

TOPIC PAGE

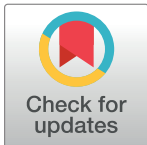
# Phylogenetic reconciliation

Hugo Menet<sup>1</sup>, Vincent Daubin<sup>1\*</sup>, Eric Tannier<sup>1,2\*</sup>

**1** Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR5558, Villeurbanne, France, **2** Inria, centre de recherche de Lyon, Villeurbanne, France

\* [vincent.daubin@univ-lyon1.fr](mailto:vincent.daubin@univ-lyon1.fr) (VD); [eric.tannier@inria.fr](mailto:eric.tannier@inria.fr) (ET)

This is a “Topic Page” article for *PLOS Computational Biology*.



## OPEN ACCESS

**Citation:** Menet H, Daubin V, Tannier E (2022) Phylogenetic reconciliation. *PLoS Comput Biol* 18(11): e1010621. <https://doi.org/10.1371/journal.pcbi.1010621>

**Editor:** Daniel Mietchen, University of Virginia, UNITED STATES

**Published:** November 3, 2022

**Copyright:** © 2022 Menet 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 (Grant ANR-19-CE45-0010 Evolution) to VD and ET. HM was the recipient of a fellowship from the Ministère de l'Éducation Supérieure et de la Recherche Français through a Contrat Doctoral Spécifique Normalien. 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.

**Wikipedia version:** [https://en.wikipedia.org/wiki/Phylogenetic\\_reconciliation](https://en.wikipedia.org/wiki/Phylogenetic_reconciliation)

## Definition

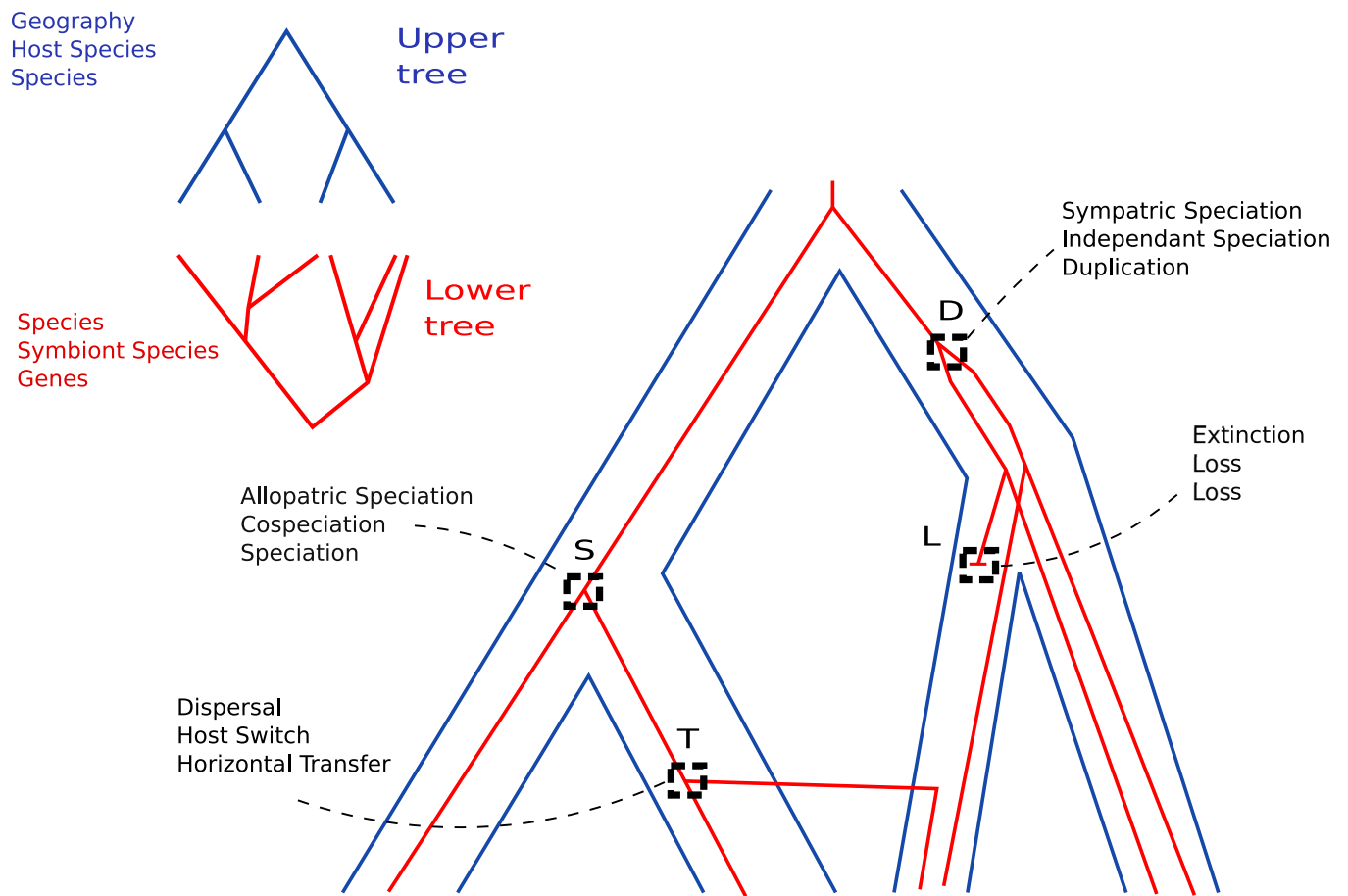
In [phylogenetics](#), reconciliation is an approach to connect the history of two or more [coevolving](#) biological entities. The general idea of reconciliation is that a [phylogenetic tree](#) representing the evolution of an entity (e.g. [homologous genes](#), [symbionts](#)...) can be drawn within another phylogenetic tree representing an encompassing entity (respectively, species, hosts) to reveal their interdependence and the evolutionary events that have marked their shared history ([Fig 1](#)). The development of reconciliation approaches started in the 1980s, mainly to depict the coevolution of a gene and a genome, and of a host and a symbiont, which can be mutualist, commensalist or parasitic. It has also been used for example to [detect horizontal gene transfer](#), or understand the dynamics of [genome evolution](#).

Phylogenetic reconciliation can account for a diversity of evolutionary trajectories of what makes life's history, intertwined with each other at all scales that can be considered, from molecules to populations or cultures. A recent avatar of the importance of interactions between levels of organization is the [holobiont](#) concept, where a macro-organism is seen as a complex partnership of diverse species. Modeling the evolution of such complex entities are one of the challenging and exciting direction of current research on reconciliation.

## Phylogenetic trees as matryoshka dolls

Phylogenies have been used for representing the diversification of life at many [levels of organization](#): macro-organisms [1], their [cells](#) throughout [development](#) [2], [micro-organisms](#) through marker genes [3], [chromosomes](#) [4], [proteins](#) [5], [protein domains](#) [6], and can also be helpful to understand the evolution of human culture elements such as languages [7] or folktales [8]. At each of these levels, phylogenetic trees describe different stories made of specific diversification events, which may or may not be shared among levels. Yet because they are structurally nested or functionally dependent, the evolution at a particular level is bound to others.

Phylogenetic reconciliation is the identification of the links between levels through the comparison of at least two associated trees. Originally developed for two trees, reconciliations for more than two levels have been recently constructed. As such, reconciliation provides evolutionary scenarios that reveal conflict and cooperation among evolving entities. These links may be unintuitive, for instance, genes present in the same genome may show uncorrelated evolutionary histories while some genes present in the genome of a symbiont may show a



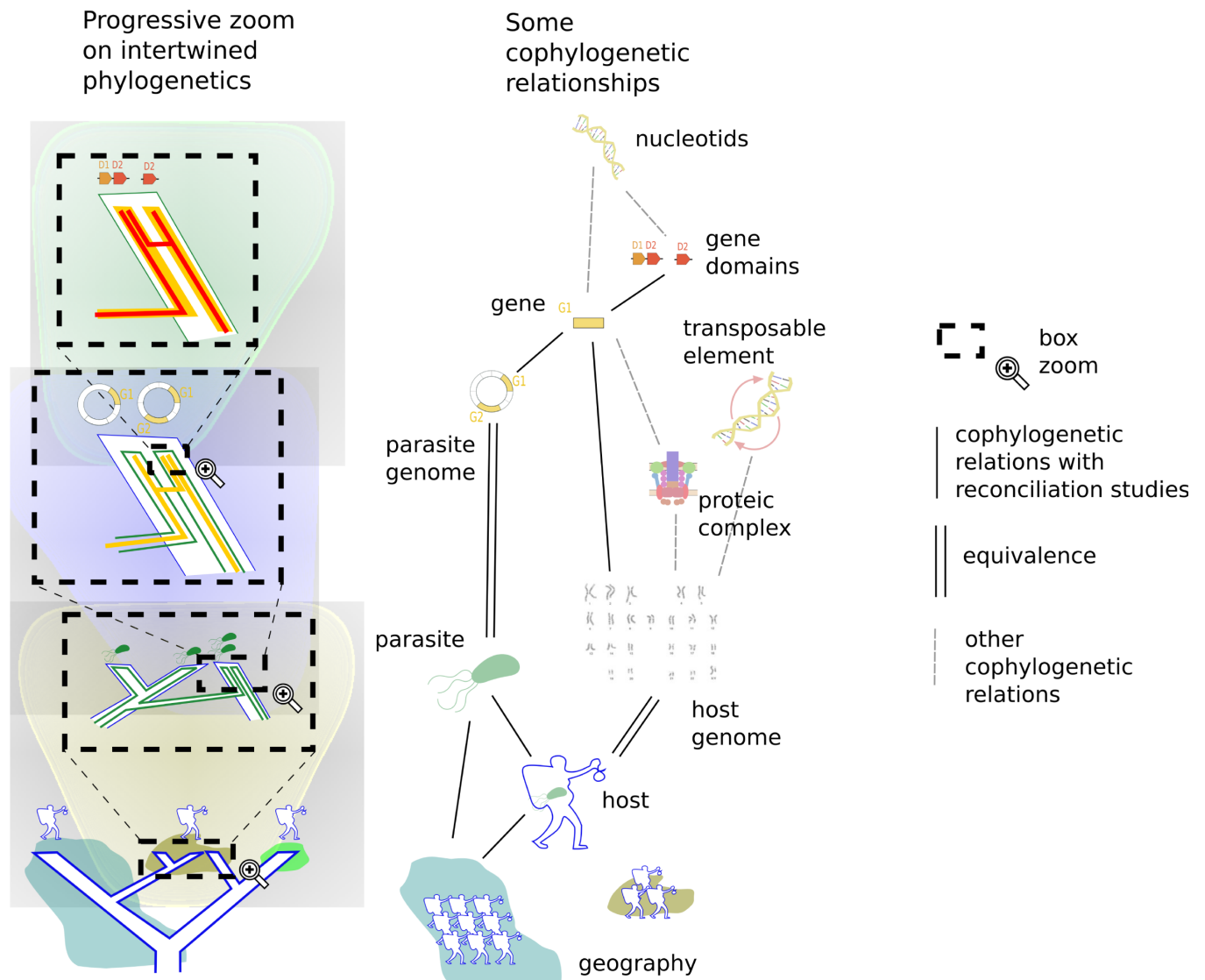
**Fig 1. Phylogenetic reconciliation.** A phylogenetic reconciliation between an upper, blue, and a lower, red, tree, with the most often used evolutionary events (S,D,T,L), and their name in phylogeography, host/symbiont and gene/species frameworks. For instance S event is called allopatric speciation when reconciling geographical areas and species, cospeciation between host and symbiont, and speciation for gene and species, but always correspond to the same co-diversification pattern.

<https://doi.org/10.1371/journal.pcbi.1010621.g001>

strong coevolution signal with the host phylogeny. Hence, reconciliation can be a useful tool to understand the constraints and evolutionary strategies underlying the assemblage that makes an holobiont.

Because all levels essentially deal with the same object, a phylogenetic tree, the same models of reconciliation, in particular those based on duplication-transfer-loss events, which are central to this article, can be transposed, with slight modifications, to any pair of connected levels [9]: an "inner", "lower", or "associate" entity (gene, symbiont species, population. . .) evolves inside an "upper", or "host" one (respectively species, host, geographical area. . .) (Fig 2). The upper and lower entities are partially bound to the same history, leading to similarities in their phylogenetic trees, but the associations can change over time, become more or less strict or switch to other partners (Fig 1).

In the following part of this text, we will give a review of DTL reconciliation methods and models, starting by an historical and methodological approach to the construction of the model. Two-level reconciliation methods, have been reviewed several times, but generally focusing on a particular pair of levels, e.g. gene/species or host/symbiont [10–16], the following parts are written with a generic voice and to confront models constructed in different frameworks. The last part of the article focus on efforts toward reconciliation with more than two levels, and a description of some biological studies that look at such models.



**Fig 2. Reconciliation and biological levels of organization.** Phylogenetic trees are intertwined at all levels of organization, integrating conflicts and dependencies within and between levels. Macro-organism populations migrate between continents, their microbe symbionts switch between populations, the genes of their symbionts transfer between microbe species, and domains are exchanged between genes (left third). This list of organization levels is not representative or exhaustive, but give a view of levels where reconciliation methods have been used. As a generic method, reconciliation could take into account numerous other levels, for instance it could consider the syntenic organization of genes [155,160], the interacting history of transposable elements and species [180], the evolution of protein complex among species [181]. The scale of evolutionary events considered can go from population events such as geographical diversification to nucleotides levels one inside genes [34], including for instance chromosome levels events inside genomes such as whole genome duplication [155].

<https://doi.org/10.1371/journal.pcbi.1010621.g002>

### History

The principle of phylogenetic reconciliation was introduced in 1979 [17] to account for differences between genes and species phylogenies. In a parsimonious setting, two evolutionary events, gene duplication and gene loss were invoked to explain the discrepancies between a gene tree and a species tree. It also described a score on gene trees knowing the species tree and an aligned sequence by using the number of gene duplication, loss, and nucleotide replacement for the evolution of the aligned sequence, an approach still central today with new models of reconciliation and phylogeny inference [18].

The name *reconciliation* has been used by Maddison, 1997 [19], as a reverse image of "phylogenetic discord" resulting from gene level evolutionary events.

Reconciliation was then developed jointly for the coevolution of host and symbiont and the diversification of species on geography. In both settings, it was important to model a horizontal event that implied parallel branches of the host tree: host switch for host and symbiont and species dispersion from one area to another in biogeography. Unlike genes and genomes, the coevolution of host and symbiont and the explanation of species diversification by geography are not always the null hypothesis. A visual depiction of the two phylogenies in a tanglegram can help assess such coevolution, although it has no statistical obvious interpretation [20].

Character methods, such as Brooks Parsimony Analysis [21], were proposed to test coevolution and reconstruct scenarios of coevolution. In these methods, one of the trees is forgotten except for its leaves, which are then used as a character evolving on the second tree.

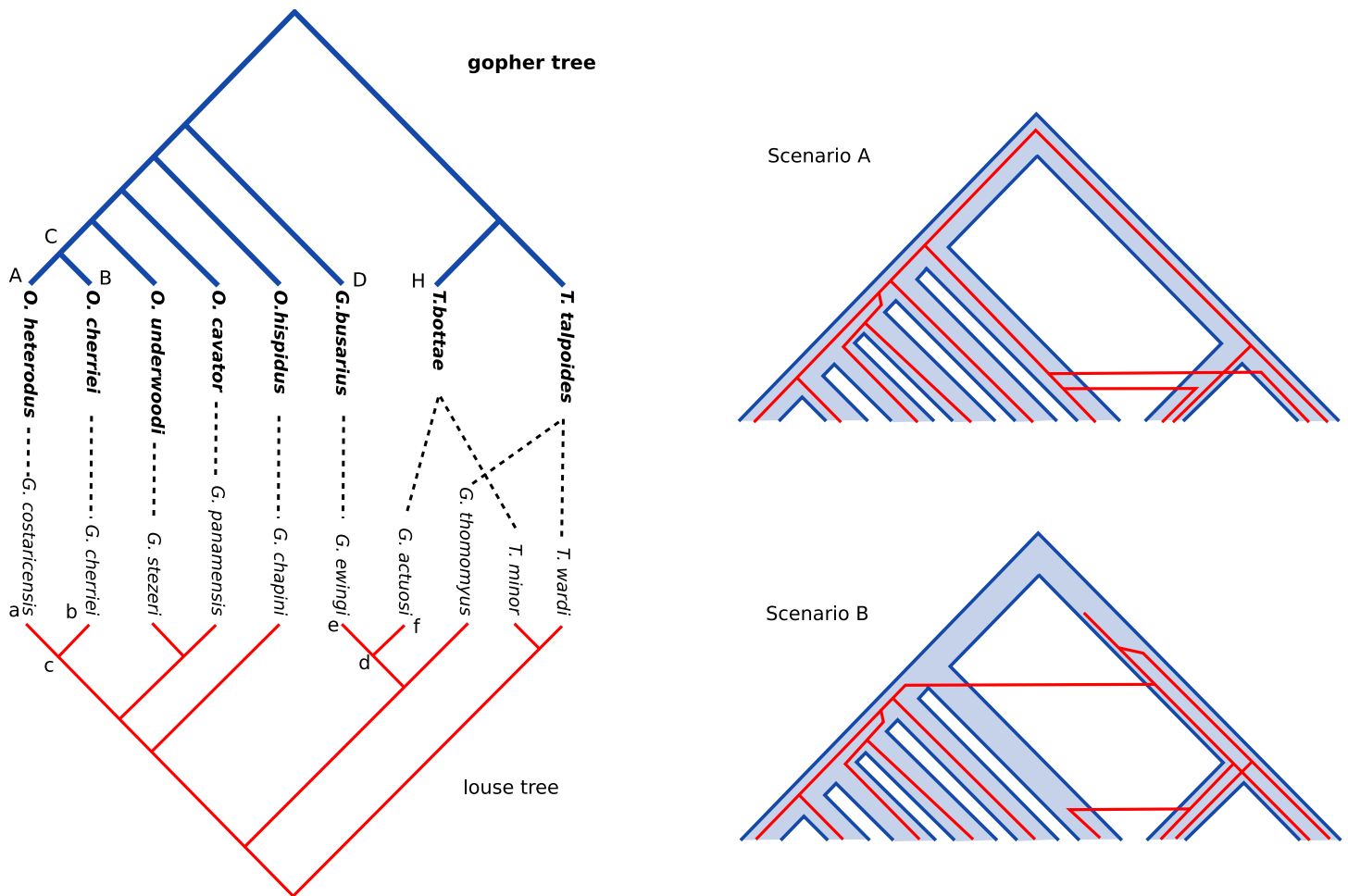
First models for reconciliation, taking explicitly into account the two topologies and using a mechanistic event-based approach, were proposed for host and symbiont and biogeography [22,23]. Debates followed, as the methods were not yet completely sound but integrated useful information in a new framework [24].

Costs for each event and a dynamic programming considering all pairs of host and symbiont nodes were then introduced in a host and symbiont approach, both of which still underlies most of the current reconciliation methods for host and symbiont, and species and genes [25]. Reconciliation returned to the framework it was introduced in, gene and species. After character models were considered for horizontal gene transfer [26], a new reconciliation model, following and improving the dynamic programming approach presented for host and symbiont, effectively introduced horizontal gene transfer to gene and species reconciliation on top of the duplication and loss model [27].

The progressive development of phylogenetic reconciliation was thus possible through exchanges between multiple communities, the host and symbiont, gene and species, and biogeography one. This story and its modern developments have been reviewed several times, generally focusing on specific pairs of levels, with a few exceptions [9,28]. New developments start to bring the different frameworks together with new integrative models.

### Pocket Gophers and their chewing lice: a classic example

Pocket gophers (*Mammalia: Rodentia*) and their chewing lice (*Insecta: Phthyraptera*) is a well studied system of host and symbiont coevolution [29]. The phylogeny of host and symbiont and the matching of their leaves are depicted on the left of Fig 3. Reconciling the two trees consists in giving a scenario with evolutionary events and matching on the ancestral nodes depicting the coevolution of the two trees. The events considered in this system are the events of the DTL model: duplication, transfer (or host switch), loss, and cospeciation, the null event of coevolution. Two scenarios were proposed in two studies [30,31], using two different frameworks which could be deemed as pre-dynamic programming DTL reconciliation. In modern DTL reconciliation frameworks, costs are assigned to events. The two scenarios were then showed to correspond to maximum parsimonious reconciliation with different cost assignments [25]. The scenario A uses 6 cospeciations, 2 duplications, 3 losses and 2 host switches to reconcile the two trees, while scenario B uses 5 cospeciations, 3 duplications, 3 losses and 2 host switches. The cost of a scenario is the sum of the cost of its events. For instance with cost of 0 for cospeciation, 2 for duplication, 1 for loss and 3 for host switch, scenario A has a cost of  $6 \times 0 + 1 \times 2 + 3 \times 1 + 1 \times 3 = 8$  and scenario B of  $5 \times 0 + 1 \times 2 + 3 \times 1 + 2 \times 3 = 11$ , and so according to a parsimonious principle, scenario A would be deemed more likely (scenario A stays more likely as long as the cost of cospeciation is less than the cost of duplication).



**Fig 3. Pocket gophers and chewing lice.** Tanglegrams and two proposed reconciliation scenario for pocket gophers and their chewing lice symbionts. For the host, O. stands for *Orthogeomys*, G. for *Geomys* and T. for *Thomomys*; for the symbiont G. stands for *Geomydoecus* and T. for *Thomoydoecus*.

<https://doi.org/10.1371/journal.pcbi.1010621.g003>

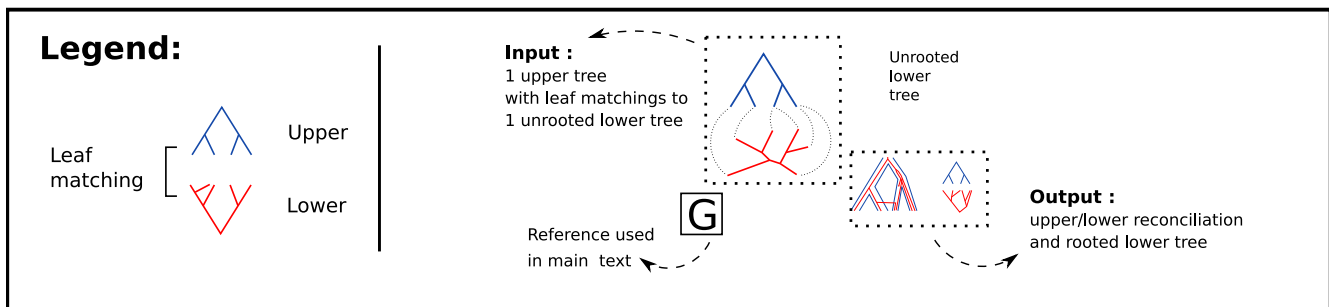
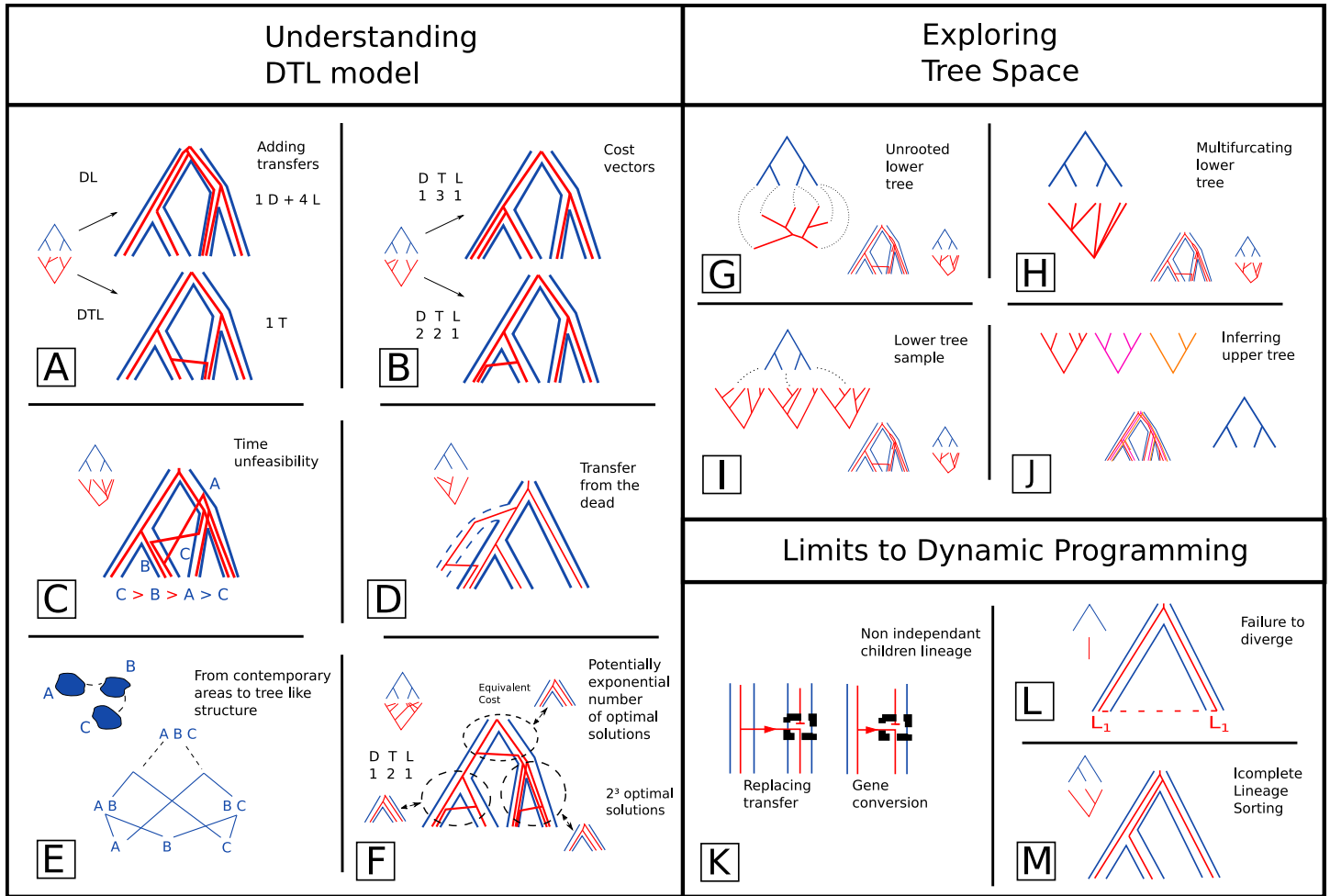
## Development of phylogenetic reconciliation models

Models and methods used today in phylogeny (Fig 4) are the result of several decades of research, made of a progressive complexification, driven by the nature of the data and the quest for biological realism on one side, and the limits and progresses of mathematical and algorithmic methods on the other. See Fig 4 for an illustration of the models and methods presented.

### Pre-reconciliation models: Characters on trees

Character methods can be used when there is no tree available for one of the levels, but only values for a character at the leaves of a phylogenetic tree for the other level. A model defines the events of character value change, their rate, probabilities or costs. For instance the character can be the presence of a host on a symbiont tree [21], the geographical region on a species tree [32], the number of genes on a genome tree [33], or nucleotides in a sequence [34]. Such methods thus aim at reconstructing ancestral characters at internal nodes of the tree [35].

Although these methods have produced results on genome evolution, the utility of a second tree appears with very simple examples. If a symbiont has recently acquired the ability to



**Fig 4. Reconciliation methods and summary.** Illustration of reconciliation events, inputs, outputs, and computational difficulties. This table is intended to serve as illustration to [Development of phylogenetic reconciliation models](#) section and can be read along it. Inputs are on the left of entries, output on the right. Upper trees are drawn in blue, lower trees in red. Adding the horizontal Transfer event add new more parsimonious solutions compared to the previous DL model (A). With this new event, costs must be assigned to D,T and L events, and different costs give different solutions (B). Not all scenarios including transfers are time feasible. Some might include time constraints incompatible with the upper tree (C). Transfer can go from a species to one of its descendant via a sister lineages that went extinct (D). In biogeography, a tree like structure can be constructed to account for the possible migrations between different geographical areas (E). In some cases, an exponential number of scenarios might be most parsimonious, for example when two equivalent patterns have the same cost (F). The lower tree can be unrooted (G), multifurcating (H), or given as a sample of potential trees (I) and reconciliation can be used to resolve those uncertainties to get a binary rooted lower tree. Reconciliation score can also be used to help construct an upper tree (J). The dynamic programming is limited, by the fact it assume independence between sister lineages, that makes it unable to consider replacing transfers or gene conversion (K), as well as Failure to diverge (L) and Incomplete Lineage Sorting (M), two population level events.

<https://doi.org/10.1371/journal.pcbi.1010621.g004>



spread in a group of species and thus it is present in most of them, character methods will wrongly indicate that the common ancestor of the hosts already had the symbiont. In contrast, a comparison of the symbiont and host trees would show discrepancies revealing horizontal transfers.

### The origins of reconciliation: The duplication loss model and the lowest common ancestor mapping

[Duplication](#) and loss were invoked first to explain the presence of multiple copies of a gene in a genome or its absence in certain species [5]. It is possible with those two events to reconcile any two trees [17] *i.e.* to map the nodes and branches of the lower and upper trees, or equivalently to give a list of evolutionary events explaining the discrepancies between the upper tree and lower tree. A most parsimonious Duplication and Loss (DL) reconciliation is computed through the Lowest Common Ancestor (LCA) mapping: proceeding from the leaves to the root, each internal node is mapped to the lowest common ancestor of the mapping of its two children.

### A Markovian model for reconciliation

The LCA mapping in the DL model follows a [parsimony principle](#): no event should be invoked if it is not necessary. However the use of this principle is debated [34] and it is commonly admitted that it is more accurate in molecular evolution to fit a probabilistic model as a [random walk](#), which does not necessarily produce parsimonious scenarios. A [birth and death Markovian model](#) is such a model that can generate a lower tree "inside" a fixed upper one from root to leaves [36]. Statistical inference provides a framework to find most likely scenarios, and in that case, a [maximum likelihood](#) reconciliation of two trees is also a parsimonious one. In addition, it is possible with such a framework to sample scenarios, or integrate over several possible scenarios in order to test different hypotheses, for example to explore the space of lower trees. Moreover probabilistic models can be integrated in larger models as probabilities simply multiply when assuming independence, for instance combining sequence evolution and DL reconciliation [37].

### Introducing horizontal transfer

Host switch, *i.e.* inheritance of a symbiont from a kin lineage, is a crucial event in the evolution of parasitic or symbiotic relationships between species. This horizontal transfer also models migration events in [biogeography](#) and became of interest for the reconciliation of gene and species trees when it appeared that many discrepancies could not simply be explained by duplication and loss and that [horizontal gene transfer](#) (HGT) was a major evolutionary process in micro-organisms evolution. This switching, or horizontal transfer, pattern can also model [admixture](#) or [introgression](#) [38]. It is considered in character methods, without information from the symbiont phylogeny [21,39]. On top of the DL model, horizontal transfer enables new very different reconciliation scenarios ([Fig 4A](#)).

### The simple yet powerful dynamic programming approach

The LCA reconciliation method yields a unique solution, which has been shown to be optimal for the problem of minimizing the weighted number of events, whatever the relative weights of duplication and loss [40]. In contrast, with Duplication, horizontal Transfer and Loss (DTL), there can be several equally parsimonious reconciliations. For instance a succession of duplications and losses can be replaced by a single transfer ([Fig 4B](#)). One of the first ideas to define a

computational problem and approach a resolution was, in a host/symbiont framework, to maximize the number of [co-speciations](#) with a heuristic algorithm [30]. Another solution is to give relative costs to the events and find a scenario that minimizes the sum of the costs of its events [25]. In the probabilistic model frameworks, the equivalent task consists in assigning rates or probabilities to events and search for maximum likelihood scenarios, or sample scenarios according to their likelihood. All these problems are solved with a dynamic programming approach.

This [dynamic programming](#) method consists in traversing the two trees in a [postorder](#). Proceeding from the leaves and then going up in the two trees, for each couple of internal nodes (one for each tree), the cost of a most parsimonious DTL reconciliation is computed [25].

In a parsimony framework, costs of reconciling a lower subtree rooted at  $l$  with an upper subtree rooted at  $U$  is initialized for the leaves with their matching:

$$c(U, l) = 0 \text{ if } l \in U \text{ else } c(U, l) = \infty$$

And then inductively, denoting  $l', l''$  the children of  $l$ ,  $U', U''$  the children of  $U$ ,  $c^S, c^D, c^T, c^L$  the costs associated to speciation, duplication, horizontal transfer and loss, respectively (with  $c^S$  often fixed to 0),

$$\begin{aligned} c(U, l) = & \min(c^S + \min(c(U', l') + c(U'', l''), c(U'', l') + c(U', l'')), \\ & c^S + c^L + \min(c(U', l) + c^L, c(U'', l) + c^L), \\ & c^D + c(U, l') + c(U, l''), \\ & c^T + \min(\min_V(c(V, l')) + c(U, l''), \min_V(c(V, l'')) + c(U, l'))) \end{aligned}$$

The costs  $\min_V(c(V, l'))$  and  $\min_V(c(V, l''))$ , because they do not depend on  $U$ , can be computed once for all  $U$ , hence achieving quadratic [complexity](#) to compute  $c$  for all couples of  $U$  and  $l$ . The cost of losses only appears in association with other events because in parsimony, a loss can always be associated with the preceding event in the tree.

The induction behind the use of dynamic programming is based on always progressing in the trees toward the roots. However some combinations of events that can happen consecutively can make this induction ill-defined. One such combination consists in a transfer followed immediately by a loss in the donor lineage (TL). Restricting the use of this TL event [41] repairs the induction. With an unlimited use it is necessary to use or add other known methods to solve systems of equations like [fixed point methods](#) [42], or numerical solving of differential equations [43]. In 2016, only two out of seven of the most commonly used parsimony reconciliation programs did handle TL events [44] although its consideration can drastically change the result of a reconciliation [12].

Unlike LCA mapping, DTL reconciliation typically yields several scenarios of minimal cost, in some cases an exponential number. The strength of the dynamic programming approach is that it enables to compute a minimum cost of coevolution of the input upper and lower tree in [quadratic time](#) [45], and to get a most parsimonious scenario through [backtracking](#). It can also be transposed to a probabilistic framework to compute the likelihood of coevolution and get a most likely reconciliation, replacing costs with rates, minimums by sums and sums by products [46]. Moreover the approach is suitable, through multiple backtracks, to enumerate all parsimonious solutions or to sample scenarios, optimal and sub-optimal, according to their likelihood.



### Estimation of event costs and rates

Dynamic programming *per se* is only a partial solution and does not solve several problems raised by reconciliation. Defining a most parsimonious DTL reconciliation requires giving costs to the different kind of events (D, T and L). Different cost assignments can yield different reconciliation scenarios (Fig 4B), so there is a need for a way to choose those costs. There is a diversity of approaches to do so. CoRe-PA [47] explores in a recursive manner the space of cost vectors, searching for a good matching with the event frequencies in reconciliations.

ALE [46] uses the same idea in a probabilistic framework to estimate the event rates by maximum likelihood. Alternatively COALA [48] is a pre-process using [approximate bayesian computation](#) with sequential Monte Carlo: simulation and statistic rejection or acceptance of parameters with successive refinement.

In the parsimony framework it is also possible to divide the space of possible event costs in areas of costs which lead to the same [Pareto optimal solution](#) [49]. Pareto optimal reconciliations are such that no other reconciliation has a strictly inferior cost for one type of event (duplication, transfer or loss), and less or equal for the others.

It is also possible to rely on external considerations in order to choose the event costs. For example the software Angst [50] chooses the costs that minimize the variation of genome size, in number of genes, between parent and children species.

### The problem of temporal feasibility

The dynamic programming method works for dated (internal nodes are totally ordered) or undated upper trees. However with undated trees there is a time feasibility issue. Indeed a horizontal transfer implies that the donor and the receiver are contemporary, therefore implying a time constraint on the tree. In consequence two horizontal transfers may be incompatible, because they imply contradicting time constraints (Fig 4C). The dynamic programming can not easily check for such incompatibilities. If the upper tree is undated, finding a time feasible most parsimonious reconciliation is [NP-hard](#) [27,51,52]. It is [fixed parameter tractable](#), which means that there are algorithms running in time bounded by an exponential of the number of transfers in the output scenarios [51].

Some solutions imply [integer linear programming](#) [53] or [branch and bound exploration](#) [9]. If the upper tree is dated, then there is no incompatibility issue because horizontal transfers can be constrained to never go backward in time. Finding a coherent optimal reconciliation is then solved in polynomial time [51], or with a speed-up in RASCAL [54,55], by testing only a fraction of nodes mapping. Most of the software taking undated trees do not look for temporal feasibility, except Jane [56] which explores the space of total orders via a genetic algorithm, or, in a post process, Notung [57] and Eucalypt [58], which search inside the set of optimal solutions for a time consistent ones. Other methods work as supplementary layers to reconciliations, correcting reconciliations [59] or returning a subset of feasible transfers [60], which can be used to date a species tree [60,61].

### Expanding phylogenies: Transfers from the dead

In phylogenetics in general, it is important to keep in mind that the species, extant and ancestral which are represented in any phylogeny are only a sparse sample of the species that currently exist or have existed. This is why one can safely assess that all transfers that can be detected using phylogenetic methods have originated in lineages that are, strictly speaking, absent from a studied phylogeny (Fig 4D) [62]. Accounting for extinct or unsampled biodiversity in phylogenetic studies can give a better understanding of these processes [63]. Originally, DTL reconciliation methods did not recognize this phenomenon and only allowed for transfer

between contemporaneous branches of the tree, hence ignoring most plausible solutions. However methods working on undated upper trees can be seen as implicitly handling the unknown diversity by allowing transfers "to the future" from the point of view of one phylogeny, that is, the donor is more ancient than the recipient. A transfer to the future can be translated into a speciation to unknown species, followed by a transfer from unknown species.

ALE [62] in its dated version explicitly takes the unknown diversity into account by adding a Moran process of speciation/extinctions of species to the dated birth/death model of gene evolution. Transfer from the dead are also handled in a parsimonious setting by Tera and ecce-TERA [44,64], showing that considering these transfers improve the capacity to reconstruct gene trees using reconciliation, and with a more explicit model in [65] and in probabilistic setting, in ALE undated [66].

### The specificity of biogeography: A tree like structure for the "evolution" of areas

In biogeography, some applications of reconciliation approaches consider as an upper tree an area cladogram with defined ancestral nodes. For instance the root can be [Pangea](#) and the nodes contemporary [continents](#). Sometimes internal nodes are not ancestral areas but the unions of the areas of their children, to account for the possibility of species evolving along the lower tree to inhabit one or several areas. In this case, the evolutionary events are migration, where one species colonizes a new area, [allopatric speciation](#), or vicariance, equivalent to co-speciation in host/symbiont comparisons ([Fig 4E](#)). Despite this does not always give a tree (if the unions AB and BC of leaves A, B, C exist, a child can have several parents) and this structure is not associated with time (it is possible for a species to go from A to AB by migration, as well as from AB to A by extinction), reconciliation methods, with events and dynamic programming, can infer evolutionary scenarios between this upper geographical structure and lower species tree. [Diva](#) [67] and [Lagrange](#) [43,68] are two reconciliation models constructing such a tree-like structure and then applying reconciliation, the first with a parsimony principle, the second in a probabilistic framework. Additionally [BioGeoBEARS](#) [69] is a biogeography inference package that reimplemente [DIVA](#) and [Lagrange](#) models and allows for new options, like distance dependent transfers [70] and discussion on statistical model selection [71].

### Graphical output

With two trees and multiple evolutionary events linking them to represent, viewing reconciled trees is a challenging but necessary question in order to make reconciliation studies more accessible. Some reconciliation software include annotation of the evolutionary events on the lower trees [57], while others [47,56,58,72] and specific packages, in [DL](#) [73] or [DTL](#) [74], trace the lower tree embedded in the upper one. One difficulty in this regard is the variety of output format for the different reconciliation software, however recently a common standard, [recphyloxml](#) [75], has been established and endorsed by part of the community with available [viewer](#).

### Addressing additional practical considerations

Applying DTL reconciliation to biological data raises several problems related to uncertainty and confidence levels of input and output. Concerning the output, the uncertainty of the answer calls for an exploration of the whole solution space. Concerning the input, phylogenetic reconciliation has to handle uncertainties in the resolution or rooting of the upper or lower trees, or even to propose roots or resolutions according to their confidence.

## Exploring the space of reconciliations

Multiple DTL reconciliation scenarios can have equal cost or tight probabilities (Fig 4E). Dynamic programming makes it possible to sample reconciliations, uniformly among optimal ones [76] or according to their likelihood. It is also possible to enumerate them in time proportional to the number of solutions [58], a number which can quickly become intractable (even only for optimal ones) (Fig 4F). Finding and presenting structure among the multitude of possible reconciliations has been at the center of recent methodological developments, especially for host and symbiont aimed methods. Several works have focused on representing a set of reconciliations in a compact way, from a uniform sample of optimal ones [76] or by constructing a graph summarizing the optimal solutions [77]. This can be achieved by giving support values to specific events based on all optimal (or suboptimal) reconciliations [78], or with the use of a consensus reconciled tree [79,80]. In a DL model it is possible to define a median reconciliation, based on shared events and to compute it in polynomial time [81].

EMPress [72] can group similar reconciliations through clustering [82], with all pairwise distance between reconciliations computable in polynomial time (independently of the number of most parsimonious reconciliations) [83]. With the same aim, Copybara [84] defines equivalence classes among reconciliations, efficiently computing representative for all classes, and outputs with linear delay a given number of reconciliations (first optimal ones, then sub optimal). The space of most parsimonious reconciliation can be expanded or reduced when increasing or decreasing horizontal transfer allowed distance [58], which is easily done by dynamic programming.

## Inferring phylogenetic trees with reconciliation

**Reconciliation and input uncertainty.** Reconciliation works with two fixed trees, a lower and an upper, both assumed correct and rooted. However, those trees are not first hand data. The most frequently used data for phylogenetics consists in aligned nucleotidic or proteic sequences. Extracting DNA, sequencing, assembling and annotating genomes, recognizing homology relationships among genes and producing multiple alignments for phylogenetic reconstruction are all complex processes where errors can ultimately affect the reconstructed tree [85]. Any topology or rooting error can be misinterpreted and cause systematic bias. For instance, in DL reconciliations, errors on the lower tree bias the reconciliation toward more duplication events closer to the root and more losses closer to the leaves [86].

On the other hand, reconciliation, as a macro evolutionary model, can work as a supplementary layer to the micro evolutionary model of sequence evolution, resolving polytomies (nodes with more than two children) or rooting trees, or be intertwined with it through integrative models in order to get better phylogenies.

Most of the works in this direction focus on gene/species reconciliations, nevertheless some first steps have been made in host/symbiont, such as considering unrooted symbiont trees [87] or dealing with polytomies in Jane [56].

**Exploring the space of lower trees with reconciliation.** Reconciliation can easily take unrooted lower trees as input (Fig 4G), which is a frequently used feature because trees inferred from molecular data are typically unrooted. It is possible to test all possible roots, or a thoughtful triple traversal of the unrooted tree allows to do it without additional time complexity [41]. In a duplication-loss model the set of roots minimizing the costs are found close to one another, forming a "plateau", [88] a property which does not generalizes to DTL [79,87].

Reconciliation can also take as input [non binary trees](#) (Fig 4H), that is, with internal nodes with more than two children. Such trees can be obtained for example by contracting branches with low statistical support. Inferring a binary tree from a non binary tree according to

reconciliation scores is solved in DL with efficient methods [57,89–92]. In DTL, the problem is NP hard [93]. Heuristics [94] and exact fixed parameter tractable algorithms [93] are possible resolutions.

Another way to handle uncertainty in lower trees is to take as input a sample of alternative lower trees instead of a single one. For example in the paper that gave reconciliation its name [17] it was proposed to consider all most likely lower trees, and choose from these trees the best one according to their DL costs, a principle also used by TreeFix-DTL [95].

The sample of lower trees can also reflect their likelihood according to the aligned sequences (Fig 4I), as obtained from bayesian [Markov chain Monte Carlo](#) methods as implemented for example in Phylobayes [96]. AngST [50], ALE [42] and EcceTERA [64] use "amalgamation", a extension of the DTL dynamic programming that is able to efficiently traverse a set of alternative lower trees instead of a single tree.

A local search in the space of lower trees guided by a joint likelihood, on the one hand from multiple sequence alignments and on the other hand from reconciliation with the upper tree, is achieved in Phyldog with a DL model [97] and in GeneRax with DTL [18]. In a DL model with sequence evolution and relaxed [molecular clock](#) the lower tree space is explored with an MCMC in [98]. MowgliNNI [99] can modify the input gene tree at poorly supported nodes to increase DTL score, similarly TreeSolve resolve the multifurcations added by collapsing poorly supported nodes [100].

Finally, integrative models, mixing sequence evolution and reconciliation, can compute a joint likelihood via dynamic programming (for both reconciliation and gene sequences evolution) [42], use Monte Carlo Markov Chain to include molecular clock to estimate branch lengths, in a DL model [36] or with a relaxed molecular clock [98], and in a DTL model [101]. These models have been applied in gene/species frameworks, not yet in host/symbiont or biogeography.

**Inferring upper trees using reconciliation.** Inferring an upper tree from a set of lower trees is a long standing question related to the supertree problem [102]. It is particularly interesting in the case of gene/species reconciliation where many (typically thousands of) gene trees are available from complete genome sequences. Supertree methods attempt to assemble a species tree based on sets of trees which may differ in terms of contemporary species sets and topology, but usually without consideration for the biological process explaining these differences. However some supertree approaches are statistically consistent for the reconstruction of the species tree if the gene trees are simulated under a DL model. This means that if the number of input lower trees generated from the true upper tree via the DL model grows toward infinity, given that there are no additional error, the output upper tree converges almost surely to the true one. This has been shown in the case of a [quartet distance](#) [103], and with a generalized [Robinson Foulds](#) multicopy distance [104], introduced in [105], with better running time but assuming gene trees do not contain bipartitions contradicting the species tree, which seems rare under a DL model.

However, reconciliation can also be used for the inference of upper tree. It is a computationally hard problem: already resolving polytomies in a non binary upper tree with a binary lower one, minimizing a DL reconciliation score, is NP-hard [106]. In particular, reconstructing the species tree giving the best DL cost for several gene trees is called the Gene Duplication problem or more generally Gene Tree parsimony. The problem was seen as a way to detect paralogy to get better species tree reconstruction [107,108]. It is NP-hard, with interesting results on the problem complexity [109,110] (Fig 4J) and the behavior of the model with different input size, structure and ILS presence [111]. Multiple solutions exists, with ILP [112] or heuristics [113,114], and with the possibility of a deep coalescence score [115].

ODTL [46] takes as input gene trees and searches a maximum likelihood species tree according to a DTL model, with a [hill-climbing](#) search. The approach produces a species tree with internal nodes ordered in time ensuring a time compatibility for the scenarios of transfer among lower trees (see paragraph [The problem of temporal feasibility](#)).

Addressing a more general problem, Phyldog [97] searches for the maximum likelihood species tree, gene trees and DL parameters from multiple family alignments via multiple rounds of local search. It thus performs the exploration of both upper and lower trees at the same time. MixTreEM [116] presents a faster solution.

## Limits of the two-level DTL model

### A limit to dynamic programming: non independent evolution of children lineages

The dynamic programming framework, like usual birth and death models, works under the hypothesis of independent evolution of children lineages in the lower tree. However this hypothesis does not hold if the model is complemented with several other documented evolutionary events, such as horizontal transfer with replacement of an homologous gene in the recipient lineage, or [gene conversion](#). Horizontal transfer with replacement is usually modeled by a rearrangement of the upper tree, called Subtree Prune and Regraft (SPR) ([Fig 4K](#) left). Reconciling under SPR is NP-hard, even in dated trees, and fixed parameter tractable regarding the output size [117,118].

Another way to model and infer replacing horizontal transfers is through maximum agreement forest, where branches are cut in the lower and upper trees in order to get two identical (or statistically indistinguishable [119]) upper and lower forests. The problem is NP-hard [120], but several approximations have been proposed [121]. Replacing transfers can be considered on top of the DL model [122]. In the same vein gene conversion can be seen as a "replacing duplication" ([Fig 4K](#) right). In this latter case, a polynomial algorithm which does not use dynamic programming and is an extension of the LCA method, can find all optimal solutions including gene conversions [118].

### Integrating population levels: Failure to diverge and incomplete lineage sorting

In host/symbiont frameworks, a single symbiont species is sometimes associated to several hosts species. This means that while a speciation or diversification has been observed in the host, the populations are indistinguishable in the symbiont. This is handled for example by additional polytomies in the symbiont tree, possibly leading to intractable inference problems, because polytomies need to be resolved. It is also modeled by an additional evolutionary event "failure to diverge" (Jane [56], Amocoala [123]) ([Fig 4L](#)). Failure to diverge can be a way to allow "free" host switch in a population, a flow of symbionts between closely related hosts. Following that vision, host switch allowed only for close hosts is considered in [58]. This idea of horizontal flow between close populations can also be applied to gene/species frameworks, with a definition of species based on a gradient of gene flow between populations [124].

Failure to diverge is one way of introducing population dynamics in reconciliation, a framework mainly adapted to the multi-species level, where populations are supposed to be well differentiated. There are other population phenomena that limit this framework, one of them being [deep coalescence](#) of lineages, leading to [Incomplete Lineage Sorting](#) (ILS), which is not handled by the DTL model [89,125]. The multi species coalescent is a classic model of alleles evolution along a species tree, with birth of alleles and sorting of alleles at speciations, that

takes into account population sizes and naturally encompass ILS [111,126–129]. In a reconciliation context, several attempts have been made in order to account for ILS without the complex integration of a population model. For example, ILS can be seen as a possible evolutionary pattern for the gene tree (Fig 4M). In that case children lineages are not independent of one another, leading to intractability results. ILS alone can be handled with LCA, but ILS + DL reconciliation is NP hard, even without transfers [130].

Notung [89] handles ILS by collapsing short branches of the species tree in polytomies and allowing ILS as a free diversification of gene trees on those polytomies. EcceTERA [131] bounds the maximum size of connected parts of the species tree where ILS can happen, proposing a fixed parameter tractable algorithm in that parameter.

ILS and DL can be considered on an upper *network* instead of tree. This models in particular introgression, with the possibility to estimate model parameters [132].

More integrative reconciliation models accounting for ILS have been proposed including both DL and multispecies coalescent [133], with DLCoal. It is a probabilistic model with a parsimony translation [134], proposing two sequential LCA-type heuristics handled via an intermediate locus tree between gene and species. However outside of the gene/species reconciliation framework ILS seems, for no particular reason, never considered in host/symbiont, nor in biogeography.

## Reconciliation in models with more than two levels

A striking aspect of reconciliation is the common methodology handling different levels of organization: it is used for comparing domain and protein trees, gene and species trees, hosts and symbiont trees, population and geographic trees. However, now that scientists tend to consider that multi-level models of biological functioning bring a novel and game changing view of organisms and their environment [135], the question is how to use reconciliation to bring phylogenetics to this holobiont era (Fig 2).

Coevolution of entities at different scales of organization is at the basis of the holobiont idea: macro-organisms, micro-organisms and their genes all have a different history bound to a common functioning in a single ecosystem. Biological system like the entanglement of host, symbionts and their genes imply functional and evolutionary dependencies between more than two levels.

## Examples of multi-level systems

**Genes coevolving beyond genome boundaries.** The holobiont concept [136] stresses the possibility of genes from different genomes to cooperate and coevolve [137–139]. For instance, certain genes in a symbiont genome may provide a function to its host, like the production of a vital compound absent from available feeding sources. An iconic example is the case for blood-feeding or sap-feeding insects, which often depend on one or several bacterial symbionts to thrive on a resource that is abundant in sugar, but lacks essential amino-acids or vitamins [140]. Another example is the association of *Fabaceae* with nitrogen-fixing bacteria. The compound beneficiary to the host is typically produced by a set of genes encoded in the symbiont genome, which throughout evolution, may be transferred to other symbionts, and/or in and out of the host genome. Reconciliation methods have the potential to reveal evolutionary links between portions of genomes from different species. A search for coevolving genes beyond the boundaries of the genomes in which they are encoded would highlight the basis for the association of organisms in the holobiont.

**Horizontal gene transfer routes depend on multiple levels.** In *intracellular mutualistic symbiont* insect systems, multiple occurrence of horizontal gene transfers have been identified, whether from host to symbiont, symbiont to host or symbiont to symbiont [141].



Transfers of endosymbiont genes involved in nutrition pathways beneficiary to the insect host have been shown to occur preferentially if the donor and recipient lineages share the same host [142–144]. This is also the case in insect with bacterial symbionts providing defensive protein [145] or in obligate leaf nodule bacterial symbionts associated with plants [146]. In the human host, gene transfers has been shown to occur preferentially among symbionts hosted in the same organs [147].

A review on horizontal gene transfers in host/symbiont systems [148] stresses the importance of supporting HGTs with multiple evidence. Notably it is argued that transfers should be considered better supported when involving symbionts sharing a habitat, a geographical area, or a same host. One should however keep in mind that most of the diversity of hosts and symbionts is unknown and that transfers may have occurred in unsampled closely related species, hosts or symbionts.

The idea that gene transfer in symbionts is constrained by the host can also be used to investigate hosts history. For instance, based on phylogeographical studies, it is now accepted that the bacteria *Helicobacter pylori* has been associated with Human populations since the origins of the human species [149,150]. Analysis of the genomes of *Helicobacter pylori* in Europe suggests that they are issued from a recombination between African and Asian *Helicobacter pylori*. This strongly implies early contacts between the corresponding human populations.

Similarly, an analysis of HGTs in coronaviruses from different mammalian species using reconciliation methods has revealed frequent contact between viruses lineages which can be interpreted as frequent host switches [151].

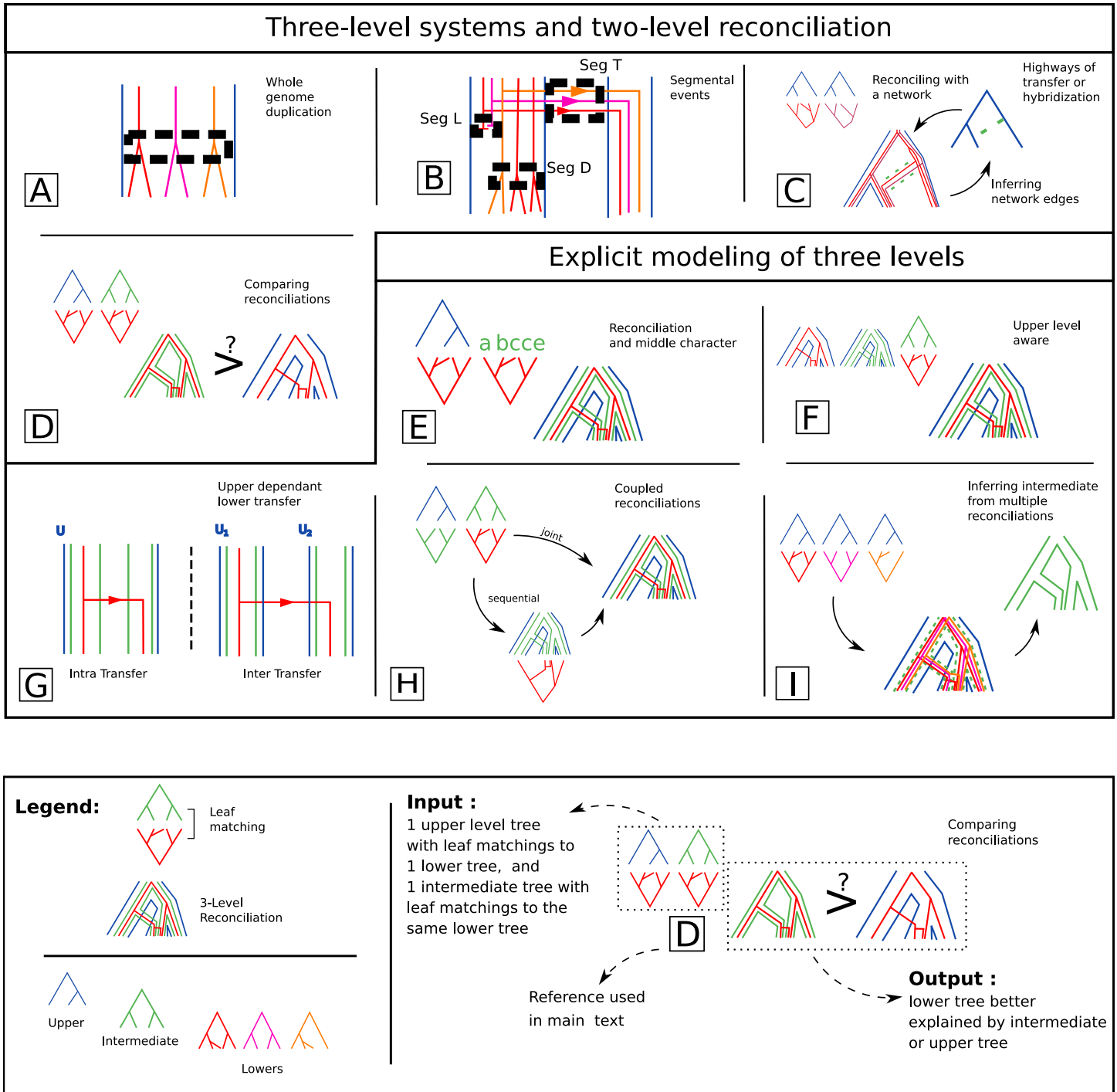
**Cultural evolution.** The evolution of elements of human culture, for instance languages and folktales, in association with human population genetics, has been studied using concepts from phylogenetics. Although reconciliation has never been used in this framework, some of these studies encompass multiple levels of organization, each represented by a tree or the evolution of a character, with a focus on the coevolution of these levels.

Language trees can be compared with population trees in order to reveal vertically transmitted folktales, via a character model on this language tree [152]. Variants in each folktales family, languages, genetic diversity, populations and geography can be compared two by two, to link folktales diversification with languages on one side and with geography on the other side [153]. As in genetics with symbionts sharing host promoting HGTs, linguistic barriers can foreclose the transmission of folktales or language elements [154].

### Investigating three-level systems using two-level reconciliation

Multi level reconciliation is not as developed as two-level reconciliation. One way to approach the evolutionary dependencies between more than two levels of organization is to try to use available standard two-level methods to give a first insight into biological system's complexity.

**Multi-gene events: Implicit consideration of an intermediate level.** At the gene/species tree level, one typically deals with many different gene trees. In this case, the hypothesis that different gene families evolve independently is made implicitly. However this needs not be the case. For instance, duplication, transfer and loss can occur for segments of a genome spanning an arbitrary number of contiguous genes. It is possible to consider such multi-gene events using an intermediate guide for lower trees inside the upper one. For instance one can compute the joint likelihood of multiple gene tree reconciliations with a dated species tree with duplication, loss and whole genome duplication [155] or in a parsimonious setting [107,156–158], and one definition of the problem is NP-hard (Fig 5A). Similarly the DL framework can be enriched with duplication and loss of chromosome segments instead of a single gene (Fig 5B). However DL reconciliation becomes intractable with that new possibility [159].



**Fig 5. Multi-scale reconciliation.** Illustration of input, output and events, of published methods which can be identified with 3-level methods. The formalism is similar to the one on Fig 4. Multiple gene lineages can undergo joint events like whole genome duplication (A) or segmental events (B), some events might be more probable than others, like specific horizontal transfers with highway of transfers or hybridization (C). Cophylogenetic patterns can be compared, to see for instance if the common pattern of a host and a symbiont are not just the common pattern of the symbiont and the geography (D). Characters can evolve on reconciled phylogeny, like gene synteny (E), or two levels can be reconciled with the constraint of an upper one (F). Transfers can be upper dependent, more likely between two intermediate entities that belong to a same upper one (G). Three levels can be reconciled together, sequentially, the intermediate in the upper before adding the lower, or trying to find a joint most parsimonious scenario for the two reconciliations (H). These multi-level models can also be used to reconstruct the intermediate phylogeny (I).

<https://doi.org/10.1371/journal.pcbi.1010621.g005>

The link between two consecutive genes can also be modeled as an evolving character, subject to gain, loss, origination, breakage, duplication and transfer [160]. The evolution of this link appears as an additional level to species and gene trees, partly constrained by the gene/species tree reconciliation, partly evolving on its own, according to genome organization. It thus models the *synteny*, or proximity between genes. At another scale it can as well model the evolution of the belonging of two domains to a protein.

The detection of "highways of transfers", the preferential acquisition of groups of genes from a specific donor, is another example of non-independence of gene histories [161], similarly multi-gene transfers can be detected [162]. It has also led to methodological developments such as reconciliations using phylogenetic networks, seen as a tree augmented with transfers edges, which can be used to constrain transfers in a DTL model [163]. Networks can also be used to model introgression and Incomplete Lineage Sorting [38,164,165] (Fig 5C).

**Detecting coevolution in multiple pairs of levels.** It is a central question to understand the evolution of an holobiont to know what are the levels that coevolve with each others, for instance between host species, host genes, symbionts and symbiont genes. It is possible to approach the multiple inter-dependencies between all levels of evolution by multiple pairwise comparisons of two evolving entities.

Reconciliation of host and symbiont on one side and geography and symbiont on the other side, can also help to identify patterns of diversification of host and symbiont that reflect coevolution on one side, and patterns that can be explained by a common geographical diversification on the other [166–169] (Fig 5D). Similarly, a study used reconciliation methods to differentiate the effect of diet evolution and *phylogenetic inertia* on the composition of mammalian *gut microbiomes*. By reconstructing ancestral diets and microbiome composition onto a mammalian phylogeny, the study revealed that both effects contribute but at different time scales [35].

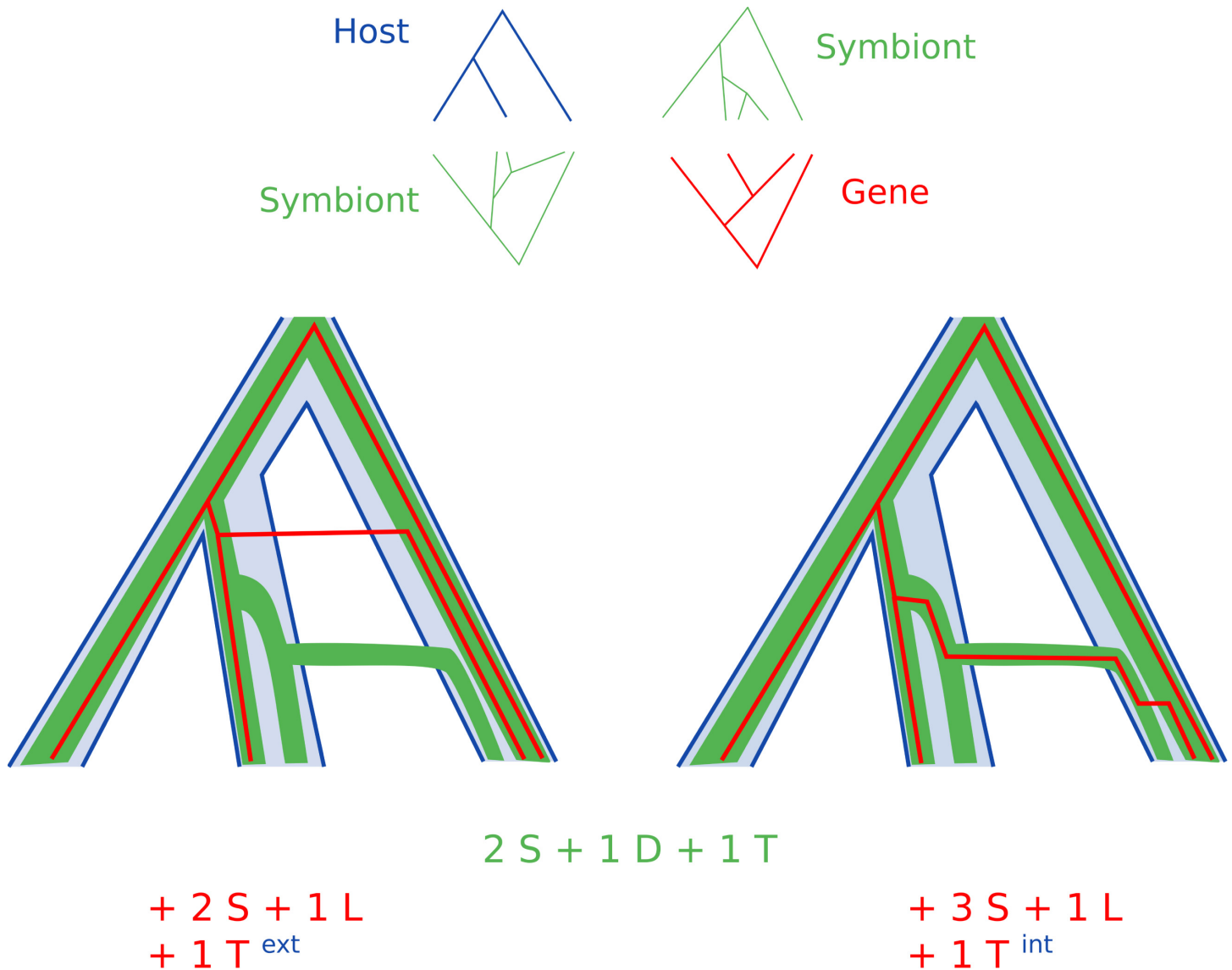
### Explicit modeling of three or more levels

In a model of a multi-level system as host/symbiont/genes, horizontal gene transfers should be more likely between two symbionts of a same host. This is invisible to a two-level gene tree/species tree or host/symbiont reconciliation: in some cases looking at any combination of two levels can lead to miss an evolutionary scenario which can only be the most likely if the information from the three trees are considered together (Fig 6).

Trying to face the limitation of these use of standard two-level reconciliations with systems involving inter-dependencies at multiple levels, a methodological effort has been done in the last decade to construct and use multi-level models. It requires the identification of at least one "intermediate" level between the upper and the lower one.

**Pre-reconciliation: Characters onto reconciled trees.** A first step towards integrated three levels model is to consider phylogenetic trees at two levels and another level represented only with characters at the leaves of one of the trees (Fig 5E). For instance a reconciliation of host and symbiont phylogenies can be informed by geographic data [170]. Ancestral geographic locations of host and symbiont species obtained through a character inference method can then be used to constraint the host/symbiont reconciliation: ancestral hosts and symbionts can only be associated if they belong to the same geographical location (Fig 5F).

At another scale the evolution at the sub-gene level can be approached with a character method [171]. Here, parts of genes (e.g. the sequence coding for protein domains) is reconciled according to a DL model with a species tree, and the genes they belong to are mentioned as characters of these parts. Ancestral genes are then reconstructed a posteriori via merge and splits of gene parts.



**Fig 6. Inter-host and intra-host horizontal gene transfers between symbionts.** Higher level of organization can shed light on lower levels reconciliation. In this example, the goal is to reconstruct the history of a gene present in a symbiont genome. A single transfer and a single loss of gene is the most parsimonious scenario for the reconciliation of the gene tree with either the host or the symbiont tree. Yet when considering the reconciliation of the symbiont and host trees, this scenario implies a gene transfer between two symbionts across branches of the host tree (left). Such an inter-host transfer should be considered unlikely because a series of hidden events are necessary for the gene to come in contact with its next recipient symbiont. Considering the three levels together puts forward a new scenario without inter-host transfer (right) which is slightly less parsimonious in two-level reconciliations, but implies a more likely event of gene transfer within host.

<https://doi.org/10.1371/journal.pcbi.1010621.g006>

**Two-level reconciliations informed by a third level.** As pointed by several studies (see paragraph [Horizontal gene transfer routes depend on multiple levels](#)), an upper level can inform a reconciliation between an intermediate and lower one, notably for horizontal transfers. Three level models can take into account these assumptions to guide reconciliations between an intermediate and lower trees with the knowledge of an upper tree. The model can for example give higher likelihoods to reconciliation scenarios where horizontal gene transfers happen between entities sharing the same habitat. It has been achieved for the first time with DTL gene/species reconciliations nested with a DTL gene domain and gene reconciliation

[125]. Different costs for inter and intra transfers depend on whether or not transfers happen between genes of the same genomes (Fig 5G and 5H).

Note that this model explicitly considers three levels and three trees, but does not yet define a real three level reconciliation, with a likelihood or score associated [125]. It relies on a sequential operation, where the second reconciliation is informed by the result of the first one.

**The reconciliation problem in multi-level models.** The next step is to define the score of a reconciliation consisting of three nested trees and to compute, given the three trees, three-level reconciliations according to their score. It has been achieved with a species/gene/domain system, where genes evolve within the species tree with a DL model and domains evolve within the gene/species system with a DTL model, forbidding domain transfers between genes of two different species (Fig 5G) [172]. Inference involves candidate scenarios with joint scores (Fig 5H joint). Computing the minimum score scenario is NP-hard, but dynamic programming or integer linear programming can offer heuristics [172,173]. Variation of the problem when multiple domains are considered [174] and a simulation framework [175] is available.

Name	Levels of interest	Platform	Command line or Graphical User Interface	Usage	Probability or parsimony	Software License
Diva ( <a href="https://sourceforge.net/projects/diva/">https://sourceforge.net/projects/diva/</a> )	Geography and species	Unix, Mac, Win	Command Line	Reconciliation inference	Parsimony	GNU GPLv2
Lagrange	Geography and species	Linux, Mac, Win		Reconciliation inference	Probability	GNU GPL 2
BioGeoBEARS ( <a href="http://phylo.wikidot.com/biogeobears">http://phylo.wikidot.com/biogeobears</a> )	Geography and species	R package	R package	Reconciliation inference, statistical model test	Probability	GPL-2 GPL-3
Jane ( <a href="https://www.cs.hmc.edu/~hadass/jane/">https://www.cs.hmc.edu/~hadass/jane/</a> )	Host and symbionts	Unix, Mac, Win	GUI or command line	Reconciliation inference, tree uncertainty	Parsimony	Proprietary, registration to download
eMPress ( <a href="https://sites.google.com/g.hmc.edu/empress/home">https://sites.google.com/g.hmc.edu/empress/home</a> )	Host and symbionts	Unix, Mac, Win	GUI or command line	Reconciliation inference, costs estimation, solution space study	Parsimony	GPL 3
Eucalypt ( <a href="https://team.inria.fr/erabelle/en/software/eucalypt/">https://team.inria.fr/erabelle/en/software/eucalypt/</a> )	Host and symbionts		Command line, graphical output with included viewer CophyTrees	Reconciliation inference, solution space study	Parsimony	Cecill
Capybara ( <a href="https://capybara-doc.readthedocs.io/">https://capybara-doc.readthedocs.io/</a> )	Host and symbionts	Linux, Mac, Win, and python package	GUI, python package	Reconciliation inference, solution space study	Parsimony	code available on github ( <a href="https://github.com/Helio-Wang/capybara">https://github.com/Helio-Wang/capybara</a> )
CoRe-PA ( <a href="http://pacosy.informatik.uni-leipzig.de/49-1-CoRe-PA.html">http://pacosy.informatik.uni-leipzig.de/49-1-CoRe-PA.html</a> )	Host and symbiont	Linux, Mac, Win	GUI, Command line, graphical svg output	Reconciliation inference, cost estimation, dated tree, statistical test	Parsimony	-
CoRe-ILP ( <a href="http://pacosy.informatik.uni-leipzig.de/217-0-CoRe-ILP.html">http://pacosy.informatik.uni-leipzig.de/217-0-CoRe-ILP.html</a> )	Host and symbiont	Linux, Mac, Win	Command line	Reconciliation inference, temporal feasibility, dated tree	Parsimony	require IBM ILOG CPLEX Optimizer (academic license can be obtained for free)
Rascal	Host and symbiont	-	-	Reconciliation inference, dated tree	Parsimony	-
Treemap ( <a href="https://sites.google.com/site/cophylogeny/treemap/">https://sites.google.com/site/cophylogeny/treemap/</a> )	Host and symbiont	Java	Command line	Reconciliation inference, statistical test	Parsimony	-
Mowgli ( <a href="http://www.atgc-montpellier.fr/Mowgli/">http://www.atgc-montpellier.fr/Mowgli/</a> )	Species and genes, Host and symbiont	Linux, Mac	Command line, graphical output with compatible viewer [www.sylvx.org Sylvx]	Reconciliation inference, tree uncertainty, ILS, geographical constraints, dated tree input	Parsimony	-
RANGER-DTL ( <a href="https://compbio.engr.uconn.edu/software/ranger-dtl/">https://compbio.engr.uconn.edu/software/ranger-dtl/</a> )	Species and genes	Linux, Mac, Win	Command line	Reconciliation inference, tree uncertainty, solution sampling, replacing transfers	Parsimony	GPL 3
Notung ( <a href="http://www.cs.cmu.edu/~durand/Notung/">http://www.cs.cmu.edu/~durand/Notung/</a> )	Species and genes	Linux, Mac, Win < 7	GUI	Reconciliation inference, tree uncertainty, gene, gene domain, species model	Parsimony	Proprietary
AnGST ( <a href="https://web.mit.edu/almlab/angst.html">https://web.mit.edu/almlab/angst.html</a> )	Species and genes	Python 2	Command line	Reconciliation inference, cost estimation, dated tree input, tree uncertainty	Parsimony	code available on github ( <a href="https://github.com/almlab/angst">https://github.com/almlab/angst</a> )
ecceTERA ( <a href="https://github.com/celinescornavacca/ecceTERA">https://github.com/celinescornavacca/ecceTERA</a> )	Species and genes	Linux, Mac, built from code	Command line, compatible with Sylvx viewer, and recphyloxml output	Reconciliation inference, cost estimation, dated, partially dated or undated species tree input, tree uncertainty, reconciliation space study, species network	Parsimony	Cecill
ALE ( <a href="https://github.com/ssolo/ALE">https://github.com/ssolo/ALE</a> )	Species and genes	Linux, Mac	Command line	Reconciliation inference, cost estimation, dated or undated species tree input, tree uncertainty	Probability	GPL 3
PrIME ( <a href="http://prime.scilifelab.se/">http://prime.scilifelab.se/</a> )	Species and gene	Linux, Mac	Command line, graphical output (PrIMETV)	Reconciliation inference, gene and species tree inference from reconciliation and aligned sequences, orthology analysis	Probability	(Open source)
JPrIME ( <a href="https://github.com/arvestad/jprime">https://github.com/arvestad/jprime</a> )	Gene and species	Java lib	Command line	Reconciliation inference, gene and species tree inference from reconciliation and aligned sequences	Probability	New BSD

**Fig 7. Reconciliation inference software.** Reconciliation software that aim at inferring reconciliation scenarios.

<https://doi.org/10.1371/journal.pcbi.1010621.g007>

Name	Levels of interest	Platform	Command line or Graphical User Interface	Usage	Probability or parsimony	Software License
iGTP ( <a href="https://genome.cs.iastate.edu/igt/home">https://genome.cs.iastate.edu/igt/home</a> )	Species and genes	Linux, Mac, Win	GUI	Gene tree correction in DL or deep coalescence	Parsimony	Source code on request
TreeSolve ( <a href="https://compbio.engr.uconn.edu/software/treesolve/">https://compbio.engr.uconn.edu/software/treesolve/</a> )	Species and genes	Linux, Win	GUI	Gene tree correction in DTL	Parsimony	Source code on request
TreeFix ( <a href="https://www.cs.hmc.edu/~yjw/software/treefix/">https://www.cs.hmc.edu/~yjw/software/treefix/</a> ), TreeFix-DTL ( <a href="https://www.cs.hmc.edu/~yjw/software/treefix-dtl/">https://www.cs.hmc.edu/~yjw/software/treefix-dtl/</a> )	Species and genes	Linux	Command line	Gene tree correction in DL and DTL	Parsimony	GNU GPLv3
Treerecs ( <a href="https://project.inria.fr/treerecs/">https://project.inria.fr/treerecs/</a> )	Species and genes	Linux, Mac, Win	GUI, integrated to Seaview ( <a href="http://dopa.prabi.fr/software/seaview">http://dopa.prabi.fr/software/seaview</a> )	DL tree correction	Parsimony, Probability	GNU Affero GPL Version 3.0-or-later
Phyldog ( <a href="https://pbil.univ-lyon1.fr/software/phyldog/">https://pbil.univ-lyon1.fr/software/phyldog/</a> )	Species and genes	Linux, docker, vm	Command line	Gene and species tree inference from reconciliation and aligned sequences	Probability	Cecill
MixTreEM ( <a href="http://prime.scilifelab.se/mixtreem/index.html">http://prime.scilifelab.se/mixtreem/index.html</a> )	Species and gene	Linux, Mac, Win (build from source)	Command line	Gene and species tree inference from reconciliation and aligned sequences	Probability	(Open source)
GeneRax ( <a href="https://github.com/BenoitMorel/GeneRax">https://github.com/BenoitMorel/GeneRax</a> )	Species and genes	Linux, Mac	Command line, graphical output with recphyloxml and thirdkind	Gene tree inference from reconciliation and aligned sequences, species tree inference	Probability	GNU Affero GPL v3.0
Coala ( <a href="https://team.inria.fr/erable/en/software/coala/">https://team.inria.fr/erable/en/software/coala/</a> )	Host and symbionts	Linux, Mac	Command line	Costs estimation	Parsimony	Cecill
Sylvx ( <a href="http://www.sylvx.org/">http://www.sylvx.org/</a> )	Species and genes, Host and symbiont	Linux, Mac, Win	GUI	Viewer, compatible with Mowgli, ecceTERA	-	-
Thirdkind ( <a href="https://github.com/simonpenel/thirdkind/wiki">https://github.com/simonpenel/thirdkind/wiki</a> )	Species and genes	Linux, Mac, Win	Command line	Viewer, compatible with recphyloxml	-	Cecill
ARTra ( <a href="https://compbio.engr.uconn.edu/software/ARTra/">https://compbio.engr.uconn.edu/software/ARTra/</a> )	Species and genes	Linux, Mac	Command line	Additive and replacing transfers inference	Parsimony	GNU GPL
DLCoal ( <a href="http://compbio.mit.edu/dlcoal/">http://compbio.mit.edu/dlcoal/</a> )	Species and genes	Linux, Mac, Win	Command line	Reconciliation inference with ILS	Parsimony	GNU GPL
SEADOG ( <a href="https://compbio.engr.uconn.edu/software/seadog/">https://compbio.engr.uconn.edu/software/seadog/</a> )	Species and genes and domains	Linux, Mac	Command line	3-level reconciliation inference	Parsimony	GNU GPL version 3

**Fig 8. Auxiliary software.** Reconciliation software which primary goal is not to infer reconciliation scenarios. Most of them are used for tree correction using reconciliation score, some are used for rates inference or graphical visualization of scenarios.

<https://doi.org/10.1371/journal.pcbi.1010621.g008>

**Inferring the intermediate tree using model of 3-level lower/intermediate/upper reconciliation.** Just like two-level reconciliation can be used to improve lower or upper phylogenies, or to help constructing them from aligned sequences, joint reconciliation models can be used in the same manner. In this vein a coupled gene/species DL, domain gene DL and gene sequence evolution model in a bayesian framework improves the reconstruction of gene trees [176] (Fig 5I).

### Software

Multiple software have been developed to implement the various models of reconciliation. Tables from Fig 7 and Fig 8 do not aim for exhaustivity but present a consequent number of software aimed at reconciling trees to infer reconciliation scenarios (Fig 7) or for other usage such as correcting or inferring trees, or testing coevolution (Fig 8). The levels of interest section detail the levels for which the software was implemented, even though it is entirely possible, for instance, to use a software made for species and gene reconciliation to reconcile host and symbionts [177]. Parsimony or probability is the underlying model that is used for the reconciliation.

### Future directions

Reconciliation is now mature as a methodological research subject, a network of researchers and labs working together is emerging, with an active research, a good diversity of available software, and cooperative initiatives like RecPhyloXML, a common standard of output of reconciliations [75]. In the future methodological advances which sustain the development of



new models will certainly play an important part in the possibilities of studies surrounding reconciliations. Notably, new approaches may depart from the dynamic programming solution for DTL which progresses along a rather narrow road: almost each new constraint or event on top of it yields intractability results.

In this article we progressed from two to three embedded trees, and there is potentially an infinity of interacting and coevolving levels to study (see four levels examples in [144,146,152,153,178,179]). Current quantitative methods obviously cannot yet handle such a complexity. In order to compare hypotheses, and assess them in a statistically grounded framework, they are still to be developed and generalized to help the understanding of multi-level evolving systems, including protein domains, genes, protein complexes, micro and macro organisms, and their ecology.

We showed that there have been multiple first steps in the modeling and methods for the embedding of three trees with lower/intermediate and intermediate/upper reconciliations. Methodological efforts could propose new hints for a joint optimization with horizontal transfers for each levels, and moreover offer a probabilistic framework.

Three level reconciliations have only been applied to domain/gene/species combinations while they could handle the classic holobiontic combination gene/symbiont/host. Models could allow the identification of the coevolving entities inside an ecosystem or a holobiont. For example, the parts of a symbiont tree which follow its hosts, while other parts escape this host but follow geography. Or, at another level, the parts of gene trees evolving with symbiont genomes, and the parts evolving with hosts, indicating at which level they are selected.

## References

1. Haeckel EHPA. Systematische phylogenie. Berlin: Georg Reimer; 1894.
2. Nair NU, Lin Y, Manasovska A, Antic J, Grnarova P, Sahu AD, et al. Study of cell differentiation by phylogenetic analysis using histone modification data. *BMC Bioinformatics* 2014 Aug 8; 15(1):269. <https://doi.org/10.1186/1471-2105-15-269> PMID: 25104072
3. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 1990 Jun; 87(12):4576–9. <https://doi.org/10.1073/pnas.87.12.4576> PMID: 2112744
4. Th Dobzhansky, Sturtevant AH. Inversions in the Chromosomes of *Drosophila Pseudoobscura*. *Genetics* 1938 Jan; 23(1):28–64. <https://doi.org/10.1093/genetics/23.1.28> PMID: 17246876
5. Zuckerkandl E, Pauling L. Molecules as documents of evolutionary history. *J Theor Biol* 1965 Mar; 8(2):357–66. [https://doi.org/10.1016/0022-5193\(65\)90083-4](https://doi.org/10.1016/0022-5193(65)90083-4) PMID: 5876245
6. Bagowski CP, Bruins W, Te Velthuis AJW. The nature of protein domain evolution: shaping the interaction network. *Curr Genomics* 2010 Aug; 11(5):368–76. <https://doi.org/10.2174/138920210791616725> PMID: 21286315
7. Gray RD, Bryant D, Greenhill SJ. On the shape and fabric of human history. *Philos Trans R Soc B Biol Sci* 2010 Dec 12; 365(1559):3923–33. <https://doi.org/10.1098/rstb.2010.0162> PMID: 21041216
8. Tehrani JJ. The Phylogeny of Little Red Riding Hood. *PLoS One* 2013 Nov 13; 8(11):e78871. <https://doi.org/10.1371/journal.pone.0078871> PMID: 24236061
9. Wieseke N, Bernt M, Middendorf M. Unifying Parsimonious Tree Reconciliation. *ArXiv13077831 Cs Q-Bio* [Internet]. 2013 Jul 30 [cited 2019 Sep 12]; Available from: <http://arxiv.org/abs/1307.7831>
10. Boussau B, Scornavacca C. Reconciling Gene trees with Species Trees. In: Scornavacca C, Delsuc F, Galtier N, editors. *Phylogenetics in the Genomic Era* [Internet]. No commercial publisher | Authors open access book; 2020 [cited 2020 Aug 20]. p. 3. 2:1–3.2:23. Available from: <https://hal.archives-ouvertes.fr/hal-02535529>
11. Szöllősi GJ, Tannier E, Daubin V, Boussau B. The Inference of Gene Trees with Species Trees. *Syst Biol* 2015 Jan; 64(1):e42–62. <https://doi.org/10.1093/sysbio/syu048> PMID: 25070970
12. Doyon JP, Ranwez V, Daubin V, Berry V. Models, algorithms and programs for phylogeny reconciliation. *Brief Bioinform* 2011 Sep; 12(5):392–400. <https://doi.org/10.1093/bib/bbr045> PMID: 21949266

13. Nakhleh L. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol Evol* 2013 Dec; 28(12):719–28. <https://doi.org/10.1016/j.tree.2013.09.004> PMID: 24094331
14. Charleston MA, Perkins SL. Traversing the tangle: Algorithms and applications for cophylogenetic studies. *J Biomed Inform* 2006 Feb 1; 39(1):62–71. <https://doi.org/10.1016/j.jbi.2005.08.006> PMID: 16226921
15. Charleston M, Libeskind-Hadas R. Event-Based Cophylogenetic Comparative Analysis. In: Garamszegi LZ, editor. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice* [Internet]. Berlin, Heidelberg: Springer; 2014 [cited 2020 Aug 20]. p. 465–80. Available from: [https://doi.org/10.1007/978-3-662-43550-2\\_20](https://doi.org/10.1007/978-3-662-43550-2_20)
16. Martínez-Aquino A. Phylogenetic framework for coevolutionary studies: a compass for exploring jungles of tangled trees. *Curr Zool* 2016 Aug; 62(4):393–403. <https://doi.org/10.1093/cz/zow018> PMID: 29491928
17. Goodman M, Czelusniak J, Moore GW, Romero-Herrera AE, Matsuda G. Fitting the Gene Lineage into its Species Lineage, a Parsimony Strategy Illustrated by Cladograms Constructed from Globin Sequences. *Syst Zool* 1979 Jun; 28(2):132.
18. Morel B, Kozlov AM, Stamatakis A, Szöllösi GJ. GeneRax: A tool for species tree-aware maximum likelihood based gene family tree inference under gene duplication, transfer, and loss [Internet]. *Bioinformatics*; 2019 Sep [cited 2020 Mar 10]. Available from: <http://biorxiv.org/lookup/doi/10.1101/779066>
19. Maddison WP. GENE TREES IN SPECIES TREES. *Syst Biol*. 1997; 14.
20. de Vienne DM. Tanglegrams Are Misleading for Visual Evaluation of Tree Congruence. *Mol Biol Evol* 2019 Jan 1; 36(1):174–6. <https://doi.org/10.1093/molbev/msy196> PMID: 30351416
21. Brooks DR. Hennig's Parasitological Method: A Proposed Solution. *Syst Zool*. 1981; 30 (3):229–249.
22. Nelson G, Platnick N. Systematics and Biogeography: Cladistics and Vicariance. *Syst Zool*. 1981;31.
23. Ronquist F, Nylin S. Process and Pattern in the Evolution of Species Associations. *Syst Biol* 1990 Dec 1; 39(4):323–44.
24. Page RDM. Component Analysis: A Valiant Failure? *Cladistics*. 1990; 6 (2):119–136. <https://doi.org/10.1111/j.1096-0031.1990.tb00532.x> PMID: 34933509
25. Charleston MA. Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Math Biosci* 1998 May; 149(2):191–223. [https://doi.org/10.1016/s0025-5564\(97\)10012-8](https://doi.org/10.1016/s0025-5564(97)10012-8) PMID: 9621683
26. Hein J. A heuristic method to reconstruct the history of sequences subject to recombination.: 10.
27. Hallett MT, Lagergren J. Efficient algorithms for lateral gene transfer problems. In: *Proceedings of the fifth annual international conference on Computational biology—RECOMB '01* [Internet]. Montreal, Quebec, Canada: ACM Press; 2001 [cited 2020 Jun 5]. p. 149–56. Available from: <http://portal.acm.org/citation.cfm?doid=369133.369188>
28. Page RDM. Maps Between Trees and Cladistic Analysis of Historical Associations among Genes, Organisms, and Areas. *Syst Biol* 1994 Mar 1; 43(1):58–77.
29. Hafner MS, Nadler SA. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 1988 Mar 17; 332(6161):258–9. <https://doi.org/10.1038/332258a0> PMID: 3347269
30. Page RDM. Parallel Phylogenies: Reconstructing the History of Host-Parasite Assemblages. *Cladistics*. 1994; 10 (2):155–173.
31. Ronquist F. Reconstructing the history of host-parasite associations using generalised parsimony. *Cladistics* 1995 Mar 1; 11(1):73–89. <https://doi.org/10.1111/j.1096-0031.1995.tb00005.x> PMID: 34920597
32. Wiley EO. Parsimony Analysis and Vicariance Biogeography. *Syst Zool*. 1988; 37 (3):271–290.
33. Csűrös M, Miklós I. Streamlining and Large Ancestral Genomes in Archaea Inferred with a Phylogenetic Birth-and-Death Model. *Mol Biol Evol* 2009 Sep; 26(9):2087–95. <https://doi.org/10.1093/molbev/msp123> PMID: 19570746
34. Felsenstein J. *Inferring Phylogenies*. Oxford, New York: Oxford University Press; 2003. 580 p.
35. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, et al. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun* 2017 Feb 23; 8(1):14319. <https://doi.org/10.1038/ncomms14319> PMID: 28230052
36. Arvestad L, Berglund AC, Lagergren J, Sennblad B. Bayesian gene/species tree reconciliation and orthology analysis using MCMC. *Bioinformatics* 2003 Jul 3; 19(suppl\_1):i7–15. <https://doi.org/10.1093/bioinformatics/btg1000> PMID: 12855432
37. Arvestad L, Berglund AC, Lagergren J, Sennblad B. Gene tree reconstruction and orthology analysis based on an integrated model for duplications and sequence evolution. In: *Proceedings of the eighth annual international conference on Research in computational molecular biology* [Internet]. San

- Diego, California, USA: Association for Computing Machinery; 2004 [cited 2020 Jun 5]. p. 326–35. (RECOMB '04). Available from: <https://doi.org/10.1145/974614.974657>
38. Yu Y, Dong J, Liu KJ, Nakhleh L. Maximum likelihood inference of reticulate evolutionary histories. *Proc Natl Acad Sci* 2014 Nov 18; 111(46):16448–53. <https://doi.org/10.1073/pnas.1407950111> PMID: 25368173
  39. Csűös M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 2010 Aug 1; 26(15):1910–2. <https://doi.org/10.1093/bioinformatics/btq315> PMID: 20551134
  40. Chauve C, El-Mabrouk N. New Perspectives on Gene Family Evolution: Losses in Reconciliation and a Link with Supertrees. In: Batzoglou S, editor. *Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer; 2009. p. 46–58. (Lecture Notes in Computer Science).
  41. Doyon JP, Scornavacca C, Gorbunov KYu, Szöllösi GJ, Ranwez V, Berry V. An Efficient Algorithm for Gene/Species Trees Parsimonious Reconciliation with Losses, Duplications and Transfers. In: Tannier E, editor. *Comparative Genomics* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010 [cited 2020 Feb 19]. p. 93–108. Available from: [http://link.springer.com/10.1007/978-3-642-16181-0\\_9](http://link.springer.com/10.1007/978-3-642-16181-0_9)
  42. Szöllösi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. Efficient Exploration of the Space of Reconciled Gene Trees. *Syst Biol*. 2013; 62(6):901–912. <https://doi.org/10.1093/sysbio/syt054> PMID: 23925510
  43. Ree RH, Smith SA. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Syst Biol* 2008 Feb 1; 57(1):4–14. <https://doi.org/10.1080/10635150701883881> PMID: 18253896
  44. Jacox E, Chauve C, Szöllösi GJ, Ponty Y, Scornavacca C. ecceTERA: comprehensive gene tree-species tree reconciliation using parsimony. *Bioinforma Oxf Engl*. 2016;01; 32(13):2056–8. <https://doi.org/10.1093/bioinformatics/btw105> PMID: 27153713
  45. Bansal MS, Alm EJ, Kellis M. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. *Bioinformatics* 2012 Jun 15; 28(12):i283–91. <https://doi.org/10.1093/bioinformatics/bts225> PMID: 22689773
  46. Szöllösi GJ, Boussau B, Abby SS, Tannier E, Daubin V. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc Natl Acad Sci* 2012 Oct 23; 109(43):17513–8. <https://doi.org/10.1073/pnas.1202997109> PMID: 23043116
  47. Merkle D, Middendorf M, Wieseke N. A parameter-adaptive dynamic programming approach for inferring cophylogenies. *BMC Bioinformatics* 2010 Jan 18; 11(Suppl 1):S60. <https://doi.org/10.1186/1471-2105-11-S1-S60> PMID: 20122236
  48. Baudet C, Donati B, Sinaimer B, Crescenzi P, Gautier C, Matias C, et al. Cophylogeny reconstruction via an approximate Bayesian computation. *Syst Biol* 2015 May; 64(3):416–31. <https://doi.org/10.1093/sysbio/syu129> PMID: 25540454
  49. Libeskind-Hadas R, Wu YC, Bansal MS, Kellis M. Pareto-optimal phylogenetic tree reconciliation. *Bioinformatics* 2014 Jun 15; 30(12):i87–95. <https://doi.org/10.1093/bioinformatics/btu289> PMID: 24932009
  50. David LA, Alm EJ. Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* 2011 Jan; 469(7328):93–6. <https://doi.org/10.1038/nature09649> PMID: 21170026
  51. Tofigh A, Hallett M, Lagergren J. Simultaneous identification of duplications and lateral gene transfers. *IEEE/ACM Trans Comput Biol Bioinform* 2011 Apr; 8(2):517–35. <https://doi.org/10.1109/TCBB.2010.14> PMID: 21233529
  52. Ovadia Y, Fielder D, Conow C, Libeskind-Hadas R. The co phylogeny reconstruction problem is NP-complete. *J Comput Biol J Comput Mol Cell Biol* 2011 Jan; 18(1):59–65. <https://doi.org/10.1089/cmb.2009.0240> PMID: 20715926
  53. Wieseke N, Hartmann T, Bernt M, Middendorf M. Cophylogenetic Reconciliation with ILP. *IEEE/ACM Trans Comput Biol Bioinform* 2015 Nov 1; 12(6):1227–35. <https://doi.org/10.1109/TCBB.2015.2430336> PMID: 26671795
  54. Drinkwater B, Charleston MA. An improved node mapping algorithm for the cophylogeny reconstruction problem. *Coevolution* 2014 Jan 2; 2(1):1–17.
  55. Drinkwater B, Charleston MA. RASCAL: A Randomized Approach for Coevolutionary Analysis. *J Comput Biol* 2016 Mar; 23(3):218–27. <https://doi.org/10.1089/cmb.2015.0111> PMID: 26828619
  56. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms Mol Biol* 2010 Feb 3; 5(1):16. <https://doi.org/10.1186/1748-7188-5-16> PMID: 20181081
  57. Durand D, Halldórsson BV, Vernet B. A Hybrid Micro–Macroevolutionary Approach to Gene Tree Reconstruction. *J Comput Biol* 2006 Mar; 13(2):320–35. <https://doi.org/10.1089/cmb.2006.13.320> PMID: 16597243

58. Donati B, Baudet C, Sinaimeri B, Crescenzi P, Sagot MF. EUCALYPT: efficient tree reconciliation enumerator. *Algorithms Mol Biol* 2015 Jan 23; 10(1):3. <https://doi.org/10.1186/s13015-014-0031-3> PMID: 25648467
59. Ma W, Smirnov D, Libeskind-Hadas R. DTL reconciliation repair. *BMC Bioinformatics* [Internet]. 2017 Mar 14 [cited 2020 Feb 18]; 18(Suppl 3). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5374596/> <https://doi.org/10.1186/s12859-017-1463-9> PMID: 28361686
60. Chauve C, Rafiey A, Davín AA, Scornavacca C, Veber P, Boussau B, et al. MaxTiC: Fast ranking of a phylogenetic tree by Maximum Time Consistency with lateral gene transfers. *bioRxiv* 2017 Oct 6;127548.
61. Davín AA, Tannier E, Williams TA, Boussau B, Daubin V, Szöllösi GJ. Gene transfers can date the tree of life. *Nat Ecol Evol* 2018 May; 2(5):904–9. <https://doi.org/10.1038/s41559-018-0525-3> PMID: 29610471
62. Szöllösi GJ, Tannier E, Lartillot N, Daubin V. Lateral gene transfer from the dead. *Syst Biol* 2013 May 1; 62(3):386–97. <https://doi.org/10.1093/sysbio/syt003> PMID: 23355531
63. Davín AA, Tricou T, Tannier E, de Vienne DM, Szöllösi GJ. Zombi: a phylogenetic simulator of trees, genomes and sequences that accounts for dead lineages. *Bioinformatics* 2020 Feb 15; 36(4):1286–8. <https://doi.org/10.1093/bioinformatics/btz710> PMID: 31566657
64. Scornavacca C, Jacox E, Szöllösi GJ. Joint amalgamation of most parsimonious reconciled gene trees. *Bioinforma Oxf Engl* 2015 Mar 15; 31(6):841–8. <https://doi.org/10.1093/bioinformatics/btu728> PMID: 25380957
65. Weiner S, Bansal MS. Improved Duplication-Transfer-Loss Reconciliation with Extinct and Unsourced Lineages. *Algorithms* 2021 Aug; 14(8):231.
66. Szöllösi GJ, Davín AA, Tannier E, Daubin V, Boussau B. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philos Trans R Soc B Biol Sci* 2015 Sep 26; 370(1678):20140335. <https://doi.org/10.1098/rstb.2014.0335> PMID: 26323765
67. Ronquist F. Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Syst Biol* 1997 Mar 1; 46(1):195–203.
68. Ree RH, Moore BR, Webb CO, Donoghue MJ. A LIKELIHOOD FRAMEWORK FOR INFERRING THE EVOLUTION OF GEOGRAPHIC RANGE ON PHYLOGENETIC TREES. 2005; 13.
69. Matzke NJ. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front Biogeogr* [Internet]. 2013 [cited 2021 Aug 30]; 5(4). Available from: <https://escholarship.org/uc/item/44j7n141>
70. Van Dam MH, Matzke NJ. Evaluating the influence of connectivity and distance on biogeographical patterns in the south-western deserts of North America. *J Biogeogr.* 2016; 43 (8):1514–1532.
71. Matzke NJ. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Syst Biol* 2014 Nov 1; 63(6):951–70. <https://doi.org/10.1093/sysbio/syu056> PMID: 25123369
72. Santichaivekin S, Yang Q, Liu J, Mawhorter R, Jiang J, Wesley T, et al. eMPress: a systematic copylogeny reconciliation tool. *Bioinformatics* [Internet]. 2020 Nov 20 [cited 2021 Jun 7]; (btaa978). Available from: <https://doi.org/10.1093/bioinformatics/btaa978>
73. Sennblad B, Schreil E, Berglund Sonnhammer AC, Lagergren J, Arvestad L. primetv: a viewer for reconciled trees. *BMC Bioinformatics* 2007 May 7; 8:148. <https://doi.org/10.1186/1471-2105-8-148> PMID: 17484781
74. Chevenet F, Doyon JP, Scornavacca C, Jacox E, Jousset E, Berry V. SylvX: a viewer for phylogenetic tree reconciliations. *Bioinformatics* 2016 Feb 15; 32(4):608–10. <https://doi.org/10.1093/bioinformatics/btv625> PMID: 26515823
75. Duchemin W, Gence G, Arigon Chifolleau AM, Arvestad L, Bansal MS, Berry V, et al. RecPhyloXML: a format for reconciled gene trees. *Bioinformatics* 2018 Nov 1; 34(21):3646–52. <https://doi.org/10.1093/bioinformatics/bty389> PMID: 29762653
76. Bansal MS, Alm EJ, Kellis M. Reconciliation Revisited: Handling Multiple Optima when Reconciling with Duplication, Transfer, and Loss. *J Comput Biol* 2013 Oct; 20(10):738–54. <https://doi.org/10.1089/cmb.2013.0073> PMID: 24033262
77. Scornavacca C, Paprotny W, Berry V, Ranwez V. Representing a set of reconciliations in a compact way. *J Bioinforma Comput Biol* 2013 Apr; 11(2):1250025.
78. Nguyen TH, Ranwez V, Berry V, Scornavacca C. Support Measures to Estimate the Reliability of Evolutionary Events Predicted by Reconciliation Methods. *PLoS One* 2013 Oct 4; 8(10):e73667. <https://doi.org/10.1371/journal.pone.0073667> PMID: 24124449

79. Kundu S, Bansal MS. On the impact of uncertain gene tree rooting on duplication-transfer-loss reconciliation. *BMC Bioinformatics* 2018 Aug 13; 19(Suppl 9):290. <https://doi.org/10.1186/s12859-018-2269-0> PMID: 30367593
80. Ma W, Smirnov D, Forman J, Schweickart A, Slocum C, Srinivasan S, et al. DTL-RnB: Algorithms and Tools for Summarizing the Space of DTL Reconciliations. *IEEE/ACM Trans Comput Biol Bioinform* 2018 Apr; 15(2):411–21. <https://doi.org/10.1109/TCBB.2016.2537319> PMID: 26955051
81. Huber KT, Moulton V, Sagot MF, Sinaimeri B. Geometric medians in reconciliation spaces of phylogenetic trees. *Inf Process Lett* 2018 Aug 1; 136:96–101.
82. Mawhorter R, Libeskind-Hadas R. Hierarchical clustering of maximum parsimony reconciliations. *BMC Bioinformatics* 2019 Nov 27; 20(1):612. <https://doi.org/10.1186/s12859-019-3223-5> PMID: 31775628
83. Santichaivekin S, Mawhorter R, Libeskind-Hadas R. An efficient exact algorithm for computing all pairwise distances between reconciliations in the duplication-transfer-loss model. *BMC Bioinformatics* [Internet] 2019 Dec 17 [cited 2020 Sep 14]; 20(Suppl 20). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6915856/> <https://doi.org/10.1186/s12859-019-3203-9> PMID: 31842734
84. Wang Y, Mary A, Sagot MF, Sinaimeri B. Capybara: equivalence CIAss enumeration of coPhylogenY event-BASed ReconciliAtions. *Bioinformatics* 2020 Aug 15; 36(14):4197–9. <https://doi.org/10.1093/bioinformatics/btaa498> PMID: 32556075
85. Boussau B, Daubin V. Genomes as documents of evolutionary history. *Trends Ecol Evol* 2010 Apr; 25(4):224–32. <https://doi.org/10.1016/j.tree.2009.09.007> PMID: 19880211
86. Hahn MW. Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. *Genome Biol* 2007 Jul 16; 8(7):R141. <https://doi.org/10.1186/gb-2007-8-7-r141> PMID: 17634151
87. Urbini L, Sinaimeri B, Matias C, Sagot MF. Exploring the Robustness of the Parsimonious Reconciliation Method in Host-Symbiont Cophylogeny. *IEEE/ACM Trans Comput Biol Bioinform* 2019 May 1; 16(3):738–48.
88. Górecki P, Eulenstein O, Tiuryn J. Unrooted Tree Reconciliation: A Unified Approach. *IEEEACM Trans Comput Biol Bioinforma IEEE ACM* 2013 Mar 19; 10:522–36. <https://doi.org/10.1109/TCBB.2013.22> PMID: 23929875
89. Stolzer M, Lai H, Xu M, Sathaye D, Vernet B, Durand D. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 2012 Sep 15; 28(18):i409–15. <https://doi.org/10.1093/bioinformatics/bts386> PMID: 22962460
90. Lafond M, Noutahi E, El-Mabrouk N. Efficient Non-Binary Gene Tree Resolution with Weighted Reconciliation Cost. 2016;12.
91. Comte N, Morel B, Hasić D, Guéguen L, Boussau B, Daubin V, et al. Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics* 2020 Sep 15; 36(18):4822–4. <https://doi.org/10.1093/bioinformatics/btaa615> PMID: 33085745
92. Zheng Y, Zhang L. Reconciliation with Non-binary Gene Trees Revisited. In: Sharan R, editor. *Research in Computational Molecular Biology*. Cham: Springer International Publishing; 2014. p. 418–32. (Lecture Notes in Computer Science).
93. Kordi M, Bansal MS. On the Complexity of Duplication-Transfer-Loss Reconciliation with Non-binary Gene Trees. In: Harrison R, Li Y, Măndoiu I, editors. *Bioinformatics Research and Applications*. Cham: Springer International Publishing; 2015. p. 187–98. (Lecture Notes in Computer Science).
94. Lai H, Stolzer M, Durand D. Fast Heuristics for Resolving Weakly Supported Branches Using Duplication, Transfers, and Losses. In: Meidanis J, Nakhleh L, editors. *Comparative Genomics*. Cham: Springer International Publishing; 2017. p. 298–320. (Lecture Notes in Computer Science).
95. Bansal MS, Wu YC, Alm EJ, Kellis M. Improved gene tree error correction in the presence of horizontal gene transfer. *Bioinformatics* 2015 Apr 15; 31(8):1211–8. <https://doi.org/10.1093/bioinformatics/btu806> PMID: 25481006
96. Lartillot N, Philippe H. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Mol Biol Evol* 2004 Jun; 21(6):1095–109. <https://doi.org/10.1093/molbev/msh112> PMID: 15014145
97. Boussau B, Szölösi GJ, Duret L, Gouy M, Tannier E, Daubin V. Genome-scale coestimation of species and gene trees. *Genome Res* 2013 Feb; 23(2):323–30. <https://doi.org/10.1101/gr.141978.112> PMID: 23132911
98. Åkerborg Ö, Sennblad B, Arvestad L, Lagergren J. Simultaneous Bayesian gene tree reconstruction and reconciliation analysis. *Proc Natl Acad Sci* 2009 Apr 7; 106(14):5714–9. <https://doi.org/10.1073/pnas.0806251106> PMID: 19299507



99. Nguyen TH, Ranwez V, Pointet S, Chifolleau AMA, Doyon JP, Berry V. Reconciliation and local gene tree rearrangement can be of mutual profit. *Algorithms Mol Biol* 2013 Apr 8; 8(1):12. <https://doi.org/10.1186/1748-7188-8-12> PMID: 23566548
100. Kordi M, Bansal MS. TreeSolve: Rapid Error-Correction of Microbial Gene Trees. In: Martín-Vide C, Vega-Rodríguez MA, Wheeler T, editors. *Algorithms for Computational Biology*. Cham: Springer International Publishing; 2020. p. 125–39. (Lecture Notes in Computer Science).
101. Sjöstrand J, Tofigh A, Daubin V, Arvestad L, Sennblad B, Lagergren J. A Bayesian Method for Analyzing Lateral Gene Transfer. *Syst Biol* 2014 May 1; 63(3):409–20. <https://doi.org/10.1093/sysbio/syu007> PMID: 24562812
102. Warnow T. Supertree Construction: Opportunities and Challenges. ArXiv180503530 Q-Bio [Internet]. 2018 May 9 [cited 2020 Aug 27]; Available from: <http://arxiv.org/abs/1805.03530>
103. Legried B, Molloy EK, Warnow T, Roch S. Polynomial-Time Statistical Estimation of Species Trees under Gene Duplication and Loss. *bioRxiv* 2020 Jan 21;821439. <https://doi.org/10.1089/cmb.2020.0424> PMID: 33325781
104. Molloy EK, Warnow T. FastMulRFS: Fast and accurate species tree estimation under generic gene duplication and loss models. *bioRxiv* 2020 May 7;835553. <https://doi.org/10.1093/bioinformatics/btaa444> PMID: 32657396
105. Chaudhary R, Burleigh JG, Fernández-Baca D. Inferring species trees from incongruent multi-copy gene trees using the Robinson-Foulds distance. *Algorithms Mol Biol* 2013 Nov 1; 8(1):28. <https://doi.org/10.1186/1748-7188-8-28> PMID: 24180377
106. Zheng Y, Wu T, Zhang L. Reconciliation of Gene and Species Trees With Polytomies. ArXiv12013995 Q-Bio [Internet]. 2012 May 2 [cited 2020 Jun 17]; Available from: <http://arxiv.org/abs/1201.3995>
107. Guigo R, Muchnik I, Smith TF. Reconstruction of Ancient Molecular Phylogeny. *Mol Phylogenet Evol* 1996 Oct 1; 6(2):189–213. <https://doi.org/10.1006/mpev.1996.0071> PMID: 8899723
108. Page RDM. Extracting Species Trees From Complex Gene Trees: Reconciled Trees And Vertebrate Phylogeny. *Mol Phylogenet Evol* 2000 Jan 1; 14(1):89–106. <https://doi.org/10.1006/mpev.1999.0676> PMID: 10631044
109. Ma B, Li M, Zhang L. From Gene Trees to Species Trees. *SIAM J Comput* 2000 Jan 1; 30(3):729–52.
110. Bansal MS, Shamir R. A Note on the Fixed Parameter Tractability of the Gene-Duplication Problem. *IEEE/ACM Trans Comput Biol Bioinform* 2011 May; 8(3):848–50. <https://doi.org/10.1109/TCBB.2010.74> PMID: 20733245
111. Maddison WP, Knowles LL. Inferring phylogeny despite incomplete lineage sorting. *Syst Biol* 2006 Feb; 55(1):21–30. <https://doi.org/10.1080/10635150500354928> PMID: 16507521
112. Chang WC, Burleigh GJ, Fernández-Baca DF, Eulenstein O. An ILP solution for the gene duplication problem. *BMC Bioinformatics* 2011 Feb 15; 12(1):S14. <https://doi.org/10.1186/1471-2105-12-S1-S14> PMID: 21342543
113. Page RD. GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics* 1998 Jan 1; 14(9):819–20. <https://doi.org/10.1093/bioinformatics/14.9.819> PMID: 9918954
114. Wehe A, Bansal MS, Burleigh JG, Eulenstein O. DupTree: a program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics* 2008 Jul 1; 24(13):1540–1. <https://doi.org/10.1093/bioinformatics/btn230> PMID: 18474508
115. Chaudhary R, Bansal MS, Wehe A, Fernández-Baca D, Eulenstein O. iGTP: A software package for large-scale gene tree parsimony analysis. *BMC Bioinformatics* 2010 Nov 23; 11(1):574. <https://doi.org/10.1186/1471-2105-11-574> PMID: 21092314
116. Ullah I, Parviainen P, Lagergren J. Species Tree Inference Using a Mixture Model. *Mol Biol Evol* 2015 Sep 1; 32(9):2469–82. <https://doi.org/10.1093/molbev/msv115> PMID: 25963975
117. Bordewich M, Semple C. On the Computational Complexity of the Rooted Subtree Prune and Regraft Distance. *Ann Comb* 2005 Jan 1; 8:409–23.
118. Hasic D, Tannier E. Gene tree reconciliation including transfers with replacement is hard and FPT. *J Comb Optim* 2019 Aug; 38(2):502–44.
119. Abby SS, Tannier E, Gouy M, Daubin V. Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests. *BMC Bioinformatics* 2010 Jun 15; 11:324. <https://doi.org/10.1186/1471-2105-11-324> PMID: 20550700
120. Hein J, Jiang T, Wang L, Zhang K. On the complexity of comparing evolutionary trees. *Discret Appl Math* 1996 Dec 5; 71(1):153–69.
121. Rodrigues EM, Sagot MF, Wakabayashi Y. The maximum agreement forest problem: Approximation algorithms and computational experiments. *Theor Comput Sci* 2007 Apr 20; 374(1):91–110.



122. Kordi M, Kundu S, Bansal MS. On Inferring Additive and Replacing Horizontal Gene Transfers Through Phylogenetic Reconciliation. In: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics [Internet]. Niagara Falls, NY, USA: Association for Computing Machinery; 2019 [cited 2020 Feb 18]. p. 514–23. (BCB '19). Available from: <https://doi.org/10.1145/3307339.3342168>
123. Urbini L. Models and algorithms to study the common evolutionary history of hosts and symbionts [Internet] [phdthesis]. Université de Lyon; 2017 [cited 2020 May 29]. Available from: <https://tel.archives-ouvertes.fr/tel-01673445>
124. Marin J, Achaz G, Crombach A, Lambert A. The genomic view of diversification. *J Evol Biol.* 2020; 33(10):1387–1404. <https://doi.org/10.1111/jeb.13677> PMID: 32654283
125. Stolzer M, Siewert K, Lai H, Xu M, Durand D. Event inference in multidomain families with phylogenetic reconciliation. *BMC Bioinformatics* 2015 Oct 2; 16(14):S8. <https://doi.org/10.1186/1471-2105-16-S14-S8> PMID: 26451642
126. Rannala B, Yang Z. Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci. *Genetics* 2003 Aug 1; 164(4):1645–56. <https://doi.org/10.1093/genetics/164.4.1645> PMID: 12930768
127. Degnan JH, Salter LA. Gene Tree Distributions Under the Coalescent Process. *Evolution.* 2005; 59(1):24–37. PMID: 15792224
128. Liu L, Pearl DK. Species Trees from Gene Trees: Reconstructing Bayesian Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions. *Syst Biol* 2007 Jun 1; 56(3):504–14. <https://doi.org/10.1080/10635150701429982> PMID: 17562474
129. Rannala B, Edwards SV, Leaché A, Yang Z. Chapter 3.3 The Multi-species Coalescent Model and Species Tree Inference.: 21.
130. Bork D, Cheng R, Wang J, Sung J, Libeskind-Hadas R. On the computational complexity of the maximum parsimony reconciliation problem in the duplication-loss-coalescence model. *Algorithms Mol Biol* 2017 Mar 14; 12(1):6. <https://doi.org/10.1186/s13015-017-0098-8> PMID: 28316640
131. Chan Y, ban, Ranwez V, Scornavacca C. Inferring incomplete lineage sorting, duplications, transfers and losses with reconciliations. *J Theor Biol.* 2017; 432:1–13. <https://doi.org/10.1016/j.jtbi.2017.08.008> PMID: 28801222
132. Du P, Ogilvie HA, Nakhleh L. Unifying Gene Duplication, Loss, and Coalescence on Phylogenetic Networks. In: Cai Z, Skums P, Li M, editors. *Bioinformatics Research and Applications*. Cham: Springer International Publishing; 2019. p. 40–51. (Lecture Notes in Computer Science).
133. Rasmussen MD, Kellis M. Unified modeling of gene duplication, loss, and coalescence using a locus tree. *Genome Res* 2012 Apr 1; 22(4):755–65. <https://doi.org/10.1101/gr.123901.111> PMID: 22271778
134. Wu YC, Rasmussen MD, Bansal MS, Kellis M. Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome Res* 2013 Dec 5;gr.161968.113.
135. Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, et al. Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. *mSystems* 2016 Mar 29; 1(2):e00028–16. <https://doi.org/10.1128/mSystems.00028-16> PMID: 27822520
136. Margulis L, Fester R, editors. *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*. Cambridge, MA, USA: MIT Press; 1991. 470 p.
137. Bordenstein SR, Theis KR. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol* 2015 Aug; 13(8):e1002226. <https://doi.org/10.1371/journal.pbio.1002226> PMID: 26284777
138. Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I. The role of microorganisms in coral health, disease and evolution. *Nat Rev Microbiol* 2007 May; 5(5):355–62. <https://doi.org/10.1038/nrmicro1635> PMID: 17384666
139. Zilber-Rosenberg I, Rosenberg E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* 2008 Aug; 32(5):723–35. <https://doi.org/10.1111/j.1574-6976.2008.00123.x> PMID: 18549407
140. Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet.* 2008; 42:165–190. <https://doi.org/10.1146/annurev.genet.41.110306.130119> PMID: 18983256
141. López-Madrigal S, Gil R. Et tu, Brute? Not Even Intracellular Mutualistic Symbionts Escape Horizontal Gene Transfer. *Genes* [Internet] 2017 Sep 29 [cited 2019 Nov 5]; 8(10). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5664097/>

142. Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Müller A, Woyke T, et al. Comparative Genomics Suggests an Independent Origin of Cytoplasmic Incompatibility in *Cardinium hertigii*. *PLoS Genet* 2012 Oct 25; 8(10):e1003012. <https://doi.org/10.1371/journal.pgen.1003012> PMID: 23133394
143. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proc Natl Acad Sci U S A* 2014 Jul 15; 111(28):10257–62. <https://doi.org/10.1073/pnas.1409284111> PMID: 24982177
144. Manzano-Marín A, D’acier AC, Clamens AL, Orvain C, Cruaud C, Barbe V, et al. Serial horizontal transfer of vitamin-biosynthetic genes enables the establishment of new nutritional symbionts in aphids’ di-symbiotic systems. *ISME J* 2019 Oct 17; 1–15.
145. Nakabachi A, Ueoka R, Oshima K, Teta R, Mangoni A, Gurgui M, et al. Defensive Bacteriome Symbiont with a Drastically Reduced Genome. *Curr Biol* 2013 Aug 5; 23(15):1478–84. <https://doi.org/10.1016/j.cub.2013.06.027> PMID: 23850282
146. Pinto-Carbó M, Sieber S, Dessein S, Wicker T, Verstraete B, Gademann K, et al. Evidence of horizontal gene transfer between obligate leaf nodule symbionts. *ISME J* 2016 Sep; 10(9):2092–105. <https://doi.org/10.1038/ismej.2016.27> PMID: 26978165
147. Jeong H, Arif B, Caetano-Anollés G, Kim KM, Nasir A. Horizontal gene transfer in human-associated microorganisms inferred by phylogenetic reconstruction and reconciliation. *Sci Rep* 2019 Apr 11; 9(1):1–18.
148. Wijayawardena BK, Minchella DJ, DeWoody JA. Hosts, parasites, and horizontal gene transfer. *Trends Parasitol* 2013 Jul 1; 29(7):329–38. <https://doi.org/10.1016/j.pt.2013.05.001> PMID: 23759418
149. Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, et al. Age of the Association between *Helicobacter pylori* and Man. *PLoS Pathog* 2012 May 10; 8(5):e1002693. <https://doi.org/10.1371/journal.ppat.1002693> PMID: 22589724
150. Achtman M. How old are bacterial pathogens? *Proc R Soc B Biol Sci*. 2016 Aug 17; 283(1836):20160990. <https://doi.org/10.1098/rspb.2016.0990> PMID: 27534956
151. Fu Y, Pistozzi M, Yang X, Lin Z. A Comprehensive Classification of Coronaviruses and Inferred Cross-Host Transmissions [Internet]. *Bioinformatics*; 2020 Aug [cited 2021 Mar 19]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.08.11.232520>
152. da Silva SG, Tehrani JJ. Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *R Soc Open Sci*. 2015; 3(1):150645.
153. Ross RM, Greenhill SJ, Atkinson QD. Population structure and cultural geography of a folktale in Europe. *Proc R Soc B Biol Sci* 2013 Apr 7; 280(1756):20123065. <https://doi.org/10.1098/rspb.2012.3065> PMID: 23390109
154. Bortolini E, Pagani L, Crema ER, Sarno S, Barbieri C, Boattini A, et al. Inferring patterns of folktale diffusion using genomic data. *Proc Natl Acad Sci* 2017 Aug 22; 114(34):9140–5. <https://doi.org/10.1073/pnas.1614395114> PMID: 28784786
155. Zwaenepoel A, Van de Peer Y. Ancient whole genome duplications and the evolution of the gene duplication and loss rate [Internet]. *Evol Biol*; 2019 Feb [cited 2020 Feb 19]. Available from: <http://biorxiv.org/lookup/doi/10.1101/556076>
156. Page RDM, Cotton JA. Vertebrate phylogenomics: reconciled trees and gene duplications. In: *Biocomputing 2002* [Internet]. WORLD SCIENTIFIC; 2001 [cited 2022 Mar 10]. p. 536–47. Available from: [https://www.worldscientific.com/doi/abs/10.1142/9789812799623\\_0050](https://www.worldscientific.com/doi/abs/10.1142/9789812799623_0050)
157. Burleigh Jg., Bansal Ms., Wehe A, Eulenstein O. Locating Large-Scale Gene Duplication Events through Reconciled Trees: Implications for Identifying Ancient Polyploidy Events in Plants. *J Comput Biol* 2009 Aug; 16(8):1071–83. <https://doi.org/10.1089/cmb.2009.0139> PMID: 19689214
158. Bansal MS, Eulenstein O. The multiple gene duplication problem revisited. *Bioinformatics* 2008 Jul 1; 24(13):i132–8. <https://doi.org/10.1093/bioinformatics/btn150> PMID: 18586705
159. Dondi R, Lafond M, Scornavacca C. Reconciling multiple genes trees via segmental duplications and losses. *Algorithms Mol Biol* 2019 Mar 20; 14(1):7. <https://doi.org/10.1186/s13015-019-0139-6> PMID: 30930955
160. Duchemin W, Anselmetti Y, Patterson M, Ponty Y, Bérard S, Chauve C, et al. DeCoSTAR: Reconstructing the Ancestral Organization of Genes or Genomes Using Reconciled Phylogenies. *Genome Biol Evol* 2017 May 1; 9(5):1312–9. <https://doi.org/10.1093/gbe/evx069> PMID: 28402423
161. Bansal MS, Banay G, Gogarten JP, Shamir R. Detecting Highways of Horizontal Gene Transfer. *J Comput Biol* 2011 Sep 1; 18(9):1087–114. <https://doi.org/10.1089/cmb.2011.0066> PMID: 21899418
162. Kloub L, Gosselin S, Fullmer M, Graf J, Gogarten JP, Bansal MS. Systematic Detection of Large-Scale Multigene Horizontal Transfer in Prokaryotes. *Mol Biol Evol* 2021 Jun 1; 38(6):2639–59. <https://doi.org/10.1093/molbev/msab043> PMID: 33565580

163. Scornavacca C, Mayol JCP, Cardona G. Fast algorithm for the reconciliation of gene trees and LGT networks. *J Theor Biol* 2017 Apr 7; 418:129–37. <https://doi.org/10.1016/j.jtbi.2017.01.024> PMID: 28111320
164. Yu Y, Nakhleh L. Fast Algorithms for Reconciliation under Hybridization and Incomplete Lineage Sorting. *ArXiv12121909 Cs Q-Bio* [Internet]. 2012 Dec 9 [cited 2019 Oct 23]; Available from: <http://arxiv.org/abs/1212.1909>
165. Yu Y, Barnett RM, Nakhleh L. Parsimonious Inference of Hybridization in the Presence of Incomplete Lineage Sorting. *Syst Biol* 2013 Sep 1; 62(5):738–51. <https://doi.org/10.1093/sysbio/syt037> PMID: 23736104
166. Nieberding C, Jousselin E, Desdevises Y. The use of co-phylogeographic patterns to predict the nature of host–parasite interactions, and vice versa. In: *The biogeography of host-parasite interactions*. 2010. p. 59–69.
167. Martínez-Aquino A, Ceccarelli FS, Eguiarte LE, Vázquez-Domínguez E, León GPP de. Do the Historical Biogeography and Evolutionary History of the Digenean *Margotrema* spp. across Central Mexico Mirror Those of Their Freshwater Fish Hosts (Goodeinae)? *PLoS One* 2014 Jul 7; 9(7):e101700. <https://doi.org/10.1371/journal.pone.0101700> PMID: 24999998
168. Weckstein JD. Biogeography Explains Cophylogenetic Patterns in Toucan Chewing Lice. *Syst Biol* 2004 Feb 1; 53(1):154–64. <https://doi.org/10.1080/10635150490265085> PMID: 14965910
169. Fountain ED, Pauli JN, Mendoza JE, Carlson J, Peery MZ. Cophylogenetics and biogeography reveal a coevolved relationship between sloths and their symbiont algae. *Mol Phylogenet Evol* 2017 May; 110:73–80. <https://doi.org/10.1016/j.ympev.2017.03.003> PMID: 28288943
170. Berry V, Chevenet F, Doyon JP, Jousselin E. A geography-aware reconciliation method to investigate diversification patterns in host/parasite interactions. *Mol Ecol Resour* 2018 Sep 1; 18(5):1173–84. <https://doi.org/10.1111/1755-0998.12897> PMID: 29697894
171. Wu YC, Rasmussen MD, Kellis M. Evolution at the subgene level: domain rearrangements in the *Drosophila* phylogeny. *Mol Biol Evol* 2012 Feb; 29(2):689–705. <https://doi.org/10.1093/molbev/msr222> PMID: 21900599
172. Li L, Bansal MS. An Integrated Reconciliation Framework for Domain, Gene, and Species Level Evolution. *IEEE/ACM Trans Comput Biol Bioinform* 2019 Jan; 16(1):63–76. <https://doi.org/10.1109/TCBB.2018.2846253> PMID: 29994126
173. Li L, Bansal MS. An Integer Linear Programming Solution for the Domain-Gene-Species Reconciliation Problem. In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* [Internet]. New York, NY, USA: ACM; 2018 [cited 2019 Sep 25]. p. 386–97. (BCB '18). Available from: <http://doi.acm.org/10.1145/3233547.3233603>
174. Li L, Bansal MS. Simultaneous Multi-Domain-Multi-Gene Reconciliation Under the Domain-Gene-Species Reconciliation Model. In: Cai Z, Skums P, Li M, editors. *Bioinformatics Research and Applications*. Springer International Publishing; 2019. p. 73–86. (Lecture Notes in Computer Science).
175. Kundu S, Bansal MS. SaGePhy: an improved phylogenetic simulation framework for gene and subgene evolution. *Bioinformatics* 2019 Sep 15; 35(18):3496–8. <https://doi.org/10.1093/bioinformatics/btz081> PMID: 30715213
176. Muhammad SA, Sennblad B, Lagergren J. Species tree-aware simultaneous reconstruction of gene and domain evolution. *bioRxiv*. 2018 Jun; 2:336453.
177. Bailly-Bechet M, Martins-Simões P, Szöllősi GJ, Mialdea G, Sagot MF, Charlat S. How Long Does *Wolbachia* Remain on Board? *Mol Biol Evol* 2017 May 1; 34(5):1183–93. <https://doi.org/10.1093/molbev/msx073> PMID: 28201740
178. Cuthill JFH, Charleston M. Wing patterning genes and coevolution of Müllerian mimicry in *Heliconius* butterflies: Support from phylogeography, cophylogeny, and divergence times. *Evolution*. 2015; 69(12):3082–3096. <https://doi.org/10.1111/evo.12812> PMID: 26552726
179. Ballinger MJ, Gawryluk RMR, Perlman SJ. Toxin and Genome Evolution in a *Drosophila* Defensive Symbiosis. *Genome Biol Evol* 2019 Jan 1; 11(1):253–62. <https://doi.org/10.1093/gbe/evy272> PMID: 30576446
180. Loreto ELS, Carareto CMA, Capy P. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* 2008 Jun; 100(6):545–54. <https://doi.org/10.1038/sj.hdy.6801094> PMID: 18431403
181. Denise R, Abby SS, Rocha EPC. Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLoS Biol* 2019 Jul 19; 17(7):e3000390. <https://doi.org/10.1371/journal.pbio.3000390> PMID: 31323028