



From Reviewer Requests to Rigorous Evaluation: Building Unbiased Benchmarks for Evolutionary Scenarios

41st TBI Winterseminar, Feb 9, 2026

José Antonio Ramírez Rafael

Annachiara Korchmaros, Katia Aviña-Padilla, Alitzel López-Sánchez, Andrea Arlette España-Tinajero, Gabriel Martinez-Medina, Alfredo J. Hernández-Álvarez, Marc Hellmuth,

Peter F. Stadler, Maribel Hernández-Rosales



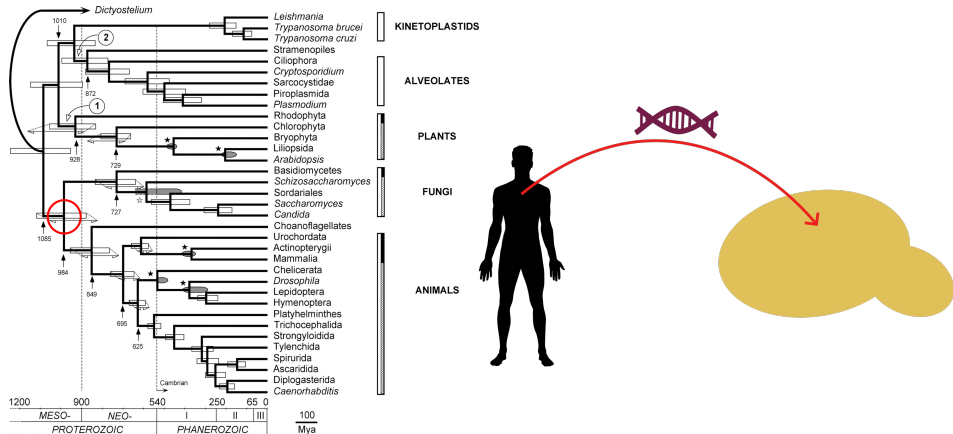
Secretaría de
Ciencia, Humanidades,
Tecnología e Innovación

MAX PLANCK INSTITUTE
FOR MATHEMATICS IN THE SCIENCES



- 1 Preliminaries
- 2 REvolutionH-tl
- 3 QfO challenge
- 4 Orthobench challenge
- 5 Summary

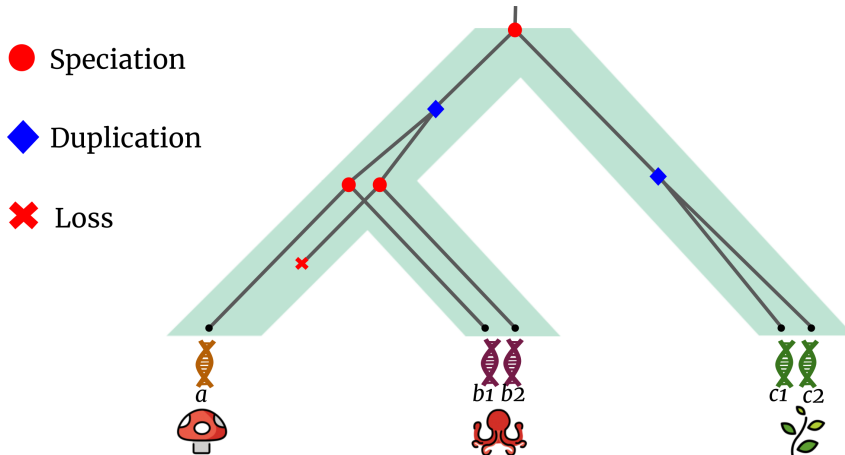
Ortholog genes share ancestral functions



[Kachroo et al., 2015, Douzery et al., 2004]

An evolutionary scenario depicts genes evolving inside species

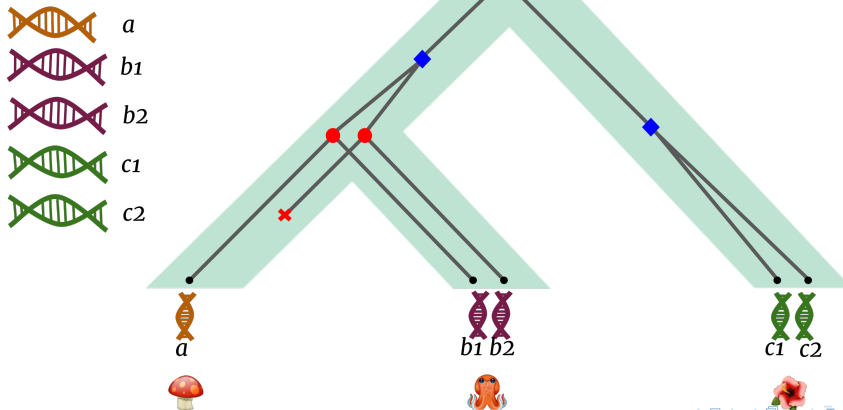
$t : V^0(\mathcal{T}) \rightarrow \{\bullet, \blacklozenge, \times\}$ assign evolutionary events to inner nodes of gene tree.



An orthogroup is part of a gene family

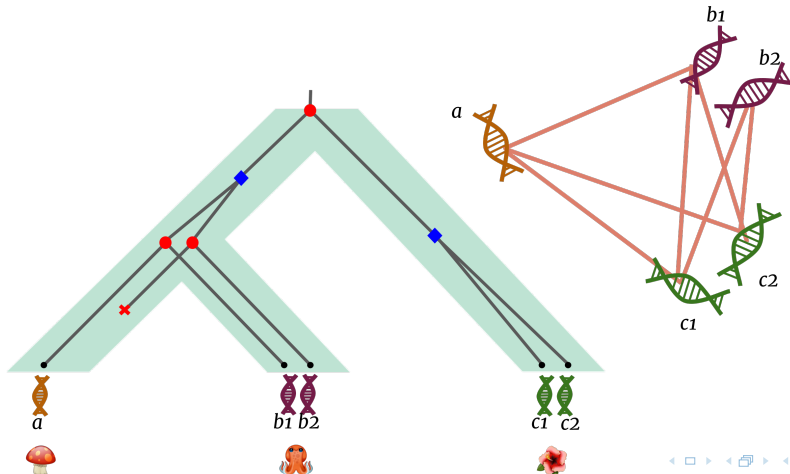
An *orthogroup* is a collection of genes sharing an evolutionary history.

Example:



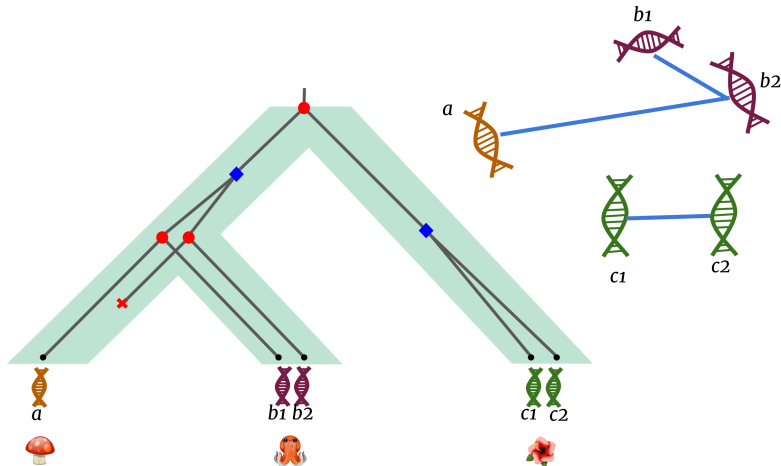
Orthology relies on speciation events

Genes x, y are *orthologs* whenever $t(\text{lca}_T(x, y)) = \bullet$.



Orthology relies on speciation events

Genes x, y are *paralogs* whenever $t(\text{lca}_T(x, y)) = \diamond$.



1 Preliminaries

2 REvolutionH-tl

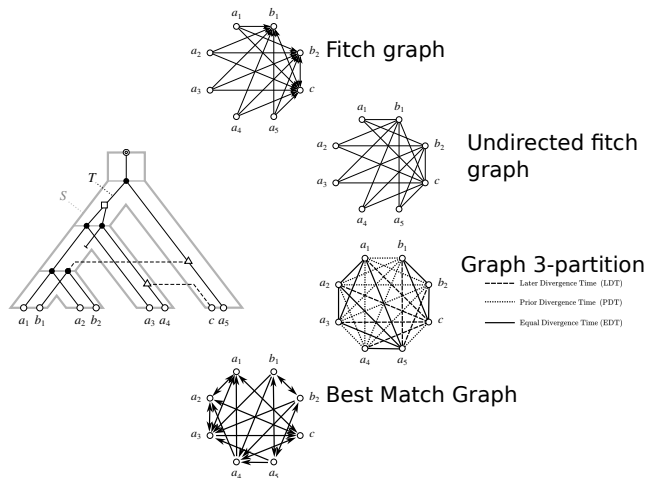
3 QfO challenge

4 Orthobench challenge

5 Summary

Species	Gene ID	Sequence	Sequences
	a	GCTTAAAACTGATTAACCTGCT	
	b1	GCTTCTGATTAACCTGCT	
	b2	GCTTCTGACTAGTCTGCT	
	c1	GCTTCTGATTTCTCCCTGCC	
	c2	GCTTCTGATTACTCCCTGCC	
			Evolutionary scenarios & HGT inference

Graph theory reveals combinatorial restrictions on gene family evolution



[Geiß et al., 2018, Hellmuth et al., 2018, Geiß et al., 2019, Schaller et al., 2021, Schaller et al., 2023, Hellmuth et al., 2024]

REvolutionH-tl: A Fast and Robust Tool for Decoding Evolutionary Gene Histories



REvolutionH-tl

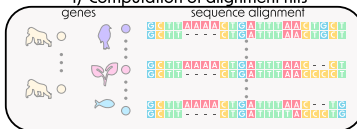
Reconstruction of Evolutionary Histories Tool

input

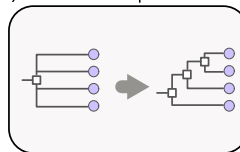


FASTA files

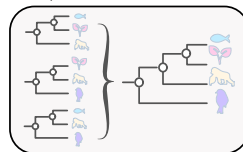
1) Computation of alignment hits



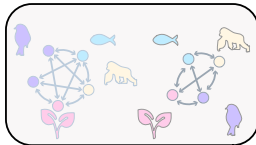
4) Resolution of duplication nodes



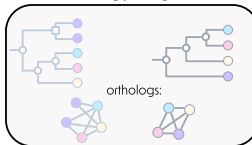
5) Species tree reconstruction



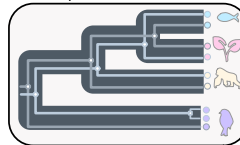
2) BMG estimation and orthogroup detection



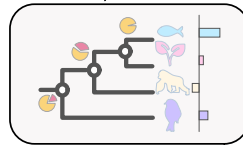
3) Gene tree reconstruction and orthology assignment



6) Tree reconciliation



7) Visualization



[Ramírez-Rafael et al., 2025]

① Preliminaries

② REvolutionH-tl

③ QfO challenge

④ Orthobench challenge

⑤ Summary

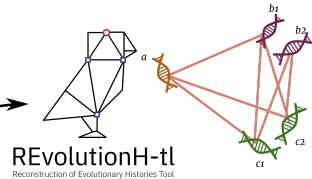
QfO Orthology benchmark pipeline

Orthology Benchmarking

Nevers Y et al.,
The Quest for Orthologs benchmark service in 2022
Nucleic Acids Res, 2022 (doi: 10.1093/nar/gkac330)



78 Genomes:
- 48 Eukaryotes
- 23 Bacteria
- 7 Archaea



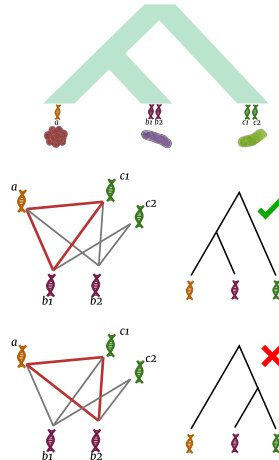
REvolutionH-tl
Reconstruction of Evolutionary Histories Tool



There is no ground truth for the evolution of living organisms

Surrogate measurements for correctness of orthology prediction:

- **Generalized species tree discordance**
- Reference gene trees
- Functional tests



There is no ground truth for the evolution of living organisms

Surrogate measurements for correctness of orthology prediction:

- Generalized species tree discordance
- **Reference gene trees**
- Functional tests

SwissTree



Treefam



There is no ground truth for the evolution of living organisms

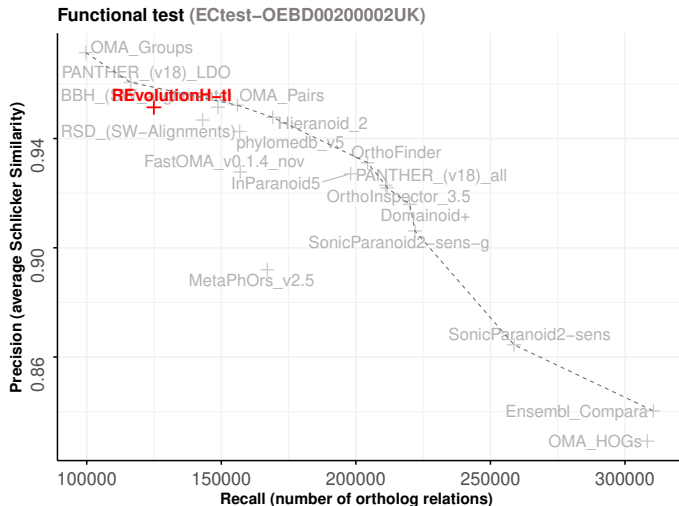
Surrogate measurements for correctness of orthology prediction:

- Generalized species tree discordance
- Reference gene trees
- **Functional tests**

ENZYME Swiss-Prot



QfO recall assessment is biologically meaningless



REvolutionH-t1 is in the top 5 most precise tools

Benchmark Type	FuncOonal		Phylogeny														Reference		
Metric	Schlicker Avg.		FracOon of incorrect trees							Robinson-Foulds Average							PPV		
Benchmark	EC	GO	GSTD2 Eukary.	GSTD2 Fungi	GSTD2 Luca	GSTD2 Vert.	STD Bacteria	STD Eukary.	STD Fungi	GSTD2 Eukary.	GSTD2 Fungi	GSTD2 Luca	GSTD2 Vert.	STD Bacteria	STD Eukary.	STD Fungi	SwissTree	TreeFam-A	VGNC
Domainoid+																			
OMA-Groups																			
BBH (SW-alignments)																			
OMA-Pairs																			
Hieranoid 2																			
OMA-HOGs																			
Ensembl-Compars																			
OrthoInspector 3.5																			
Phylomedb-v5																			
PANTHER-v18-LDO																			
PANTHER-v18-all																			
SonicParanoid2-sens-g																			
InParanoid5																			
SonicParanoid2-sens																			
MetaPhOrs-v2.5																			
RSD (SW-Alignments)																			
FastOMA-v0.1.4-nov																			
OrthoFinder																			
REvolutionH-t1	5	5	2	2	8	3	2	16	3	2	4	1	16	2	19	4	18	16	19

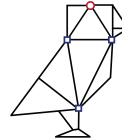
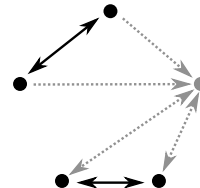
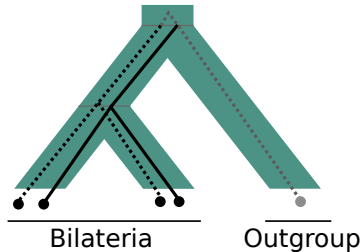
- 1 Preliminaries
- 2 REvolutionH-tl
- 3 QfO challenge
- 4 Orthobench challenge
- 5 Summary

REvolutionH-tl provide reference orthogroups

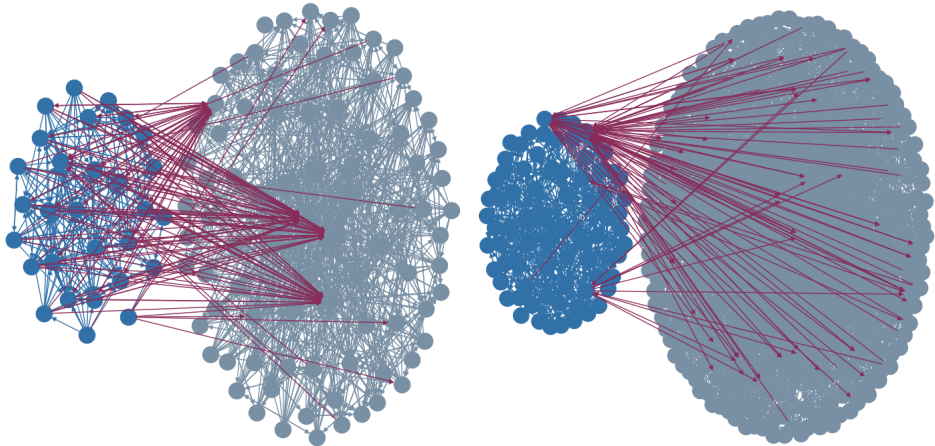
Orthobench

**17 Genomes:**

- 12 Bilateria
- 3 Outgroup
- 2 Ingroup

**REvolutionH-tl**
Reconstruction of Evolutionary Histories Tool

We have many false positives with respect to orthobench reference



Orthobench dataset is old



JOURNAL ARTICLE

Benchmarking Orthogroup Inference Accuracy: Revisiting Orthobench

David M Emms , Steven Kelly *Genome Biology and Evolution*, Volume 12, Issue 12, December 2020, Pages 2258–2266,<https://doi.org/10.1093/gbe/evaa211>Published: 06 October 2020 [Article history ▼](#)

BioEssays

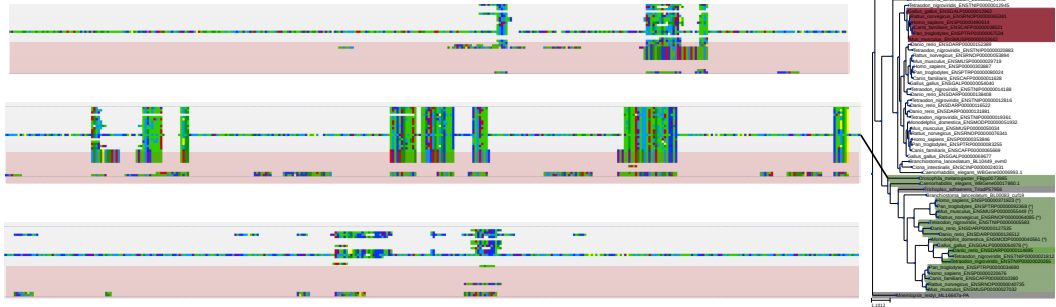
Prospects & Overviews |  Open Access

Orthology prediction methods: A quality assessment using curated protein families[†]

[Kalliopi Trachana](#), [Tomas A. Larsson](#), [Sean Powell](#), [Wei-Hua Chen](#), [Tobias Doerks](#), [Jean Muller](#), [Peer Bork](#) First published: 19 August 2011 | <https://doi.org/10.1002/bies.201100062> | [VIEW METRICS](#)

Orthobench contains miss annotated sequences

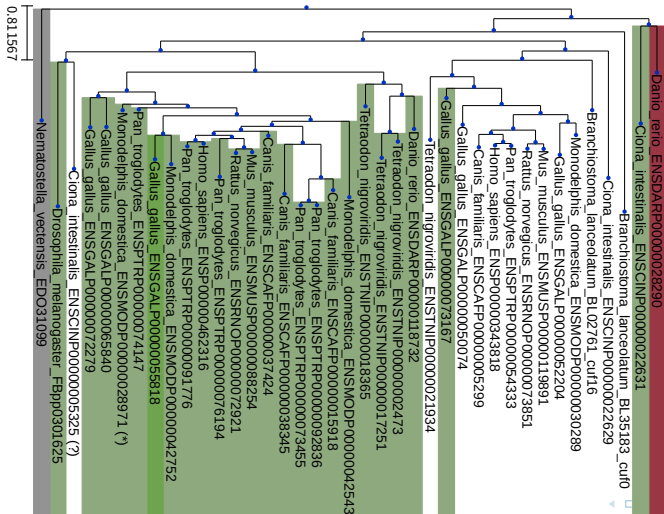
RefOG002 + false positives



Orthobench contains outdated sequences

- **Retired/deprecated database entries** – Benchmarks use obsolete ENSEMBL records no longer present in current genome annotations.
- **Extreme fragment contamination** – Reference sets include very short sequences (<500 aa) labeled as fulllength proteins.
- **Sequenceannotation length mismatches** – Input sequences exceed ENSEMBLannotated lengths, suggesting unannotated introns/UTRs.
- **Domainlevel overmerging** – Groups defined by a single conserved domain capture unrelated fulllength proteins (e.g., kinase domains in RefOG027).

Orthobench contains inconsistent reference orthogroups

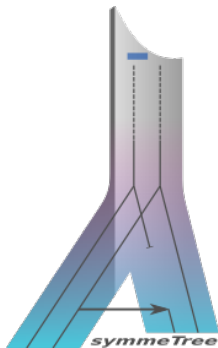


- 1 Preliminaries
- 2 REvolutionH-tl
- 3 QfO challenge
- 4 Orthobench challenge
- 5 Summary

Take-home notes

- Reconstruction of the complex evolutionary scenarios is crucial for understanding the diversity of life.
- High quality ground-truth datasets are needed to asses and improve pipelines of evolutionary analysis.
- State of the art methodologies provide a framework to compare prediction of different tools with real genomes.
- State of the art methodologies may provided biased metrics and should be reviewed.

Next steps: publish a complex synthetic dataset and revised orthobenchd



- Species trees with 5, 10, ... 50 species
- Different rates for duplication, loss, and transfer events in gene trees.
- Provide true reconciliation.
- Relaxed molecular clock scenarios.
- Domain evolution across gene sequences.

Thanks for your attention

Advisors:

Peter F. Stadler,
Maribel Hernández Rosales

Collaborators:

Annachiara Korchmaros, Alitzel López-Sánchez,
Katia Aviña-Padilla, Marc Hellmuth, Alfredo J.
Hernández-Álvarez, Andrea Arlette
España-Tinajero, Gabriel Martinez-Medina

Contact:

jose.ramirezra@cinvestav.mx



MAX PLANCK INSTITUTE
FOR MATHEMATICS IN THE SCIENCES



BIOINFORMATICS
Leipzig



UNIVERSITÄT
LEIPZIG

Call for the 5th Retreat on Bioinformatics and Complex Networks



6 References

- [Douzery et al., 2004] Douzery, E. J. P., Snell, E. A., Bapteste, E., Delsuc, F., and Philippe, H. (2004).
The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils?
Proc. Natl. Acad. Sci. U. S. A., 101(43):15386–15391.
- [Geiß et al., 2018] Geiß, M., Anders, J., Stadler, P. F., Wieseke, N., and Hellmuth, M. (2018).
Reconstructing gene trees from fitch's xenology relation.
J. Math. Biol., 77(5):1459–1491.
- [Geiß et al., 2019] Geiß, M., Laffitte, M. G., Sánchez, A. L., Valdivia, D., Hellmuth, M., Rosales, M. H., and Stadler, P. (2019).
Best match graphs and reconciliation of gene trees with species trees. corr.
arXiv preprint arXiv:1904.12021.
- [Hellmuth et al., 2018] Hellmuth, M., Long, Y., Geiß, M., and Stadler, P. F. (2018).
A short note on undirected fitch graphs.
Art Discrete Appl. Math., 1(1):#1.08.
- [Hellmuth et al., 2024] Hellmuth, M., Stadler, P. F., and Thekkumpadan Puthiyaveedu, S. (2024).
Fitch graph completion.
In *Lecture Notes in Computer Science*, Lecture notes in computer science, pages 225–237. Springer Nature Switzerland, Cham.
- [Kachroo et al., 2015] Kachroo, A. H., Laurent, J. M., Yellman, C. M., Meyer, A. G., Wilke, C. O., and Marcotte, E. M. (2015).
Evolution. systematic humanization of yeast genes reveals conserved functions and genetic modularity.
Science, 348(6237):921–925.
- [Ramírez-Rafael et al., 2025] Ramírez-Rafael, J. A., Korchmaros, A., Aviña-Padilla, K., López-Sánchez, A., Martínez-Medina, G., Hernández-Álvarez, A. J., Hellmuth, M., Stadler, P. F., and Hernández-Rosales, M. (2025).
REvolutionH-tl: A fast and robust tool for decoding evolutionary gene histories.

- [Schaller et al., 2023] Schaller, D., Hellmuth, M., and Stadler, P. F. (2023).
Orientation of fitch graphs and reconciliation-free inference of horizontal gene transfer in gene trees.
SIAM J. Discrete Math., 37(3):2172–2207.
- [Schaller et al., 2021] Schaller, D., Lafond, M., Stadler, P. F., Wieseke, N., and Hellmuth, M. (2021).
Indirect identification of horizontal gene transfer.
Journal of Mathematical Biology, 83(1):10.