



## Dating with constraints

**Cécile Ané** based on reviews by David Duchêne and 1 anonymous reviewer

A recommendation of:

### Relative time constraints improve molecular dating

Gergely J Szollosi, Sebastian Hoehna, Tom A Williams, Dominik Schrempf, Vincent Daubin, Bastien Boussau (2021), *bioRxiv*, 2020.10.17.343889, ver. 8 recommended and peer-reviewed by Peer Community in Evolutionary Biology [10.1101/2020.10.17.343889](https://doi.org/10.1101/2020.10.17.343889)

## Open Access

Submitted: 21 October 2020, Recommended: 04 May 2021

### Cite this recommendation as:

Cécile Ané (2021) Dating with constraints. *Peer Community in Evolutionary Biology*, 100127. [10.24072/pci.evolbiol.100127](https://doi.org/10.24072/pci.evolbiol.100127)

### Recommendation

Published: 17 May 2021

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

Estimating the absolute age of diversification events is challenging, because molecular sequences provide timing information in units of substitutions, not years. Additionally, the rate of molecular evolution (in substitutions per year) can vary widely across lineages. Accurate dating of speciation events traditionally relies on non-molecular data. For very fast-evolving organisms such as SARS-CoV-2, for which samples are obtained over a time span, the collection times provide this external information from which we can learn the rate of molecular evolution and date past events (Boni et al. 2020). In groups for which the fossil record is abundant, state-of-the-art dating methods use fossil information to complement molecular data, either in the form of a prior distribution on node ages (Nguyen & Ho 2020), or as data modelled with a fossilization process (Heath et al. 2014).

Dating is a challenge in groups that lack fossils or other geological evidence, such as very old lineages and microbial lineages. In these groups, horizontal gene transfer (HGT) events have been identified as informative about relative dates: the ancestor of the gene's donor must be older than the descendants of the gene's recipient. Previous work using HGTs to date phylogenies have used methodologies that are ad-hoc (Davín et al 2018) or employ a small number of HGTs only (Magnabosco et al. 2018, Wolfe & Fournier 2018).

Szöllősi et al. (2021) present and validate a Bayesian approach to estimate the age of diversification events based on relative information on these ages, such as implied by HGTs. This approach is flexible because it is modular: constraints on relative node ages can be combined with absolute age information from fossil data, and with any substitution model of molecular evolution, including complex state-of-art models. To ease the computational burden, the authors also introduce a two-step approach, in which the complexity of estimating branch lengths in substitutions per site is decoupled from the complexity of timing the tree with branch lengths in years, accounting for uncertainty in the first step. Currently, one limitation is that the tree topology needs to be known, and another limitation is that constraints need to be certain. Users of this

method should be mindful of the latter when hundreds of constraints are used, as done by Szöllősi et al. (2021) to date the trees of Cyanobacteria and Archaea.

Szöllősi et al. (2021)'s method is implemented in RevBayes, a highly modular platform for phylogenetic inference, rapidly growing in popularity (Höhna et al. 2016). The RevBayes [tutorial page](#) features a step-by-step tutorial "Dating with Relative Constraints", which makes the method highly approachable.

#### References:

Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, Rambaut A, Robertson DL (2020) Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology*, 5, 1408–1417. <https://doi.org/10.1038/s41564-020-0771-4>

Davín AA, Tannier E, Williams TA, Boussau B, Daubin V, Szöllősi GJ (2018) Gene transfers can date the tree of life. *Nature Ecology & Evolution*, 2, 904–909. <https://doi.org/10.1038/s41559-018-0525-3>

Heath TA, Huelsenbeck JP, Stadler T (2014) The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111, E2957–E2966. <https://doi.org/10.1073/pnas.1319091111>

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F (2016) RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Systematic Biology*, 65, 726–736. <https://doi.org/10.1093/sysbio/syw021>

Magnabosco C, Moore KR, Wolfe JM, Fournier GP (2018) Dating phototrophic microbial lineages with reticulate gene histories. *Geobiology*, 16, 179–189. <https://doi.org/10.1111/gbi.12273>

Nguyen JMT, Ho SYW (2020) Calibrations from the Fossil Record. In: *The Molecular Evolutionary Clock: Theory and Practice* (ed Ho SYW), pp. 117–133. Springer International Publishing, Cham. [https://doi.org/10.1007/978-3-030-60181-2\\_8](https://doi.org/10.1007/978-3-030-60181-2_8)

Szollósi, G.J., Höhna, S., Williams, T.A., Schrempf, D., Daubin, V., Boussau, B. (2021) Relative time constraints improve molecular dating. *bioRxiv*, 2020.10.17.343889, ver. 8 recommended and peer-reviewed by Peer Community in Evolutionary Biology. <https://doi.org/10.1101/2020.10.17.343889>

Wolfe JM, Fournier GP (2018) Horizontal gene transfer constrains the timing of methanogen evolution. *Nature Ecology & Evolution*, 2, 897–903. <https://doi.org/10.1038/s41559-018-0513-7>

## Reviews

Toggle reviews

---

## *Revision round #2*

2021-04-05

### *Author's Reply*

[Download author's reply \(PDF file\)](#)

## *Decision round #2*

Szollosi et al. made a thorough revision of their manuscript, with an expanded simulation study to address the concerns raised in round 1, and a new analysis to study what makes a constraint informative. I agree with the reviewer. The revision to Figure 1 is very nice.

I believe that the model description needs another revision to be accurate: please see my attached technical comments, for details about the first two equations.

I look forward to receiving a final revision of this interesting paper!

[Download recommender's annotations \(PDF\)](#)

Preprint DOI: <https://doi.org/10.1101/2020.10.17.343889>

## *Reviewed by anonymous reviewer, 2021-04-05 18:52*

The authors have reasonably addressed the major issues I raised in the previous version of the manuscript. I also thank the authors for clarifying the misunderstanding about the empirical source of these constraining HGT events, which is now more clear in the text. The additional SI figures and expanded analysis of the impact of individual HGT events has also greatly improved the manuscript.

---

## *Revision round #1*

2020-12-22

### *Author's Reply*

[Download author's reply \(PDF file\)](#)

#### *Decision round #1*

In this preprint, Szollosi et al. present an method to date a tree using relative age constraint, such as implied by horizontal gene transfer events, and a two-step approach to ease the computational burden. The usefulness and the ease of using the method are exciting.

Both reviewers are positive. The first review is high-level. The second review made excellent suggestions. In particular, one concern is that the simulated trees had modest rate variation and are close to being ultrametric. Looking at the materials on [github](#), one simulated tree looks far from ultrametric to me. The authors could clarify, compare with non-ultrametricity in real trees, and perhaps consider the addition of simulations in which rate transformations are more drastic.

Reviewer 2 made valuable comments about some results interpretation, such as the marked improvement from 4 to 5 constraints, and the value added by proximal vs distal constraints. I very much agree. About distal constraints: I find the authors' conclusion that distal constraints are more informative than proximal constraints counterintuitive. Intuitively, a distal constraint corresponds to a proximal constraint after information loss. For example, a proximal constraint implies distal constraints between the "older" node and any descendant of the "younger" node. As another example, the donor and recipient of a HGT need to have the same age (proximal event), but would provide a distal constraint due to extinction or a lack of speciation events (or lack of sampling) along the lineages "around" the HGT. Like reviewer 2, I invite the authors for more discussion, and I wonder if some other factor is at play in the authors' simulation.

Another comment is about the sparse documentation of the empirical data analysis. I concur: documentation needs to be greatly expanded, to help understanding and increase reproducibility. For example, the lack of documentation made it hard to understand some information and annotations in figures 6 & 7 (e.g.: is the "95% HPD for Viridiplantae" in fig. 7 based on the authors' analysis, or from some other source?).

I attached technical / minor comments and suggestions from my own reading.

[Download recommender's annotations \(PDF\)](#)

Preprint DOI: <https://doi.org/10.1101/2020.10.17.343889>

*Reviewed by [David Duchêne](#), 2020-12-04 23:11*

In their manuscript, Szollosi et al. report an implementation and detailed exploration of a new approach of time-calibration based on relative node times. The approach is intuitive, and to my knowledge has not been described or tested in previous research. The description is very clear and the explorations using simulations and empirical data are thorough. In particular I commend the authors for exploring various widths of calibration and for using such realistic simulation schemes. The method is valuable and I believe that any comments on the methods or manuscript would be a matter of personal preference, rather than academic rigour. For these reasons I wish to recommend this piece in its present form.

*Reviewed by anonymous reviewer, 2020-12-14 20:44*

[Download the review \(PDF file\)](#)