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