



HAL
open science

Host-symbiont-gene phylogenetic reconciliation

Hugo Menet, Alexia Nguyen Trung, Vincent Daubin, Eric Tannier

► **To cite this version:**

Hugo Menet, Alexia Nguyen Trung, Vincent Daubin, Eric Tannier. Host-symbiont-gene phylogenetic reconciliation. 2022. hal-03781023v1

HAL Id: hal-03781023

<https://hal.science/hal-03781023v1>

Preprint submitted on 25 Nov 2022 (v1), last revised 22 May 2023 (v2)

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

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

Host-symbiont-gene phylogenetic reconciliation

Hugo Menet¹, Alexia Nguyen Trung¹, Vincent Daubin¹, and Eric Tannier^{1,2,*}

¹*Univ Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Évolutive UMR5558, F-69622 Villeurbanne, France*

²*Inria, centre de recherche de Lyon, 69622 Villeurbanne*

**To whom correspondence should be addressed: eric.tannier@inria.fr*

Abstract

Motivation: Biological systems are made of entities organized at different scales (*e.g.* macro-organisms, symbionts, genes...) which evolve in interaction. These interactions range from independence or conflict to cooperation and coevolution, which results in them having a common history. The evolution of such systems is approached by phylogenetic reconciliation, which describes the coevolution of two different levels, genes and species, or hosts and symbionts for example. The limit to two levels hides the multi-level inter-dependencies that characterize complex systems.

Results: We present a probabilistic model of evolution of three nested levels of organization which can account for the coevolution of hosts, symbionts and their genes. This model allows gene transfer as well as host switch, gene duplication as well as symbiont diversification inside a host, gene or symbiont loss. It handles the possibility of ghost lineages as well as temporary free-living symbionts.

Given three phylogenetic trees, we devise a Monte Carlo algorithm which samples evolutionary scenarios of symbionts and genes according to an approximation of their likelihood in the model. We evaluate the capacity of our method on simulated data, notably its capacity to infer horizontal gene transfers, and its ability to detect host-symbiont co-evolution by comparing host/symbiont/gene and symbiont/gene models based on their estimated likelihoods. Then we show in a aphid enterobacter system that some reliable transfers detected by our method, are invisible to classic 2-level reconciliation. We finally evaluate different hypotheses on human population histories in the light of their coevolving *Helicobacter pylori* symbionts, reconciled together with their genes.

Availability: Implementation is available on GitHub <https://github.com/hmenet/TALE>. Data are available on Zenodo <https://doi.org/10.5281/zenodo.6782794>.

1 Introduction

The toolbox of evolutionary biology largely relies on the assumption of statistical independence of biological objects at any level of organization: organisms from different

35 species are isolated from a biological system based on their genomes, genomes are cut
36 into independent genes, and inside genes, nucleotides are evolving independently from
37 each other [18].

38 Yet the essence of living systems lies in dependence: constraint, cooperation or
39 conflict [48]. Symbiotic micro-organisms coevolve with animals or plants [50]. The
40 ensemble they form is gathered under the holobiont concept. It allows to see genes as
41 entities not only following their own interest, not only participating to the functioning
42 of the genome they are hosted by, but also participating to, and probably evolving
43 with, a larger biological system.

44 A powerful tool to study these inter-dependencies is phylogenetic reconciliation: an
45 ensemble of models and methods explaining the differences and similarities between
46 phylogenies of two coevolving entities. Gene/species systems have been studied by
47 phylogenetic reconciliation, accounting for events of gene duplication, horizontal gene
48 transfer and gene loss (DTL model) [14, 38, 56, 9, 32]. The same model can be ap-
49 plied with little to no modification to symbiont/host [11, 47, 13], protein domain/gene
50 coevolution [42, 51], or biogeography [28, 45, 46]. DTL models have also been used to
51 reconstruct genome histories [16], detect highways of lateral gene transfers in bacteria,
52 archaea or eukaryota [7], assess the relative role of duplication and gene transfer in the
53 evolution of genomes [49], infer ancient symbiotic relationships [5], reconstruct histories
54 of gene fusion and fission [15], model endosymbiotic gene transfer [4].

55 A limitation of reconciliation methods is their separate application on molecular
56 studies on one side (gene/species coevolution), and ecological studies on the other
57 (host/symbiont coevolution). The striking methodological unity of the two (the same
58 DTL model is applied on both the molecular and ecological systems) and the growing
59 interest for multi-level systems integrating molecular and ecological inter-dependencies
60 (e.g. the holobiont concept) calls for a unique model for host, symbiont, gene coevo-
61 lution. In support of this claim, a number of empirical studies already rely on host
62 symbiont histories when proposing horizontal gene transfers between symbionts [41, 39,
63 27, 37], when often, only symbiont gene/species comparisons do not provide enough
64 statistical support for them [59, 43].

65 Three level reconciliations have been introduced by Stolzer et al. [51] and applied
66 to protein domain, gene and species. They describe two embedded DTL models and
67 an inference method by parsimony. The inference method first reconciles genes and
68 species trees in a DTL model. Then, knowing which genes are present in which species,
69 it reconciles the protein domains with the genes. This defines two kinds of horizon-
70 tal protein domain transfers between genes, depending on whether the genes are in
71 the same species (which we will call "intra" transfer) or not ("inter" transfer), with
72 a different cost for those two events. Further efforts in this direction have been pub-
73 lished by Li and Bansal [25] with a duplication/loss model between gene and species
74 and a DTL model, forbidding inter species transfers, between protein domains and
75 genes. They show NP-hardness of inferring the most parsimonious couple of nested
76 reconciliations [25] and propose different heuristics and problem variants [26, 24]. A
77 probabilistic model without transfers has been proposed by Muhammad, Sennblad, and
78 Lagergren [35]. It aims at inferring dated gene trees from protein domain alignments
79 using Markov Chain Monte Carlo. These attempts prove that it is possible to jointly

80 handle three nested levels in a single computational model, but none of them can yet
81 handle host/symbiont/gene systems in a statistical framework, because of specific limi-
82 tations of each of them (parsimony framework, no transfer or no inter-host transfer, no
83 joint inference between levels of organization, no explicit handling of absent lineages).

84 We propose a probabilistic model that describes the evolution of three nested coe-
85 volving entities at three different scales, adapted to a host/symbiont/gene system. In
86 our model a symbiont tree is generated by a DTL model inside the host, with a possibil-
87 ity of evolving temporarily outside the host phylogeny. A gene is generated by a DTL
88 model inside the symbiont, where gene transfer is more likely between symbionts that
89 share a common host ("intra" transfer) than for those that do not ("inter" transfer).

90 Based on this model we propose an inference method extending the two-level rec-
91 onciliation "ALE" software [55, 54]. It takes three trees as input, constructs joint
92 scenarios and estimates event rates and likelihoods according to the model. Our im-
93 plementation also features the possibility to infer a symbiont species tree if only the
94 host tree and several symbiont gene trees are given as input. In addition a comparison
95 of the likelihood of two-level and three-level reconciliations can be used as a test for
96 multi-scale coevolution.

97 We report a benchmark test of the inference method on simulated data, using an
98 external simulator [23], showing that under the hypothesis that gene transfers are
99 more likely between symbionts of a same host, the three-level reconciliation represent
100 a significant gain compared to the two-level one in terms of the capacity to retrieve the
101 symbiont donors and receivers of horizontal gene transfers.

102 We use the inference method to identify horizontal gene transfers between *Cinara*
103 aphid symbionts that are detected by expertise [27] but missed by two-level reconcili-
104 ations.

105 Finally we show on genes of *Helicobacter pylori* from human populations how likeli-
106 hood computations can be used to compare different hypotheses on the diversification
107 of a host, given the genes of its symbionts, taking into account the coevolution between
108 all three scales.

109 2 2-level reconciliation, definitions and prelim- 110 inaries

111 We denote by G, S, H respectively the gene tree, symbiont (or species) tree and host
112 tree. Given a tree T , $|T|$ is the number of nodes of T .

113 We briefly describe in this section a two-level DTL reconciliation model, based on
114 the undated version of the "ODT" model as implemented in "ALE undated" [54]. It
115 is a birth and death like model generating a rooted phylogenetic tree G inside S , with
116 speciation at all speciation nodes of S , and duplications, transfers and loss specific to
117 G along the branches of S . We thus have three rates for duplication, transfer and loss
118 events, concerning the evolution of genes inside their species. A gene tree can originate
119 in any branch of the species tree with a uniform prior.

120 The input of 2-level reconciliation inference is one gene tree, and one species tree,
121 with a many to one matching of the leaves. Both trees are assumed undated, binary

122 and rooted.

123 We call reconciliation scenario a list of events of kinds D,T,L, or S for each internal
124 gene tree node, that can be the result of the birth and death process (the events
125 happening on a branch are assigned to its child node). These lists transcribe into a
126 mapping of the gene tree nodes to the species tree nodes it evolves in. We note $R_{G,S}$
127 the set of all possible reconciliation scenarios between G and S .

128 We denote by p^S, p^D, p^T, p^L the probabilities for a gene evolving on a species tree
129 branch during the process to undergo the S,D,T,L events, with $p^S + p^D + p^T + p^L = 1$.
130 These probabilities are constant along the species tree and for all gene families. When
131 confusion is possible we add a S index for symbiont/gene reconciliation, and H index for
132 host/symbiont one. When a transfer occurs, the receiver branch is chosen according to
133 a uniform probability, avoiding ancestor branches of the donor. This avoids impossible
134 transfers but is not sufficient to guarantee that the overall scenario is time feasible.
135 The likelihood of a scenario r , $P(r|S)$ is the product of the probabilities of all events.

136 Summing over all possible scenarios for one gene tree and species tree we obtain
137 the likelihood of the gene tree $P(G|S)$. We do not have to enumerate all scenarios to
138 compute that sum, because we can compute this likelihood using dynamic program-
139 ming, considering matching all couples of gene and species sub-trees, starting from the
140 leaves, and enumerating all possible events to get each match. This in return enables
141 us to sample scenarios according to their likelihood, or finding the most likely scenario,
142 by backtracking through the table constructed.

143 We will call such a reconciliation of a gene tree and symbiont tree, "2-level" rec-
144 onciliation, in comparison with the host/symbiont/gene 3-level reconciliation that we
145 introduce in the following section.

146 3 3-level reconciliation, likelihood estimation and 147 scenario inference

148 3.1 Elements of the probabilistic model

149 We use an undated framework similar to the one implemented in ALE undated [54]
150 presented in previous section. A rooted binary host phylogenetic tree H is given,
151 without branch lengths. We do not include the generation parameters in our model
152 and consider instead the host tree as a parameter. We model the evolution of one
153 or multiple symbiont trees S with the DTL model [54] (see previous section) adding
154 the possibility for a symbiont to live temporarily in an unknown host (this feature is
155 described in more detail in Section 3.8).

156 We then model the evolution of genes in the symbiont trees with duplication, loss
157 and intra horizontal transfer, meaning that horizontal transfer is possible only between
158 symbiont branches that are present in the same host branch. We thus have six rates in
159 our model, three for the duplication, transfer and loss between host and symbiont, and
160 three additional ones for duplication, intra-transfer and loss events concerning genes
161 inside their species (note that both p_H^S and p_S^S the speciation probabilities for the
162 host/symbiont and symbiont/gene reconciliations are not free parameters as in both

cases $p^S = 1 - p^D - p^T - p^L$). An illustration of the realization of such a model, as

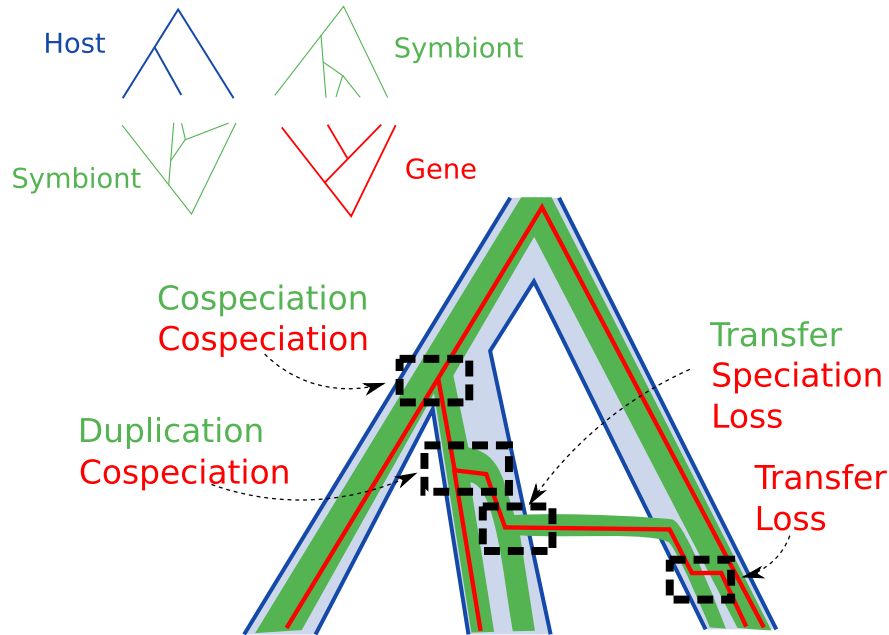


Figure 1: An example of a 3-Level reconciliation input (top left, with three trees and associations between the leaves of two couples of trees) and a possible reconciliation scenario for this input. Events of the host/symbiont co-evolution are written in red, while events of the symbiont/gene reconciliation are written in green.

163
164

well as the input for the inference part, is given in Figure 1.

165
166
167

This model can be immediately used for simulations, but we chose to use an external simulator for our tests [23], to minimize the similarities in models and implementation between simulation and inference.

168

3.2 Monte Carlo approximation of the likelihood

169
170
171
172
173
174
175
176

The inference consists in computing the probability that G have been generated by the model: $P(G|S, H)$, from a set of input trees (one rooted host tree H , one or several rooted symbiont trees S , one or several unrooted gene trees G), and given the DTL probabilities for the two reconciliation levels. It is then possible to estimate the evolutionary rates and sample among reconciliation scenarios. We also propose a way to propose one likely symbiont tree if only the host tree and some gene trees are given. All given trees are supposed to be binary, and branch lengths are not taken into account.

177
178
179
180

Because a similar computation in a parsimonious framework is NP-hard [25], it is probably not possible to exactly and quickly compute $P(G|S, H)$. We thus apply an approximation technique based on sampling reconciliations. The probability of a gene tree can indeed be decomposed by summing over all possible host/symbiont

181 reconciliation scenarios $r_{S,H}$:

$$P(G|S, H) = \sum_{r_{S,H} \in R_{S,H}} P(G|S, H, r_{S,H})P(r_{S,H}|S, H) \quad (1)$$

182 The number of reconciliations in this sum is at least exponential in the size of the input
 183 (and even the number of scenarios maximizing $P(r_{S,H}|S, H)$ can be exponential [13]),
 184 so we use a Monte Carlo approach to estimate it, sampling a reasonable number N of
 185 symbiont/host reconciliations:

$$P(G|S, H) \simeq \frac{1}{N} \sum_{n=1}^N P(G|S, H, r_n) \quad (2)$$

186 where r_n is sampled in the set $R_{S,H}$ of all reconciliations according to its likelihood
 187 $P(r_n|S, H)$.

188 3.3 Reconciliation inference and ghost lineages

189 Sampling reconciliations in $R_{S,H}$ can be done with the dynamic programming algorithm
 190 implemented in "ALE undated" and is a two-level reconciliation problem [56].

191 Given $r_n \in R_{S,H}$, the probability $P(G|S, H, r_n)$ can be computed with an adapta-
 192 tion of the same dynamic programming algorithm. It consists in checking, during the
 193 dynamic programming process, for all gene transfer possibilities, if the donor symbiont
 194 i and receiver one j share a host in r_n . If they do, then it is an "intra" transfer and
 195 the transfer has the probability defined by the transfer rate.

196 In our model gene transfer can only occur between two symbiont species inside
 197 a same host. However transfer between two symbionts in different hosts is possible
 198 through ghost species. Indeed it is always reasonable to assume that a major part of
 199 species are extinct or unsampled and gene transfers are often "from the dead" [52, 19,
 200 61]. In consequence in the model a transfer can occur from a donor that is now extinct.
 201 This transfer is traced back to an ancestor of this extinct donor that is not in the same
 202 host than the receiver. See in Figure 2 how an "inter" transfer between i and j (on
 203 the left) can be modelled (on the right) by an extinct sister lineage to symbiont i , that
 204 switched host, and transferred a gene to j while being in the same host. As the sister
 205 lineage goes extinct, in the inferred gene history it is transferred from i to j .

206 We denote by $P_S^T(i \rightarrow j)$ the probability for a gene present in symbiont i to undergo
 207 a horizontal transfer to symbiont j , and $P_H^T(e \rightarrow h)$ the probability for a gene present
 208 in a symbiont associated to host e to transfer to a symbiont associated to host h . Let
 209 H_i (H_j) be the set of host branches that contain symbiont i (resp. j). We go from P_H^T
 210 to P_S^T by summing over all possibles hosts h of the receiver symbiont j and all hosts e
 211 of the donor symbiont i :

$$P_S^T(i \rightarrow j) = \sum_{e \in H_i, h \in H_j} P_H^T(e \rightarrow h) \quad (3)$$

212 At fixed h we rewrite with $P_e = P^T(e \rightarrow h)$. Recall p_S^T are the probability of horizontal
 213 transfer in the symbiont/gene reconciliation, and $p_H^S, p_H^D, p_H^T, p_H^L$ the probabilities of

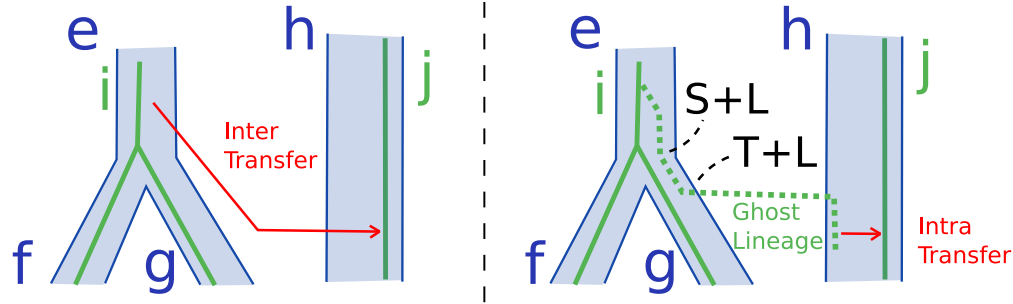


Figure 2: Computation of inter transfer rate from intra transfer rate and ghost species: the left inter transfer can be modeled by multiple scenarios without inter gene transfer but implying ghost symbiont lineages, such as the one on the right that implies first a speciation and a loss, and then a transfer and a loss before the extinction of the species.

214 speciation, duplication, transfer and loss in the host/symbiont reconciliation. Let E_e
 215 be the probability of extinction, that is, the probability that a gene is present in a
 216 branch e of the host tree and absent from all the leaves. Let $|S_h|$ be the number of
 217 symbiont branches matched to host h in the host/symbiont reconciliation scenario.
 218 The initial case in our inductive definition of $P_e = P(e \rightarrow h)$ is the case where $e = h$,
 219 so when the donor symbiont is in one of the receiver symbiont host, in that case the
 220 probability to transfer to that one symbiont of h , is uniform among the $|S_h|$ symbionts
 221 present in h . Then, for the induction, we rewrite the undated reconciliation equations,
 222 to progress a symbiont in the host tree from any host e to host h of the receiver symbiont
 223 and such that the symbiont species we invoke then goes extinct. The notations are
 224 similar to those used in the undated ALE description in [33], or figure 2: we denote by
 225 f, g the children of a host e .

$$\begin{cases} P_e = \frac{1}{|S_h|} p_S^T & \text{if } e = h \\ P_e = p_H^S (P_f E_g + P_g E_f) + 2p_H^D P_e E_e + \sum_{k \in H} \frac{p_H^T}{|H|} P_k E_e \end{cases} \quad (4)$$

226 This equation has a self dependency due to the Transfer/Loss event, which is already
 227 accounted for in reconciliation methods [20, 55]. We forbid successions of several
 228 Transfer/Loss events to break this self dependency and solve this equation.

229 3.4 Sequential and 2-level estimation of the likelihood

230 Because the Monte Carlo approach can be computationally heavy, we devised an al-
 231 ternative "Sequential" heuristic. Instead of sampling scenarios randomly like in the
 232 Monte Carlo, we select the one that maximises the marginal likelihood [60]. That is,
 233 at each step of the backtracking of the dynamic programming procedure we select the
 234 maximum likelihood position.

235 This approach is similar to the one of Stolzer et al. [51], but in a probabilistic setting,
 236 using marginal likelihood, and with a way of computing the inter transfer probabilities

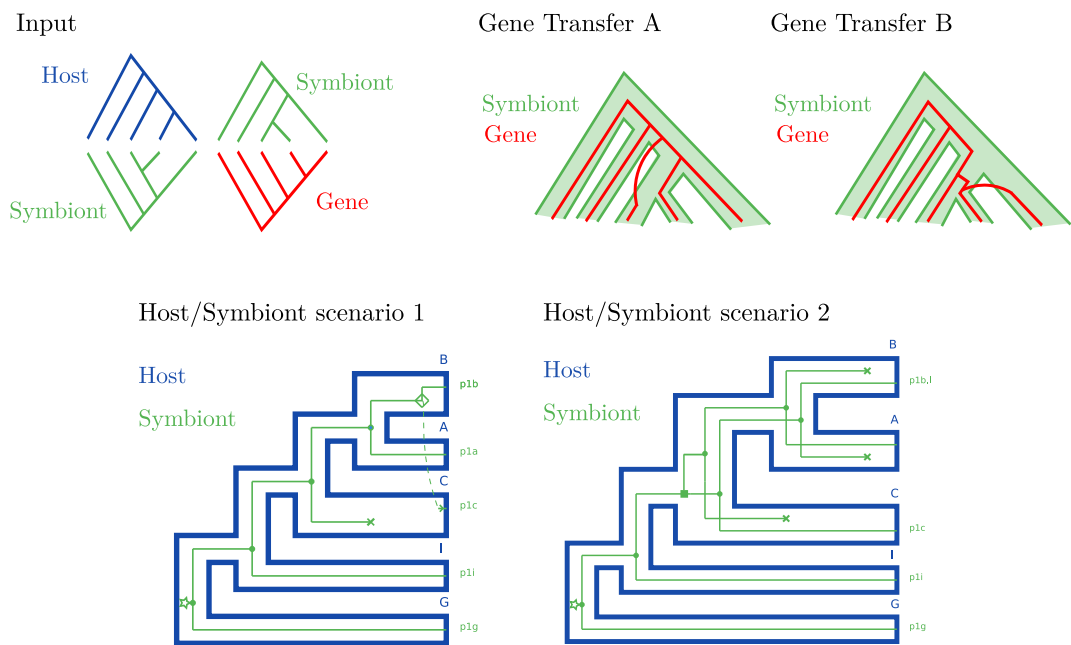


Figure 3: An example of input where the Sequential heuristic is less robust than the Monte Carlo one. We compare the support for two gene transfer scenarios, scenario A and B. There are two main possible host/symbiont reconciliation scenarios, scenario 1 and scenario 2. In scenario 1, gene transfer A is more likely, and in scenario 2, gene transfer B is more likely. The support for both gene transfers for the Sequential and Monte Carlo heuristics are presented in table 1.

237 from the host/symbiont and symbiont/gene DTL reconciliation parameters instead of
 238 using an additional parameter.

239 We study the differences between the Sequential heuristic, the Monte Carlo sam-
 240 pling, and a classic 2-level approach, in order to measure the importance of taking into
 241 account the dependency between host symbiont and symbiont gene reconciliations.
 242 The description of the 3-level model is also important from a theoretical point of view
 243 to get a better understanding of what we compute in the different heuristic inference
 244 methods. Writing the equations also enables us to see the different parameters that
 245 can be inferred, using a maximum likelihood framework for instance. An example is
 246 the possibility to infer a host/symbiont reconciliation aware of the genes.

247 The faster Sequential heuristic may not be as robust as the Monte Carlo one. Li
 248 and Bansal in [26] present an example where the sequential approach cannot propose a
 249 solution at all, in a parsimony model where inter horizontal gene transfer are forbidden.
 250 In figure 3 we present another illustration, with this time an emphasis on the "not
 251 continuous" aspect of the Sequential heuristic in regard to the host and symbiont
 252 reconciliation events rates.

253 A small change in the transfer rate of the host and symbiont makes a big difference
 254 for the gene and symbiont reconciliation with the Sequential heuristic, but a small one
 255 for the Monte Carlo one, see the results in table 1.

Heuristic	Gene transfer A	Gene transfer B
Host Symbiont rates T 0.006 D 0.1 L 0.1		
Monte Carlo	0.43	0.27
Sequential	0.90	< 0.05
2-level	0.18	0.21
Host Symbiont rates T 0.005 D 0.1 L 0.1		
Monte Carlo	0.35	0.33
Sequential	< 0.05	0.49
2-level	0.19	0.23

Table 1: Comparison of the support for the two gene transfer scenarios in the example presented in figure 3.

3.5 Time complexity and tractability

DTL reconciliation methods use a dynamic programming approach to compute the probability of the coevolution of two trees (and all of their subtrees) [10]. Then backtracking produces reconciliation scenarios. In a gene/species reconciliation, if all transfers have the same probability, *i.e.* this probability is independent from the donor-receiver couple, DTL reconciliation can be computed in quadratic time [6].

We denote h , s , g the number of nodes of the host, symbiont, and gene tree respectively.

The first part of our algorithm, is to reconcile host and symbiont, a classic 2-level reconciliation, with quadratic complexity $O(hs)$. In our implementation sampling a host symbiont reconciliation scenario is then done in cubic complexity (as we extend the transfer sum) however it might be possible to also get a quadratic complexity here, though as we do not consider all couples, it is in general faster than the forward computation.

Then we compute the gene transfer probabilities between all couple of symbiont nodes, this is done with a dynamic programming similar to the one for reconciliation in $O(hs)$, and a final sum (equation 3) over all hosts of the considered symbionts in $O(h^2s^2)$, which in the reasonable case where the number of symbiont nodes per host nodes (in the reconciliation scenario) is below a constant k , we get $O(h^2k^2 + hs)$ for this part.

Finally we can compute the host aware gene/symbiont reconciliation. The difference with classic 2-level reconciliation is that transfer rates depend on the donor-receiver couple. In consequence we cannot use the efficient computation trick used for uniform rates. For each couple of gene and symbiont subtrees, we must explicitly consider transfers toward all symbiont nodes, yielding a cubic complexity of $O(s^2g)$ for host aware symbiont/gene reconciliation.

For the Monte Carlo approach, we repeat all steps except the initial host/symbiont reconciliation, leading to a total complexity of $O(N(hs + h^2k^2 + s^2g))$ where h , s , g are the size of the host, symbiont and gene trees, k is a bound on the number of symbiont per host in the sampled reconciliations (s in the worst case), and N is the number of samples in the Monte Carlo approach.

287 The datasets presented give a good idea of the size of the data we can consider with
288 this new method. Computation on the Cinara aphid dataset, with a size of 25 leaves
289 for the symbiont tree, 9 leaves for the host, and 13 gene families takes about 3 minutes
290 on a single laptop core, including the rate estimation steps. This is a dataset on which
291 it would be possible to use the Monte Carlo approach. The pylori dataset is larger, the
292 symbiont has 119 leaves, the host 7 leaves, and there are 1034 gene families, of which
293 322 have 119 leaves. Reconciliation, with fixed rates (without rate estimation) took
294 just under a day using 8 cores.

295 **3.6 Symbiont tree inference**

296 In case the symbiont tree is unknown, we devised an option to infer the symbiont tree
297 by amalgamation [12, 55] of universal unicopy gene trees, guided by the host tree.

298 Clade prior probabilities are computed from universal unicopy gene trees, and dy-
299 namic programming is used to compute the likelihood. A symbiont tree is sampled in
300 the backtracking phase at the same time as the host/symbiont reconciliation scenario.

301 This amalgamation is also implemented for the symbiont/gene part, to account for
302 gene tree being unrooted, and to be able to include uncertainty in gene tree topology,
303 just like in 2-level reconciliations[20, 55].

304 **3.7 Rates estimation and likelihood comparison**

305 In our model, the data is the gene trees, and the free parameters are the three DTL
306 probabilities of the symbiont/gene reconciliation. We consider the host/symbiont DTL
307 parameters as fixed, *i.e.* estimated without knowing the data. This makes it possible
308 to compare, based on the likelihood, our approach and a 2-level one (symbiont/gene
309 reconciliation, unaware of the host), because they have the same free parameters, and
310 because they both define a probability distribution on the same space of gene trees
311 associated to the symbiont tree.

312 In practice we estimate the host/symbiont DTL parameters, as done in ALE [56],
313 with an expectation maximization method, and then fix these parameters. Then we
314 run the Monte Carlo or sequential approach multiple times to estimate rates for the
315 symbiont/gene reconciliation with the same expectation maximisation method.

316 **3.8 Free living relatives of symbionts**

317 In the course of their evolutionary history, some symbiont may live outside a host, or
318 within an unknown host.

319 This is particularly important for us because we invoke unknown hosts in the case
320 of inter host horizontal gene transfers (section 3.3). In order to consider these cases
321 we added the possibility for a symbiont to be "free living", meaning associated to no
322 host.

323 We did that by adding the symbiont tree as a possible host tree, and matching
324 the symbiont leaves with no host to themselves. In that way, we see transfer between
325 free living as less likely than when a common host is known. The utility of this model

326 addition is visible in the *Cinara* aphids example developed in the Results section (see
327 Fig. 6).

328 **3.9 Output format and solution visualization**

329 Our implementation can output a sample of full scenarios, both for symbiont/genes
330 and the corresponding host/symbiont reconciliations. The scenarios are given in Rec-
331 PhyloXML, a common standard for reconciliation output endorsed by a significant part
332 of the gene/species reconciliation community [17]. The scenarios can be visualised us-
333 ing Thirdkind¹ [40], a reconciliation viewer that handles 3-level reconciliations. We
334 also output event frequencies based on the reconciliation scenario sampling. Indeed
335 we sample a number (100 by default) of symbiont/gene reconciliations and observe the
336 frequency of each event in these replicas, thus obtaining an estimate of the posterior
337 probability of events. It is this result that we use to evaluate the ability of our method
338 to infer specific events, such as receptors and donors of horizontal symbiont transfers,
339 which we compare to simulated scenarios or previously proposed scenarios on aphids
340 *Cinara*.

341 **4 Experimental results**

342 **4.1 Simulated datasets**

343 **4.1.1 Description of the simulation process**

344 Our probabilistic model can be used for simulation, however in order to test our
345 method, we chose to use an exterior simulation framework. We used the available
346 software Sagephy developed by Kundu and Bansal [23]. Sagephy generates three em-
347 bedded trees and allows replacing transfers on top of additive ones. We used the
348 parameters proposed by the same team in another article [22], as representative of
349 small (D 0.133, T 0.266, L 0.266), medium (D 0.3, T 0.6, L 0.6) and high (D 0.6, T 1.2,
350 L 1.2) transfer rates, without replacing transfers. The software enables to specify an
351 inter transfer rate, corresponding to the probability for a gene transfer to be between
352 symbionts hosted by different hosts ("inter" transfer). When a horizontal transfer
353 is chosen during generation of the gene tree (inside a symbiont tree and knowing a
354 host/symbiont reconciliation), the transfer is chosen to be an inter host one with the
355 inter transfer rate. So an inter transfer rate of 0 corresponds to our inference model
356 of only intra transfer, and of 1 corresponds to a case where transfers are only between
357 symbionts in separate hosts.

358 We constructed two simulated datasets, one with a combination of the different
359 rates for the DTL parameters, and one with only medium rates but with different inter
360 transfer rates. For the first dataset, we used all 9 combinations of small, medium and
361 high rates for the symbiont generation and the gene generation, with only intra host
362 gene transfer (*i.e.* an inter transfer rate of zero). For the second dataset, we used only

¹<https://crates.io/crates/thirdkind>

363 medium rates for both symbiont and genes generation, but we used 6 inter transfer
364 rates going from 0 to 1.

365 For both datasets, and for each set of rates, we generated 50 instances consisting
366 of 1 host tree with 100 leaves, 1 symbiont tree and 5 gene trees, each generated in the
367 pruned version of the other trees (branches that do not reach present are pruned before
368 the generation of the next tree). We then selected host leaves with a probability of
369 0.08 to simulate unexhaustive sampling, resulting in host trees with an average size of
370 8 leaves. We thus simulate extinct lineages, and even with a simulation inter transfer
371 rate of 0, some gene transfers will be inter. This ended up to 399 instances for the first
372 dataset and 226 instances for the second one, and at least 29 instances of 5 genes for
373 each set of parameters.

374 We compared the results from three approaches. (1) The "2-level" heuristic which
375 is a 2-level reconciliation between the gene and symbiont trees, ignorant of the host
376 tree. (2) The "Sequential" heuristic, which consists in computing the most likely
377 host/symbiont DTL reconciliation and doing the symbiont/gene reconciliation, given
378 that host/symbiont reconciliation. (3) The full 3-level "Monte Carlo" method, sum-
379 ming the results of the gene reconciliations over 50 sampled host/symbiont reconcilia-
380 tion scenarios. We let our approaches estimate evolutionary rates.

381 We measured first the capacity of the three methods to infer the correct symbiont
382 donor and recipient of gene transfers (with precision and recall), and second, the like-
383 lihood they attribute to each symbiont/gene coevolution. Identifying the exact donor
384 and recipient of simulated transfers is usually considered a hard task for reconciliation
385 algorithms. Usually reconciliation studies are not evaluated with this strong criterion
386 [36], but with the inference of ancestral characters [58], the number of transfers [53],
387 the ability to infer better trees [8], or the ability to map the correct event type to each
388 gene node [22]. We chose to look at the capacity to infer specific transfers because
389 we feel that it is in this task that our model has the capacity to show its utility. It
390 can infer more precise gene transfers because transfers are constrained by additional
391 elements compared to other methods.

392 Our probabilistic reconciliation approaches output estimates of the posterior proba-
393 bilities of evolutionary events, so we used these probabilities as weights for our precision
394 and recall definition in Figure 4 for the detection of horizontal gene transfer donor and
395 receiver symbionts. Denoting by $L_{t,sim}$ the list of simulated transfers and $L_{t,obs}$ the list
396 of observed transfers, and $P_{obs}(T)$ the estimation of our approach for the probability
397 of transfer T .

$$398 \text{ Precision} = \frac{\sum_{T \in L_{t,sim}} P_{obs}(T)}{\sum_{T \in L_{t,obs}} P_{obs}(T)} \text{ and Recall} = \frac{\sum_{T \in L_{t,sim}} P_{obs}(T)}{\sum_{T \in L_{t,sim}} 1} \quad (5)$$

399 **4.1.2 The 3-level method infers more true transfers than the 2-level** 400 **method**

401 Overall the Monte Carlo and sequential approaches give similar results on these sim-
402 ulated datasets, and better results (in particular for recall and to a lesser extent for
403 precision) than the 2-level approach (Figure 4). In most cases, the faster Sequential
404 heuristic can advantageously replace the Monte Carlo one because they have the same

405 recall and precision. In a few case, that might be the more interesting ones, the Monte
 406 Carlo has a slight advantage, and though it is more computationally costly, it is also
 407 theoretically more robust.

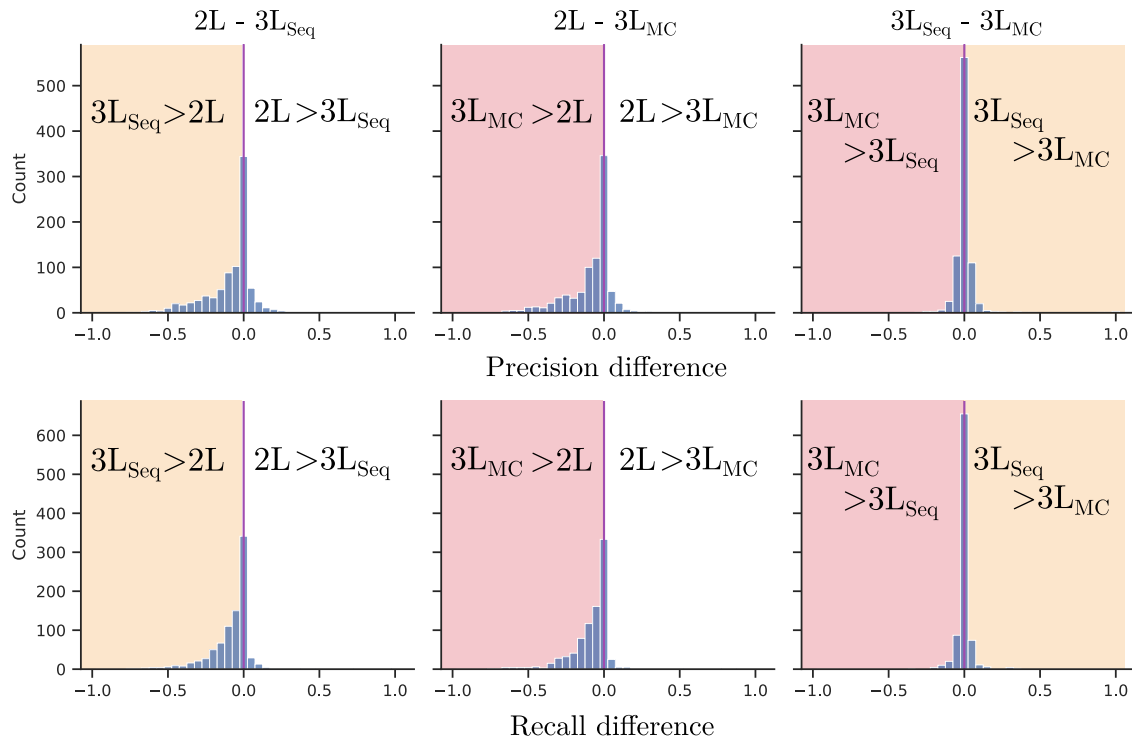


Figure 4: Distribution of differences of precision and recall on the inference of horizontal gene transfers for all combinations of two approaches: 2-level (2L), 3-level with the Monte Carlo heuristic ($3L_{MC}$) and 3-level with the Sequential heuristic ($3L_{Seq}$), centered on 0, and for all 874 gene families of the 3-level simulation, with no inter host gene transfer, that undergo at least one transfer.

408 4.1.3 A host-symbiont co-evolution test

409 The reconciliation likelihood difference between 3-level inference and 2-level inference
 410 is a marker of host-symbiont co-evolution. Indeed, Figure 5 (A) shows that when the
 411 simulation model is less dependent from the host phylogeny (inter transfer rates of 0.6,
 412 0.8 and 1.0), the likelihood difference between the 2-level and 3-level inference methods
 413 are mostly in favor of the 2-level. It happens for almost all instances in the 1.0 model
 414 with no intra transfers. For all these instances a preference for 2-level reconciliation
 415 (according to the likelihood) is more likely when few transfers are inferred (we sum
 416 over 1 to 5 gene families generated for each host and symbiont instance). This is a sign
 417 of the precision of the method to not classify 2-level instances as 3-level ones.

418 In a model with only intra transfers (inter transfer rate of 0), we have a very good
 419 recall for the detection of the 3-level model, almost all only intra transfer instances are
 420 classified as 3-level as they should be. A more detailed exam of this recall is presented

421 in Figure 5 (B) with the first simulated dataset, with only intra transfer and varying
422 DTL parameters.

423 Figure 5 (B) shows the likelihood difference when only intra transfers occur in
424 the simulations. We see that when the number of transfers is higher, the likelihood
425 difference better reflects the mode of simulation. In practice a way to increase the
426 number of transfers is to increase the number of gene families considered.

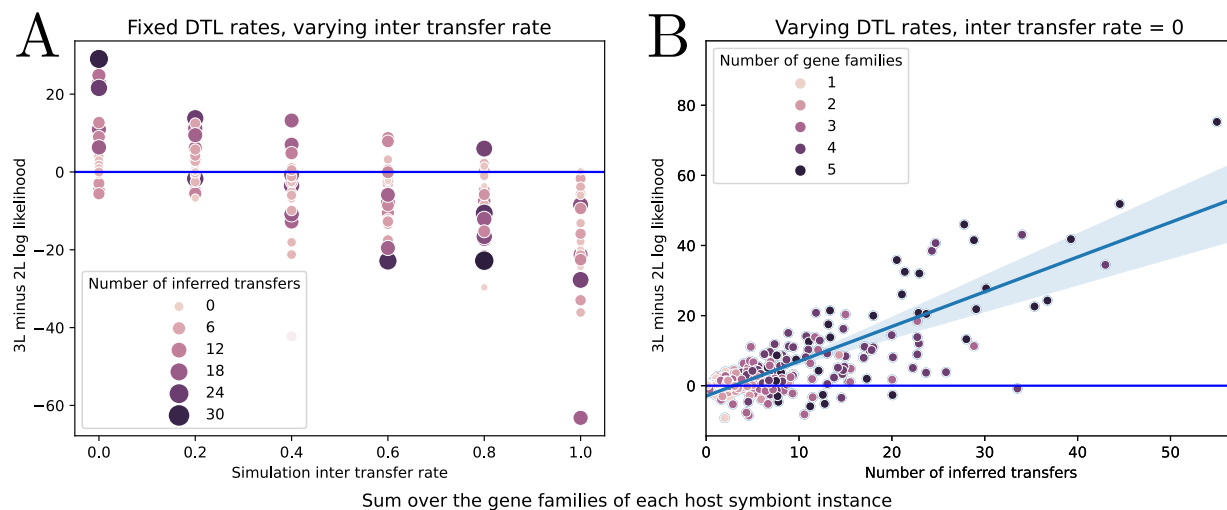


Figure 5: A test of host symbiont co-evolution. We measure the difference of likelihood between the 3-level model and the 2-level model, using the estimation of these likelihoods provided by our "2-level" and "3-level Sequential" heuristics, in order to differentiate instances where gene trees are generated in a 3-level host/symbiont/gene model or in a 2-level symbiont/gene model. Each instance is composed of a host tree, a symbiont tree, and 1 to 5 gene families. For one instance we sum the differences over all gene families. (A) Sensitivity of the likelihood difference to the value of the inter host gene transfer probability in Sagephy. As expected, the more an inter transfer rate is probable (independent from the host phylogeny), the less we detect host-symbiont co-evolution with the likelihood difference measure. Colors indicate the number of inferred transfers. (B) Sensitivity of the likelihood difference to the number of inferred transfers (dataset with only intra transfers). Colors depict the number of gene families considered in the host and symbiont instance. Because transfers carry the co-evolution signal, the sensitivity of the method increases with the number of transfers, which are higher if we increase the number of gene families.

427 4.2 Precise identification of a gene transfer in enterobac- 428 teria symbiotic of *Cinara* aphids

429 A recent study on *Cinara* aphids enterobacteria systems [27] identified one host switch
430 and two horizontal gene transfers, one intra-host from *Erwinina* to *Hamiltonella* and

431 one inter-host from *Sodalis* to *Erwinia*. The genes transferred (thi) and some others
432 (bioa,d,b) were first inherited through gene transfers, probably from *Sodalis* related
433 symbionts. Moreover, those genes transferred are part of functions to complement the
434 lack in the sap-feeding host nutrition. It seems that a new endosymbiont acquires the
435 genes of another one to sustain the host. This exemplifies a case where a symbiont
436 gene can co-evolve with the symbiont host, more than with the symbiont itself. We
437 reproduced this scenario in Figure 6 (A), and a representative gene tree witnessing the
438 transfers is reproduced in Figure 6 (B).

439 Gene trees including *Cinara* endosymbionts and other enterobacteria species were
440 available from the supplementary material made available by Manzano-Marín et al.
441 [27]. *Cinara* and their endosymbionts phylogenies show exact correspondences on the
442 studied period. We kept all enterobacteria associated to a *Cinara* aphid (of *Erwinia*
443 and *Hamiltonella* genus), and chose a representative subset of the other enterobacteria
444 present in the gene trees, notably containing *Sodalis* species, closest identified parent
445 to one of the transferred genes, and other *Erwinia* and *Hamiltonella* genus species.
446 We used the phylogeny proposed in Annotree for these species [31], to complement the
447 *Cinara* aphids symbionts phylogeny proposed in [27]. We used our 3-level reconciliation
448 on the host tree and symbiont tree, using the possibility of our method to take into
449 account these "free living" bacteria. As the host and symbiont (apart from the free
450 living) are identical, we used the sequential heuristic.

451 We tested the capacity of the 3-level method compared to a 2-level one to detect
452 the gene transfers identified by Manzano-Marín et al. [27]. The intra transfer from
453 *Erwinia* to *Hamiltonella* is retrieved in around 80 percent of the scenarios sampled by
454 the 3-level method, and both are better retrieved than in the method that does not take
455 the host into account (Figure 6 (D)). A theoretical explanation using a toy example is
456 given in Figure 6 (C). An alternative transfer, in the other direction, from *Hamiltonella*
457 to *Erwinia* is slightly more likely but the configuration of the host evolution supports
458 the intra transfer.

459 This exemplifies how multi-scale dependencies can only be captured by 3-level mod-
460 els.

461 **4.3 *Helicobacter pylori* genes as documents for human** 462 **migrations**

463 *Helicobacter pylori* is a bacterial symbiont of a significant proportion of humans, which
464 has been supposed to be a marker of human migrations across the Earth [1]. Bacterial
465 strains have been divided in different populations corresponding to geographical areas
466 (Africa 1, Africa 2, Asia 2, East Asia, North East Africa, Europe) [57, 29].

467 The supposed coevolving complex made by humans, bacterial symbiont and their
468 genes makes it an accessible system for the host/symbiont/gene reconciliation method.
469 In particular gene transfers should be more probable between *Helicobacter* strains if
470 they are hosted by a same human population.

471 We collected available current strains of *H. pylori* from the NCBI which have a
472 genetic population assigned by MLST allelic profile [2, 21]. A phylogenetic tree was
473 built based on the concatenation of universal-unicopy genes (322 gene families), and

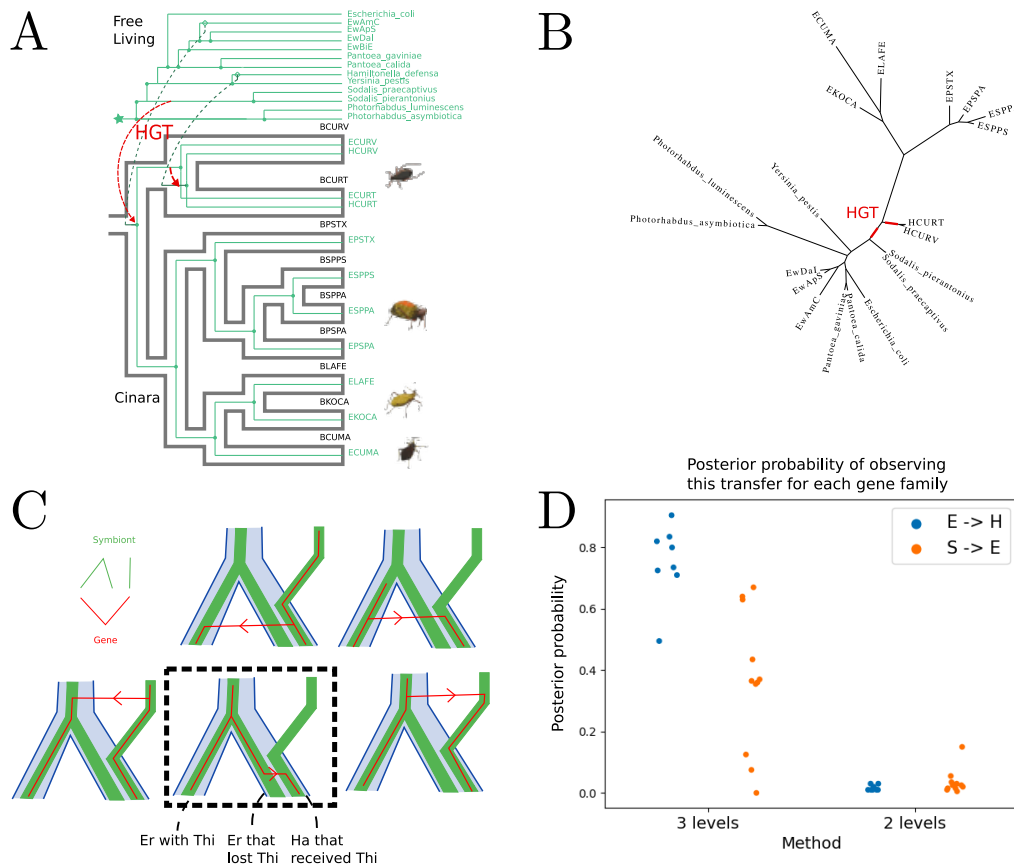


Figure 6: The evolution of *Cinara* and their enterobacteria symbionts. (A) The coevolutionary scenario identified by Manzano-Marín et al. [27]. The reconciliation of the hosts (*Cinara* aphids) and symbionts (bacteria) are depicted along with the position of the horizontal gene transfers (in red). (B) Phylogenetic tree of one gene with the position of the two transfers. (C) Theoretical explanation of the difference between the results of the 2-level and 3-level reconciliation methods. The two top reconciliations are a bit more likely in a 2-level framework, as they require a single transfer while the bottom ones require a transfer and a loss, but one of the bottom one (with the dotted square) is better in a 3-level model, as it allows an intra-host transfer. (D) Support (a posteriori probability of the transfer, computed from its observed frequency in the reconciliation sample) for the identified HGTs, from *Erwinia* to *Hamiltonella*, and from *Sodalis* to *Erwinia*, for 3-level and 2-level reconciliations.

474 a sample of 113 strains representing the diversity of *H. pylori* in the old world (ex-
475 cluding strains from the Americas) was obtained using Treemmer [30]. Then, 6 non
476 *pylori* strains were added (*H. hepaticus*, *H. acinonychis*, *H. canadensis*, *H. felis*, *H.*
477 *bizzozeronii*, *H. cetorum*), as external groups.

478 In this study we considered the 1034 gene families, including 322 universal unicopy
479 families, that displayed strains from the external groups and from at least 3 continents.

480 We then considered four different population trees (host trees) containing the geo-
481 graphical areas as leaves, coherent with the scientific literature [57, 29]. 322 universal
482 unicopy gene trees were used, and the strain (symbiont) tree was amalgamated from
483 gene trees with the population trees as a guide (see subsection 3.6). As strains were
484 much more numerous than populations, and subject to a more complex diversification
485 than DTL events, we allowed an additional event, named I, that consists in a duplica-
486 tion followed by a speciation and loss of one of the copies, with a specific rate, inferior
487 to the combination of these three events. This event allows a strain to be present in
488 a population and one of its descendants, and is used as one of the default events in
489 biogeography reconciliation frameworks [44].

490 We then applied our sequential approach and compared the likelihood of the gene-
491 /strains aware of the host reconciliation to compare the population trees. The results
492 are depicted in Figure 7 (A). The likelihood of the systems according to the population
493 tree is reported, divided into two components: the likelihood of the population/strain
494 comparison, and the likelihood of the gene/strain aware of the population comparison.
495 The population tree on the left column is the most likely given the model, the method
496 and the used data. Assessing the robustness of the result would require a sensibility
497 study which is out of the scope of this contribution.

498 Figure 7 (B) is an illustration of a reconciliation scenario for the maximum likelihood
499 host tree with Thirdkind [40]. We see the host tree and the amalgamated strain tree
500 reconciled (I events are represented as transfers from a parent node to one of its child).
501 On top of these two embedded trees red lines represent the aggregation of gene transfers
502 depending on the host of the donor and receiver strains. The opacity of the transfer
503 lines are proportional to the number of times a certain kind of transfer is observed
504 across the 1034 gene families in one sampled scenario.

505 5 Discussion

506 In a review on horizontal gene transfer in host symbiont systems [59] the authors
507 highlight the need of plurality of evidence to robustly assess the existence of transfers.
508 Evidence can be of multiple types, gene trees, donor receiver ecology, or host symbiont
509 association. We provide a framework where these multiple evidence can be gathered,
510 and the proof of concept that it can work, on *Cinara* aphids and their enterobacteria.

511 Our method uses a probabilistic framework that enables rate estimation, tree infer-
512 ence, tree comparison and model comparison. We also introduced a method to compute
513 the inter transfer rate from the intra transfer one and the modeling of ghost lineages
514 in the host symbiont reconciliation. We introduced a Monte Carlo approach that en-
515 ables to estimate event probabilities and likelihood, by sampling through multiple host
516 symbiont scenarios in a double DTL model.

A

Pylori putative population tree				
Symbiont tree	Amalgamation of the 322 universal unicopy gene trees			
Host/symbiont Log likelihood	- 308	- 320	- 322	- 320
Gene trees	1034 gene trees			
Symbiont/gene Log likelihood	-653×10^3	-658×10^3	-658×10^3	-676×10^3

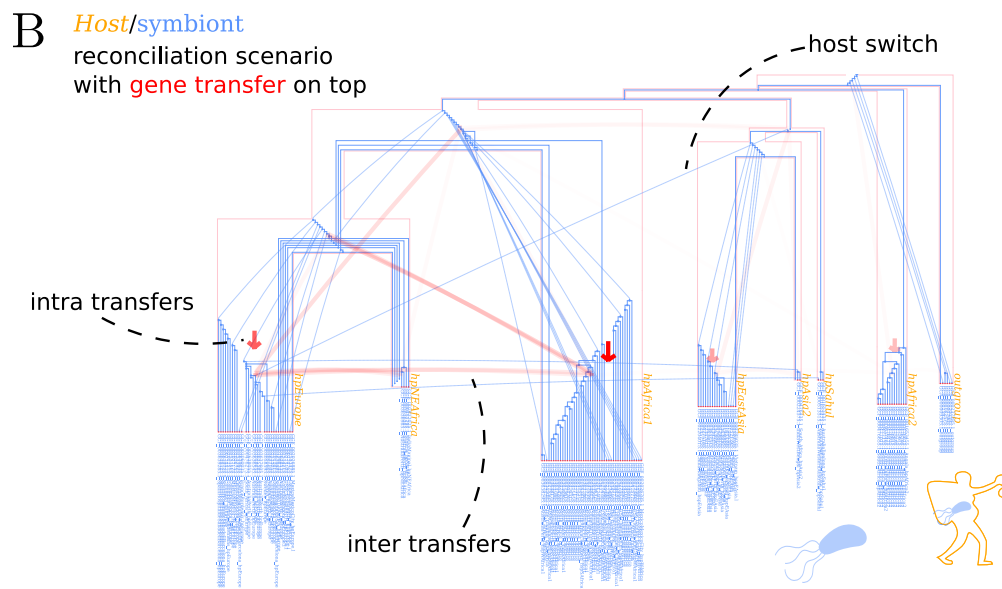


Figure 7: Co-evolution of human populations and *Helicobacter pylori*. (A) Log likelihood of the different population trees. (B) The representation with Third-Kind [40] of one possible reconciliation scenario of *Helicobacter pylori* strain tree and the population tree maximizing likelihood. Aggregated gene transfers are depicted on top of the DTL reconciliation, with the opacity corresponding to the number of time the transfers were seen across the 1034 gene families.

517 While our intuition is that the Monte Carlo approach is more robust than the
518 sequential one, notably in cases where gene events happen around uncertain host
519 symbiont reconciliation nodes, our evaluation on simulated data did not show a big
520 difference in most cases. We think that in biological data, we can expect more in-
521 teraction between the events of the host symbiont reconciliation and the ones of the
522 gene symbiont one, which are independent in our simulation. Developing new simula-
523 tion frameworks that can model such dependencies, for instance by increasing the loss
524 rates when multiple genes or symbionts are present, or using a functional approach to
525 the evolution of genes, could be important to the understanding of these multi-level
526 models.

527 The ability of our inference methods to be used for model comparison seems promis-
528 ing. We saw that with an increasing number of gene families we could increase our
529 confidence in the answer. However the different gene families must contain a part of
530 independent information, as is the case in the simulation where all families dependence
531 are completely in the host and symbiont trees. For instance in the *Cinara* aphids
532 dataset, the genes considered are mostly similar, and do not really make the number of
533 independent transfers increase, and with only one intra transfer, that necessitates an
534 additional loss to occur, the 2-level model displays a better likelihood than the 3-level.
535 If more independent transfers were present, we can suppose that some of them might
536 not necessitate such a loss and the test would favor a host symbiont co-evolution.

537 All these features deserve further tests to know their domain of validity and to
538 draw biological conclusions. In particular, the inference of the symbiont tree, with the
539 use of amalgamation, from an input distribution of universal unicopy gene tree would
540 deserve to be tested against other standard methods as concatenate or species tree
541 reconstruction with 2-level reconciliation model as it is implemented in SpeciesRax
542 [34].

543 An interesting future direction in this line would be to construct, instead of a
544 symbiont tree, compartment trees, which would depict the coevolution of genes that
545 are not necessarily in the same species.

546 A comparison of the inference method to similar ones [51, 25, 35] could also be
547 undertaken. However in an host/symbiont/gene framework, horizontal transfer in the
548 host/symbiont reconciliation are crucial, and only the model of Stolzer et al. [51] takes
549 these events into account. Moreover the sequential heuristic is simply a rewriting of
550 this model in a probabilistic framework.

551 More generally, the model is not bound to host/symbiont/gene systems, but any set
552 of three nested coevolving entities can be studied with it: species/gene/protein domain
553 as it was done in previous studies [51, 24, 35], or geography/species/gene, and so on.
554 As the scales of biological observation are probably infinite, so are the combination of
555 three nested scales.

556 Examples presented in this article show the possibilities of the method, but still
557 derive no biologically significant breakthrough. However the necessity of such a method,
558 detecting multilevel co-evolution, could arise with the more and more numerous studied
559 biological systems that fit into this multi-scale coevolution framework, notably with an
560 increasing interest for hologenomics [3].

561

6 Acknowledgements

562

This work was performed using the computing facilities of the CC LBBE/PRABI.

563

7 Funding

564

This work was supported by the French National Research Agency (Grant ANR-19-CE45-0010 Evolutthon).

565

566

References

567

[1] Mark Achtman. “How old are bacterial pathogens?” In: *Proceedings of the Royal Society B: Biological Sciences* 283.1836 (Aug. 17, 2016), p. 20160990. DOI: 10.1098/rspb.2016.0990. URL: <https://royalsocietypublishing.org/doi/full/10.1098/rspb.2016.0990> (visited on 09/12/2019).

568

569

570

571

[2] Mark Achtman, Takeshi Azuma, Douglas E. Berg, Yoshiyuki Ito, Giovanna Morelli, Zhi-Jun Pan, Sebastian Suerbaum, Stuart A. Thompson, Arie Van Der Ende, and Leen-Jan Van Doorn. “Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions”. In: *Molecular Microbiology* 32.3 (1999), pp. 459–470. DOI: <https://doi.org/10.1046/j.1365-2958.1999.01382.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2958.1999.01382.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.1999.01382.x>.

572

573

574

575

576

577

578

579

580

[3] Antton Alberdi, Sandra B. Andersen, Morten T. Limborg, Robert R. Dunn, and M. Thomas P. Gilbert. “Disentangling host–microbiota complexity through hologenomics”. In: *Nature Reviews Genetics* 23.5 (May 2022). Number: 5 Publisher: Nature Publishing Group, pp. 281–297. ISSN: 1471-0064. DOI: 10.1038/s41576-021-00421-0. URL: <https://www.nature.com/articles/s41576-021-00421-0> (visited on 06/01/2022).

581

582

583

584

585

586

[4] Yoann Anselmetti, Nadia El-Mabrouk, Manuel Lafond, and Aïda Ouan-graoua. “Gene tree and species tree reconciliation with endosymbiotic gene transfer”. In: *Bioinformatics* 37 (Supplement_1 July 1, 2021), pp. i120–i132. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btab328. URL: <https://doi.org/10.1093/bioinformatics/btab328> (visited on 10/20/2021).

587

588

589

590

591

[5] Marc Bailly-Bechet, Patricia Martins-Simões, Gergely J. Szöllősi, Gladys Mialdea, Marie-France Sagot, and Sylvain Charlat. “How Long Does *Wolbachia* Remain on Board?” In: *Molecular Biology and Evolution* 34.5 (May 1, 2017), pp. 1183–1193. ISSN: 0737-4038. DOI: 10.1093/molbev/msx073. URL: <https://academic.oup.com/mbe/article/34/5/1183/2992913> (visited on 12/02/2019).

592

593

594

595

596

- 597 [6] Mukul S. Bansal, Eric J. Alm, and Manolis Kellis. “Efficient algorithms for
598 the reconciliation problem with gene duplication, horizontal transfer and
599 loss”. In: *Bioinformatics* 28.12 (June 15, 2012). Publisher: Oxford Aca-
600 demic, pp. i283–i291. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/
601 bts225. URL: [https://academic.oup.com/bioinformatics/article/
602 28/12/i283/269262](https://academic.oup.com/bioinformatics/article/28/12/i283/269262) (visited on 06/01/2020).
- 603 [7] Mukul S. Bansal, Guy Banay, J. Peter Gogarten, and Ron Shamir. “Detect-
604 ing Highways of Horizontal Gene Transfer”. In: *Journal of Computational
605 Biology* 18.9 (Sept. 1, 2011), pp. 1087–1114. DOI: 10.1089/cmb.2011.0066.
606 URL: <https://www.liebertpub.com/doi/abs/10.1089/cmb.2011.0066>
607 (visited on 02/20/2020).
- 608 [8] Mukul S. Bansal, Yi-Chieh Wu, Eric J. Alm, and Manolis Kellis. “Improved
609 gene tree error correction in the presence of horizontal gene transfer”. In:
610 *Bioinformatics* 31.8 (Apr. 15, 2015), pp. 1211–1218. ISSN: 1367-4803. DOI:
611 10.1093/bioinformatics/btu806. URL: [https://academic.oup.com/
612 bioinformatics/article/31/8/1211/212713](https://academic.oup.com/bioinformatics/article/31/8/1211/212713) (visited on 11/18/2019).
- 613 [9] Bastien Boussau and Celine Scornavacca. “Reconciling Gene trees with
614 Species Trees”. In: *Phylogenetics in the Genomic Era*. Ed. by Celine Scor-
615 navacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher
616 — Authors open access book, 2020, 3.2:1–3.2:23. URL: [https://hal.
617 archives-ouvertes.fr/hal-02535529](https://hal.archives-ouvertes.fr/hal-02535529) (visited on 08/20/2020).
- 618 [10] Michael Charleston. “Jungles: a new solution to the host/parasite phy-
619 logeny reconciliation problem”. In: *Mathematical Biosciences* 149.2 (May
620 1998), pp. 191–223. ISSN: 0025-5564. DOI: 10.1016/S0025-5564(97)
621 10012-8.
- 622 [11] Michael Charleston and Ran Libeskind-Hadas. “Event-Based Cophyloge-
623 netic Comparative Analysis”. In: *Modern Phylogenetic Comparative Meth-
624 ods and Their Application in Evolutionary Biology*. Springer Berlin Heidel-
625 berg, 2014, pp. 465–480. DOI: 10.1007/978-3-662-43550-2_20. URL:
626 https://doi.org/10.1007/978-3-662-43550-2_20.
- 627 [12] Lawrence A. David and Eric J. Alm. “Rapid evolutionary innovation during
628 an Archaean genetic expansion”. In: *Nature* 469.7328 (Jan. 2011), pp. 93–
629 96. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature09649. URL: [http:
630 //www.nature.com/articles/nature09649](http://www.nature.com/articles/nature09649) (visited on 02/19/2020).
- 631 [13] Beatrice Donati, Christian Baudet, Blerina Sinaimer, Pierluigi Crescenzi,
632 and Marie-France Sagot. “EUCALYPT: efficient tree reconciliation enu-
633 merator”. In: *Algorithms for Molecular Biology* 10.1 (Jan. 23, 2015), p. 3.
634 ISSN: 1748-7188. DOI: 10.1186/s13015-014-0031-3. URL: [https://doi.
635 org/10.1186/s13015-014-0031-3](https://doi.org/10.1186/s13015-014-0031-3) (visited on 02/18/2020).

- 636 [14] Jean-Philippe Doyon, Vincent Ranwez, Vincent Daubin, and Vincent Berry.
637 “Models, algorithms and programs for phylogeny reconciliation”. In: *Brief-*
638 *ings in Bioinformatics* 12.5 (Sept. 2011), pp. 392–400. ISSN: 1477-4054.
639 DOI: 10.1093/bib/bbr045.
- 640 [15] Wandrille Duchemin, Yoann Anselmetti, Murray Patterson, Yann Ponty,
641 S everine B erard, Cedric Chauve, Celine Scornavacca, Vincent Daubin, and
642 Eric Tannier. “DeCoSTAR: Reconstructing the Ancestral Organization of
643 Genes or Genomes Using Reconciled Phylogenies”. In: *Genome Biology and*
644 *Evolution* 9.5 (May 1, 2017), pp. 1312–1319. DOI: 10.1093/gbe/evx069.
645 URL: <https://academic.oup.com/gbe/article/9/5/1312/3200386>
646 (visited on 09/13/2019).
- 647 [16] Wandrille Duchemin, Vincent Daubin, and Eric Tannier. “Reconstruction
648 of an ancestral *Yersinia pestis* genome and comparison with an ancient se-
649 quence”. In: *BMC Genomics* 16.10 (Oct. 2, 2015), S9. ISSN: 1471-2164. DOI:
650 10.1186/1471-2164-16-S10-S9. URL: [https://doi.org/10.1186/1471-](https://doi.org/10.1186/1471-2164-16-S10-S9)
651 [2164-16-S10-S9](https://doi.org/10.1186/1471-2164-16-S10-S9) (visited on 01/05/2022).
- 652 [17] Wandrille Duchemin, Guillaume Gence, Anne-Muriel Arigon Chifolleau,
653 Lars Arvestad, Mukul S. Bansal, Vincent Berry, Bastien Boussau, Fran ois
654 Chevenet, Nicolas Comte, Adri an A. Dav ın, Christophe Dessimoz, David
655 Dylus, Damir Hasic, Diego Mallo, R emi Planel, David Posada, Celine Scor-
656 navacca, Gergely Sz oll osi, Louxin Zhang,  Eric Tannier, and Vincent Daubin.
657 “RecPhyloXML: a format for reconciled gene trees”. In: *Bioinformatics*
658 34.21 (Nov. 1, 2018), pp. 3646–3652. ISSN: 1367-4803. DOI: 10.1093/
659 [bioinformatics/](https://academic.oup.com/bioinformatics/article/34/21/3646/4995844)
660 [bty389](https://academic.oup.com/bioinformatics/article/34/21/3646/4995844). URL: [https://academic.oup.com/bioinformatics/](https://academic.oup.com/bioinformatics/article/34/21/3646/4995844)
[article/34/21/3646/4995844](https://academic.oup.com/bioinformatics/article/34/21/3646/4995844) (visited on 09/12/2019).
- 661 [18] Joseph Felsenstein. *Inferring Phylogenies*. Oxford, New York: Oxford Uni-
662 versity Press, Sept. 4, 2003. 580 pp. ISBN: 978-0-87893-177-4.
- 663 [19] Gregory P. Fournier, Jinling Huang, and J. Peter Gogarten. “Horizontal
664 gene transfer from extinct and extant lineages: biological innovation and
665 the coral of life”. In: *Philosophical Transactions of the Royal Society B:*
666 *Biological Sciences* 364.1527 (Aug. 12, 2009), pp. 2229–2239. ISSN: 0962-
667 8436. DOI: 10.1098/rstb.2009.0033. URL: [https://www.ncbi.nlm.nih.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2873001/)
668 [gov/pmc/articles/PMC2873001/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2873001/) (visited on 09/21/2021).
- 669 [20] Edwin Jacox, Cedric Chauve, Gergely J. Sz oll osi, Yann Ponty, and Celine
670 Scornavacca. “ecceTERA: comprehensive gene tree-species tree reconcilia-
671 tion using parsimony”. In: *Bioinformatics (Oxford, England)* 32.13 (2016),
672 pp. 2056–2058. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btw105.
- 673 [21] Keith A. Jolley, James E. Bray, and Martin C. J. Maiden. “Open-access
674 bacterial population genomics: BIGSdb software, the PubMLST.org web-
675 site and their applications”. In: *Wellcome Open Research* 3 (2018), p. 124.
676 ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.14826.1.

- 677 [22] Misagh Kordi, Soumya Kundu, and Mukul S. Bansal. “On Inferring Addi-
678 tive and Replacing Horizontal Gene Transfers Through Phylogenetic Rec-
679 onciliation”. In: *Proceedings of the 10th ACM International Conference on*
680 *Bioinformatics, Computational Biology and Health Informatics*. BCB ’19.
681 Niagara Falls, NY, USA: Association for Computing Machinery, Sept. 4,
682 2019, pp. 514–523. ISBN: 978-1-4503-6666-3. DOI: 10.1145/3307339.3342168. URL: <https://doi.org/10.1145/3307339.3342168> (visited
683 on 02/18/2020).
684
- 685 [23] Soumya Kundu and Mukul S. Bansal. “SaGePhy: an improved phyloge-
686 netic simulation framework for gene and subgene evolution”. In: *Bioin-*
687 *formatics* 35.18 (Sept. 15, 2019), pp. 3496–3498. ISSN: 1367-4803. DOI:
688 10.1093/bioinformatics/btz081. URL: [https://academic.oup.com/](https://academic.oup.com/bioinformatics/article/35/18/3496/5305633)
689 [bioinformatics/article/35/18/3496/5305633](https://academic.oup.com/bioinformatics/article/35/18/3496/5305633) (visited on 09/25/2019).
- 690 [24] Lei Li and Mukul S. Bansal. “An Integer Linear Programming Solution
691 for the Domain-Gene-Species Reconciliation Problem”. In: *Proceedings of*
692 *the 2018 ACM International Conference on Bioinformatics, Computational*
693 *Biology, and Health Informatics*. BCB ’18. event-place: Washington, DC,
694 USA. New York, NY, USA: ACM, 2018, pp. 386–397. ISBN: 978-1-4503-
695 5794-4. DOI: 10.1145/3233547.3233603. URL: [http://doi.acm.org/10.](http://doi.acm.org/10.1145/3233547.3233603)
696 [1145/3233547.3233603](http://doi.acm.org/10.1145/3233547.3233603) (visited on 09/25/2019).
- 697 [25] Lei Li and Mukul S. Bansal. “An Integrated Reconciliation Framework for
698 Domain, Gene, and Species Level Evolution”. In: *IEEE/ACM Transactions*
699 *on Computational Biology and Bioinformatics* 16.1 (Jan. 2019), pp. 63–76.
700 DOI: 10.1109/TCBB.2018.2846253.
- 701 [26] Lei Li and Mukul S. Bansal. “Simultaneous Multi-Domain-Multi-Gene Rec-
702 onciliation Under the Domain-Gene-Species Reconciliation Model”. In: *Bioin-*
703 *formatics Research and Applications*. Ed. by Zhipeng Cai, Pavel Skums,
704 and Min Li. Lecture Notes in Computer Science. Springer International
705 Publishing, 2019, pp. 73–86. ISBN: 978-3-030-20242-2.
- 706 [27] Alejandro Manzano-Marín, Armelle Coeur D’acier, Anne-Laure Clamens,
707 Céline Orvain, Corinne Cruaud, Valérie Barbe, and Emmanuelle Jouselin.
708 “Serial horizontal transfer of vitamin-biosynthetic genes enables the estab-
709 lishment of new nutritional symbionts in aphids’ di-symbiotic systems”. In:
710 *The ISME Journal* (Oct. 17, 2019), pp. 1–15. ISSN: 1751-7370. DOI: 10.
711 1038/s41396-019-0533-6. URL: [https://www.nature.com/articles/](https://www.nature.com/articles/s41396-019-0533-6)
712 [s41396-019-0533-6](https://www.nature.com/articles/s41396-019-0533-6) (visited on 11/05/2019).
- 713 [28] Andrés Martínez-Aquino. “Phylogenetic framework for coevolutionary stud-
714 ies: a compass for exploring jungles of tangled trees”. In: *Current Zoology*
715 62.4 (Aug. 2016), pp. 393–403. ISSN: 1674-5507. DOI: 10.1093/cz/zow018.
716 URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5804275/>
717 (visited on 09/13/2019).

- 718 [29] Francis Mégraud, Philippe Lehours, and Filippa F. Vale. “The history of
719 *Helicobacter pylori*: from phylogeography to paleomicrobiology”. In: *Clinical Microbiology and Infection* 22.11 (Nov. 1, 2016), pp. 922–927. ISSN:
720 1198-743X. DOI: 10.1016/j.cmi.2016.07.013. URL: <https://www.sciencedirect.com/science/article/pii/S1198743X16302373> (vis-
721 ited on 09/21/2021).
722
723
- 724 [30] Fabrizio Menardo, Chloé Loiseau, Daniela Brites, Mireia Coscolla, Sebast-
725 tian M. Gygli, Liliana K. Rutaihwa, Andrej Trauner, Christian Beisel, Sonia
726 Borrell, and Sebastien Gagneux. “Treemmer: a tool to reduce large phylo-
727 genetic datasets with minimal loss of diversity”. In: *BMC Bioinformatics*
728 19.1 (May 2, 2018), p. 164. ISSN: 1471-2105. DOI: 10.1186/s12859-018-
729 2164-8. URL: <https://doi.org/10.1186/s12859-018-2164-8> (visited
730 on 01/10/2022).
- 731 [31] Kerrin Mendler, Han Chen, Donovan H Parks, Briallen Lobb, Laura A
732 Hug, and Andrew C Doxey. “AnnoTree: visualization and exploration of a
733 functionally annotated microbial tree of life”. In: *Nucleic Acids Research*
734 47.9 (May 21, 2019), pp. 4442–4448. ISSN: 0305-1048. DOI: 10.1093/nar/
735 gkz246. URL: <https://doi.org/10.1093/nar/gkz246> (visited on
736 09/20/2021).
- 737 [32] Hugo Menet, Vincent Daubin, and Eric Tannier. “Phylogenetic reconcil-
738 iation”. June 2021. URL: [https://hal.archives-ouvertes.fr/hal-
739 03258402](https://hal.archives-ouvertes.fr/hal-03258402) (visited on 10/20/2021).
- 740 [33] Benoit Morel, Alexey M. Kozlov, Alexandros Stamatakis, and Gergely
741 J. Szöllösi. *GeneRax: A tool for species tree-aware maximum likelihood
742 based gene family tree inference under gene duplication, transfer, and loss*.
743 preprint. Bioinformatics, Sept. 26, 2019. DOI: 10.1101/779066. URL: [http:
744 //biorxiv.org/lookup/doi/10.1101/779066](http://biorxiv.org/lookup/doi/10.1101/779066) (visited on 03/10/2020).
- 745 [34] Benoit Morel, Paul Schade, Sarah Lutteropp, Tom A. Williams, Gergely
746 J. Szöllösi, and Alexandros Stamatakis. *SpeciesRax: A tool for maximum
747 likelihood species tree inference from gene family trees under duplication,
748 transfer, and loss*. Company: Cold Spring Harbor Laboratory Distributor:
749 Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory
750 Section: New Results Type: article. Mar. 29, 2021, p. 2021.03.29.437460.
751 DOI: 10.1101/2021.03.29.437460. URL: [https://www.biorxiv.org/
752 content/10.1101/2021.03.29.437460v1](https://www.biorxiv.org/content/10.1101/2021.03.29.437460v1) (visited on 01/06/2022).
- 753 [35] Sayyed Awn Muhammad, Bengt Sennblad, and Jens Lagergren. “Species
754 tree-aware simultaneous reconstruction of gene and domain evolution”. In:
755 *bioRxiv* (June 2, 2018), p. 336453. DOI: 10.1101/336453. URL: [https://
756 www.biorxiv.org/content/10.1101/336453v1](https://www.biorxiv.org/content/10.1101/336453v1) (visited on 09/25/2019).

- 757 [36] Agnieszka Mykowiecka, Anna Muszewska, and Paweł Górecki. “Inferring
758 time-consistent and well-supported horizontal gene transfers”. In: 2018
759 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
760 IEEE Computer Society, Dec. 1, 2018, pp. 79–83. ISBN: 978-1-5386-5488-
761 0. DOI: 10.1109/BIBM.2018.8621558. URL: <https://www.computer.org/csdl/proceedings-article/bibm/2018/08621558/17D45Xi9rWz>
762 (visited on 10/20/2021).
- 764 [37] Atsushi Nakabachi, Reiko Ueoka, Kenshiro Oshima, Roberta Teta, Alfonso
765 Mangoni, Mihaela Gurgui, Neil J. Oldham, Gerhild van Echten-Deckert,
766 Keiko Okamura, Kohei Yamamoto, Hiromitsu Inoue, Moriya Ohkuma, Yuichi
767 Hongoh, Shin-ya Miyagishima, Masahira Hattori, Jörn Piel, and Takema
768 Fukatsu. “Defensive Bacteriome Symbiont with a Drastically Reduced Genome”.
769 In: *Current Biology* 23.15 (Aug. 5, 2013), pp. 1478–1484. ISSN: 0960-9822.
770 DOI: 10.1016/j.cub.2013.06.027. URL: [https://www.cell.com/current-biology/abstract/S0960-9822\(13\)00752-5](https://www.cell.com/current-biology/abstract/S0960-9822(13)00752-5) (visited on 11/06/2019).
- 772 [38] Luay Nakhleh. “Computational approaches to species phylogeny inference
773 and gene tree reconciliation”. In: *Trends in ecology and evolution* 28 (2013),
774 pp. 719–728.
- 775 [39] Naruo Nikoh, Takahiro Hosokawa, Minoru Moriyama, Kenshiro Oshima,
776 Masahira Hattori, and Takema Fukatsu. “Evolutionary origin of insect–Wolbachia
777 nutritional mutualism”. In: *Proceedings of the National Academy of Sci-
778 ences of the United States of America* 111.28 (July 15, 2014), pp. 10257–
779 10262. ISSN: 0027-8424. DOI: 10.1073/pnas.1409284111. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4104916/> (visited on
780 11/06/2019).
- 782 [40] Simon Penel, Hugo Menet, Théo Tricou, Vincent Daubin, and Eric Tannier.
783 “Thirdkind: displaying phylogenetic encounters beyond 2-level reconcilia-
784 tion”. In: *Bioinformatics (Oxford, England)* (Feb. 9, 2022), btac062. ISSN:
785 1367-4811. DOI: 10.1093/bioinformatics/btac062.
- 786 [41] Thomas Penz, Stephan Schmitz-Esser, Suzanne E. Kelly, Bodil N. Cass,
787 Anneliese Müller, Tanja Woyke, Stephanie A. Malfatti, Martha S. Hunter,
788 and Matthias Horn. “Comparative Genomics Suggests an Independent Ori-
789 gin of Cytoplasmic Incompatibility in *Cardinium hertigii*”. In: *PLOS Ge-
790 netics* 8.10 (Oct. 25, 2012), e1003012. ISSN: 1553-7404. DOI: 10.1371/
791 journal.pgen.1003012. URL: [https://journals.plos.org/plosgenetics/
792 article?id=10.1371/journal.pgen.1003012](https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1003012) (visited on 11/06/2019).
- 793 [42] Matthew D. Rasmussen and Manolis Kellis. “Unified modeling of gene du-
794 plication, loss, and coalescence using a locus tree”. In: *Genome Research*
795 22.4 (Apr. 1, 2012), pp. 755–765. ISSN: 1088-9051. DOI: 10.1101/gr.
796 123901.111. URL: [http://genome.cshlp.org/cgi/doi/10.1101/gr.
797 123901.111](http://genome.cshlp.org/cgi/doi/10.1101/gr.123901.111) (visited on 10/23/2019).

- 798 [43] Matt Ravenhall, Nives Škunca, Florent Lassalle, and Christophe Dessimoz.
799 “Inferring Horizontal Gene Transfer”. In: *PLOS Computational Biology*
800 11.5 (May 28, 2015), e1004095. ISSN: 1553-7358. DOI: 10.1371/journal.
801 pcbi.1004095. URL: [https://journals.plos.org/ploscompbiol/
802 article?id=10.1371/journal.pcbi.1004095](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004095) (visited on 02/19/2020).
- 803 [44] Richard H Ree, Brian R Moore, Campbell O Webb, and Michael J Donoghue.
804 “A likelihood framework for inferring the evolution of geographic range on
805 phylogenetic trees”. In: (2005), p. 13.
- 806 [45] Richard H. Ree and Stephen A. Smith. “Maximum Likelihood Inference of
807 Geographic Range Evolution by Dispersal, Local Extinction, and Cladogen-
808 esis”. In: *Systematic Biology* 57.1 (Feb. 1, 2008), pp. 4–14. ISSN: 1063-5157.
809 DOI: 10.1080/10635150701883881. URL: [https://academic.oup.com/
810 sysbio/article/57/1/4/1703014](https://academic.oup.com/sysbio/article/57/1/4/1703014) (visited on 10/02/2019).
- 811 [46] Fredrik Ronquist. “Dispersal-Vicariance Analysis: A New Approach to the
812 Quantification of Historical Biogeography”. In: *Systematic Biology* 46.1
813 (Mar. 1, 1997), pp. 195–203. ISSN: 1063-5157. DOI: 10.1093/sysbio/46.
814 1.195. URL: [https://academic.oup.com/sysbio/article/46/1/195/
815 1685511](https://academic.oup.com/sysbio/article/46/1/195/1685511) (visited on 10/03/2019).
- 816 [47] Santi Santichaivekin, Qing Yang, Jingyi Liu, Ross Mawhorter, Justin Jiang,
817 Trenton Wesley, Yi-Chieh Wu, and Ran Libeskind-Hadas. “eMPress: a
818 systematic cophylogeny reconciliation tool”. In: *Bioinformatics* (btaa978
819 Nov. 20, 2020). ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btaa978.
820 URL: <https://doi.org/10.1093/bioinformatics/btaa978> (visited on
821 06/07/2021).
- 822 [48] Jan Sapp. *Evolution by association*. Oxford University Press, 1994.
- 823 [49] Joel Sjöstrand, Ali Tofigh, Vincent Daubin, Lars Arvestad, Bengt Sennblad,
824 and Jens Lagergren. “A Bayesian Method for Analyzing Lateral Gene
825 Transfer”. In: *Systematic Biology* 63.3 (May 1, 2014), pp. 409–420. ISSN:
826 1063-5157. DOI: 10.1093/sysbio/syu007. URL: [https://doi.org/10.
827 1093/sysbio/syu007](https://doi.org/10.1093/sysbio/syu007) (visited on 01/06/2022).
- 828 [50] Justin L. Sonnenburg and Erica D. Sonnenburg. “Vulnerability of the in-
829 dustrialized microbiota”. In: *Science* 366.6464 (2019), eaaw9255. DOI: 10.
830 1126/science.aaw9255. eprint: [https://www.science.org/doi/pdf/
831 10.1126/science.aaw9255](https://www.science.org/doi/pdf/10.1126/science.aaw9255). URL: [https://www.science.org/doi/abs/
832 10.1126/science.aaw9255](https://www.science.org/doi/abs/10.1126/science.aaw9255).
- 833 [51] Maureen Stolzer, Katherine Siewert, Han Lai, Minli Xu, and Dannie Du-
834 rand. “Event inference in multidomain families with phylogenetic reconcil-
835 iation”. In: *BMC Bioinformatics* 16.14 (Oct. 2, 2015), S8. ISSN: 1471-2105.
836 DOI: 10.1186/1471-2105-16-S14-S8. URL: [https://doi.org/10.1186/
837 1471-2105-16-S14-S8](https://doi.org/10.1186/1471-2105-16-S14-S8) (visited on 09/25/2019).

- 838 [52] Gergely J. Szöllösi, Eric Tannier, Nicolas Lartillot, and Vincent Daubin.
839 “Lateral gene transfer from the dead”. In: *Systematic Biology* 62.3 (May 1,
840 2013), pp. 386–397. ISSN: 1076-836X. DOI: 10.1093/sysbio/syt003.
- 841 [53] Gergely J. Szöllösi, Bastien Boussau, Sophie S. Abby, Eric Tannier, and
842 Vincent Daubin. “Phylogenetic modeling of lateral gene transfer recon-
843 structs the pattern and relative timing of speciations”. In: *Proceedings of*
844 *the National Academy of Sciences* 109.43 (Oct. 23, 2012). Publisher: Na-
845 tional Academy of Sciences Section: Biological Sciences, pp. 17513–17518.
846 ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1202997109. URL: <https://www.pnas.org/content/109/43/17513> (visited on 06/05/2020).
- 848 [54] Gergely J. Szöllösi, Adrián Arellano Davín, Eric Tannier, Vincent Daubin,
849 and Bastien Boussau. “Genome-scale phylogenetic analysis finds extensive
850 gene transfer among fungi”. In: *Philosophical Transactions of the Royal*
851 *Society B: Biological Sciences* 370.1678 (Sept. 26, 2015). Publisher: Royal
852 Society, p. 20140335. DOI: 10.1098/rstb.2014.0335. URL: <https://royalsocietypublishing.org/doi/full/10.1098/rstb.2014.0335>
853 (visited on 10/20/2021).
- 855 [55] Gergely J. Szöllösi, Wojciech Rosikiewicz, Bastien Boussau, Eric Tannier,
856 and Vincent Daubin. “Efficient Exploration of the Space of Reconciled Gene
857 Trees”. In: *Systematic Biology* 62.6 (Nov. 2013), pp. 901–912. ISSN: 1063-
858 5157. DOI: 10.1093/sysbio/syt054. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3797637/> (visited on 02/18/2020).
- 860 [56] Gergely J. Szöllösi, Eric Tannier, Vincent Daubin, and Bastien Boussau.
861 “The Inference of Gene Trees with Species Trees”. In: *Systematic Biology*
862 64.1 (Jan. 2015), e42–e62. ISSN: 1063-5157. DOI: 10.1093/sysbio/syu048.
863 URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4265139/>
864 (visited on 12/02/2019).
- 865 [57] Langgeng Agung Waskito and Yoshio Yamaoka. “The Story of *Helicobacter*
866 *pylori*: Depicting Human Migrations from the Phylogeography”. In: *Helicobacter pylori in Human Diseases: Advances in Microbiology, Infectious*
867 *Diseases and Public Health Volume 11*. Ed. by Shigeru Kamiya and Steffen
868 Backert. Advances in Experimental Medicine and Biology. Cham: Springer
869 International Publishing, 2019, pp. 1–16. ISBN: 978-3-030-21916-1. DOI: 10.
870 1007/5584_2019_356. URL: https://doi.org/10.1007/5584_2019_356
871 (visited on 09/21/2021).
- 873 [58] Nicolas Wieseke, Tom Hartmann, Matthias Bernt, and Martin Middendor-
874 f. “Cophylogenetic Reconciliation with ILP”. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 12.6 (Nov. 1, 2015),
875 pp. 1227–1235. ISSN: 1545-5963. DOI: 10.1109/TCBB.2015.2430336.
876 URL: <http://ieeexplore.ieee.org/document/7103015/> (visited on
877 10/28/2019).
878

- 879 [59] Bhagya K. Wijayawardena, Dennis J. Minchella, and J. Andrew DeWoody.
880 “Hosts, parasites, and horizontal gene transfer”. In: *Trends in Parasitol-*
881 *ogy* 29.7 (July 1, 2013), pp. 329–338. ISSN: 1471-4922, 1471-5007. DOI:
882 10.1016/j.pt.2013.05.001. URL: [https://www.cell.com/trends/](https://www.cell.com/trends/parasitology/abstract/S1471-4922(13)00073-1)
883 [parasitology/abstract/S1471-4922\(13\)00073-1](https://www.cell.com/trends/parasitology/abstract/S1471-4922(13)00073-1) (visited on 11/05/2019).
- 884 [60] Ziheng Yang. *Computational molecular evolution*. Oxford series in ecol-
885 ogy and evolution. Oxford: Oxford University press, 2006. ISBN: 978-0-19-
886 856699-1 978-0-19-856702-8.
- 887 [61] Olga Zhaxybayeva and J. Peter Gogarten. “Cladogenesis, coalescence and
888 the evolution of the three domains of life”. In: *Trends in genetics: TIG* 20.4
889 (Apr. 2004), pp. 182–187. ISSN: 0168-9525. DOI: 10.1016/j.tig.2004.02.
890 004.