We thank the editor for their consideration of our manuscript for GBE and appreciate the helpful comments. We have tried to address the comments as best as we could in the context of our manuscript. Below in blue, we outline specifically the changes made to the manuscript.

Comments to the Author:

I have been asked to consider this manuscripts for GBE, following a transfer from MBE. As such, I have evaluated the comments from the AE at MBE (R. Rogers), as well as those from two reviewers. Both reviewers were largely positive, primarily requesting clarification. I find that this revised manuscript would be appropriate for GBE, and I am recommending a Minor Revision in order to give the authors an opportunity to address the following points:

It seems problematic (e.g., in Table 1) to discuss these in terms of false positive and false negative rates, though the authors do add: "assuming the gene tree topology is the correct topology". In general, one would think of false positives and false negatives in scenarios when a ground-truth is known (e.g., in simulation-based testing). As another example, in Figure 6, it is not inherently clear to me that the description of "tree misspecification leads to erroneous results in tests for positive selection" is necessarily true. While it is clear that it leads to differing results, as far as I can see here it is not known with certainty what fraction (if any) of these genes are indeed evolving under rapid and recurrent positive selection (again, as this is empirical data). As such, it would appear possible that they are all erroneous results, and thus 'tree misspecification' may not necessarily be the driving reason.

In sum, the authors may want to reconsider the wording around this issue throughout the manuscript, or point out if I am mis-interpreting something.

We agree that it is not possible to determine with absolute certainty that tree misspecification is the root cause of the disagreement, especially since either tree could be inferred erroneously, so we have updated the language throughout the manuscript to avoid terms such as false positive, false negative, correct/incorrect tree.

Specifically:

Lines XX:

Replaced

“high error rates”

With

“substantial deviation from the results with gene trees”

Lines XX: Removed the text “when the incorrect species tree is used”

Lines XX:

Replaced

“We found evidence that tree misspecification likely induces both false positive (type I) and false negative (type II) errors. For example, many genes were inferred as having experienced positive directional selection when using a single species tree, but not when using local gene trees and vice versa (Figure 6). Assuming the locally inferred gene tree is more accurate than the single tree inferred from concatenation of all gene sequences, this resulted in varying rates and types of error (Table 1).”

With

“We found that many genes were inferred as having experienced positive directional selection when using a single species tree, but not when using gene trees and vice versa (Figure 6). The extent to which the single species tree differed from the gene trees for the different types of selection test is documented in Table 1.”

Lines: XX and XX: Removed the parentheticals “(likely false negatives)” and “(likely false positives)”

Lines XX:

Replaced

“Here, we use empirical data in mice to show that these errors result in false positive (detected signal for selection only when using the gene tree) and false negative results (detected signal for selection only when using the species tree).”

With

“Here, we use empirical data in mice to document the extent that such incorrect placement of substitutions can lead to either a detected signal for selection when only using the gene tree or a detected signal for selection when only using the species tree.”

Lines XX:

Replaced

“inflated false positive and false negative rates”

With

“incorrect inferences”

Lines XX:

Replaced

“On the other hand, models that only allow rates to vary among sites, such as PAML’s M1a vs. M2a test, showed an increase in the number of putative false positives inferred when using the wrong tree.”

With

“On the other hand, models that only allow rates to vary among sites, such as PAML’s M1a vs. M2a test, showed an increase in the number of cases where there was positive selection detected with the gene trees but not with the species tree.”

Table 1 caption:

Replaced

“Rates and types of error when using a single species tree for gene based selection tests (Assuming the gene tree topology is the correct topology).”

With

“Incidence where a single species tree does not match the gene tree expectation in three different tests for positive selection, either not detecting positive selection inferred when using the gene tree (Undetected selection) or by detecting positive selection that is not inferred when using the gene tree (Newly detected selection).”

Table 1 column headings:

Replaced

“False positive rate” “False negative rate”

With

“Undetected selection” “Newly detected selection”

Figure 6 caption: Removed the sentence “Tree misspecification leads to erroneous results in tests for positive selection.”

More generally speaking, I felt that there was a lost opportunity here to guide the readers on the differences and interpretations of coalescent trees/histories relative to phylogenetic trees/histories, as that is in many ways at the heart of some of these points, particularly when comparing between very closely related 'species' (or, possibly, diverged populations, depending on how one interprets species concepts). For example, I believe from a population genetics view when discussing 'gene tree' discordance surrounding a selective sweep for example, the authors are really invoking the expected and localized star-shaped coalescent history potentially generated by such a beneficial fixation. Thus, I think that the manuscript would be of great value to readers if it elaborated upon this point, as the phylogenetics literature seemingly often confuses itself on this topic (e.g., it surprises no one that coalescent trees differ along a recombining genome, but when comparing closely related species / distantly related populations it suddenly surprises many phylogeneticists that 'species trees' differ along a genome?). Said another way, there are simply these processes operating in natural populations (genetic drift, gene flow, natural selection, etc), and thus phylogenetic patterns are the accumulation of these population genetic processes over long time scales, and that transition between 'coalescent histories' and 'phylogenetic histories' can be blurred over particular time periods. On this point, the authors may find the discussions in the two following papers to be helpful:

Louca & Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature.

Johri et al. 2023. Recommendations for improving statistical inference in population genomics. PLoS Biology.

We agree with the editor that it is important to link the effect (phylogenetic discordance) with the cause (population level processes). As the editor points out, many people that do phylogenetics at the species level tend to forget that discordance can still be present at deeper timescales and fail to account for or even quantify it in their analyses. One of our main goals with this paper was to hammer this point home so there could be something concrete that shows the genomic context and downstream effects of discordance. As such, we didn’t focus on the underlying population processes besides recombination as much as is perhaps warranted. Since the paper is already quite lengthy, we have added a bit of text to the Conclusions paragraph in the Discussion to mention the population genetics aspect more explicitly, and to point the readers to other helpful papers:

Lines XX: “. More generally, our results highlight how processes operating among natural populations are still relevant even over longer phylogenetic timescales when making inferences between species. Phylogenetic patterns of discordance are the result of the accumulation of coalescence, genetic drift, natural selection, and other forces that act on populations and neither the discordance or these patterns are captured or modeled in a single species tree. This is akin to the recent conclusion that phylogenies alone are insufficient for estimating diversification rates (Louca and Pennell 2020). Likewise, if one’s goal is to study molecular evolution between species, the underlying population genetic forces must be accounted for, or at least mitigated, most easily by using individual gene trees rather than a single species tree. Discordance and uncertainty of underlying population processes are discussed further in Johri, et al. (2022), Mirarab, et al. (2021), and Steenwyk, et al. (2023).”

We hope this is sufficient in the context of this manuscript. Johri, et al. 2022, Louca and Pennell 2020, Mirarab, et al. 2021 and Steenwyk, et al. 2023 have been added to the Reference list.

minor: "Smith and Haigh 1974" is incorrect, it should read "Maynard Smith and Haigh 1974"

Corrected, thank you.