Recommender's comments:

This is an excellent paper, and the revision is much improved. The author provided substantive responses to all comments in the previous reviews. I think some of this is important enough to include in the text of the article rather than being relegated to the author's response. For this reason, I'm going to ask for a third draft. Details follow.

Thanks for the positive assessment and for offering to further improve the manuscript.

100: The text here fails to mention that t1 + 2Ne is not really the expected coalescence time for the A|B split. I pointed this out in my previous review, and the author acknowledged it in his response. The text should make clear that the present formulation is an approximation that should work well when t2 − t1 >> 2Ne but may introduce bias otherwise.

I agree. I have not modified the Methods section of the manuscript so as not to alter the flow, but have added a paragraph in the discussion (lines 433-437) discussing this approximation and the underlying assumption.

I take the opportunity to correct an error in my previous reply to your comment (previous round of review). The formula you provided for the conditional average coalescence time does not imply estimating an extra parameter, like I said, since $z=(t_2-t_1)/2N_e$ can be rewritten as $z=2(\tau_2-\tau_1)/\theta$. So this suggestion is totally relevant and should be implemented/investigated in a forthcoming version of the program.

193: Section 2.6 should mention the point, raised by reviewer 2, that a genomic segment long enough to include mutations will probably also include recombinations. This issue is addressed at length later, but I would anticipate it briefly here.

The revised version specifies the condition that non-recombining segments include relevant phylogenetic information, and have been correctly identified previously (line 195-196).

210–211: It's worth mentioning that Aphid performs well in these simulations even though the parameters chosen make Aphid's approximation (see comment above on line 100) rather poor. For these parameter values, the expected coalescence time for the A|B split is t1 + 0.69 × 2Ne, which is substantially smaller than the value (t1 + 2Ne) assumed by Aphid. I would have expected this to generate substantial bias. Yet Aphid does a good job of recovering the simulation parameters. This should be advertized a bit.

This is now highlighted at the end of the newly-added paragraph on the corresponding approximation (line 437-439).

264–267: The difficulty in anotating ILS trees might arise from the problem discussed in my comment on line 100. If so, these problems should be greatest when z = 2(τ2 − τ1)/θ is small. (For details, see my comment on lines 97–99 in the previous review.)

You might well be right. I note however that we are here considering conflicting topologies only, whereas the problem you're mentioning concerns in the first place a confusion between the no-event and ((AB)C)-ILS scenarios. Yet the approximation could affect the reliability or parameter estimation, and in turn gene tree annotation, as you were suggesting. We need to explore this deeper.

374–375: I've never seen scientific notation written this way. I would write 1.15×10^{-2} rather than 1.15 10−2. While I'm on typography, in this paper dashes look like "-" rather than like "—". The latter is conventional. To get longer dashes in LATEX, use "---". These are just suggestions.

Corrected as suggested.

459–460: Legofit has also been used to study chimpanzees: Colin M. Brand et al. "Estimating bonobo (Pan paniscus) and chimpanzee (Pan troglodytes) evolutionary history from nucleotide site patterns". Proceedings of the National Academy of Sciences, USA 119.17 (2022), e2200858119. doi: 10.1073/pnas.2200858119.

Now cited.