



HAL
open science

Thirdkind: displaying phylogenetic encounters beyond 2-level reconciliation

Simon Penel, Hugo Menet, Théo Tricou, Vincent Daubin, Eric Tannier

► To cite this version:

Simon Penel, Hugo Menet, Théo Tricou, Vincent Daubin, Eric Tannier. Thirdkind: displaying phylogenetic encounters beyond 2-level reconciliation. *Bioinformatics*, 2022, 38 (8), pp.2350-2352. 10.1093/bioinformatics/btac062 . hal-03563666

HAL Id: hal-03563666

<https://hal.science/hal-03563666v1>

Submitted on 30 Mar 2022

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.

Thirdkind: displaying phylogenetic encounters beyond 2-level reconciliation.

Simon Penel 1 * , Hugo Menet 1 ,Théo Tricou 1 , Vincent Daubin 1 and Eric Tannier 1,2

Laboratoire de Biométrie et Biologie Evolutive/UMR5558, CNRS/UCBL, Villeurbanne, 69622, France.
Centre de Recherche Inria Lyon, Villeurbanne, 69622,France.

Abstract

Motivation: Reconciliation between a host and its symbiont phylogenies or between a species and a gene phylogenies is a prevalent approach in evolution, however no simple generic tool (i.e. virtually usable by all reconciliation software, from host/symbiont to species/gene comparisons) is available to visualise reconciliation results. Moreover there is no tool to visualise 3-levels reconciliations, i.e. to visualise 2 nested reconciliations as for example in a host/symbiont/gene complex.

Results: Thirdkind is a light and easy to install command line software producing svg files displaying reconciliations, including 3-levels reconciliations. It takes a standard format recPhyloXML as input, and is thus usable with most reconciliation software.

Availability: <https://github.com/simonpenel/thirdkind/wiki>

Contact: simon.penel@univ-lyon1.fr

1 Introduction

Phylogenetic reconciliation consists in linking the history of two (or more) co-evolving biological entities, at two different levels of organization. It is achieved by embedding a phylogenetic tree into another, pointing at the dependencies between evolutionary histories and the level-specific events. For example, gene phylogenetic trees can be mapped into species trees, explaining the topological differences by speciation, duplication, loss, horizontal transfer. This can also apply for the comparison of host and symbiont co-evolution, or the evolution of protein domains inside genes, etc...

Several methods and available software are dedicated to constructing reconciliations (for a review, see Menet et al., 2021). The XML format “recPhyloXML” has been proposed as a standard to describe phylogenetic reconciliations (Duchemin et al., 2018) and is now produced directly or via available translation scripts by a majority of reconciliation software.

Visualisation of phylogenetic reconciliations are proposed by various programs and interfaces as NOTUNG (Chen et al., 2000), SylvX (Chevenet et al., 2016), Treerecs (Comte et al., 2020), Jane (Conow et al., 2010), eMPress (Santichaivekin et al., 2021) and Capybara (Wang et al., 2020).

However at the exception of SylvX, all are integrated in a specific reconciliation software and cannot visualise reconciliations produced by others. None of these software is handling recPhyloXML input files (at the exception of the unpublished Recphylovisu web interface <http://phylariane.univ-lyon1.fr/recphyloxml/recphylovisu> of which Thirdkind can be considered as an update), and none of them is generic to any kind of reconciliation (for example SylvX does not allow temporary free living symbionts, as it is not allowed for genes to live outside a genome) nor can handle multiple horizontal transfer (i.e. several genes transferred with the same donor and recipient) and the consideration of numerous possible scenarios.

DoubleRecViz (Kuitche et al., 2021) uses a derived version of recPhyloXML, adding a transcript level to gene and species format but without support for horizontal transfers.

Eventually there is no software able to combine two nested reconciliations i.e. to get in a single representation the gene/symbiont reconciliation and the symbiont/host reconciliation, despite the recent interest of methodologists in these 3-levels systems (Stolzer et al., 2015; Kundu et al., 2019; Muhammad et al., 2018) answering current and significant questions in biology (Menet et al., 2021).

Thirdkind is a light command-line software allowing the user to generate a svg from recPhyloXML files with a large choice of options (orientation, police size, branch length, multiple trees, redundant transfers handling, etc.) and to handle the visualisation of 2 nested reconciliations. Trees can be dated via their branch lengths or undated.

2 Installation

Thirdkind is written in Rust (<https://crates.io/crates/thirdkind>) and thus very easy to install: install cargo and then type 'cargo install thirdkind'.

Source code is available at <https://github.com/simonpenel/thirdkind>.

3 Usage

3.1 Input files (option -f)

Thirdkind is dedicated to read recPhyloXML format files, but it can read newick or phyloXML files if needed. The option -f is used to indicate the name of the input file, whatever the format. The format is guessed from the extension of the file, or it can be chosen with the -F option.

Newick is a simple parenthesed tree format, phyloXML is a xml format dedicated to phylogeny, recPhyloXML derives from phyloXML and is dedicated to reconciliations. A phyloXML file contains only 1 tree (reconciled or not). A recPhyloXML file contains at least one "upper" tree (the species tree in the species/gene complex, or the host in a host/symbiont complex) and one "lower" reconciled tree (respectively the gene, or symbiont) mapped to (one of) the upper tree(s). A clade presents several tags: a name, a location, a type of event, etc. Each node of the lower tree(s) has a "location", which should correspond to the "name" of one of the clades in the upper tree(s). It is possible to have multiple lower trees, and multiples upper trees in a single file.

3.2 Output styles

Thirdkind allows to generate svg ouput files according to 2 different styles:

1) The "recphyloxml style" where the lower tree(s) and upper tree(s) are displayed, the lower tree(s) being embedded into the upper tree(s). The output consists of one ore several reconciled lower trees drawn as lines inside or outside one or several upper trees, drawn as tubes. Lower trees have symbols at their nodes (a square for a duplication, a circle for a speciation and a cross for a loss). Transfers (or host switches) are Bezier spotted lines ending with an arrow. If there is more than one lower tree, they are displayed with different colors (Figure 1).

2) The "phyloxml style" where a unique tree is displayed, with symbol within the tree representing evolution events: a square for a duplication, a circle for a speciation, a cross for a loss, transfer is a spotted line between an orange diamond and a yellow triangle (Figure 1, above and suppl.mat.). If the input format is phyloXML or newick, the style of the output is "phyloxml". If the input format is recPhyloXML, the style of the output is "recphyloxml", or "phyloxml" with the options -S (display the upper tree only) and -G (display one of the reconciled lower tree). Note that if the file format is recPhyloXML but the file extension is phyloXML, the output style will be the upper tree in "phyloxml" style.

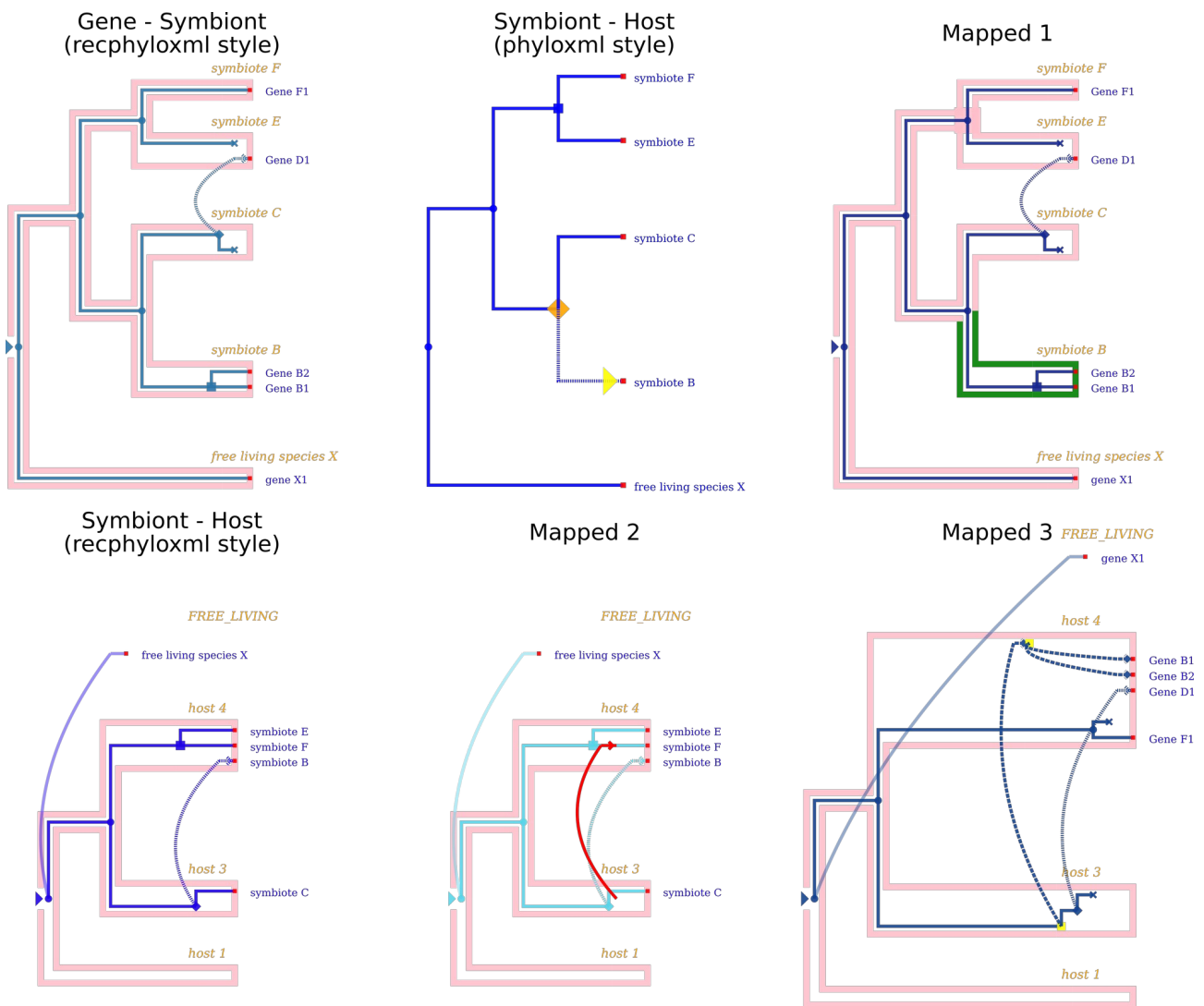


Fig. 1. The 6 main outputs of a 3-level gene/symbiont/host reconciliation. one “recphyloxml style” svg for each of the two input, a “phyloxml style” svg of the reconciled symbiont tree (from -f file) and three “mapped” svg files describing the gene/symbiont/host reconciliation. The first “mapped” svg is a modified version of the “recphyloxml style” svg of the gene/symbiont reconciliation: the drawing of the symbiont tree as a tube presents features describing its reconciliation with the host (a big square for a duplication node, an additional tube branch colored in black for a loss and the tube segments between the start and end of a transfer are colored in green). The second “mapped” svg is a modified version of the “recphyloxml style” svg of the symbiont/host reconciliation in which gene transfers are mapped to the host nodes and displayed in red: a gene transfer between the symbiont “C” present in host “3” and the symbiont “E” present in host “4” is displayed as a red Bezier line between host “3” and host “4” in the tube host tree. The third “mapped” is a mapping of the genes trees over the host tree through the symbiont. If genes “B1” and “B2” are associated to the symbiont “B”, and the symbiont B is associated to host “4”, the genes B1, B2 are associated to host “4” in the svg. If a gene is transferred between hosts via a symbiont transfer, the transfer start with a yellow square and the stippling is different. A gene transfer across symbionts which is not affected by a transfer of the symbiont across hosts is displayed as a classic gene transfer.

3.3 Dated trees (option -l)

The -l option allows to use the branch lengths to display the tree. It applies to the upper tree in “recphyloxml” style context, and to the current tree otherwise.

3.4 Minimising transfer crossings (option -O)

Since minimising transfer crossings is a NP-hard problem (Klavitter and Stumpf, 2020) we used a simple heuristic : explore the upper tree from the start node and the end node of a transfer until the ancestor of these 2 nodes, giving to each node on the way a score reflecting how it should be oriented to reduce the distance between start and end nodes. This is mainly useful when studying a single tree with few transfers (with 1 transfer the heuristic finds the best solution).

3.5 Multiple recPhyloXML input files (option -m)

It is possible to use a list of recPhyloXML files instead a single recPhyloXML file. This option is useful to handle large sets of gene histories inside species histories as the ones generated by GeneRax (Morel et al., 2020).

3.6 Dealing with redundant transfers (options -t , -T and -J)

In case of multiple gene histories it may be useful to enlighten redundant gene transfers (i.e. with the same donor and recipient), when trying to identify highways of transfer for instance. Option -t draws only one gene history without the transfers and then in red the transfers of all the histories according to their abundance: only the transfers with an abundance higher that the threshold are drawn and the opacity reflects the abundance. The option -T allows to choose the gene history to display. The option -J displays the abundance of the transfer. Thirdkind can visualise two kind of redundancy: multiple separate scenarios for one gene family, in which case the different transfers seen can belong to different reconciliation histories, and multiple gene families with one scenario for each family as seen in Figure 2.

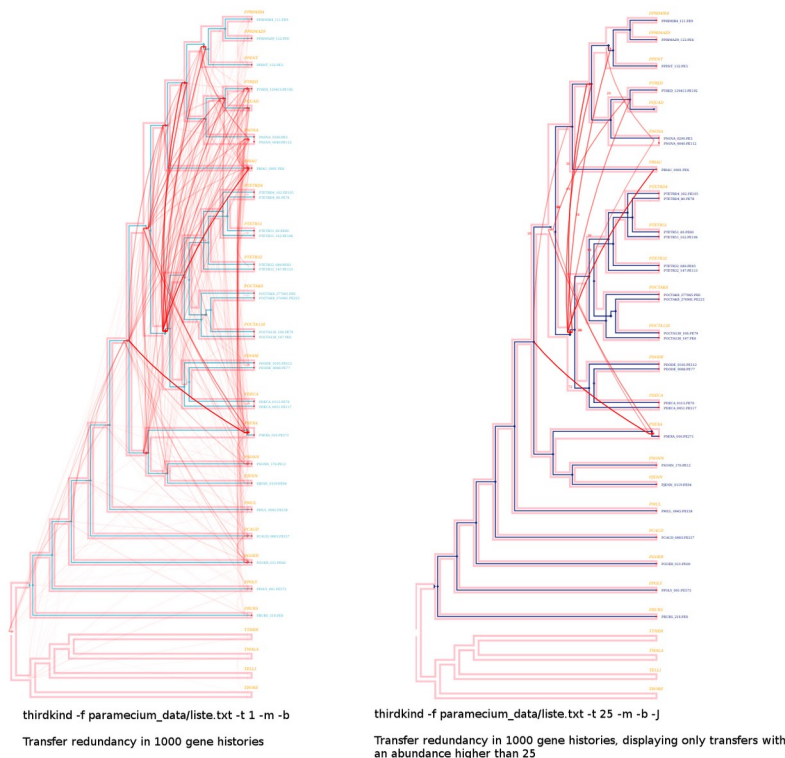


Figure 2. Using -m and -t options to display redundant gene transfers.

3.7 Dealing with 'free living' symbionts (options -e)

In the history of a micro-organism, some taxa may be free living species and some others may have evolved to be a symbiont of a host. In this case, free living organisms should have a "location" indicating "FREE LIVING" instead of the name of a host. Thirdkind draws the free living part of the symbiont path tree outside the upper tree (Figure 3).

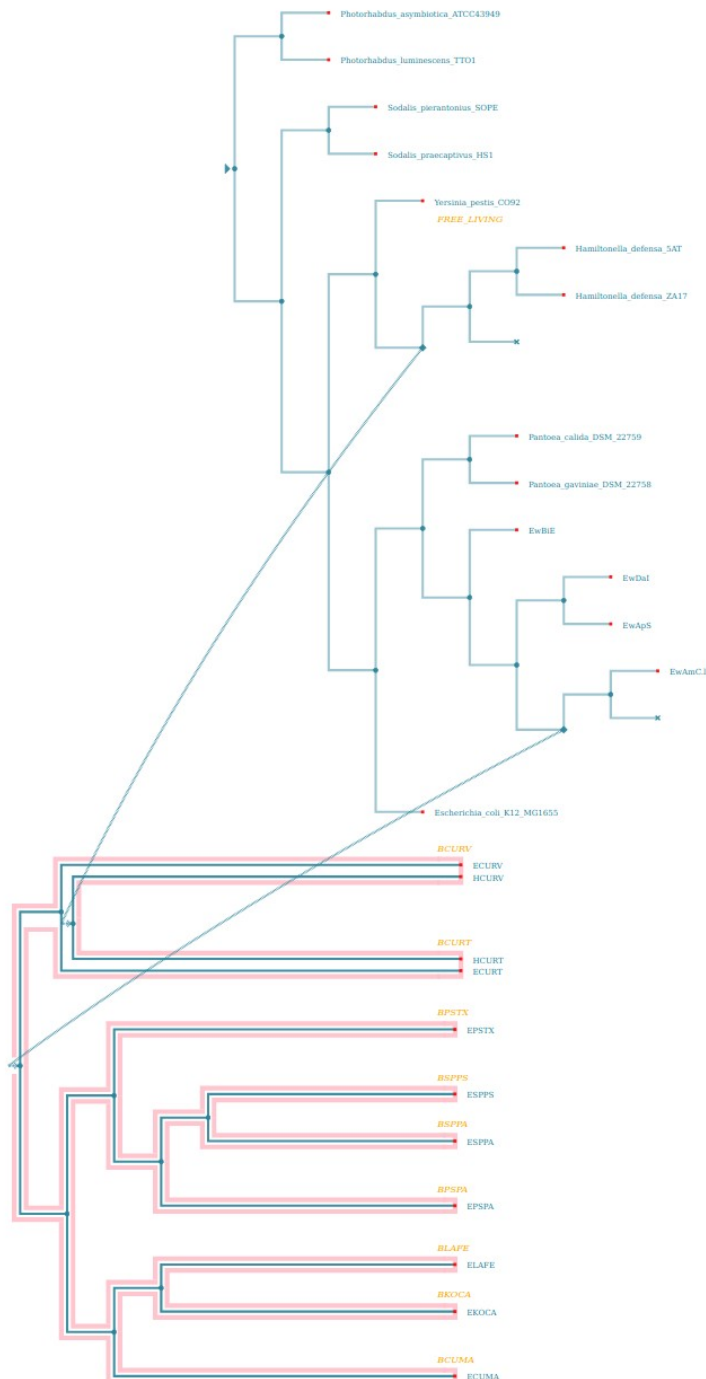


Figure 3. Using -e option to display free living species and symbionts.

3.8 Nested recPhyloXML files (options -g and -f)

It is possible to combine two reconciliations as for example a gene/species reconciliation and a symbiont/host reconciliation, in which the symbiont of the second reconciliation is the species of the first one. This is valid for any variant of a 3-levels co-evolution, as geography/species/genes, or species/gene/domains. For clarity of the exposition we adopt the host/symbiont/gene vocabulary, keeping in mind the genericity of the method. The -g option indicates the gene/species file, -f indicates the symbiont/host file. The software generates several svg files: one “recphyloxml style” svg for each of the two input, a “phyloxml style” svg of the reconciled symbiont tree (from -f file), a “phyloxml style” svg of each reconciled gene trees (from -g file) and three “mapped” svg files describing the gene/symbiont/host reconciliation (Figure 1).

3.9 Execution time and readability

Thirdkind was able to process 5,000 reconciled trees of 50 nodes in 2 seconds and to process a tree of 7,000 nodes in 1 second. The readability depends on the ability to handle the resulting svg (desktop computer, mobile, poster, etc).

Acknowledgements. Grant from Agence Nationale pour la Recherche, ANR-19-CE45-0010.

References

- Chen,K. et al. (2010) NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.*, 7, 429–447.
- Chevenet,F. et al. (2016) SylvX: a viewer for phylogenetic tree reconciliations. *Bioinformatics*, 32, 608–610.
- Comte,N. et al. (2020) Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics*, 36, 4822–4824.
- Conow,C. et al. (2010) Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms Mol Biol.*, DOI: 10.1186/1748-7188-5-16.
- Duchemin,W. et al. (2018) RecPhyloXML: a format for reconciled gene trees. *Bioinformatics*, 34, 3646–3652.
- Klavitter,J. and Stumpf, P. (2020) Drawing Tree-Based Phylogenetic Networks with Minimum Number of Crossings, <https://arxiv.org/abs/2008.08960>
- Kuitche,E. et al. (2021) DoubleRecViz: a web-based tool for visualizing transcript–gene–species tree reconciliation. *Bioinformatics*, 37, 1920–1922.
- Kundu,S. and Bansal,M.S. (2019) SaGePhy: an improved phylogenetic simulation framework for gene and subgene evolution. *Bioinformatics*, 35, 3496–3498.
- Menet,H. et al. (2021) Phylogenetic reconciliation. <https://hal.archives-ouvertes.fr/hal-03258402>.
- Morel,B. et al. (2020) GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss. *Molecular Biology and Evolution*, 37, 2763–2774.
- Muhammad,S.A. et al. (2018) Species tree-aware simultaneous reconstruction of gene and domain evolution. *bioRxiv*, <https://doi.org/10.1101/336453>.
- Santichaivekin,S. et al. (2021) eMPress: a systematic cophylogeny reconciliation tool. *Bioinformatics*, 37, 2481–2482.
- Stolzer,M. et al. (2015) Event inference in multidomain families with phylogenetic reconciliation. *BMC Bioinformatics*, 16, S8.
- Wang,Y. et al. (2021) Capybara: equivalence CLASS enumeration of coPhylogenY event-BASed ReconciliAtions. *Bioinformatics*, 36, 4197–4199.