

1

2 **A geography aware reconciliation method to investigate diversification patterns in host/parasite**
3 **interactions**

4

5 **V. Berry¹, F. Chevenet^{1&2}, J-P. Doyon¹, E. Jousset³**

6 ¹ Institut de Biologie Computationnelle, LIRMM, UMR 5506 CNRS, Univ Montpellier, France

7

8 ² MIVEGEC, CNRS 5290, IRD 224, Univ Montpellier, France

9

10 ³ CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, France

11

12 **Corresponding authors: V. Berry vberry@lirmm.fr and E. Jousset**
13 **emmanuelle.jousset@inra.fr**

14

15 **Running title (45 characters): Geography aware reconciliation**

16

17 **ABSTRACT (250 WORDS)**

18 Cospeciation studies aim at investigating whether hosts and symbionts speciate simultaneously or
19 whether the associations diversify through host shifts. This problem is often tackled through
20 reconciliation analyses that map the symbiont phylogeny onto the host phylogeny by mixing different
21 types of diversification events. These reconciliations can be difficult to interpret and not always
22 biologically realistic. Researchers have underlined that the biogeographic histories of both hosts and
23 symbionts influence the probability of cospeciation and host switches, but up to now no reconciliation
24 software integrates geographic data. We present a new functionality in the *Mowgli* software that
25 bridges this gap. The user can provide geographic information on both the host and symbiont extant
26 and ancestral taxa. Constraints in the reconciliation algorithm have been implemented to generate
27 biologically realistic codiversification scenarios. We apply our method to the fig/fig wasp association
28 and infer diversification scenarios that differ from reconciliations ignoring geographic information. In
29 addition, we updated the reconciliation viewer *SylvX* in order to visualize ancestral character states on
30 the phylogenetic trees and highlight zones that are geographically inconsistent in reconciliations
31 computed without geographic constraints. We suggest that the comparison of reconciliations obtained
32 with and without constraints can help solving ambiguities in the biogeographic histories of the
33 partners. With the development of robust methods in historical biogeography and the advent of next-
34 generation sequencing that leads to better-resolved trees, a geography aware reconciliation method
35 represents a substantial advance that is likely to be useful to researchers studying the evolution of
36 biotic interactions and biogeography.

37

38 **Keywords: Ancestral trait, biogeography, cophylogeny, host/parasite, software, reconciliation,**
39 **tree visualization.**

40 1) INTRODUCTION

41 Biotic interactions play a prominent role in species diversification. Interactions that result into long-
42 term associations persisting over evolutionary time scales can sometimes lead to cospeciation, i.e. the
43 concomitant occurrence of speciation in lineages that are ecologically associated (Brooks 1981; Page
44 1990, 1991). The idea that such a pattern can occur first stemmed from parasitological studies
45 suggesting that parasite classifications reflect the phylogenetic relationships of their hosts (Fahrenholz
46 1913). Hafner & coll. (Hafner & Nadler 1988; Hafner *et al.* 1994) were the first authors to thoroughly
47 test this assertion. They used the association between pocket gophers and their lice as a model system
48 and provided a clear demonstration that the phylogenies of the two interacting lineages were parallel.
49 This study spurred further research on cospeciation. The developments of specific methods that aimed
50 at testing the congruence of the phylogenetic histories of interacting organisms have since played an
51 important role in the study of cospeciation. It is indeed these methods that moved cospeciation studies
52 beyond visual comparisons of phylogenetic trees and ad-hoc narratives for these visualizations. It soon
53 became apparent that the study of the concordance between phylogenetic trees could be applied to
54 reconciling gene trees and species trees (Page & Charleston 1997; Page & Charleston 1998) which
55 further enhanced the interest of evolutionary biologists for methodological developments in this field.

56 Reviews on cospeciation methods (Brooks *et al.* 2004; de Vienne *et al.* 2013; Doyon *et al.*
57 2011; Johnson & Clayton 2004; Martínez-Aquino 2016; Paterson & Banks 2001; Stevens 2004) all
58 emphasize the diversity and the complexity of the scenarios that must be explored when testing for the
59 congruence of speciation events in two interacting lineages. To compare host and parasite phylogenies,
60 Brooks & coll. (Brooks 1981; Brooks & McLennan 1991) first developed a parsimony method (the
61 Brooks Parsimony Analysis, BPA). In this method the associations between hosts and their parasites
62 are transformed into a matrix of host characters and the parsimony tree reconstructed from such a
63 matrix is then compared to the host phylogeny. A decade later, Page and collaborators developed a
64 fundamentally different method, called “tree reconciliation”, a term first coined in the work of
65 Goodman *et al.* (1979) that compared gene and species trees. This method attempts to reconcile the
66 phylogenetic history of the parasite with that of their hosts: the parasite phylogeny is “mapped” onto
67 the host phylogeny (i.e. each node in the parasite tree is assigned to a node or a branch in the host
68 phylogeny). In such a map, the diversification events of the parasites are linked to their host
69 phylogenetic history and four types of events are considered: cospeciation events, host switches,
70 sorting events and duplication events (Page 1994a; Page 1994b) (see material and method for a
71 description of each event). When graphically displayed, reconciliation maps greatly ease our
72 understanding of the evolution of biotic interactions.

73 Algorithms to optimize reconciliations are numerous. One of the first reconciliation software,
74 *TreeMap 2*, uses an algorithm called “Jungles” (Charleston 1998) where each event is assigned a cost:
75 the chosen reconciliations are the ones that have minimum costs. However it generates in the process
76 an exponential number of scenarios. Recent methods have proposed algorithms that are more efficient
77 and can also just search for an optimal reconciliation: e.g. *Tarzan*, (Merkle & Middendorf, 2005);
78 *Jane*, (Conow *et al.* 2010); *Core-PA* (Merkle *et al.*, 2010), *Mowgli*, (Doyon *et al.* 2010), *COALA*
79 (Baudet *et al.* 2015), *ecceTERA* (Jacox *et al.* 2016), *Notung* (Stolzer *et al.* 2012), *EUCALYPT* (Donati
80 *et al.* 2015) and *ILPEACE* (van Iersel *et al.* 2014). Recently, the *RASCAL* software proposed to infer
81 suboptimal scenarios to reduce computing times (Drinkwater & Charleston 2016). Cospeciation is
82 witnessed on a reconciliation map whenever a speciation node in the parasite phylogeny is mapped
83 onto a speciation on the host phylogeny. Another requirement for demonstrating that two interacting
84 lineages have cospeciated is to provide evidence of the temporal congruence of the cospeciation event
85 in the host and parasite phylogenies (Page 1991). Though reconciliation algorithms do not strictly
86 enforce the simultaneity of cospeciation events, they can enforce time consistency in the sequence of

87 evolutionary events, meaning that the parasite cannot switch back in time onto a host that no longer
88 exists (*i.e.* transfers cannot occur towards a node in the host phylogeny that has already split into child
89 species at the time of the transfer event) (Merkle & Middendorf 2005; Nojgaard *et al.* 2017). This
90 constraint is explicit in *Mowgli* (Doyon *et al.* 2011; Doyon *et al.* 2010), *ecceTera* (Jacox *et al.* 2016)
91 and *RASCAL* (Drinkwater & Charleston 2016). Hence, reconciliation methods have greatly improved
92 in the last decade; algorithms are now efficient and some have solved the time consistency issue that
93 affected some of the first methodological developments in the field. However, interpreting the
94 scenarios that emerge from these inferences remains a difficult task. It is generally challenging to
95 identify biologically realistic reconciliations. Much remains to be done to improve these inferences
96 and translate them into evolutionary scenarios that give insights into the biological factors that govern
97 the evolution of interspecific associations.

98 Some key information that could significantly improve our inferences but are overlooked in
99 codiversification methods are the geographic locations of extant and ancestral nodes. Indeed, the
100 biogeographic histories of interacting lineages necessarily constrain their common part of evolutionary
101 history (Martinez-Aquino *et al.* 2014; Nieberding *et al.* 2010). Obviously, a cospeciation event can
102 only happen between taxa that co-occur in the same area.. The geographic context of both hosts and
103 parasites also influences host switch events. In biotic interactions where the parasites can undergo long
104 dispersal events, transfers can happen between allopatric hosts (*i.e.* hosts that do not live in the same
105 geographic area). However, they are only possible if the geographic locations of the “sending host”
106 (the host from which the switch is initiated) and the “receiving host” coincide with a dispersal event
107 along the corresponding branch in the parasite phylogeny. Therefore, a more accurate mapping of
108 cospeciation and host switch events can be obtained if the geographic locations of both hosts and
109 parasites are known prior to conducting the reconciliation.

110 Methods for inferring historical biogeography from phylogenetic reconstructions have greatly
111 improved in the last two decades. Early developments in historical biogeography aimed at
112 reconstructing “area cladograms” that reflected the history of connections between areas of endemism
113 for the group of organisms under study and used analytical tools that were very similar to the tools
114 developed for the study of cospeciation using parsimony as the optimization criterion (*e.g.*, BPA, see
115 Morrone, 2009 for a review on cladistic biogeography and its methodological developments). More
116 recent probabilistic methods in the field of historical biogeography aim at reconstructing ancestral
117 geographic range of focal lineages from current species distribution and a dated phylogenetic tree.
118 They model the evolution of geographic areas on a phylogenetic tree using Maximum Likelihood
119 optimization or Bayesian inference and incorporate divergence times into the inference process: the
120 longer the phylogenetic branch, the higher the probability of geographic range shifts and the larger the
121 uncertainty in the ancestral range estimates. Geographic areas can be treated as simple categorical
122 characters that are reconstructed on the tree using for instance a stochastic Markov model of evolution.
123 More biologically realistic and widely applied methods in historical biogeography, such as DEC
124 (*Dispersal, Extinction, Cladogenesis*) (Ree *et al.* 2005; Ree & Smith 2008), model range evolution
125 using different parameters for each biogeographic process (dispersal, range expansion or extinction).
126 In addition to modelling these key processes in range evolution, the main innovation of DEC consists
127 in incorporating a time-dependent transition matrix that defines the movements between geographic
128 areas, at different time intervals in order to reflect how dispersal opportunities changed through time
129 (*e.g.* changes in the continents configuration for instance) (see Ree & Sanmartin 2009; Ronquist *et al.*
130 2011, for reviews on parametric biogeography). Fossil distribution and information on the climatic
131 preferences of ancestral lineages can also be incorporated as constraints to improve biogeographic
132 inferences (Meseguer *et al.* 2015). Several conceptual and computational improvements have been

133 implemented since the initial version of DEC (DEC + J, Matzke 2014; DECX, Beeravolu Reddy &
134 Condamine 2016). Different biogeographic models have also been proposed (GeoSSE, Goldberg *et al.*
135 2011; BayArea Landis *et al.* 2013). As a result, robust biogeographic scenarios are now available for
136 numerous lineages. Ancestral areas inferred by these methods can then serve as input for reconciliation
137 analyses. In this paper we build on these advances to provide a geography-aware reconciliation
138 method, pushing further the realism of scenarios proposed by such methods.

139 We first describe the constraints we enforce to ensure geographic consistency in
140 reconciliations and how they were implemented in the *Mowgli* reconciliation software (Doyon *et al.*
141 2010). We also updated the *SylvX* reconciliation viewer (Chevenet *et al.* 2016), in order to integrate
142 and visualize annotations (e.g. geographic areas) at ancestral nodes for the host and parasite
143 phylogenies and highlight inconsistent zones in the reconciliation. We then test these new
144 developments on a mock dataset and on a ‘textbook’ example of cospeciation, namely the interaction
145 between figs (*Ficus*) and their pollinating fig wasps (Cruaud *et al.* 2012; Rønsted *et al.* 2005; Wiebes
146 1979).

147 2) METHODS

148 **Extending *Mowgli* to account for geographic information**

149 In this section we first recall the reconciliation model followed by *Mowgli* (Doyon *et al.*
150 2010).

151 Only rooted parasite and host trees are considered; their leaf nodes (tips) are each labelled by a taxon
152 name. The host tree is dated, meaning that either each branch length represents an amount of time (the
153 tree is thus ultrametric) or that the age of each internal node is provided (e.g. in million years). Internal
154 nodes usually have two descendants, but an internal node can also have a single child also when the
155 evolution of an ancestral lineage living a relatively long period of time is decomposed into a set of
156 consecutive time periods called *slices* (see Fig. 3 of *Mowgli* Manual). This slicing of branches is a
157 transparent artefact that allows reconciliation methods to achieve fast computing times while still
158 ensuring time consistency of host switches (see Doyon *et al.* 2010; Jacox *et al.* 2016; Libeskind-Hadas
159 & Charleston 2009).

160 Let P and H denote respectively a parasite and a host tree, x and x_p are nodes (or extant
161 species, *i.e.*, leaves) of H and u and u_p are nodes (or extant species) of P . Reconciliation algorithms
162 usually consider each current and ancestral host to be associated with one or several specific parasites
163 at any time (*e.g.* in *Mowgli*, *TreeMap*, *Jane*). However, the identity of the host can vary over time, *e.g.*
164 after a *host switch*. This evolutionary event is one of the four types of *events* considered in
165 cospeciation studies:

166 - a *host switch*, also known as a transfer (T event), occurs when a parasite lineage from a
167 source host is transferred to a destination host. The transfer of the parasite must be time consistent,
168 that is the “sending” branch (x_p, x) and the “receiving” branch (x'_p, x'), where the host switch is
169 mapped, must belong to the same time slice;

170 - a *cospeciation* (S event) happens when the speciation of a parasite shortly follows or
171 coincides with the speciation of its host. This is considered by *Mowgli* as a joint speciation of both
172 parasite and its host;

173 - a *within host speciation* also known as duplication (D event), models a speciation of a
174 parasite u of P , where both descendant species continue to live on the host that u lived on. This is
175 represented by u evolving along a (x_p, x) branch of H and then splitting into two new lineages in (x_p, x) ;

176 - a parasite *loss* (L event) occurs when a parasite lineage goes extinct while its host persists.

177 An illustration of these events can be found the *Mowgli* Manual.

178 *Mowgli* also sometimes considers combinations of events in order to speed up computations. A SL
179 event occurs shortly after a cospeciation (S): one of the parasite child lineages is quickly lost (L) in the
180 host phylogeny on a child lineage of the involved speciation node. A TL event occurs when a parasite
181 u evolving on a branch (x_p, x) is lost (L) on this branch shortly after having switched (T) to another host
182 (x'_p, x') .

183 As explained above, accounting for geographic information can lead to more realistic diversification
184 scenarios. We first integrate such information by assigning a set of *areas* to each node of P and H . For
185 an extant taxon this means that a population of the corresponding species is reported to live in *each* of
186 the assigned areas. In contrast, when an internal taxon is assigned to one or several areas, this means
187 that populations of this now extinct taxon are inferred to have lived back in time in one or several of
188 these geographical zones.

189 In order to compute biogeographically meaningful reconciliations between the P and H trees, specific
190 constraints have to be implemented in reconciliation algorithms. We detail below how we model these
191 constraints in the context of the four D/T/L/S events or combinations thereof. First, note that areas of a
192 node and its parent in the host or parasite tree can be different, due to dispersal and vicariance
193 events. During the reconciliation process, the time period represented by a branch between nodes x_p
194 and x of the host tree is considered to be assigned the union of the areas of x_p and x . If a species
195 changes area along the branch from one area assigned to x_p to a different one assigned to x , we do not
196 know exactly when it happened, so we consider that at any time between x_p and x , part of the
197 population of the evolving species can live in any area proposed for x_p or x .

198 Considering nodes of the trees, we denote by $area(x)$ the set of geographic areas where an extant
199 species x is observed (at the tip of a tree). Areas proposed for an internal node x , that is for an extinct
200 species, are also denoted $area(x)$. However, as indicated above, the meaning is somewhat different as
201 $area(x)$ represents in this case the set of areas where x could have lived. Because of the incertitude in
202 the historical biogeography inferences, we do not enforce that x lived in each of these areas. Similarly,
203 considering branches (x_p, x) of the H tree, $area(x_p, x)$ denotes the set of areas where the species might
204 have lived during this period: this is the union of $areas(x_p)$ and $area(x)$. Note that each area in which
205 exactly one of the two species x and x_p is present corresponds to a migration or extinction event that
206 has occurred along this branch. In addition, only (x_p, x) branches being one slice higher are considered
207 for H , as *Mowgli* operates on this level of detail.

208 We now detail which geographic constraints apply so that the reconciliation between a parasite tree P
209 and a host tree H is geographically consistent. Recall that a reconciliation is a mapping of P 's nodes
210 and branches onto those of H .

211 - An extant parasite u can be mapped onto an extant host x , only if $area(u) \subseteq area(x)$ (Fig. 1 A). If this
212 constraint is not fulfilled then *Mowgli* cannot compute a reconciliation.

213 - We allow the mapping of an ancestral parasite u at a speciation node x in the host tree, only if
214 $area(u)$ and $area(x)$ have a non-empty intersection, *i.e.* when there is at least one area where the
215 parasite and the host were able to meet (Fig. 1 B).

216 -A parasite node u can be mapped into a branch (x_p, x) of H due to a duplication or host switch event
217 (Fig. 1 C), and in those cases, we also require that $area(u) \cap area(x_p, x) \neq \emptyset$. Note, that this constraint
218 does not prevent parasite dispersal events during host switches.

219 - If a branch (u_p, u) of the parasite tree is mapped for all or part of it onto a host branch (x_p, x) (Fig. 1
220 D), then we also require that $area(u_p, u) \cap area(x_p, x) \neq \emptyset$.

221 - Last, if a branch (u_p, u) of the parasite is going through a node x of the host tree (which happens when
222 the host speciates into two descendant hosts but the parasite sticks to only one of them – an SL event),
223 then the $area(x)$ and $area(u_p, u)$ must have common elements (Fig. 1 E).

224 Note that when part of the reconciliation mapping traverses an artificial node x in H , then no particular
225 constraint applies locally: the possibility of such a scenario is directed by constraints ensured with
226 respect to the branch (x_p, x) of H to which x belongs.

227 When respecting the above constraints, *Mowgli* will propose a scenario that is geographically
228 consistent. This scenario can have a higher cost than those obtained when not accounting for
229 geographic information. This simply results from the fact that the search space contains geographically
230 inconsistent scenarios that are possibly less costly. *Mowgli*'s extension described above, allows
231 choosing the less costly scenario among those that are geographically consistent.

232 ***SylvX*'s new functionalities**

233 We extended the *SylvX* editor in order to visualize current and ancestral geographic areas of hosts and
234 symbionts. Pie charts can be used to display alternative areas for each node of the tree and/or the
235 reconciliation. Area colour sets can be dynamically updated and tuned using the Hue, Saturation and
236 Value scales. Thresholds are available to simplify views. *SylvX* also contains a new tool in the
237 *Annotation* panel to highlight reconciliation parts that do not respect geographical constraints (when
238 such constraints have not been enforced when computing the reconciliation). This is done by loading
239 an annotation file generated by *Mowgli* (constraintsPBM.csv).

240 **Implementation**

241 *Mowgli* takes as input a “host tree” and a “parasite tree” stored in files in a Newick format. A list of
242 nodes with their geographic areas (or other annotations) can be given in the same files. Biogeographic
243 inferences typically generate probability or likelihood values for each character state (area) at each
244 node. *Mowgli* can accept a single area or a set of areas at each node. To run *Mowgli* and obtain a
245 reconciliation respecting geographical constraints, the `-a` flag must be added in the command
246 launching the program. Adding the `-y` flag instead computes a reconciliation independently of the
247 indicated constraints but pinpoints the places where the mapping violates these constraints (in
248 `mapping.mpr` and `constraintPBMs.csv` files, see the provided manual for details). This
249 allows users to identify inconsistencies between the most parsimonious reconciliations and the hosts
250 and parasites respective biogeographic histories.

251 *SylvX* takes a host tree in Newick format with node *id* numbers and a reconciliation (with symbiont
252 tree node *id*). The host tree (`outputSpeciesTree.mpr`) and reconciliation obtained with *Mowgli*
253 can be directly imported into *SylvX*. The latter also supports input files from other reconciliation

254 software, *e.g.*, ecceTERA (Jacox *et al.* 2016). Annotation files for the host and parasite phylogenies
255 giving node information can be imported in a CSV format. As many annotations as needed can be
256 added in the annotation files and it is up to the user to choose which ones to plot onto the species tree
257 and the reconciliation map through *SylvX*'s interface.

258 In order to seamlessly pass a user annotation file in `csv` format into both *Mowgli* and *SylvX*, we
259 provide a *Perl* script that can be run through the command line in order to: 1) obtain tree node
260 identifiers that will be used by both programs and 2) merge input trees and corresponding annotations
261 files into *Mowgli*'s input format. Files can be generated so that a single (most likely) ancestral range
262 can be specified or alternative geographic areas can be assigned to all nodes (see Supplementary
263 Material 1 for a description of the full procedure to generate files, set a threshold value above which to
264 keep alternative areas and perform a complete analysis).

265 3) WORKED EXAMPLE

266 **Datasets**

267 To demonstrate the method and its utility, we tested it on two datasets. We first created a mock
268 dataset: two phylogenetic trees with nine tips for a hypothetical host/parasite interaction in which
269 extant and ancient geographic areas for each lineage are informed. The dataset was generated by hand
270 so that: 1) present-day geographic areas of associated taxa are consistent (*i.e.*, hosts and associated
271 parasites live in the same area); 2) the two phylogenies are not perfectly parallel but show some
272 cospeciation events; 3) some geographic locations at nodes that we would like to cospeciate do not
273 coincide in the parasite and host phylogenies. We ran *Mowgli* on this dataset successively with and
274 without enforcing geographic constraints using in both cases the default parameters (cost 0 for a
275 cospeciation, 1 for a loss and 1 for a host switch, 1 for duplication, not enforcing the root of the
276 parasite tree to be mapped on the root of the species tree). In order to measure the impact of cost
277 settings on the reconciliation scenarios, we ran this dataset using alternative costs for host switches
278 and losses.

279 As a second dataset, we used a subset of the data from the latest phylogenetic investigation of
280 figs (*Ficus*) and their pollinating wasps (Cruaud *et al.* 2012). For both partners of the association,
281 biogeographic scenarios were available for phylogenies of 200 taxa. From the complete phylogenetic
282 trees (available in <http://datadryad.org>, doi: 10.5061/dryad.hr620), we derived two trees of 23 taxa
283 each, that included a couple of representative species for each *Ficus* main taxonomic subdivision. We
284 excluded one of the fig subgenera (*Pharmacosycea*) and its associated pollinators (*Tetrapus* spp.)
285 whose phylogenetic positions are still debated. We have not tested our method on the total dataset
286 presented in Cruaud *et al.* (2012) as some uncertainties remain concerning the root of the phylogenetic
287 trees, which could lead to spurious interpretations. The most likely ancestral geographic areas of each
288 node were directly derived from the biogeographic reconstructions of Cruaud *et al.* 2012, obtained
289 with Maximum Likelihood Optimization in Mesquite (Maddison & Maddison 2006). We ran *Mowgli*
290 on this dataset with and without enforcing geographic constraints (using default event costs and not
291 enforcing the root of the parasite tree to be mapped on the root of the species tree), and explored how
292 these reconstructions shed light on the biogeographic history of the association. In order, to investigate
293 how incertitude on ancestral geographic ranges impacts the reconciliation, we ran the reconciliation on
294 the dataset including alternative ancestral areas for both *Ficus* and their associated pollinators. For
295 each node of the pollinator and the *Ficus* phylogenies, the geographic areas which proportional
296 likelihood was above 0.15 were kept and assigned to their respective nodes.

297 **Results**

298 Figure 2 represents the reconstruction obtained on the mock dataset. When not taking
299 geographic constraints into account (Fig. 2A), a cospeciation event at a node where the two associates
300 do not live in the same area was retrieved (node S1 of the host tree in Fig. 2A). The transfer T1
301 preceding this cospeciation event is also geographically impossible as it suggests a dispersal (the
302 donor host lives in Asia or Africa, and the receiving host lives in America) while the parasite actually
303 stays in Africa. The scenario obtained when enforcing geographic constraints is more costly (Fig. 2B):
304 it entails one more transfer and consequently one less cospeciation event but is biologically more
305 realistic. When using different cost vectors (i.e. using a cost of 3 for losses), the reconciliation where
306 geographic constraints are not taken into account includes additional transfers to avoid losses (Fig.
307 S2A); those are all geographically inconsistent. The reconciliation with geographic constraints also
308 changes (Fig. S2C) and necessitates 5 transfers to ensure geographic consistency of the diversification
309 events in hosts and parasites without inferring any parasite losses. When we increased the cost of
310 transfers (cost T=10, Fig. S2B), the reconciliation without geographic constraints infers several early
311 duplications and losses in order to avoid a costly transfer. On the other hand the results of the
312 reconciliation under constraints (Fig. S2D) did not change comparatively to the one obtained with
313 default cost settings. Hence, in this particular case, adding biological constraints into the reconciliation
314 process stabilizes the reconciliation and makes it less dependent on cost settings.

315 On the fig/fig wasp dataset (Fig. 3), not accounting for geographical constraints leads to
316 geographic inconsistency in one node (cospeciation S1 in the host tree of Figure 3A). The transfer that
317 precedes it (T5) is impossible and the mapping of the fig wasp phylogeny onto the fig phylogeny from
318 node S1 to event T6 is geographically inconsistent ((Fig. 3A). Enforcing geographic constraints when
319 a single (most likely) area is specified for each node generates a reconciliation scenario that is more
320 costly (Fig. 3B, one more transfer is necessary to reconcile the two phylogenies) but coherent with the
321 figs and the fig wasp biogeographic histories. This scenario suggests that the fig wasps independently
322 colonized figs in the Neotropics and in the Afrotropics through two distinct host switches from Asia
323 rather than accompanied the speciation of their hosts, as was suggested by Figure 3A (and node 29 of
324 Fig. S12 in Cruaud *et al.* 2012). The annotation of ancestral geographic areas on the reconciliation
325 map also shows that host switches occurred both in “sympatric” settings (within the same geographic
326 areas as broadly defined in our dataset) and allopatric settings (i.e. host switches occur between two
327 geographically distant hosts). Overall, four switches out of seven occurred in sympatry (T1, T2, T4,
328 T5) while the remaining three switches (T3, T6, T7) correspond to long distance dispersal events (Fig.
329 3B). Adding uncertainty in ancestral geographic range, generates a reconciliation that matches the one
330 obtained without constraint (Fig. 3C), as geographic areas of node S1 of the host figs now includes
331 Asia among its potential geographic areas. This matches the ancestral geographic area of the inferred
332 associated fig wasps. In that scenario a single host switch is associated with a long dispersal event of
333 the fig wasps (T3: from Asia to Australasia), all other host switches occur in sympatric settings
334 (within Asia) and fig wasp geographic range evolution merely mirrors the one of their hosts.

335 4 DISCUSSION

336 We provide here significant extensions for a reconciliation tool (*Mowgli*) and a visualization
337 tool (*Sylvx*) to infer co-diversification scenarios that, for the first time, can take the historical
338 biogeographies of the associated lineages into account. The extension of the *Mowgli* software
339 precludes geographic inconsistency during the reconciliation process. The resulting reconciliations can
340 then be visualized and edited in the *SylvX* updated graphical interface that now integrates annotations
341 of ancestral nodes. *Mowgli* is already one of the few time-consistent efficient methods that build
342 optimal reconciliations. With the integration of geographic constraints in its algorithm, this tool now
343 provides more realistic codiversification scenarios than other reconciliation methods. Producing

344 biologically realistic scenarios can ease their interpretation. In addition, geography-aware
345 reconciliations can reveal whether host switches occur in sympatry or whether they are associated with
346 dispersal events of the symbionts/parasites: this helps unravelling the evolutionary processes
347 underlying host switches

348 In the particular example of the fig/fig wasp interaction presented here, the geographic
349 inconsistency revealed at one of the cospeciating nodes in the analysis (Fig 3A) ran without
350 constraints might actually point out some ambiguity in the biogeographic history of the *Ficus* hosts.
351 According to the inference conducted in Cruaud *et al.* (2012) the most likely area for the common
352 ancestor of Neotropical figs (belonging to the *Americana* section) and Afrotropical figs (belonging to
353 the *Galoglychia* section) is Africa while the proposed cospeciating pollinators lived in Asia (S1; Fig.
354 3A). In order to respect geographic consistency (when only the most likely area is kept for each
355 ancestral species, Fig. 3B), our geography-aware reconciliation suggests that the current association of
356 figs wasps with *Galoglychia* in Africa, resp. *Americana* in the Neotropics, is the result of two
357 independent switches (Fig. 3B, events T6 and T7) of the pollinators from an Asian fig ancestor (the
358 ancestor of the *Conosycea* figs). However, the biogeographic analysis of Cruaud *et al.* (2012) also
359 suggested that the node S1 of *Ficus* could be situated in Asia (though with a much lower likelihood
360 than the Afrotropics). When specifying alternative geographic areas (Fig. 3C), including Asia for the
361 conflicting node in the *Ficus* phylogeny, we obtain a reconciliation that matches the one obtained
362 without constraints (therefore entailing one less transfer and one more cospeciation event). This result
363 suggests that the common ancestor of the African figs of section *Galoglychia* and the new world figs
364 from the section *Americana* could have been located in Asia. Under this latter scenario most of the
365 host switches observed happen in sympatric settings. We will not conclude on the biogeographic
366 history of the fig/fig wasp association as the purpose of our study is not to explore alternative
367 scenarios for this association. The above discussion mainly demonstrates the utility of our method in
368 revealing inconsistency between biogeographic scenarios and a cospeciation hypothesis and therefore
369 proposing alternative scenarios that conciliate both. As in all ancestral character state inferences that
370 rely on present day data, biogeographic reconstructions entail some incertitude. In particular, they are
371 highly sensitive to missing data (species that have not been sampled and/or extinct species). It is
372 therefore important to compute reconciliations with alternative ancestral ranges to investigate
373 biogeographic scenarios.

374 **Perspectives**

375 The tools developed in this study can be applied to all interspecific interactions for which
376 biogeographic scenarios are available for both partners. Fast developments in sequencing technologies
377 generate more accurate and more exhaustive phylogenies and methods in historical biogeography have
378 also improved. Therefore, we can hope that numerous datasets will be available in the near future and
379 that cospeciation could be tested on more systems (Cruaud & Rasplus 2016). For instance, robust
380 phylogenies and biogeographic scenarios are now available for groups of lice that have been model
381 systems in coevolutionary studies (Boyd *et al.* 2017). Once comprehensive phylogenies of the hosts
382 are available, our method could be used to better understand the geographic context of host switches in
383 this model system. Geography-aware reconciliation could also be applied to explore the diversification
384 history of the numerous parasitic wasps that are part of the microfauna exploiting figs: several lineages
385 of parasitic wasps have been shown to partly cospeciate with their host figs (Jousselin *et al.* 2008;
386 Jousselin *et al.* 2006) and biogeographic scenarios for some lineages are available (Cruaud *et al.*
387 2011). These developments could also be applied to specific sections of the genus *Ficus* in order to
388 shed light on their complex biogeographic histories (e.g. section *Urostigma* that has experienced
389 several dispersal events between Africa and Asia, Chantarasuwan *et al.* 2016). Other nursery

390 pollination/mutualisms such as the interaction between *Yucca* and their pollinating moths are also
391 good candidates for including geographic constraints into coevolutionary scenarios, as some studies
392 have questioned the respective role of geography and host-plant association in driving the
393 diversification of *Yucca* moths (Althoff *et al.* 2012). Plant/pollinator systems (Hutchinson *et al.* 2017),
394 parasitoid/host insect associations (Deng *et al.* 2013; Wilson *et al.* 2012) , herbivorous insect/plant
395 interactions (*e.g.* McLeish *et al.* 2007; Percy *et al.* 2004) and various vertebrate/parasite associations
396 (*e.g.* Badets *et al.* 2011; Bentz *et al.* 2006; Weckstein 2004) for which researchers have investigated
397 the relative role of geography and biotic interactions in shaping cophylogenetic signals could also be
398 studied.

399 Furthermore, the approach presented in this paper does not only apply to geographic
400 information and could be extended to other biological traits. For instance in systems where the species
401 are partitioned into different habitats (*e.g.* forest canopy species vs savannah species), geographic
402 areas could be replaced by traits related to the ecological niches; constraints that are similar to the ones
403 applied for geography could then be easily transferrable. Informing ancestral characters for habitats on
404 the host and the symbiont phylogenetic trees and using “*Mowgli* with constraints” would result in
405 constraining cospeciation and host switches to associates sharing the same ecological habitats. In a
406 similar way, the respective climatic niches of associated organisms could also be used when parasite
407 (or symbiont) distributions are known to be strongly constrained by thermal tolerance (see Singh *et al.*
408 2017, for a recent study showing that climatic conditions influences the patterns of association
409 between fungi and their algal partners). In many specialized interactions, such as host/obligate
410 bacterial endosymbionts (*e.g.* Jousselin *et al.* 2009, Rosenblueth *et al.* 2012) or host/viruses
411 associations (Ramsden *et al.* 2009; Garamszegi 2009), inferring ancestral character states for some
412 ecological traits for the “symbiotic” lineages (the parasite) independently of their hosts is not always
413 straightforward. However, the evolution of these obligate associations and their maintenance are still
414 governed by some phenotype matching between the partners. For instance in host/bacterial symbiont
415 associations, the metabolic complementarity of the host and the symbiont (Zientz *et al.* 2004) could be
416 reconstructed and used to constrain the reconciliations. In host/virus associations, information about
417 the host immune system and viruses adaptations could be used (Longdon *et al.* 2014). The extension
418 of *Mowgli* proposed here could probably be adapted to fit the biological properties of these
419 associations

420 Independently of the new functionality implemented in *Mowgli*, the concomitant update of
421 *SylvX* allows the comparison of ancestral states for any character of the hosts and/or the symbionts.
422 This can help interpreting reconciliations by replacing them in their biological context. One of the
423 most useful functionalities of *SylvX* is now to be able to visualize whether host switches are associated
424 with evolutionary transitions in character states in both the parasite and/or the host. It can therefore
425 help understanding the biological processes that are associated with these transfers. Mapping
426 characters of the associates throughout the reconciliation can also help investigating whether there is
427 correlated trait evolution in host and parasites. Until now, such correlations could only be investigated
428 on one of the associate phylogeny (*e.g.* Sorci *et al.* 2003; Jousselin *et al.* 2003). Looking at
429 simultaneous transitions in character states in both partners throughout a host/parasite reconciliation
430 might help identifying co-adapted traits that constrain the association.

431 In conclusion, we provide here a framework that can integrate the character histories of the
432 associates into the reconciliation process. It can take into account incertitude in the character states and
433 allows recovering biologically realistic scenarios. It can also shed light on character history inferences
434 by pointing out inconsistencies between the character states of the two associates on the reconciliation
435 map. The new developments made in *SylvX* facilitate these interpretations. A more integrative

436 approach than the one presented here, would consist in co-optimizing the reconciliation and the
437 biogeographical inference simultaneously. However this would require using the same optimization
438 criterion for both inferences and setting adequate parameters for these very different processes in a
439 single model. When conducted, this work should probably rely on Maximum Likelihood optimization
440 as in the ALE reconciliation software (Szöllösi *et al.* 2012). For now, we believe that the use of
441 “constraint-aware” reconciliations is preferable to current practices that consist in elaborating ad-hoc
442 narratives once the reconciliations are obtained and compared with the character histories of the
443 associates.

444 **ACKNOWLEDGMENTS**

445 *Funding:* ANR PhyloSPACE. We thank A. Cruaud, JY Rasplus & coll. for sharing the fig/fig-
446 wasp dataset.

447 **AUTHOR CONTRIBUTION**

448 VB, JPD and EJ designed the study. VB and JPD developed *Mowgli*, FC developed *SylvX*. VB and EJ
449 wrote the manuscript with contributions of FC.

450 **SOFTWARE AVAILABILITY**

451 *Mowgli* is available on <http://www.atgc-montpellier.fr/Mowgli/> with manual, *GeoRecHelper* and
452 example files. It runs on OSX (Mac) and Linux systems.

453 *SylvX* is available on www.sylvx.org with manual and example files and can be installed on any
454 platforms.

455

456 **REFERENCES**

- 457 Althoff DM, Segraves KA, Smith CI, Leebens-Mack J, Pellmyr O (2012) Geographic isolation trumps
458 coevolution as a driver of yucca and yucca moth diversification. *Molecular Phylogenetics and*
459 *Evolution* **62**, 898-906.
- 460 Badets M, Whittington I, Lalubin F, *et al.* (2011) Correlating Early Evolution of Parasitic
461 Platyhelminths to Gondwana Breakup. *Systematic Biology* **60**, 762-781.
- 462 Baudet C, Donati B, Sinimeri B, *et al.* (2015) Cophylogeny Reconstruction via an Approximate
463 Bayesian Computation. *Systematic Biology* **64**, 416-431.
- 464 Beeravolu Reddy C, Condamine F (2016) An Extended Maximum Likelihood Inference of
465 Geographic Range Evolution by Dispersal, Local Extinction and Cladogenesis. *BioRxiv*.
- 466 Bentz S, Sinnappah-Kang ND, Lim L-HS, *et al.* (2006) Historical biogeography of amphibian
467 parasites, genus *Polystoma* (Monogenea: Polystomatidae). *Journal of Biogeography* **33**, 742-
468 749.
- 469 Boyd BM, Allen JM, Nguyen N-P, *et al.* (2017) Phylogenomics using Target-Restricted Assembly
470 Resolves Intrageneric Relationships of Parasitic Lice (Phthiraptera: Columbicola). *Systematic*
471 *Biology*, **66**, 896-911.
- 472 Brooks D (1981) Hennig's parastological method: a proposed solution *Systematic Zoology* **30**, 229-
473 249.
- 474 Brooks D, McLennan DE (1991) *Phylogeny, Ecology, and Behavior*, 441 p., University of Chicago
475 Press.

- 476 Brooks DR, Dowling APG, van Veller MGP, Hoberg EP (2004) Ending a decade of deception: a
477 valiant failure, a not so-valiant failure, and a success story. *Cladistics* **20**, 32-46.
- 478 Chantarasuwan B, Rønsted N, Kjellberg F, Sungkaew S, van Welzen PC (2016) Palaeotropical
479 intercontinental disjunctions revisited using a dated phylogenetic hypothesis with nearly
480 complete species level sampling of *Ficus* subsect. *Urostigma* (Moraceae). *Journal of*
481 *Biogeography* **43**, 384-397.
- 482 Charleston MA (1998) Jungles: a new solution to the host/parasite phylogeny reconciliation problem.
483 *Mathematical Biosciences* **149**, 191-223.
- 484 Chevenet F, Doyon JP, Scornavacca C, *et al.* (2016) SylvX: a viewer for phylogenetic tree
485 reconciliations. *Bioinformatics* **32**, 608-613.
- 486 Conow C, Fielder D, Ovidia Y, Libeskind-Hadas R (2010) Jane: a new tool for the cophylogeny
487 reconstruction problem. *Algorithms for Molecular Biology* **5**.
- 488 Cruaud A, Jabbour-Zahab R, Genson G, *et al.* (2011) Out of Australia and back again: the world-wide
489 historical biogeography of non-pollinating fig wasps (Hymenoptera: Sycophaginae). *Journal*
490 *of Biogeography* **38**, 209-225.
- 491 Cruaud A, Rasplus J-Y (2016) Testing cospeciation through large-scale cophylogenetic studies.
492 *Current Opinion in Insect Science* **18**, 53-59.
- 493 Cruaud A, Ronsted N, Chantarasuwan B, *et al.* (2012) An Extreme Case of Plant-Insect
494 Codiversification: Figs and Fig-Pollinating Wasps. *Systematic Biology* **61**, 1029-1047.
- 495 de Vienne DM, Refregier G, Lopez-Villavicencio M, *et al.* (2013) Cospeciation vs host-shift
496 speciation: methods for testing, evidence from natural associations and relation to coevolution.
497 *New Phytologist* **198**, 347-385.
- 498 Deng J, Yu F, Li H-B, *et al.* (2013) Cophylogenetic relationships between *Anicetus* parasitoids
499 (Hymenoptera: Encyrtidae) and their scale insect hosts (Hemiptera: Coccidae). *BMC*
500 *Evolutionary Biology* **13**, 275.
- 501 Donati B, Baudet C, Sinaimer B, Crescenzi P, Sagot M-F (2015) EUCALYPT: efficient tree
502 reconciliation enumerator. *Algorithms for Molecular Biology* **10**, 3.
- 503 Doyon J-P, Ranwez V, Daubin V, Berry V (2011) Models, algorithms and programs for phylogeny
504 reconciliation. *Briefings in Bioinformatics* **12**, 392-400.
- 505 Doyon JP, Scornavacca C, Gorbunov KY, *et al.* (2010) An Efficient Algorithm for Gene/Species
506 Trees Parsimonious Reconciliation with Losses, Duplications and Transfers. In: *Comparative*
507 *Genomics* (ed. Tannier E), pp. 93-108.
- 508 Drinkwater B, Charleston MA (2016) RASCAL: A Randomized Approach for Coevolutionary
509 Analysis. *Journal of Computational Biology* **23**, 218-227.
- 510 Fahrenholz H (1913) Ectoparasiten und Abstammungslehre. *Zool. Anz.* **41**, 371-374.
- 511 Garamszegi LZ (2009) Patterns of co-speciation and host switching in primate malaria parasites.
512 *Malaria Journal* **8**, 110.
- 513 Goldberg EE, Lancaster LT, Ree RH (2011) Phylogenetic Inference of Reciprocal Effects between
514 Geographic Range Evolution and Diversification. *Systematic Biology* **60**, 451-465.
- 515 Goodman M, Czelusniak J, Moore GW, Romero-Herrera AE, Matsuda G (1979) Fitting the Gene
516 Lineage into its Species Lineage, a Parsimony Strategy Illustrated by Cladograms Constructed
517 from Globin Sequences. *Systematic Zoology* **28**, 132-163.
- 518 Hafner MS, Nadler SA (1988) Phylogenetic trees support the coevolution of parasites and their hosts.
519 *Nature* **332**, 258-259.
- 520 Hafner MS, Sudman PD, Villablanca FX, *et al.* (1994) Disparate rates of molecular evolution in
521 cospeciating hosts and parasites. *Science* **265**, 1087-1090.
- 522 Hutchinson MC, Cagua EF, Stouffer DB (2017) Cophylogenetic signal is detectable in pollination
523 interactions across ecological scales. *Ecology* **98**, 2640-2652.

- 524 Jacox E, Chauve C, Szollosi GJ, Ponty Y, Scornavacca C (2016) ecceTERA: comprehensive gene
525 tree-species tree reconciliation using parsimony. *Bioinformatics* **32**, 2056-2058.
- 526 Johnson KP, Clayton DH (2004) Untangling coevolutionary history. *Systematic Biology* **53**, 92-94.
- 527 Jousselin E, Desdevises Y, Coeur d'acier A (2009) Fine-scale cospeciation between *Brachycaudus* and
528 *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in
529 aphids. *Proceedings of the Royal Society B: Biological Sciences* **276**, 187-196.
- 530 Jousselin E, Rasplus JY, Kjellberg F (2003) Convergence and coevolution in a mutualism: Evidence
531 from a molecular phylogeny of *Ficus*. *Evolution* **57**, 1255-1269.
- 532 Jousselin E, van Noort S, Berry V, *et al.* (2008) One Fig to Bind Them All: Host Conservatism in a
533 Fig Wasp Community Unraveled by Cospeciation Analyses among Pollinating and
534 Nonpollinating Fig Wasps. *Evolution* **62**, 1777-1797.
- 535 Jousselin E, Van Noort S, Rasplus JY, Greeff JM (2006) Patterns of diversification of Afrotropical
536 Otiteselline fig wasps: phylogenetic study reveals a double radiation across host figs and
537 conservatism of host association. *Journal of Evolutionary Biology* **19**, 253-266.
- 538 Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP (2013) Bayesian Analysis of Biogeography when
539 the Number of Areas is Large. *Systematic Biology* **62**, 789-804.
- 540 Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM (2014) The Evolution and Genetics of
541 Virus Host Shifts. *Plos Pathogens* **10**, e1004395.
- 542 Maddison WP, Maddison DR (2006) Mesquite: a modular system for evolutionary analysis.
- 543 Martinez-Aquino A, Ceccarelli FS, Eguiarte LE, Vazquez-Dominguez E, de Leon GPP (2014) Do the
544 Historical Biogeography and Evolutionary History of the Digenean *Margotrema* spp. across
545 Central Mexico Mirror Those of Their Freshwater Fish Hosts (Goodeinae)? *Plos One* **9**.
- 546 Matzke NJ (2014) Model Selection in Historical Biogeography Reveals that Founder-Event Speciation
547 Is a Crucial Process in Island Clades. *Systematic Biology* **63**, 951-970.
- 548 McLeish MJ, Crespi BJ, Chapman TW, Schwarz MP (2007) Parallel diversification of Australian gall-
549 thrips on *Acacia*. *Molecular Phylogenetics and Evolution* **43**, 714-725.
- 550 Merkle D, Middendorf M (2005) Reconstruction of the cophylogenetic history of related phylogenetic
551 trees with divergence timing information. *Theory in Biosciences* **123**, 277-299.
- 552 Merkle D, Middendorf M, Wieseke N (2010) A parameter-adaptive dynamic programming approach
553 for inferring cophylogenies. *Bmc Bioinformatics* **11**, S60.
- 554 Meseguer AS, Lobo JM, Ree R, Beerling DJ, Sanmartin I (2015) Integrating Fossils, Phylogenies, and
555 Niche Models into Biogeography to Reveal Ancient Evolutionary History: The Case of
556 *Hypericum* (Hypericaceae). *Systematic Biology* **64**, 215-232.
- 557 Morrone JJ (2009) *Evolutionary biogeography. An integrative approach with case studies*. 301 p.
558 Columbia University Press, New York.
- 559 Nieberding C, Jousselin E, Desdevises Y (2010) The use of co-phylogeographic patterns to predict the
560 nature of host-parasite interactions, and vice versa. In: *The Biogeography of Host-parasite*
561 *interactions* (eds. Morand S, Krasnov BR), pp. 59-70. Oxford University Press, New York.
- 562 Nøjgaard N, Geiß M, Stadler PF *et al.* (2017) Forbidden time travel : chracterization of Time-
563 consistent tree reconciliation map. The LIPIcs Leibniz International Proceedings in
564 Informatics: "Algorithms in Bioinformatics", 17th International Workshop, WABI
565 2017, Boston MA, USA.
- 566 Page RD (1994a) Maps between trees and cladistic analysis of historical associations among genes,
567 organisms, and areas. *Systematic Biology* **43**, 58-77.
- 568 Page RD, Charleston MA (1997) From gene to organismal phylogeny: reconciled trees and the gene
569 tree/species tree problem. *Molecular Phylogenetics and Evolution* **7**, 231-240.

- 570 Page RD, Charleston MA (1998) Trees within trees: phylogeny and historical associations. *Trends in*
571 *Ecology and Evolution* **13**, 356-359.
- 572 Page RDM (1990) Temporal congruence and cladistic analysis of biogeography and cospeciation.
573 *Systematic Zoology* **39**, 205-226.
- 574 Page RDM (1991) Clocks, clades, and cospeciation: comparing rates of evolution and timing of
575 cospeciation events in host-parasite assemblages. *Systematic Zoology* **40**, 188-198.
- 576 Page RDM (1994b) Parallel phylogenies: reconstructing the history of host-parasite assemblages.
577 *Cladistics* **10**, 155-173.
- 578 Paterson AM, Banks J (2001) Analytical approaches to measuring cospeciation of host and parasites:
579 through a glass, darkly. *International Journal for Parasitology* **31**, 1012-1022.
- 580 Percy DM, Page RD, Cronk QCB (2004) Plant-insect interactions: double-dating associated insect and
581 plant lineages reveals asynchronous radiations. *Systematic Biology* **53**, 120-127.
- 582 Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus Evolution in Relation to Its Rodent and
583 Insectivore Hosts: No Evidence for Codivergence. *Molecular Biology and Evolution* **26**, 143-
584 153.
- 585 Ree RH, Moore BR, Webb CO, Donoghue MJ (2005) A likelihood framework for inferring the
586 evolution of geographic range on phylogenetic trees. *Evolution* **59**, 2299-2311.
- 587 Ree RH, Sanmartin I (2009) Prospects and challenges for parametric models in historical
588 biogeographical inference. *Journal of Biogeography* **36**, 1211-1220.
- 589 Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal,
590 local extinction, and cladogenesis. *Systematic Biology* **57**, 4-14.
- 591 Ronquist F, Sanmartin I, Futuyma DJ, Shaffer HB, Simberloff D (2011) Phylogenetic methods in
592 biogeography. *Annual Review of Ecology, Evolution, and Systematics* **42**, 441-464.
- 593 Rønsted N, Weiblen GD, Cook JM, *et al.* (2005) 60 million years of co-divergence in the fig-wasp
594 symbiosis. *Proceedings of the Royal Society B-Biological Sciences* **272**, 2593-2599.
- 595 Rosenblueth M, Sayavedra L, Samano-Sanchez H, Roth A, Martinez-Romero E (2012) Evolutionary
596 relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts
597 (Hemiptera: Coccoidea). *Journal of Evolutionary Biology* **25**, 2357-2368.
- 598 Singh G, Dal Grande F, Divakar PK, *et al.* (2017) Fungal-algal association patterns in lichen
599 symbiosis linked to macroclimate. *New Phytologist* **214**, 317-329.
- 600 Sorci G, Skarstein F, Morand S, Hugot JP (2003) Correlated evolution between host immunity and
601 parasite life histories in primates and oxyurid parasites. *Proceedings of the Royal Society B-*
602 *Biological Sciences* **270**, 2481-2484.
- 603 Stevens J (2004) Computational aspects of host-parasite phylogenies. *Brief Bioinform* **5**, 339-349.
- 604 Stolzer M, Lai H, Xu ML, *et al.* (2012) Inferring duplications, losses, transfers and incomplete lineage
605 sorting with nonbinary species trees. *Bioinformatics* **28**, I409-I415.
- 606 Szöllösi GJ, Boussau B, Abby SS, Tannier E, Daubin V (2012) Phylogenetic modeling of lateral gene
607 transfer reconstructs the pattern and relative timing of speciations. *Proceedings of the National*
608 *Academy of Sciences* **109**, 17513-17518.
- 609 van Iersel L, Scornavacca C, Kelk S. (2014). Exact reconciliations of undated trees.
610 <https://arxiv.org/abs/1410.7004>
- 611 Weckstein JD (2004) Biogeography explains: cophylogenetic patterns in toucan chewing lice.
612 *Systematic Biology* **53**, 154-164.
- 613 Wiebes JT (1979) Co-evolution of figs and their insect pollinators. *Annual Review of Ecology and*
614 *Systematics* **10**, 1-12.

- 615 Wilson JS, Forister ML, Dyer LA, *et al.* (2012) Host conservatism, host shifts and diversification
616 across three trophic levels in two Neotropical forests. *Journal of Evolutionary Biology* **25**,
617 532-546.
- 618 Zientz E, Dandekar T, Gross R (2004) Metabolic interdependence of obligate intracellular bacteria and
619 their insect hosts. *Microbiology and Molecular Biology Reviews* **68**, 745-770.
- 620

621 Fig. 1: Description of how geographical constraints are handled by the *Mowgli* software. Plain lines
622 and nodes represent branches and nodes of the parasite tree, while cylinders and dashed ellipses
623 represent branches and nodes of the host tree. The colours of a node correspond to geographical areas,
624 these areas are observed (hence enforced) for extant taxa but inferred for ancestral nodes. A) The
625 parasite tip u can be mapped to a tip x of the host tree if the areas of the host contain all areas of the
626 parasite. B) *Mowgli* accepts that a parasite node u cospeciates with a host at a node x , if the two nodes
627 share an area. Here, it was inferred that the ancestral parasite u lived in green and/or red areas, but
628 cospeciated with a host x that lived in blue and/or red areas, we conclude that u lived only in the red
629 area at this time and that x lived at least in the red area. C) To map an ancestral parasite inside a branch
630 (x_p, x) – to represent the source of a host switch (upper part of the figure), or a duplication of the
631 parasite (lower part) -- *Mowgli* requires that the parasite has potentially lived in an area of x_p or x . In
632 this example, u shares an area with x_p and v shares an area with x . D) To map a branch (u_p, u) of the
633 parasite tree inside a branch (x_p, x) of the host tree, *Mowgli* requires that the parasite mapped on the
634 host branch (u in the left part of the figure showing the destination of a switch and v_p in the right part
635 showing the departure due to a switch) has potentially lived in any area assigned to node x_p or to node
636 x . This is the case here for node u that was indicated as having lived in the red area (also assigned to x)
637 and for v_p , assigned to the green area, also proposed for x_p . Note that mapping v_p into (x_p, x) would have
638 also been correct if v_p had been assigned to the red area, indicating that it changed from the green to
639 the red area, together with its host, before switching to another host. E) When a parasite lineage (u_p, u)
640 living on an ancestral host remains with one descending child of this host after its speciation at node x :
641 *Mowgli* requires that the area at which the host speciation occurred is also found among the areas
642 inferred for u_p or u . The mapping in this example indicates that the parasite changed area with its host,
643 before the host speciation event.

644 Fig. 2: Results of the reconciliations obtained on a simulated data set with default cost settings: A)
645 without enforcing geographic constraints (7 coSpeciation events, 1 Transfer, 1 Loss), purple dashed
646 branches indicate parts of the reconciliation where geographic constraints are not fulfilled B) when
647 enforcing geographic constraints (6 S, 2 T, 1 L). In both cases, the big pie charts correspond to the
648 host ancestral geographic areas while small pie charts correspond to the symbiont ancestral geographic
649 areas. The leaves of the species tree are also coloured according to the current geographic distribution
650 of the associates. Annotations files given for the host tree and the symbiont tree specified a single most
651 likely area at each node.

652 Fig. 3: Results of the reconciliations inferred by *Mowgli* on the fig/fig wasp data set with default cost
653 settings: A) using a single most likely area for ancestral species and without enforcing geographic
654 constraints (events: 17 S, 6 T, 1 L), purple dashed branches indicate parts of the reconciliation where
655 geographic constraints are not fulfilled; B) using a single most likely area when enforcing geographic
656 constraints (events: 16 S, 7 T, 1 L). The leaves of the phylogenetic trees are coloured according to
657 current geographic distribution of the associates. In both cases, big pie charts correspond to the *Ficus*
658 ancestral geographic areas, small pie charts correspond to the pollinator ancestral geographic areas; C)
659 Reconciliation obtained when alternative ancestral areas are considered (namely those with
660 probability > 0.15). Sections of the pies are proportional to the probability of the associated states.
661 *Ficus* taxonomic subdivisions are reported on the right inside of the figure.

662 Supplementary Material 1: Description of the pipeline to generate trees and annotation files that can be
663 taken as inputs for both *Mowgli* and *Sylvx*.

664 Supplementary Material 2: Reconciliations obtained under different cost settings on the mock dataset.

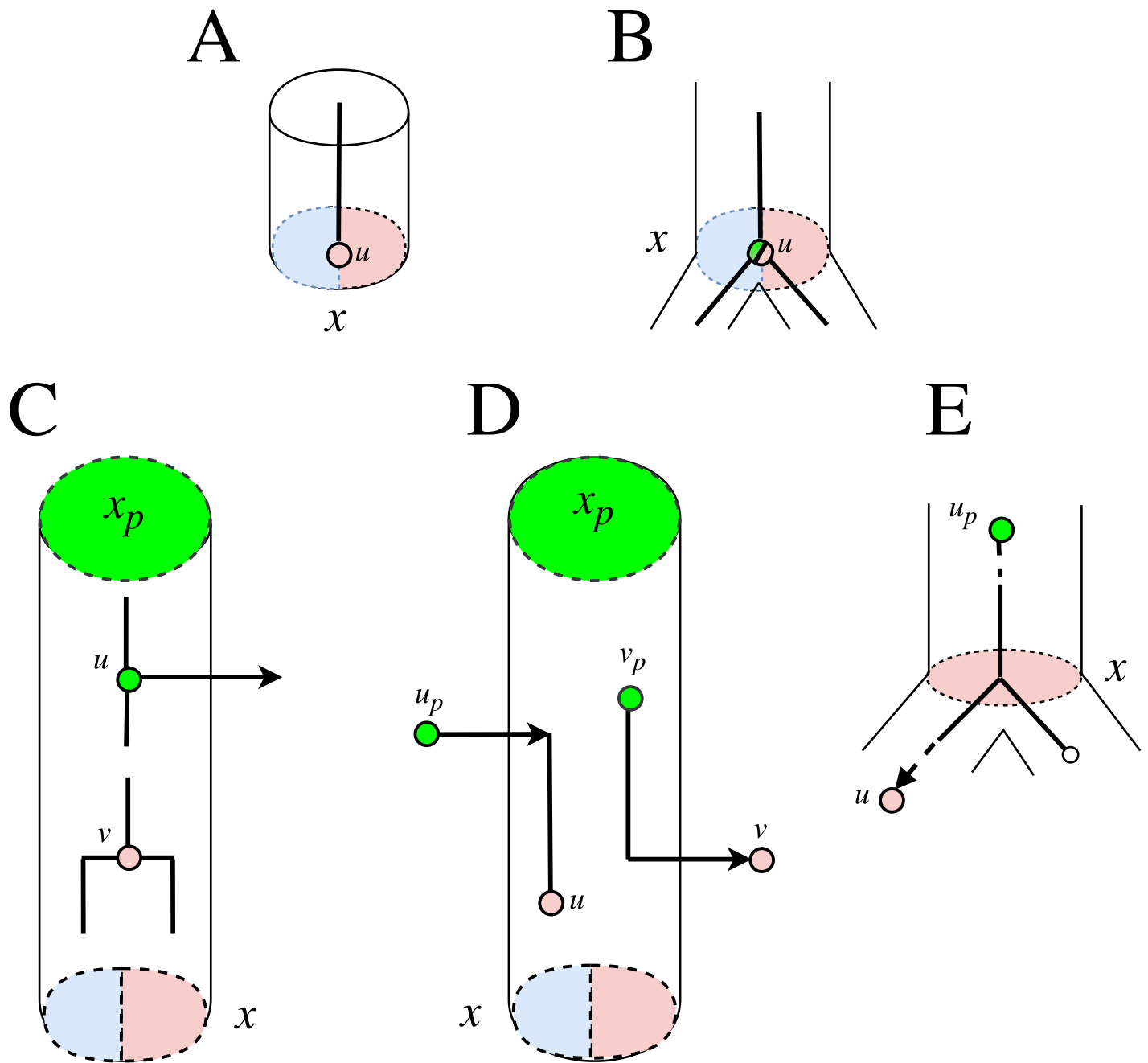
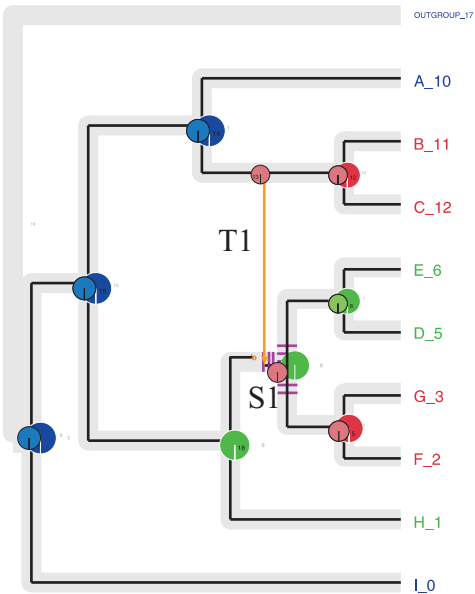


Figure 1



● Asia ● Africa ● America

Figure 2A

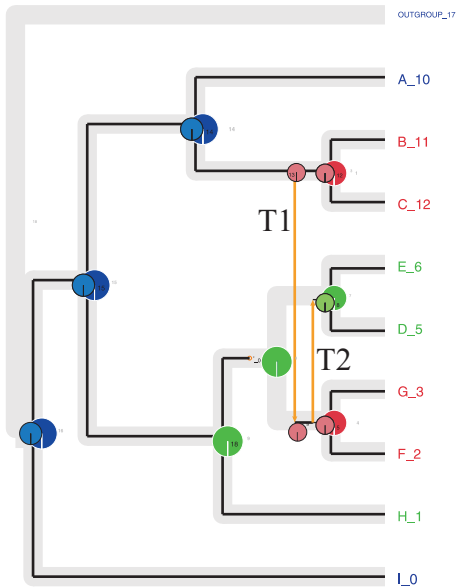
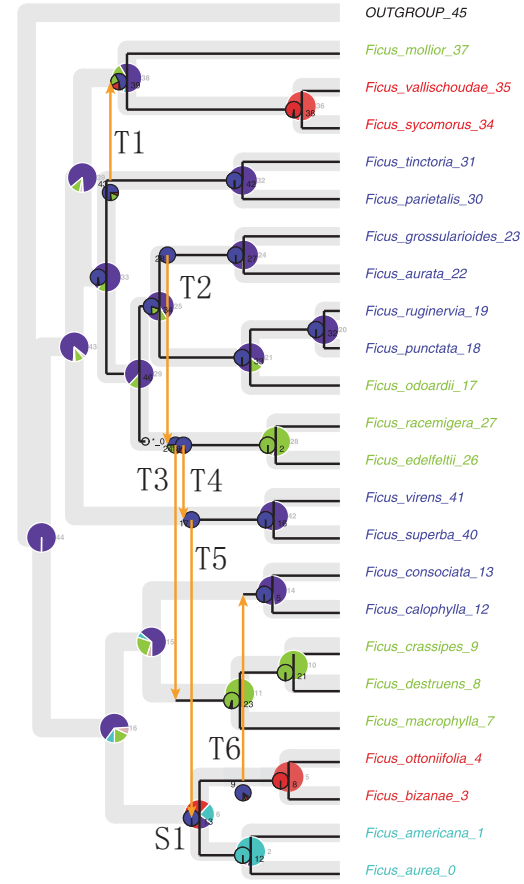
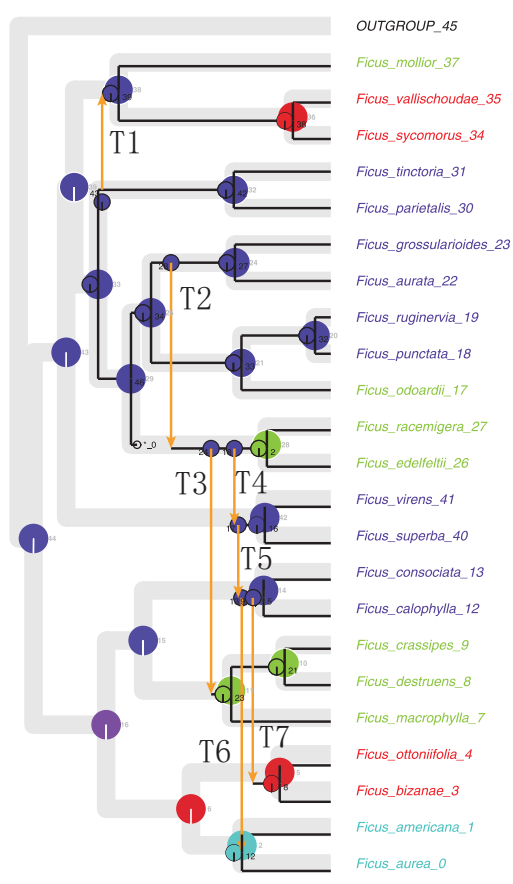
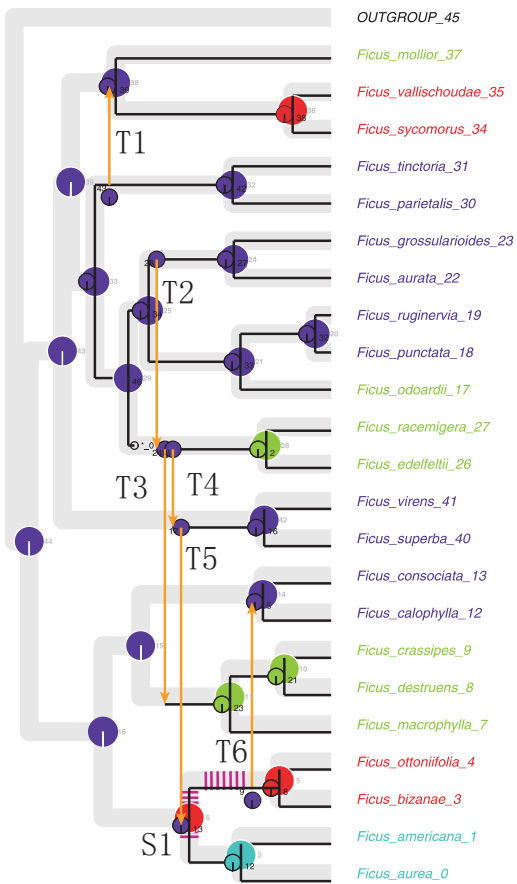


Figure 2B



Adenosperma

Sycomorus

Sycidium

Ficus

Synoecia

Oreosycea

Urostigma

Conosycea

Malvanthera

Galoglychia

Americana

● Neotropics ● Asia ● Afrotropics ● Australasia

Figure 3A

Figure 3B

Figure 3C