Natural selection and the probability of parallel genetic evolution

Journal:	*Journal of Evolutionary Biology*
Manuscript ID	JEB-2016-00117
Manuscript Type:	Research Papers
Keywords:	Theory, Ecological genetics, Adaptation, Natural selection

Abstract:

Parallel evolution is often assumed to result from repeated adaptation to novel, yet ecologically similar, environments. Here we develop and analyze a mathematical model that predicts the probability of parallel genetic evolution as a function of the strength of phenotypic selection and constraints imposed by genetic architecture. Our results show that the probability of parallel genetic evolution increases with the strength of natural selection and effective population size, and is particularly likely to occur for genes with large phenotypic effects. Building on these results, we develop a Bayesian framework for estimating the strength of parallel phenotypic selection from genetic data. Using extensive individual based simulations, we show that our Bayesian estimator is robust across a wide range of genetic and evolutionary scenarios and provides a useful tool for rigorously testing the hypothesis that parallel genetic evolution is the result of adaptive evolution.

Key Words: Bayesian, QTL, ecological selection, genetic architecture, ecological genetics, adaptation

Introduction:

As the availability of genome sequences has increased, interest in understanding how genomic architecture shapes adaptation has grown substantially (Stapley et al., 2010). How and which genes respond to selection is a complex result of many factors, including allelic effect sizes, epistatic interactions, linkage disequilibrium, and pleiotropy. Significant work using natural populations (Nadeau & Jiggins, 2010), experimental evolution (Qi et al., 2016, Wichman et al., 1999), and evolutionary theory has been devoted to elucidating how these many factors interact to shape adaptation. One particularly useful natural system for addressing such questions are those exhibiting parallel evolution. Characterized by repeated phenotypic adaptation in isolated populations in response to similar environmental conditions, parallel evolution (Hohenlohe et al., 2010) offers natural experimental replicates for studying genetic adaptation.

Developing theoretical predictions is key to understanding how the many facets of genomic architecture shape adaptation. Two main theoretical studies have addressed how genetic factors affect the probability of parallel evolution. Modeling a single gene under strong selection, Orr (2005), calculated the probability that 1 of k possible alleles arises from mutation and fixes repeatedly under strong selection. His results highlight two key factors, selection and genetic constraint, that have long been hypothesized to play important roles in repeated gene use. Namely, gene reuse increases with the strength of selection, and decreases with the number of possible alleles, k, at the locus. These results are supported by experimental adaptation of the bacteriophage $\phi X174$ to high temperatures (Wichman et al., 1999) as well as in the adaptation of antifungal drug resistance in *Saccharomyces cerevisiae* (Anderson et al., 2003)

The second theoretical study by Chevin et al. (2010) extends Orr's model in two important ways. First they generalize the assumption of strong selection. Using a genetically explicit version of Fisher's

geometric model they calculate the probability of parallel genetic evolution under multi-variate stabilizing selection. This extension demonstrates several additional factors contributing to parallel evolution: the extent of pleiotropy and the curvature of the fitness surface. Specifically, parallel evolution is most likely to occur when pleiotropy is weak. In addition, contrary to Orr's conclusion, when pleiotropy is strong and the fitness surface curved, parallel evolution may decrease with the strength of selection. Their second extension of Orr's work is by placing his conclusions about genetic constraint in a genomic context. Calculating the probability of repeated gene use given a set of potential loci each with multiple possible alleles, they once again demonstrate that genetic constraint is an important factor in parallel genetic evolution. Importantly however, the probability of parallel evolution depends on the distribution of allelic effects at any given locus as well as the genome wide distribution of effects.

The results of Orr (2005) and Chevin et al. (2010) help us to understand the many interconnected factors that contribute to repeated gene use in response to selection. Their results are however, difficult to apply to the type of data regularly collected within natural populations. In natural systems parallel genetic evolution is characterized by one or more gene substitutions appearing independently with limited information on the order in which these mutations arose. As the same substitutions may have occurred in different genetic backgrounds, and hence at different points on the fitness surface, there are limitations to theoretical predictions which hold genetic background and selection constant. A complementary approach, and one that is critical for understanding how selection and genetic architecture shape gene use, is to assess the probability of parallel genetic evolution at multiple loci over the course of adaptation.

This work extends Orr's and Chevin's results to assess parallel genetic evolution in natural populations. The focus of this paper is two-fold; first we will develop a multi-locus theory of parallel adaptation that allows us to predict the probability of parallel evolution in terms of quantities that are

regularly measured in natural populations. Using this model we analyze how several key factors, in particular the allelic effect size distribution, shapes the probability of parallel evolution. The second focus of our paper is to demonstrate how this theory can be applied to natural populations. To this end, we develop a framework for estimating the strength of natural selection driving parallel adaptation using available genetic data. We use this result to make recommendations on experimental designs for studying parallel genetic evolution.

The Model

Biological Scenario:

We envision a scenario (see Figure 1A) where individuals from an ancestral population colonize two or more novel environments and establish new populations. After this initial colonization we assume gene flow between the ancestral and descendent populations is negligible and that individuals within populations mate at random. The descendent populations then experience identical patterns of phenotypic selection that cause population mean phenotypes to diverge in parallel from the ancestral population; for example the repeated reduction in body armor in freshwater sticklebacks from their common marine ancestors (Colosimo et al., 2004). Next we envision that the genetic basis of parallel phenotypic evolution is studied using one of two commonly used experimental designs (Conte et al., 2012).

Figure 1B illustrates the first experimental design where parallel genetic evolution is assessed at a set of candidate genes. To identify possible candidate genes, individuals from at least one descendent population (descendent population 1 in Figure 1B) are crossed with ancestral individuals and the resulting offspring are scanned for divergent QTL's. The remaining populations (descendent population 2 in Figure 1B) are then tested for the candidate genes using a variety of approaches such as genetic complementation tests. This method, which we will call the candidate gene test, has been used in

human populations to identify the genetic basis of the multiple independent origins of lactose tolerance (Ingram et al., 2009, Enattah et al., 2008, Tishkoff et al., 2007). Alternatively, in the second design (shown in Figure 1C), descendent populations are searched independently for the genes responsible for the repeated phenotypic divergence from the ancestral population. This is done by performing independent QTL scans in each descendent population. This "QTL method" was used to identify separate genes responsible for a change in developmental rate in two populations of *Oncorhyncus mykiss,* (Robison et al., 2001, Nichols et al., 2007, Sundin et al., 2005).

Analytical Model

Our model assumes the trait experiencing parallel selection is controlled by n additive loci. Each locus, i, has two possible alleles A_i and a_i and a phenotypic effect equal to b_i, such that the phenotype of an individual is described by:

$$z = \bar{z} + \sum_{i=1}^{n} b_i(X_i - p_i) \quad (1)$$

where X_i is an indicator variable taking the value 1 if the individual carries the A allele at locus i and the value 0 if the individual carries the a allele at locus i, p_i is the frequency of the A allele at locus i, and $\bar{z}$ is the average phenotype of the population. We assume that the phenotype of the ancestral population is small, meaning that the frequency of the A allele is low at all loci, and equal to p_{0_i}. Within the new environments, individuals experience selection for large phenotypes, favoring an increase in frequency of the A alleles. Specifically, we assume the relationship between an individual's phenotype and its fitness, $W(z)$, is linear and described by the expression:

$$W(z) = \beta z + \alpha \quad (2)$$

where the parameters β and α represent the slope and intercept of the fitness surface respectively.

To make our multi-locus model analytically tractable we further assume that the strength of linear directional selection is weak $\mathcal{O}(\epsilon)$ and the rate of recombination between loci is relatively high.

Under these assumptions, linkage disequilibrium changes very quickly relative to allele frequencies, and a quasi-linkage-equilibrium (QLE) is reached where linkage disequilibrium is small, also $\mathcal{O}(\epsilon)$ (Nagylaki, 1993, Nagylaki et al., 1999). Using the expression for the phenotypic trait z, given in equation (1), as well as the expression for fitness, given by equation (2), we can use the multi-locus methods developed by Barton and Turelli (1991) and expanded by Kirkpatrick et al. (2002) to derive the change in the frequency of the A_i allele at QLE over a single generation (*See supplementary material*):

$$\Delta p_i = \frac{\beta}{\alpha} b_i p_i (1 - p_i) \quad (3)$$

Equation (3) reveals that under conditions of linear directional selection and QLE, the effects of epistasis and linkage disequilibrium are small and hence loci evolve independently; an assumption we will later relax with individual based simulations.

The independent evolution of loci enables us to utilize a classic result of the Wright-Fisher model describing the probability of fixation for an allele with initial frequency p_o in a population of constant size N. This probability is given by:

$$P_{fix} = \frac{e^{2Ns(1-p_0)}\left(e^{2Nsp_o}-1\right)}{e^{2Ns}-1} \quad (4)$$

(Kimura, 1957, Karlin & Taylor, 1981). In equation (4) s is the strength of selection acting on the allele, and p_0 is its initial frequency. Under our assumption of linear directional selection, $s = \frac{\beta}{\alpha} b_i$, equation (5) can be re-written as:

$$P_{fix}(i) = \frac{e^{2N\frac{\beta}{\alpha}b_i(1-p_{o_i})}\left(e^{2N\frac{\beta}{\alpha}b_i p_{o_i}}-1\right)}{e^{2N\frac{\beta}{\alpha}b_i}-1} \quad . \quad (5)$$

Equation (5) reveals that the probability of fixation depends on initial allele frequency, local population size, the strength of phenotypic selection, and the phenotypic effect of the locus. In the next section we will use this result to explore how these important parameters influence the extent of parallel evolution.

The probability of parallel genetic evolution at a single locus:

We begin by analyzing the simplest possible scenario: a single genetic locus. For this simple case, parallel adaptation entails the repeated fixation of the same allele in multiple descendent populations. The probability of this occurring can be calculated by using equation (5) to express the probability that the A_i allele fixes independently in each of m populations:

$$P_{\parallel} = \left(\frac{e^{2N\frac{\beta}{\alpha}b_i\left(1-p_{o_i}\right)}\left(e^{2N\frac{\beta}{\alpha}b_i p_{o_i}}-1\right)}{e^{2N\frac{\beta}{\alpha}b_i}-1} \right)^m \quad (6)$$

Equation (6) highlights four important factors that will influence the probability of observing parallel genetic evolution. First, (6) shows that the probability of repeated fixation of an allele increases with its initial frequency, p_{o_i}. Second, (6) reveals that large effect alleles, those with large b_i, are more likely to fix in parallel. This relationship between effect size and parallel evolution is shown in Figure 2. Third, equation (6) shows that parallel evolution is more likely to occur when evolution is driven more by the deterministic force of natural selection than the stochastic force of random genetic drift. Specifically, the probability of parallel evolution increases with the product of population size and the phenotypic selection gradient in derived populations, $N\frac{\beta}{\alpha}$. This product, which we will denote collectively as η, captures the balance between drift and selection and shows that parallel evolution is more likely in large populations experiencing strong natural selection as shown by the three curves in figure 2.

The probability of parallel genetic evolution at multiple loci:

Although the single locus results of the previous section are insightful, they fall short of capturing the genetic richness of real populations where the extent of parallel evolution must be assessed across multiple loci. Fortunately, calculating the probability of parallel evolution across multiple loci is straightforward, and yields the following formula:

$$P_{\parallel} = \prod_{i=1}^{n} \left(\frac{e^{2N\frac{\beta}{\alpha}b_i\left(1-p_{o_i}\right)}\left(e^{2N\frac{\beta}{\alpha}b_i p_{o_i}}-1\right)}{e^{2N\frac{\beta}{\alpha}b_i}-1} \right)^{m} \quad (7)$$

where the product is carried over the number of loci. Not surprisingly, equation (7) shows that the factors enhancing the probability of parallel evolution at a single locus (e.g., large population size, strong selection, etc.) also increase the probability of parallel evolution across multiple loci. In addition, equation (7) yields several novel insights that emerge only when multiple loci are considered.

The first, and most obvious, insight to emerge from (7) is that perfectly parallel genetic evolution, where all loci are fixed for the selectively favored A_i alleles in all descendent populations, becomes less and less likely as the number of loci increases. This is a simple result of the product rule of probabilities, and arises because the overall probability of parallel evolution decreases as each additional locus is required to fix in parallel in the m descendent populations. The second insight that emerges from equation (7) is that when selection is relatively weak, population sizes are relatively small, and adaptive alleles initially infrequent, it is quite surprising to observe parallel evolution at anything other than a single locus with large phenotypic effects (Figure 3). As selection becomes stronger, population sizes larger, or adaptive alleles initially more frequent, however, it becomes increasingly likely that parallel evolution will occur at multiple loci, including loci with only moderate phenotypic effects (Figure 3). These results are, for the most part, relatively insensitive to the particular distribution of effect sizes across loci except in cases of strong selection and high initial allele frequency (Figure 3). Together, these results suggest that the extent to which the observation of parallel genetic evolution at any particular number of loci is surprising depends heavily on the value of the parameter η.

Bayesian inferences of parallel phenotypic selection:

The results we derived in the previous section demonstrate a strong connection between the parameter η and the probability of observing parallel genetic evolution. In this section we develop a

method for rigorously estimating the value of this key parameter using a Bayesian framework that capitalizes on equation (7). Our goal is to provide a methodology that allows support for a hypothesis of adaptive parallel evolution to be rigorously assessed using data collected in empirical studies of parallel genetic evolution. Specifically, by estimating η it becomes possibly to rigorously distinguish between parallel genetic evolution caused by random genetic drift, $\eta = 0$, and parallel genetic evolution caused by adaptation.

Our Bayesian approach will rely on genetic data described by a matrix, $\mathcal{D}$, where rows represent descendent populations and columns loci. Each element of $\mathcal{D}$ takes a value of 0 or 1 depending on which allele has fixed at a particular locus in a given population (Figure 1A). Using equation (7) we can develop a likelihood function specifying the probability of observing the data, $\mathcal{D}$, given a particular value of the parameter η, and empirical estimates for the effect sizes b_i and initial allele frequencies p_0. Effect sizes can (and frequently are) estimated using QTL scans (Conte et al., 2015, Broman & Sen, 2009, Lynch & Walsh, 1998) and initial allele frequencies can be estimated by measuring the allele frequencies in the ancestral population. This likelihood expression consists of a product of terms, one for each locus in each population. If the A allele has fixed at a locus it contributes a term P_{fix}, as defined by equation (5). Alternatively, if the A allele is lost, it contributes a term $\left(1 - P_{fix}\right)$. Thus, for m populations and n loci, the likelihood of observing the data, $\mathcal{D}$, is given by the following product:

$$\mathcal{L}(\mathcal{D}) = \prod_{j=1}^{m} \prod_{i=1}^{n} P_{fix}(\eta, i)^{\mathcal{D}_{ij}} \left(1 - P_{fix}(\eta, i)\right)^{1-\mathcal{D}_{i,j}} \quad (8)$$

where i is an index over loci and j an index over populations.

The likelihood, (8), can be used to develop a Bayesian tool for estimating the posterior distribution of the key parameter η. Specifically, Bayes' theorem enables us to formulate estimates for η in the form of the posterior distribution $p(\eta|\mathcal{D})$ that is biologically meaningful for all possible genetic outcomes, $\mathcal{D}$:

$$p(\eta|\mathcal{D}) = \frac{\pi(\eta)\mathcal{L}(\mathcal{D})}{\int_{\eta}\mathcal{L}(\mathcal{D})} \quad (9)$$

where $\pi(\eta)$ is our prior distribution for the parameter η. The denominator of this expression is the integral over the likelihood surface, and cannot be easily evaluated. For this reason, we employ Markov Chain Monte Carlo simulation methods to sample from the posterior distribution and generate an estimate of the most probable value of η for the given genetic data $\mathcal{D}$. We label this estimate $\hat{\eta}$. We take two approaches to evaluating the performance of this Bayesian estimator. First we analyze its performance under the assumptions of the analytical model by generating the genetic data $\mathcal{D}$ using a Wright-Fisher model. Next, we test the robustness of our Bayesian estimator to violations of the assumptions of our analytical model by generating the genetic data $\mathcal{D}$ using multi-locus individual based simulations.

Wright-Fisher Simulation

We simulated the data $\mathcal{D}$ for two populations under the Wright-Fisher model by drawing a random number for each locus and population and setting $\mathcal{D}_{i,j}$ to 1 if the random number was less than p_{fix}, given by equation (5), and to 0 otherwise. The value of p_{fix} depends on the initial allele frequency at each locus,p_{0_i}, the allelic effect sizes of each locus, b_i, as well as the parameter η. For each simulation we drew the values of these parameters independently and at random. Initial allele frequencies ranged between 0 and 0.1 and were drawn independently for each locus. Because our model envisions divergence of descendent populations from a common ancestor we assumed that the initial frequency at any one locus was the same in both populations. Allelic effect sizes were drawn independently for each locus from a uniform distribution between 0 and 1. The value of η for each run was drawn from a uniform distribution ranging between 0 to 50. The genetic outcome $\mathcal{D}$ simulated in this manner may not, however, resemble what would be measured using experimental methods. For example, using current genomic techniques it is not possible to identify loci that have not diverged from

the ancestral state. To address how experimental methodologies effect our Bayesian estimates we considered two modified forms of $\mathcal{D}$ that resemble sampling under the two experimental methods described previously (see Figure 1). The first of these methods, the candidate gene method, (Figure 1B) assesses parallel genetic evolution at candidate genes which are known to have generated the phenotypic divergence in the first descendent population. This is often done by performing a cross between individuals from one of the divergent populations with the ancestral population and assessing the genetic variation in the F1's. Since the second divergent population is not independently assessed for divergent QTL's, under this method we only consider the columns of $\mathcal{D}$ (ie loci) where the A allele has fixed in the first population. The effective number of loci under this method is denoted by n_{CG}. The second experimental method, the QTL method, (Figure 1C) independently assesses divergent loci in all populations. $\mathcal{D}$ under this method therefore contains all columns (loci) which have fixed in at least one population. We denote the effective number of loci under this method by n_{QTL}.

For each simulated $\mathcal{D}$, as well as for $\mathcal{D}$ modified by the two experimental methods, we estimated η using a metropolis algorithm as described in the supplementary material. For the prior $\pi(\eta)$ we used a uniform distribution on the interval $\eta = \pm 60$. To analyze the performance of the Bayesian estimator we ran a regression of the estimated values of $\hat{\eta}$ on the true values η, using 200 data points. Overall, this analysis revealed that the Bayesian estimator was quite accurate, explaining between 30% and 60% of the variation (see Table S1). In addition, our analysis showed that the accuracy of the estimates increases with the number of loci. This trend holds regardless of the experimental method used. However, the effective number of loci under the candidate gene method is always greater than when QTL scans are carried out in all descendent populations. The results of these simulations suggest our estimator performs quite well when data meet the assumptions of our analytical model; however, this may not be the case for real data. In the next section, we explore the

performance of our estimator using individual based simulations that allow us to violate key assumptions of our analytical model such as weak selection and frequent recombination.

Individual Based Simulation

Our individual based simulations (IBS) consider two allopatric populations, each of which has a constant size of $N = 1000$ individuals. Initial allele frequencies and effect sizes at each locus, as well as the value of η, were drawn randomly as described above under the Wright-Fisher model. Individuals within each population undergo a two stage life cycle. During the first stage, "selection", the probability that an individual survives is given by its fitness, with fitness computed using either equation (2) which describes linear selection or an expression for stabilizing selection described below. Surviving individuals then enter the second life cycle stage, "reproduction", which consists of generating an offspring population from the remaining parental population. This is done by drawing a pair of parents at random from the pool of surviving individuals and producing an offspring from these parents by recombining the parental genomes at a specified rate r and allowing mutation between the two allelic states at a per locus mutation rate of $\mu = 10^{-6}$. This process is continued with replacement of parents until the offspring population reaches the pre-selection size of N. This life cycle is repeated until all loci approach fixation or loss (allele frequencies > 0.99 or < 0.01) at which point the simulations were terminated and the matrix of genetic data $\mathcal{D}$ filled by rounding. As in the previous section, we formulate modified versions of $\mathcal{D}$ that resemble sampling under the two experimental methods. Then, using the metropolis algorithm, we compute estimates for the value of η using the original outcome $\mathcal{D}$ as well as the two modified forms of $\mathcal{D}$ (*see Supplementary Material*).

We used the Individual based simulations to test the robustness of the Bayesian estimator when selection is strong and/or non-linear. To test the effect of non-linear selection, individual based simulations were run where an individual's fitness was determined by one of two alternative forms of

selection: linear directional selection described by (2), or stabilizing selection toward a phenotypic optimum:

$$W(z) = e^{-\gamma(z-\theta)^2} \quad (10)$$

where θ is the phenotypic optima and γ is the strength of stabilizing selection. Including simulations where selection is stabilizing is important because it relaxes our previous assumption that loci evolve independently. Stabilizing selection is particularly useful in testing this assumption because the extent of interdependence between loci can be manipulated by changing the value of the phenotypic optima, θ. Specifically, when θ is greater than the largest possible phenotype, z_{max}, loci remain relatively independent as directional selection predominates over epistatic selection. However when $\theta < z_{max}$ this is no longer true as epistatic selection now dominates. Therefore, when $\theta > z_{max}$ evolution is much more likely to resemble linear selection as our analytical model assumed. We simulated these two forms of stabilizing selection respectively, by either requiring that θ be larger than z_{max} or slightly smaller than z_{max} (*see Supplementary material*). As expected, analysis of simulated data shows that the accuracy of our Bayesian estimates depends on the form of selection. Specifically, estimates for η are most accurate under linear selection, somewhat less accurate under stabilizing selection toward a distant optimum($\theta > z_{max}$), and least accurate under stabilizing selection toward a close optimum ($\theta < z_{max}$) (See Figure 4 and Table S2). In addition to assuming that selection is linear we also assumed that selection is weak. By computing the variance about the regression line as η increased we were able to confirm that, for the data shown in figure 4, the accuracy of our estimates decreases with increasing selection. Next, we used our simulations to explore the sensitivity of our estimator to infrequent recombination among candidate loci (See Table S3). Not surprisingly, these simulations revealed that our Bayesian estimator performs better under free recombination ($r = 0.5$) than when recombination is constrained ($r = 0.05$). The effect of constrained recombination is more drastic for stabilizing selection

than linear selection, and is particularly pronounced when $\theta < z_{max}$. This is expected since this latter scenario generates the strongest epistatic selection and thus has the greatest potential to cause linkage disequilibrium to accumulate. Finally, we used the individual based simulations to test the accuracy of our Bayesian estimator when recurrent gene flow occurs between ancestral and descendent populations, when the strength of selection differs in the two descendent populations, and when estimates of the parameters p_0 and b_i are imprecise (*see Supplementary Material and Table S4-S6*). These simulations reveal that the Bayesian estimator is indeed robust to error in the estimates of these parameters, particularly for error in initial allele frequency p_0.

Up to this point we have focused on using the Bayesian estimator to provide single point estimates for η. What makes the Bayesian approach particularly powerful, however, is that it provides a posterior probability distribution for the key parameter, η. Having access to the full posterior distribution allows us to calculate a 95% credible interval for the parameter η and determine whether or not it overlaps with zero. From an empirical standpoint, being able to rule out $\eta = 0$ allows us to reject the hypothesis that observed levels of parallel genetic evolution resulted from random genetic drift alone. Figure 5 illustrates two such posterior probability distributions and accompanying 95% credible intervals. Only in 5A where the credible interval does not overlap $\eta = 0$ can we reject drift as the cause of parallel genetic evolution.

Discussion

It has long been understood that natural selection, parallel genetic evolution, and genomic architecture are inexorably linked (Orr, 2005, Schluter, 2009, Chevin et al., 2010). The models we study here formalize this connection over the course of adaptation. We began our investigation by calculating the probability of parallel evolution at a single locus and showed that parallel evolution is most likely when phenotypic selection is strong, standing genetic variation for adaptive alleles is appreciable, adaptive alleles have large phenotypic effects, and population sizes are large. Next, we extended our

315 analyses to multiple loci, demonstrating that the number of loci that evolve in parallel depends on the
316 product of phenotypic selection and local population size (η). If selection is relatively weak, or
317 population sizes small, we expect parallel evolution at no more than a single locus. In contrast, when
318 selection is relatively strong, or population sizes very large, parallel evolution may occur across multiple
319 loci. These results demonstrate that without information on the strength of phenotypic selection and
320 population size, we have no way to assess whether the amount of parallel genetic evolution we observe
321 in an empirical study is interesting, or just another piece of genetic natural history. To remedy this
322 problem, and better connect studies of parallel genetic evolution to the evolutionary processes they
323 imply, we developed a Bayesian approach that capitalizes on available genetic data to estimate the
324 product of phenotypic selection and local population size (η). In the following paragraphs we explore
325 several of the key results in more detail and discuss their implications for past, present, and future
326 studies of parallel genetic evolution.

327 The first important result that emerges from all of our models is that parallel evolution is most
328 likely to be observed at loci with large phenotypic effects on traits experiencing strong phenotypic
329 selection in novel environments. This result receives at least some support from empirical studies of
330 parallel genetic evolution. For example the large effect gene Eda has been found in eight fresh-water
331 descendent populations of three-spine stickleback, *Gasterosteus aculeatus,* and is largely responsible for
332 the parallel reduction in lateral plate number in these populations. In contrast the small effect locus LG7
333 has been confirmed in only two of the 8 descendent populations (Schluter et al., 2004, Colosimo et al.,
334 2004, Conte et al., 2012). It seems likely that this example — where a stark ecological shift from salt to
335 fresh water has occurred — corresponds to a case where natural selection is likely quite strong. Another
336 important result of equation (6), however, is that probability of parallel evolution at a single locus is a
337 function of both effect size and initial allele frequency. This may help explain why, contrary to the results
338 described above, a recent comprehensive survey of allelic effects involved in parallel adaptation in two

339 stickleback populations has found no correlation between effect size and probability of repeated gene
340 use (Conte et al., 2015).

341 By integrating multi-locus genetics into a model of adaptation, we were also able to derive
342 expressions for the probability of observing parallel genetic evolution at various numbers of loci over the
343 course of adaptation. The most important result to emerge from this analysis is that in the absence of
344 information about the likely strength of phenotypic selection in derived populations and the number of
345 individuals composing these derived populations, there is no way to assess the significance of observing
346 parallel evolution at any particular number of loci. Put differently, if natural selection in novel
347 environments is quite strong or population sizes in novel environments quite large, observing parallel
348 evolution at multiple genetic loci is not too terribly surprising. If, however, natural selection is weak or
349 population sizes very small, observing this same level of genetic parallelism would be rather unexpected.
350 This suggests that if we are to more rigorously interpret the results of empirical studies of parallel
351 genetic evolution, we must do better than simply counting up the number of parallel genetic changes
352 observed. Our Bayesian tool accomplishes this goal by providing a methodology for rigorously tying
353 information on the extent of parallel genetic information to underlying evolutionary processes.

354 Such a Bayesian estimator is only useful if it produces accurate predications across a range of
355 parameter space. Indeed, our individual based simulations reveal that our Bayesian estimator can be
356 quite accurate and robust although there are limitations. For example accurate estimation requires data
357 from at least 8 total loci, be that 4 loci in two populations, 2 loci in four populations, or some
358 intermediate combination. Whether data is gathered at fewer loci in many populations or many loci in
359 few populations should, in principle, have no effect on the accuracy of the estimator. Many of the
360 studies discussed above however, have far fewer than this. For example studies of parallel
361 pigmentation changes in a variety of species, from beach mice to cave fish, focus primarily on one or
362 two loci in somewhere between 2 and 6 populations. Therefore if future studies hope to understand

the role of natural selection in driving parallel evolution it is important that they focus on acquiring data from as many loci and as many populations as possible. A natural consequence of the increasing accuracy derived from information on a larger number of loci is that certain experimental approaches are more powerful than others for studying parallel evolution. Specifically, we have explored two alternative experimental methods (see Figure 1 B and C), that differ predictably in the number of loci that are detected. Because the QTL method always detects parallel evolution at a larger number of loci, we recommend its use over the candidate gene method. The accuracy of the Bayesian estimate is influenced not only be the amount of available data but the accuracy of the estimated parameter values b_i and p_{0_i}. We have tested the robustness of the estimate to error in these parameters (*See the supplementary Material*). Because we may often be uncertain about the exact value of p_{0_i} due to sampling error or stochastic variation in small populations (Hermisson & Pennings, 2005), in some case it may be more appropriate to run the Bayesian estimator with a prior distribution for the parameter p_{0_i} rather than a single point estimate.

In addition to requiring information on parallel genetic evolution drawn from a reasonably large number of loci, our Bayesian estimator relies on several assumptions that may affect its accuracy. For example, our approach assumes that population size is constant across time, and thus does not allow for extreme bottlenecks or extensive founder effects. This assumption may prove particularly important in cases of repeated evolution of reduced skin pigmentation in European and Asian human populations for which there is evidence for extensive bottlenecks (Schmegner et al., 2005, Amos & Hoffman, 2010). Finally, our approach assumes that selection/population size is identical in each population and that recurrent gene flow does not occur. Although these assumptions may ultimately prove important in some cases, our individual based simulations show that they have only a limited impact on the accuracy of estimates in most cases (*See Supplementary Material*).

386 Combined, our analyses of single and multi-locus models show that it is difficult to draw
387 conclusions about the biological significance of parallel genetic evolution without information on the
388 strength of parallel phenotypic selection and local population size. Our Bayesian estimator provides a
389 robust statistical methodology for translating observed levels of genetic parallelism into an estimate for
390 the product of phenotypic selection and local population size. As a consequence, our Bayesian approach
391 provides a much needed tool for distinguishing between adaptive and non-adaptive hypotheses for
392 observed levels of parallel genetic evolution. Applying our approach to existing and emerging data on
393 parallel genetic evolution may thus provide novel insights into the importance of adaptive evolution in
394 natural populations.
395

Acknowledgments:

We thank Paul Joyce, Richard Gomulkiewicz, Lyudmila Barannyk, and Paul Hohenlohe for their helpful suggestions and discussions on this work. Funding was provided by NSF grant DEB 1118947 and DEB 1450653 and University of Idaho IBEST.

References:

Amos, W. & Hoffman, J. I. 2010. Evidence that two main bottleneck events shaped modern human genetic diversity. *Proceedings of the Royal Society B-Biological Sciences* **277**: 131-137.

Anderson, J. B., Sirjusingh, C., Parsons, A. B., Boone, C., Wickens, C., Cowen, L. E. & Kohn, L. M. 2003. Mode of selection and experimental evolution of antifungal drug resistance in Saccharomyces cerevisiae. *Genetics* **163**: 1287-1298.

Barton, N. H. & Turelli, M. 1991. Natural and Sexual Selection on Many Loci. *Genetics* **127**: 229-255.

Broman, K. W. & Sen, S. 2009. *A guide to QTL mapping with R/qtl*. Springer, Dordrecht.

Chevin, L. M., Martin, G. & Lenormand, T. 2010. Fisher's Model and the Genomics of Adaptation: Restricted Pleiotropy, Heterogenous Mutation, and Parallel Evolution. *Evolution* **64**: 3213-3231.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D. & Kingsley, D. M. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLOS Biology* **2**: 635-641.

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., Schluter, D. & Peichel, C. L. 2015. Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. *Genetics*.

Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B-Biological Sciences* **279**: 5039-5047.

Enattah, N. S., Jensen, T. G. K., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., El-Shanti, H., Seo, J. K., Alifrangis, M., Khalil, I. F., Natah, A., Ali, A., Natah, S., Comas, D., Mehdi, S. Q., Groop, L., Vestergaard, E. M., Imtiaz, F., Rashed, M. S., Meyer, B., Troelsen, J. & Peltonen, L. 2008. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *American Journal of Human Genetics* **82**: 57-72.

Hermisson, J. & Pennings, P. S. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**: 2335-2352.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. & Cresko, W. A. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *Plos Genetics* **6**.

Ingram, C. J. E., Mulcare, C. A., Itan, Y., Thomas, M. G. & Swallow, D. M. 2009. Lactose digestion and the evolutionary genetics of lactase persistence. *Human Genetics* **124**: 579-591.

Karlin, S. & Taylor, H. M. 1981. *A second course in stochastic processes*. Academic Press, New York.

Kimura, M. 1957. Some Problems of Stochastic-Processes in Genetics. *Annals of Mathematical Statistics* **28**: 882-901.

Kirkpatrick, M., Johnson, T. & Barton, N. 2002. General models of multilocus evolution. *Genetics* **161**: 1727-1750.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Mass.

Nadeau, N. J. & Jiggins, C. D. 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics* **26**: 484-492.
Nagylaki, T. 1993. The Evolution of Multilocus Systems under Weak Selection. *Genetics* **134**: 627-647.
Nagylaki, T., Hofbauer, J. & Brunovsky, P. 1999. Convergence of multilocus systems under weak epistasis or weak selection. *Journal of Mathematical Biology* **38**: 103-133.
Nichols, K. M., Broman, K. W., Sundin, K., Young, J. M., Wheeler, P. A. & Thorgaard, G. H. 2007. Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in Oncorhynchus myhiss. *Genetics* **175**: 335-347.
Orr, H. A. 2005. The probability of parallel evolution. *Evolution* **59**: 216-220.
Qi, Q., Toll-Riera, M., Heilbron, K., Preston, G. M. & MacLean, R. C. 2016. The genomic basis of adaptation to the fitness cost of rifampicin resistance in Pseudomonas aeruginosa. *Proceedings of the Royal Society B-Biological Sciences* **283**.
Robison, B. D., Wheeler, P. A., Sundin, K., Sikka, P. & Thorgaard, G. H. 2001. Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (Oncorhynchus mykiss). *Journal of Heredity* **92**: 16-22.
Schluter, D. 2009. Evidence for Ecological Speciation and Its Alternative. *Science* **323**: 737-741.
Schluter, D., Clifford, E. A., Nemethy, M. & McKinnon, J. S. 2004. Parallel evolution and inheritance of quantitative traits. *American Naturalist* **163**: 809-822.
Schmegner, C., Hoegel, J., Vogel, W. & Assum, G. 2005. Genetic variability in a genomic region with long-range linkage disequilibrium reveals traces of a bottleneck in the history of the European population. *Human Genetics* **118**: 276-286.
Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P. & Slate, J. 2010. Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**: 705-712.
Sundin, K., Brown, K. H., Drew, R. E., Nichols, K. M., Wheeler, P. A. & Thorgaard, G. H. 2005. Genetic analysis of a development rate QTL in backcrosses of clonal rainbow trout, Oncorhynchus mykiss. *Aquaculture* **247**: 75-83.
Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. & Deloukas, P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* **39**: 31-40.
Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. 1999. Different trajectories of parallel evolution during viral adaptation. *Science* **285**: 422-424.

Figure Legends:

Figure 1: Schematic of biological scenario. Panel A depicts two descendent populations diverging in parallel from a common ancestral population. The a allele predominates at all four loci in the ancestral population whereas the A allele fixes at various loci in the two descendent populations. Panel B and C depict two methods for deducing the underlying genetics of reduced body size in the two descendent populations depicted in panel A. Panel B shows the GC method which relies on a genome wide scan of

progeny from a cross between the first descendent population and the ancestral population and subsequent candidate gene search in the second descendent population. Panel C shows the GG method which involves two genome wide scans, one in each population. Compared to the GC method, the GG method uncovers an additional locus driving divergence in population 2.

Figure 2: The probability of parallel evolution as a function of allelic effect size, $\boldsymbol{b}$. For a given strength of selection the probability of fixation, and hence parallel evolution, increases with allelic effect size. The rate of increase is non-linear and depends on the strength of selection s and the population size N which are given by the compound parameter $\eta = Ns$. The initial alele frequency for the three curves was held constant at $p_0 = 0.01$.

Figure 3: The probability of parallel evolution at $\boldsymbol{n}$ loci. For a trait determined by the effects of 10 total loci, the probability of observing parallel evolution at exactly n loci, depends on the strength of selection, which varies from near 0 to a value of $\eta = 50$ and the intial allele frequency which is either low (0.01) or high (0.1). The probability of parallel evolution may also depend on the underlying effect size distribution dipicted here as three different gamma probability distributions with different shape and scale parameters (Red: $k = 1, \theta = \frac{1}{2}$, Blue: $k = 2, \theta = \frac{1}{4}$, Green: $k = 5, \theta = \frac{1}{5}$,) but with the same mean effect size ($\mu = \frac{1}{2}$).

Figure 4: Regression Fit of IBS Data under the three forms of selection. Data and linear regression fit (dashed), and perfect fit (solid) between the time averaged values of η and the Bayesian estimate $\hat{\eta}$ for 200 replicates of the individual based simulation. Sub-panels differ in the number of loci (ranging form 2 to 8) and the form of natural selection (Linear, Stabilizing with $\theta > z_{max}$, and Stabilizing with $\theta < z_{max}$). The Bayesian estimator uses genetic data filtered to resemble sampling using the GG experimental method.

499 **Figure 5: Posterior Probabiltiy distribution and credible intervals.** Posterior probabiltiy distributions
500 simulated for two genetic data sets D at four loci, given the initial allele frquencies p_0, and effect sizes b.
501 Shading the 95% credible interval for the estimate of η alows us to identify when we can and cannot
502 reject drift, η=0, 5A and 5B respectively.

Figure 1:

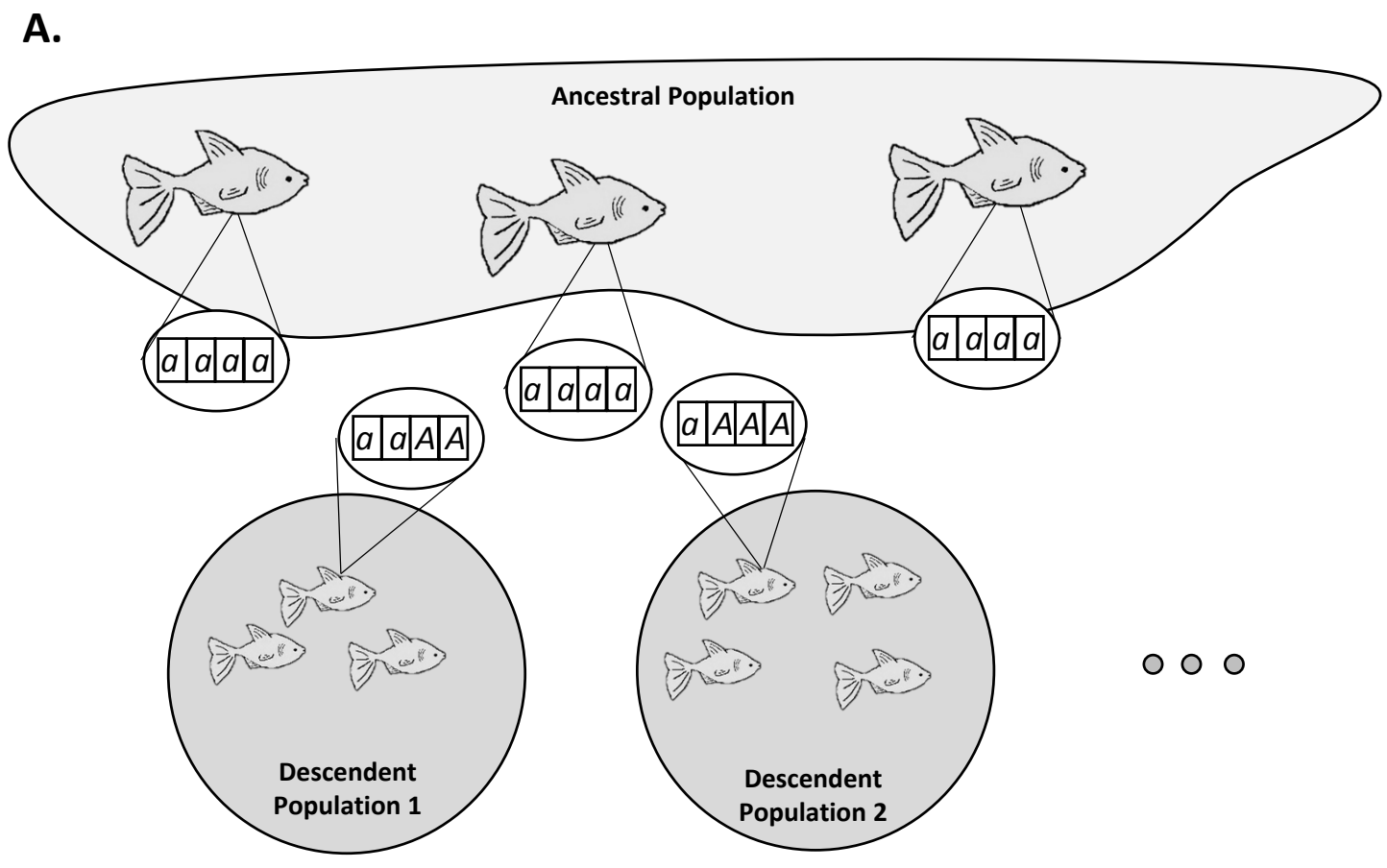

A.
Ancestral Population
a a a a
a a a a
a a a a
a a A A
a A A A
Descendent Population 1
Descendent Population 2
B.
Candidate Gene experimental method:
GWAS of Population 1xAncestral cross
Observed Alleles
X
A A
Candidate gene scan in Population 2
A A
A A
C.
QTL experimental method:
GWAS of Population 2xAncestral cross
Observed Alleles
X
A A A
Lost allele becomes apparent through comparison
GWAS of Population 1xAncestral cross
X
a A A

Figure 2:

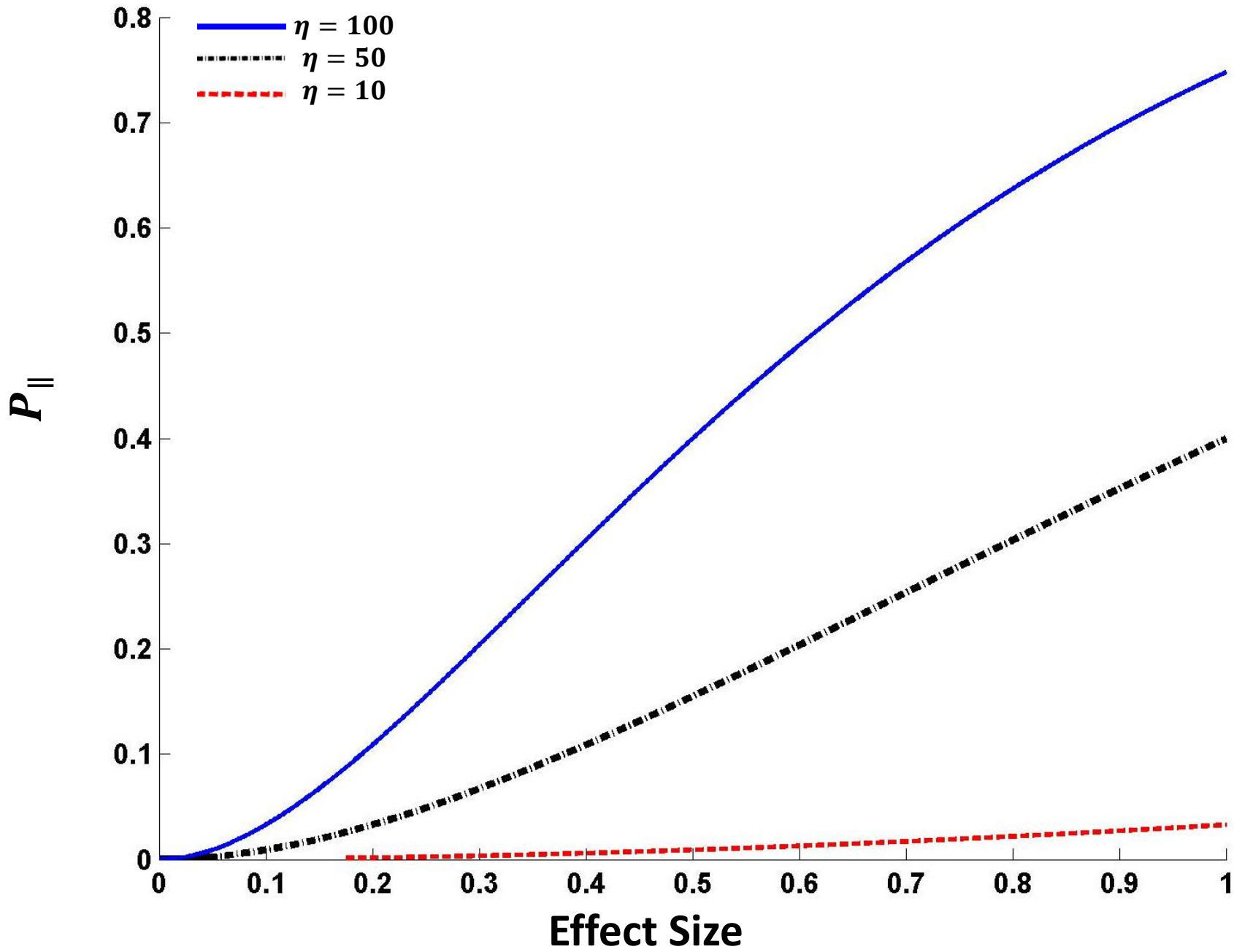

$\eta = 100$
$\eta = 50$
$\eta = 10$
$P_{=}$
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Effect Size

Figure 3:

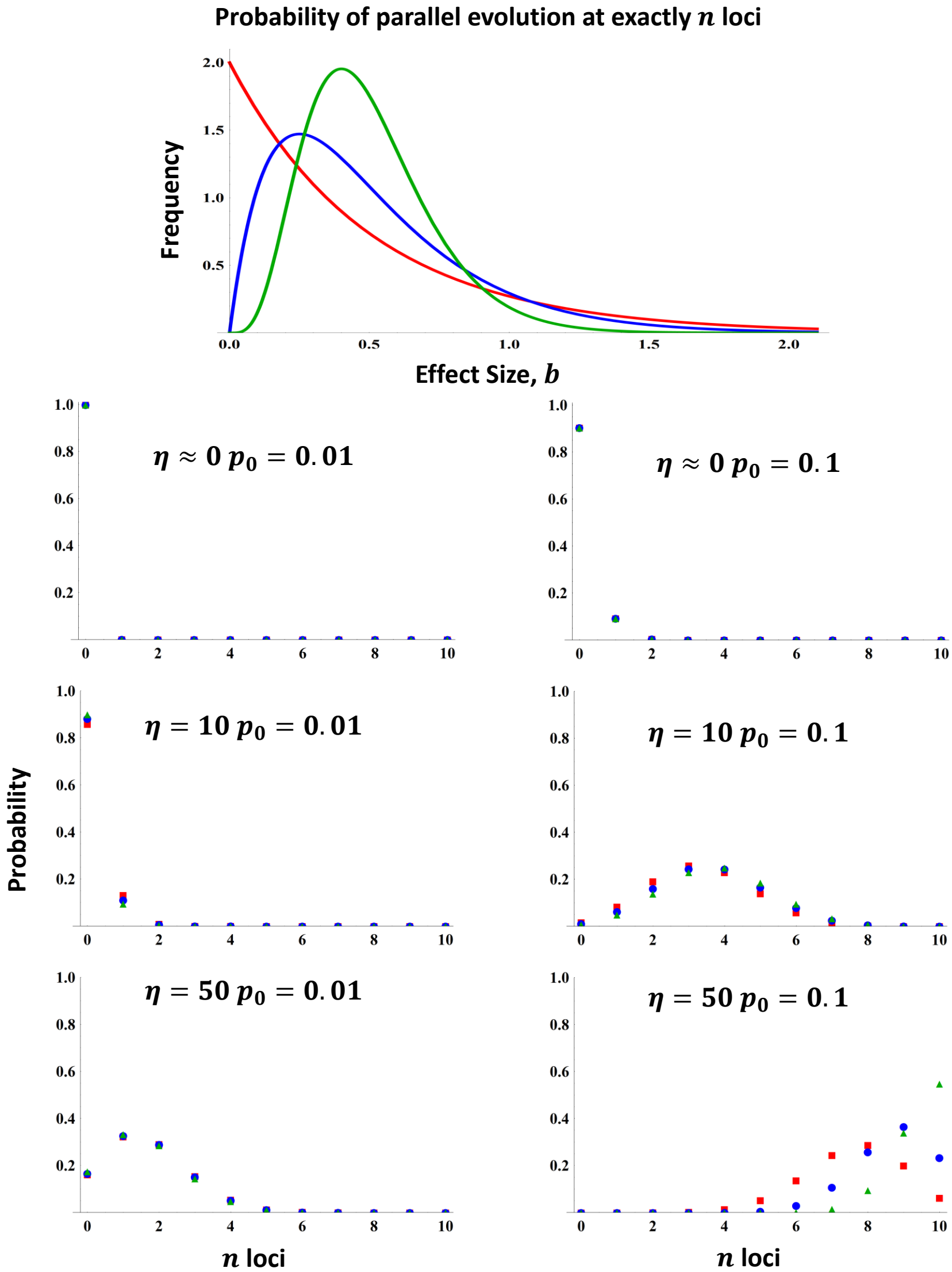

Probability of parallel evolution at exactly n loci
Frequency
Effect Size, b
$\eta \approx 0\ p_0 = 0.01$
$\eta \approx 0\ p_0 = 0.1$
$\eta = 10\ p_0 = 0.01$
$\eta = 10\ p_0 = 0.1$
$\eta = 50\ p_0 = 0.01$
$\eta = 50\ p_0 = 0.1$
Probability
n loci
n loci

Figure 4:

Time Averaged vs. MCMC Estimate for η using the GG experimental method

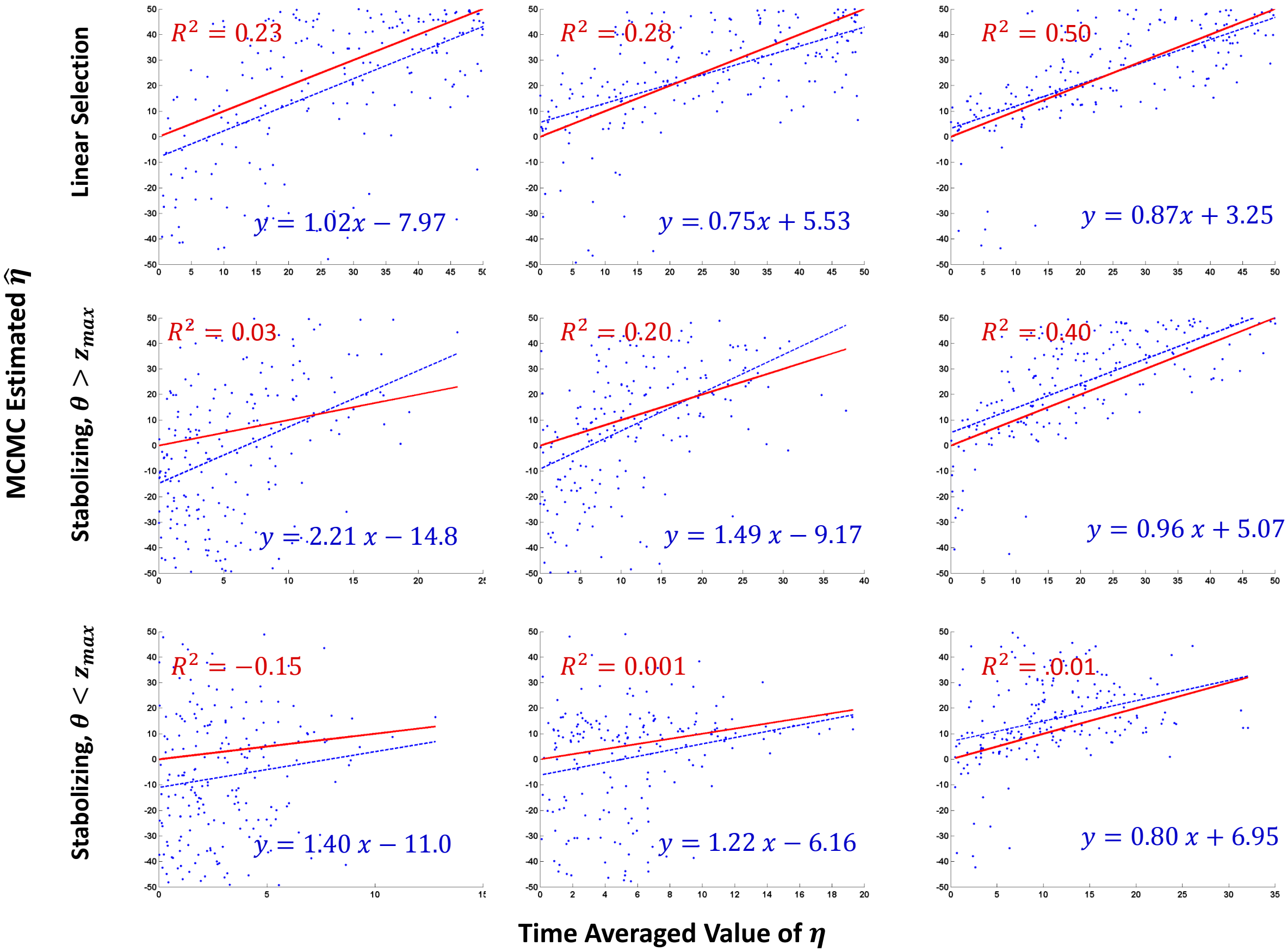
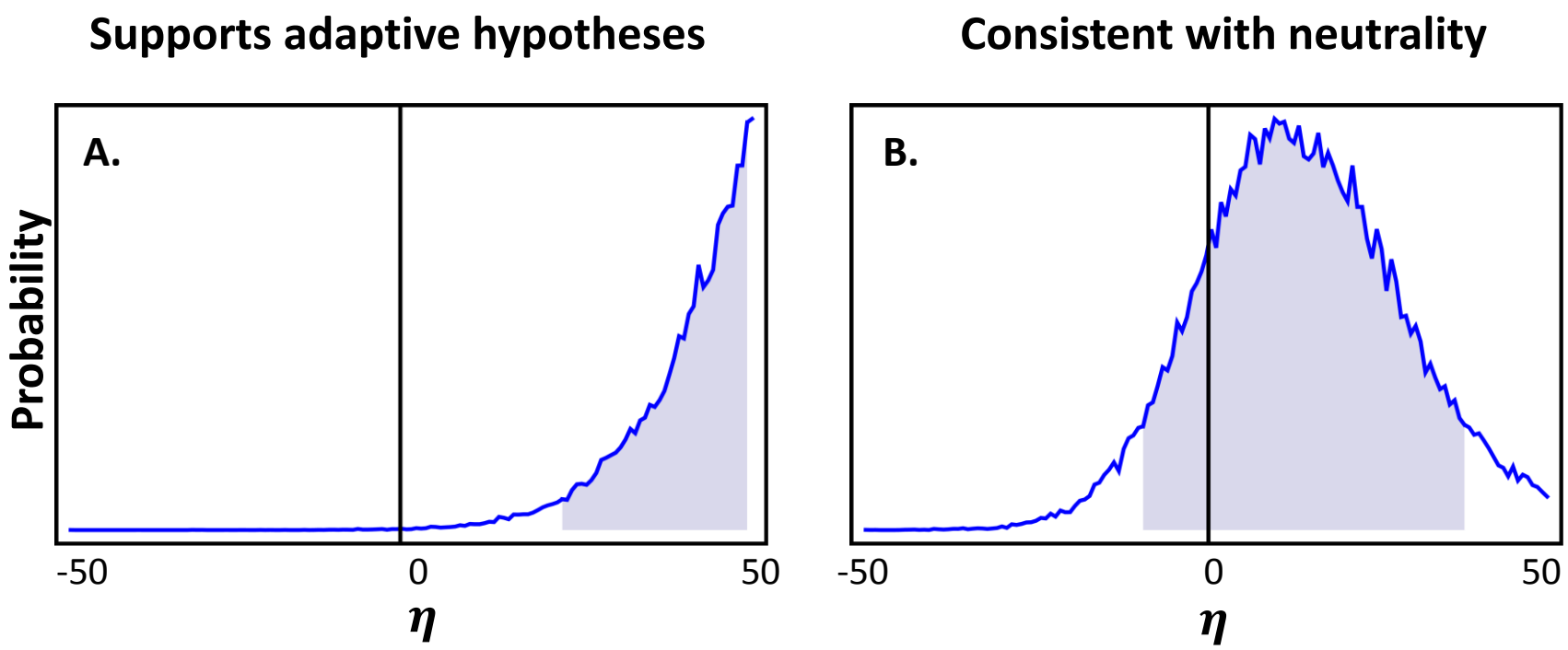


Figure 5:


Supports adaptive hypotheses
Consistent with neutrality
A.
B.
Probability
-50
0
50
η
-50
0
50
η

Supplementary Material:

A: *Evolution of Allele Frequencies and Linkage Disequilibrium under directional selection*

As discussed in the main text, our model studies parallel phenotypic evolution in two or more descendent populations which were colonized by a single common ancestral population (Figure 1). Within the ancestral population, selection is assumed to favor small values of the phenotype, z; larger values of the phenotype z are favored in the descendent populations. We further assume the phenotype z is determined by the additive action of n diallelic loci with alleles A and a such that:

$$z = \bar{z} + \sum_{i=1}^{n} b_i \zeta_i + e_z \quad \text{(S1)}$$

where $\bar{z}$ is the average phenotype of the population, b_i is the effect of the A allele relative to the a allele, $\zeta_i = (X_i - p_i)$ where X_i is an indicator variable which takes on the value 1 if the individual carries the A_i allele and a value of 0 if it carries the a_i allele, and p_i is the allele frequency of the A allele. The variable e_z describes the random environmental component of the phenotype. For simplicity we assume there is no environment effect and hence $e_z = 0$. Given this simplification equation (S1) reduces to equation (1) of the main text. Because we assume selection favors a small value of the phenotype, z, in the ancestral population and ignore mutation, allele frequencies, p, within the ancestral population will be near zero. We assume that the approximate values of these allele frequencies are known.

We begin our analysis by focusing on the simplest possible selective scenario capable of generating parallel phenotypic evolution: directional selection of identical strength within each of the descendent populations. Specifically, we assume that selection is linear such that absolute fitness as a function of the phenotype is given by the line:

$$W(z) = \beta z + \alpha \quad \text{(S2)}$$

Here β and α describe the slope and intercept of the selection surface respectively. Equation (S2) is the same as equation (2) of the main text. Averaging (S2) over individuals gives us the following expression for the average fitness of a population.

$$\bar{W} = \beta\bar{z} + \alpha$$

Where $\bar{z}$ is the average phenotype of the population as defined in equation (S1). The relative fitness of an individual with phenotype z is given by the ratio $w(z) = \frac{W(z)}{\bar{W}}$. The resulting expression for relative fitness is simplified by assuming that selection is weak and Taylor expanding about $\beta = 0$:

$$w(z) = \frac{\beta z+\alpha}{\beta\bar{z}+\alpha} \approx 1 + \frac{\beta}{\alpha}(z - \bar{z}) + \mathcal{O}(\beta^2) \quad \text{(S3)}$$

Substituting in the definition of the trait z from equation (S1) we get an expression for relative fitness as a function of the individual effect of each locus:

$$w(z) = 1 + \sum_{i=1}^{n} \frac{\beta}{\alpha} b_i \zeta_i \quad \text{(S4)}$$

Expression (S4) is very useful as it allows us to determine the selection on each locus. To do so we begin with equation (7) from Kirkpatrick et al. (2002) which gives a general expression for relative fitness:

$$w(z) = 1 + \sum_{i}^{n} a_i(\zeta_i) + \sum_{i}^{n} \sum_{j<i}^{n} a_{i,j}\left(\zeta_{ij} - D_{ij}\right) + \cdots \quad \text{(S5)}$$

where D_{ij} is the linkage disequilibrium between the alleles at the i^{th} and j^{th} loci, a_i is the selection coefficient on the A_i allele, and $a_{i,j}$ is a selection coefficient describing epistatic selection acting on the combination of the A_i and A_j alleles. We can solve for these selection coefficients (the $a's$) by comparing like terms between (S4) and (S5). This reveals that:

$$a_i = \frac{\beta}{\alpha} b_i \quad and \quad a_{i,j} = 0 \quad \text{(S6)}$$

With these selection coefficients in hand we can describe the change in the allele frequency of the A_i allele over a single generation using equation (10) from Kirkpatrick et al. (2002).

$$\Delta p_i = a_i p_i(1 - p_i) + \sum_{j \neq i}^{n} a_j D_{ij} \quad \text{(S7)}$$

Where D_{ij} is the linkage disequilibrium between loci i and j. To further simplify this expression we assume that recombination is frequent relative to the strength of selection, an assumption which allows the populations to reach a state known as quasi-linkage-equilibrium (QLE). At QLE, the D_{ij} terms are small and on the same order as the selection, $\mathcal{O}(\beta)$ (Nagylaki 1993; Nagylaki et al. 1999). This allows us to simplify the change in allele frequency in (S7) to:

$$\Delta p_i \approx a_i p_i(1 - p_i) + \mathcal{O}(\beta^2) \quad \text{(S8)}$$

which simplifies to equation (3) of the main text upon substituting the value of a_i from equation (S6).

Because (S8) shows that at QLE the evolution of allele frequencies across loci is independent, we can make use of several classical results derived from the Wright-Fisher model to study the balance between selection and drift within the colonizing populations. Specifically, this can be seen by comparing equation (S8) to the classical expression for the change in allele frequency under the Wright-Fisher model:

$$\Delta p = s\, p(1 - p) \quad \text{(S9)}$$

(Hartl and Clark 2007). Comparison of (S8) to (S9) reveals that under linear selection and at QLE $s = \frac{\beta}{\alpha} b_i$. We can use this expression for s to utilize another classical result of the Wright-Fisher model, the probability of fixation of an allele under linear directional selection (Karlin and Taylor 1981):

$$P_{fix} = \frac{e^{2Ns(1-p_0)}\left(e^{2Nsp_o} - 1\right)}{e^{2Ns} - 1} = \frac{e^{2N\frac{\beta}{\alpha}b(1-p_0)}\left(e^{2N\frac{\beta}{\alpha}bp_o} - 1\right)}{e^{2N\frac{\beta}{\alpha}b} - 1} \quad \text{(S10)}$$

which is equation (4) and (5) of the main text.

Extension of equation (S10) to the probability of parallel evolution at a single locus (equation 6 of the main text) is straight-forward. Parallel genetic evolution at a single locus in m independent populations requires the independent fixation of the derived allele in the m populations and is therefore simply the m^{th} power of equation S10. The extension of (S10) to the multi-locus probability of parallel evolution given by equation 7 of the main text is similarly straight-forward. The only additional complications are that the allelic effect size b_i and the initial allele frequency, p_0, now require a subscript to denote the locus and that the probability of parallel evolution now represents the product a single locus parallel evolution events across loci.

B: *Markov Chain Monte Carlo Simulation of Posterior Distributions.*

At the conclusion of the individual based simulation we have the genetic data $\mathcal{D}$, which consists of a $2 \text{ X } n$ matrix of 0's and 1's indicating the fixation or loss of alleles at the n loci in the two descendent populations. As inputs to our simulation we also know the values for the allelic effect sizes and initial allele frequencies. We use the following Metropolis algorithm to sample the posterior distribution, $p(\eta|\mathcal{D})$ where $\eta = N\frac{\beta}{\alpha}$. The basic algorithm can be described in 7 steps, the first of these steps initializes the algorithm, the 2nd through 4th steps are recursive and generate samples from the posterior (Marjoram et al. 2003), and the 5th and 6th steps address the convergence and termination of the algorithm respectively. Convergence to the true posterior distribution can be computed by simulating multiple independent sequences of points and then comparing the variance between versus within these sequences; we will simulate $m = 5$ sequences to assess convergence. One important feature of MCMC sampling of the posterior distribution is called "burn in", a phenomena describing how the beginning portion of a sequence of sampled points depends largely on the initial starting point rather than on the posterior distribution, therefore the first simulated points are often uninformative in

describing the posterior. To eliminate "burn in" effects, when testing convergence and computing the estimated posterior distribution it is a common practice to only use the last half of the simulated points in a sequence (Gelman 2004).

Initialization

1. Draw m estimates of η from the prior distribution. These estimates will serve as the starting points for each of the m sequences.

Recursive algorithm: (Repeat for each of the m sequences)

2. From the current estimate, η, propose a move to a new point, η^*, where $\eta^*$is drawn from the jump distribution $J(\eta, \eta^*)$.
3. Calculate the probability of accepting the point η^*, h.

$$h = \min\left(1, \frac{P(D|\eta^*)\pi(\eta^*)J(\eta, \eta^*)}{P(D|\eta)\pi(\eta)J(\eta^*, \eta)}\right)$$

4. Move to $\eta^*$with probability h, otherwise stay at η.

Assessing Convergence: (After simulating 2n points in each of the m sequences)

5. Denote these m sequences by $\psi_{i,j}$ where i is the index over points and ranges from 1 to $2n$ and j is the index over the independent sequences and ranges between 1 and m. Discard the first n points in each sequence for "burn in" so that the index i now ranges between 1 and n. Calculate the following ratio of the variance between versus within each sequence:

$$R = \sqrt{\frac{\frac{n-1}{n}W + \frac{1}{n}B}{W}}$$

Where:

$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}\psi_{i,j} + \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}\psi_{i,j}\right)^2$ and

$W = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n-1}\sum_{i=1}^{n}\left(\psi_{i,j} - \frac{1}{n}\sum_{i=1}^{n}\psi_{i,j}\right)^2$

6. If $R < 1$ simulating additional points will not significantly improve the estimation of the peak of the posterior distribution (Gelman 2004).

After enough points have been simulated to reach a ratio of $R < 1$, we use the last n points in each of the $m = 5$ sequences to generate a single histogram, consisting of 100 bins, which approximates the posterior distribution. The estimate for η, $\hat{\eta}$, was given by the histogram bin containing the most points.

C: *Sensitivity of the Bayesian Estimator to migration from the ancestral population*

Our analytical model assumes the descendent populations diverge from their ancestor in perfect allopatry. Although this is likely true in many well-studied cases, in others recurrent gene flow from the ancestral population may occur. We investigate this scenario by integrating recurrent gene flow into our individual based simulations and then using these simulations to test the accuracy of the Bayesian estimator. We found that the Bayesian estimator is robust even in the presence of recurrent gene flow, remaining quite accurate when even an average of 2 migrants per generation arrives in the descendent populations (Table S4).

D: *Sensitivity of Bayesian Estimator to differing selection gradients in descendent populations.*

An additional assumption of our analytical model and Bayesian estimator is that the strength of natural selection is identical in the descendent populations. Although this assumption is likely to be at least approximately true in many well-studied cases, it is also possible the strength of selection differs among the descendent populations. We tested this possibility by integrating variable selection gradients into our individual based simulations and then using these simulations to evaluate the accuracy of the

Bayesian estimator. Specifically, we compared our Bayesian estimate of η to the average value of η actually experienced by the two populations when the selection gradient between the two populations differed by up to 20%. We saw no significant difference between the accuracy of the estimator under differing selection compared to when selection was identical (Table S5).

E: *Sensitivity of Bayesian Estimator to error in parameters*

Thus far we have discussed the robustness of the Bayesian estimator to many violations in the assumptions of the analytical model; these tests, however, were performed with an assumption of their own. Specifically, the Bayesian estimate of η was found under the assumption that initial allele frequencies and effect sizes of the alleles were known without error. Although these values can be estimated in natural systems, estimates will likely be associated with substantial error. To investigate how error in the estimates of these key parameters influenced the accuracy of our Bayesian estimator we ran individual based simulations where, rather than using the true values for the initial frequencies and effect sizes, we estimated $\hat{\eta}$ using parameter estimates accompanied by a specified amount of error. Specifically, we drew initial allele frequencies and effect sizes from a Gaussian distribution centered about the true value and with a variance of either 0% or 10% of the true value. The results of these simulations are shown in (Table S6), and reveal that the Bayesian estimator was relatively insensitive to error in initial allele frequency and only moderately sensitive to error in effect size.

F: References:

Gelman, A. 2004. Bayesian data analysis. Chapman & Hall/CRC, Boca Raton, Fla.

Hartl, D. L. and A. G. Clark. 2007. Principles of population genetics. Sinauer Associates, Sunderland, Mass.

Karlin, S. and H. M. Taylor. 1981. A second course in stochastic processes. Academic Press, New York.

Kirkpatrick, M., T. Johnson, and N. Barton. 2002. General models of multilocus evolution. Genetics 161:1727-1750.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavare. 2003. Markov chain Monte Carlo without likelihoods. P Natl Acad Sci USA 100:15324-15328.

Nagylaki, T. 1993. The Evolution of Multilocus Systems under Weak Selection. Genetics 134:627-647.

Nagylaki, T., J. Hofbauer, and P. Brunovsky. 1999. Convergence of multilocus systems under weak epistasis or weak selection. J Math Biol 38:103-133.

Tables:

Table S1:

	All Loci				GG Method				GC Method			
# of loci	**Slope**	**Itercept**	**R^2**	**Ideal R^2**	**Slope**	**Itercept**	**R^2**	**Ideal R^2**	**Slope**	**Itercept**	**R^2**	**Ideal R^2**
2	1.147795	-12.2237	0.382516	0.285826	1.058062	-8.8708	0.374565	0.295216	1.111731	-11.5976	0.370363	0.267604
4	0.959192	-2.23407	0.439945	0.41532	0.949658	-1.76086	0.403426	0.383155	0.962404	-1.79253	0.416365	0.39995
8	0.907914	1.589759	0.592764	0.584333	0.922304	1.730936	0.557142	0.552914	0.929305	0.649544	0.595217	0.586888

Table S2:

		All Loci				GG Method				GC Method			
Selection	**# of loci**	**Slope**	**Itercept**	**R^2**	**Ideal R^2**	**Slope**	**Itercept**	**R^2**	**Ideal R^2**	**Slope**	**Itercept**	**R^2**	**Ideal R^2**
Linear	2	0.993931	-7.84718	0.323874	0.227409	1.022567	-7.97479	0.307461	0.233554	1.022533	-9.10763	0.300105	0.20418
Linear	4	0.756274	5.979153	0.325097	0.29129	0.746356	5.526915	0.314167	0.276214	0.779655	5.043194	0.328181	0.301432
Linear	8	0.853359	3.459289	0.531232	0.515331	0.867688	3.253798	0.51223	0.500288	0.889206	3.235636	0.54589	0.536879
$\theta > z_{max}$	2	2.625239	-18.7313	0.192086	0.0166	2.205828	-14.7943	0.154519	0.026395	2.360828	-17.1241	0.166472	0.003039
$\theta > z_{max}$	4	1.568844	-11.0517	0.254115	0.189913	1.490425	-9.16855	0.246104	0.199309	1.619483	-10.0341	0.25449	0.205522
$\theta > z_{max}$	8	1.015979	3.549194	0.442131	0.404899	0.960228	5.07128	0.447337	0.396366	0.955678	5.586419	0.451979	0.38915
$\theta < z_{max}$	2	1.30543	-12.8406	0.016797	-0.23195	1.403294	-11.0413	0.019578	-0.15149	1.320669	-12.6884	0.016353	-0.21251
$\theta < z_{max}$	4	1.301843	-6.36213	0.061018	0.013012	1.220801	-6.16244	0.05796	0.001686	1.309492	-6.76073	0.062507	0.006473
$\theta < z_{max}$	8	0.860807	6.626172	0.122763	0.018462	0.796777	6.950119	0.114301	0.012296	0.842249	6.075352	0.118696	0.040224

Table S3:

		All Loci				GG Method				GC Method			
Selection	# of loci	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2
Linear	2	0.805834	-5.51332	0.161113	0.013794	0.797929	-5.00893	0.160271	0.018121	0.908339	-6.71297	0.207498	0.100172
Linear	4	0.815067	2.26572	0.289378	0.263764	0.958235	-1.58709	0.341933	0.328472	0.884206	0.834166	0.326031	0.312265
Linear	8	0.789538	4.896994	0.437137	0.405046	0.793404	5.857202	0.49175	0.457663	0.795886	5.478371	0.476069	0.444676
$\theta > z_{max}$	2	2.035355	-14.9953	0.137879	-0.01712	1.799011	-15.4037	0.111434	-0.08901	1.805776	-14.7436	0.105895	-0.0622
$\theta > z_{max}$	4	1.359543	-3.1936	0.28387	0.262606	1.409364	-3.54603	0.275801	0.250532	1.358227	-2.87746	0.241728	0.222614
$\theta > z_{max}$	8	0.878473	4.582206	0.413538	0.392837	0.884301	3.971124	0.388928	0.375328	0.88283	4.164366	0.384784	0.369638
$\theta < z_{max}$	2	1.952207	-15.0071	0.033728	-0.24192	2.456438	-16.5704	0.052944	-0.23382	2.231008	-15.3975	0.038509	-0.19063
$\theta < z_{max}$	4	1.632467	-8.57066	0.083332	0.032068	1.621518	-8.61235	0.100455	0.036049	1.541532	-8.15362	0.07568	0.024201
$\theta < z_{max}$	8	1.053753	3.842943	0.175806	0.096223	1.145484	3.171672	0.19544	0.105922	1.08128	3.842346	0.193132	0.097805

Table S4:

				All Loci				GG Method				GC Method			
Selection	# of loci	$Error\ in\ p_0$	$Error\ in\ b$	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2
Linear	2	10%	10%	0.750731	-1.85462	0.194334	0.071629	0.783463	-2.20934	0.207322	0.103721	0.808416	-3.66876	0.223495	0.102703
Linear	4	10%	10%	0.863989	0.851631	0.345397	0.321882	0.889694	0.120192	0.355744	0.334789	0.900137	0.368497	0.362652	0.348127
Linear	8	10%	10%	0.765458	5.73031	0.5079	0.460212	0.851563	3.581057	0.511715	0.496155	0.823923	4.112699	0.528926	0.504624
Linear	2	10%	0%	0.843997	-3.84964	0.20902	0.123974	0.972201	-9.05249	0.250067	0.139702	0.820594	-3.38814	0.208808	0.11374
Linear	4	10%	0%	1.003107	-2.68517	0.423377	0.410097	1.035959	-4.47945	0.456793	0.4312	1.058345	-5.2665	0.450801	0.422826
Linear	8	10%	0%	0.815542	4.778413	0.522613	0.495865	0.883528	3.016746	0.548438	0.538903	0.899576	2.553102	0.555071	0.548153
Linear	2	0%	10%	0.607291	5.723086	0.10027	0.026871	0.700849	1.604204	0.130831	0.050792	0.625651	4.367798	0.105088	0.023792
Linear	4	0%	10%	0.852635	-0.9344	0.330718	0.278685	0.916524	-4.28903	0.306573	0.243258	0.89235	-1.62596	0.365417	0.323737
Linear	8	0%	10%	0.819625	2.728602	0.565961	0.527694	0.816885	3.087144	0.531526	0.497547	0.812521	2.79454	0.531662	0.49165

Table S5:

			All Loci				GG Method				GC Method			
Selection	# of loci	# migrants	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2
Linear	2	1	0.730709	5.011167	0.145798	0.124712	0.766462	3.6703	0.159232	0.141348	0.794994	3.597031	0.176274	0.16328
Linear	4	1	0.760797	7.528352	0.263227	0.235217	0.74217	8.804608	0.275457	0.235172	0.756406	8.168807	0.290141	0.254792
Linear	8	1	0.792802	6.373698	0.412981	0.382409	0.814237	5.717994	0.382511	0.360918	0.795089	6.060887	0.379673	0.353311
Linear	2	2	1.052674	-13.5859	0.260787	0.10436	1.082817	-14.3412	0.292574	0.125103	1.055074	-14.0939	0.266068	0.095364
Linear	4	2	1.074735	-11.5165	0.364182	0.23557	1.059636	-10.0412	0.362969	0.259513	1.110429	-11.01	0.387357	0.292205
Linear	8	2	1.187219	-10.4852	0.450421	0.386139	1.106158	-7.25574	0.44789	0.406645	1.120747	-8.13183	0.438202	0.389363

Table S6:

		difference	All Loci				GG Method				GC Method			
Selection	# of loci	in gradient	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2	Slope	Itercept	R^2	Ideal R^2
Linear	2	10%	0.852154	-1.84687	0.184609	0.146097	0.784362	-0.11105	0.169786	0.122322	0.832906	-1.27668	0.185073	0.144329
Linear	4	10%	0.776742	5.972979	0.230891	0.211672	0.712757	8.431943	0.223098	0.184411	0.80189	6.111742	0.257836	0.240062
Linear	8	10%	0.780095	8.886883	0.379939	0.319288	0.718599	10.05027	0.349604	0.269802	0.764475	9.308508	0.356151	0.292058

Table Ledgends:

Table S1: Wright-Fisher Simulations. The estimate $\hat{\eta}$ is calculated from data simulated using the Wright-Fisher model with $Ns = \eta$. The table records the slope, intercept and R^2 values for the regression fit for the simulated data as well as the R^2 values for the data fit to the ideal $\hat{\eta} = \eta$ line. Simulations were performed for between 2 and 8 loci and estimates for the strength of selection were

made using the data from all loci, as well as data filtered to resemble the two experimental methods, GG and GC.

Table S2: Evaluating accuracy across different forms of selection. The table shows the slope, intercept, and R^2 values for the linear-regression fit of predicted values of $\hat{\eta}$ to the true value in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, GG and GC. Accuracy was evaluated for 2, 4, and 8 loci and three forms of natural selection; linear directional, stabilizing with $\theta < z_{max}$, and stabilizing with $\theta > z_{max}$. The loci underwent free recombination, $r = 0.5$.

Table S3: Variable forms of selection with constrained recombination. The table shows the slope, intercept, and R^2 values for the linear-regression fit of predicted values of $\hat{\eta}$ to the true value in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, GG and GC. Accuracy was evaluated for 2, 4, and 8 loci and three forms of natural selection; linear directional, stabilizing with $\theta < z_{max}$, and stabilizing with $\theta > z_{max}$. In contrast to table S2 recombinaton was constrained at a rate of $r = 0.05$

Table S4: Estimation of η with recurrent gene flow from the ancestral population. The table shows the slope, intercept, and R^2 values for the linear-regression fit of predicted values of $\hat{\eta}$ to the true value in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, GG and GC. Accuracy was evaluated for 2, 4, and 8 loci under linear selection.

Each generation each descedent population received a set number of migrants (1 or 2) from the ancestral population.

Table S5: Estimation of η with differing selection in descendent populations. The table shows the slope, intercept, and R^2 values for the linear-regression fit of predicted values of $\hat{\eta}$ to the true value in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, GG and GC. Accuracy was evaluated for 2, 4, and 8 under linear selection. The strength of linear selection was allowed to vary up to 10% between the two populations.

Table S6: Estimation of η with error in $\boldsymbol{p_0}$ and $\boldsymbol{b}$. The table shows the slope, intercept, and R^2 values for the linear-regression fit of predicted values of $\hat{\eta}$ to the true value in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, GG and GC. Accuracy was evaluated for 2, 4, and 8 loci under linear selection. The estemated values of $\hat{\eta}$ where calculated using initial allele requencies, p_0, and alellic effect sizes, b, that differed from their true values.