The probability of parallel genetic evolution from standing genetic variation

Journal:	*Journal of Evolutionary Biology*
Manuscript ID	JEB-2016-00344
Manuscript Type:	Research Papers
Keywords:	Theory, Ecological genetics, Adaptation, Natural selection

June 15, 2016

Dear Dr. Gardner,

Thank you for your response to our manuscript and the many helpful comments from the reviewers. We have revised our manuscript in response to these suggestions and believe it is greatly improved as a result. Many of the reviewers' comments concerned how the work fits in the context of the existing literature. With this in mind, we have clarified our definition of 'parallel evolution'. Specifically, we now state explicitly that we model parallel evolution arising from standing genetic variation rather than from new mutations. Another concern raised by the reviewers was the generality of our simulations. To address this, we ran additional simulations covering a broader range of parameters. These additional simulations allowed us to more fully evaluate the robustness of our analytical results. In the following paragraphs we provide a more detailed response to each of the specific suggestions made by the reviewers. Our responses are in italics throughout.

Reviewer 1

1 - The authors need to make much clearer from the onset of the paper what it is they are interested in. There are different levels of parallelism, at the phenotypic level, the gene level, the molecular level, etc… (as investigated empirically for instance by Tenaillon et al 2012 Science, and discussed in length in Lenormand, Chevin & Bataillon 2016, available on ResearchGate). Here the focus is on parallel genetic evolution from standing variation, for genes underlying a specific trait. It is important to be precise, because this has implication for the results and conclusions that are drawn from them.

First, adaptation from standing variation means that the authors study parallelism in the fixation, not in the origination of alleles by mutation. In contrast, earlier studies (cited in the introduction) focused on parallelism of both origin by mutation AND fixation. One could easily argue that recurrent use in different populations of an allele that is already present in their ancestor (and hence identical by descent) is not parallel evolution, since the allele did not arise independently in these populations (discussed in point 3.2 in Lenormand et al 2016); arguments can be found for the reverse, but clearly this issue needs to be addressed quite explicitly here. More importantly, when focusing only on fixation, it is somewhat obvious that stronger selection and larger initial frequency will lead to higher parallelism, while this is not the case when parallelism of mutation is included. So discrepancies between the present results and those of earlier studies are largely due to this assumption, which should be made clear. This also makes it clearer why their assumption of only two alleles (say, an ancestral and derived one) is reasonable.

Second, specifying early on that they study parallelism only with respect the genetics of a focal trait (rather than parallel genetic evolution in general as studied in other theoretical papers) will make it clearer to the reader why they try estimate a component of phenotypic selection (their parameter eta) using data of phenotypic effect size, rather than just a selection coefficient.

We now include a discussion of the many definitions of parallel evolution, including parallel genetic vs. parallel phenotypic evolution, and parallel evolution form standing genetic variation vs. new mutation. We now make it very explicit that we focus on parallel genetic evolution from standing genetic variation, ***paragraph on Ln 31, ln 331-332.***

2 - The definition used for parallelism in the paper is too restrictive. Equation (6) measures the probability of fixation in all m populations. This does not seem to be a satisfying measurement of parallelism whenever m >2, because it doesn't count any event where fixation occurred in parallel in less than m populations. For instance if 100 populations have been sampled and fixation occurred in parallel

in 99 of these populations, you wouldn't count this as parallelism with your approach. This is clearly not what people describe as parallelism in real data; think for instance of Lenski's long-term evolution experiment, where some mutations fixed in a few out of 12 replicated lines, which is still considered as strong evidence for parallelism. In your model, the number of populations where fixations occurred has a simple binomial distribution with parameters m (the total number of sampled populations) and Pfix. The probabilities in this binomial distribution thus define the probability of parallelism of "order k", that is, fixation in k populations out of m. In particular, the total probability of parallel evolution is

P// = 1 - [Pr(no fixation in any population) + Pr(fixation in only one population)]

= 1- [(1-Pfix(i))^m + m*Pfix(i)(1-Pfix(i))^(m-1)]

Your use of a very restrictive definition of parallel evolution explains why the probability of parallelism decreases with the number of populations m in eq 6, instead of increasing as it should (and does with the formula above), since sampling more populations makes it more likely that parallelism occurs between at least some of them.

We recognize that our definition of the probability of parallel evolution in (6) is indeed the most restrictive possible definition. However because multiple alternative definitions exist, for example repeated fixation in at least 2 populations or repeated fixation in exactly n population, we chose to keep this restrictive definition. These alternative definitions can easily be derived from equation (5), and we now provide a derivation for one such less restrictive definition in the supplemental online material. Importantly, for the case of $m = 2$ *on which we focus almost exclusively in this paper, these definitions reduce to the same expression. We have addressed this in detail on* ***ln 143-146.***

3 – Some important details of the model are not well explained.

For instance, this is a haploid model, although it is not stated anywhere, and the empirical examples are taken from diploids. That the model is haploid can be seen in eq. (1) for the phenotype, or eq (4) for the fixation probability. For diploids, you would only get this formula for fixation probability under codominant selection, and assuming that s is the difference in relative fitness between the two homozygotes, but this would lead to deltap = (s/2) p (1-p) rather than s p (1-p) as you have here. If the difference in relative fitness between the two homozygotes is instead 2s in diploids (which leads to deltap = s p (1-p) as here), then a factor 4 instead of 2 should be used in the exponential (e.g. p 425 in Crow & Kimura, eq. 5.23 in Hartl & Clark 2007). Besides, please use the more compact and more usual formula:

P_fix = [1-Exp(- 2 N s p_0)] / [1-Exp(- 2 N s)]

which is equivalent to yours, as can be seen by multiplying both the numerator and denominator of your formula by Exp(- 2 N s).

*We now explicitly state that we are using a haploid model (****Ln 74****) and use the more common (although mathematically identical) form of the probability of fixation,* ***equations 4,5,6.***

Another issue about modeling is that in my opinion, the Barton-Turelli-Kirkpatrick framework introduces unnecessary complexity in this particular study - especially if the bottom line is that LD = 0 at QLE and there is no epistasis -, and may confuse more than help the reader. For instance in the appendix, zeta_ij is undefined (S5). And for a reader not very familiar with this literature, Dij being on the same order as selection (say, as a_j and a_i) does not necessarily imply that the term of correlated response (hitchhiking term) is negligible relative to the direct response. Looking at (S7), this will depend on how D_ij compares to p_i*(1-p_i), which may also be small, especially since you assume that the initial frequency is low. If you want to keep using this framework, please provide more explanation.

We now include a more detailed explanation and motivation for our QLE approach, focusing more on key assumptions and less on the results of Barton and Turelli and Kirkpatrick, ***paragraph on Ln 108.***
In addition, we have modified the equations in the supplementary material to clarify the notation, ***equation S5.***

4 - There are also several issues regarding phenotypic selection. The most crucial one is that it is unclear what selection gradient was used for calculating the "real" eta parameter in simulations. In the legend of fig 4 and the results, you mention time averaging of eta, but you don't explain why this time averaging was needed in the first place, and how this was done. I suspect you used this average because stabilizing selection with an optimum produces epistatic selection, as mentioned in the paper. This causes the selection gradient (and selection coefficient of a mutation affecting the trait) to decrease in time under multi-locus adaptation (e.g. Chevin & Hospital 2008 Genetics, Matuszewski et al 2015 Evolution).

We now include a description of how selection gradients were calculated for stabilizing selection simulations and why we needed to average eta over time (the reason is as the reviewer suggests), ***ln 291-299.***

It is important that you explain this averaging well, because it strongly bears on a puzzling result you don't comment much on, namely that the eta estimated statistically is always LARGER than the actual one, under selection for an optimum (slopes larger than 1 in fig 4 and Suppl tables). One possible explanation for this result is that the selection gradient that influences fixation probability is the early one - when the allele is at low frequency - which is larger than the time-averaged value of a decreasing selection gradient. A very similar argument was put forward for the hitchhiking effect in Chevin & Hospital (2008).

Thank you for pointing this out. We believe this actually occurred because we relied on too narrow a prior. Increasing the range of the uniform distribution defining the prior for eta resulted in slopes no longer larger than 1, ***figure 4.***

A less crucial point relating to phenotypic selection is that it is somewhat unfortunate that a linear fitness function was used. First this function allows for negative finesses, but more importantly it has to rely on the QLE approximation to yield independent evolution of different loci. In contrast, an exponential fitness function would cause this without any assumption, as it produces a selection gradient (derivative of log mean fitness with respect to the mean trait) that does not depend on the current mean phenotype, and hence generates no epistasis (e.g Lande 1983, Heredity). In many ways it is thus a better null model for pure directional selection.

Although we understand the mathematical convenience of an exponential fitness function, we felt the linear fitness function provided a stronger tie to empirical studies where selection is often estimated using phenotypic selection gradients. In the end, of course, it does not matter for our qualitative results and is only a matter of presentation.

Minor points:
39-51: another difference between Chevin et al (2010) and Orr (2005) is that the latter focused on parallel use of the same gene, not of the same allele at this gene.

We have significantly reworded our presentation of Orr's and Chevin et al.'s results to help clarify the definitions of parallel evolution used in each case.

98: please mention that you assume all b_i are positive, since alleles A_i increases the trait.

We have done this, ***ln 102-103.***

141-142: important to say here that this statement only holds when there is no optimum. Whenever there is an optimum, there is necessarily an effect size beyond which alleles will overshoot this optimum.

We have done this, ***ln 149-150.***

187: can BE (and frequently are)

We have made this change, ***ln 198.***

199: I think a prior is missing from the denominator of Eq (9) to conform to Bayes formula

The conditioning on the prior in the likelihood function was missing in the numerator, ***ln 210.***

275: when introducing zmax, please mention that you impose a constraint on adaptation here, which is a limit of the model. You could also give the formula for zmax as the sum of allelic effects of all loci.

As suggested, we added the definition of z_{max} as the sum of allelic effects, ***ln 286.***

309-310: about genomic architecture, the role of variation in mutation rates across loci was also stressed by Chevin et al (2010) and Streisfeld & Rausher (2011)

As we are focusing on parallel evolution from standing genetic variation rather than new mutation this is beyond the scope of this work.

334-335: "likely" repeated twice in the sentence

We have fixed this, ***ln 356.***

489-492: please specify that these distributions of effect sizes are illustrated in the upper panel

We have fixed this, ***ln 534.***

Reviewer 2

1) The expression used for the fixation probability in equation (4) does not look familiar to me. According to Kimura (1957), it should read Pfix = (1 - exp(-4Nsp_0))/(1 - exp(-4Ns)) for diploid organisms, where s is the advantage of a heterozygous carrier. As you seem to deal with haploid individuals, you should use Pfix = ((1 - exp(-2Nsp_0))/(1 - exp(-2Ns))), where s is the selective advantage of allele A over a. The factor of 2 is dropped because the variance in allele frequency is x(1-x)/N in a haploid population, rather than x(1-x)/(2N), where x is the allele frequency.

We have reformulated our solution to fit the more common form, see reviewer 1 comment ***Eq 4,5,6.***

2) I suggest investigating the role of gene flow among descendent populations (not only from ancestral to descendent ones), as this could greatly inflate your estimates of parallel evolution.

We now include additional supplementary simulations in which migration occurs between descendent populations. At a rate of 1 migrant per generation on average there is little effect. Introducing 2 migrants per generation, as expected, increases the estimates of η relative to the true values as gene flow helps bolster the frequency of the derived alleles, ***supplementary table S6.***

3) Given that the distribution of allelic effect sizes is a focal key factor (l. 66) of the current analysis, I'd have expected a more comprehensive analysis of its effect and a more detailed discussion. Figure 3 shows that the shape and rate parameters of the gamma distribution do not have a strong influence on the probability of parallel genetic evolution unless selection (eta) is substantial and the initial allele frequency (p_0) about 10% or more. The qualitative differences between the three distributions evident in the bottom right panel need to be addressed. What feature of the distribution drives this pattern? Is there an effect of kurtosis, or the relative abundance of small vs. large-effect alleles?

We now include a short description of the effect of distribution shape, in particular that the distribution mode drives the pattern seen in figure 3, ***ln 179-182.***

4) Important results apparently supporting the robustness of the approach are currently presented in multiple supporting tables and it is hard to figure out trends and differences. It would be nice if the authors could work these out a bit more verbally, or even come up with a way of presenting them graphically. In particular, the meaning of the intercept (and strong differences between various settings) are not motivated and can lead to confusion. I also found it difficult to convince myself that the "GG" method outperforms the "GC" method.

We have added a supplementary figure highlighting the major results across tables ***Figure S1.*** *In addition we now consistently label the two experimental methods as the "QTL method" and the "candidate gene method".*

6) The link to previous works by Orr (2005) and Chevin et al. (2010) is made in the Introduction. It would be nice if the authors could come back to these articles in the Discussion and summarize what we learn from this novel analysis. I also missed a mentioning of Ralph and Coop (2015; PLoS Genet) who studied convergent evolution in a spatially more complex setting, allowing for gene flow as well as explicitly incorporating the effect of recurrent mutation.

We now include a discussion of Ralph and Coop's results in connection with the probability of parallel evolution from standing genetic variation versus new mutation, ***ln 62-65.***

7) For highly quantitative traits, a substantial level of adaptation may be reached even when not every single underlying locus is fixed for the favoured allele. Yet, the authors apply a strict definition of parallel genetic adaptation as the situation where all underlying alleles are fixed (l. 159-163). I wonder how the conclusions would be affected by two relaxations, namely that i) fixation does not need to occur at all, but only a proportion, of underlying loci, and ii) fixation does not need to be complete - which would be the case if there were gene flow from the ancestral population or other sources.

We now explicitly state this result is for the special case of perfect parallel genetic evolution, ***ln 167-169.*** *Our model, as well as our individual based simulations, focus on the probability of fixation but do not require that alleles to be completely fixed. In our individual based simulation, an allele is considered "fixed" if its frequency surpasses a specified threshold (allele frequencies* > 0.99 *or* < 0.01*) see* ***ln 271.*** *This threshold corresponds to the limits of detection for polymorphism within natural populations. In addition, as shown in the supplementary table (S4) ancestral migration or migration between descendent populations (Table S5) has little effect on the accuracy of the Bayesian estimator.*

Minor comments
l. 8: Insert "that parallel genetic evolution" after "effective population size, and".

We kept the original wording

l. 19-10: Replace "how genomic architecture shapes adaptation" by "how genomic architecture impacts adaptation".

We kept the original wording

l. 27: Is this reference to Hohenlohe et al. (2010) at the correct position, i.e. is this really where "parallel evolution) is defined/reviewed for the first time?

We have clarified the reference to Hohenlohe et al. 2010, ***ln 29-30.***

l. 30: Replace "shape" by "influence". As a comment: the genetic architecture is also expected to evolve in response to selection pressure, but most likely on a longer time scale.

This section has been significantly revised, ***ln 44.***

l. 32: Are all the k alleles adaptive? If so, please insert "adaptive" or "beneficial" after "possible".

We now explicitly state that the alleles were assumed to be beneficial, ***ln 45.***

l. 34: I suggest introducing "gene reuse" in a sentence; it is a bit confusing as it also applies to "allele reuse" in your context.

This section has been significantly reworded and no longer refers to "gene reuse".

l. 49-51: It would be nice to have one more sentence summarizing Chevin et al.'s (2010) findings on the distribution of allelic effect sizes a bit more.

To understand Chevin et al.'s results concerning the effect size distributions would require significant explanation of their model and these results are somewhat tangential to our explanation as they focus primarily on pleiotropy which we do not discuss here.

l. 55: Insert comma after "systems".

This section has been significantly reworded.

l. 62: I find "extends" not exactly matching and suggest "complements". Then I would replace "complementary" by "alternative" in l. 59 to avoid repetition.

This section has been significantly reworded.

l. 64: Insert "genetic" after "parallel".

This section has been significantly reworded.

l. 65-70: Given the outline of discussing the effects of the allelic effect size distribution and the importance of the experimental design, I would have expected more than only a sentence about these topics in the Results/Discussion.

We now discuss in more detail the effect of the effect size distribution on the probability of parallel evolution, particularly in the context of figure 3, ***ln 180.***

l. 79: Insert comma after "example".

We have done this, ***ln 79.***

l. 87-88: Please explain "genetic complementation tests" or add a reference.

We have added a reference, ***ln 89.***

l. 88/93: Please introduce the abbreviations for the two tests eventually used later on in the text and tables.

We now consistently refer to the two experimental methods as the "QTL method" and the "Candidate gene method".

l. 96: Please specify that you consider a haploid model.

We have done this, ***ln 74.***

l. 98: Insert "associated with allele A" after "b_i" to make clear that allele A is the one that increases the trait value.

We have added this, ***ln 102.***

l. 99-100: Delete ":" after "by" and insert comma immediately after the equation.

We have done this.

l. 104: It is a bit misleading that you set the initial frequencies equal for all descendent populations, but nevertheless use an index i for p_0. I realise that this is to clarify the notation when multiplying in equation (7). Perhaps mention that you keep the index i for clarity.

The i index is over the loci within *each population not across populations.*

l. 107-108: Delete ":" after "expression" and insert comma immediately after the equation.

We have done this, ***ln 99***

l. 118: Insert a full-stop immediately after equation (3).

We have done this, ***ln 122.***

l. 124: Delete ":" after "by".

We have done this, ***ln 129.***

l. 128-129: Delete ":" after "as" and insert comma immediately after the equation.

We have done this, ***ln 133.***

l. 122-129: This could be written more compactly; you only need to show what is currently equation (5); what is currently equation (4) is shown in the SI, which is sufficient.

We kept the original phrasing, as we feel it helps the development of the model.

l. 134: Delete the second "simple".

We have done this, ***ln 138.***

l. 143: Insert "genetic" after "parallel".

We added this, ***ln 151.***

l. 148: Change "figure" to "Figure".

We have changes this, ***ln 156.***

l. 169-170: An interpretation/explanation is missing. See comment 3) above.

We have now added a detailed explanation ***ln 179-182.***

l. 173: Overall, I think that the fact that the approach is Bayesian is over emphasized (e.g. it is unnecessary to say that it is Bayesian on l. 354), given that you do not explore alternative choices of the prior of eta. The main result of this paper is to provide the likelihood function.

We have tried to remove the emphasis on the Bayesian aspect of the estimator, shifting the focus instead to the posterior distribution and distinguishing $\hat{\eta}$ from 0. See the revised ***figures 4 and 5****.*

l. 180: I agree that the limit of eta -> 0 theoretically corresponds to the case of no selection. However,

the diffusion approximation states that drift will be dominating if eta < ~1. So, the biologically relevant threshold is not eta = 0, but eta ~ 1.

We see the reviewer's point but believe eta = 0 is the correct criterion. Specifically, although it is true that only once eta > 1 does selection become the dominant evolutionary force, any value of eta > 0 indicates that natural selection has played some role, and thus that parallel adaptive evolution has occurred to some extent. A strength of the approach we advocate here, is that a posterior distribution for eta is returned, allowing the investigator to adhere to whichever threshold they wish. We now clarify this issue on ***ln 348-350*** *.*

l. 193: Insert comma immediately after the equation.

We have done this, ***ln 204.***

l. 209 ff.: When describing the inference procedure, please state the maximum n (no. of loci) that you used (Figure 3 implies 10, but I was not sure). More importantly, I suspect that your approach scales very well with the number of loci and populations. If so, it would seem worth emphasizing that (e.g. in the Discussion).

Figure 3 illustrates the probability of parallel evolution at n out of 10 total loci, ***ln 534****, whereas the Individual based simulations (as shown in figure 4) were run with either 2,4 or 8 loci,* ***ln 540****. Indeed, the number of loci depends both on the number of populations and the number of loci per population as described in the discussion,* ***ln 382-383.***

l. 214-215: Did you draw the allele frequencies from a uniform distribution between 0 and 0.1? Please clarify.

We state that the effect sizes were drawn uniformly between 0 and 1, ***ln 228.***

l. 220-222: It was not clear to me what you mean by this.

We have reworded this more explicitly to help clarify how the two experimental methods were modeled in the simulation, ***ln 229-242***

l. 222: Replace "effect" by "affect".

We have done this, ***ln 232.***

l. 228/229: Delete the apostrophes after "F1" and "QTL".

We have done this, ***ln 238,239.***

l. 230/233: You do not seem to return to n_CG and (CG or GC?) and n_QTL. Is it necessary to introduce these variables?

We now consistently refer the two experimental methods and no longer discuss the effective number of loci.

l. 232: "D under this method..." -> "Under this method, D...".

We have significantly reworded this section.

l. 235: The rejection criterion in the supplementary material suggests you are using a Metropolis-Hastings algorithm, not a Metropolis algorithm – unless the jump distribution is the same for all steps, in which case you should specify this in the supplementary material. Metropolis would start with an uppercase letter.

We have fixed this, ***ln 246.***

l. 241: Shouldn't this be the other way round, i.e the number of candidate genes under the QTL method is at least as large as the one under the candidate genes method?

Yes, we have fixed clarified the wording, ***ln 251-253.***

l. 243: Insert "the" before "data".

We have done this, ***ln 255.***

l. 254: Fix the starting quotes for "reproduction", which are currently typeset as ending quotes.

We have done this, ***ln 265.***

l. 259-260: I suggest tracking the fixation times when you repeat the analyses, and report them. Is it realistic to assume that there is enough time for fixation in reality?

Under linear selection the average time to near fixation (p=0.1, p=0.9) of 8 loci was 1186 generations, whereas the average time to near fixation of 2 loci was only 586 generations. In both cases, the distribution of fixation times was highly left skewed. This suggests sufficient time has passed for fixation in many well-studied cases of parallel evolution, even those that are considered to be the result of relatively recent divergence. For instance, marine and freshwater three-spine stickleback are thought to have diverged less than 10,000 years ago (Barrett and Schluter 2008). Using the average lifespan of 3.6 years as a conservative estimate for generation time of stickleback (DeFaveri and Merila, 2013), this suggests a bare minimum of 2778 stickleback generations have occurred since divergence of marine and freshwater forms, a number far in excess of the time required for fixation in our simulations.

l. 263: See comment to l. 235.

We have fixed this, ***ln 274.***

l. 265: "Individual based" -> "individual based".

We have changed the capital letter, ***ln 276***

l. 270: Insert a comma immediately after the equation.

We added this, ***ln 280.***

l. 271: "optima" -> "optimum" (singular).

We have changed this, ***ln 281.***

l. 274: See comment to l. 271. No need to repeat "theta".

We have fixed this, ***ln 284.***

l. 284: Insert comma after "linear".

We have added this, ***ln 399.***

l. 293-295: Please also assess the sensitivity to gene flow among descendent populations (see comment 2 above).

We now include results of simulations with migration among descendent populations.

l. 296: Please correct the references to Tables S4-S6 if necessary.

The tables and table legends have been redone.

l. 299-303: This is because there is a likelihood function, not because the framework is Bayesian!

We remove the reference to Bayesian, ***ln 320.***

l. 301-307: I suggest thinking in terms of "biological", not statistical, significance, i.e. eta > 1 should be your criterion. See comment to l. 180 above.

See previous comment.

l. 310: Perhaps "inherently" instead of "inexorably"?

We have changed this, ***ln 328.***

l. 311: I wonder if you want to be a bit more conservative in your formulation, as you only partially formalize the connection between the genetic architecture and parallel genetic evolution. You basically ignore linkage, assuming quasi-linkage equilibrium.

We have reworded this and included references to Orr and Chevin to clarify our contribution, ***ln 327-333.***

l. 319-321: It is not clear what you mean by "another piece of genetic natural history". Do you mean "demographic history"? Please clarify.

*We have clarified the text**, ln 345***

l. 332: Insert comma after "In contrast".

*We have done this, **ln 354***

l. 335-337: This result falls short of being trivial and basically directly follows from Barton and Turelli (1991) and Kirkpatrick et al. (2002).

*We now clarify that this result was expected, **ln 357-358***

l. 336: Insert "the" before "probability".

This section has been reworded.

l. 337-340: Are you implying that in this stickleback example variation in initial frequency prevented detection of a signal? Please add a clarifying sentence.

*Yes, variation in initial allele frequency could mask any correlation between effect size and repeated gene use, **ln 364-365***

l. 348: Delete "terribly".

We have done this.

l. 356: Add a comma after "For example".

*We have done this, **ln 379.***

l. 357: "8 total loci" -> "8 loci in total". Also, would it not make more sense to talk about "data points" rather than "loci" here?

We kept the original wording.

l. 361: Please add references for the beach mice and cave fish examples.

*We have added the references, **ln 384.***

l. 363: Insert a comma after "evolution".

*We have added this, **ln 386.***

l. 370: "be" -> "by".

We have done this.

l. 389: "estimate for" -> "estimate of".

We have changed this, ***ln 412.***

l. 390: Delete "Bayesian".

We have done this ***ln 413.***

l. 392: "our approach" -> "this method".

We have changed this ***ln 415.***

l. 392-393: "on parallel genetic evolution" -> "from multiple populations with a common ancestry."

We have changed this ***ln 415.***

l. 404-405: Species names should be italic. Also applies to l. 447, l. 450-451, and l. 462-663.

We have changed this.

l. 406: Use lowercase initial letters. Also applies to l. 414-415, l. 426-427, l. 432-433, l. 439, and l. 452.

We have changed this.

l. 476/478: Remind the reader of the meaning of GC and GG; they have not been introduced before and I found it difficult to relate GC and GG to "candidate gene method/test" and "QTL method". Would, e.g., "CG" and "IQ" for "candidate gene" and "independent QTL test" perhaps be better options?

We no longer refer to the GC and GG methods.

l. 483: The non-linear increase of the probability of parallel evolution is not surprising given the logistic form of Eq. (6).

We agree.

l. 487: Delete comma after "n loci".

We have done this, ***ln 531.***

l. 488: Insert commas after "n = 50" and after "frequency".

We have done this, ***ln 532.***

l. 490: "dipicted" -> depicted

We have done this, ***ln 534.***

Figure 3: Please make clear that the first panel shows the three effect-size distributions that are

considered. Please provide an interpretation of the pattern in the bottom-right panel (eta = 500, p_0 = 0.1). Why do the curves for different input distributions look the way they do?

We have added a reference to the top and bottom panels both in the text and the figure legend. See reviewer 1's comment.

Figure 5: It is not clear what the difference between panels A and B is in terms of parameters. Please specify the values of eta, b and p_0 in the caption.

We have changed figure 5 significantly.

Supplementary Material:

Middle of p. 1: Insert a comma after "Given this simplification".

We have done this.

Around equation S4: Replace ":" by a comma after "each locus". Add a full-stop after the equation. Replace "=" by an approximately equal sign.

We have done this.

First line after equation (S5): Delete "the alleles at"; "loci" -> "locus".

We have done this.

Around equation (S6): Delete ":" after "reveals that"; put the two equations on separate lines and typeset "and" in regular, not italic font.

We have done this.

Around equation (S7): Replace the full-stop after "Kirkpatrick et al. (2002)" by a comma; add a comma after the equation; change "Where" to "where", and typeset the locus indices i and j in italic.

We have changed this.

Two lines above equation (S8): Insert "coefficient" after "selection".

We have added this.

Around equation (S8): Delete ":" after "in (S7) to"; add a comma after the equation; "which simplifies to..." -> "which further simplifies to..."; "upon" -> "after".

We have done this.

Equation (S10): add a comma immediately after the equation.

We have done this.

First line on page 4: Adjust according to the comment to l. 122-129 above.

We have rewritten equation S10 into the more classic form, but have kept this equation in the supplementary material for clarity.

Last sentence before "B: Markov Chain Monte Carlo Simulations of Posterior Distributions:" Please fix the following formulation: "the product a single locus parallel evolution events across loci".

We have clarified this.

First paragraph of "B: Markov Chain Monte Carlo Simulations of Posterior Distributions:": Insert an opening sentence. "At the conclusion of the individual based simulation we have..." -> "The individual based simulations provide...; "consists" -> "consist". "2 X n" -> "2 x n". Remove apostrophes after "0" and "1". If appropriate, replace "Metropolis algorithm" by "Metropolis-Hastings algorithm". It would be good to tell the reader that details will follow later. E.g. insert "(see below for details)" after "...of the algorithm respectively". "can be computed" -> "can be quantified". Use another variable than m for the number of chains, as m is already used for the number of loci. Omit the entire part starting with "One important feature..." and ending with "... in a sequence (Gelman 2004)". Just say "We treated the first half of the sequences as burn-in period".

We have made these changes, using the capital letter M to denote the number of sequences.

Before the description of the algorithm, add a transitional sentence.

Algorithm: Step 3: Please specify the jump distribution. Step 5: "now ranges between 1 and n" -> "now ranges from 1 to n". Add a comma after the equation for R; replace "Where:" by "where". Give references for the equations for B and W, derive them, or make clear they are also given in Gelman (2004).

We have clarified our use of a uniform jump distribution and made these changes.

Second sentence in "C: Sensitivity of the Bayesian Estimator to migration from the ancestral population": I disagree with "Although this is likely true in many well-studied cases". In most cases, there is probably quite some gene flow from the ancestral population, as well as admixture among derived populations.

We have extended our simulations to include higher migration rates as well as migration between descendent populations.

Last two sentences on page 6: Omit the sentence starting with "Although this assumption is...". "We tested this possibility by..." -> "We assessed the effect of varying strengths of selection by...".

First paragraph of page 7: I found that quite interesting. At what stage do you see differences, i.e. an effect of different selection gradients among populations? Did you also compare different absolute magnitudes of the selection coefficient?

We ran additional simulations with twice as much discrepancy between the selection gradients and see only very slight decreases in the accuracy of the estimates hence it is likely that even not very similar "similar selective environments" can lead to parallel evolution and accurate estimation.

Section "E: Sensitivity of Bayesian Estimator to error in parameters": "to many violations" -> "to violations"; "with an assumption of their own" -> "under an overarching assumption"; "centered about"-> "centered around".

We have made these changes.

Tables S1 to S6: Please make clearer in the captions that only Table S1 is based on the Wright-Fisher simulations, and all the others on the individual-based simulations. Please remind the reader of the meaning of the intercept, as it varies substantially among comparisons.

We have reworded the table legends.

The caption to Table S4 seems to be missing.
The captions to Tables S5 and S6 seem to be confounded.

We have made new Tables and legends.

We hope that our revisions adequately address the concerns of the reviewers and associate editor. Please let us know if there are any other changes we can make to improve our paper.

Sincerely,

Ailene MacPherson

And

Scott L. Nuismer

Abstract:

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

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

Introduction:

As the availability of genome sequences has increased, interest in understanding how genomic architecture shapes adaptation at both the genetic and phenotypic levels has grown substantially (Stapley et al., 2010). How and which genes respond to selection is a complex result of many aspects of the genotype to phenotype map, including allelic effect sizes, epistatic interactions, linkage disequilibrium, and pleiotropy. Significant work using natural populations (Nadeau & Jiggins, 2010), experimental evolution (Qi et al., 2016, Wichman et al., 1999), and evolutionary theory (Orr, 2005, Chevin et al., 2010) has been devoted to elucidating how these many factors interact to shape adaptation. Particularly useful natural systems for addressing such questions are those exhibiting parallel evolution. Many striking examples of repeated phenotypic and genetic changes exist (Conte et al., 2012, Stern, 2013, Martin & Orgogozo, 2013), putatively as a consequence of adaptation to similar selective environments (Schluter, 2009). These systems can be viewed as natural experimental replicates for understanding the interplay of selection and genetic architecture in shaping patterns of adaptation (Hohenlohe et al., 2010).

There are many definitions for "parallel evolution", a phenomena which may or may not be distinguished from "convergent evolution". These many definitions all share the common theme of repeated evolution in two or more populations, but differ in two major ways. First, in terms of whether or not these populations originated from a recent common ancestral population or are only distantly related. Second, definitions differ in the biological level at which repeated evolution occurs, ranging from the genetic to the phenotypic (Lenormand & Chevin, *In press.*). Here we focus on parallel genetic evolution defined as the repeated fixation of identical alleles in multiple descendent populations. We assume these alleles were initially segregating within the ancestral population at low levels such that we focus on parallel evolution from standing genetic variation. Our interest in this definition is motivated by

a number of biological systems where adaptation to a novel environment is thought to result from standing genetic variation present in the ancestral population (Hoekstra et al., 2006, Steiner et al., 2007, Colosimo et al., 2005) . In contrast to de novo mutation, adaptation from standing genetic variation is likely rapid (Barrett & Schluter, 2008) and may lead to distinct genomic signatures of parallel adaptation (Roesti et al., 2014)

Understanding the conditions that promote parallel genetic evolution has been facilitated by theoretical studies. For instance, Orr (2005) calculated the probability that 1 of k beneficial mutations arises and fixes repeatedly and found that parallel evolution becomes more likely as the strength of selection increases and the number of possible alleles, k, decreases. These results are supported by experimental adaptation of the bacteriophage $\phi X174$ to high temperatures (Wichman et al., 1999) and adaptation of antifungal drug resistance in *Saccharomyces cerevisiae* (Anderson et al., 2003). Taking a different approach, Chevin et al. (2010) calculated the probability that a beneficial mutation fixes at the same genetic locus in independent populations. By allowing mutations to influence multiple phenotypes simultaneously, this work demonstrated that the probability of parallel evolution is greatest when pleiotropy is weak. In addition, this work demonstrated that when mutations have pleiotropic effects, the probability of parallel evolution is greater when populations are relatively close to their adaptive optima (i.e., not too maladapted). Together, these previous theoretical studies provide a solid framework for understanding the likelihood of parallel evolution arising from the fixation of novel mutations.

Although understanding the contribution of new mutations to parallel evolution is inarguably important, in some systems it may be more relevant to understand the likelihood of parallel evolution from standing genetic variation. For instance, in the stickleback, *Gasterosteus aculeatus*, repeated adaptation to freshwater is thought to involve genes already segregating at low frequencies within the marine populations (Colosimo et al., 2005). In cases like these, the presence of adaptive alleles in the

ancestral population can have a significant effect of the probability of parallel evolution, influencing both the long term probability of parallel adaptation as well as the rate at which adaptation occurs (Ralph & Coop, 2015). Our focus here is to further develop our understanding of parallel evolution by developing a genetically explicit multi-locus framework for predicting the probability of parallel evolution from standing genetic variation. We have two specific goals: First we will develop a multi-locus theory of parallel genetic adaptation that allows us to predict the probability of parallel evolution in terms of quantities that are regularly measured in natural populations. Second, we will develop a statistical framework that uses routinely collected genetic data to estimate the historical strength of parallel selection.

The Model

Biological Scenario:

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

Figure 1B illustrates the first experimental design where parallel genetic evolution is assessed at a set of candidate genes. To identify possible candidate genes, individuals from at least one descendent population (descendent population 1 in Figure 1B) are crossed with ancestral individuals and the resulting offspring are scanned for divergent QTL's. The remaining populations (descendent population

2 in Figure 1B) are then tested for the candidate genes using a variety of approaches such as genetic complementation tests (Hartl, 2005). This method, which we will call the candidate gene test, has been used in human populations to identify the genetic basis of the multiple independent origins of lactose tolerance (Ingram et al., 2009, Enattah et al., 2008, Tishkoff et al., 2007). Alternatively, in the second design (shown in Figure 1C), descendent populations are searched independently for the genes responsible for the repeated phenotypic divergence from the ancestral population. This is done by performing independent QTL scans in each descendent population. This "QTL method" was used to identify separate genes responsible for a change in developmental rate in two populations of *Oncorhyncus mykiss,* (Robison et al., 2001, Nichols et al., 2007, Sundin et al., 2005).

Analytical Model

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

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

where X_i is an indicator variable taking the value 1 if the individual carries the A allele at locus i and the value 0 if the individual carries the a alleleat locus i. We assume $b_i$is positive for all i implying that the A_i allele always increases the phenotypic trait z. The frequency of the A_i allele is given by $p_i$and $\bar{z}$ denotes the average phenotype of the population. We assume the phenotype of the ancestral population is small, meaning that the frequency of the A_i allele is low at all loci, and initially equal to p_{0_i}. Within the new environments, individuals experience selection for large phenotypes, favoring an increase in frequency of the A alleles.

The biggest challenge to modeling evolution across multiple loci is that epistasis and linkage disequilibrium make it extremely difficult to formulate analytical predictions for the probability of fixation at individual loci. Two key assumptions, however, make calculating the probability of fixation

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

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

Second, we assume the strength of linear directional selection, β, is weak $O(\epsilon)$ and the rate of recombination between loci relatively high. Under these conditions, recombination breaks apart linkage disequilibrium more quickly than it can be built up by selection, allowing a quasi-linkage-equilibrium (QLE) to be reached where linkage disequilibrium is also small, $O(\epsilon)$ (Nagylaki, 1993, Nagylaki et al., 1999). Using the expression for the phenotypic trait z, given in equation (1), as well as the expression for fitness, given by equation (2), we can use the multi-locus methods developed by Barton and Turelli (1991) and expanded by Kirkpatrick et al. (2002) to derive the change in the frequency of the A_i allele at QLE over a single generation (*See supplementary material*):

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

Equation (3) reveals that at QLE, the effects of epistasis and linkage disequilibrium are negligible and hence loci evolve independently. Later, using individual based simulations, we will relax these key assumptions and evaluate the robustness of this analytical approximation.

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

$$P_{fix} = \frac{(1 - e^{2Nsp_0})}{1 - e^{2Ns}} \quad (4)$$

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

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

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

The probability of parallel genetic evolution at a single locus:

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

$$P_{\parallel} = \left(\frac{\left(1-e^{2N\frac{\beta}{\alpha}b_i p_{0_i}}\right)}{1-e^{2N\frac{\beta}{\alpha}b_i}}\right)^m \qquad (6)$$

Requiring repeated fixation in all m populations represents a very restrictive definition of parallel genetic evolution and in some cases a less restrictive definition may be preferable. In such cases, it is straightforward to develop expressions for the probability of repeated fixation in any subset of m populations using (5). An example of the calculations for a less restrictive definition are provided in the online supplemental material.

Equation (6) highlights four important factors that will influence the probability of observing parallel genetic evolution. First, the probability of repeated fixation of an allele increases with its initial frequency, p_{0_i}. Second, large effect alleles, those with large b_i, are more likely to fix in parallel under directional selection. This relationship between effect size and parallel evolution is shown in Figure 2. Third, parallel genetic evolution is more likely to occur when evolution is driven more by the

deterministic force of natural selection than the stochastic force of random genetic drift. Specifically, the probability of parallel evolution increases with the product of population size and the phenotypic selection gradient in derived populations, $N\frac{\beta}{\alpha}$. This product, which we will denote collectively as η, captures the balance between drift and selection and shows that parallel evolution is more likely in large populations experiencing strong natural selection as shown by the three curves in Figure 2.

The probability of parallel genetic evolution at multiple loci:

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

$$P_{\parallel} = \prod_{i=1}^{n}\left(\frac{\left(1-e^{2N\frac{\beta}{\alpha}b_i p_{0_i}}\right)}{1-e^{2N\frac{\beta}{\alpha}b_i}}\right)^{m} \quad (7)$$

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

The first, and most obvious, insight to emerge from (7) is that in the special case of perfectly parallel genetic evolution, where all loci are fixed for the selectively favored A_i alleles in all descendent populations, becomes less and less likely as the number of loci increases. This is a simple result of the product rule of probabilities, and arises because the overall probability of parallel evolution decreases as each additional locus is required to fix in parallel in the m descendent populations. The second insight that emerges from equation (7) is that when selection is relatively weak, population sizes are relatively

small, and adaptive alleles initially infrequent, it is quite surprising to observe parallel evolution at anything other than a single locus with large phenotypic effects (Figure 3). As selection becomes stronger, population sizes larger, or adaptive alleles initially more frequent, however, it becomes increasingly likely that parallel evolution will occur at multiple loci, including loci with only moderate phenotypic effects (Figure 3). These results are, for the most part, relatively insensitive to the particular distribution of effect sizes across loci. Only in cases of strong selection and high initial allele frequency does the effect size distribution contribute significantly (bottom right panel). In such cases, the probability of parallel evolution at a large number of loci increases with the mode of the distribution. In other words is most likely when the effect size distribution is not skewed toward small effect loci (Figure 3). Together, these results suggest that the extent to which the observation of parallel genetic evolution at any particular number of loci is surprising depends heavily on the value of the parameter η.

Bayesian inferences of parallel phenotypic selection:

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

Our Bayesian approach will rely on genetic data described by a matrix, $\mathcal{D}$, where rows represent descendent populations and columns loci. Each element of $\mathcal{D}$ takes a value of 0 or 1 depending on which allele has fixed at a particular locus in a given population (Figure 1A). Using equation (7) we can

develop a likelihood function specifying the probability of observing the data, $\mathcal{D}$, given a particular value of the parameter η, and empirical estimates for the effect sizes b_i and initial allele frequencies p_0. Effect sizes can be (and frequently are) estimated using QTL scans (Conte et al., 2015, Broman & Sen, 2009, Lynch & Walsh, 1998) and initial allele frequencies can be estimated by measuring the allele frequencies in the ancestral population. This likelihood expression consists of a product of terms, one for each locus in each population. If the A allele has fixed at a locus it contributes a term P_{fix}, as defined by equation (5). Alternatively, if the A allele is lost, it contributes a term $\left(1 - P_{fix}\right)$. Thus, for m populations and n loci, the likelihood of observing the data, $\mathcal{D}$, is given by the following product

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

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

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

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

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

Wright-Fisher model. Next, we test the robustness of the estimator to violations of the assumptions of our analytical model by generating the genetic data $\mathcal{D}$ using multi-locus individual based simulations.

Wright-Fisher Simulation

We simulated the data $\mathcal{D}$ for two populations under the Wright-Fisher model by drawing a random number for each locus and population and setting $\mathcal{D}_{i,j}$ to 1 if the random number was less than p_{fix}, given by equation (5), and to 0 otherwise. The value of p_{fix} depends on the initial allele frequency at each locus,p_{0_i}, the allelic effect sizes of each locus, b_i, as well as the parameter η. For each simulation we drew the values of these parameters independently and at random. Initial allele frequencies were drawn independently at each locus from a uniform distribution between 0 and 0.1. Because our model envisions divergence of descendent populations from a common ancestor we assumed that the initial frequency at any one locus was the same in both populations. Allelic effect sizes were drawn independently for each locus from a uniform distribution between 0 and 1. The value of η for each run was drawn from a uniform distribution ranging between 0 to 50. The genetic outcome $\mathcal{D}$ simulated in this manner may not, however, resemble what would be measured using experimental methods. For example, using current genomic techniques it is not possible to identify loci that have not diverged from the ancestral state. To address how experimental methodologies affect our Bayesian estimates we considered two modified forms of $\mathcal{D}$ that resemble sampling under the two experimental methods described previously (see Figure 1). The first of these methods, the candidate gene method, (Figure 1B) assesses parallel genetic evolution at candidate genes which are known to have generated the phenotypic divergence in the first descendent population. This is often done by performing a cross between individuals from one of the divergent populations with the ancestral population and assessing the genetic variation in the F1s. Since the second divergent population is not independently assessed for divergent QTLs, under this method we only consider the columns of $\mathcal{D}$ (i.e., loci) where the A allele has fixed in the first population. The second experimental method, the QTL method, (Figure 1C)

independently assesses divergent loci in all populations. Under this method, $\mathcal{D}$ therefore contains all columns (loci) which have fixed in at least one population and hence the effective number of loci identified using this method will always be greater than or equal to the number found by the less thorough candidate gene method.

For each simulated $\mathcal{D}$, as well as for $\mathcal{D}$ modified by the two experimental methods, we estimated η using a Metropolis-Hastings algorithm as described in the supplementary material. For the prior $\pi(\eta)$ we used a uniform distribution on the interval $\eta = \pm 60$. To analyze the performance of the estimator we ran a regression of the estimated values of $\hat{\eta}$ on the true values η, using 200 data points. Overall, this analysis revealed that the estimator was quite accurate, explaining between 30% and 60% of the variation (see Table S1). In addition, our analysis showed that the accuracy of the estimates increases with the number of loci. This trend holds regardless of the experimental method used. However, the effective number of loci under the QTL method is always greater than when candidate genes are first identified in one population and then subsequently searched for in the other. The results of these simulations suggest our estimator performs quite well when the data meet the assumptions of our analytical model; however, this may not be the case for real data. In the next section, we explore the performance of our estimator using individual based simulations that allow us to violate key assumptions of our analytical model such as weak selection and frequent recombination.

Individual Based Simulation

Our individual based simulations (IBS) consider two allopatric populations, each of which has a constant size of $N = 1000$ individuals. Initial allele frequencies and effect sizes at each locus, as well as the value of η, were drawn randomly as described above under the Wright-Fisher model. Individuals within each population undergo a two stage life cycle. During the first stage, "selection", the probability that an individual survives is given by its fitness, with fitness computed using either equation (2) which describes linear selection or an expression for stabilizing selection described below. Surviving

individuals then enter the second life cycle stage, "reproduction", which consists of generating an offspring population from the remaining parental population. This is done by drawing a pair of parents at random from the pool of surviving individuals and producing an offspring from these parents by recombining the parental genomes at a specified rate r and allowing mutation between the two allelic states at a per locus mutation rate of $\mu = 10^{-6}$. This process is continued with replacement of parents until the offspring population reaches the pre-selection size of N. This life cycle is repeated until all loci approach fixation or loss (allele frequencies > 0.99 or < 0.01) at which point the simulations were terminated and the matrix of genetic data $\mathcal{D}$ filled by rounding. As in the previous section, we formulate modified versions of $\mathcal{D}$ that resemble sampling under the two experimental methods. Then, using the Metropolis-Hastings algorithm, we compute estimates for the value of η using the original outcome $\mathcal{D}$ as well as the two modified forms of $\mathcal{D}$ (*see Supplementary Material*).

We used the individual based simulations to test the robustness of the estimator when selection is strong and/or non-linear. To test the effect of non-linear selection, individual based simulations were run where an individual's fitness was determined by one of two alternative forms of selection: linear directional selection described by (2), or stabilizing selection toward a phenotypic optimum:

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

where θ is the phenotypic optimum and γ is the strength of stabilizing selection. Including simulations where selection is stabilizing is important because it relaxes our previous assumption that loci evolve independently. Stabilizing selection is particularly useful in testing this assumption because the extent of interdependence between loci can be manipulated by changing the value of the phenotypic optimum. Specifically, the extent of interdependence between loci will depend on the value of the optima relative to the largest possible phenotype $z_{max} = \sum_i b_i$. When θ is greater than the largest possible phenotype, z_{max}, loci remain relatively independent as directional selection predominates over epistatic selection.

However when $\theta < z_{max}$ this is no longer true as epistatic selection now dominates. Therefore, when $\theta > z_{max}$ evolution is much more likely to resemble linear selection as our analytical model assumed. We simulated these two forms of stabilizing selection respectively, by either requiring that θ be larger than z_{max} or slightly smaller than z_{max} (*see Supplementary material*). Under stabilizing selection the strength of selection, in the form of η, changes as the population adapts, decreasing as the population approaches the optimum (Chevin & Hospital, 2008, Matuszewski et al., 2015). Therefore we computed a "realized" strength of linear selection be averaging the selection gradient, $\frac{Cov(z,w)}{Var(z)}$ over all time points for which $Var(z) \neq 0$. When $\theta \ll z_{max}$ the resulting time averaged selection gradient will be significantly smaller than under linear selection. When $\theta > z_{max}$ the realized selection gradient will depend on the distance between θ and z_{max} as well as the strength of stabilizing selection, but in the individual based simulations is generally only slightly smaller than selection under linear directional selection.

As expected, analysis of simulated data shows that the accuracy of our Bayesian estimates depends on the form of selection. Specifically, estimates for η are most accurate under linear selection, somewhat less accurate under stabilizing selection toward a distant optimum($\theta > z_{max}$), and least accurate under stabilizing selection toward a close optimum ($\theta < z_{max}$) (See Figure 4 and Table S2). In addition to assuming that selection is linear, we also assumed that selection is weak. By computing the variance about the regression line as η increased we were able to confirm that, for the data shown in figure 4, the accuracy of our estimates decreases with increasing selection. Next, we used our simulations to explore the sensitivity of our estimator to infrequent recombination among candidate loci (See Table S3). Not surprisingly, these simulations revealed that our estimator performs better when recombination is frequent ($r = 0.5$) than when recombination is rare ($r = 0.05$). The effect of infrequent recombination is more drastic for stabilizing selection than linear selection, and is particularly pronounced when $\theta < z_{max}$. This is expected since this latter scenario generates the strongest epistatic

selection and thus has the greatest potential to cause linkage disequilibrium to accumulate. Finally, we used the individual based simulations to test the accuracy of our estimator when recurrent gene flow occurs from ancestral to descendent populations when gene flow occurs among descendent populations, when the strength of selection differs in the two descendent populations, and when estimates of the parameters p_0 and b_i are imprecise (*see Supplementary Material and Table S4-S6*). These simulations reveal that our estimates of η are robust to error in estimates of other model parameters including error in the estimated initial allele frequency, p_0.

Up to this point we have focused on using the Bayesian estimator to provide single point estimates for η. Having access to the full posterior distribution allows us to calculate a 95% credible interval for the parameter η and determine whether or not it overlaps with zero. From an empirical standpoint, being able to rule out $\eta = 0$ allows us to reject the hypothesis that observed levels of parallel genetic evolution can be explained by random genetic drift alone. The closed versus open circles in Figure 4 represent data points for which credible intervals drawn from the posterior distribution do, or do not, overlap zero. Figure 5 shows how the probability of rejecting 0 (filled bars) increases with η.

Discussion

It has long been understood that natural selection, parallel genetic evolution, and genomic architecture are inherently linked (Orr, 2005, Schluter, 2009, Chevin et al., 2010). Previous theoretical work has focused on repeated genetic evolution from new mutation and found that key components of genetic architecture, such as the number of possible beneficial mutations (Orr, 2005) and the distribution of mutational effects (Chevin et al., 2010), influence the probability of parallel evolution. Here we have used a multi-locus model of parallel evolution from standing genetic variation to continue to formalize these connections. We began our investigation by calculating the probability of parallel evolution at a single locus and showed that parallel evolution is most likely when phenotypic selection is strong, standing genetic variation for adaptive alleles is appreciable, adaptive alleles have large

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

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

involved in parallel adaptation in two stickleback populations has found no correlation between effect size and probability of repeated gene use (Conte et al., 2015). Our results suggest that the lack of correlation may be the result of highly variable initial allele frequencies among loci.

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

Such a Bayesian estimator is only useful if it produces accurate predications across a range of parameter space. Indeed, our individual based simulations reveal that our estimator can be quite accurate and robust although there are limitations. For example, accurate estimation requires data from at least 8 total loci, be that 4 loci in two populations, 2 loci in four populations, or some intermediate combination. Whether data is gathered at fewer loci in many populations or many loci in few populations should, in principle, have no effect on the accuracy or efficiency of the estimator. Many of the studies discussed above however, have far fewer than this. For example, studies of parallel pigmentation changes in a variety of species, from beach mice (Hoekstra et al., 2006) to cave fish (Protas

et al., 2006, Gross et al., 2009), focus primarily on one or two loci in somewhere between 2 and 6 populations. Therefore if future studies hope to understand the role of natural selection in driving parallel evolution, it is important that they focus on acquiring data from as many loci and as many populations as possible. A natural consequence of the increasing accuracy derived from information on a larger number of loci is that certain experimental approaches are more powerful than others for studying parallel evolution. Specifically, we have explored two alternative experimental methods (see Figure 1 B and C), that differ predictably in the number of loci that are detected. Because the QTL method always detects parallel evolution at a larger number of loci, we recommend its use over the candidate gene method. The accuracy of the Bayesian estimate is influenced not only by the amount of available data but the accuracy of the estimated parameter values b_i and p_{0_i}. We have tested the robustness of the estimate to error in these parameters (*See the supplementary Material*). Because we may often be uncertain about the exact value of p_{0_i} due to sampling error or stochastic variation in small populations (Hermisson & Pennings, 2005), in some case it may be more appropriate to run the Bayesian estimator with a prior distribution for the parameter p_{0_i} rather than a single point estimate.

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

411 Combined, our analyses of single and multi-locus models show that it is difficult to draw
412 conclusions about the biological significance of parallel genetic evolution without information on the
413 strength of parallel phenotypic selection and local population size. Our Bayesian estimator provides a
414 robust statistical methodology for translating observed levels of genetic parallelism into an estimate of
415 the product of phenotypic selection and local population size. As a consequence, our approach provides
416 a much needed tool for distinguishing between adaptive and non-adaptive hypotheses for observed
417 levels of parallel genetic evolution. Applying this method to existing and emerging data from multiple
418 populations with common ancestry may thus provide novel insights into the importance of adaptive
419 evolution in natural populations.

References:

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

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

Barrett, R. D. & Schluter, D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol* **23**: 38-44.

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

Chevin, L. M. & Hospital, F. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **180**: 1645-1660.

Chevin, L. M., Martin, G. & Lenormand, T. 2010. Fisher's model and the genomics of adaptation: Restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**: 3213-3231.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D. & Kingsley, D. M. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**: 1928-1933.

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

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., Schluter, D. & Peichel, C. L. 2015. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. *Genetics*.

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

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

Gross, J. B., Borowsky, R. & Tabin, C. J. 2009. A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *Plos Genetics* **5**.

Hartl, D. J., E.W. 2005. *Genetics: Analysis of genes and genomes*. Jones & Bartlett Learning.

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

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101-104.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. & Cresko, W. A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics* **6**.

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

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

Kimura, M. 1957. Some problems of stochastic-processes in genetics. *Annals of Mathematical Statistics* **28**: 882-901.

Lenormand, T. & Chevin, L. M. B., T.M. (*In press.*) Parallel evolution: what does it (not) tell us and why is it (still) interesting? In: *Chance in Evolution*, (Ramsey, G. P., C.H., ed.). pp. Univ. Chicago Press.

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

Martin, A. & Orgogozo, V. 2013. The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution* **67**: 1235-1250.

Matuszewski, S., Hermisson, J. & Kopp, M. 2015. Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* **200**: 1255-74.

Nadeau, N. J. & Jiggins, C. D. 2010. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics* **26**: 484-492.

Nichols, K. M., Broman, K. W., Sundin, K., Young, J. M., Wheeler, P. A. & Thorgaard, G. H. 2007. Quantitative trait loci x maternal cytoplasmic environment interaction for development rate in *Oncorhynchus myhiss*. *Genetics* **175**: 335-347.

Orr, H. A. 2005. The probability of parallel evolution. *Evolution* **59**: 216-220.

Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., Zon, L. I., Borowsky, R. & Tabin, C. J. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics* **38**: 107-111.

Qi, Q., Toll-Riera, M., Heilbron, K., Preston, G. M. & MacLean, R. C. 2016. The genomic basis of adaptation to the fitness cost of rifampicin resistance in *Pseudomonas aeruginosa*. *Proceedings of the Royal Society B-Biological Sciences* **283**.

Ralph, P. L. & Coop, G. 2015. The role of standing variation in geographic convergent adaptation. *American Naturalist* **186**: S5-S23.

Robison, B. D., Wheeler, P. A., Sundin, K., Sikka, P. & Thorgaard, G. H. 2001. Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). *Journal of Heredity* **92**: 16-22.

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W. & Berner, D. 2014. The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology* **23**: 3944-3956.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science* **323**: 737-741.

Schluter, D., Clifford, E. A., Nemethy, M. & McKinnon, J. S. 2004. Parallel evolution and inheritance of quantitative traits. *American Naturalist* **163**: 809-822.

Schmegner, C., Hoegel, J., Vogel, W. & Assum, G. 2005. Genetic variability in a genomic region with long-range linkage disequilibrium reveals traces of a bottleneck in the history of the European population. *Human Genetics* **118**: 276-286.

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P. & Slate, J. 2010. Adaptation genomics: The next generation. *Trends in Ecology & Evolution* **25**: 705-712.

Steiner, C. C., Weber, J. N. & Hoekstra, H. E. 2007. Adaptive variation in beach mice produced by two interacting pigmentation genes. *Plos Biology* **5**: 1880-1889.

Stern, D. L. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**: 751-764.

Sundin, K., Brown, K. H., Drew, R. E., Nichols, K. M., Wheeler, P. A. & Thorgaard, G. H. 2005. Genetic analysis of a development rate QTL in backcrosses of clonal rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* **247**: 75-83.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A. & Deloukas, P. 2007. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* **39**: 31-40.

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

513

514 **Figure Legends:**

515 **Figure 1: Schematic of biological scenario.** Panel A depicts two descendent populations diverging in
516 parallel from a common ancestral population. The a allele predominates at all four loci in the ancestral
517 population whereas the A allele fixes at various loci in the two descendent populations. Panel B and C
518 depict two methods for deducing the underlying genetics of reduced body size in the two descendent
519 populations depicted in panel A. Panel B shows the candidate gene method which relies on a genome
520 wide scan of progeny from a cross between the first descendent population and the ancestral
521 population and subsequent candidate gene search in the second descendent population. Panel C shows
522 the QTL method which involves two genome wide scans, one in each population. Compared to the
523 candidate genet method, the QTL method uncovers an additional locus driving divergence in population
524 2.

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

530 **Figure 3: The probability of parallel evolution at n loci.** For a trait determined by the effects of 10 total
531 loci the probability of observing parallel evolution at exactly n loci depends on the strength of selection,
532 which varies from near 0 to a value of $\eta = 50$, and the intial allele frequency, which is either low (0.01)
533 or high (0.1). The probability of parallel evolution may also depend on the underlying effect size
534 distribution depicted here (top panel) as three different gamma probability distributions with different

shape and scale parameters (Red: $k = 1, \theta = \frac{1}{2}$, Blue: $k = 2, \theta = \frac{1}{4}$, Green: $k = 5, \theta = \frac{1}{5}$,) but with the same mean effect size ($\mu = \frac{1}{2}$).

Figure 4: Regression Fit of IBS Data under the three forms of selection. Data and linear regression fit (blue), and perfect fit (red) between the time averaged values of η and the Bayesian estimate $\hat{\eta}$ for 200 replicates of the individual based simulation. Closed (open) points indicate estimates where the 95% credible interval does not (do) overlap $\eta = 0$. Sub-panels differ in the number of loci (ranging from 2 to 8) and the form of natural selection (Linear, Stabilizing with $\theta > z_{max}$, and Stabilizing with $\theta < z_{max}$). The Bayesian estimator uses genetic data filtered to resemble sampling using the QTL experimental method.

Figure 5: Significance of $\widehat{\boldsymbol{\eta}}$ estimates. The fraction of estimates that differ significantly (black) and do not differ significantly (white) from $\widehat{\boldsymbol{\eta}}$ =0 with 95% confidence across the range of time average η values.

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:

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

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

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

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

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.

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

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

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

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

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

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 relative fitness:

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

where D_{ij} is the linkage disequilibrium between the i^{th} 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

$$a_i = \frac{\beta}{\alpha} b_i \quad \text{(S6a)}$$

$$a_{i,j} = 0. \quad \text{(S6b)}$$

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),

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

Where D_{ij} is the linkage disequilibrium between loci i and j. To further simplify this expression we assume that recombination is frequent relative to the strength of selection, an assumption which allows the populations to reach a state known as quasi-linkage-equilibrium (QLE). At QLE, the D_{ij} terms are small and 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

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

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:

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

(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):

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

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

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

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

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

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 $2 \times n$ 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. One important feature of MCMC sampling of the posterior distribution is called "burn in", a phenomena describing how the beginning portion of a sequence of sampled points depends largely on the initial starting point rather than on the posterior distribution, therefore the first simulated points are often uninformative in describing the posterior. To eliminate these effects, we treated the first half of the sequences as a burn-in period (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 $\eta^*$is drawn from the jump distribution $J(\eta, \eta^*)$. 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 weighting of the variances between versus within each sequence:

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

where:

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

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

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

After enough points have been simulated to reach a ratio of $R < 1$, we use the last n points in each of the $M = 5$ sequences to generate a single histogram, consisting of 100 bins, which approximates

the posterior distribution. The estimate for η, $\hat{\eta}$, was given by the histogram bin containing the most points.

C: *Sensitivity of the Bayesian Estimator to migration from the ancestral population 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 Bayesian estimator. We found that the Bayesian estimator is fairly robust to the presence of recurrent gene flow, remaining quite accurate even when up to $Nm = 2$ migrants move among populations per generation regardless of the source (ancestral or descendent) of these migrants (Table S4, S5). 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, $Nm = 2$, (Table S5).

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

An additional assumption of our analytical model and Bayesian estimator is that the strength of natural selection is identical in the descendent populations. 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 Bayesian estimator. Specifically, we compared our Bayesian estimate of η to the average value of η actually experienced by the two populations when the selection gradient between the two populations differed by up to 20%. We saw no significant difference between the accuracy of the estimator under differing selection compared to when selection was identical (Table S6).

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

Thus far we have discussed the robustness of the Bayesian 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 Bayesian estimator we ran individual based simulations where, rather than using the true values for the initial frequencies and effect sizes, we estimated $\hat{\eta}$ using parameter estimates accompanied by a specified amount of error. Specifically, we drew initial allele frequencies and effect sizes from a Gaussian distribution centered around the true value and with a variance of either 0% or 10% of the true value. The results of these simulations are shown in (Table S7), and reveal that the Bayesian estimator was relatively insensitive to error in initial allele frequency and only moderately sensitive to error in effect size.

F: References:

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

	Candidate Gene Method				QTL Method			
# of loci	**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:

		Candidate Gene Method				QTL Method			
Selection	**# of loci**	**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:

		Candidate Gene Method				QTL Method			
Selection	# of loci	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:

			Candidate Gene Method				QTL Method			
Selection	# of loci	# migrants	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:

			Candidate Gene Method				QTL Method			
Selection	# of loci	# migrants	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:

		difference	Candidate Gene Method				QTL Method			
Selection	# of loci	in gradient	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:

				Candidate Gene Method				QTL Method			
Selection	# of loci	$Error\ in\ p_0$	$Error\ in\ b$	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 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 in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, 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 in the

simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, 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 in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, 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 in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, 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.

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 in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, 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 in the simulation η, and the R^2 values for the line of perfect fit $\hat{\eta} = \eta$. Estimates for the strength of selection were made using the data from all loci, as well as data filtered to resemble the two experimental methods, 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.

Figure Legends:

Figure S1: Ideal R^2 values across robustness tests. Ideal R^2 values for QTL method (blue) and candidate gene (red) estimates using data from 2 (circle), 4 (square), or 8 (triagnle) loci across different robustness tests. Wright Fisher (WF) simulations represent an ideal case. Explinations of different robustness tests are given in main text as well, supplimentary sections C,D, and E, as well as table legends.

References

Gelman, A. 2004. *Bayesian data analysis,* 2nd ed. Chapman & Hall/CRC, Boca Raton, Fla.

Hartl, D. L. & Clark, A. G. 2007. *Principles of population genetics,* 4th ed. Sinauer Associates, Sunderland, Mass.

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

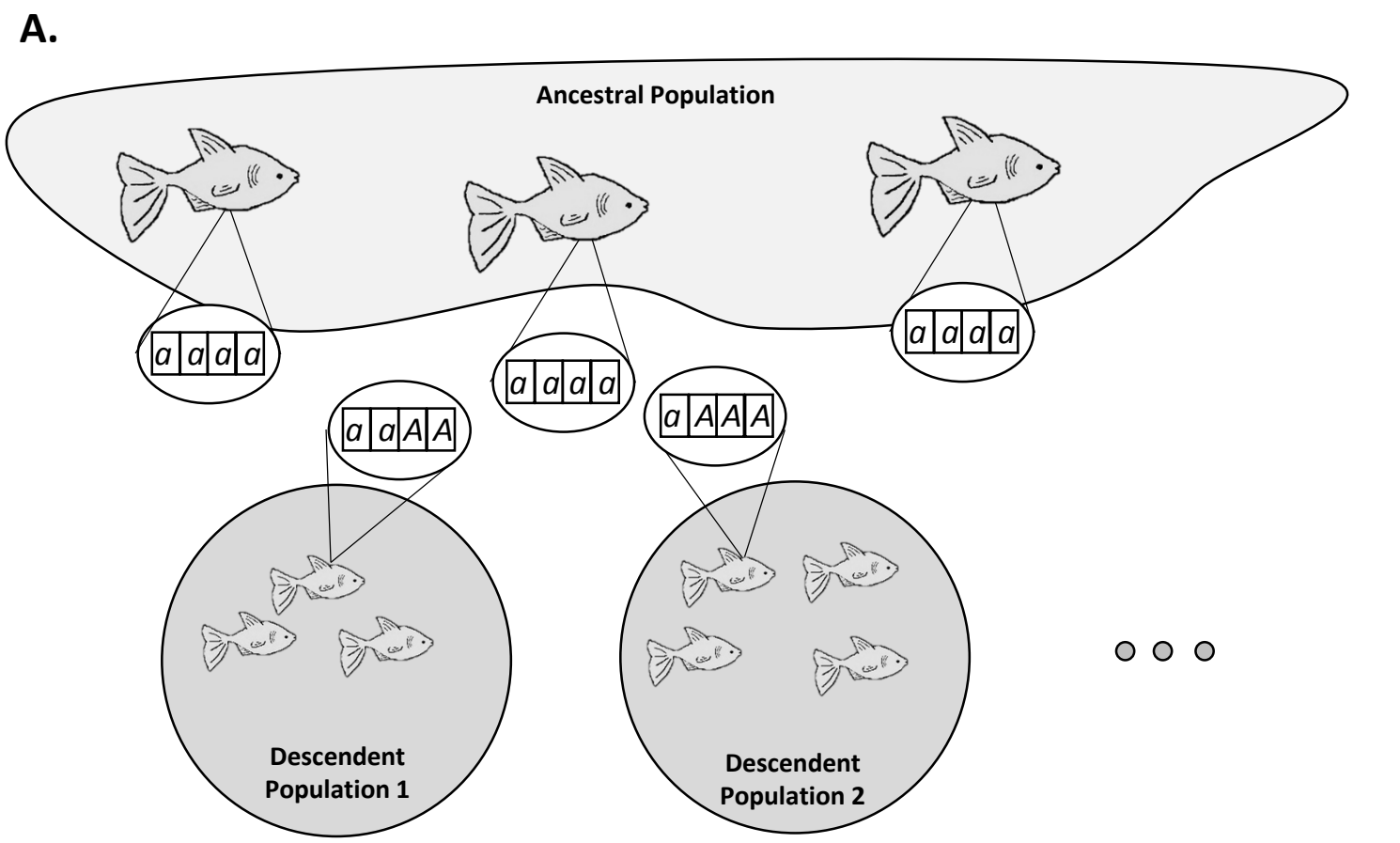
Kirkpatrick, M., Johnson, T. & Barton, N. 2002. General models of multilocus evolution. *Genetics* **161**: 1727-1750.

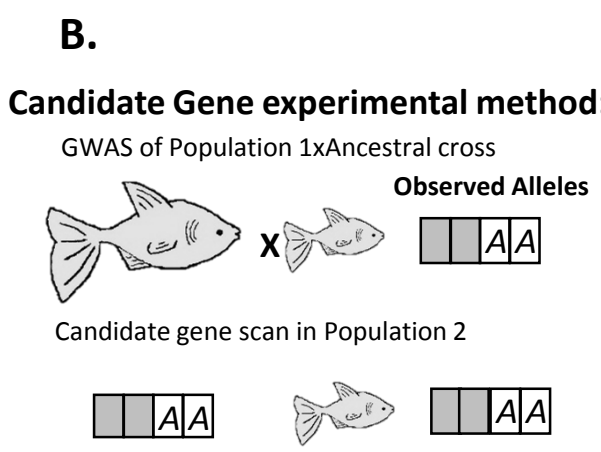
Marjoram, P., Molitor, J., Plagnol, V. & Tavare, S. 2003. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 15324-15328.

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

Nagylaki, T., Hofbauer, J. & Brunovsky, P. 1999. Convergence of multilocus systems under weak epistasis or weak selection. *Journal of Mathematical Biology* **38**: 103-133.

Figure 1:


A.
Ancestral Population
a a a a
a a a a
a a A A
a A A A
a a a a
Descendent Population 1
Descendent Population 2


B.
Candidate Gene experimental method:
GWAS of Population 1xAncestral cross
Observed Alleles
X
A A
Candidate gene scan in Population 2
A A
A A

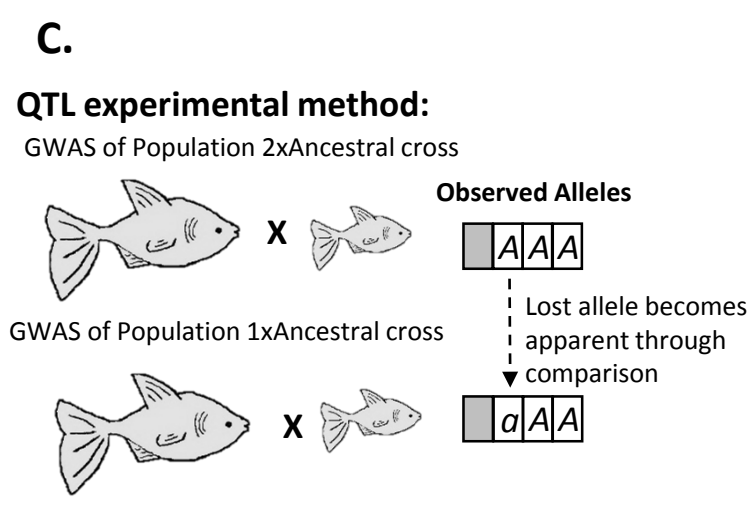

C.
QTL experimental method:
GWAS of Population 2xAncestral cross
Observed Alleles
X
A A A
Lost allele becomes apparent through comparison
GWAS of Population 1xAncestral cross
X
a A A

Figure 2:

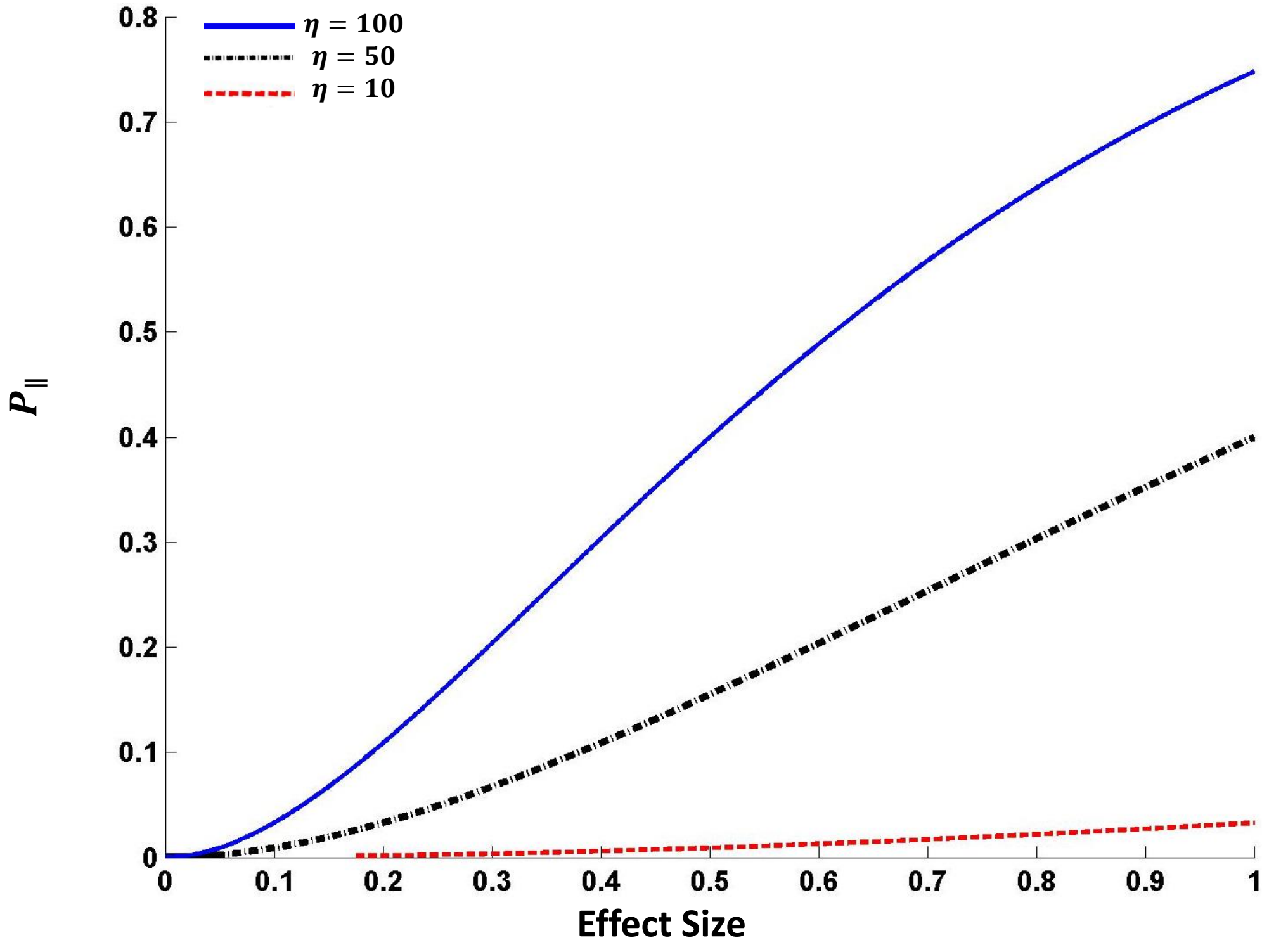

$\eta = 100$
$\eta = 50$
$\eta = 10$
$P_{\parallel}$
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Effect Size

Figure 3:

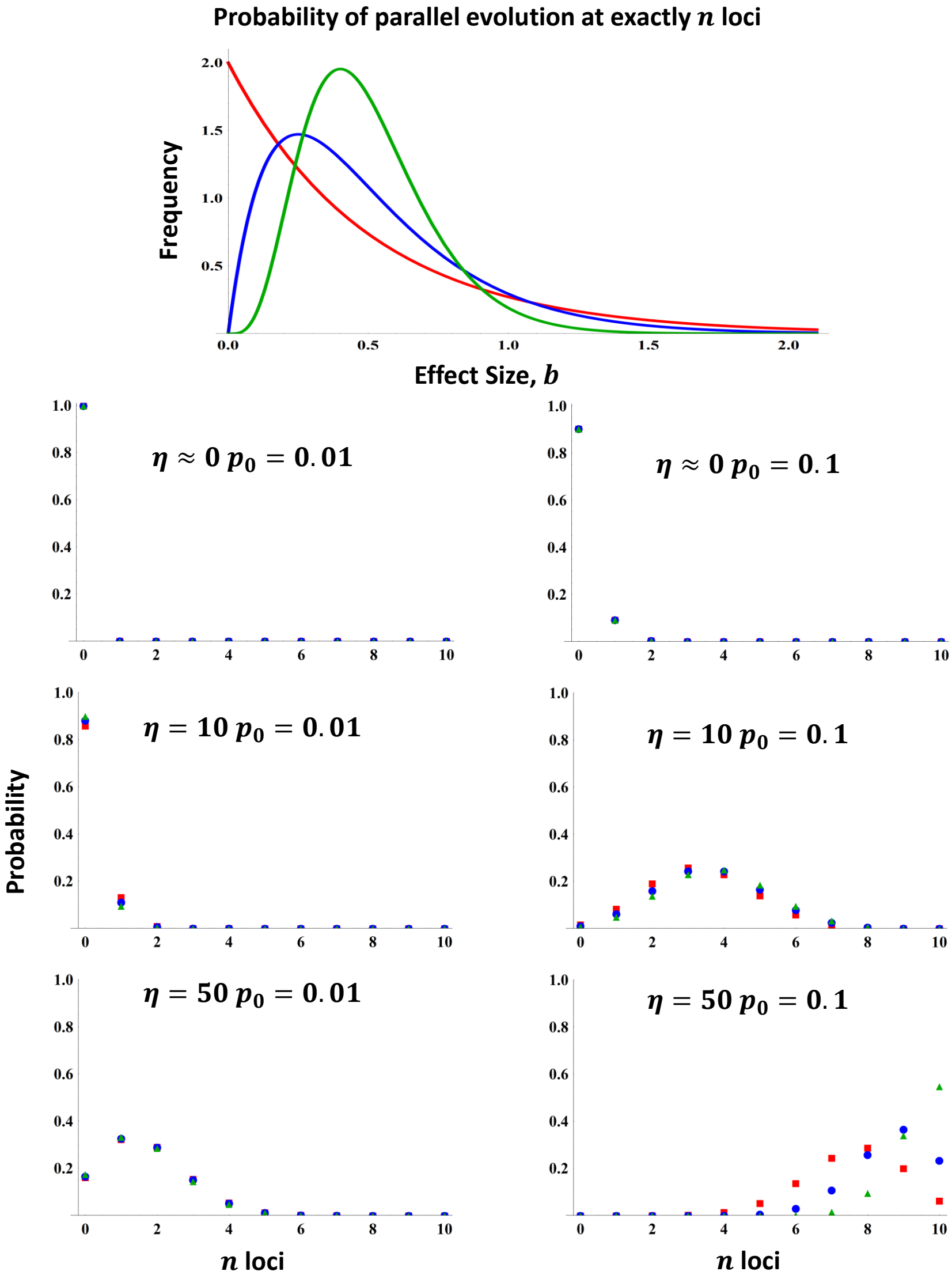

Probability of parallel evolution at exactly n loci
Frequency
Effect Size, b
$\eta \approx 0\ p_0 = 0.01$
$\eta \approx 0\ p_0 = 0.1$
$\eta = 10\ p_0 = 0.01$
$\eta = 10\ p_0 = 0.1$
$\eta = 50\ p_0 = 0.01$
$\eta = 50\ p_0 = 0.1$
Probability
n loci
n loci

Figure 4

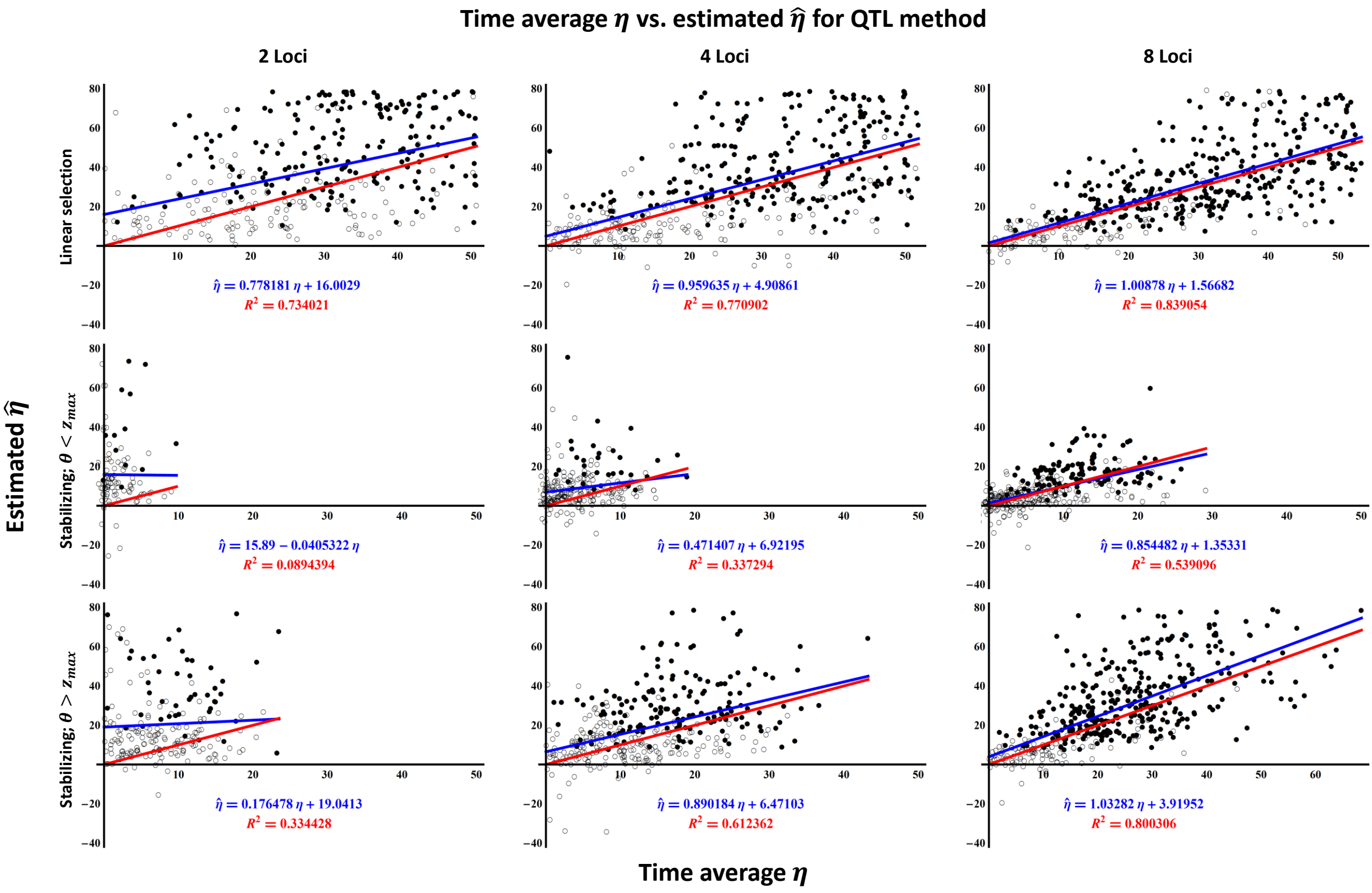

Time average η vs. estimated $\hat{\eta}$ for QTL method
2 Loci
4 Loci
8 Loci
Linear selection
Stabilizing; $\theta < z_{max}$
Stabilizing; $\theta > z_{max}$
$\hat{\eta} = 0.778181\,\eta + 16.0029$
$R^2 = 0.734021$
$\hat{\eta} = 0.959635\,\eta + 4.90861$
$R^2 = 0.770902$
$\hat{\eta} = 1.00878\,\eta + 1.56682$
$R^2 = 0.839054$
$\hat{\eta} = 15.89 - 0.0405322\,\eta$
$R^2 = 0.0894394$
$\hat{\eta} = 0.471407\,\eta + 6.92195$
$R^2 = 0.337294$
$\hat{\eta} = 0.854482\,\eta + 1.35331$
$R^2 = 0.539096$
$\hat{\eta} = 0.176478\,\eta + 19.0413$
$R^2 = 0.334428$
$\hat{\eta} = 0.890184\,\eta + 6.47103$
$R^2 = 0.612362$
$\hat{\eta} = 1.03282\,\eta + 3.91952$
$R^2 = 0.800306$
Estimated $\hat{\eta}$
Time average η

Figure 5:

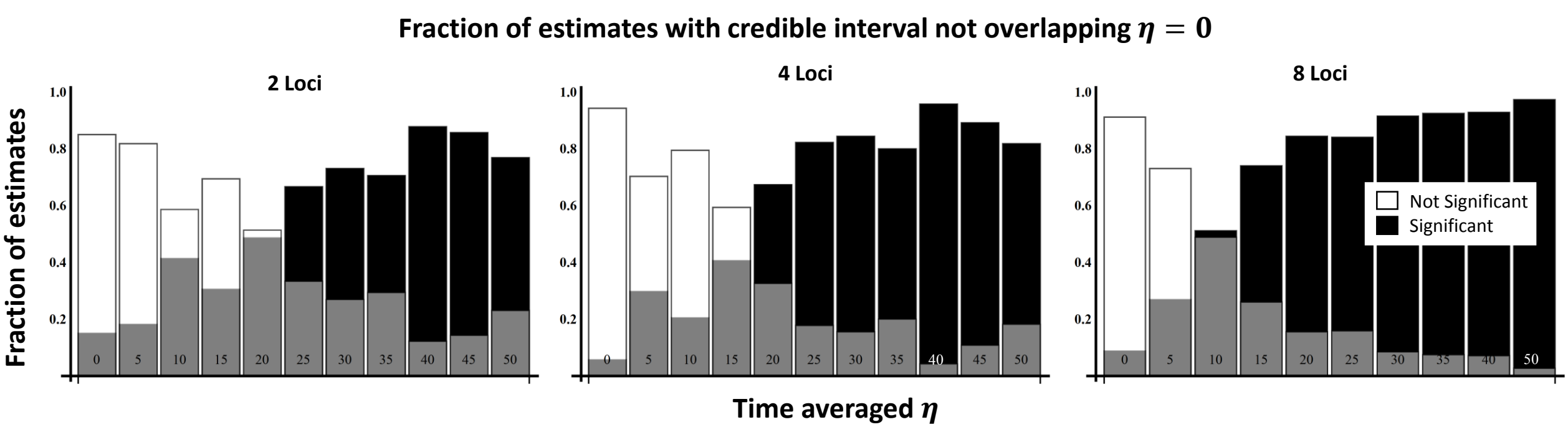
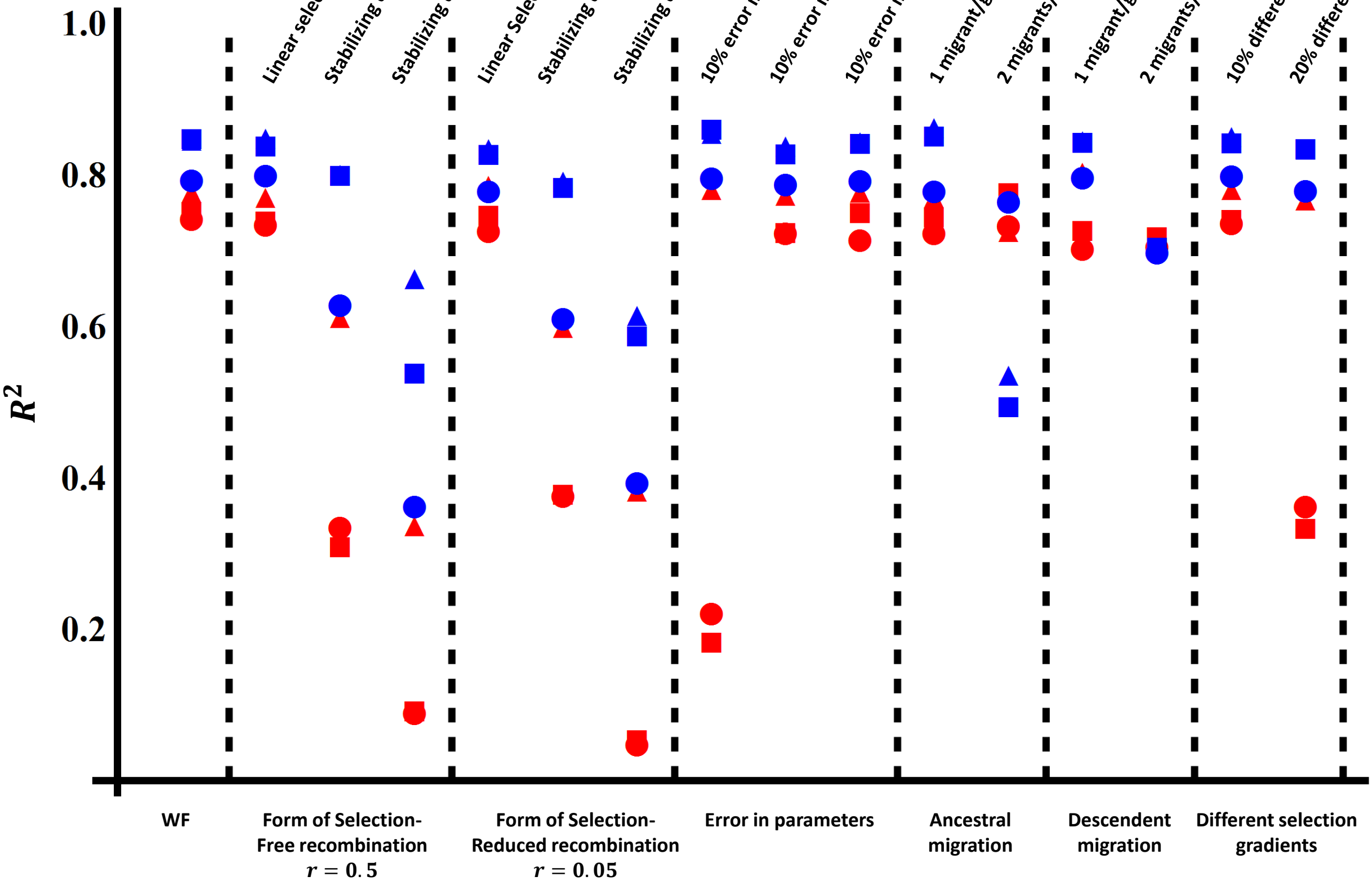

Fraction of estimates with credible interval not overlapping $\eta = 0$
2 Loci
4 Loci
8 Loci
Fraction of estimates
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30
35
40
45
50
Not Significant
Significant
Time averaged η

Figure S1:


1.0
0.8
0.6
0.4
0.2
R^2
Linear selection
Stabilizing $\theta > z_{max}$
Stabilizing $\theta < z_{max}$
Linear Selection
Stabilizing $\theta > z_{max}$
Stabilizing $\theta < z_{max}$
10% error in p_0 and b
10% error in p_0 only
10% error in b only
1 migrant/generation
2 migrants/ generation
1 migrant/generation
2 migrants/generation
10% different
20% different
WF
Form of Selection- Free recombination $r = 0.5$
Form of Selection- Reduced recombination $r = 0.05$
Error in parameters
Ancestral migration
Descendent migration
Different selection gradients