



**HAL**  
open science

## Response to “On the impact of incomplete taxon sampling on the relative timing of gene transfer events”

Théo Tricou, Eric Tannier, Damien de Vienne

### ► To cite this version:

Théo Tricou, Eric Tannier, Damien de Vienne. Response to “On the impact of incomplete taxon sampling on the relative timing of gene transfer events”. *PLoS Biology*, 2024, 22 (3), pp.e3002557. 10.1371/journal.pbio.3002557. hal-04515190

**HAL Id: hal-04515190**

**<https://hal.science/hal-04515190v1>**

Submitted on 3 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



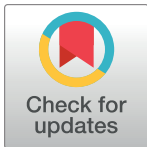
Distributed under a Creative Commons Attribution 4.0 International License

## FORMAL COMMENT

## Response to “On the impact of incomplete taxon sampling on the relative timing of gene transfer events”

Théo Tricou<sup>1</sup>, Eric Tannier<sup>2,3</sup>, Damien M. de Vienne<sup>2\*</sup>

**1** Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, F-63122 Saint-Genès-Champanelle, France, **2** Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR5558, F-69622 Villeurbanne, France, **3** INRIA Grenoble Rhône-Alpes, F-38334 Montbonnot, France

\* [damien.de-vienne@univ-lyon1.fr](mailto:damien.de-vienne@univ-lyon1.fr)

In a previous article [1], we used simulations to show how acknowledging the existence of extinct and unsampled lineages, or “ghosts,” could substantially influence conclusions regarding gene flow derived from gene tree branch lengths analyses. We mentioned that this could in particular challenge the claim by Pittis and Gabaldon [2,3] that genes of alpha-proteobacterial origin are more recently transferred than those originating from actinobacteria.

Bernabeu and colleagues [4] propose a critique of our results based on 2 additional interesting analyses.

The first one, summarized in their Fig 1C, is a description of the conditions under which the conclusions of Pittis and Gabaldon [2,3] can actually be reversed (“conditions to shift” in their terminology), thus acknowledging this possibility. Bernabeu and colleagues conclude that these conditions are “rather restrictive” or “highly constrained,” suggesting that they are unlikely. This calls for 3 comments. One is that the list of conditions they give can be slightly relaxed. Indeed, it is not necessary that the earlier transfer (from alpha-proteobacteria) originates from a ghost lineage (as in Fig 1C in [4]). The shift is also obtained if this transfer comes directly from a non-ghost alpha-proteobacteria ( $\alpha 1$  in their Fig 1) at an earlier time than the transfer from a ghost actinobacteria. This decreases the restrictivity of the conditions to shift. Another comment is that the conditions are not necessarily independent. Constraining the events to happen in a limited timeframe (the FECA-to-LECA period) is presented as a condition that decreases the chance that it occurs at all [4], but according to our analyses (Fig 6B in [1]) it may at the same time increase the probability that it leads to a shift. A last comment is that “rather restrictive” and “highly constrained” are vague and subjective qualifications. An opposite intuition is arguable: if we consider that most of the microbial diversity is unknown [5], and if we observe that numerous complete clades (such as the CPR, which represents a large part of the ToL presented as Fig 1A in Ref [4]) are probably missing from such an analysis, the conditions for shift seem sufficiently plausible to question the robustness of the stem-length based claims. Further work is needed to transform these intuitions into a probability and assess the likelihood that the conclusions of Pittis and Gabaldon [2,3] can be reversed (or not) when acknowledging the overwhelming presence of ghosts. Our simulations [1] were an attempt to quantify this effect. Surely they are not sufficient to give a definitive answer, because they were indeed not informed by all the knowledge we can have about the tree of life. At least, and in the light of the work of Bernabeu and colleagues [4], the question remains, in our view, largely open.

**OPEN ACCESS**

**Citation:** Tricou T, Tannier E, de Vienne DM (2024) Response to “On the impact of incomplete taxon sampling on the relative timing of gene transfer events”. *PLoS Biol* 22(3): e3002557. <https://doi.org/10.1371/journal.pbio.3002557>

**Received:** December 15, 2023

**Accepted:** February 21, 2024

**Published:** March 19, 2024

**Copyright:** © 2024 Tricou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the French National Research Agency Grants ANR-18-CE02-0007-01 to DIMDV and ANR-19-CE45-0010 to ET. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

The second analysis of Bernabeu and colleagues [4] explores the sensitivity of the probability to reverse the conclusions of Pittis and Gabaldon [2,3] to the extinction rate. Using new simulations, they show that when 1% of the total number of species are sampled, increasing this rate to 0.9 reduces the proportion of incorrect predictions to 32%, compared with approximately 40% and 50% for rates of 0.5 and 0, respectively (Fig S3 in ref [4]). This sensitivity analysis is interesting, as it usefully complements our simulations (for which we recognize the noted error in the report of the simulation parameters in the text). It confirms that many combinations of parameters can influence the probability of shifts in one direction or the other. This is convergent with our claims that high uncertainties are associated with results based on the ghost-unaware stem-length method, even the smallest probability of shift equal to 0.32.

Finally, Bernabeu and colleagues [4] rightly recall that the possible effect of ghosts has been recognised before. Indeed, Hahn and colleagues [6] call for a “cautious” interpretation of the results; Susko and colleagues [7] recognise a possible “caveat”; Pittis and Gabaldon [2] test the possible effect of ghosts (in Section 4 of their supplementary material) by uniformly sampling alpha-proteobacteria and find no effect on the results. We hope that our work [1], along with the response by Bernabeu and colleagues [6], and a growing body of other studies ([8–10], to only cite a few), will help ghosts to gain a place in scientific articles that better reflects their thought importance in nature; that ghosts will not be confined to “supplementary data,” “additional remarks,” and quick control experiments anymore, but will find their place in main texts, figures, and hypotheses of future research.

## References

1. Tricou T, Tannier E, de Vienne DM. Ghost lineages can invalidate or even reverse findings regarding gene flow. *PLoS Biol.* 2022; 20:1–16. <https://doi.org/10.1371/journal.pbio.3001776> PMID: 36103518
2. Pittis AA, Gabaldón T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature.* 2016; 531:101–104. <https://doi.org/10.1038/nature16941> PMID: 26840490
3. Vosseberg J, van Hooff JJE, Marcet-Houben M, van Vlimmeren A, van Wijk LM, Gabaldón T, et al. Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat Ecol Evol.* 2020:1–9. <https://doi.org/10.1038/s41559-020-01320-z> PMID: 33106602
4. Bernabeu M, Manzano-Morales S, Gabaldón T. On the impact of incomplete taxon sampling on the relative timing of gene transfer events. *PLoS Biol.* 2024; 22: e3002460. <https://doi.org/10.1371/journal.pbio.3002460>
5. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci.* 2016; 113:5970–5975. <https://doi.org/10.1073/pnas.1521291113> PMID: 27140646
6. Hahn MW, Hibbins MS. A Three-Sample Test for Introgression. *Mol Biol Evol.* 2019; 36:2878–2882. <https://doi.org/10.1093/molbev/msz178> PMID: 31373630
7. Susko E, Steel M, Roger AJ. Conditions under which distributions of edge length ratios on phylogenetic trees can be used to order evolutionary events. *bioRxiv.* 2021; 2021.01.16.426961. <https://doi.org/10.1016/j.jtbi.2021.110788> PMID: 34097914
8. Ottenburghs J. Ghost Introgression: Spooky Gene Flow in the Distant Past. *Bioessays.* 2020; 42:2000012. <https://doi.org/10.1002/bies.202000012> PMID: 32227363
9. Tricou T, Tannier E, de Vienne DM. Ghost lineages highly influence the interpretation of introgression tests. *Syst Biol.* 2022; 71:1147–1158. <https://doi.org/10.1093/sysbio/syab011> PMID: 35169846
10. Pang X-X, Zhang D-Y. Impact of Ghost Introgression on Coalescent-Based Species Tree Inference and Estimation of Divergence Time. *Syst Biol.* 2023; 72:35–49. <https://doi.org/10.1093/sysbio/syab047> PMID: 35799362