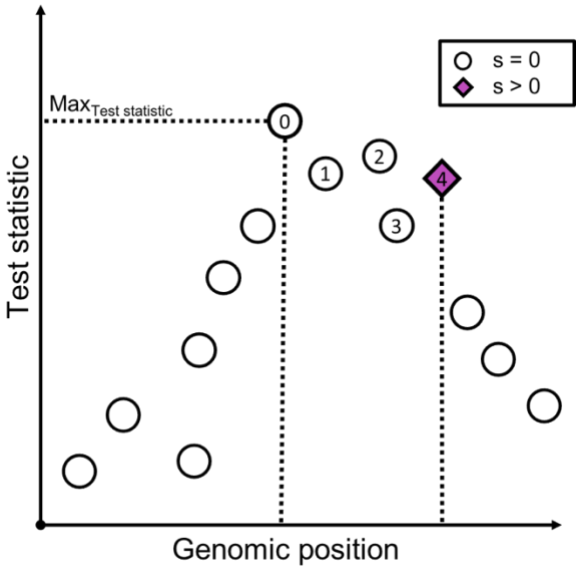
217

218 Based on this ranking, we used two different approaches to classify simulations being either
219 successful, or unsuccessful. First, we considered a simulation to be only successful if the true
220 target of selection (*i.e.,* the focal SNP with a selection coefficient > 0) was the SNP with the
221 highest test statistic (success-A). In a second analysis step, we classified a simulation as
222 success, if the true target of selection was not more than 100 SNPs away from the SNP with
223 the highest test statistic (Figure 2, success-B). We acknowledge that success-B is dependent
224 on the maximum distance that we allow between the target of selection and the SNP with
225 the highest CMH test statistic (Figure 2). However, varying the maximum distance threshold
226 did not alter the relative performance of different experimental designs (data not shown).

227 The power of an experimental design is defined as the proportion of simulations that were

228 able to detect the true target of selection.

229



230
231 Figure 2 Schematic overview of the definition of success-B in a secondary Evolve and Resequence simulation. All
232 SNPs (neutral = circle, beneficial = purple diamond) are tested for an allele frequency change with the Cochran-
233 Mantel-Haenszel-test, and are ranked based on their test statistic (y-axis). If the true selection target (purple
234 diamond) is less than 100 SNPs away from the SNP with the highest test statistic, the simulation run is deemed
235 a success. In the example depicted, the distance in number of SNPs between the true target of selection, and
236 the SNP with the highest test statistic is 4, and the simulation is classified as success.

237

238 *Resolution*

239 For simulations where the selection target was detected (success-B), we determined the

240 resolution of fine mapping of the selection target by counting the number of SNPs between

241 the true selection target, and the SNP with the highest CMH test statistic.

242

243 *Assessment of experimental and population genetic parameters*

244 We used logistic regressions with a binomial error structure (ε) and a logit link function (35)

245 to test how different experimental and population genetic parameters affect success and

246 failure to identify targets of selection, i.e. success (Y) is treated as a binary response encoded

in 0 and 1 of secondary E&R. We fitted two different models with R function *glm* with μ being the overall mean per model*:*

(1) $Y_{ijklmn} = \mu + cross_i + h_j + s_k + af_l + N_m + (cross{:}h)_{ij} + (h{:}af)_{jl} + (cross{:}N)_{im} + (h{:}N)_{jm} + (cross{:}h{:}N)_{ijm} + \varepsilon_{ijklmn}$

(2) $Y_{ijklm} = \mu + cross_i + h_j + s_k + af_l + (cross{:}h)_{ij} + (h{:}af)_{jl} + \varepsilon_{ijklm}$

Model 1 (equation (1)) contained (i) five explanatory variables as main effects; crossing scheme ($cross_i$), a fixed categorical effect with 2 levels - 1:1 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 4 levels - 300; 1,200; 4,800; 19,200, (ii) an interaction term between dominance coefficient and mean starting allele frequency ($(h{:}af)_{jl}$), (iii) and a triple interaction between crossing scheme, dominance coefficient, and population size ($(cross{:}h{:}N)_{ijm}$), and all pairwise interaction terms of effects involved in the triple interaction term ($(cross{:}h)_{ij}$; $(cross{:}N)_{im}$; $(h{:}N)_{jm}$). We included interaction terms into the model that have population genetic interpretations.

Prior to model fitting, the two covariates selection coefficient and mean starting allele frequency were multiplied by 100, and z-transformed to a mean of zero and standard deviation of one for easier interpretable estimates (36). Data analyzed with model 1 contained 16,000 observations, namely 2,000 independent simulation runs for each crossing scheme (1:1, dil:mt), and each of the four different population sizes (300; 1,200; 4,800;

271 19,200) (Table 1). To avoid potential bias, we randomly chose 4 sets of focal/non-focal
272 founder lines for the 1:1 crossing scheme and performed 500 simulation runs per set. The
273 chosen set of founder lines did not have a significant effect on the success of secondary E&R
274 and is thus not included in the final model (LRT full-null model comparison; data not shown).
275
276 Model 2 (equation (2)) contained (i) four explanatory variables; crossing scheme ($cross_i$), a
277 fixed categorical effect with 5 levels - 1:1; 1:few; 1:many; dil:st; dil:mt, dominance coefficient
278 (h_j), selection coefficient (s_k), and mean starting allele frequency (af_l) over replicated
279 populations as main effects, as described for model 1, (ii) an interaction term between
280 crossing scheme and dominance coefficient ($(cross{:}\,h)_{ij}$), (iii) and an interaction term
281 between dominance coefficient and mean starting allele frequency ($(h{:}\,af)_{jl}$),. We analyzed
282 10,000 samples (2,000 independent simulation runs for each crossing scheme) (Table 1). For
283 crossing schemes using only 6 different founder lines (1:1, 1:few), the simulation runs are
284 based on 4 randomly chosen sets of founder lines each (500 simulation runs per set). Again,
285 the set of founder lines does not have a significant effect on secondary E&R success and is
286 thus not included in the final model (LRT full-null model comparison; data not shown). We
287 performed additional analysis with 500 observations (100 samples per crossing scheme) to
288 determine the influence of the number of replicates on the power of secondary E&R (Table
289 1).
290
291 We performed all diagnostic checks required for logistic regression. Absence of collinearity
292 was confirmed by computing the generalized Variance Inflation Factors (37) using function *vif*
293 in R package *car* (v3.0-8 (38)). Model stability was checked with the R function *dfbeta*. For
294 visualization, linear predictors (LP) were back-transformed to success probabilities using the

295 inverse logit transformation: $p_{sucess} = \frac{e^{LP}}{1+e^{LP}}$. 95 % confidence intervals of the fitted values
296 were investigated with the R function *predict.glm*. Significance of single explanatory variables
297 was tested with a Type III Sum of Squares using function *Anova* in R package car (38).
298 Significance of explanatory variables including their modeled interactions was tested with a
299 likelihood ratio test (LRT) comparing the full model with a nested reduced model with the
300 same structure as the full model, but lacking the assessed explanatory variable (and its
301 interactions). Significance is declared at an alpha cut off-off of 5%. We used Nagelkerke's R^2
302 index (39) to calculate the improvement of each model parameter upon the prediction of a
303 null model.
304
305 Data availability
306 Information regarding the accessibility of raw sequence reads, phased haplotypes of the
307 ancestral *D. simulans* haplotypes as well as MimicrEE2 ready text files for the *D. simulans*
308 recombination map (file used in this project: Dsim_recombination_map_LOESS_100kb_1.txt)
309 can be found in (29). MimicrEE2 ready input files of the different experimental designs, the
310 simulated selection regimes, processed simulation results and all scripts that are necessary to
311 reproduce the results are available at SourceForge (link TBD).
312
313
314

315 **Results**

316 We used forward simulations to assess the influence of different experimental and population
317 genetic parameters, more specifically crossing scheme, population size, dominance
318 coefficient, selection strength, and mean starting allele frequency on the success to detect
319 and fine map selection targets in secondary E&R experiments. We used a Cochran-Mantel-
320 Haenszel (CMH) test to identify SNPs rising in allele frequency. The CMH test allows to test
321 for independence of matched categorical data (33), and compares favorably to other
322 statistical methods in reliably identifying possible targets of selection in E&R setups (40). A
323 simulation run was considered successful, if the true target of selection was the SNP with the
324 highest CMH test statistic (success-A). In a second analysis step, we considered simulation
325 runs as successful, if the true target of selection was not more than 100 SNPs away from the
326 SNP with the highest CMH test statistic (Figure 2; success-B). We used logistic regression to
327 assess which experimental and population genetic parameters have a significant effect on the
328 success of secondary E&R.

329

330 First, we evaluated two crossing schemes that can be easily implemented in empirical studies,
331 more specifically the 1:1 (Figure 1A) and dil:mt (Figure 1E) crossing schemes. The 1:1 crossing
332 scheme is solely based on founder lines (Figure 1A). Using information about the selected
333 haplotype from a primary E&R study, it is possible to determine which founder lines carry a
334 selection target (12), which requires them to be sequenced. Crossing a focal founder line with
335 the selected haplotype with a non-focal line without known selection targets offers the
336 advantage of reducing the number of selection targets dramatically. Using different non-focal
337 lines without known strong selection targets in each replicate reduces the potential of
338 confounding effects of unidentified selection targets in these lines. On the other hand, dil:mt

339 (Figure 1E) represents an entirely different approach. Dil:mt, which has multiple targets of
340 selection at different frequencies, is probably the simulated crossing scheme with the most
341 straight forward implementation (12, 27). It is based on a "dilution" approach, where evolved
342 individuals from a primary E&R experiment are crossed to the ancestral founder population
343 from the same primary E&R (27). In contrast to 1:1, a dil:mt crossing scheme requires both –
344 ancestral and evolved – populations of the primary E&R, but relatively limited information
345 about selection targets on individual founder lines.
346
347 To evaluate the 1:1 and dil:mt crossing scheme, we simulated secondary E&R consisting of 5
348 replicated populations with constant population size (300; 1,200; 4,800, or 19,200 individuals
349 per replicate) that evolve for 60 generations, and used logistic regression to assess the effects
350 of model parameters on secondary E&R success (Model 1, *Assessment of experimental and*
351 *population genetic parameters* in Material & Methods). While selection strength, dominance
352 coefficient, and the interaction between dominance coefficient and mean starting allele
353 frequency all have a significant effect on the power of secondary E&R (Table S2 (success-A),
354 Table S3 (success-B)), our analysis reveals that crossing scheme and population size have by
355 far the strongest influence on secondary E&R success. Both the crossing scheme (LRT full-null
356 model comparison (Model 1): $\chi^2 = 3510.1$, df = 8, p<0.001 (success-A); $\chi^2 = 3211.3$, df = 8,
357 p<0.001 (success-B)); and the population size (LRT full-null model comparison (Model 1): $\chi^2 =$
358 2633.4, df = 12, p<0.001 (success-A); $\chi^2 = 2142.9$, df = 12, p<0.001 (success-B)) have a
359 significant effect on the success of secondary E&R, and are the only parameters with a
360 Nagelkerke's R^2 index above 0.2 (Table 2).
361

362 While a 1:1 crossing scheme results in higher power values than the dil:mt crossing scheme
363 independently of the population size, the difference to the dil:mt crossing scheme is more
364 pronounced in bigger experimental populations (Figure 3). This does not hold only for the
365 power, but also for the resolution (= the distance in SNPs between the true target of selection,
366 and the SNP with the highest CMH test statistic, Figure 4). A potential disadvantage of the 1:1
367 crossing scheme is that the low number of different founder lines (n=6) causes more linkage
368 disequilibrium, indicated by the number of neighboring SNPs with identical test statistics
369 (=ties) and thus broadens signatures compared to dil:mt crossing scheme that has more
370 founder lines (n=100) (Figure S1). However, this is outweighed by superior power of the 1:1
371 crossing scheme at every population size investigated (Figure 3, Figure 4).

372

373

374 Table 2 Nagelkerke's R^2-index of each simulation parameter including its interaction in the logistic regression (2
375 crossing schemes (1:1; dil:mt), 4 population sizes (300; 1,200; 4,800; 19,200 individuals); Figure 3). If the reduced
376 model explains the data as well as a full model that includes the evaluated effect, Nagelkerke's R^2-index is 0.

effect	**R^2 (success-A)**	**R^2 (success-B)**
crossing scheme	0.305	0.264
dominance coefficient	0.077	0.069
selection coefficient	0.007	0.007
mean starting allele frequency	0.003	0.002
population size	0.204	0.224

377

378

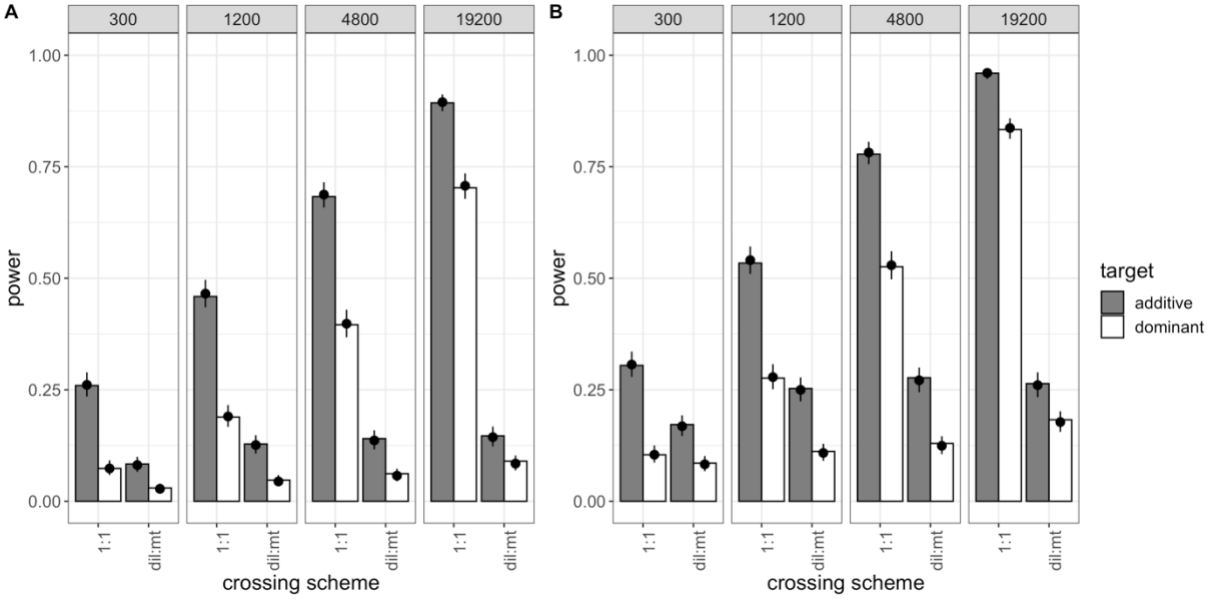


Figure 3 Power of the 1:1 and dil:mt crossing scheme at different population sizes (2,000 simulations/experimental design). Bars show the power (*i.e.,* the proportion of successful simulations) separately for each combination of crossing scheme (1:1, dil:mt), population size (300; 1,200; 4,800; 19,200 individuals), and dominance coefficient (additive in grey, 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 tests).

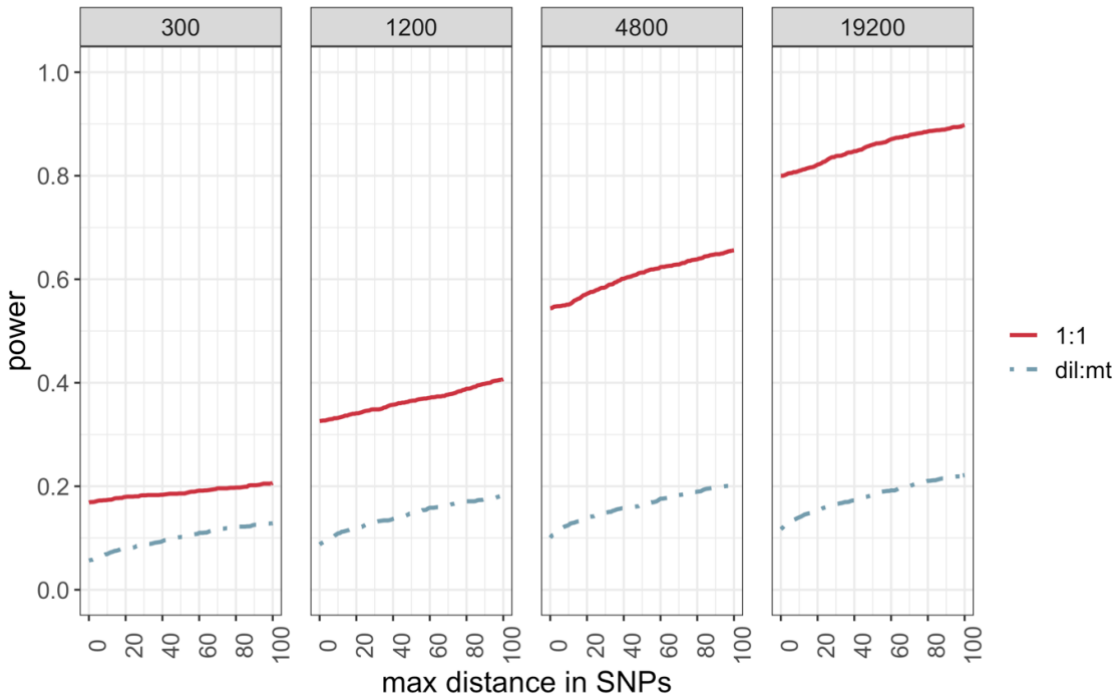


Figure 4 Resolution of the 1:1 and dil:mt crossing scheme at different population sizes. Proportion of simulations (y-axis) that do not exceed a maximum distance in SNPs (x-axis) between the SNP with the highest Cochran-Mantel-Haenszel test statistic and the true target of selection. Each panel shows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

395 Experimental duration

396 Our analysis has shown that after 60 generations of adaptation, only experimental designs
397 with a population size of at least 1,200 individuals exceed power values of 50 % (Figure 3).
398 The 1:1 crossing scheme clearly outperforms dil:mt regardless of the population size, and
399 reaches power above 75% only with a population size of at least 4,800 individuals. However,
400 the maintenance of 5 replicates for 60 generations is labor intensive and can be quite time
401 consuming for most sexual organisms. We explored the possibility that shorter experiments
402 with less than 60 generations may already be sufficient to achieve satisfactory power values.
403 We thus reanalyzed the power of the 1:1, and dil:mt crossing scheme after 20 generations of
404 adaptation (Table 1).

405

406 Consistent with previous computer simulation studies (23–25) and empirical results (41), we
407 observe reduced power for experimental designs with shorter experimental duration (Figure
408 S2). As observed after 60 generations, the crossing scheme (LRT full-null model comparison
409 (Model 1): $\chi^2 = 2672.7$, df = 8, p<0.001 (success-A); $\chi^2 = 2234.4$, df = 8, p<0.001 (success-
410 B)) and the population size (LRT full-null model comparison (Model 1): $\chi^2 = 3176.4$, df = 12,
411 p<0.001 (success-A); $\chi^2 = 3375.3$, df = 12, p<0.001 (success-B)) have the biggest effects on
412 the success of secondary E&R (Table S4 (success-A), Table S5 (success-B), Table S6). In contrast
413 to our analysis after 60 generations of adaptation, the dil:mt crossing scheme results in higher
414 power than the 1:1 crossing scheme for populations with the smallest simulated population
415 size (300 individuals) after 20 generations of adaptation (Figure S2). A possible explanation
416 for this is that the dil:mt crossing scheme has multiple beneficial targets. Because we simulate
417 additive selection, linked selection targets can act synergistically and increase the frequency
418 of the focal SNP over shorter time scales. This phenomenon of pronounced allele frequency

419 increase due to linked selection will be especially important if the population size is small (*i.e.,*
420 drift is not neglectable), and the experimental duration is short. This is reflected in a higher
421 Nagelkerke's R^2 index for population size in secondary E&R with an experimental duration of
422 20 generations compared to 60 generations of adaptation (Table S6, Table 2). With increasing
423 population size and thus, reduced genetic drift, the 1:1 crossing scheme outcompetes dil:mt,
424 similar to our analysis after 60 generations (Figure S2, Figure S3). Although shorter
425 experimental duration results in reduced power, our analysis after 20 generations of
426 adaptation highlights that if the maintenance of experimental populations for many
427 generations is not feasible, shorter experiments can achieve similar power if they are
428 maintained at larger population sizes (Figure 3, Figure S2).
429
430 Alternative crossing schemes
431 Since large population sizes can be challenging to maintain, we also evaluated additional
432 crossing schemes which may perform better even for smaller population sizes (Model 2,
433 *Assessment of experimental and population genetic parameters* in Material & Methods).
434 1:few (Figure 1B) is a modification of the 1:1 crossing scheme, which combines all non-focal
435 lines in each replicate - this provides a consistent genetic composition in each replicate. The
436 1:many crossing scheme (Figure 1C) uses 1 focal, and 99 non-focal lines without selection
437 target. We reasoned that the larger number of segregating variants (970,466 SNPs compared
438 to max 527,771 SNPs in 1:1 crossing scheme, Table S1) potentially provides a higher mapping
439 resolution. Finally, we modified the dilution crossing scheme by simulating only one selection
440 target in the founder population (dil:st, Figure 1D). Hence, no additional selection targets can
441 create potentially confounding selection signatures. For each experimental design, we

442 simulated a population size of 300 individuals, 60 generations of adaptation, and 5 replicates
443 (Table 1).
444 Regardless of the experimental design, the simulated focal selection targets experienced a
445 pronounced allele frequency increase (Figure S4). In “dilution” crossing schemes (dil:mt,
446 dil:st; Figure 1D-E) the frequency trajectories of the focal SNPs were highly variable because
447 of the heterogeneous starting frequency. The crossing schemes where a single focal line with
448 a starting frequency of 50 % carries the beneficial allele typically had a superior performance
449 than dilution crossing schemes (Figure 5). Notably, the significant influence of crossing
450 scheme on secondary E&R success (Table S7 (success-A), Table S8 (success-B), Table S9)
451 cannot be explained by a loss of the focal SNP in dilution crossing schemes, which occurred in
452 less than 1 % of the simulations.
453
454 We found that the dil:st crossing scheme is still inferior to the 1:1 crossing scheme for additive
455 loci (Figure 5). Also, the modifications of the 1:1 crossing scheme (1:few, 1:many) do not
456 provide a substantial improvement (Figure 5). Surprisingly, the 1:many crossing scheme
457 performed very poorly. This may be at least partly attributed to the larger number of SNPs/kb,
458 which will affect the SNP-based success rate (success-B). Given that the three additional
459 crossing schemes do not provide a clear advantage (Figure 5, Figure 6), and are not easier to
460 execute experimentally, we did not evaluate them with larger population sizes.
461

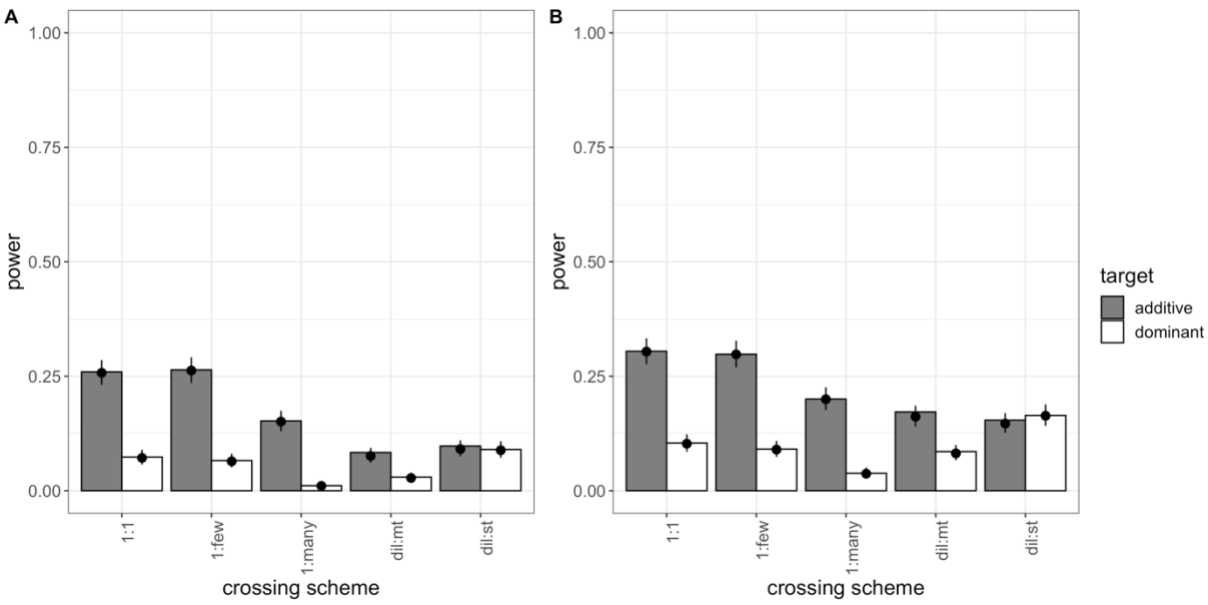


Figure 5 Power of five different crossing schemes (population size = 300 individuals; 2,000 simulations/experimental design). Bars show the power (*i.e.,* proportion of successful simulations) separately for each combination of crossing scheme and dominance coefficient (additive in grey; dominant in white). The dots with error bars display the estimate from the fitted model 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).

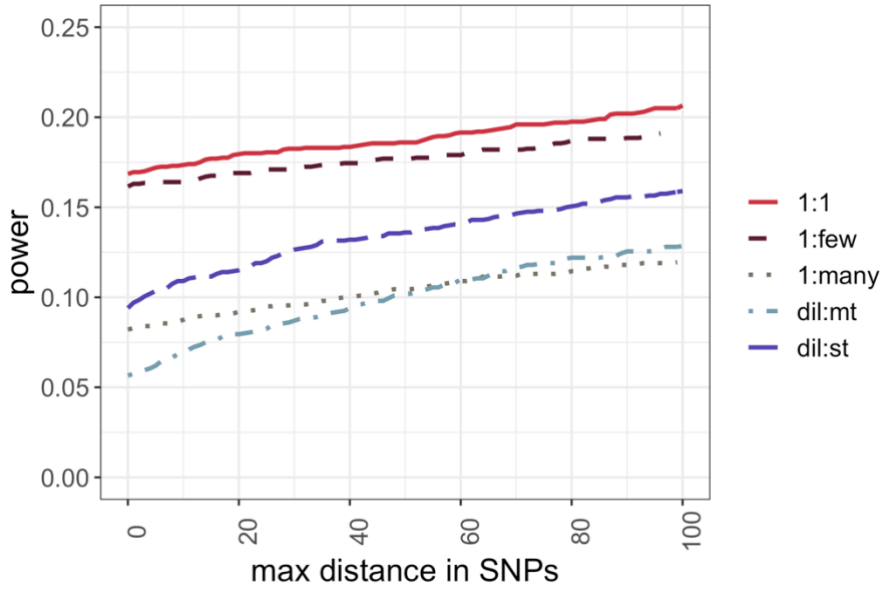


Figure 6 Resolution of five different crossing schemes. Proportion of simulations (y-axis) that do not exceed a maximum distance in SNPs (x-axis) between the SNP with the highest Cochran-Mantel-Haenszel test statistic and the true target of selection.

480 Number of replicates

481 Given the superior performance of experimental designs with larger population sizes (Figure
482 3), we were also interested whether more replicates with a smaller population size may
483 provide further improvements. We performed additional simulations for all five different
484 crossing schemes with a population size of 300 individuals per replicate, and 30 replicates
485 (Table 1). Consistent with other studies (23, 25), more replicates result in increased power
486 (Figure S5). The power of secondary E&R with 30 replicates (Table S10 (success-A), Table S11
487 (success-B)), is only significantly affected by the crossing scheme (LRT full-null model
488 comparison (Model 2): $\chi^2 = 37.744$, df = 8, p<0.001 (success-A); $\chi^2 = 16.692$, df = 8, p=0.033
489 (success-B); Table S12). The 1:1 crossing scheme with 30 replicates also has the highest
490 resolution (Figure S6A). As expected, the impact of linkage disequilibrium, indicated by the
491 number of ties, is dramatically reduced for all crossing schemes when more replicates are
492 simulated (Figure S6B compared to Figure S7). However, similar improvements as seen with
493 30 replicates can be achieved if 5 replicates of the 1:1 crossing scheme are used with a
494 population size of 1,200 individuals (Figure 3). Thus, the same power can be achieved while
495 maintaining about 33 % fewer individuals.

496

497 Dominance coefficient, Selection coefficient, and mean starting allele frequency

498 As expected, dominance coefficient (h), selection coefficient (s), and average starting allele
499 frequency (af) have a significant effect on the success of secondary E&R in all simulation
500 scenarios, except for data with 30 replicates. We observed that selection targets with a higher
501 selection coefficient and/or a higher mean starting allele frequency are easier to fine map.
502 The lack of significant effects of h, s, and af on secondary E&R success for the dataset with 30
503 replicates can probably be explained by the smaller number of conducted simulations (100

504 for 30 replicates and 2,000 for 5 replicates see Table 1). After crossing scheme and population
505 size, the dominance coefficient is the parameter with the highest Nagelkerke's R^2 index in our
506 analysis, with additive loci being easier to fine map. This is caused by the fact that
507 heterozygotes and target homozygotes with a dominant beneficial SNP have the same fitness,
508 resulting in less efficient selection that becomes especially apparent if the selected allele has
509 already reached a high frequency in the population, and non-selected allele homozygotes
510 become rare (indicated by a significant negative effect of the interaction term between
511 dominance coefficient and mean starting allele frequency in most of our statistical models).
512
513 **Discussion**
514 This work was inspired by the difficulty of most E&R studies with sexually reproducing
515 organisms to pinpoint selection targets, mostly due to numerous neutral hitchhikers resulting
516 in large haplotype blocks. Secondary E&R – a follow-up EE validating putative selection targets
517 of a primary E&R under an identical selection regime – has been recently suggested as
518 experimental approach for selection target confirmation (27). We used extensive computer
519 simulations to evaluate how experimental and population genetics parameter shape the
520 power and resolution of secondary E&R. As expected, dominance coefficient, selection
521 strength, and mean starting allele frequency of the selected target all have a significant effect
522 on the success of E&R, where dominant selection targets at high frequency are particularly
523 challenging to detect. However, population size and crossing scheme emerged as the most
524 influential parameters in our analysis.
525
526
527

528 The crossing scheme has a pronounced effect on secondary E&R success

529 We show that a simple crossing scheme, which only requires that (a subset of) the founder
530 lines are sequenced and that founder lines with and without the focal adaptive allele can be
531 distinguished, has the best power and resolution of the five crossing schemes tested. The 1:1
532 crossing scheme is particularly well-suited when many selection targets are detected in the
533 evolved populations of the primary E&R, because it uses only lines from the non-adapted
534 founder population. Ancestral haplotypes from the non-adapted founder population will
535 harbor on average less selection targets than evolved haplotypes that can acquire multiple
536 selection targets through recombination events during the primary E&R (42). Genotypes
537 harboring multiple selection targets make it challenging to fine-map single selection targets,
538 because selected alleles will not propagate independently from each other. Nevertheless,
539 highly polygenic architectures where selected haplotype blocks contain multiple, weakly
540 selected alleles are not well-suited for secondary E&R, because recombination will eventually
541 separate linked favorable alleles and no strongly selected allele can be identified.

542

543 Sufficient E&R power requires large population sizes

544 Our analyses also indicate that rather large population sizes are crucial for being able to
545 identify the causative variant with sufficient confidence using a secondary E&R approach. This
546 observation explains why an empirical secondary E&R study in *D. simulans* with a population
547 size of 1,250 flies per replicated population failed to narrow down a pronounced candidate
548 region from a primary E&R in a secondary E&R after 30 generations, and only confirmed the
549 presence of selected alleles (27).

550

551 Our results suggest that large population sizes become especially crucial for short-term
552 secondary E&R. The previously reported attenuated power of short-term E&R due to reduced
553 experimental duration (25) can be increased by a higher experimental population size. Our
554 analyses show that to achieve satisfactory power after only 20 generations requires
555 population sizes of tens of thousands of individuals– dimensions that are rather rare in
556 current EE designs with strictly sexually reproducing organisms, where population sizes are
557 mostly limited by the capacities of the lab. Similar to previous simulation studies (23, 25), we
558 demonstrated that increasing the number of replicates results in more powerful secondary
559 E&R. If maintaining large population sizes is not feasible, maintaining many replicates at small
560 population size can help to boost secondary E&R performance. Additionally, the loss of one
561 replicate in a highly replicated setup has a smaller impact than in an experimental setup with
562 very few replicates. However, we show that additional replication cannot completely
563 compensate the advantage of large population sizes in secondary E&R.
564
565 Choice of model organism and selection regime
566 Our simulations are parameterized for *D. simulans* because this species is better suited for
567 E&R studies than *D. melanogaster* (22). While other species, such as *S. cerevisiae* (7) or *C.*
568 *remanei* (43) have been used to study adaptive response from standing genetic variation in
569 outcrossing species, *Drosophila* is the most popular organism. Independent of the actual
570 experimental parameters used, we anticipate that our results can be generalized to other
571 species, but the availability of inbred founder lines is probably a more severe restriction,
572 which limits the wide-spread use of secondary E&R.
573

Finally, we simulated a rather simple selection scenario of directional selection, not considering more complex scenarios that include pleiotropy and/or epistasis. While 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. While this reduces the power of the 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.

Acknowledgements

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. This work was supported by the Austrian Science Funds (FWF, W1225, P-32935), and by the European Research Council (ERC) grant “ArchAdapt”.

Authors Contribution

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.

603 **References**

604 1. T. J. Kawecki, *et al.*, Experimental evolution. *Trends Ecol. Evol.* **27**, 547–560 (2012).

605 2. T. Garland, M. Rose, *Experimental Evolution: Concepts, Methods, and Applications of*
606 *Selection Experiments* (2009).

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

610 4. T. L. Turner, A. D. Stewart, A. T. Fields, W. R. Rice, A. M. Tarone, Population-Based
611 Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of
612 Body Size Variation in Drosophila melanogaster. *PLoS Genet.* **7**, e1001336 (2011).

613 5. A. Long, G. Liti, A. Luptak, O. Tenaillon, Elucidating the molecular architecture of
614 adaptation via evolve and resequence experiments. *Nat. Publ. Gr.* **16**, 567–582
615 (2015).

616 6. A. M. Johansson, M. E. Pettersson, P. B. Siegel, Ö. Carlborg, Genome-Wide Effects of
617 Long-Term Divergent Selection. *PLOS Genet.* **6**, e1001188 (2010).

618 7. M. K. Burke, G. Liti, A. D. Long, Standing Genetic Variation Drives Repeatable
619 Experimental Evolution in Outcrossing Populations of Saccharomyces cerevisiae. *Mol.*
620 *Biol. Evol.* **31**, 3228–3239 (2014).

621 8. A. Papkou, *et al.*, The genomic basis of Red Queen dynamics during rapid reciprocal
622 host–pathogen coevolution. *Proc. Natl. Acad. Sci.* **116**, 923–928 (2019).

623 9. N. E. Martins, *et al.*, Host adaptation to viruses relies on few genes with different
624 cross-resistance properties. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5938–43 (2014).

625 10. H. Teotónio, I. M. Chelo, M. Bradić, M. R. Rose, A. D. Long, Experimental evolution
626 reveals natural selection on standing genetic variation. *Nat. Genet.* **41**, 251–257

(2009).

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

12. N. Barghi, *et al.*, Genetic redundancy fuels polygenic adaptation in Drosophila. *PLOS Biol.* **17**, 1–31 (2019).

13. R. Tobler, *et al.*, Massive Habitat-Specific Genomic Response in *D. melanogaster* Populations during Experimental Evolution in Hot and Cold Environments. *Mol. Biol. Evol.* **31**, 364–375 (2014).

14. S. U. Franssen, V. Nolte, R. Tobler, C. Schlötterer, Patterns of Linkage Disequilibrium and Long Range Hitchhiking in Evolving Experimental *Drosophila melanogaster* Populations. *Mol. Biol. Evol.* **32**, 495–509 (2015).

15. S. V Nuzhdin, T. L. Turner, Promises and limitations of hitchhiking mapping. *Curr. Opin. Genet. Dev.* **23**, 694–699 (2013).

16. A. Iranmehr, A. Akbari, C. Schlötterer, V. Bafna, CLEAR : Composition of Likelihoods for Evolve and. *Genetics* **206**, 1011–1023 (2017).

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

18. J. K. Kelly, K. A. Hughes, Pervasive Linked Selection and Intermediate-Frequency Alleles Are Implicated in an Evolve-and-Resequencing Experiment of Drosophila simulans . *Genetics* **211**, genetics.301824.2018 (2018).

19. K. Spitzer, M. Pelizzola, A. Futschik, Modifying the Chi-square and the CMH test for population genetic inference: Adapting to overdispersion. *Ann. Appl. Stat.* **14**, 202–

651 220 (2020).

652 20. G. I. Lang, *et al.*, Pervasive genetic hitchhiking and clonal interference in forty
653 evolving yeast populations. *Nature* **500**, 571–574 (2013).

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

656 22. N. Barghi, R. Tobler, V. Nolte, C. Schlötterer, *Drosophila simulans* : A Species with
657 Improved Resolution in Evolve and Resequence Studies. *G3:*
658 *Genes|Genomes|Genetics* **7**, 2337–2343 (2017).

659 23. D. Kessner, J. Novembre, Power analysis of artificial selection experiments using
660 efficient whole genome simulation of quantitative traits. *Genetics* **199**, 991–1005
661 (2015).

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

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

667 26. C. Vlachos, R. Kofler, Optimizing the Power to Identify the Genetic Basis of Complex
668 Traits with Evolve and Resequence Studies. *Mol. Biol. Evol.* **36**, 2890–2905 (2019).

669 27. C. Burny, *et al.*, Secondary Evolve and Resequencing: An Experimental Confirmation
670 of Putative Selection Targets without Phenotyping. *Genome Biol. Evol.* **12**, 151–159
671 (2020).

672 28. S. J. Gratz, *et al.*, Genome engineering of Drosophila with the CRISPR RNA-guided
673 Cas9 nuclease. *Genetics* **194**, 1029–1035 (2013).

674 29. J. M. Howie, R. Mazzucco, T. Taus, V. Nolte, C. Schlötterer, DNA motifs are not general

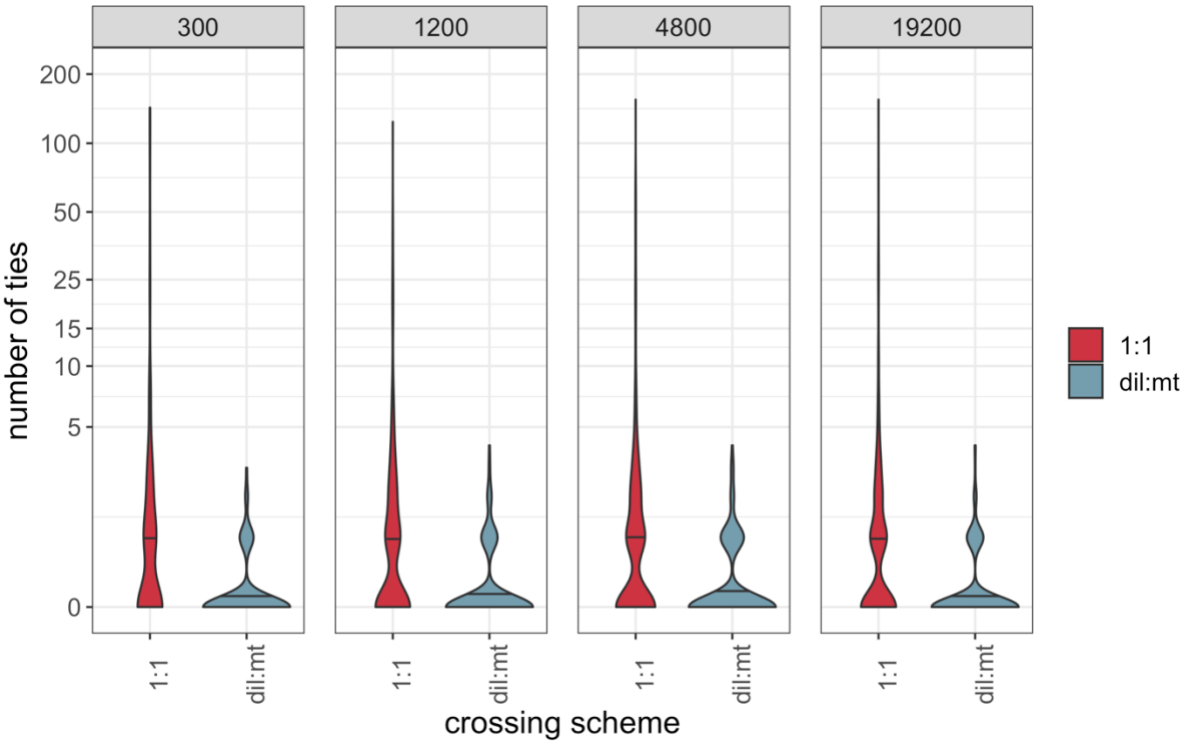
675 predictors of recombination in two drosophila sister species. *Genome Biol. Evol.* **11**,
676 1345–1357 (2019).
677 30. C. Vlachos, R. Kofler, MimicrEE2: Genome-wide forward simulations of Evolve and
678 Resequencing studies. *PLOS Comput. Biol.* **14**, e1006413 (2018).
679 31. R Core Team 3.5.3, A Language and Environment for Statistical Computing. *R Found.*
680 *Stat. Comput.* **2**, https://www.R-project.org (2019).
681 32. T. Taus, A. Futschik, C. Schlötterer, Quantifying selection with pool-seq time series
682 data. *Mol. Biol. Evol.* **34**, 3023–3034 (2017).
683 33. A. Agresti, M. Kateri, "Categorical Data Analysis" in *International Encyclopedia of*
684 *Statistical Science*, M. Lovric, Ed. (Springer Berlin Heidelberg, 2011), pp. 206–208.
685 34. M. Dowle, A. Srinivasan, data.table: Extension of `data.frame` (2019).
686 35. R. H. Baayen, *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*
687 (Cambridge University Press, 2008) https:/doi.org/10.1017/CBO9780511801686.
688 36. H. Schielzeth, Simple means to improve the interpretability of regression coefficients.
689 *Methods Ecol. Evol.* **1**, 103–113 (2010).
690 37. J. Fox, G. Monette, Generalized Collinearity Diagnostics. *J. Am. Stat. Assoc.* **87**, 178–
691 183 (1992).
692 38. J. Fox, S. Weisberg, *An R Companion to Applied Regression*, Third (Sage, 2019).
693 39. N. Nagelkerke, A note on a general definition of the coefficient of determination.
694 *Biometrika* **78**, 691–692 (1991).
695 40. C. Vlachos, *et al.*, Benchmarking software tools for detecting and quantifying
696 selection in evolve and resequencing studies. *Genome Biol.* **20**, 169 (2019).
697 41. A. M. Langmüller, C. Schlötterer, Low concordance of short-term and long-term
698 selection responses in experimental Drosophila populations. *Mol. Ecol.* **29**, 3466–

3475 (2020).

42. K. A. Otte, C. Schlötterer, Detecting selected haplotype blocks in evolve and resequence experiments. *Mol. Ecol. Resour.* **21**, 93–109 (2021).

43. H. Teotónio, S. Estes, P. C. Phillips, C. F. Baer, Experimental evolution with Caenorhabditis nematodes. *Genetics* **206**, 691–716 (2017).

706 **Supplementary Figures**



707

708 Figure S1 Violin plots of the number of ties (y axis = $\log_{10}$ (number of ties + 1)) for simulations where the true
709 target of selection has the highest Cochran-Mantel-Haenszel (CMH) test statistic (success-A). The black
710 horizontal lines in the violin plots display the median number of observed ties. Each panel shows the results for
711 one particular simulated population size. Ties are defined as neighboring SNPs that have the same CMH test
712 statistic as the target of selection.

713

714

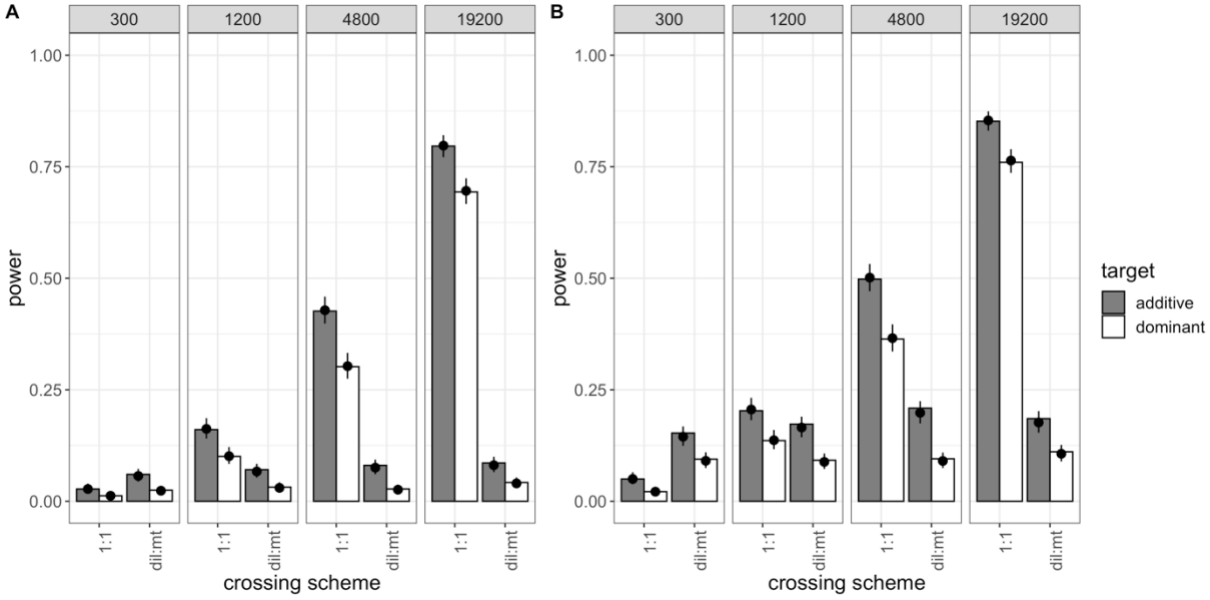


Figure S2 Power of the 1:1 and dil:mt crossing scheme at different population sizes (2,000 simulations/experimental design) after 20 generations of adaptation. Bars show the power (*i.e.,* proportion of successful simulations) separately for each combination of crossing scheme (1:1, dil:mt), population size (300; 1,200; 4,800; 19,200 individuals), and dominance coefficient (additive in grey, dominant in white). The dots with error bars display the model fit 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).

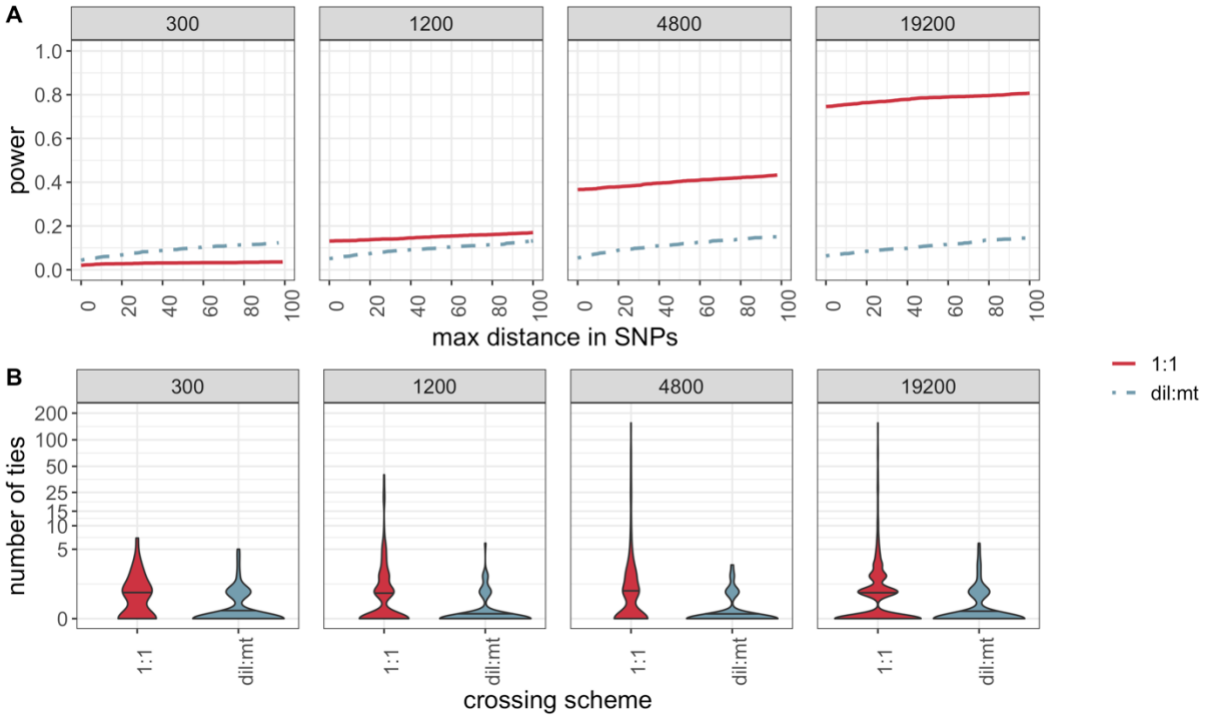


Figure S3 Resolution for the 1:1 and dil:mt crossing scheme simulated with different population sizes after 20 generations of adaptation. (A) Proportion of simulations (y-axis) that do not exceed a maximum distance in SNPs (x-axis) between the SNP with the highest Cochran-Mantel-Haenszel (CMH) test statistic and the true target of selection (B) Violin plots of the number of ties (y axis = $\log_{10}$ (number of ties + 1)) for simulations where the true target of selection has the highest CMH test statistic (success-A). The black horizontal lines in the violin plots display the median number of observed ties. Ties are defined as neighboring SNPs that have the same CMH test statistic as the target of selection.

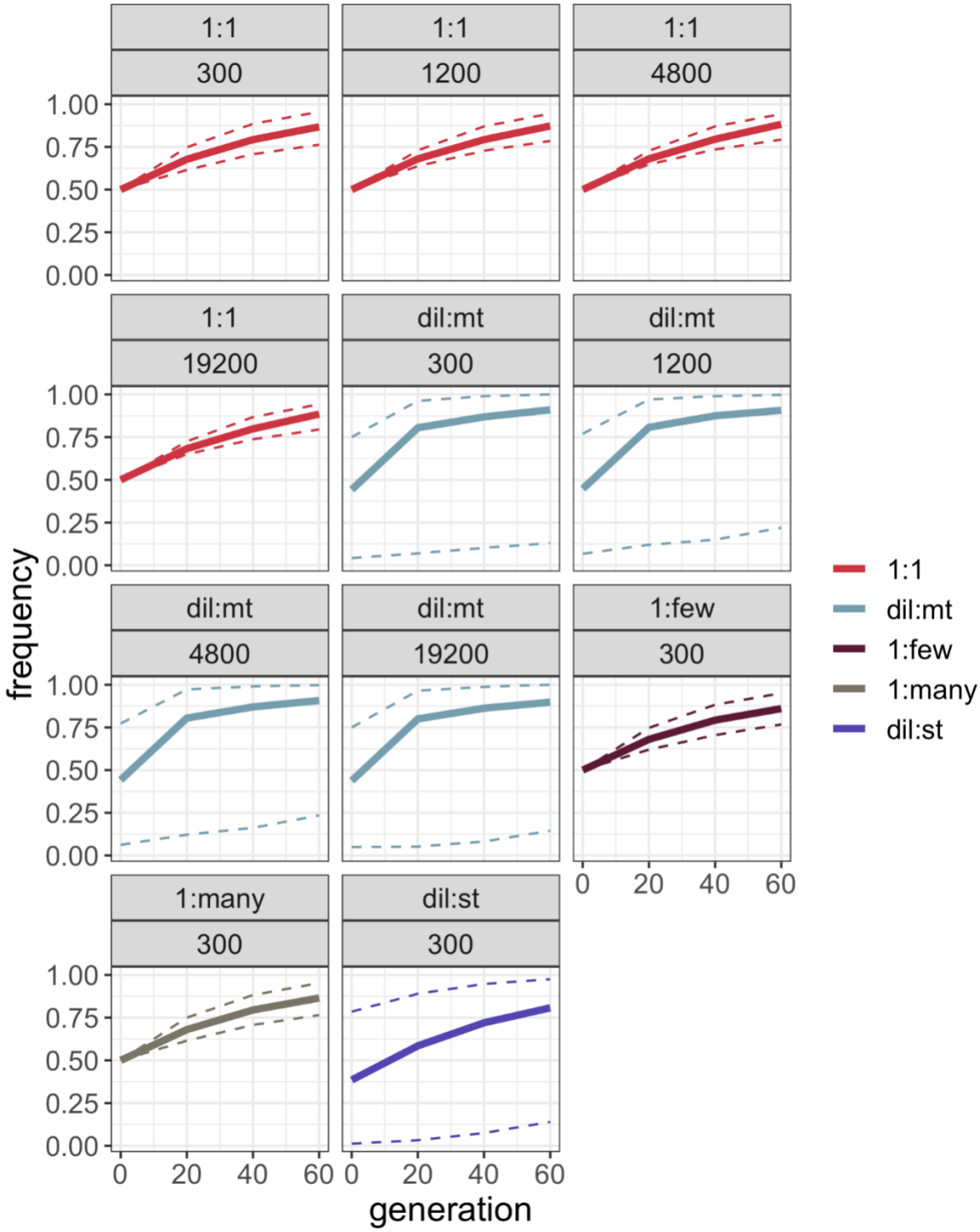


Figure S4 Median allele frequency trajectory of focal targets in different experimental designs. For each single simulation, the focal target frequency of a distinct generation was averaged over five replicates. Solid lines show the median allele frequency trajectory over 2,000 independent simulations per experimental design. Dashed lines show the 5 and 95 percentiles of the allele frequency trajectories. Each panel shows the median allele frequency trajectory of one particular crossing scheme (1:1, dil:mt, 1:few, 1:many, dil:st) and population size (300; 1,200; 4,800; 19,200 individuals) combination.

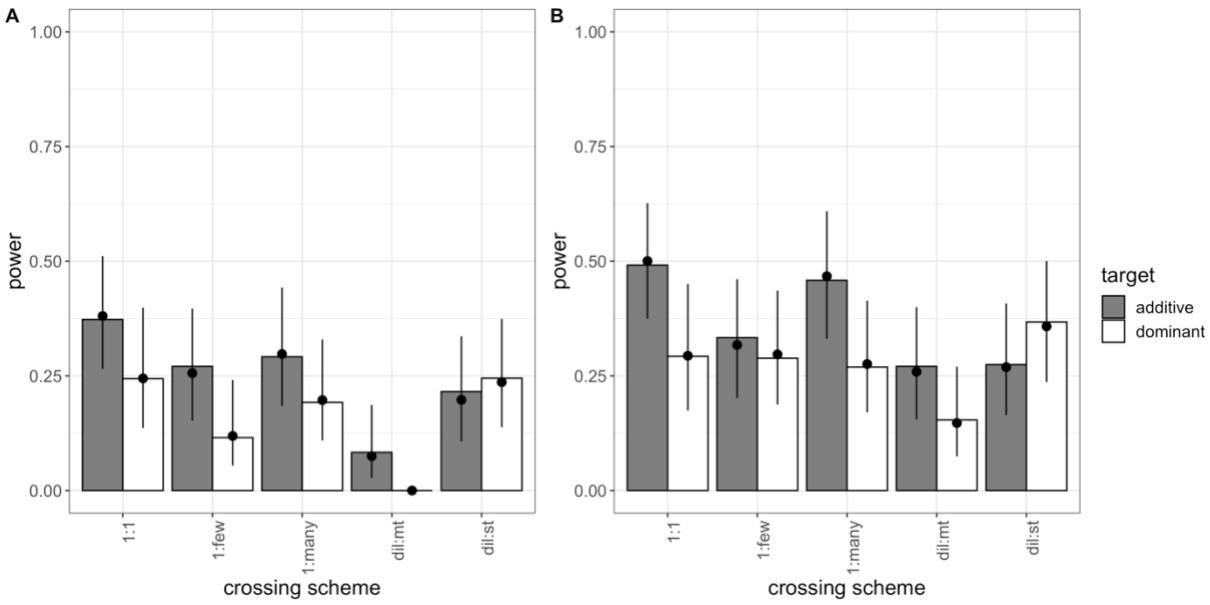


Figure S5 Power of five different crossing schemes (population size = 300 individuals; 30 replicates; 100 simulations/experimental design). Bars show the power (*i.e.* proportion of successful simulations) separately for each combination of crossing scheme and dominance coefficient (additive in grey; dominant in white). The dots with error bars display the estimate from the fitted model 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).

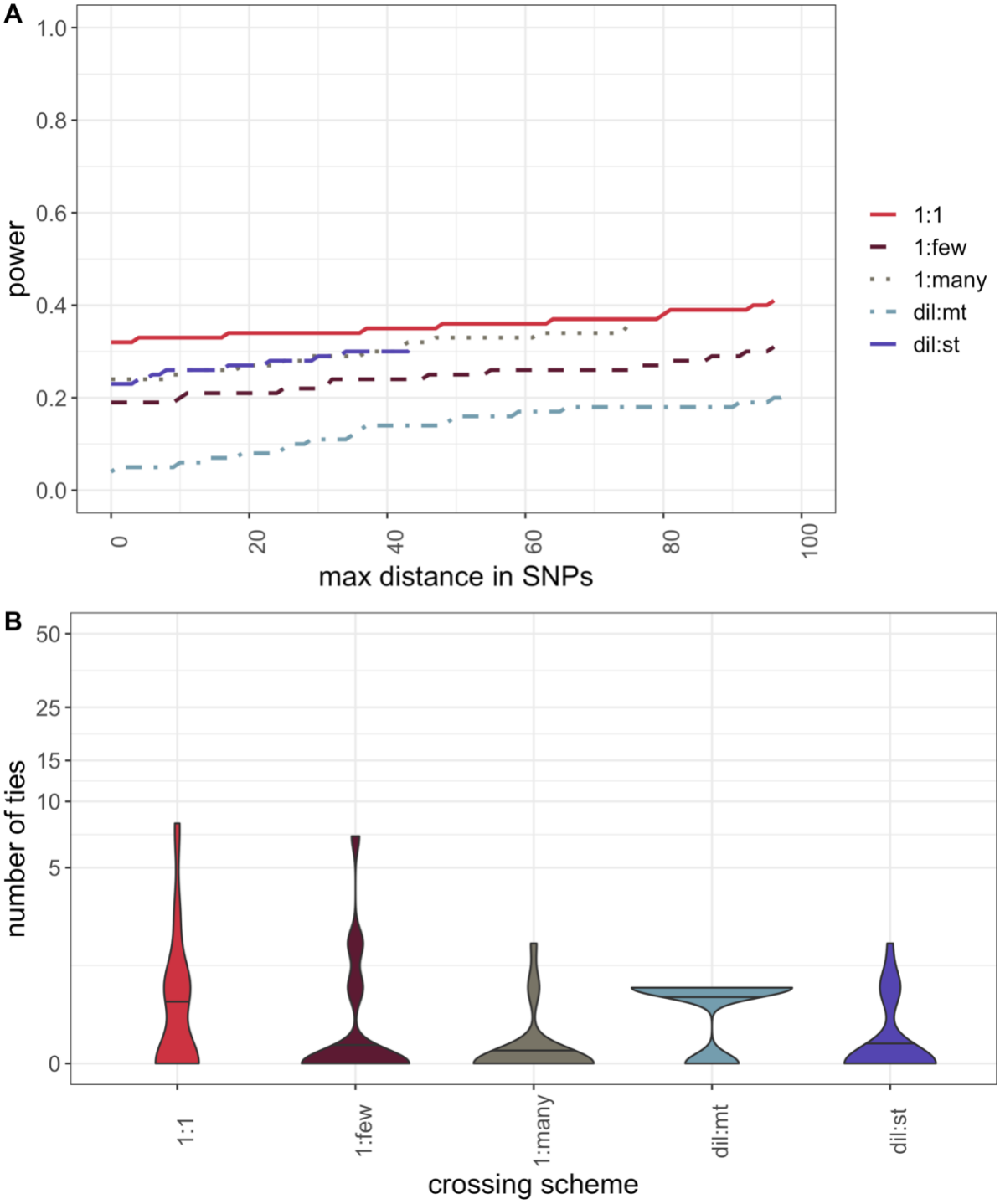


Figure S6 Resolution for five different crossing schemes (population size = 300 individuals; 30 replicates; 100 simulations/experimental design). (A) Proportion of simulations (y-axis) that do not exceed a maximum distance in SNPs (x-axis) between the SNP with the highest Cochran-Mantel-Haenszel (CMH) test statistic and the true target of selection (B) Violin plots of the number of ties (y axis = $\log_{10}$ (number of ties + 1)) for simulations where the true target of selection has the highest CMH test statistic (success-A). The black horizontal lines in the violin plots display the median number of observed ties. Ties are defined as neighboring SNPs that have the same CMH test statistic as the target of selection.

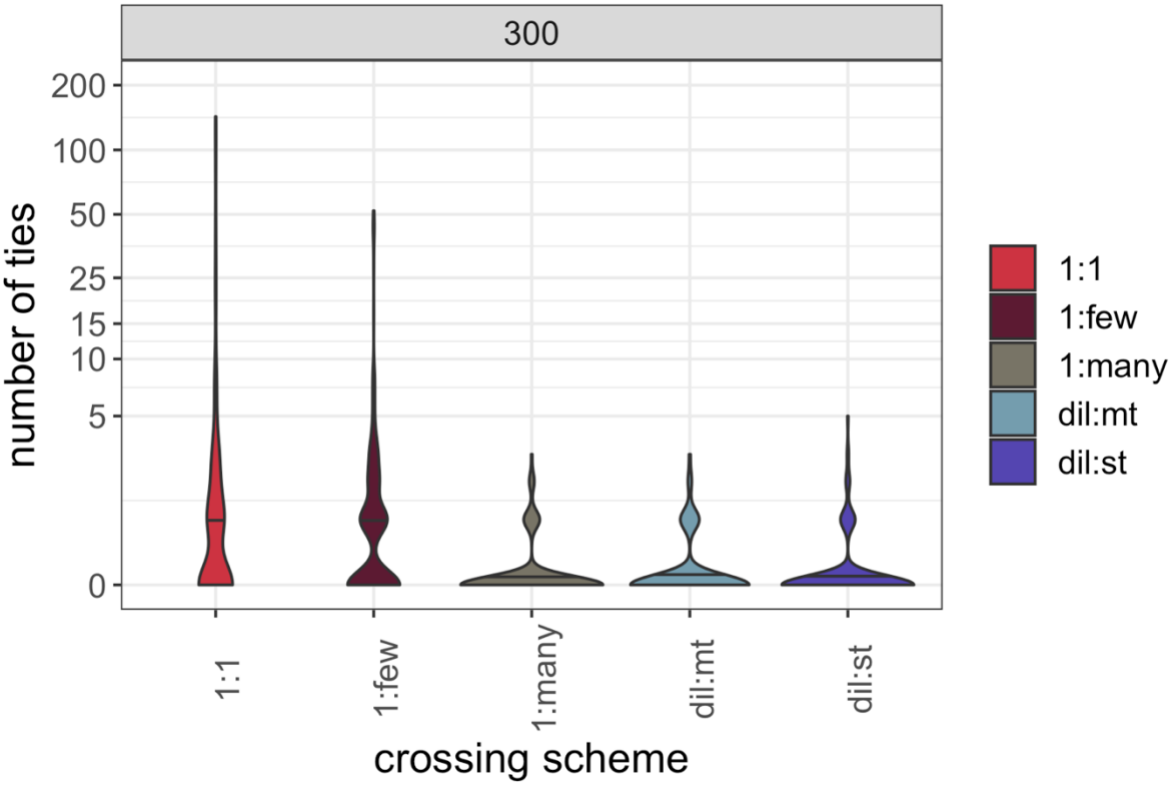


Figure S7 Violin plots of the number of ties (y axis = $\log_{10}$ (number of ties + 1)) for simulations where the true target of selection has the highest Cochran-Mantel-Haenszel (CMH) test statistic (success-A, population size = 300 individuals; 5 replicates; 2,000 simulations/experimental design). The black horizontal lines in the violin plots display the median number of observed ties. Ties are defined as neighboring SNPs that have the same CMH test statistic as the target of selection.

771 **Supplementary Tables**

772

773 Table S1 Number of SNPs per crossing scheme on chromosome-arm 2L.

774

crossing scheme	median	range
1:1	523,008	479,507 – 527,771
1:few	526,510	522,470 – 530,230
1:many	970,466	970,466 – 970,466
dil:mt	970,466	970,466 – 970,566
dil:st	970,466	970,466 – 970,466

775

776

Table S2 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 60 generations of adaptation (success-A; Model 1). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 5 replicates; N= population size; coss:h = interaction term between crossing scheme and dominance coefficient; h:af = interaction term between dominance coefficient and mean starting allele frequency; cross:N = interaction term between crossing scheme and population size ; h:N = interaction term between dominance coefficient and population size; cross:h:N = interaction term between crossing scheme, dominance coefficient, and population size.

term	χ^2	**df**	***P***
cross	117.75	1	<0.001
h	113.81	1	<0.001
s [1]	67.38	1	<0.001
af [1]	0.10	1	0.751
N	1023.53	3	<0.001
cross:h	0.37	1	0.544
h:af	18.51	1	<0.001
cross:N	222.05	3	<0.001
h:N	3.05	3	0.384
cross:h:N	3.78	3	0.286

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value was 0.463/0.130 for the starting allele frequency, and 0.086/0.009 for the selection coefficient respectively

Table S3 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 60 generations of adaptation (success-B; Model 1). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 5 replicates; N= population size; coss:h = interaction term between crossing scheme and dominance coefficient; h:af = interaction term between dominance coefficient and mean starting allele frequency; cross:N = interaction term between crossing scheme and population size ; h:N = interaction term between dominance coefficient and population size; cross:h:N = interaction term between crossing scheme, dominance coefficient, and population size.

term	χ^2	**df**	***P***
cross	50.81	1	<0.001
h	115.74	1	<0.001
s [1]	73.39	1	<0.001
af [1]	1.95	1	0.162
N	1225.22	3	<0.001
cross:h	4.18	1	0.041
h:af	16.67	1	<0.001
cross:N	426.90	3	<0.001
h:N	6.03	3	0.110
cross:h:N	16.12	3	0.001

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value was 0.463/0.130 for the starting allele frequency, and 0.086/0.009 for the selection coefficient respectively

Table S4 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 20 generations of adaptation (success-A; Model 1). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 5 replicates; N= population size; coss:h = interaction term between crossing scheme and dominance coefficient; h:af = interaction term between dominance coefficient and mean starting allele frequency; cross:N = interaction term between crossing scheme and population size ; h:N = interaction term between dominance coefficient and population size; cross:h:N = interaction term between crossing scheme, dominance coefficient, and population size.

term	χ^2	**df**	***P***
cross	14.79	1	<0.001
h	6.07	1	0.0138
s [1]	18.21	1	<0.001
af [1]	30.24	1	<0.001
N	1714.41	3	<0.001
cross:h	0.01	1	0.942
h:af	0.02	1	0.898
cross:N	378.43	3	<0.001
h:N	0.64	3	0.887
cross:h:N	1.29	3	0.732

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value was 0.463/0.130 for the starting allele frequency, and 0.086/0.009 for the selection coefficient respectively

Table S5 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 20 generations of adaptation (success-B; Model 1). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 5 replicates; N= population size; coss:h = interaction term between crossing scheme and dominance coefficient; h:af = interaction term between dominance coefficient and mean starting allele frequency; cross:N = interaction term between crossing scheme and population size ; h:N = interaction term between dominance coefficient and population size; cross:h:N = interaction term between crossing scheme, dominance coefficient, and population size.

term	χ^2	**df**	***P***
cross	65.33	1	<0.001
h	11.80	1	<0.001
s [1]	58.96	1	<0.001
af [1]	60.39	1	<0.001
N	1789.63	3	<0.001
cross:h	1.60	1	0.206
h:af	0.36	1	0.551
cross:N	657.22	3	<0.001
h:N	1.88	3	0.598
cross:h:N	4.68	3	0.197

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value was 0.463/0.130 for the starting allele frequency, and 0.086/0.009 for the selection coefficient respectively

Table S6 Nagelkerke's R^2-index of each simulation parameter including its interaction in the logistic regression (Model 1; 2 crossing schemes (1:1, dil:mt), 4 population sizes (300; 1,200; 4,800; 19,200 individuals), 20 generations, 5 replicates; Figure S2). If the reduced model explains the data equally well as a full model that includes the evaluated effect, Nagelkerke's R^2-index is 0.

effect	**R^2 (success-A)**	**R^2 (success-B)**
crossing scheme	0.283	0.295
dominance	0.019	0.021
selection coefficient	0.002	0.007
starting allele frequency	0.005	0.009
population size	0.322	0.211

Table S7 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 60 generations of adaptation (success-A; Model 2). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 5 replicates; cross:h= interaction term between crossing scheme and dominance coefficient; h:af= interaction term between dominance coefficient and mean starting allele frequency.

term	χ^2	**df**	***P***
cross	197.66	4	<0.001
h	115.11	1	<0.001
s [1]	76.20	1	<0.001
af [1]	11.94	1	<0.001
cross:h	77.19	4	<0.001
h:af	7.56	1	0.006

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value were 0.463/0.133 for the starting allele frequency, and 0.085/0.009 for the selection coefficient respectively

Table S8 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 60 generations of adaptation (success-B; Model 2). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 5 replicates; cross:h= interaction term between crossing scheme and dominance coefficient; h:af= interaction term between dominance coefficient and mean starting allele frequency.

term	χ^2	**df**	***P***
cross	97.39	4	<0.001
h	116.20	1	<0.001
s [1]	72.09	1	<0.001
af [1]	19.69	1	<0.001
cross:h	97.2	4	<0.001
h:af	7.07	1	0.008

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value were 0.463/0.133 for the starting allele frequency, and 0.085/0.009 for the selection coefficient respectively

Table S9 Nagelkerke's R^2-index of each simulation parameter including its interaction in the logistic regression (5 crossing schemes, population size = 300 individuals, experimental duration = 60 generations; 5 replicates per simulation; Model 2). If the reduced model explains the data equally well as a full model that includes the evaluated effect, Nagelkerke's R^2-index is 0.

effect	**R^2 (success-A)**	**R^2 (success-B)**
crossing scheme	0.061	0.033
dominance	0.095	0.076
selection coefficient	0.002	0.001
starting allele frequency	0.003	0.004

Table S10 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 60 generations of adaptation (success-A; Model 2). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 30 replicates; cross:h= interaction term between crossing scheme and dominance coefficient; h:af= interaction term between dominance coefficient and mean starting allele frequency.

term	χ^2	**df**	***P***
cross	13.89	4	0.008
h	1.30	1	0.255
s [1]	2.74	1	0.098
af [1]	2.82	1	0.093
cross:h	5.95	4	0.203
h:af	1.72	1	0.189

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value were 0.463/0.133 for the starting allele frequency, and 0.085/0.009 for the selection coefficient, respectively

Table S11 Type III Sum of Squares of explanatory variables for the analysis of success in secondary Evolve and Resequence studies after 60 generations of adaptation (success-B; Model 2). cross= crossing scheme; h= dominance coefficient; s= selection coefficient; af= mean starting allele frequency of the focal target over 30 replicates; cross:h= interaction term between crossing scheme and dominance coefficient; h:af= interaction term between dominance coefficient and mean starting allele frequency.

term	χ^2	**df**	***P***
cross	9.86	4	0.043
h	3.55	1	0.059
s [1]	3.65	1	0.056
af [1]	0.77	1	0.381
cross:h	5.41	4	0.247
h:af	1.01	1	0.315

[1] multiplied by 100, and z-transformed to mean = 0, and sd =1; Mean/standard deviation of the original value were 0.463/0.133 for the starting allele frequency, and 0.085/0.009 for the selection coefficient, respectively

Table S12 Nagelkerke's R^2-index of each simulation parameter including its interaction in the logistic regression (5 crossing schemes, population size = 300 individuals, experimental duration = 60 generations; 30 replicates per simulation; Figure S5; Model 2). If the reduced model explains the data equally well as a full model that includes the evaluated effect, Nagelkerke's R^2-index is 0.

effect	**R^2 (success-A)**	**R^2 (success-B)**
crossing scheme	0.116	0.046
dominance	0.046	0.034
selection coefficient	0.009	0.010
starting allele frequency	0.010	0.003