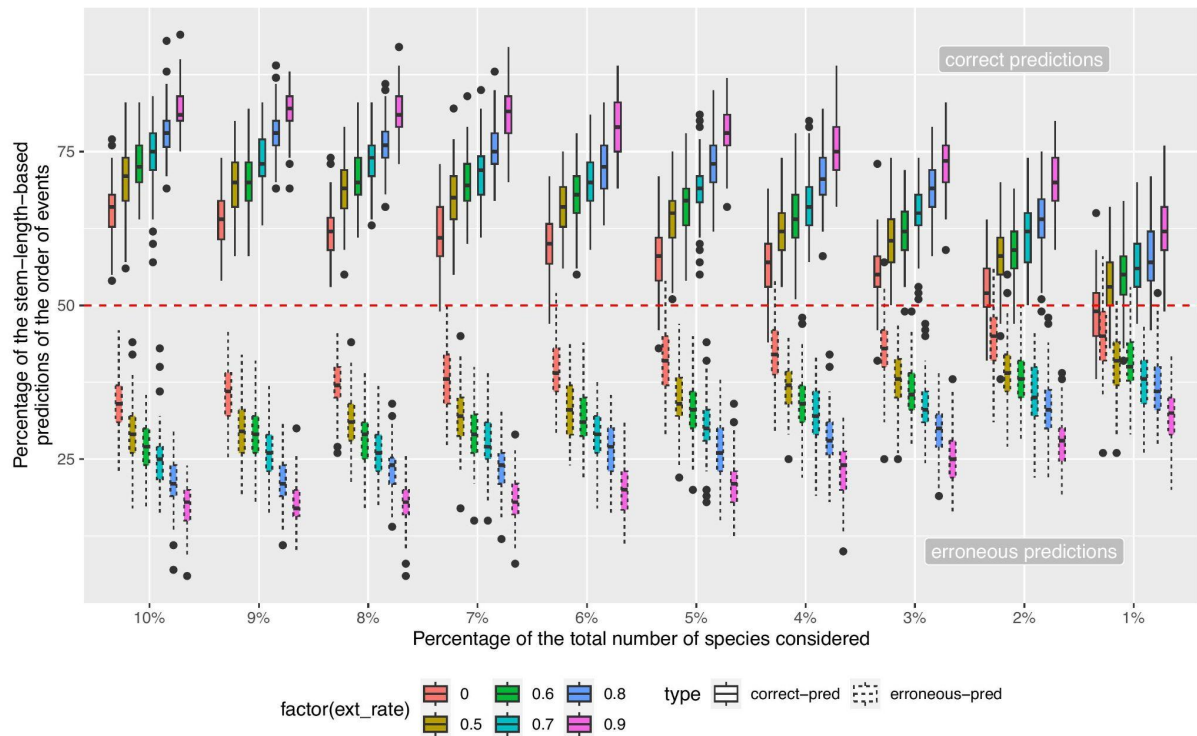
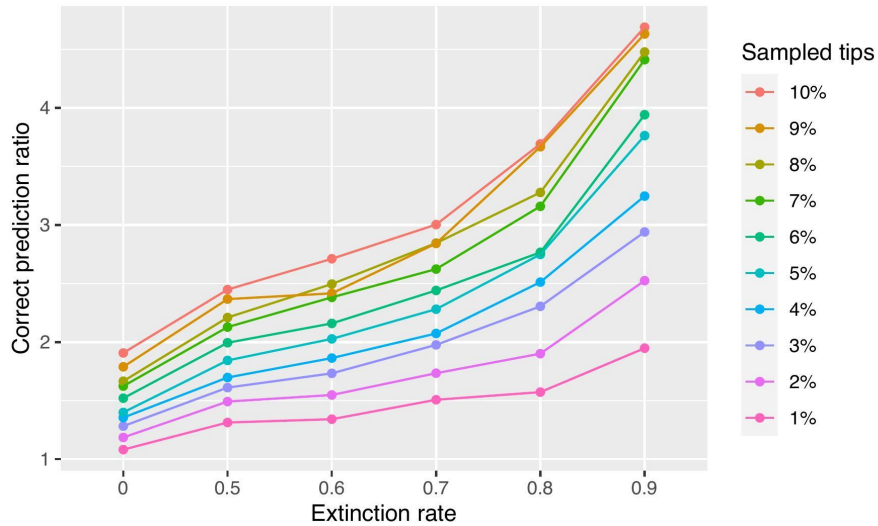


Supplementary material for: On the impact of incomplete taxon sampling on the relative timing of gene transfer events

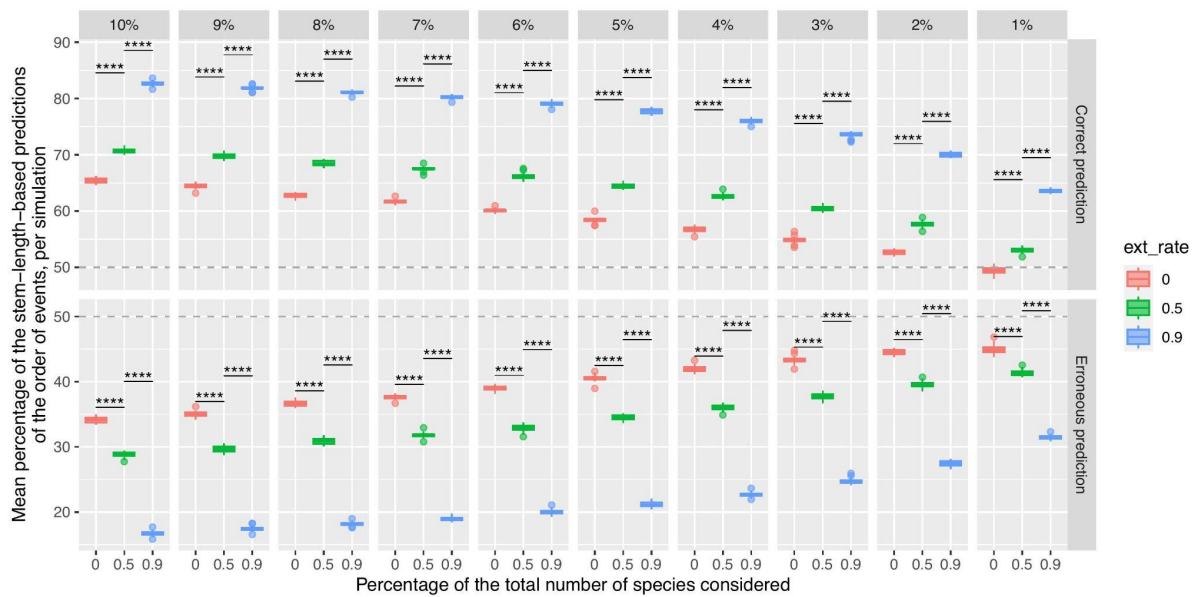
Moisès Bernabeu, Saioa Manzano-Morales, Toni Gabaldón



S1 Fig. Percentage of correct and shifted conclusions depending on the percentage of considered species. Dashed boxplots are the erroneous predictions and colours show the extinction rate used for the tree simulations.



S2 Fig. Correct prediction ratio against the extinction rate. The correct prediction ratio was calculated by dividing the fraction of correct inferences divided by the fraction of incorrect inferences. Colours show the percentage of sampled tips for inferring the shifts.



S3 Fig. Boxplots for the correct and erroneous predictions. We performed 24 sets of simulations for each percentage of considered species and extinction rate and calculated the mean of the correct and erroneous predictions to assess the intra-simulation variance. We assessed the difference among the extinction rate groups (0 vs. 0.5 and 0.5 vs. 0.9) using a Wilcoxon test.

S4 Text. Discussion about the extinction rate parameters used

To assess the impact of the extinction rate in the ghost lineages, we repeated the Tricou et al. analyses using their code (<https://doi.org/10.5281/zenodo.6901799>) with different extinction rates (0, 0.5, 0.6, 0.7, 0.8 and 0.9). We reproduced the Fig. 6 in [5] (S1 Fig) adding the different extinction rates (in colours). Considering just 1% of the total number of species from the initial tree, the percentage of correct predictions is 55% using an extinction rate of 0.9. Tricou et al. [5] describe that they used an extinction rate of 0.5, however, we re-ran the analysis for this value, and we obtained higher proportions for the correct predictions of the sorting. Investigating the GitHub repository code, we noticed that the uploaded version has an extinction rate of 0.

These results confirm the idea that lower extinction rates provide longer internal branches. When using a rate of 0.5, the proportion of shifts is on average 40.6%. This average proportion of shifts differs from the one obtained by Tricou et al. [5] (> 45%). Moreover, using an extinction rate of 0.9 (as in all the other examples done in the paper), the proportion of shifts has a mean of 62.46%.

Thus, lower extinction rates make the conclusion shift more likely in this simulation framework as we argue in the formal comment. S2 Fig. shows the correct-incorrect prediction ratio per extinction rate, the higher the extinction rate, the more difference we observe favouring the correct predictions, being always higher than 1, this means that the correct predictions proportion is always higher than the incorrect predictions proportion. These results (S1 Fig., S2 Fig. and S3 Fig.), altogether, show that the rate used in [5] was not 0.5, but 0, as their code repository show.