The probability of parallel genetic evolution from standing genetic variation

SCHOLARONE™
Manuscripts

Dear Editor,

We would like to submit a revised version of our paper, "The probability of parallel genetic evolution from standing genetic variation" for potential publication in The Journal of Evolutionary Biology. We appreciate the helpful comments provided by associate editor Gardner and the second reviewer, and outline how we have changed our manuscript in response to their specific recommendations below.

Associate Editor:
I would like to see a few more words added to the manuscript on the scope and limitations of the study, in line with Reviewer 2's comments from the previous round. This is to ensure maximum clarity of exposition.

*We now more thoroughly discuss key limitations and assumptions of our study within a single cohesive paragraph in the discussion **(ln 512-542)**.*

Reviewer 2:
l. 42–44: This still sounds as if you focus on evolution from standing genetic variation because you (want to, for mathematical convenience?) assume that alleles were initially segregating at low frequency. I think you should turn this around and say that you are interested in studying adaptation from standing variation, which allows you to make the assumption of initially low frequencies. Therefore, I also suggest replacing "Our interest in this definition is motivated..." on l.44–45 by "Our decision is motivated...".

*We have revised the wording in response to this suggestion **(ln 46-47)**.*

l. 85–87: The second clause ("for example, repeated selection...") is still missing a verb.

*We have reworded this phrase **(ln 93-94)**.*

l. 121–125: Saying that beta is weak and that it is of order epsilon, where epsilon is small, is redundant. I suggest getting rid of the "O(epsilon)" at all from the text, and just rely on an accurate verbal description.

*We have removed this redundancy by eliminating ϵ **(ln 130-33)**.*

l. 116-123: You introduce O(epsilon) without definition of epsilon, so the paragraph is strictly speaking meaningless. You still do not seem to be addressing reviewer 1's point about what happens if p_i (1 - p_i) is of the same order of magnitude (i.e. as small) as the sum over the D_ij terms (cf. Eq. S7).

We no longer use ϵ (see response to previous comment).

Strictly speaking, the equal signs in several equations after Eq. (4) should be replaced by approximately equal signs. You could avoid this by replacing "This probability is given by" before Eq. (4) by "This probability can be approximated as". Then, you can convert the approximately equal sign in Eq. (4) back to an equal sign.

*We have changed the wording as suggested **(ln 137)**.*

l.192: "figure 3" -> "Figure 3"

We have fixed this (ln 194).

l.197: It is not necessary to say "the value of".

We have removed this (ln199).

l. 209–210: "...value for the parameter eta,..." -> "...value of the parameter eta,..."?

We have fixed this (ln 212).

l.237: "using the Wright–Fisher model" -> "under the Wright–Fisher model"

We have changed this (Ln 239).

l. 261: The comma should be after, not before, "(Figure 1C)".

We have changed this (ln 263).

l.319: Replace full stop by comma after "distant optimum"

We have fixed this (ln 323).

l.322: "figure 4" -> "Figure 4"? Please check the journal's requirements

We have changed this (ln 326).

l.348: Insert comma after "overlap zero"

We have added this (ln 351).

l.364: "population sizes very large" -> "population sizes are very large"

We have fixed this (ln 367).

l. 405: My previous comment to what used to be l.384 (see below) has been addressed by including simulations with more than two populations (Figure S1, right-most column). Citing from my previous review: "l. 384: I did not spend much time thinking about this: But is it proven that it does not matter for the accuracy and efficiency of the estimator whether the data comprise few loci in many populations or vice versa? Without a proof, I suggest backing away from this point. You assessed the performance of the estimator only for two populations. I am fine with a statement about the number of loci, but not with a statement about the number of populations, as far as the performance of the estimator is concerned. I suspect that the effect of the number of populations will strongly depend on how you define parallel evolution. Since you use the most restrictive definition possible (fixation in all m populations), any conclusion from this would not necessarily be very generic."

I was surprised by the similarity between the right-most column in Figure S1 and the one for "Errors in parameters". Please provide an explanation if appropriate.

Thank you for pointing this out. We re-ran the additional population simulations and have corrected Figure S1.

l.413: Delete comma before "that".

We have reworded this section.

l.436: "provides" -> "is" or "offers" (to avoid repeating "provides"; see l.434)

We have changed this (In 454).

l.449: Please write the full journal name.

We have changed this (In 464).

l.473: Please check if it should say "MC1R" or "Mc1r".

Mc1r is correct.

l.490: Is the page range missing?

This references is in an unpublished book and is therefore yet to have page numbers.

l.511–512: For consistence, change first letters of words to lower case. Please fix/check the page range.

We have fixed this (In 527).

l.563: Please change "Fit" and "Data" to lowercase.

We have fixed this (In 578).

l.565: Perhaps: "Closed points" -> "Filled points"?

We have changed this (In 580).

l.566: "does not (do) overlap" -> "does (not) overlap"

We have fixed this (In 581).

l.571: Please change hat(eta) from bold to regular font, and introduce a space after the "equal to" sign.

We have fixed this (In 585).

Figure 3: Please increase the font size for the axis tick labels. Moreover, I suggest reducing the number of different font types and shapes used.

We increased the font sizes and made the math and letter text consistent across the figure.

Figure 4: Thanks for correcting this to match the values from the QTL method. Please increase the font size for the axis tick labels and the summary of results in red and blue.

We have increased the font sizes.

You do not seem to have addressed the following previous comment:
"In Figure 4, please revise the blue text below the panel in the middle row in the column for 2 loci: check if intercept and slope are in the correct order and that the order is consistent with the other panels. It seems as if the former is correct, but the latter needs to be rectified."

The slope and intercept are correct in this panel.

Supplementary Material:
Section A

- Did you address this previous point? "On the first line, it says "parallel phenotypic evolution", but given point 1) of reviewer 1 and the authors' response to it, I think it should now read "parallel genetic evolution (from standing genetic variation)". "Parallel phenotypic evolution" still appears multiple times at other places in the Supplementary Material. Please adjust as appropriate."

We use "parallel phenotypic evolution" and "parallel genetic evolution" to distinguish between observations of parallelism at the phenotypic level vs. the genetic level. The reference to parallel phenotypic selection refers to the observation of shared phenotypes motivating a study of the underlying genetics.

- Equation (S4): Please add a comma after "effect of locus" just before the equation. Please rectify the the position of the equation label "(S4)".

We have fixed this.

- Before Eq. (S5): "...for relative fitness" -> "...for the relative fitness,"

We have added this.

- Please add a comma immediately after Eq. (S6a).

We have added this.

- Given that there is a term "$O(a\_i^2)$" in Eq. (S8), you should replace the "approximately equal to" by an "equal to" sign.

We have changed this.

- If I am not mistaken, the issue of what happens if p_i (1 - p_i) is of the same order as D_{ij} is still not addressed (see a point of reviewer 1 from the first round of review).

Although D_{ij} can theoretically be on the same order as $p_i(1 - p_i)$, as the allele frequency approaches fixation this is unlikely to occur as it requires both a very high correlation between the two loci and for locus j to be highly polymorphic. To test if this ever occurs, we ran several simulations using the various fitness functions and recombination rates and calculated $\frac{p_i(1-p_i)}{D_{ij}}$ over the course of the simulation. In short, $p_i(1 - p_i)$ was typically at least 2 orders of magnitude greater than linkage disequilibrium. The two quantities were never of the same order for more than a single generation during the whole course of adaptation and therefore this issue is unlikely to significantly affect the accuracy of our QLE assumption.

- Before Eq. (S9), change "Wright-Fisher model" -> "Wright–Fisher model,"

We have added this.

- Replace the first "equal to" sign in Eq. (S10) by an "approximately equal to" sign.

We have changed this.

- Third-to-last line before Eq. (S11): "equation 5" -> "equation (5)".

We have changed this.

- Before Eq. (S11): Please give a reference for the alternative definition as "the repeated fixation in x of the m descendent populations".

We derived the alternative definition for the repeated fixation in m populations. The derivation is similar to that for the definition in the main text and hence is not included.

"The probability of parallel evolution defined this way is given by" -> "The probability of parallel evolution defined this way can be approximated as"

We have changed this.

Figure S1 (caption): "...with variable number of populations" -> "with a variable number of populations"; replace comma after "number of populations" by colon (":"); delete "respectively"; "triagnle" -> "triangle" (THIS OCCURS TWICE); "Wright Fisher" -> "Wright–Fisher"; "Explinations" -> "Explanations"; "given in main text" -> "given in the main text"; "supplimentary" -> "supplementary".

We have addressed these.

My previous comments have only been partially addressed:
"Thanks for introducing Figure S1. I suggest improving the colour scheme and reducing the size of the

symbols. It is currently hard to distinguish points based on features other than colour. Sometimes the red square, circle, and triangle overlap in such a way that it is hard to discern them. I suggest using a graphical legend. Caption: "method" missing after "candidate gene"? "Explinations" -> "Explanations". "supplimentary" -> "supplementary". Please improve the language/grammar in the last sentence of the caption (e.g. repetition of "as well"). Why does introducing 10% error in p_0 and b result in higher R^2 than introducing 10% error in only one of p_0 or b? Please increase the font size for text on top and at the bottom of the figure."

We have increased the font sizes and addressed the changes to the caption- see previous comment.

My following previous comment has not been addressed:
"I wonder if it could help to add a similar figure showing the slope, rather than the R^2. This is because I find it hard to see why R^2 alone should be the statistic of choice to judge the performance of your method. You could have a very shallow slope, but still explain a decent proportion of the variance, and the other way round. Figure 4 illustrates this problem to some extent."

Figure S1 shows the ideal R^2 value when the line is forced to have zero intercept and slope of 1 ($\eta = \eta^*$). Fitting the data to this ideal line rather than the best fit regression enables us to capture the effect of both the slope and intercept. So although a very shallow regression slope may indeed explain a lot of the variance (high least squares regression R^2) it will not lead to a high ideal R^2 value.

Tables S2 to S8: Thanks for adding this! Caption: "...to the true value in the simulation eta, and the..." -> "...to the true values of eta from simulations, and the..."; change "or" from italic to regular font on the second-to-last line.

We have corrected this.

Table S5: I still wonder about the following: "For 8 loci and 2 migrants per generation, the slopes appear to be considerably larger than 1. Is this due to an underestimation of eta when the true eta is low, or due to an overestimation when it is high? Or both?

With large amounts of gene flow between decedent populations will inflate the estimation of η for both large and small true values of η. This is reflected both by the large intercept and the slope greater than 1.

We hope that our responses satisfactorily address the concerns of associate editor Gardner and the reviewer. If there are any additional issues that it would be helpful to address, please let us know.

Thank you.

1   **Abstract**
2
3
4   Parallel evolution is often assumed to result from repeated adaptation to novel, yet ecologically

5   similar, environments. Here we develop and analyze a mathematical model that predicts the probability

6   of parallel genetic evolution from standing genetic variation as a function of the strength of phenotypic

7   selection and constraints imposed by genetic architecture. Our results show that the probability of

8   parallel genetic evolution increases with the strength of natural selection and effective population size,

9   and is particularly likely to occur for genes with large phenotypic effects. Building on these results, we

10   develop a Bayesian framework for estimating the strength of parallel phenotypic selection from genetic

11   data. Using extensive individual based simulations, we show that our estimator is robust across a wide

12   range of genetic and evolutionary scenarios and provides a useful tool for rigorously testing the

13   hypothesis that parallel genetic evolution is the result of adaptive evolution. An important result that

14   emerges from our analyses is that existing studies of parallel genetic evolution frequently rely on data

15   that is insufficient for distinguishing between adaptive evolution and neutral evolution driven by

16   random genetic drift. Overcoming this challenge will require sampling more populations and the

17   inclusion of larger numbers of loci.

18

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

20 **Introduction**
21
22
23 As the availability of genome sequences has increased, interest in understanding how genomic

24 architecture shapes adaptation at both the genetic and phenotypic level has grown substantially

25 (Stapley et al., 2010). How and which genes respond to selection is a complex result of many aspects of

26 the genotype to phenotype map, including allelic effect sizes, epistatic interactions, linkage

27 disequilibrium, and pleiotropy. Significant work using natural populations (Nadeau & Jiggins, 2010),

28 experimental evolution (Qi et al., 2016, Wichman et al., 1999), and evolutionary theory (Orr, 2005,

29 Chevin et al., 2010) has been devoted to elucidating how these many factors interact to shape

30 adaptation. Particularly useful natural systems for addressing such questions are those exhibiting

31 parallel evolution. Many striking examples of repeated phenotypic and genetic change exist (Conte et

32 al., 2012, Stern, 2013, Martin & Orgogozo, 2013), putatively as a consequence of adaptation to similar

33 selective environments (Schluter, 2009). These systems can be viewed as natural experimental

34 replicates for understanding the interplay of selection and genetic architecture in shaping patterns of

35 adaptation (Hohenlohe et al., 2010).

36 There are many definitions for "parallel evolution", a phenomenon which may or may not be

37 distinguished from "convergent evolution". These many definitions all share the common theme of

38 repeated evolution in two or more populations, but differ in two major ways. First, in terms of whether

39 or not these populations originated from a recent common ancestral population or are only distantly

40 related. Second, definitions differ in the biological level at which repeated evolution occurs, ranging

41 from the genetic to the phenotypic level (Lenormand et al., *In press.*). Here we focus on parallel genetic

42 evolution defined as the repeated fixation of identical alleles in multiple descendent populations. We

43 further restrict our study to cases where parallel genetic evolution results from standing genetic

44 variation rather than from new mutations. Our interest in this specific scenario is motivated by a

2

45   number of biological systems where adaptation to a novel environment is thought to result from

46   standing genetic variation present in the ancestral population (Hoekstra et al., 2006, Steiner et al., 2007,

47   Colosimo et al., 2005). In contrast to *de novo* mutation, adaptation from standing genetic variation is

48   likely rapid (Barrett & Schluter, 2008) and may lead to distinct genomic signatures of parallel adaptation

49   (Roesti et al., 2014)

50   Understanding the conditions that promote parallel genetic evolution has been facilitated by

51   theoretical studies. For instance, Orr (2005) calculated the probability that 1 of k *de novo* beneficial

52   mutations arises and fixes repeatedly and found that parallel evolution becomes more likely as the

53   strength of selection increases and the number of possible alleles, k, decreases. These results are

54   supported by experimental adaptation of the bacteriophage $\phi X 174$ to high temperatures (Wichman et

55   al., 1999) and adaptation of antifungal drug resistance in *Saccharomyces cerevisiae* (Anderson et al.,

56   2003). Taking a different approach, Chevin et al. (2010) calculated the probability that a beneficial *de*

57   *novo* mutation fixes at the same genetic locus in independent populations. By allowing mutations to

58   influence multiple phenotypic traits simultaneously, this work demonstrated that the probability of

59   parallel evolution is greatest when pleiotropy is weak. In addition, this work demonstrated that when

60   mutations have pleiotropic effects, the probability of parallel evolution is greater when populations are

61   relatively close to their adaptive optima (i.e., not too maladapted). Together, these previous theoretical

62   studies provide a solid framework for understanding the likelihood of parallel evolution arising from the

63   fixation of novel mutations.

64   Although understanding the contribution of new mutations to parallel evolution is inarguably

65   important, in some systems it may be more relevant to understand the likelihood of parallel evolution

66   from standing genetic variation. For instance, in the stickleback, *Gasterosteus aculeatus*, repeated

67   adaptation to freshwater is thought to involve genes already segregating at low frequencies within the

68   marine populations (Colosimo et al., 2005). In cases like these, the presence of adaptive alleles in the

69   ancestral population can have a significant effect on the probability of parallel evolution, influencing

70   both the long term probability of parallel adaptation as well as the rate at which adaptation occurs

71   (Ralph & Coop, 2015). Our focus here is to enhance our understanding of parallel evolution by

72   developing a genetically explicit multi-locus framework for predicting the probability of parallel

73   evolution from standing genetic variation. We have two specific goals: first, we will predict the

74   probability of parallel evolution in terms of quantities that are regularly measured in natural populations

75   using a multi-locus model of parallel genetic adaptation that assumes weak selection and rapid

76   recombination. Second, we will develop a statistical framework for estimating the historical average

77   strength of parallel selection by coupling our multi-locus model to routinely collected genetic data.

78   **The Model**

79   _Biological Scenario_

80   We envision a scenario where haploid individuals from an ancestral population colonize two or

81   more novel environments and establish new populations (see Figure 1A). After this initial colonization

82   we assume gene flow between the ancestral and descendent populations is negligible and that

83   individuals within populations mate at random. The descendent populations then experience identical

84   patterns of phenotypic selection causing population mean phenotypes at the focal trait to diverge in

85   parallel from the ancestral population; for example, repeated selection resulting in reduced body armor

86   in multiple freshwater stickleback populations relative to their common marine ancestral phenotype

87   (Colosimo et al., 2004).

88   We next envision that the genetic basis of the trait undergoing parallel phenotypic evolution is

89   studied using one of two commonly used experimental designs (Conte et al., 2012). Figure 1B illustrates

90   the first experimental design where parallel genetic evolution is assessed at a set of candidate genes. To

91   identify possible candidate genes, individuals from at least one descendent population (descendent

92   population 1 in Figure 1B) are crossed with ancestral individuals and the resulting offspring are scanned

4

93   for divergent QTL's influencing the focal trait. The remaining populations (descendent population 2 in

94   Figure 1B) are then tested for the candidate genes using a variety of approaches such as genetic

95   complementation tests (Hartl, 2005).   This method, which we will call the "*candidate gene method*", has

96   been used in human populations to identify the genetic basis of the multiple independent origins of

97   lactose tolerance (Ingram et al., 2009, Enattah et al., 2008, Tishkoff et al., 2007). Alternatively, in the

98   second design (shown in Figure 1C), descendent populations are searched independently for the genes

99   responsible for the repeated phenotypic divergence from the ancestral population.   This is done by

100   performing independent QTL scans in each descendent population.   This *"QTL method"* was used to

101   identify separate genes responsible for a change in developmental rate in two populations of

102   *Oncorhyncus mykiss,* (Robison et al., 2001, Nichols et al., 2007, Sundin et al., 2005).

103   *Analytical Model*

104   Our model assumes the trait experiencing parallel selection is controlled by n additive loci. Each

105   locus, denoted with the index i, has two possible alleles A_i and a_i and a phenotypic effect equal to b_i

106   associated with the A_i allele, such that the phenotype of an individual is described by

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

108   where X_i is an indicator variable taking the value 1 if the individual carries the A allele at locus i and the

109   value 0 if the individual carries the a allele at locus i . We assume b_i is positive for all i implying that the

110   A_i allele always increases the value of the trait z. The frequency of the A_i allele is given by p_i and $\bar{z}$

111   denotes the average phenotype of the population. We assume the average phenotype of the ancestral

112   population is small, meaning that the frequency of the A_i allele is low at all loci, and initially equal to p_{0_i}.

113   Within the new environments, individuals experience selection for large phenotypes, favoring an

114   increase in frequency of the A alleles.

115   The biggest challenge to modeling evolution across multiple loci is that epistasis and linkage

116   disequilibrium make it extremely difficult to formulate analytical predictions for the probability of

117   fixation at individual loci. Two key assumptions, however, make calculating the probability of fixation

118   tractable. First, we assume the relationship between an individual's phenotype, z, and its fitness, $W(z)$,

119   is linear:

120      $$W(z) = \beta z + \alpha. \tag{2}$$

121   and thus defined only by its intercept (α) and slope (β). Second, we assume the strength of linear

122   directional selection, β, is weak, and that the rate of recombination between loci relatively high. Under

123   these conditions, recombination breaks apart linkage disequilibrium more quickly than it can be built up

124   by selection, and a quasi-linkage-equilibrium (QLE) is reached where linkage disequilibrium is also small,

125   and of the same order as β (Nagylaki, 1993, Nagylaki et al., 1999). Using the expression for the

126   phenotypic trait z, given in equation (1), as well as the expression for fitness, given by equation (2), we

127   can use the multi-locus methods developed by Barton and Turelli (1991) and expanded by Kirkpatrick et

128   al. (2002) to derive the change in the frequency of the A_i allele at QLE over a single generation

129      $$\Delta p_i \approx \frac{\beta}{\alpha} b_i p_i (1 - p_i) \tag{3}$$

130   (*See supplementary material for a full derivation*). Because we have assumed linear selection and that

131   the population is at quasi-linkage equilibrium (QLE), equation (3) does not depend on linkage

132   disequilibrium or the frequencies of alleles at other loci; instead, each locus evolves independently.

133   Later, using individual based simulations, we will relax these key assumptions and evaluate the

134   robustness of this analytical approximation.

135   The independent evolution of loci enables us to utilize a classic result of the Wright-Fisher model

136   describing the probability of fixation for an allele with initial frequency p_0 in a population of constant

137   size N. This probability can be approximated as

138      $$P_{fix} = \frac{\left(1 - e^{2Nsp_0}\right)}{1 - e^{2Ns}} \tag{4}$$

139   (Kimura, 1957, Karlin & Taylor, 1981) where s is the strength of selection acting on the allele and p_0 is its

140   initial frequency. Under our assumption of linear directional selection, strength of selection acting on

141   locus i is $s = \frac{\beta}{\alpha} b_i$, and equation (4) can be re-written as

142     $$P_{fix}(i) = \frac{\left(1 - e^{2N\frac{\beta}{\alpha}b_i p_{0_i}}\right)}{1 - e^{2N\frac{\beta}{\alpha}b_i}} \ . \tag{5}$$

143   Equation (5) reveals that the probability of fixation depends on initial allele frequency, local population

144   size, the strength of phenotypic selection, and the phenotypic effect of the locus. In the next section we

145   will use this result to explore how these important parameters influence the extent of parallel evolution.

146   *The probability of parallel genetic evolution at a single locus*

147   We begin by analyzing the simplest possible scenario: a single genetic locus. For this case,

148   parallel evolution entails the repeated fixation of the same allele in multiple descendent populations.

149   The probability of this occurring can be calculated by using equation (5) to find the probability that at

150   the locus of interest, i, the A_i allele fixes independently in each of m populations:

151     $$P_{\parallel} = \left(p_{fix}(i)\right)^m . \tag{6}$$

152   Requiring repeated fixation in all m populations represents a very restrictive definition of parallel

153   genetic evolution and in some cases a less restrictive definition may be preferable. In such cases, it is

154   straightforward to develop expressions for the probability of repeated fixation in any subset of m

155   populations using (5). An example of the calculations for a less restrictive definition is provided in the

156   online supplemental material.

157   Equations (5,6) highlight three important factors that will influence the probability of observing

158   parallel genetic evolution. First, the probability of repeated fixation of an allele increases with its initial

159   frequency, p_{0_i}. Second, large effect alleles, those with large b_i, are more likely to fix in parallel under

160   directional selection. This relationship between effect size and parallel evolution is shown in Figure 2.

161   Third, parallel genetic evolution is more likely to occur when evolution is driven primarily by the

162   deterministic force of natural selection rather than the stochastic force of random genetic drift.

163   Specifically, the probability of parallel evolution increases with the product of population size and the

164   phenotypic selection gradient in derived populations, $N\frac{\beta}{\alpha}$. This product captures the balance between

165   drift and selection and shows that parallel evolution is more likely in large populations experiencing

166   strong natural selection as shown by the three curves in Figure 2. As this term arises repeatedly in the

167   derivation to follow we will denote it with the composite parameter η

168    $$\eta = N\frac{\beta}{\alpha}.$$ (7)

169   *The probability of parallel genetic evolution at multiple loci*

170   Although the single locus results of the previous section are insightful, they fall short of

171   capturing the genetic richness of real populations where the extent of parallel evolution must be

172   assessed across multiple loci. Fortunately, calculating the probability of parallel evolution across

173   multiple loci is straightforward, and yields the following formula:

174    $$P_{\parallel} = \prod_{i=1}^{n} \left(p_{fix}(i) \right)^{m},$$ (8)

175   where the product is carried over the number of loci. Not surprisingly, equation (8) shows that the

176   factors enhancing the probability of parallel evolution at a single locus (e.g., large population size, strong

177   selection, etc.) also increase the probability of parallel evolution across multiple loci. What distinguishes

178   one locus from the next is the initial allele frequency and the allelic effect size. Therefore, the probability

179   of parallel evolution across multiple loci will depend on the distribution of allelic effects. Equation (8)

180   clarifies the connection between the effect size distribution and parallel evolution yielding several novel

181   insights that emerge only when multiple loci are considered.

182   The first and most obvious insight to emerge from (8) is that perfectly parallel genetic evolution,

183   where all loci are fixed for the selectively favored A_i alleles in all descendent populations, becomes less

184   and less likely as the number of loci increases. This is a simple result of the product rule of probabilities,

185   and arises because the overall probability of parallel evolution decreases as each additional locus is

186   required to fix in parallel in the m descendent populations. The second insight that emerges from

187   equation (8) is that when selection is relatively weak, population sizes are relatively small, and adaptive

188   alleles initially infrequent, it is unlikely to observe parallel evolution at more than a single locus with

189   large phenotypic effect (Figure 3). As selection becomes stronger, population sizes larger, or adaptive

190   alleles initially more frequent, however, it becomes increasingly likely that parallel evolution will occur

191   at multiple loci, including loci with moderate phenotypic effects (Figure 3). These results are, for the

192   most part, relatively insensitive to the particular distribution of effect sizes across loci. Only in cases of

193   strong selection and high initial allele frequency (when evolution becomes more deterministic) does the

194   effect size distribution contribute significantly (bottom right panel of Figure 3). In such cases, the

195   probability of parallel evolution at a large number of loci increases with the mode of the effect size

196   distribution. In other words, parallel evolution at a large number of loci is most likely when the effect

197   size distribution is not skewed toward small effect loci (Figure 3). Together, these results suggest that

198   the likelihood of observing parallel genetic evolution at any particular number of loci depends heavily on

199   the parameter η.

200   _Bayesian inferences of parallel phenotypic selection_

201   The results derived in the previous section demonstrate a strong connection between the

202   parameter η and the probability of observing parallel genetic evolution. In this section we develop a

203   method for estimating the value of this key parameter using a Bayesian framework that capitalizes on

204   equation (8). Our goal is to provide a methodology that allows support for a hypothesis of adaptive

205   parallel evolution to be assessed using data collected in empirical studies of parallel genetic evolution.

9

206 Specifically, by estimating η it becomes possible to distinguish between parallel genetic evolution

207 caused by random genetic drift, $\eta = 0$, and parallel genetic evolution caused by natural selection.

208 Our Bayesian approach will rely on genetic data described by a matrix, $\mathcal{D}$, where rows represent

209 descendent populations and columns loci. Each element of $\mathcal{D}$ takes a value of 0 or 1 depending on

210 which allele has fixed at a particular locus in a given population (Figure 1A). Using equation (8) we can

211 develop a likelihood function specifying the probability of observing the data, $\mathcal{D}$, given a particular value

212 of the parameter η, empirical estimates of the effect sizes b_i and initial allele frequencies p_0. The effect

213 sizes can be (and frequently are) estimated using QTL scans (Conte et al., 2015, Broman & Sen, 2009,

214 Lynch & Walsh, 1998) whereas initial allele frequencies can be estimated by measuring the allele

215 frequencies in the ancestral population. Accurate estimation of the genetic data $\mathcal{D}$, the allelic effect

216 sizes, and the initial allele frequencies will require sufficient sample sizes from each descendent

217 population and the ancestral population. We will later address the consequences of uncertainty in the

218 estimation of the initial allele frequencies and allelic effect sizes using individual based simulations.

219 Given estimates for these parameters, the likelihood expression consists of a product of terms, one for

220 each locus of the focal trait in each population. If the A allele has fixed at a locus it contributes a term

221 P_{fix}, as defined by equation (5). Alternatively, if the A allele is lost, it contributes a term $\left(1 - P_{fix}\right)$.

222 Thus, for m populations and n loci, the likelihood of observing the data, $\mathcal{D}$, is given by the following

223 product

224 $$\mathcal{L}(\mathcal{D}) = \prod_{j=1}^{m} \prod_{i=1}^{n} P_{fix}(\eta, i)^{\mathcal{D}_{ij}} \left(1 - P_{fix}(\eta, i)\right)^{1 - \mathcal{D}_{i,j}} , \qquad (9)$$

225 where i is an index over loci and j an index over populations. The likelihood for η as a function of the

226 genetic data $\mathcal{D}$ and the genetic architecture of the trait under selection is based on principles similar to

227 those developed by Rice and Townsend (2012). A key difference, however, is that here the likelihood of

228   parallel genetic evolution is based on only the loci influencing a single phenotypic trait, rather than the

229   entire mutational effect distribution.

230   The likelihood, (9), can be used in a Bayesian setting to estimate a posterior distribution for the

231   key parameter η. Specifically, Bayes' theorem enables us to formulate estimates for η in the form of the

232   posterior distribution $p(\eta|\mathcal{D})$ that is biologically meaningful for all possible genetic outcomes, $\mathcal{D}$,

233     $p(\eta|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\eta)\pi(\eta)}{\int_\eta \mathcal{L}(\mathcal{D}|\eta)\pi(\eta)}$,   (10)

234   where $\pi(\eta)$ is our prior distribution for the parameter η. The denominator of this expression is the

235   integral over the likelihood surface, and cannot be easily evaluated. For this reason, we use a Markov

236   Chain Monte Carlo algorithm to sample from the posterior distribution and generate an estimate of the

237   most probable value of η for the given genetic data $\mathcal{D}$. We label this estimate $\hat{\eta}$. We take two

238   approaches to evaluating the performance of this estimator. First we analyze its performance under the

239   assumptions of the analytical model by generating the genetic data $\mathcal{D}$ under the Wright-Fisher model.

240   Next, we test the robustness of the estimator to violations of the assumptions of our analytical model by

241   generating the genetic data $\mathcal{D}$ using multi-locus individual based simulations.

242   *Wright-Fisher Simulation*

243   We simulated the data $\mathcal{D}$ for two populations under the Wright-Fisher model by drawing a

244   random number between 0 and 1 for each locus and population and setting $\mathcal{D}_{i,j}$ to 1 if the random

245   number was less than p_{fix} from equation (5) and to 0 otherwise. The value of p_{fix} depends on the initial

246   allele frequency at each locus,p_{0_i}, the allelic effect sizes of each locus, b_i, as well as the parameter η.

247   For each simulation we drew the values of these parameters independently and at random. Initial allele

248   frequencies were drawn independently at each locus from a uniform distribution between 0 and 0.1.

249   Because our model envisions divergence of descendent populations from a common ancestor we

250   assumed that the initial frequency at any one locus was the same in both populations. Allelic effect sizes

251   were drawn independently for each locus from a uniform distribution between 0 and 1. The value of η

252   for each run was drawn from a uniform distribution ranging between 0 to 50. The genetic outcome $\mathcal{D}$

253   simulated in this manner may not, however, resemble what would be measured using experimental

254   methods. For example, using current genomic techniques it is not possible to identify loci that have not

255   diverged from the ancestral state. To address how experimental methodologies affect our Bayesian

256   estimates we considered two modified forms of $\mathcal{D}$ that resemble sampling under the two experimental

257   methods described previously (see Figure 1). The first of these methods, the candidate gene method

258   (Figure 1B), assesses parallel genetic evolution at candidate genes which are known to have generated

259   the phenotypic divergence in the first descendent population. This is often done by performing a cross

260   between individuals from one of the divergent populations with the ancestral population and assessing

261   the genetic variation in the F1s. Since the second divergent population is not independently assessed

262   for divergent QTLs, under this method we only consider the columns of $\mathcal{D}$ (i.e., loci) where the A allele

263   has fixed in the first population. The second experimental method, the QTL method (Figure 1C),

264   independently assesses divergent loci in all descendent populations. Under this method, $\mathcal{D}$ therefore

265   contains all columns (loci) which have fixed in at least one population. Hence the effective number of

266   loci identified using this method will always be greater than or equal to the number found by the less

267   thorough candidate gene method.

268   For each simulated $\mathcal{D}$, as well as for $\mathcal{D}$ modified by the two experimental methods, we

269   estimated η using a Metropolis-Hastings algorithm as described in the supplementary material. For the

270   prior $\pi(\eta)$ we used a uniform distribution on the interval $\eta = \pm 80$. To analyze the performance of the

271   estimator we ran a regression of the estimated values of $\hat{\eta}$ on the true values η, using 200 data points.

272   Overall, this analysis revealed that the estimator was quite accurate, explaining between 80% and 85%

273   of the variation (see Table S1). In addition, our analysis showed that the accuracy of the estimates

274   increases with the number of loci. This trend holds regardless of the experimental method used.

275   However, the effective number of loci under the QTL method is always greater than when candidate

276   genes are first identified in one population and then subsequently searched for in the other. The results

277   of these simulations suggest our estimator performs quite well when the data meet the assumptions of

278   our analytical model; however, this may not be the case for real data. In the next section, we explore the

279   performance of our estimator using individual based simulations. These simulations allow us to evaluate

280   the consequences of violating key assumptions of our analytical model such as the weak selection and

281   frequent recombination required for our quasi linkage equilibrium approximation.

282   *Individual Based Simulation*

283   Our individual based simulations consider two allopatric populations, each of which has a

284   constant size of $N = 1000$ individuals. Initial allele frequencies and effect sizes at each locus, as well as

285   the value of η, were drawn randomly as described above under the Wright-Fisher model. Individuals

286   within each population undergo a two stage life cycle. During the first stage, "selection", the probability

287   that an individual survives is given by its fitness, with fitness computed using either equation (2) which

288   describes linear selection or an expression for stabilizing selection described below. Surviving

289   individuals then enter the second life cycle stage, "reproduction", which consists of generating an

290   offspring population from the surviving parental population. This is done by drawing a pair of parents at

291   random from the pool of surviving individuals and producing an offspring from these parents by

292   recombining the parental genomes at a specified rate r and allowing mutation between the two allelic

293   states at a per locus mutation rate of $\mu = 10^{-6}$. This process is continued with replacement of parents

294   until the offspring population reaches the pre-selection size of N. This life cycle is repeated until all loci

295   approach fixation or loss (allele frequencies > 0.99 or < 0.01) at which point the simulations were

296   terminated and the matrix of genetic data $\mathcal{D}$ filled by rounding the allele frequency to 1 or 0. As in the

297   previous section, we formulate modified versions of $\mathcal{D}$ that resemble sampling under the two

298   experimental methods. Then, using the Metropolis-Hastings algorithm, we compute estimates for the

13

299   value of η using the original outcome $\mathcal{D}$ as well as the two modified forms of $\mathcal{D}$ (*see Supplementary*

300   *Material*).

301   We used the simulations to test the robustness of the estimator when selection is strong and/or

302   non-linear. To test the effect of non-linear selection, simulations were run where an individual's fitness

303   was determined by one of two alternative forms of selection: linear directional selection described by

304   (2), or stabilizing selection toward a phenotypic optimum:

305     $W(z) = e^{-\gamma(z-\theta)^2}$,   (11)

306   where θ is the phenotypic optimum and γ is the strength of stabilizing selection. Including simulations

307   where selection is stabilizing is important because it relaxes our previous assumption that loci evolve

308   independently. Stabilizing selection is particularly useful in testing this assumption because the relative

309   importance of interdependence between loci can be manipulated by changing the value of the

310   phenotypic optimum. Specifically, the relative importance of interactions between loci will depend on

311   the value of the optimum relative to the largest possible phenotype $z_{max} = \sum_i b_i$. When θ is greater

312   than the largest possible phenotype, z_{max}, loci remain relatively independent as directional selection

313   predominates over epistatic selection. However when $\theta < z_{max}$ this is no longer true as epistatic

314   selection now dominates. Therefore, when $\theta > z_{max}$ evolution is much more likely to resemble linear

315   selection as our analytical model assumes. We simulated stabilizing selection under these two different

316   scenarios, by either requiring that θ be larger than z_{max} or slightly smaller than z_{max} (*see*

317   *Supplementary material*). Under stabilizing selection η changes as the population adapts, decreasing as

318   the population approaches the optimum (Chevin & Hospital, 2008, Matuszewski et al., 2015).

319   Therefore, we computed a "realized" strength of linear selection by averaging the selection gradient,

320     $\frac{Cov(z,w)}{Var(z)}$, over all time points for which $Var(z) \neq 0$.

14

321   As expected, analysis of simulated data shows that the accuracy of our estimates depends on

322   the form of selection. Specifically, estimates for η are most accurate under linear selection, somewhat

323   less accurate under stabilizing selection toward a distant optimum, $\theta > z_{max}$, and least accurate under

324   stabilizing selection toward a close optimum, $\theta < z_{max}$ (See Figure 4 and Table S2). In addition to

325   assuming that selection is linear, we also assumed that selection is weak. By computing the variance

326   about the regression line as η increased we were able to confirm that, for the data shown in Figure 4,

327   the accuracy of our estimates decreases with increasing selection. Next, we used our simulations to

328   explore the sensitivity of our estimator to infrequent recombination among candidate loci (See Table

329   S3). Not surprisingly, these simulations revealed that our estimator performs better when

330   recombination is frequent ($r = 0.5$) than when recombination is rare ($r = 0.05$). The effect of

331   infrequent recombination is more drastic for stabilizing selection than linear selection, and is particularly

332   pronounced when $\theta < z_{max}$. This is expected since this latter scenario generates the strongest epistatic

333   selection and thus has the greatest potential to cause linkage disequilibrium to accumulate.

334   Finally, we used the individual based simulations to test the accuracy of our estimator when

335   several assumptions of the biological scenario envisioned above are violated. These violations included

336   recurrent gene flow from the ancestral to the descendent populations, gene flow among descendent

337   populations, selection that differs in strength across descendent populations, and estimates of the

338   parameters p_0 and b_i that are imprecise (*see Supplementary Material and Table S4-S6 and Figure S1*).

339   These simulations reveal that our estimates of η are robust to violations in many of these assumptions.

340   Specifically, the estimator performs well when migration occurs between descendent populations or

341   from the ancestral population as long as rates remain below 2 migrants per generation. In addition,

342   estimates for the average η in the descendent populations remain robust even when the selection

343   gradients differ among descendent populations by up to 20%. Lastly, estimates for η are relatively

344   insensitive to modest levels of error (<10%) in the estimated values of the parameters p_0 and b_i. Up to

345   this point we have focused on using the Bayesian estimator to provide single point estimates for η.

346   Having access to the full posterior distribution, however, allows us to calculate a 95% credible interval

347   for the parameter η and determine whether or not it overlaps with zero. From an empirical standpoint,

348   being able to rule out $\eta = 0$ allows us to reject the hypothesis that observed levels of parallel genetic

349   evolution can be explained by random genetic drift alone. The open versus closed circles in Figure 4

350   represent data points for which credible intervals drawn from the posterior distribution do, or do not,

351   overlap zero, respectively. Figure 5 shows how the probability of rejecting 0 (filled bars) increases with

352   η.

353   **Discussion**

354   It has long been understood that natural selection, parallel genetic evolution, and genomic

355   architecture are inherently linked (Orr, 2005, Schluter, 2009, Chevin et al., 2010). Previous theoretical

356   work has focused on repeated genetic evolution from new mutation and found that key components of

357   genetic architecture, such as the number of possible beneficial mutations (Orr, 2005) and the

358   distribution of mutational effects (Chevin et al., 2010) influence the probability of parallel evolution.

359   Here we have used a multi-locus model of parallel evolution from standing genetic variation to further

360   formalize these connections. We began our investigation by calculating the probability of parallel

361   evolution at a single locus and showed that parallel evolution is most likely when phenotypic selection is

362   strong, standing genetic variation for adaptive alleles is appreciable, adaptive alleles have large

363   phenotypic effects, and population sizes are large. Next, we used a quasi-linkage equilibrium

364   approximation to extend our analyses to multiple loci, demonstrating that the number of loci that

365   evolve in parallel depends on the product of phenotypic selection and local population size (η). If

366   selection is relatively weak, or if population sizes are small, we expect parallel evolution at no more than

367   a single locus. In contrast, when selection is relatively strong, or population sizes are very large, parallel

368   evolution may occur across multiple loci. These results demonstrate that without information on the

16

369   strength of phenotypic selection and population size, we have no way to assess whether the amount of

370   parallel genetic evolution we observe in an empirical study is beyond what would be expected under

371   neutrality. To remedy this problem, and better connect studies of parallel genetic evolution to the

372   evolutionary processes they imply, we developed a Bayesian approach that capitalizes on available

373   genetic data to estimate the product of phenotypic selection and local population size (η). In the

374   following paragraphs we explore several of the key results in more detail and discuss their implications

375   for past, present, and future studies of parallel genetic evolution.

376   The first important result that emerges from all of our models is that parallel evolution is most

377   likely to be observed at loci with large phenotypic effects on traits experiencing strong phenotypic

378   selection in novel environments. This result receives at least some support from empirical studies of

379   parallel genetic evolution. For example the large effect gene Eda has been found in eight fresh-water

380   descendent populations of three-spine stickleback, *Gasterosteus aculeatus,* and is largely responsible for

381   the parallel reduction in lateral plate number in these populations. In contrast, the small effect locus LG7

382   has been confirmed in only two of the 8 descendent populations (Schluter et al., 2004, Colosimo et al.,

383   2004, Conte et al., 2012). It seems likely that this example — where a stark ecological shift from salt to

384   fresh water has occurred — corresponds to a case where natural selection is quite strong. Another

385   important, albeit unsurprising, result of equation (6) is that the probability of parallel evolution at a

386   single locus also depends on effect size and initial allele frequency. This may help explain why, contrary

387   to the results described above, a recent comprehensive survey of allelic effects involved in parallel

388   adaptation in two stickleback populations found no correlation between effect size and probability of

389   repeated gene use (Conte et al., 2015).   Our results suggest that the lack of correlation may be the

390   result of highly variable initial allele frequencies among loci.

391   By integrating multi-locus genetics into a model of adaptation, we were also able to derive

392   expressions for the probability of observing parallel genetic evolution at various numbers of loci over the

17

393   course of adaptation. The most important result to emerge from this analysis is that in the absence of

394   information about the likely strength of phenotypic selection in derived populations and the number of

395   individuals composing these derived populations, there is no way to assess the significance of observing

396   parallel evolution at any particular number of loci. Put differently, if natural selection in novel

397   environments is quite strong or population sizes in novel environments quite large, observing parallel

398   evolution at multiple genetic loci is not too surprising. If, however, natural selection is weak or

399   population sizes very small, observing this same level of genetic parallelism would be rather unexpected.

400   This suggests that if we are to more rigorously interpret the results of empirical studies of parallel

401   genetic evolution, we must do better than simply counting up the number of parallel genetic changes

402   observed. Our Bayesian tool accomplishes this goal by providing a methodology for tying information

403   on the extent of parallel genetic information to underlying evolutionary processes.

404   For our Bayesian approach to be broadly useful, it must produce reliable estimates across a

405   broad range of parameter space. For this reason, we used individual based simulations to assess the

406   accuracy and robustness of our approach when key assumptions of the underlying model are violated.

407   These simulations demonstrate that our estimator is indeed both accurate and robust, although there

408   are limitations. For example, accurate estimation requires data from at least 8 total loci, be that 4 loci in

409   two populations, 2 loci in four populations, or some intermediate combination. Whether data is

410   gathered at fewer loci in many populations or many loci in few populations should, in principle, have no

411   effect on the accuracy or efficiency of the estimator. Many of the studies discussed above however,

412   have far fewer than this. For example, studies of parallel pigmentation changes in a variety of species,

413   from beach mice (Hoekstra et al., 2006) to cave fish (Protas et al., 2006, Gross et al., 2009), focus

414   primarily on one or two loci in somewhere between 2 and 6 populations. Therefore if future studies

415   hope to understand the role of natural selection in driving parallel evolution, it is important that they

416   focus on acquiring data from as many loci and as many populations as possible. Because increasing the

18

417   number of loci at which parallel genetic evolution is assessed will almost certainly require studying loci

418   with smaller phenotypic effects on the focal trait, there may be hard limits to the number of loci that

419   can be usefully included. Fortunately, no such limitation exists with respect to the number of

420   populations that can be included.

421   The demonstration that increasing the number of loci at which parallel evolution is assessed is

422   important, suggests that some experimental approaches may perform better than others. For instance,

423   we explored two alternative experimental methods (see Figure 1 B and C) that differ predictably in the

424   number of loci that are detected: the QTL method and the candidate gene method. Because the QTL

425   method always detects parallel evolution at an equal or larger number of loci, we recommend its use

426   over the candidate gene method. Finally, our results show that the accuracy of the Bayesian estimate is

427   influenced not only by the amount of available data but by the accuracy of the estimated parameter

428   values b_i and p_{0_i}. Fortunately, however, as long as error in these parameters remains modest, our

429   Bayesian approach continues to produce reliable estimates (*See the supplementary Material*).

430   Alternatively, because the approach we have developed is Bayesian, it is also possible to input a prior

431   distribution for the parameter p_{0_i} rather than a single point estimate in cases where the exact value of

432   p_{0_i} is in doubt due to sampling error or stochastic variation in small populations (Hermisson & Pennings,

433   2005).

434   In addition to requiring information on parallel genetic evolution drawn from a reasonably large

435   number of loci or populations, our estimator relies on several key assumptions that affect its accuracy.

436   The most important of these assumptions is that selection is weak and linear. When combined with the

437   assumption that recombination is frequent, these key assumptions allowed us to utilize a quasi-linkage

438   equilibrium approximation. Another potentially important assumption of our approach is that

439   population size is constant across time. As a result, we implicitly ignore potentially important impacts of

440   sporadic population bottlenecks or founder effects. This assumption may prove particularly important

441   in cases of repeated evolution of reduced skin pigmentation in European and Asian human populations

442   for which there is evidence for extensive bottlenecks (Schmegner et al., 2005, Amos & Hoffman, 2010).

443   Finally, our approach assumes that selection/population size is identical in each population and that

444   recurrent gene flow does not occur. Although these assumptions may ultimately prove important in

445   some cases, our individual based simulations show that they have only a limited impact on the accuracy

446   of estimates in most cases (*See Supplementary Material*).

447   Combined, our analyses of single and multi-locus models show that it is difficult to draw

448   conclusions about the biological significance of parallel genetic evolution without information on the

449   strength of parallel phenotypic selection and local population size. We have overcome this hurdle by

450   developing a robust statistical methodology for translating observed levels of genetic parallelism into an

451   estimate of the product of phenotypic selection and local population size. This statistical approach

452   provides a much needed tool for distinguishing between adaptive and non-adaptive hypotheses for

453   observed levels of parallel genetic evolution. Applying this method to existing and emerging data from

454   multiple populations with common ancestry may thus offers novel insights into the importance of

455   adaptive evolution in natural populations.

456

457

458   **References**

459   Amos, W. & Hoffman, J. I. 2010. Evidence that two main bottleneck events shaped modern human
460   genetic diversity. *Proceedings of the Royal Society B-Biological Sciences* **277**: 131-137.
461   Anderson, J. B., Sirjusingh, C., Parsons, A. B., Boone, C., Wickens, C., Cowen, L. E. & Kohn, L. M. 2003.
462   Mode of selection and experimental evolution of antifungal drug resistance in *Saccharomyces*
463   *cerevisiae*. *Genetics* **163**: 1287-1298.
464   Barrett, R. D. & Schluter, D. 2008. Adaptation from standing genetic variation. *TRENDS in Ecology and*
465   *Evolution* **23**: 38-44.
466   Broman, K. W. & Sen, S. 2009. *A guide to QTL mapping with R/qtl*. Springer, Dordrecht.
467   Chevin, L. M. & Hospital, F. 2008. Selective sweep at a quantitative trait locus in the presence of
468   background genetic variation. *Genetics* **180**: 1645-1660.
469   Chevin, L. M., Martin, G. & Lenormand, T. 2010. Fisher's model and the genomics of adaptation:
470   Restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**: 3213-3231.
471   Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J.,
472   Myers, R. M., Schluter, D. & Kingsley, D. M. 2005. Widespread parallel evolution in sticklebacks
473   by repeated fixation of ectodysplasin alleles. *Science* **307**: 1928-1933.
474   Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D. & Kingsley, D. M.
475   2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS*
476   *Biology* **2**: 635-641.
477   Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., Schluter, D. & Peichel, C. L.
478   2015. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine
479   stickleback. *Genetics*.
480   Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. 2012. The probability of genetic parallelism and
481   convergence in natural populations. *Proceedings of the Royal Society B-Biological Sciences* **279**:
482   5039-5047.
483   Enattah, N. S., Jensen, T. G. K., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., El-Shanti, H., Seo,
484   J. K., Alifrangis, M., Khalil, I. F., Natah, A., Ali, A., Natah, S., Comas, D., Mehdi, S. Q., Groop, L.,
485   Vestergaard, E. M., Imtiaz, F., Rashed, M. S., Meyer, B., Troelsen, J. & Peltonen, L. 2008.
486   Independent introduction of two lactase-persistence alleles into human populations reflects
487   different history of adaptation to milk culture. *American Journal of Human Genetics* **82**: 57-72.
488   Gross, J. B., Borowsky, R. & Tabin, C. J. 2009. A novel role for Mc1r in the parallel evolution of
489   depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genetics*
490   **5**.
491   Hartl, D. J., E.W. 2005. *Genetics: Analysis of genes and genomes*. Jones & Bartlett Learning.
492   Hermisson, J. & Pennings, P. S. 2005. Soft sweeps: Molecular population genetics of adaptation from
493   standing genetic variation. *Genetics* **169**: 2335-2352.
494   Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. 2006. A single amino acid
495   mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101-104.
496   Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. & Cresko, W. A. 2010. Population
497   genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS*
498   *Genetics* **6**.
499   Ingram, C. J. E., Mulcare, C. A., Itan, Y., Thomas, M. G. & Swallow, D. M. 2009. Lactose digestion and the
500   evolutionary genetics of lactase persistence. *Human Genetics* **124**: 579-591.
501   Karlin, S. & Taylor, H. M. 1981. *A second course in stochastic processes*. Academic Press, New York.
502   Kimura, M. 1957. Some problems of stochastic-processes in genetics. *Annals of Mathematical Statistics*
503   **28**: 882-901.

504 Lenormand, T., Chevin, L. M. & Bataillon, T. M. (*In press.*) Parallel evolution: what does it (not) tell us and
505   why is it (still) interesting? In: *Chance in Evolution*, (Ramsey, G. P., C.H., ed.)*.* pp. Univ. Chicago
506   Press.
507 Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, Mass.
508 Martin, A. & Orgogozo, V. 2013. The loci of repeated evolution: A catalog of genetic hotspots of
509   phenotypic variation. *Evolution* **67**: 1235-1250.
510 Matuszewski, S., Hermisson, J. & Kopp, M. 2015. Catch me if you can: Adaptation from standing genetic
511   variation to a moving phenotypic optimum. *Genetics* **200**: 1255-74.
512 Nadeau, N. J. & Jiggins, C. D. 2010. A golden age for evolutionary genetics? Genomic studies of
513   adaptation in natural populations. *Trends in Genetics* **26**: 484-492.
514 Nichols, K. M., Broman, K. W., Sundin, K., Young, J. M., Wheeler, P. A. & Thorgaard, G. H. 2007.
515   Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in
516   *Oncorhynchus myhiss*. *Genetics* **175**: 335-347.
517 Orr, H. A. 2005. The probability of parallel evolution. *Evolution* **59**: 216-220.
518 Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R. &
519   Tabin, C. J. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of
520   albinism. *Nature Genetics* **38**: 107-111.
521 Qi, Q., Toll-Riera, M., Heilbron, K., Preston, G. M. & MacLean, R. C. 2016. The genomic basis of
522   adaptation to the fitness cost of rifampicin resistance in *Pseudomonas aeruginosa*. *Proceedings*
523   *of the Royal Society B-Biological Sciences* **283**.
524 Ralph, P. L. & Coop, G. 2015. The role of standing variation in geographic convergent adaptation.
525   *American Naturalist* **186**: S5-S23.
526 Rice, D. P. & Townsend, J. P. 2012. A test for selection employing quantitative trait locus and mutation
527   accumulation data. *Genetics* **190**: 1533-+.
528 Robison, B. D., Wheeler, P. A., Sundin, K., Sikka, P. & Thorgaard, G. H. 2001. Composite interval mapping
529   reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus*
530   *mykiss*). *Journal of Heredity* **92**: 16-22.
531 Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W. & Berner, D. 2014. The genomic signature of
532   parallel adaptation from shared genetic variation. *Molecular Ecology* **23**: 3944-3956.
533 Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science* **323**: 737-741.
534 Schluter, D., Clifford, E. A., Nemethy, M. & McKinnon, J. S. 2004. Parallel evolution and inheritance of
535   quantitative traits. *American Naturalist* **163**: 809-822.
536 Schmegner, C., Hoegel, J., Vogel, W. & Assum, G. 2005. Genetic variability in a genomic region with long-
537   range linkage disequilibrium reveals traces of a bottleneck in the history of the European
538   population. *Human Genetics* **118**: 276-286.
539 Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D.,
540   Beckerman, A. P. & Slate, J. 2010. Adaptation genomics: The next generation. *Trends in Ecology*
541   *& Evolution* **25**: 705-712.
542 Steiner, C. C., Weber, J. N. & Hoekstra, H. E. 2007. Adaptive variation in beach mice produced by two
543   interacting pigmentation genes. *PLoS Biology* **5**: 1880-1889.
544 Stern, D. L. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**: 751-764.
545 Sundin, K., Brown, K. H., Drew, R. E., Nichols, K. M., Wheeler, P. A. & Thorgaard, G. H. 2005. Genetic
546   analysis of a development rate QTL in backcrosses of clonal rainbow trout, *Oncorhynchus*
547   *mykiss*. *Aquaculture* **247**: 75-83.
548 Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K.,
549   Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B.,
550   Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. & Deloukas, P. 2007. Convergent adaptation
551   of human lactase persistence in Africa and Europe. *Nature Genetics* **39**: 31-40.

552   Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M. & Bull, J. J. 1999. Different trajectories of
553   parallel evolution during viral adaptation. *Science* **285**: 422-424.

554

555   **Figure Legends**

556   **Figure 1: Schematic of biological scenario.** Panel A depicts two descendent populations diverging in

557   parallel from a common ancestral population. The a allele predominates at all four loci in the ancestral

558   population whereas the A allele fixes at various loci in the two descendent populations. Panel B and C

559   depict two methods for deducing the underlying genetics of reduced body size in the two descendent

560   populations depicted in panel A. Panel B shows the candidate gene method which relies on a genome

561   wide scan of progeny from a cross between the first descendent population and the ancestral

562   population and subsequent candidate gene search in the second descendent population. Panel C shows

563   the QTL method which involves two genome wide scans, one in each population. Compared to the

564   candidate gene method, the QTL method uncovers an additional locus driving divergence in population

565   2.

566   **Figure 2: The probability of parallel evolution as a function of allelic effect size, b**. For a given strength

567   of selection the probability of fixation, and hence parallel evolution, increases with allelic effect size.

568   The rate of increase is non-linear and depends on the strength of selection s and the population size N

569   which are given by the compound parameter $\eta = Ns$. The initial alele frequency for the three curves

570   was held constant at $p_0 = 0.01$.

571   **Figure 3: The probability of parallel evolution at n loci.** For a trait determined by the effects of 10 total

572   loci the probability of observing parallel evolution at exactly n loci depends on the strength of selection,

573   which varies from near 0 to a value of $\eta = 50$, and the intial allele frequency, which is either low (0.01)

574   or high (0.1). The probability of parallel evolution may also depend on the underlying effect size

575   distribution depicted here (top panel) as three different gamma probability distributions with different

576 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

577 same mean effect size ($\mu = \frac{1}{2}$).

578 **Figure 4: Regression fit of IBS data under the three forms of selection.** Data and linear regression fit

579 (blue), and perfect fit (red) between the time averaged values of η and the Bayesian estimate $\hat{\eta}$ for 200

580 replicates of the individual based simulation. Open (filled) points indicate estimates where the 95%

581 credible interval does (not) overlap $\eta = 0$. Sub-panels differ in the number of loci (ranging from 2 to 8)

582 and the form of natural selection (Linear, Stabilizing with $\theta > z_{max}$, and Stabilizing with $\theta < z_{max}$). The

583 Bayesian estimator uses genetic data filtered to resemble sampling using the QTL experimental method.

584 **Figure 5: Significance of $\hat{\eta}$ estimates.** The fraction of estimates that differ significantly (black) and do

585 not differ significantly (white) from $\hat{\eta} = 0$ with 95% confidence across the range of time average η
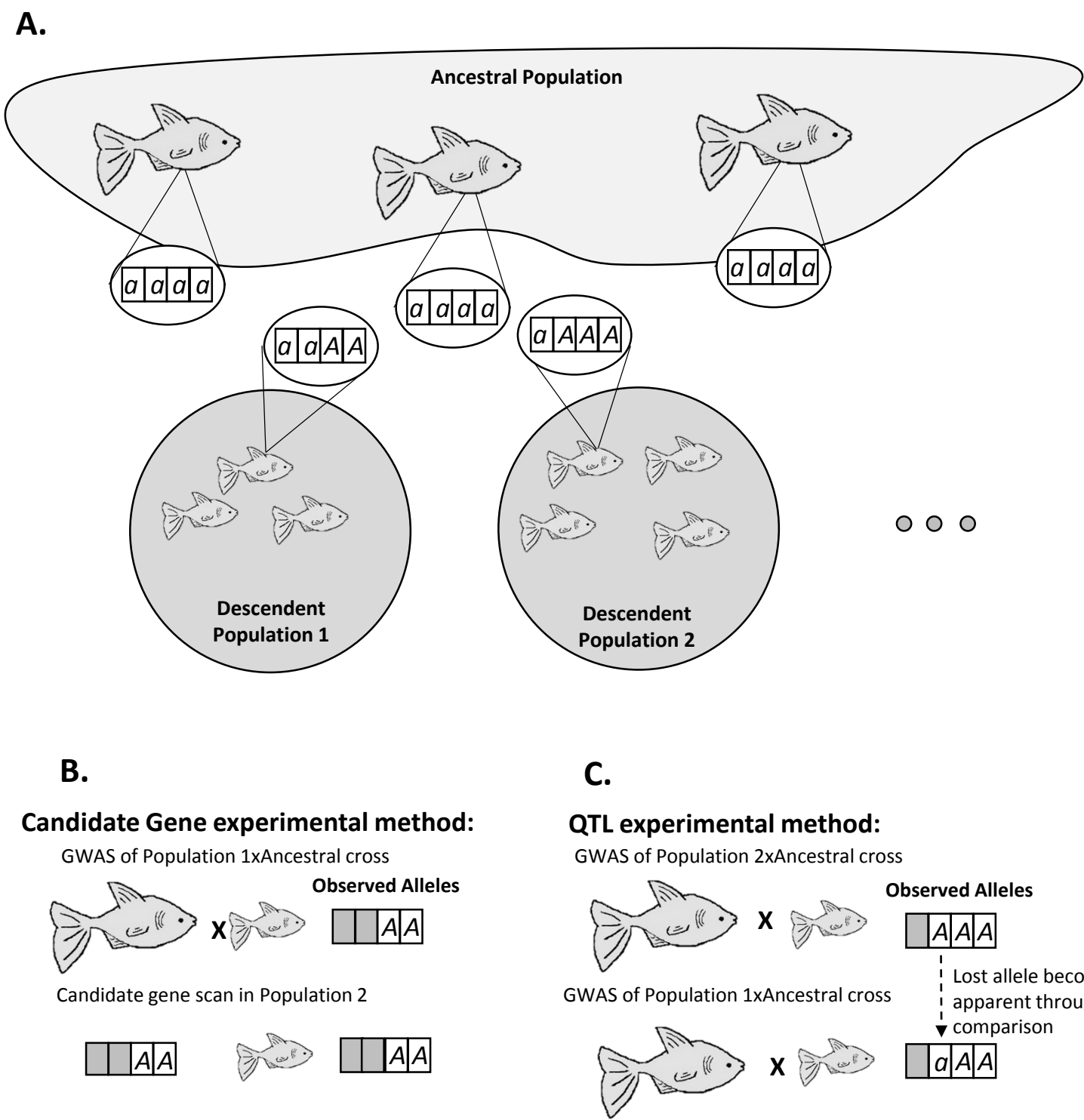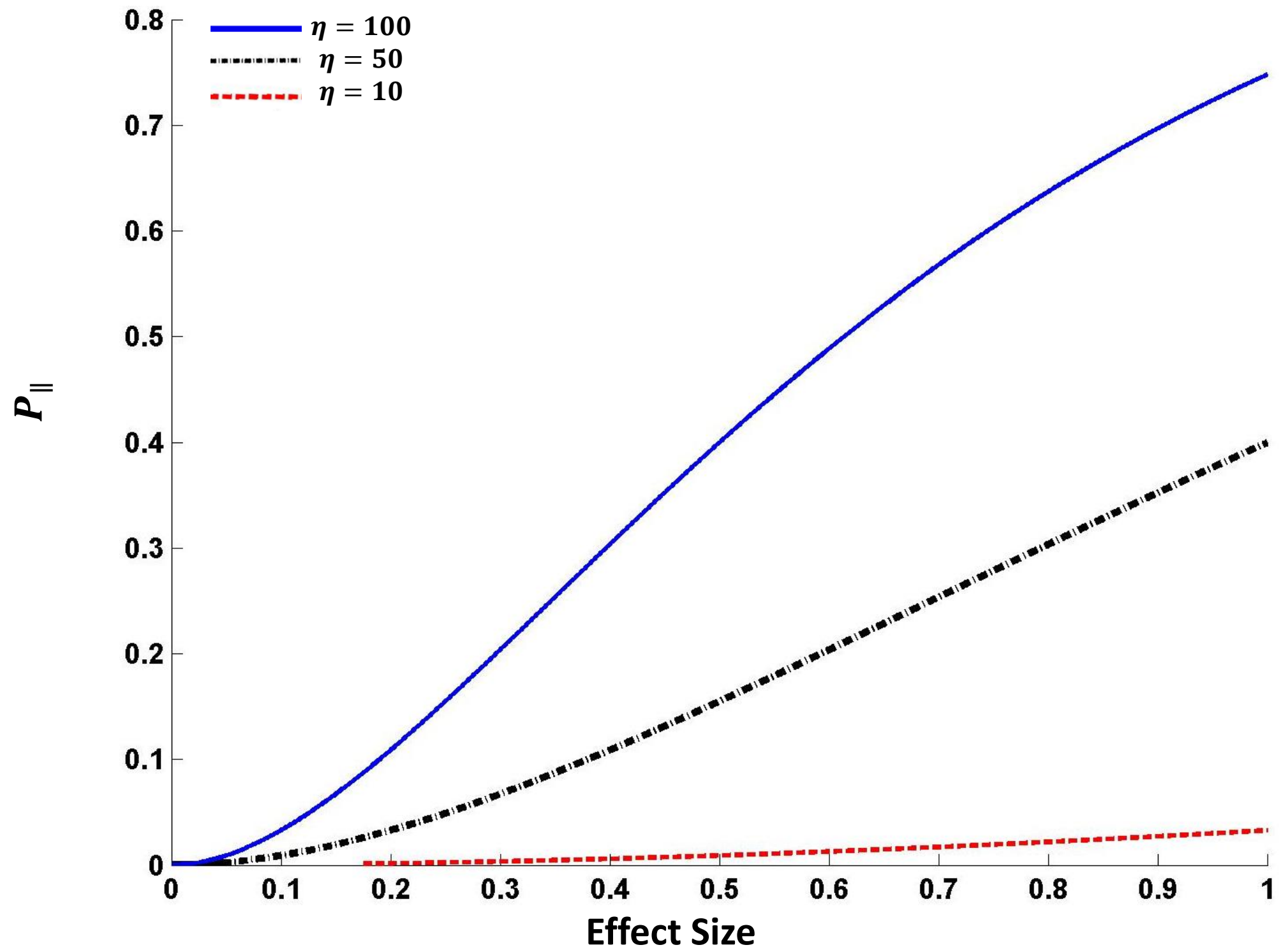
586 values. Overlapping portions are shown in grey.

24

Figure 1:

A.

B.

Candidate Gene experimental method:

GWAS of Population 1xAncestral cross

Candidate gene scan in Population 2

C.

QTL experimental method:

GWAS of Population 2xAncestral cross

GWAS of Population 1xAncestral cross

Figure 2:

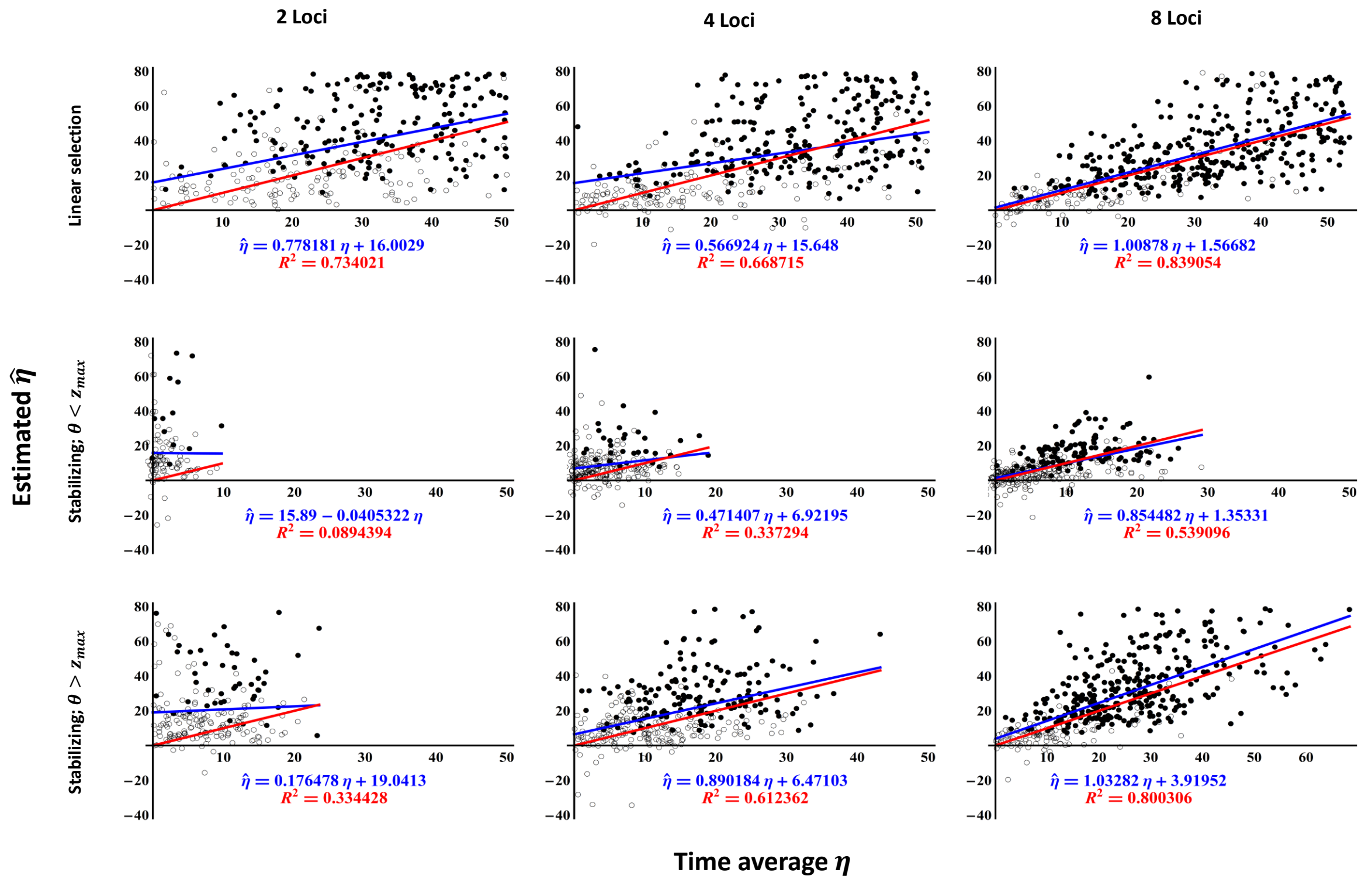# Probability of parallel evolution at exactly n loci

Figure 3:
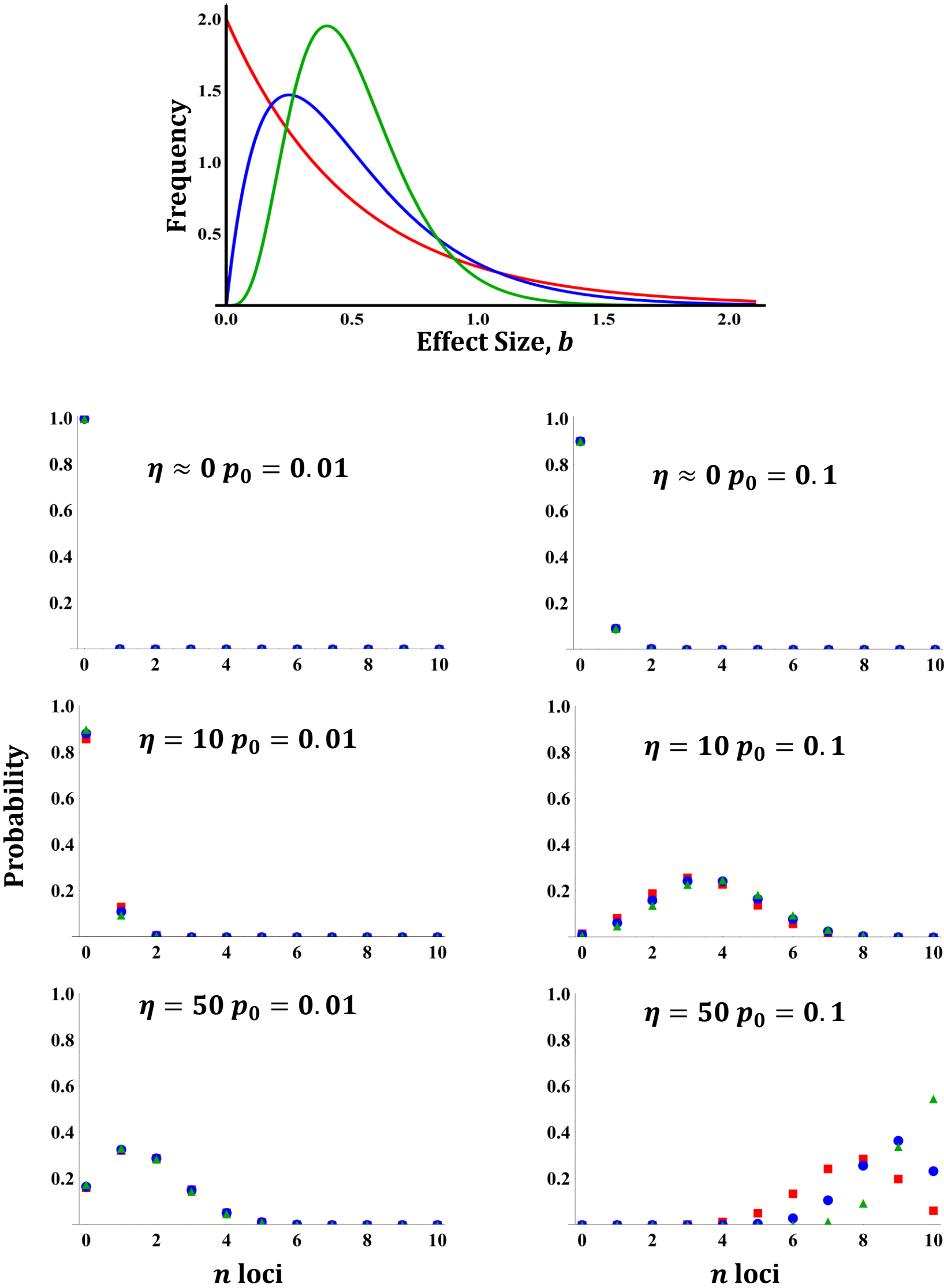
Figure 4

Time average η vs. estimated $\widehat{\eta}$ for QTL method

Figure 5:

Fraction of estimates with credible interval not overlapping $\eta = 0$
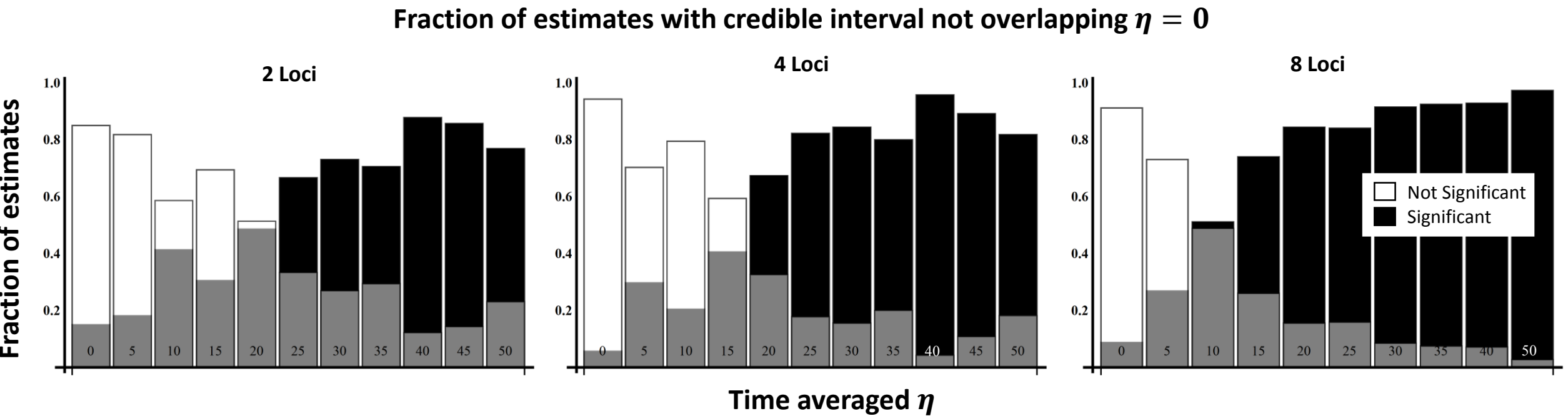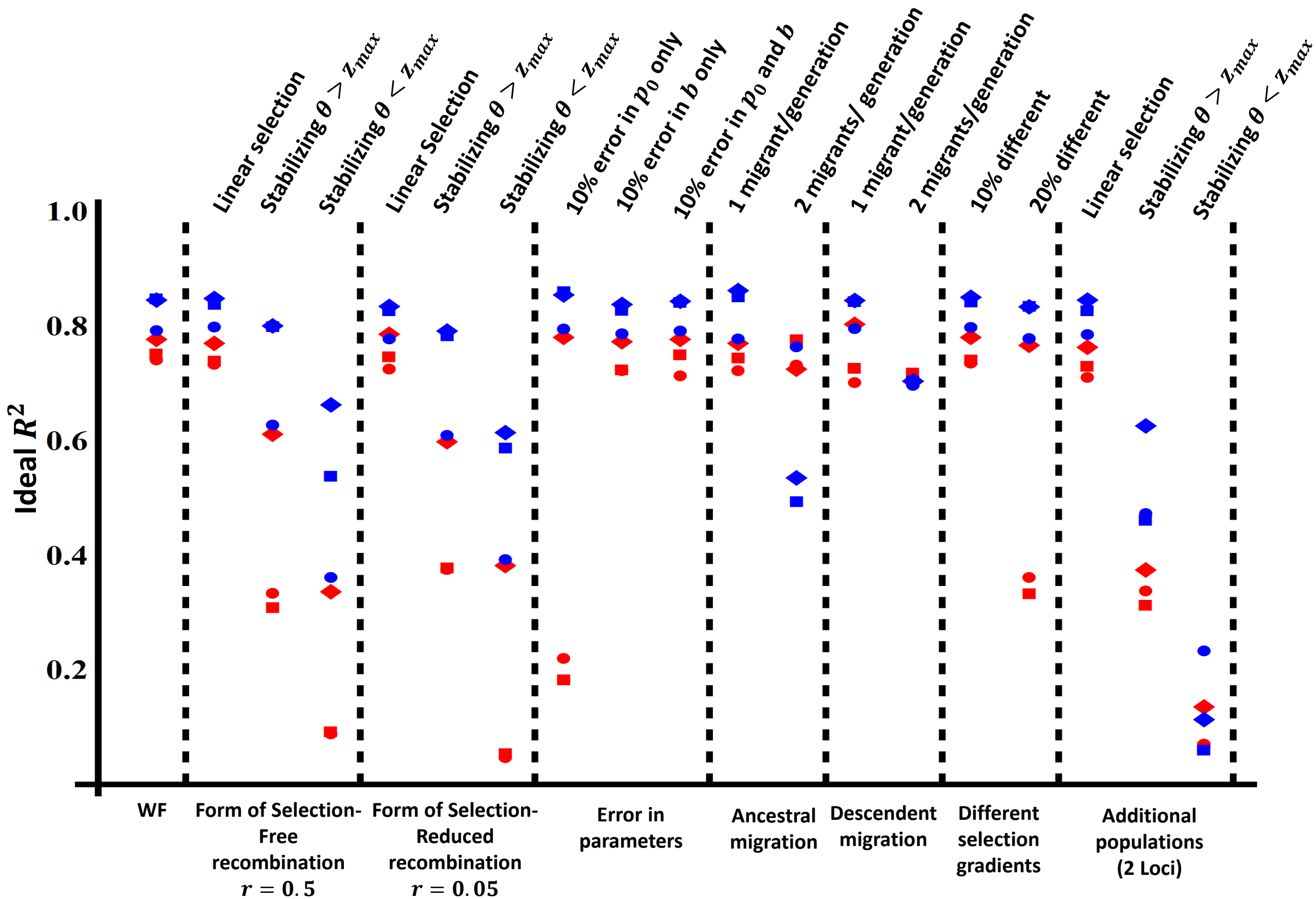
Figure S1:

Supplementary Material:

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

We study 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

$$z = \bar{z} + \sum_{i=1}^{n} b_i \zeta_i + e_z, \tag{S1}$$

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 phenotype is given by

$$W(z) = \beta z + \alpha, \tag{S2}$$

where 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,

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

Here $\bar{z}$ is the average phenotype. 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 $\beta = 0$

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

Substituting equation (S1) into equation (S3) results in an expression for relative fitness as a function of the individual effect of each locus,

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

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

$$w(z) = 1 + \sum_{i}^{n} a_i(\zeta_i) + \sum_{i}^{n} \sum_{j<i}^{n} a_{i,j}(\zeta_i \zeta_j - D_{ij}) + \cdots, \tag{S5}$$

where D_{ij} is the linkage disequilibrium between the ith and j^{th} locus, 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

$$a_i = \frac{\beta}{\alpha} b_i, \tag{S6a}$$

$$a_{i,j} = 0. \tag{S6b}$$

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),

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

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 of the same order as the selection coefficient, $\mathcal{O}(a_i)$ (Nagylaki, 1993, Nagylaki et al., 1999). This allows us to simplify the change in allele frequency in (S7) to

$$\Delta p_i = a_i p_i (1 - p_i) + \mathcal{O}(a_i^2), \tag{S8}$$

which further simplifies to equation (3) of the main text after substituting the value of a_i from equation (S6a).

Because (S8) shows that evolution of allele frequencies across loci is independent at QLE, 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,

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

(Hartl & 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 & Taylor, 1981):

3

$$P_{fix} \approx \frac{\left(1-e^{2Nsp_0}\right)}{1-e^{2Ns}} = \frac{\left(1-e^{2N\frac{\beta}{\alpha}p_o}\right)}{1-e^{2N\frac{\beta}{\alpha}}}, \tag{S10}$$

which corresponds to 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 mth 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 they refer to. The resulting probability of parallel evolution across multiple loci now represents a product of the single locus result (S10) across loci, denoted by the i subscript.

As mentioned in the main text, equations (6) and (7) represent a very restrictive definition of parallel evolution, requiring repeated adaptation in *all m* descendent populations. Alternative definitions of parallel evolution can however be easily derived from equation (5) of the main text. For example, one plausible alternative definition of parallel evolution is the repeated fixation in *x* of the m descendent populations. The probability of parallel evolution defined in this way can be approximated as

$$P_{\parallel} = \binom{m}{x}\left(\frac{\left(1-e^{2N\frac{\beta}{\alpha}b_i p_{0_i}}\right)}{1-e^{2N\frac{\beta}{\alpha}b_i}}\right)^{\mathrm{x}}\left(1-\left(\frac{\left(1-e^{2N\frac{\beta}{\alpha}b_i p_{0_i}}\right)}{1-e^{2N\frac{\beta}{\alpha}b_i}}\right)\right)^{m-x}, \tag{S11}$$

where $\binom{m}{x}$ is the binomial coefficient and the second term is the probability of parallel evolution in x populations.

4

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

At the conclusion of the individual based simulations, adaptation in the descendent populations

is summarized by the genetic data $\mathcal{D}$, a 2 x n matrix of 1's and 0's indicating the fixation or loss of alleles

at the n loci in the two descendent populations. Simulations assume the values for the allelic effect

sizes and initial allele frequencies are known and these are used as inputs. We use the following

Metropolis-Hastings 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 (see below). 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. We treated the first half of the sequences as a burn-in period as the

values at these points depend primarily on the initial conditions of each sequence and not the posterior

distribution (Gelman, 2004).

The Metropolis-Hastings algorithm can be described by the following steps.

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 η^* is drawn

    from the jump distribution $J(\eta, \eta^*)$, a distribution describing the probability of

moving from η to η^*. The jump distribution used here was a uniform distribution

ranging from $\eta \pm 100$.

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 from 1 to M.

    Discard the first n points in each sequence for "burn in" so that the index i now

    ranges from 1 to n. Calculate the following expression, which is a weighted

    comparison of the variance between versus within each sequence:

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

    where:

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

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

6. If $R < 1$ simulating additional points will not significantly improve the estimation of

    the mode 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 and to migration*

among descendent populations.

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 or among descendent populations 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 estimator. We found that the estimator is fairly robust to the presence of

recurrent gene flow occurring at rate ν, remaining quite accurate even when up to $N\nu = 2$ migrants

move among populations per generation regardless of the source (ancestral or descendent) of these

migrants (Table S4, S5). Simulations with $N\nu = 3$ and $N\nu = 4$ confirm that the estimator is very robust

to migration amongst descendent populations but performs poorly under migration from the ancestral

population at these high levels. As expected however, migration among descendent populations does

increase the frequency of the selectively favored allele within these populations, leading to slight

overestimates of the strength of selection when gene flow becomes appreciable, $N\nu = 2$ Table (S5).

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

An additional assumption of our analytical model and estimator is that the strength of natural

selection is identical in the descendent populations. We tested the effect of differing selection gradients

among descendent populations with our individual based simulations and then used these simulations

to evaluate the accuracy of the 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 (S6).

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

Thus far we have discussed the robustness of the estimator to 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 by assuming 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 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 estimated initial allele frequencies and effect sizes from a Gaussian distribution centered around the true value and with a variance of either 0% or 10% of the true value, the estimated initial allele frequencies were assumed to be the same in both descendent populations. The results of these simulations are shown in Table (S7), and reveal that the 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. *pg 289-297.*
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:

# of loci	Candidate Gene Method				QTL Method			
	Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
2	15.16692	0.847294	0.798549	0.742167	19.8606	0.780984	0.829471	0.752435
4	4.313604	0.992298	0.789878	0.778282	7.161146	0.945582	0.813414	0.793367
8	0.880424	1.007933	0.849744	0.84871	1.318149	1.015741	0.848798	0.846376

Table S2:

Selection	# of loci	Candidate Gene Method				QTL Method			
		Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	16.00289	0.778181	0.786338	0.734021	20.77678	0.730456	0.815123	0.740056
Linear	4	4.908612	0.959635	0.781458	0.770902	6.543477	0.934036	0.813979	0.79955
Linear	8	1.566819	1.008783	0.84166	0.839054	3.592519	0.956031	0.853823	0.849199
$\theta > z_{max}$	2	19.04129	0.176478	0.570297	0.334428	23.47881	0.157239	0.613157	0.309989
$\theta > z_{max}$	4	6.471029	0.890184	0.653319	0.612362	7.974566	0.909744	0.686446	0.628065
$\theta > z_{max}$	8	3.919524	1.032816	0.819669	0.800306	5.610145	0.982669	0.823882	0.801448
$\theta < z_{max}$	2	15.89003	-0.04053	0.416807	0.089439	15.75335	-0.12212	0.373713	0.093054
$\theta < z_{max}$	4	6.921948	0.471407	0.474124	0.337294	7.862538	0.445404	0.501893	0.362634
$\theta < z_{max}$	8	1.353306	0.854482	0.544687	0.539096	2.607046	0.897374	0.681847	0.663801

Table S3:

Selection	# of loci	Candidate Gene Method				QTL Method			
		Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	15.87892	0.775401	0.77533	0.725998	19.81205	0.735114	0.810594	0.747602
Linear	4	6.138736	0.92206	0.798193	0.786955	9.198878	0.861285	0.796594	0.77863
Linear	8	3.654056	0.893438	0.830906	0.827941	4.976391	0.88004	0.840244	0.835328
$\theta > z_{max}$	2	13.61869	0.580798	0.554989	0.376406	14.42015	0.871419	0.587351	0.378834
$\theta > z_{max}$	4	6.962834	1.005779	0.663432	0.59926	8.065374	1.035488	0.692149	0.610299
$\theta > z_{max}$	8	6.550021	0.906787	0.80325	0.7844	7.121709	0.922785	0.816295	0.792451
$\theta < z_{max}$	2	17.98822	-0.40386	0.430633	0.047842	20.97166	0.293755	0.504912	0.05447
$\theta < z_{max}$	4	4.724858	0.706142	0.471843	0.383263	6.840519	0.609338	0.532616	0.393311
$\theta < z_{max}$	8	1.064135	0.991435	0.594402	0.58825	2.302309	0.919698	0.631262	0.614892

Table S4:

Selection	# of loci	# migrants	Candidate Gene Method				QTL Method			
			Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	1	9.462552	1.000405	0.768803	0.72351	12.86251	0.973032	0.808819	0.745215
Linear	4	1	5.85923	0.865046	0.7767	0.771027	7.986257	0.865206	0.791305	0.77893
Linear	8	1	1.388972	1.043136	0.857805	0.852324	2.034207	1.039227	0.87068	0.863085
Linear	2	2	9.651092	0.702307	0.7416	0.733154	12.04375	0.744005	0.788795	0.777064
Linear	4	2	-2.08238	0.868612	0.76941	0.726144	-2.95146	0.941102	0.787093	0.764882
Linear	8	2	-5.43055	0.795074	0.785624	0.494597	-3.55962	0.742378	0.805221	0.535897

Table S5:

Selection	# of loci	# migrants	Candidate Gene Method				QTL Method			
			Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	1	18.87484	0.594862	0.755336	0.702178	20.06695	0.652962	0.786618	0.727683
Linear	4	1	5.095263	1.016475	0.822661	0.804144	9.645867	0.915799	0.823622	0.796352
Linear	8	1	1.580135	1.002535	0.846162	0.843802	2.158628	1.006954	0.850158	0.845805
Linear	2	2	20.29168	0.707668	0.789092	0.704971	24.31354	0.682712	0.827893	0.718936
Linear	4	2	18.34987	1.089859	0.866779	0.704424	21.89371	0.996555	0.869351	0.697563
Linear	8	2	19.42666	1.171849	0.898938	0.705512	20.60433	1.133726	0.900089	0.705511

Table S6:

Selection	# of loci	difference in gradient	Candidate Gene Method				QTL Method			
			Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	10%	14.13968	0.80753	0.774469	0.736087	18.18815	0.841374	0.813448	0.741789
Linear	4	10%	5.052786	0.915427	0.787541	0.781258	5.393386	0.947227	0.808594	0.798983
Linear	8	10%	1.634081	0.989398	0.844835	0.843117	3.400554	0.928276	0.853845	0.85118
Linear	2	20%	38.23155	0.091405	0.741292	0.362411	41.03655	0.08503	0.772002	0.333705
Linear	4	20%	4.2148	0.984006	0.777632	0.767716	6.646766	0.932552	0.793882	0.779343
Linear	8	20%	0.17403	1.054042	0.838123	0.835511	1.531522	1.008696	0.837089	0.834684

Table S7:

Selection	# of loci	Error in p_0	Error in b	Candidate Gene Method				QTL Method			
				Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	10%	10%	34.83995	0.102524	0.743151	0.220846	39.56384	0.080656	0.786806	0.183825
Linear	4	10%	10%	5.392651	0.957575	0.793362	0.78166	7.479929	0.938856	0.814904	0.795675
Linear	8	10%	10%	0.34253	1.002631	0.860966	0.860806	1.26521	1.003183	0.85683	0.855334
Linear	2	10%	0%	19.93728	0.664966	0.785625	0.723271	23.69046	0.652236	0.807125	0.724296
Linear	4	10%	0%	7.29575	0.918863	0.790391	0.774034	8.508657	0.907506	0.807808	0.787979
Linear	8	10%	0%	1.90241	0.931618	0.829667	0.828801	4.181844	0.875054	0.841528	0.838672
Linear	2	0%	10%	16.70543	0.709101	0.754913	0.713937	17.49327	0.819431	0.814171	0.750978
Linear	4	0%	10%	6.935046	0.946104	0.796294	0.777997	7.947352	1.002682	0.825405	0.79228
Linear	8	0%	10%	2.06413	0.978143	0.844339	0.842551	3.512649	0.948235	0.847899	0.84454

Table S8:

Selection	# of pops	Candidate Gene Method				QTL Method			
		Intercept	Slope	R^2	Ideal R^2	Intercept	Slope	R^2	Ideal R^2
Linear	2	20.17902	0.704162	0.778937	0.711608	24.19928	0.62949	0.813061	0.731029
Linear	4	7.330957	0.92165	0.782471	0.763815	9.809639	0.983656	0.82849	0.786108
Linear	8	3.548657	0.985801	0.835579	0.827861	4.840564	0.973264	0.857317	0.846613
$\theta > z_{max}$	2	20.66103	-0.02662	0.597459	0.338874	20.21517	0.145328	0.539055	0.313713
$\theta > z_{max}$	4	9.066178	0.446497	0.47196	0.37569	7.222451	0.998224	0.585238	0.473687
$\theta > z_{max}$	8	1.706663	0.8564	0.46868	0.462201	0.693301	1.01186	0.630704	0.627024
$\theta < z_{max}$	2	18.89695	-0.59417	0.443024	0.071341	22.44341	-1.60471	0.460389	0.065313
$\theta < z_{max}$	4	5.255587	0.617513	0.288124	0.136686	5.110795	1.482254	0.402424	0.234122
$\theta < z_{max}$	8	2.001947	0.484536	0.087474	0.060867	1.274492	0.6845	0.124804	0.114393

Table Legends:

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, QTL and candidate gene.

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 of η from the simulations, 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, QTL and candidate gene. 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 of η from

the simulations, 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, QTL and candidate gene. 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 of

η from the simulations, 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, QTL and candidate gene. 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 recurrent gene flow between the 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 of η from the simulations, 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, QTL and candidate gene. 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 other descendent population.

12

Table S6: 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 of η from the simulations, 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, QTL and candidate gene. 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 S7: Estimation of η with error in p_0 and 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 of η from the simulations, 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, QTL and candidate gene. 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.

Table S8: Estimation of η with 2, 4, and 8 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 of η from the simulations, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. All simulations where run with $n = 2$ loci and a variable number of populations, $m = 2, 4,$ or 8. Estimates for the strength of selection were made using the data from 2 loci, as well as data filtered to resemble the two experimental methods, QTL and candidate gene.

Figure Legends:

13

Figure S1: Ideal R^2 values across robustness tests. Ideal R^2 values for QTL method (blue) and candidate gene method (red) estimates using data from 2 (circle), 4 (square), or 8 (triangle) loci across different robustness tests, except for the "additional populations**"** points which are all simulated with 2 loci but with a variable number of popualtions: 2 (circle), 4 (square), or 8 (triangle) popualtions. Wright-Fisher (WF) simulations represent an ideal case. Explanations of different robustness tests are given in main text as well as the supplementary sections C,D, and E, as well as table legends.