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. While we feel it is reasonable that the locally inferred gene tree is likely to provide a more accurate inference of the history of that genomic region, we also acknowledge that either tree could be inferred with error and that the underlining truth is not known for empirical data. We have updated the language throughout the manuscript to avoid terms such as false positive, false negative, correct/incorrect tree.

Specifically:

Line 36:
Replaced "high error rates" with "substantial deviation from the results with gene trees".

Line 52:
Removed the phrase "when the incorrect species tree is used."

Lines 255:

Replaced Results sub-section title "Consequences of tree misspecification on analyses of molecular evolution" with "Consequences of tree specification on analyses of molecular evolution"

Lines 271-274:
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: 276 and 279: Removed the parentheticals "(likely false negatives)" and "(likely false positives)"

Lines 411-413:
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 using local gene trees versus assuming an overall species tree may shape inferences of positive directional selection on protein-coding sequences."

Line 421:
Replaced "inflated false positive and false negative rates" with "can often affect inferences".

Line 425:
Removed the parenthetical "(i.e., the correct tree, assuming no errors in gene tree reconstruction)".

Lines 427-430:
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."

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.

Thanks for these suggestions. We agree that it is important to link the effect (phylogenetic discordance) with the cause (population-level processes). As the editor points out, some phylogeneticists do seem to assume that incomplete lineage sorting and discordance in general mostly affect recent histories and fail to account for or even quantify it in their analyses. It is important to underscore that discordance can still be present at deeper timescales. 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. Although we do not feel the current study is sufficient to suggest a general framework for a baseline model, we have added a final synthesis paragraph to the Conclusions that hopefully does a better job of underscoring these points and to point the readers to other helpful papers:

Lines 470-486: Our results help illuminate the complexities of phylogenomic datasets and the need to accommodate phylogenetic discordance in genome-wide analyses. Genomic data now dominate the study of both population genetics and phylogenetics, and these once disparate fields are increasingly unified. Species tree phylogenies are an emergent pattern of the genome-wide accumulation of stochastic and directional population-level processes that cannot be fully captured or modeled by a single history (Steenwyk, et al. 2023). Importantly, phylogenetic discordance is not limited to closely related populations or species and is expected to leave persistent signals over deep evolutionary timescales (Oliver 2013). In turn, the use of tree-based frameworks for studying evolution (at any timescale) must incorporate the population-level processes that shape phylogenetic discordance. There appear to be relatively few tree-based applications where the use of a single evolutionary history is appropriate. Indeed, failure to account for phylogenetic discordance can lead to spurious inferences of molecular evolution (Figure 6; Mendes, et al. (2016)), trait evolution (Avise and Robinson 2008; Hahn and Nakhleh 2016), and even species diversification (Louca and Pennell 2020). Similar to the need for robust baseline models in population genomic inference (Johri, et al. 2022), understanding the causes and landscape of phylogenetic discordance constitutes a critical first step in phylogenomic analysis (Mirarab, et al. 2021; Steenwyk, et al. 2023).

We hope this is sufficient in the context of this manuscript. Oliver 2013, 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.