

# Power Analysis of Artificial Selection Experiments Using Efficient Whole Genome Simulation of Quantitative Traits

Darren Kessner\* and John Novembre<sup>†,1</sup>

\*Bioinformatics Interdepartmental Program, University of California, Los Angeles, California 90095, and <sup>†</sup>Department of Human Genetics, University of Chicago, Chicago, Illinois 60637

**ABSTRACT** Evolve and resequence studies combine artificial selection experiments with massively parallel sequencing technology to study the genetic basis for complex traits. In these experiments, individuals are selected for extreme values of a trait, causing alleles at quantitative trait loci (QTL) to increase or decrease in frequency in the experimental population. We present a new analysis of the power of artificial selection experiments to detect and localize quantitative trait loci. This analysis uses a simulation framework that explicitly models whole genomes of individuals, quantitative traits, and selection based on individual trait values. We find that explicitly modeling QTL provides qualitatively different insights than considering independent loci with constant selection coefficients. Specifically, we observe how interference between QTL under selection affects the trajectories and lengthens the fixation times of selected alleles. We also show that a substantial portion of the genetic variance of the trait (50–100%) can be explained by detected QTL in as little as 20 generations of selection, depending on the trait architecture and experimental design. Furthermore, we show that power depends crucially on the opportunity for recombination during the experiment. Finally, we show that an increase in power is obtained by leveraging founder haplotype information to obtain allele frequency estimates.

**KEYWORDS** artificial selection; evolve and resequence; forward simulation; power analysis; quantitative traits

**S**OME of the first artificial selection experiments, performed in the early 1900s, were designed to settle ongoing debates about the nature of selection. In particular, early researchers hoped to answer questions about whether selection on continuous variation was even possible and how to reconcile this with the Mendelian viewpoint of genes as discrete heritable units (see Falconer and Mackay 1996, Chap. 11, for theoretical background on artificial selection; Falconer 1992 for a history of early selection experiments; and Garland and Rose 2009 for a comprehensive overview of experimental evolution).

More recently, researchers have used artificial selection experiments to study the genetic basis for complex traits by analyzing allele frequency changes in evolved populations.

This technique has been used in a wide variety of organisms, including for example maize (Laurie *et al.* 2004), yeast (Ehrenreich *et al.* 2010), fruit flies (Nuzhdin *et al.* 2007; Teotonio *et al.* 2009), chickens (Johansson *et al.* 2010), and mice (Keightley and Bulfield 1993).

Massively parallel (also known as next-generation) sequencing of pooled samples has enabled researchers to obtain genome-wide allele frequency estimates for a population in a cost-effective manner (Futschik and Schlotterer 2010). These technological advances have led to the development of the evolve and resequence (E&R) method for mapping traits (Burke *et al.* 2010; Parts *et al.* 2011; Turner *et al.* 2011; Orozco-terWengel *et al.* 2012; Remolina *et al.* 2012). In E&R studies, artificial selection is followed by pooled sequencing of genomic DNA from multiple individuals. A site exhibiting a large allele frequency change in the selected population suggests the presence of a nearby quantitative trait locus (QTL) (Figure 1).

Due to limitations on resources and time, researchers will face many design decisions as they set up artificial selection experiments. First, the size of experimental populations affects the degree of genetic drift and the efficacy of selection.

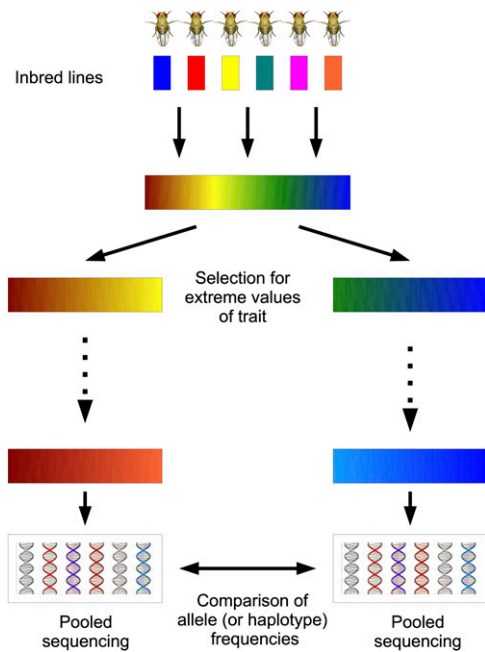
Copyright © 2015 by the Genetics Society of America  
doi: 10.1534/genetics.115.175075

Manuscript received October 13, 2014; accepted for publication February 5, 2015; published Early Online February 10, 2015.

Available freely online through the author-supported open access option.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.175075/-/DC1>.

<sup>1</sup>Corresponding author: University of Chicago, 920 E 58th Street, Chicago, IL 60637.  
E-mail: [jnovembre@uchicago.edu](mailto:jnovembre@uchicago.edu)



**Figure 1** Evolve and resequence experiment. After initial neutral mixing of founders, individuals with extreme values of the trait are selected to create the next generation. After several generations of selection, populations are sequenced and analyzed for allele frequency differences.

Next, the strength of selection is determined by the proportion of individuals selected each generation to create the next generation. In addition, the length of the experiment must be chosen in conjunction with the strength of selection to maximize allele frequency differences at QTL, while at the same time minimizing differences at neutral loci due to drift. For example, extremely strong selection over a large number of generations will result in the fixation of variants across the genome, which will stifle any ability to distinguish QTL from neutral loci. An additional factor to consider is that the selection can be performed on a single population that will be compared to a control population; alternatively, selection can be divergent, where two populations selected in opposite directions will be compared. Finally, replication is a key consideration in any experiment, and especially in experimental evolution, where random genetic drift plays a large role: observation of a large allele frequency difference at a locus in multiple replicate experiments will increase confidence that the difference is due to selection at the locus rather than drift.

E&R experiments lie at the interface between population genetics and quantitative genetics. The selection pressure on a QTL can be parameterized using concepts from quantitative genetics such as effect sizes, genetic variance, and heritability, in addition to experimental design parameters such as the proportion of individuals selected each generation. Furthermore, as we will see, the effect of a QTL allele on an individual's fitness depends strongly on interaction with the other QTL alleles carried by the individual. As noted by Felsenstein (1974, 1987), finite populations necessarily experience random linkage disequilibrium between any two

polymorphic loci. This linkage disequilibrium results in interference, where the effect of selection on each locus is decreased (Hill and Robertson 1966).

Previous studies of the power of artificial selection experiments to detect trait loci have taken the traditional population genetics approach in which selection is parameterized using selection coefficients that remain constant each generation. For example, previous work by Kim and Stephan (1999) analyzed a single locus under a constant selection coefficient, using a diffusion approximation, and two recent simulation studies employed forward simulations with loci parameterized by constant selection coefficients to obtain power estimates (Baldwin-Brown *et al.* 2014; Kofler and Schlotterer 2014). While there are ways to translate truncation selection intensity parameters to selection coefficients (see, for example, Falconer and Mackay 1996, Chap. 11), we seek to complement existing efforts by using an explicit quantitative genetic parameterization and considering a finite genome length.

We introduce a new simulation framework to investigate the power of artificial selection experiments to detect and localize QTL contributing to a quantitative trait. Our simulations employ a whole-genome quantitative genetic model of loci underlying a trait, and we explicitly model artificial selection of individuals each generation based on trait values. As in Baldwin-Brown *et al.* (2014) and Kofler and Schlotterer (2014), we model *Drosophila melanogaster* individuals, and we use experimental parameters similar to those used by Turner and Miller (2012).

A common assumption used in theoretical studies (*e.g.*, Thornton *et al.* 2013) and justified by theoretical arguments (for example, Gillespie 1991, p. 266) is that the fitness effect sizes of newly arising mutations are exponentially distributed. In addition, Orr (1998) showed that the distribution of fitness effect sizes of alleles fixed during adaptation is approximately exponential and that this result is nearly independent of the distribution of mutational effects. In light of these considerations, we follow the rationale of Otto and Jones (2000) and model effect sizes of QTL alleles as exponentially distributed. Additional support for this assumption comes from empirical studies of several quantitative traits whose genetic variation has been shown to depend on a few loci of large effect and many loci of small effect (see Mackay 2010 for a review). Because we assume no relationship between starting allele frequency and effect size, this implicitly assumes that the trait variation in the founder population is neutral or effectively neutral prior to the onset of artificial selection. For traits under stabilizing selection, we might expect variants of large effect to have reduced frequencies. The simulation framework we have developed easily generalizes to such alternative cases, but as a starting point we begin with the simplifying assumption of a neutral trait with exponentially distributed QTL effect sizes.

Our simulations show that forward simulations of a locus assuming a constant selection coefficient do not fully capture the allele frequency dynamics of a QTL under artificial selection on a quantitative trait. In contrast, explicit quantitative genetic

modeling of the trait leads to insights regarding the effect of the trait architecture on the allele frequency trajectory of a QTL. For instance, by simulating the entire genome of individuals, we demonstrate the important role that recombination plays in decreasing interference between QTL, as well as reducing linkage disequilibrium between QTL and neighboring neutral loci. Our results emphasize that designing the artificial selection experiment to allow more opportunity for recombination increases the ability to detect and localize QTL.

Finally, previous work has suggested that when founder sequence information is available, one can obtain more accurate allele frequency estimates by estimating local haplotype frequencies from pooled read data (Long *et al.* 2011; Kessner *et al.* 2013). We show that these improved allele frequency estimates, when compared to estimates calculated directly from read counts, can lead to an increase in power, although the magnitude of improvement is condition dependent.

## Methods

### Forward simulation

We used the program *forqs* (Kessner and Novembre 2014) for all forward simulations. *forqs* simulates whole genomes of individuals efficiently by tracking the haplotype chunks that are inherited from the founder individuals in the initial generation. *forqs* allows the specification of quantitative traits and provides flexible fitness functions for the simulation of complex evolutionary scenarios such as artificial selection.

In our simulations of artificial selection on a quantitative trait, individuals had three chromosomes, with lengths matching *Drosophila* chromosomes X, 2, and 3. For simplicity, we set the recombination rate to be uniform along each chromosome, with rate = 2 cM/Mb, which is similar to recombination rates reported previously (Comeron *et al.* 2012). For each set of experimental parameters we simulated populations using 200 replicates each of 12 different canonical architectures: 2, 5, 10, and 100 QTL, at initial heritability levels 0.2, 0.5, and 0.8. For a given number of QTL and heritability level, QTL positions and effect sizes were generated randomly for each simulation run (see next section). The random trait generation was implemented in an auxiliary program that produces trait description files, which are included by *forqs* configuration files that specify the experimental setup. Our simulation scenarios represented selection on standing variation where the expected waiting time  $1/(2N\mu)$  for a mutation to occur at a QTL is much larger than the number of generations in the experiment. Hence, we did not include *de novo* mutations in our simulations.

Each generation, *forqs* calculates trait values for all individuals based on the effect sizes of the alleles they carry at QTL and a random environmental effect (with variance determined by the heritability of the trait). Artificial selection was simulated using the *forqs* module *FitnessFunction\_TruncationSelection*, which assigns a fitness value of 1 to those individuals whose

trait value lies above (or below) a threshold (and 0 otherwise), where the threshold is determined each generation by the distribution of trait values in the population and the user-specified proportion of individuals selected to create the next generation.

All configuration files and analysis scripts for all simulations are freely available online at [https://bitbucket.org/dkessner/artificial\\_selection\\_pipeline](https://bitbucket.org/dkessner/artificial_selection_pipeline).

### Generation of random trait architectures

To investigate how the genetic architecture of a trait affects the behavior of QTL allele frequencies under artificial selection, we developed a method to generate random trait architectures with specified parameters. In particular, we were interested in how the number of QTL contributing to the trait and the heritability of the trait affect the power to detect QTL. In addition, we wanted to investigate how the trait architecture affects a focal QTL with a specified effect size and initial allele frequency. Finally, we wanted to ensure that linkage disequilibrium in the simulated starting population is similar to that found in populations used in experimental settings. In the following, we describe our procedure for generating the founding population and trait architecture in our simulations.

Starting with founder individuals for which we have full-genome haplotypes, we run neutral forward simulations with *forqs* to recombine the haplotypes and expand the population size. This results in a mixed population that is two to three times larger than the desired starting population for the selection experiment. For our founder individuals, we used the publicly available SNP data from 162 *Drosophila* inbred lines representing Freeze 1 of the *Drosophila* Genetic Reference Panel (DGRP) project (Mackay *et al.* 2012). Our procedure is similar to the experimental procedure used by Turner and Miller (2012), where individuals from the DGRP inbred lines are allowed to mate randomly for several generations to create a mixed population with genetic variation and linkage disequilibrium similar to those of the natural population from which the inbred lines were derived.

In some cases we additionally specify a focal QTL, for which we specify the locus, effect size, and initial allele frequency. We create a starting population by randomly selecting a subset of individuals from the mixed population. We ensure that the focal QTL has the desired allele frequency by choosing individuals in Hardy-Weinberg proportions according to their focal QTL genotype.

From the heritability and total variance parameters that we specify, we calculate a target genetic variance for the trait. We choose the remaining QTL positions uniformly at random across the genome, until we have the specified number of QTL. We choose effect sizes according to a standard exponential distribution, with random sign for positive/negative effect on the trait. From the individuals' haplotypes and QTL effect sizes, we calculate the current genetic variance of the trait in the population. We then scale the QTL effect sizes so that the genetic variance is equal to the target genetic variance.

In the case where we have a focal QTL, we keep the focal QTL effect size constant and scale the other QTL effect sizes—this requires iteration until the genetic variance is close to the target genetic variance, within a specified tolerance.

### Realized selection coefficient

We first recall the deterministic model of selection on a single locus in a diploid population. Suppose a locus has two alleles  $A_0$  and  $A_1$ . If  $p$  is the frequency of  $A_1$ , then the allele frequency change is given by

$$\Delta p = p(1-p) \frac{w_1^* - w_0^*}{\bar{w}},$$

where  $w_0^*$  and  $w_1^*$  are the marginal fitnesses of the  $A_0$  and  $A_1$  alleles, respectively, and  $\bar{w}$  is the mean fitness of the population (Rice 2004).

In the special case where  $A_1$  has additive selection coefficient  $s$ , this becomes

$$\Delta p = p(1-p) \frac{s}{1+2ps}.$$

We define the *realized selection coefficient* for a given  $p$  and  $\Delta p$  to be the selection coefficient  $s$  that would result in this allele frequency change in the deterministic case. By solving the above equation for  $s$ , we obtain a formula for the realized selection coefficient:

$$s_{\text{realized}}(p, \Delta p) = \frac{\Delta p}{p(1-p) - 2p\Delta p}.$$

### Calculation of power and false-positive rate

We explored several methods for calculating the power and false-positive rate associated with the detection of QTL by allele frequency differences between populations.

First, we note that calculating the false-positive rate as the proportion of neutral (non-QTL) variants detected does not capture the effect of detected variants clustering due to linkage disequilibrium.

In addition, significantly diverged variants are not necessarily causal, but indicate that the region nearby may contain a causal variant. To address these issues, we define a detection region to include all variants within a specified radius of any variant whose absolute allele frequency difference value exceeds a given threshold (see *Results, Measurement of power to detect and localize QTL*). This leads to a natural definition of false-positive rate: the proportion of the neutral genome covered by the detection region (see Figure 2 for an illustration).

We calculated the false-positive rate in this way with radii of 10 kb, 100 kb, and 1 Mb. We found that using a radius of 100 kb or 1 Mb led to true regions that covered a substantial portion of the genome in the cases where there were a large number of QTL. Using a 10-kb radius for the detection region gave the most interpretable results.

With regards to power, the simplest method is to calculate the proportion of QTL detected. However, we feel that a more

relevant measure of power is the proportion of the genetic variance in the initial population that is explained by the detected QTL. This measure appropriately gives greater weight to QTL responsible for more of the genetic variance. Figure 3, A and B illustrates the difference between these two measures of power: at a false-positive rate of  $10^{-4}$ , only 50% of the QTL are detected, but these QTL are responsible for .75% of the genetic variance of the trait.

Going forward, we use the 10-kb radius for the detection region, together with measuring power as the proportion of variance explained, to provide the most interpretable results.

### Analysis pipeline

The forward simulator forqs efficiently simulates the entire genome of each individual by tracking haplotype chunks. forqs outputs data files that represent each chromosome of each individual as a mosaic of founder haplotypes. To obtain the neutral variants carried by an individual, the neutral variation on founder chromosomes must be propagated to the individual's mosaic chromosomes.

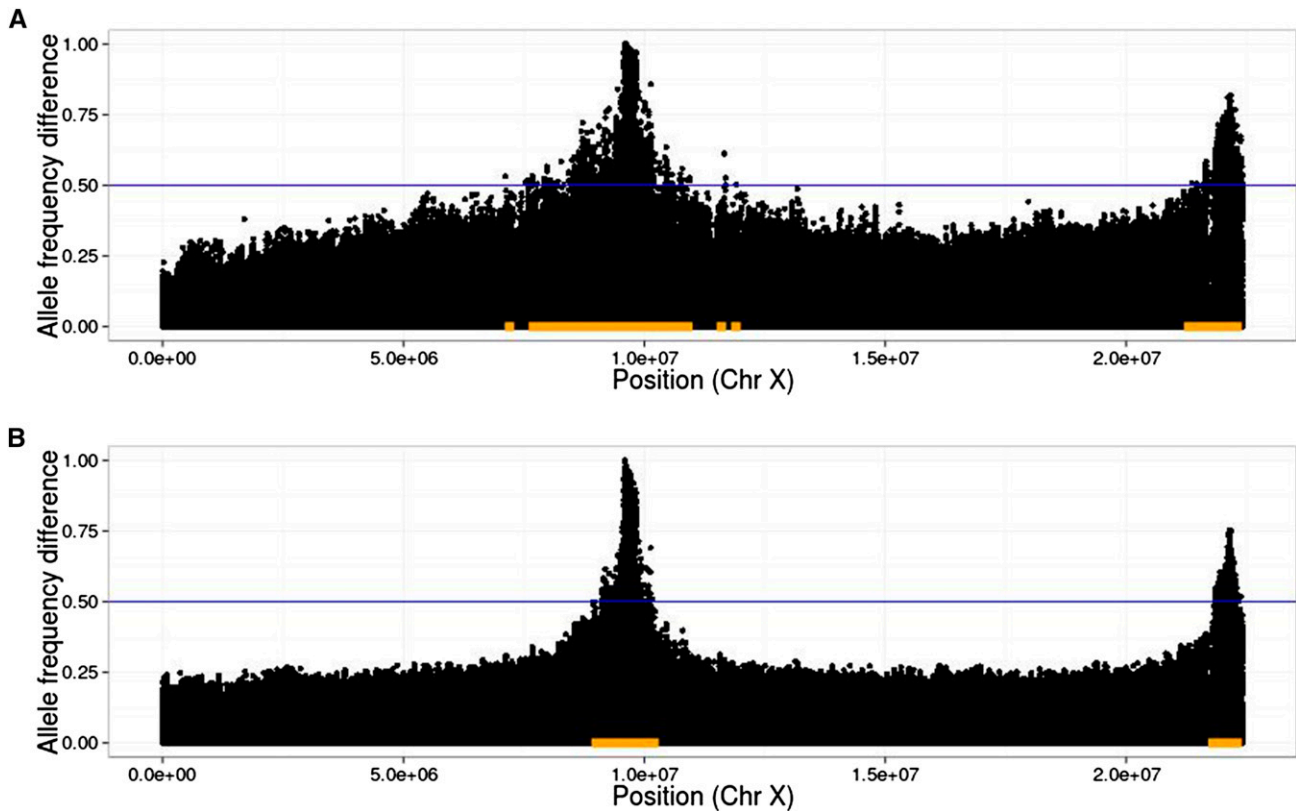
In our simulation framework, we use two forward simulations. The first simulation creates a mixed population from the founder haplotypes, after which we generate the random trait architecture. The second simulation represents the selection experiment. Individuals in the final populations are mosaics of individuals in the mixed population, which are in turn mosaics of the founders. We implemented a custom program to handle this two-step propagation of neutral variation. Given founder sequences, the mixed population, and the final population, the program calculates the allele frequency in the final population of each variant in the genome. We note that while this step does not require a large amount of computation, it is intensive for disk input/output because it uses the full DGRP variant data set for the founder sequences, which consists of 162 haplotypes at  $\sim 5$  million loci.

After calculating allele frequencies for each population, we calculate the allele frequency difference  $D$  for each high-low population pair under consideration (multiple pairs for the replication analyses). After sorting the  $D$  values, we begin with the highest  $D$  value and iteratively decrease the threshold to obtain data equivalent to a receiver operating characteristic (ROC) curve (power and false-positive rates) for that simulation run. We note that this step depends on the method for calculating power and false-positive rate, so we performed it once for each method we described above in *Calculation of power and false-positive rate*.

We obtain average ROC curves by calculating the average power over replicate simulation runs at regularly spaced false-positive rates, where the power for a particular run at a given false-positive rate is obtained by linear interpolation between points on its ROC curve.

### Empirical error distributions

To investigate the power increase due to the use of haplotype-based allele frequency estimates, we needed to simulate



**Figure 2** Illustration of linkage disequilibrium between QTL and neutral loci. In plots of allele frequency differences between high and low populations, QTL peaks are narrower when extra generations of neutral mixing are introduced. As an example, we indicate a threshold of 0.5 with the blue lines and the corresponding detection regions with orange bars. Note that the true region, consisting of sites within 10 kb of either SNP, is too small to be seen at this scale; thus, the size of the orange bars represents the (local) false-positive rate at this threshold. Also note that the threshold of 0.5 was chosen for the purpose of illustrating the detection region, and in practice a much higher threshold will typically be chosen (10 QTL,  $h^2 = 0.5$ ,  $N = 1000$ , 20% selected, 20 generations of initial mixing, shown are average  $D$  values over 20 replicates). (A) Twenty generations of selection; (B) 20 generations of selection, 4 generations per selection event (80 generations total).

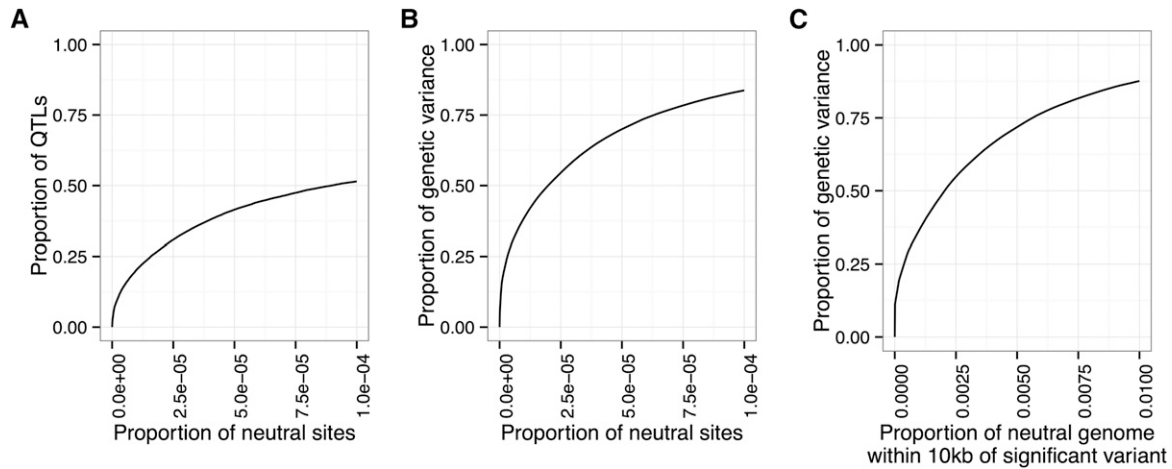
pooled sequence reads from a population, followed by haplotype frequency estimation in sliding windows across the genome, using the harp method and software detailed in Kessner *et al.* (2013). Because this procedure is computationally expensive, and due to the large number of simulations involved in this study, it was not feasible to do this for each simulated experiment.

As an alternative, we obtained empirical error distributions, which we later used to add random errors to true allele frequencies. Because errors in the haplotype frequency estimation depend on the length scale of recombination, we ran replicate neutral simulations for varying numbers of generations (from 40 to 400). Haplotypes surrounding selected QTL are expected to be longer than in neutral regions, so our empirical error distributions are conservative. We simulated pooled sequence reads from the final populations, using a new program `simreads_forqs` based on the read simulator used in Kessner *et al.* (2013). Haplotype frequency estimation was performed with harp in overlapping sliding 200-kb windows within a single 1-Mb region. We then calculated allele frequencies at variant sites within the region, using read counts. By considering allele frequencies in bins of size 0.05, we obtained

a frequency-dependent empirical error distribution. Similarly, we also derived allele frequency estimates from the local haplotype frequencies, from which we obtained frequency-dependent empirical error distributions for each generation count. We found that after  $\sim 200$  generations, the haplotype-based estimates were no better than the read-count-based estimates; this behavior is expected as the recombination length scale approaches the window size used for haplotype frequency estimation (200 kb in this analysis). Hence, for this analysis we considered only experimental scenarios lasting  $< 200$  generations.

## Results

We performed forward simulations of populations, using the program `forqs` (Kessner and Novembre 2014), which models whole genomes of individuals and selection on quantitative traits. QTL were chosen randomly from variant sites reported in the DGRP (Mackay *et al.* 2012). Thus, initial allele frequencies of QTL were randomly distributed according to the allele frequency distribution of DGRP variant sites. Similarly, linkage disequilibrium patterns reflect the patterns observed in the DGRP populations. To simulate various trait architectures,



**Figure 3** Comparison of three methods for calculating and interpreting power and false-positive rate. (A) Power is measured by the proportion of QTL detected, and false-positive rate is measured by the proportion of neutral variant sites detected. (B) Power is measured by the proportion of genetic variance in the founder population explained by the detected QTL, and false-positive rate is measured as in A. (C) Power is measured as in B, and false-positive rate is measured by the proportion of the neutral genome covered by the detection region. In this case, the ROC curve shows that >75% of the genetic variance is explained by QTL in a detection region that covers 1% of the neutral genome.

we considered 12 scenarios by simulating 2, 5, 10, or 100 QTL at initial heritability levels of 0.2, 0.5, or 0.8. Following previous theoretical and empirical studies (Orr 1998; Otto and Jones 2000; Mackay 2010; Thornton *et al.* 2013), we drew QTL effect sizes randomly from an exponential distribution. These scenarios span settings that are relatively straightforward for genetic mapping such as an oligenic trait with 2 QTL and a high heritability of 0.8 to a more challenging scenario of a complex trait with 100 QTL and heritability of 0.2. Our simulations began with several generations of neutral mixing, emulating laboratory procedures that use inbred founder lines to create larger experimental populations with increased genetic variation and reduced linkage disequilibrium (*e.g.*, Turner and Miller 2012). Using this procedure we simulated populations of various sizes, which we used as the initial populations for the artificial selection simulations.

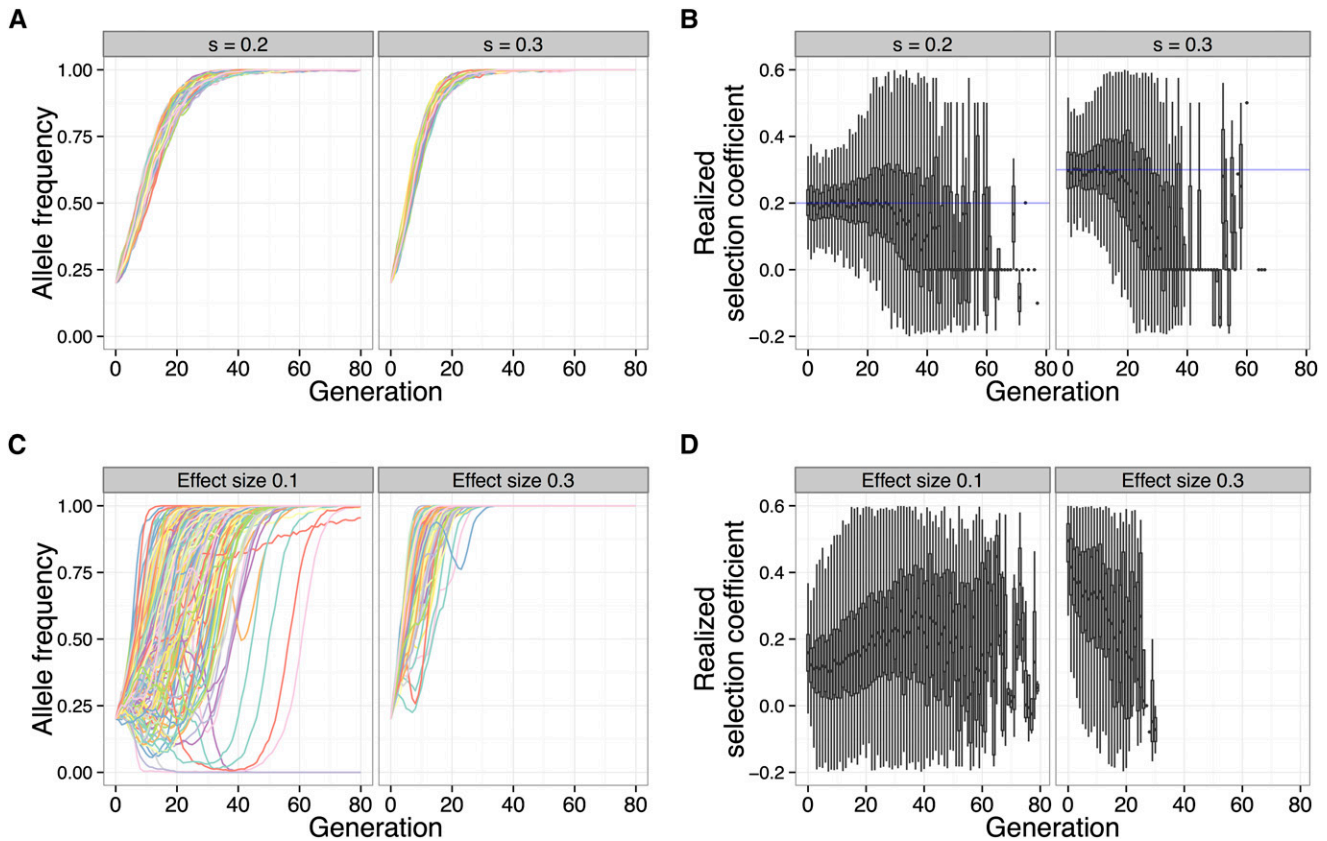
#### **Qualitative comparison of selection coefficient simulations and explicit quantitative genetic modeling**

To first illustrate the importance of modeling quantitative traits explicitly, we show examples of allele frequency trajectories of a focal QTL contributing to a quantitative trait under truncation selection in comparison to a single locus with two alleles with a constant additive selection coefficient. For these simulations, we depart from our general procedure (described above) and fix the starting allele frequency of a focal QTL to 0.2 and track its trajectory. The remaining QTL are chosen with a random distribution on starting allele frequencies and effect sizes (see *Methods*). We simulated populations of size  $N = 5000$  with the top 20% of individuals selected based on their trait values to create the next generation, so that the effective population size was  $N_e = 1000$ . To match the predicted degree of drift at neutral loci, we simulated populations of size  $N = 1000$  for the selection coefficient simulations. Selection is assumed to be in the direction of increasing trait values.

As seen in Figure 4, A and C, the allele frequency trajectories of the focal QTL under truncation selection are qualitatively different from the trajectories under a constant selection coefficient. While a strong selection coefficient leads to nearly deterministic allele frequency trajectories that monotonically increase, trajectories of a focal QTL under strong truncation selection are dependent on the underlying trait architecture. In particular, the focal QTL may experience interference from other linked QTL. This effect can be seen in the trajectories where the focal QTL decreases in frequency at first, due to repulsion linkage disequilibrium (*i.e.*, QTL of opposite effect are positively associated), but eventually increases in frequency once linkage disequilibrium has decreased sufficiently through recombination.

In addition, once an allele with a constant selection coefficient reaches high frequency, it only gradually increases in the final generations before finally going to fixation. In contrast, the focal QTL under truncation selection tended to become fixed quickly after reaching high frequency in the population. This behavior is not surprising, because after a few generations of selection, the upper tail of the population trait value distribution is highly enriched for individuals carrying high-effect variants.

To further illustrate these qualitative differences, we analyzed the realized selection coefficient of the trajectories, which represents the selection coefficient that would result in a given single-generation allele frequency change under a deterministic model (see *Methods*). Under a constant selection coefficient, the mean realized selection coefficient tracks the true selection coefficient closely during the selection phase, after which it decreases to zero during the drift phase (Figure 4B). Under truncation selection, the behavior of the mean realized selection coefficient depends on the underlying genetic architecture of the trait. When the effect size is low, the realized selection coefficient increases each generation—this is



**Figure 4** Qualitative differences between fixed selection coefficient and truncation selection on a quantitative trait. A focal QTL exhibits fundamentally different behavior under a constant selection coefficient, compared to truncation selection on a quantitative trait. Shown are allele frequency trajectories (A) and realized selection coefficient distributions (B) of a locus under two different selection coefficients (0.2, 0.3) ( $N = 1000$ , 80 generations). C and D show the same for a focal QTL (effect sizes 0.1, 0.3) under truncation selection on the trait ( $N = 5000$ , 20% selected, 80 generations, 100 QTL,  $h^2 = 0.8$ ,  $\sigma_{\text{trait}}^2 = 1$ ).

because selection acts on the larger-effect QTL first and then has a greater effect on the focal QTL after the larger-effect QTL have reached fixation. On the other hand, when the effect size is higher, the focal QTL experiences very strong selection initially, decreasing as the focal QTL rises to fixation (Figure 4D).

Another effect of the underlying trait architecture can be seen in the fixation times of the focal QTL (Figure 5). For a given effect size and heritability, the fixation time of the focal QTL increases with total number of QTL due to interference. While these two forms of simulation (explicit QTL simulation vs. single-locus fixed- $s$  simulation) cannot be exactly calibrated to match one another, the qualitative result that the explicit QTL simulation is capturing more complex and realistic dynamics motivated our use of QTL-based simulations for assessing power and experimental design choices going forward.

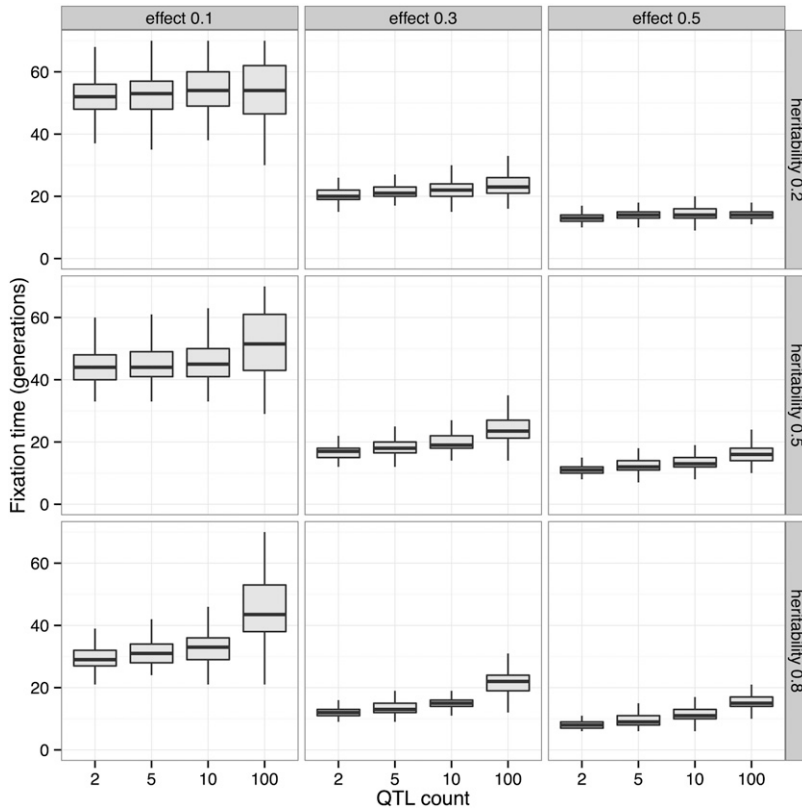
#### Measurement of power to detect and localize QTL

In all of the following analyses, we examine the power to detect QTL through allele frequency differences at variant sites. There is currently no consensus regarding the choice of test statistic for the analysis of artificial selection experiments.

Both for simplicity and because of its use in practice (Parts *et al.* 2011; Turner and Miller 2012), we chose to base our analysis on the absolute allele frequency difference  $D = |p_1 - p_2|$ , where  $p_1$  and  $p_2$  are the allele frequencies of the variant in populations 1 and 2, respectively. We also note that this statistic, also called the difference in derived allele frequencies ( $\Delta\text{DAF}$  or  $\text{DDAF}$ ), is used in several related or composite methods for detecting selection (Grossman *et al.* 2010; Turner *et al.* 2011; Utsunomiya *et al.* 2013).

We calculate  $D$  for each variant site in the genome, and we call a site detected if the  $D$  value exceeds a threshold value. By varying the threshold, we obtain ROC curves showing the relationship between power (true positive rate) and the false-positive rate.

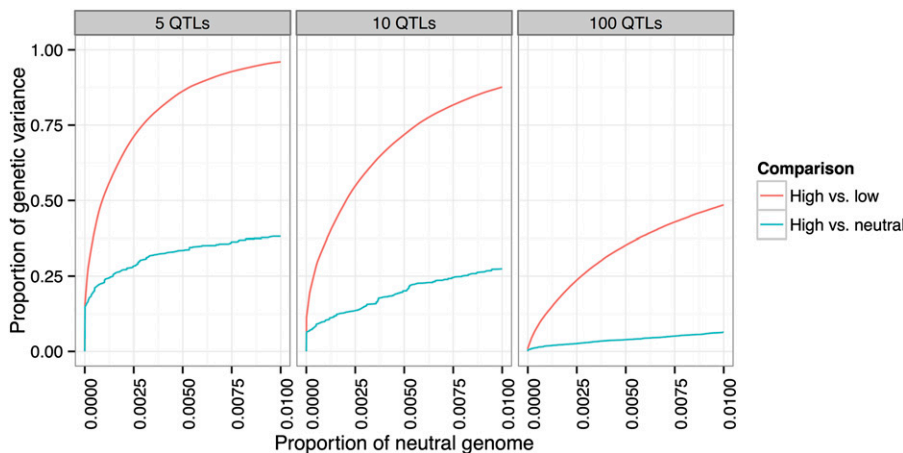
Due to linkage and strong selection, detected QTL will generally have neighboring neutral variants whose allele frequency differences also exceed the detection threshold. Because of this, detection and localization of a QTL are necessarily intertwined. In an actual experimental setting, the entire genomic region surrounding the significantly diverged loci would often be chosen for follow-up studies.



**Figure 5** Effect of genetic architecture on fixation times. Fixation times of a focal QTL for a trait under truncation selection decrease with increasing effect size and heritability. For a fixed effect size and heritability, the fixation times increase with the number of QTL contributing to the trait due to interference ( $N = 5000$ , 20% selected, 80 generations).

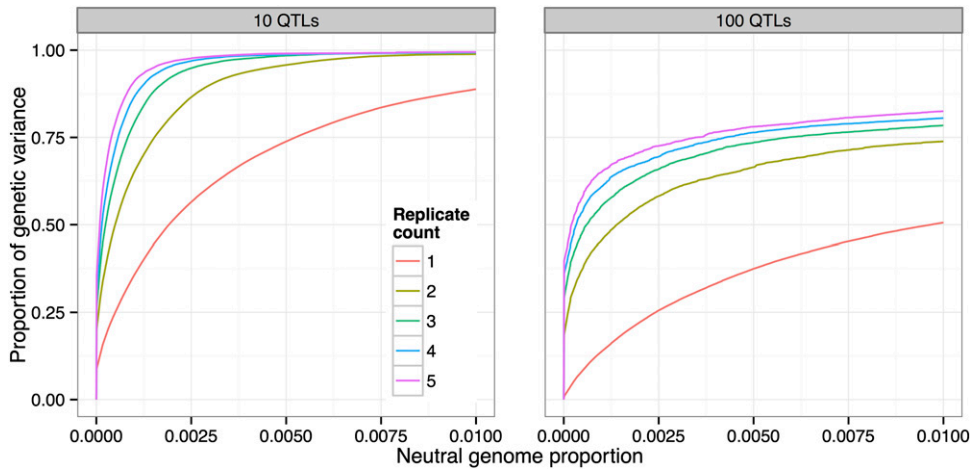
We explored several methods for calculating and interpreting power and false-positive rate. We present our results using the method that we found to be most interpretable, which we summarize here (see *Methods* for full details on the different methods). For a given  $D$ -value threshold, we determine a *detection region* that consists of all variants within a specified radius of any variant above the threshold (10 kb for the results presented here; see Figure 2 for an illustration). Power is calculated as the proportion of genetic variance (in the founder population) explained by the QTL located within the detection region. We define the *true region*

to consist of all variants within the specified radius of any QTL (*i.e.*, 10 kb here), and we define the *neutral genome* to be the rest of the genome (outside the true region). The false-positive rate is calculated as the proportion of the neutral genome covered by the detection region. Thus, a false-positive rate of 0.01 can be interpreted to mean that 1% of the genome would be incorrectly flagged for follow-up studies. In *D. melanogaster*, this would correspond to a 1.4-Mb region. We note also that the false-positive rate represents the combined size of regions surrounding loci above the threshold; the detection region surrounding a single locus will be one to



**Figure 6** Increase in power due to divergent selection. Comparison of two populations divergently selected for extreme values of a trait has greater power to detect QTL than comparison between selected and control populations. Three populations of 1000 individuals (high, low, neutral) originating from a single founder population were simulated for 20 generations, with 20% selected each generation in the high and low populations. Shown are the scenarios with 5, 10, and 100 QTL, with  $h^2 = 0.5$ .





**Figure 7** Increase in power due to replicate populations. Adding replicate pairs of divergently selected populations increases power to detect QTL. Five pairs of populations originating from a single founder population were simulated ( $h^2 = 0.5$ ,  $N = 1000$ , 20% selected, 20 generations).

two orders of magnitude smaller than this (*i.e.*,  $\sim 10\text{--}100$  kb in *D. melanogaster*), depending on the trait architecture.

### Advantages of divergent artificial selection

In artificial selection studies of quantitative traits, one commonly used technique is divergent (also called bidirectional) selection, where the first “high” population is obtained by selecting from the upper tail of the trait value distribution each generation, and the second “low” population is obtained by selecting from the lower tail (for example, Johansson *et al.* 2010 and Turner and Miller 2012). Use of this technique presumes that allele frequency differences between the high and low lines will be more pronounced at QTL contributing to the trait than, for example, differences between the high line and a control population that has been evolving neutrally.

To compare the power obtained by a divergent selection experiment to the power obtained by selection in a single direction, we simulated three populations originating from a single founder population, where one population was selected for high values, one selected for low values, and one allowed to evolve neutrally. Each population had 1000 individuals, of which 20% were selected each generation in the high and low populations, for 20 generations total.

We ran the simulations for each of our 12 canonical architectures (see *Methods*), and we show the results in Figure 6 for some of these ( $h^2 = 0.5$ , with 5, 10, and 100 QTL). Comparison between the high and low populations leads to a substantial increase in power over the comparison between the high and neutral populations. In particular, at a 1% false-positive rate, roughly half of the genetic variance of the trait is detected in the high vs. low comparison but not in the high vs. neutral comparison.

### Extent of increased power due to replication

Another technique available in artificial selection experiments is the use of replicate high and low populations to increase confidence that allele frequency differences between diverged populations are due to selection rather than genetic drift. A natural generalization of the allele frequency difference  $D$  for

replicate populations is  $D = \left| \sum_{i=1}^R p_{i,\text{high}} - p_{i,\text{low}} \right|$ , where  $R$  is the number of replicate population pairs.

We investigated the effects of evolving replicate populations by simulating five pairs of populations originating from a single founder population (10 QTL,  $h^2 = 0.5$ ,  $N = 1000$ , 20% selected, 20 generations). We then calculated the average power to detect QTL, using subsets of the data representing population replicates from one to five pairs.

We found that using two replicate populations substantially increases power to detect QTL (Figure 7). For example, at the low false-positive rate (neutral genome proportion) of 0.0005, the proportion of genetic variance detected increases from 25% to 50%. Adding further replicate populations continues to increase power, but with diminishing returns.

### Effect of population size

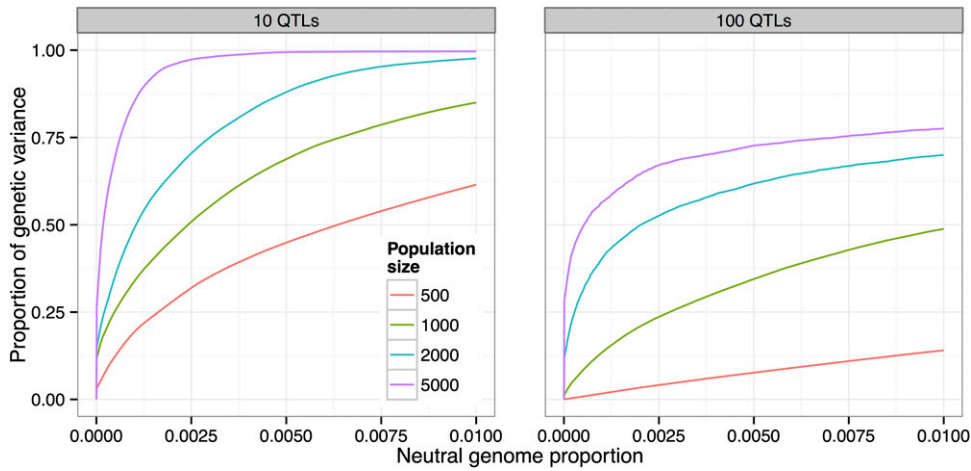
It is well known that selection on a single locus, defined by a selection coefficient  $s$ , acts more efficiently in larger populations, as can be seen in the dependence of fixation probabilities and fixation times on the population-scaled selection coefficient  $4Ns$  (Ewens 2004).

To investigate how the population size affects artificial selection experiments, we performed simulations of populations of different sizes ( $N = 500, 1000, 2000, 5000$ ) under identical experimental conditions (20% selected, 20 generations), with each of our 12 canonical trait architectures.

In all cases we found a substantial increase in power to detect and localize QTL as we increased the population size. For example, in the case of 10 QTL and  $h^2 = 0.5$ , increasing the population size from 500 to 1000 resulted in an additional 25% of the genetic variance detected at the 0.5% false-positive rate (Figure 8). Moreover, while  $<50\%$  of the genetic variance was detected with  $N = 500$  at that false-positive rate, nearly 100% was detected with  $N = 5000$ .

### Effects of the length of experiment and proportion selected

We next investigated the effects of the length of the experiment and the strength of selection on the power to detect QTL. To do this, we simulated four different selection



**Figure 8** Increase in power due to population size. Using larger population sizes substantially increases power to detect QTL ( $h^2 = 0.5$ , 20% selected, 20 generations).

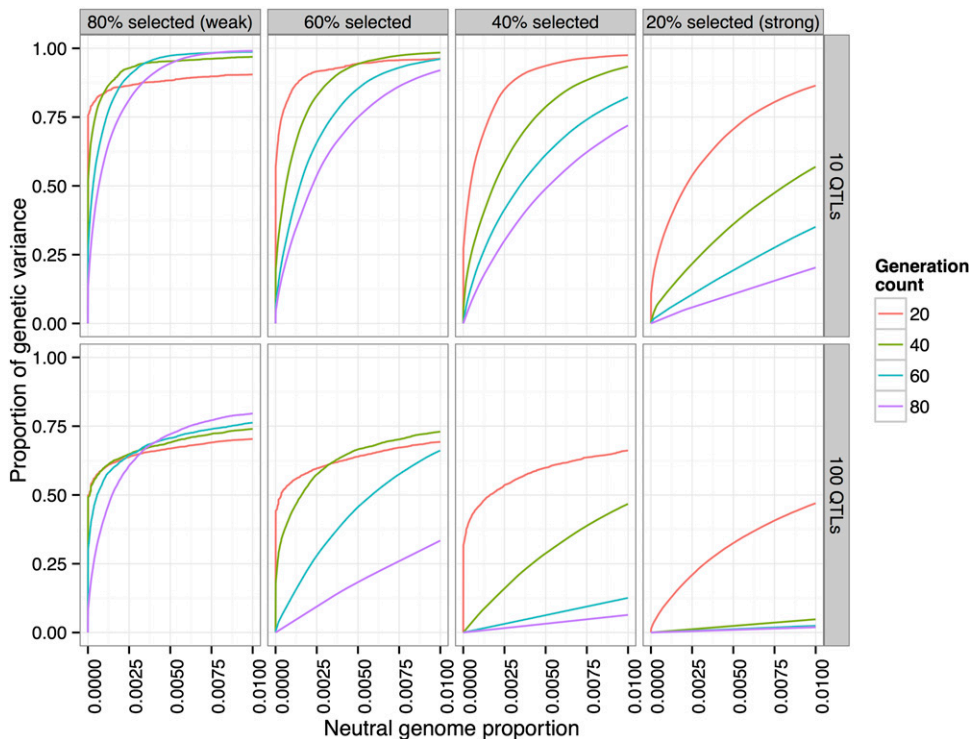
scenarios, from weak selection (top/bottom 80% selected each generation) to strong selection (20% selected). Each simulation ran for 80 generations, and we examined snapshots of each population at generations 20, 40, 60, and 80. In these simulations, the populations consisted of 1000 individuals, and the trait had 10 QTL, with a heritability of 0.5.

One particularly interesting finding was that under strong selection, power can actually decrease with the number of generations, depending on the false-positive rate threshold chosen (for example, 20% or 40% selected, at false-positive rate 0.01 in Figure 9). In these scenarios, most of the QTL have differentiated by generation 20 (>75% of the genetic variance was detected at false-positive rate 1%). However, the lower

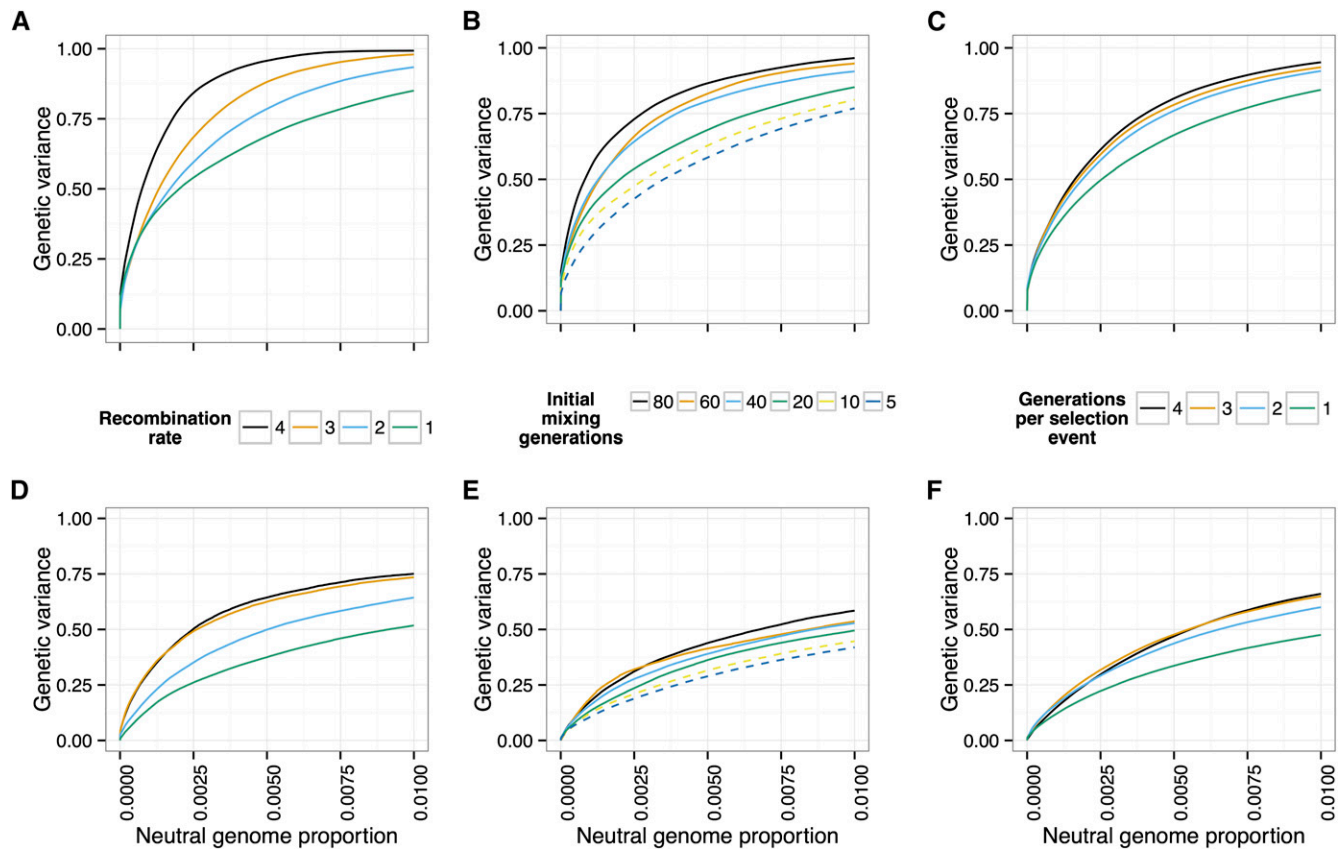
effective population size induced by strong selection results in the fixation of many neutral variants, which are then falsely detected as QTL.

Conversely, by lowering the selection pressure (80% selected each generation) the maximum power is actually increased. In addition, letting the experiment run for a greater number of generations at this lower selection pressure increases the maximum power attained.

These observations suggest that recombination plays a large role in the power to detect and localize QTL: reducing the selection pressure and increasing the number of generations should allow more recombination, which reduces linkage disequilibrium between QTL and neighboring neutral variants. Increased recombination will also reduce interference between



**Figure 9** Effects of length of experiment and selection strength. Experiments with weak selection (top/bottom 80% selected each generation) over a longer number of generations have higher power than those with strong selection (20% selected) ( $N = 1000$ ,  $h^2 = 0.5$ ).



**Figure 10** Effect of recombination. The power to detect and localize QTL depends on the extent of recombination experienced by the populations. (A and D) Power increases with the recombination rate of the organisms under selection. (B and E) Additional generations of initial mixing and (C and F) additional generations between rounds of selection similarly increase power. [Top row (A–C), 10 QTL; bottom row (D–F), 100 QTL;  $h^2 = 0.5$ ,  $N = 1000$ , 20% selected, 20 generations of selection, and 20 generations of mixing in A, C, D, and F.]

QTL, which should allow lower-effect QTL to be selected and detected. This led us to investigate the effects of recombination further in our subsequent analyses.

### Effect of recombination

To investigate the effects of recombination on the power to detect and localize QTL, we first considered the effect of increasing the recombination rate of the simulated individuals. We simulated populations of 1000 individuals, with each of our 12 canonical trait architectures, with recombination rates at 1×, 2×, 3×, and 4× our standard recombination rate (see *Methods*) (20 generations of neutral mixing, 20 generations of selection, 20% selected).

We found that increasing the recombination rate does indeed increase the power to distinguish QTL from neutral variants, for all trait architectures (Figure 10, A and D). While in practice it is not feasible to experimentally increase the recombination rate of individuals, the experiment can be designed to increase the opportunity for individual chromosomes to recombine and thus decrease linkage between QTL and neighboring neutral variants.

One way to decrease linkage disequilibrium in the population is to allow more generations of initial neutral mixing in the founder population, before selection starts (*e.g.*, Turner

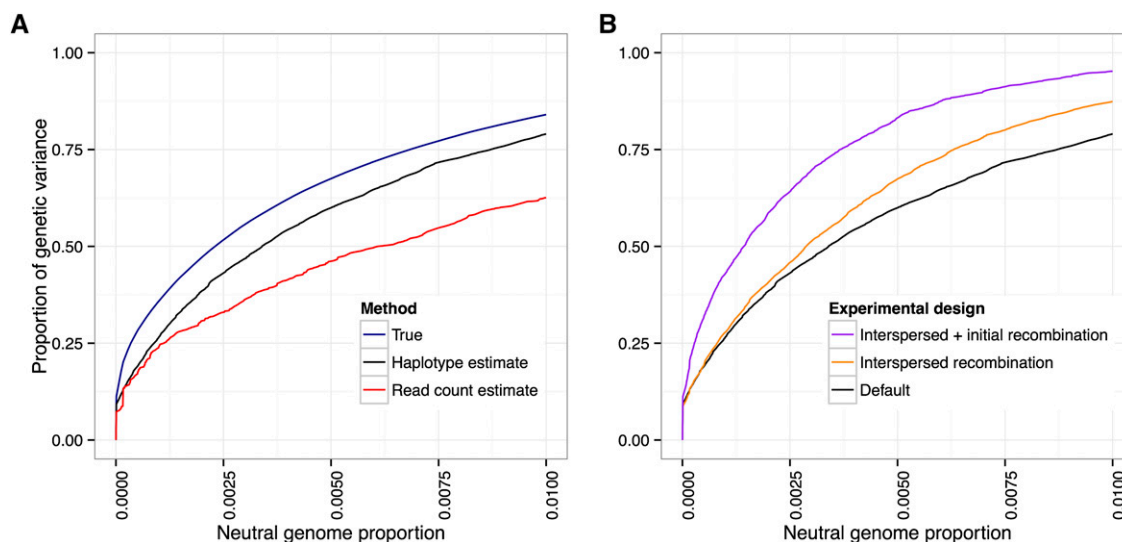
and Miller 2012). We simulated scenarios with varying numbers of generations of neutral mixing (20, 40, 60, 80), with the other parameters as above. We found that increasing the number of generations of initial neutral mixing increases power (Figure 10, B and E).

An alternative way to decrease linkage disequilibrium is to intersperse extra generations of neutral mixing between rounds of selection. To test this idea, we simulated scenarios where selection was carried out every one, two, three, or four generations, with random mating during the generations where selection did not take place (other parameters as above). Again, we found that the additional recombination events increased power to detect QTL (Figure 10, C and F). We further illustrate the effect of extra recombination on allele frequency differences in Figure 2.

Taken together, these results show that the ability to detect and localize QTL depends crucially on recombination, both to decrease interference between QTL and to break up linkage between QTL and nearby neutral variants.

### Haplotype-based inference of allele frequencies increases power

In all of our power analyses, we have assumed that we are able to calculate the allele frequencies at all variant sites



**Figure 11** Using founder haplotype information estimation improves power. (A) If founder sequence information is available, estimating local haplotype frequencies leads to better allele frequency estimates and an increase in power over that of estimates obtained from raw read counts. (B) Error in local haplotype frequency estimates increases with the number of generations due to recombination; however, there is still a net gain in power from extra generations of neutral mixing. Default: 20 generations mixing, 20 generations selection. Interspersed: 20 generations mixing, 4 generations per selection event, 20 selection events (100 generations total). Initial + interspersed: as interspersed, with 70 generations mixing (150 generations total). (Ten QTL,  $h^2 = 0.5$ .)

perfectly. However, in actual experiments, the allele frequency at a locus will be estimated by considering the sequence read counts that cover the site. These estimates are prone to errors from two sources: random base errors and stochasticity in the number of reads covering the site. The error in measurement of allele frequencies will lead to a loss of power to detect and localize QTL.

In the case where founder sequences are known, previous studies (Long *et al.* 2011; Kessner *et al.* 2013) have shown that one can leverage haplotype information from the founders to obtain local haplotype frequencies and that these estimates are robust to both sequence read errors and uneven coverage. This work suggested that improved allele frequency estimates could be derived from local haplotype frequency estimates.

We investigated the effect of allele frequency estimation error on power by first obtaining empirical error distributions for the two estimation methods (see *Methods*). We then simulated artificial selection experiments with our standard experimental parameters ( $N = 1000$ , 20% selected, 20 generations) and canonical trait architectures, followed by the random introduction of errors based on the empirical error distributions. In all cases, we found that the haplotype-based allele frequency estimates led to improved power over the read-count-based estimates. However, at very low false-positive rates ( $< 0.1\%$  neutral genome detected), the improvement is small, which suggests that founder haplotype information may not be very useful in these cases. Figure 11A shows the results from the case with 10 QTL and  $h^2 = 0.5$ .

Our analyses from the previous section showed that increasing the amount of recombination led to an increase

in power. On the other hand, increased recombination results in shorter haplotypes. This leads to greater errors in local haplotype frequency estimates and the allele frequency estimates derived from them, with a subsequent decrease in power. To investigate the relative magnitude of these two counteracting effects, we added two scenarios: one where we interspersed rounds of neutral mating between generations of selection (4 generations per selection event, 100 generations total for the experiment) and one where we also added an extra 50 generations of initial neutral mixing (150 generations total). We introduced random errors as above, using empirical error distributions that take into account the increased number of generations.

We found that recombination's positive effect of breaking up linkage between QTL and neutral sites outweighed the negative effect on haplotype frequency estimation (Figure 11B). However, for experiments lasting  $\geq 200$  generations, our haplotype-derived allele frequency estimates were no better than read-count-derived estimates (see *Methods* for parameter assumptions). Thus, we do not expect haplotype-derived allele frequency estimates to lead to increased power in scenarios where the typical length of haplotype chunks is shorter than the window size used for the local haplotype frequency estimation.

## Discussion

We have presented a new analysis of the power of artificial selection experiments to detect and localize loci contributing to a quantitative trait. In this analysis, we explicitly model whole genomes of individuals, quantitative traits, and selection

based on individual trait values, using a novel simulation framework.

We showed that population genetic simulations based on loci with constant selection coefficients do not fully capture the dynamics of QTL contributing to a trait under artificial selection and that the trait architecture plays a large role in these dynamics. In addition, explicit modeling of selection on a quantitative trait has several other advantages. For example, simulated experiments can be parameterized and results can be reported using the standard quantitative genetics concepts of effect size, genetic variance, and heritability. Also, scenarios such as divergent selection can be simulated in a straightforward manner. Finally, the behavior of a QTL under artificial selection is dependent both on the experimental design (proportion of individuals selected each generation) and on the trait architecture (effect size and linkage to other QTL). While the selection coefficient conflates these parameters into a single number, our simulation framework allows these parameters to be separated and investigated independently.

Our results show the important role that recombination plays in the ability to identify QTL. Recombination not only reduces interference between QTL, but also decreases linkage disequilibrium between QTL and neighboring neutral loci. In fact, one can view the artificial selection experiment as a classification problem. From this viewpoint, the ROC curve, which measures the ability to classify a locus as QTL vs. neutral, contains information about how well the experiment can break linkage disequilibrium between causal and neutral loci (Figure 3).

The classification viewpoint is useful when thinking about how to design a selection experiment. For example, we found that experiments allowing more opportunity for recombination, either during initial neutral mixing of the founder haplotypes or between selection events, have greater power to detect and localize QTL. Similarly, experiments with weaker selection (greater proportion selected each generation) over a longer period of time will have greater power than shorter experiments with strong selection. We note that the improved mapping resolution afforded by additional recombination is analogous to the use of recombinant inbred lines (see Crow 2007 for a history) or advanced intercross lines (Darvasi and Soller 1995) in traditional QTL mapping studies.

While recombination increases the ideal power of an artificial selection experiment, it also decreases the ability to use founder haplotype information to obtain more accurate allele frequency estimates. We showed that the haplotype-based estimates still result in a net increase in power in experimental scenarios where the scale of recombination is larger than the window used for estimating local haplotype frequencies. This suggests that when choosing the window size for such an analysis, one should take into consideration the expected scale of recombination based on the experimental design.

Similar to the findings of Kofler and Schlotterer (2014) and Baldwin-Brown *et al.* (2014), we found that increasing

population sizes and number of replicates leads to an increase in power. Additionally, our simulation framework allowed us to quantify the increase in power due to bidirectional selection. We also note that adding generations of initial neutral mixing in the founder population is in some ways similar to increasing the number of founder haplotypes, in that it places QTL on multiple genetic backgrounds. Our results regarding the increase in power due to additional initial mixing are thus consistent with the findings of both of these groups that increasing the number of founder haplotypes increases power to detect and localize QTL.

Also in agreement with Baldwin-Brown *et al.* (2014), but in contrast to Kofler and Schlotterer (2014), we found that the effect of increasing the length of the experiment is not uniformly beneficial, but rather depends on the strength of selection. From the classification viewpoint again, power depends on the ability of the experiment to allow QTL to differentiate in selected populations while keeping allele frequencies at neutral loci constant. Continuing the experiment beyond the point where the majority of QTL have differentiated will lead to increased fixation of linked neutral loci and hence lower power.

In contrast to both Kofler and Schlotterer (2014) and Baldwin-Brown *et al.* (2014), whose results suggest that hundreds of generations are necessary to obtain reasonable power, we found that artificial selection experiments can detect QTL explaining most of the genetic variance of the trait in as little as 20 generations, under reasonable assumptions about the trait architecture (100 loci,  $h^2 = 0.5$ , exponentially distributed effect sizes). We caution that the high power observed here may not apply to other experimental designs, such as extremely weak selection on a quantitative trait or artificial selection based on survival in altered environmental conditions. On the other hand, it is not obvious how to interpret selection coefficients in the context of an artificial selection on a quantitative trait, where we feel it is more natural to use parameters representing the trait architecture and experimental design.

Future directions for this work would include relaxing the assumption that effect size and initial frequency are unrelated in the founder population. This implicitly assumes neutrality or near neutrality, whereas for traits under stabilizing selection one might expect large-effect mutations to be found only at low frequencies. Initializing the joint distribution of allele frequencies and effect sizes with a distribution that captures this expected inverse relationship would be one possible alternative approach.

A related future change is to relax the assumption of exponential distribution of effect sizes. As mentioned above, this assumption is sensible for neutral traits. One impact of this assumption is that the leading factor of a simulated trait often accounts for a nontrivial fraction of the genetic variance of the trait (*e.g.*, often 20–25% in traits with 100 QTL; [Supporting Information, Figure S1](#)) and that genome-wide association studies (GWAS) have high power to detect this leading factor as long as the heritability is modest ([Figure S2](#)). While such

major-effect loci are found for some complex traits in *Drosophila* [e.g., alcohol dehydrogenase (King *et al.* 2012) and susceptibility to viral infection (Magwire *et al.* 2012)] (see Figure S1), there are many other traits where such major-effect loci are not present. For such traits, we might expect the levels of interference during selection to be reduced from the values seen in our simulations, as the expected transit of the major-effect variant to high frequency will not interfere with selection at other QTL.

Another avenue for future research is to investigate alternatives to increasing recombination during the experiment. In particular, the use of a larger number of inbred founder lines may decrease linkage disequilibrium in the initial experimental populations, leading to higher power and better resolution without adding additional rounds of recombination.

In summary, we have shown that it is feasible to do whole-genome simulations of artificial selection with explicit quantitative trait modeling. The scope of this study was limited to the use of *D. melanogaster* as our model species and simple trait architectures with no dominance or pleiotropic effects. However, we believe that our simulation methodology can be applied to a wide variety of species, trait architectures, and experimental protocols. We emphasize that the opportunity for recombination is a key factor in the power to detect and localize QTL and that this should be taken into account by future designers of artificial selection experiments.

## Acknowledgments

The authors thank Alex Platt, Charleston Chiang, Eunjung Han, and Diego Ortega Del Vecchio for helpful comments and discussion. This work was supported by the National Institutes of Health (R01 HG007089 to J.N.), the National Science Foundation (EF-0928690 to J.N.), and the University of California, Los Angeles (dissertation year fellowship to D.K.).

## Literature Cited

- Baldwin-Brown, J. G., A. D. Long, and K. R. Thornton, 2014 The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Mol. Biol. Evol.* 31(4): 1040–1055.
- Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467(7315): 587–590.
- Cameron, J. M., R. Ratnappan, and S. Bailin, 2012 The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet.* 8(10): e1002905.
- Crow, J. F., 2007 Haldane, Bailey, Taylor and recombinant-inbred lines. *Genetics* 176: 729–732.
- Darvasi, A., and M. Soller, 1995 Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141: 1199–1207.
- Ehrenreich, I. M., N. Torabi, Y. Jia, J. Kent, S. Martis *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464(7291): 1039–1042.
- Ewens, W. J., 2004 *Mathematical Population Genetics 1: Theoretical Introduction* (Interdisciplinary Applied Mathematics, Vol. 1), Ed. 2. Springer-Verlag, Berlin/Heidelberg, Germany/New York.
- Falconer, D. S., 1992 Early selection experiments. *Annu. Rev. Genet.* 26: 1–14.
- Falconer, D. S., and T. F. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Benjamin-Cummings, Menlo Park, CA.
- Felsenstein, J., 1974 The evolutionary advantage of recombination. *Genetics* 78: 737–756.
- Felsenstein, J., 1987 Sex and the evolution of recombination, pp. 74–86 in *The Evolution of Sex: An Examination of Current Ideas*, edited by R. E. Michod, and B. R. Levin. Sinauer Associates, Sunderland, MA.
- Futschik, A., and C. Schlotterer, 2010 The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186: 207–218.
- Garland, T., and M. R. Rose (Editors), 2009 *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*. University of California Press, Berkeley, California.
- Gillespie, J. H., 1991 *The Causes of Molecular Evolution* (Oxford Series in Ecology and Evolution). Oxford University Press, London/New York/Oxford.
- Grossman, S. R., I. Shlyakhter, I. Shlyakhter, E. K. Karlsson, E. H. Byrne *et al.*, 2010 A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327(5967): 883–886.
- Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* 8(3): 269–294.
- Johansson, A. M., M. E. Pettersson, P. B. Siegel, and O. Carlborg, 2010 Genome-wide effects of long-term divergent selection. *PLoS Genet.* 6(11): e1001188.
- Keightley, P. D., and G. Bulfield, 1993 Detection of quantitative trait loci from frequency changes of marker alleles under selection. *Genet. Res.* 62(3): 195–203.
- Kessner, D., and J. Novembre, 2014 forqs: forward-in-time simulation of recombination, quantitative traits and selection. *Bioinformatics* 30(4): 576–577.
- Kessner, D., T. L. Turner, and J. Novembre, 2013 Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. *Mol. Biol. Evol.* 30(5): 1145–1158.
- Kim, Y., and W. Stephan, 1999 Allele frequency changes in artificial selection experiments: statistical power and precision of QTL mapping. *Genet. Res.* 73: 177–184.
- King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hooper, S. Sen *et al.*, 2012 Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* 22(8): 1558–1566.
- Kofler, R., and C. Schlotterer, 2014 A guide for the design of evolve and resequencing studies. *Mol. Biol. Evol.* 31(2): 474–483.
- Laurie, C. C., S. D. Chasalow, J. R. LeDeaux, R. McCarroll, D. Bush *et al.*, 2004 The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168: 2141–2155.
- Long, Q., D. C. Jeffares, Q. Zhang, K. Ye, V. Nizhynska *et al.*, 2011 Poolhap: inferring haplotype frequencies from pooled samples by next generation sequencing. *PLoS ONE* 6(1): e15292.
- Mackay, T. F., 2010 Mutations and quantitative genetic variation: lessons from *Drosophila*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365(1544): 1229–1239.
- Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482(7384): 173–178.
- Magwire, M. M., D. K. Fabian, H. Schweyen, C. Cao, B. Longdon *et al.*, 2012 Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLoS Genet.* 8(11): e1003057.
- Nuzhdin, S. V., L. G. Harshman, M. Zhou, and K. Harmon, 2007 Genome-enabled hitchhiking mapping identifies QTLs

- for stress resistance in natural *Drosophila*. *Heredity* 99(3): 313–321.
- Orozco-terWengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt *et al.*, 2012 Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol. Ecol.* 21(20): 4931–4941.
- Orr, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52(4): 935–949.
- Otto, S. P., and C. D. Jones, 2000 Detecting the undetected: estimating the total number of loci underlying a quantitative trait. *Genetics* 156: 2093–2107.
- Parts, L., F. A. Cubillos, J. Warringer, K. Jain, F. Salinas *et al.*, 2011 Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21(7): 1131–1138.
- Remolina, S. C., P. L. Chang, J. Leips, S. V. Nuzhdin, and K. A. Hughes, 2012 Genomic basis of aging and life-history evolution in *Drosophila melanogaster*. *Evolution* 66(11): 3390–3403.
- Rice, S. H., 2004 *Evolutionary Theory*. Sinauer Associates, Sunderland, MA.
- Teotonio, H., and I. M. Chelo, M. Bradić, M. R. Rose, and A. D. Long, 2009 Experimental evolution reveals natural selection on standing genetic variation. *Nat. Genet.* 41(2): 251–257.
- Thornton, K. R., A. J. Foran, and A. D. Long, 2013 Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. *PLoS Genet.* 9(2): e1003258.
- Turner, T. L., and P. M. Miller, 2012 Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. *Genetics* 191: 633–642.
- Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet.* 7(3): e1001336.
- Utsunomiya, Y. T., A. M. Perez O'Brien, T. S. Sonstegard, C. P. Van Tassell, A. S. do Carmo *et al.*, 2013 Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS ONE* 8(5): e64280.

Communicating editor: E. A. Stone

# GENETICS

**Supporting Information**

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.175075/-/DC1>

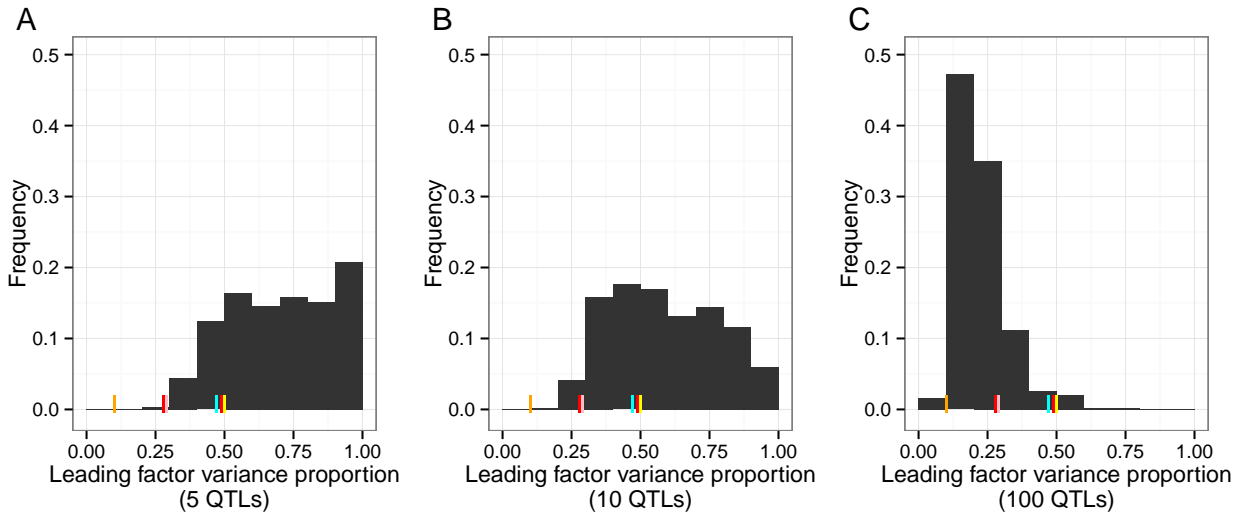
## **Power Analysis of Artificial Selection Experiments Using Efficient Whole Genome Simulation of Quantitative Traits**

**Darren Kessner and John Novembre**

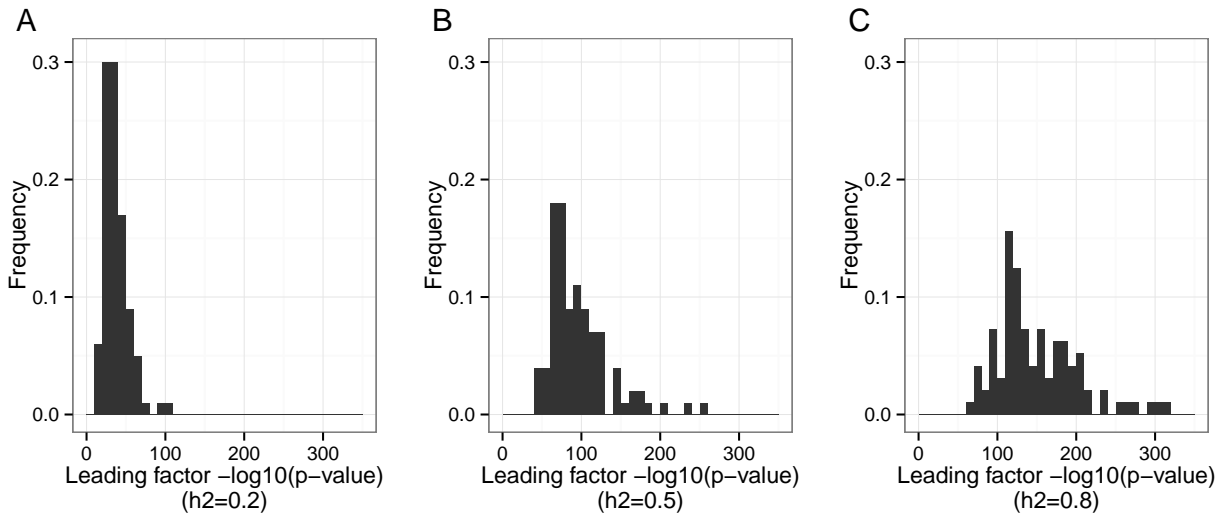


# Supplementary Information: Power analysis of artificial selection experiments

Darren Kessner and John Novembre



**Figure S1. Proportion of variance explained by leading factor.** Above are histograms of the proportion of variance explained by the QTL of largest effect, for random traits generated based on the DGRP SNPs and exponentially distributed effect sizes. Shown are histograms for random traits with A) 5 QTLs, B) 10 QTLs, and C) 100 QTLs (500 random trait simulations each). For comparison, we have included colored segments to indicate the proportion of variance explained by the leading factor of several quantitative traits as reported by selected empirical studies. [*Adh* activity, 2 experimental lines (.49, .28, red) (King *et al.*, 2012); susceptibility to viral infection, Drosophila C virus (.47, cyan), Sigma virus (.29, pink) (Magwire *et al.*, 2012); larval nicotine resistance (.50, yellow) (Marriage *et al.*, 2014); bristle number (.1, orange) (Mackay and Lyman, 2005)]. Note that the calculation of variance proportion assumes that QTLs segregate independently in the population (no linkage disequilibrium), which is generally not true for natural populations nor for the simulated populations.



**Figure S2. The  $-\log_{10}(\text{p-values})$  for leading factors in simulated GWASs using the DGRP lines.** Above are histograms of the  $-\log_{10}(\text{p-value})$  for the QTL of largest effect, for random traits generated based on the DGRP SNPs, 100 randomly chosen QTL loci with exponentially distributed effect sizes, and the specified heritability value. We assume a GWAS is conducted using the DGRP founder lines with 20 replicate flies per line and using a standard ANOVA for computing p-values. Shown are histograms for random traits with A)  $h^2 = 0.2$ , B)  $h^2 = 0.5$ , and C)  $h^2 = 0.8$  (100 random trait simulations each). (For smaller numbers of QTLs [not shown], the leading factor  $-\log_{10}(\text{p-value})$ 's are larger than those shown here). The results show that under the assumed genetic architecture used in this study the leading factor should be readily detectable by a GWAS unless heritability is low. In practical cases, where a leading effect is not detected in a GWAS for a trait with modest or high heritability, it may be because the leading effect is smaller and/or at lower frequency than the simulated values used here. Together with Supplementary Figure S1 these results highlight that for some traits the results of this simulation study may not be applicable.

## References

- King, E. G., Merkes, C. M., McNeil, C. L., Hooper, S. R., Sen, S., Broman, K. W., Long, A. D., and Macdonald, S. J. (2012). Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.*, **22**(8), 1558–1566.
- Mackay, T. F. and Lyman, R. F. (2005). *Drosophila* bristles and the nature of quantitative genetic variation. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **360**(1459), 1513–1527.
- Magwire, M. M., Fabian, D. K., Schweyen, H., Cao, C., Longdon, B., Bayer, F., and Jiggins, F. M. (2012). Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLoS Genet.*, **8**(11), e1003057.
- Marriage, T. N., King, E. G., Long, A. D., and Macdonald, S. J. (2014). Fine-mapping nicotine resistance loci in *Drosophila* using a multiparent advanced generation inter-cross population. *Genetics*, **198**(1), 45–57.