Do closely related species interact with similar partners? Testing for phylogenetic signal in bipartite interaction networks

Benoît Perez-Lamarque [1,2], Odile Maliet [1], Marc-André Selosse [2,3], Florent Martos [2], and Hélène Morlon [1]

[1] *Institut de biologie de l'École normale supérieure (IBENS), École normale supérieure, CNRS, INSERM, Université PSL, 46 rue d'Ulm, 75 005 Paris, France*

[2] *Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'histoire naturelle, CNRS, Sorbonne Université, EPHE, UA, CP39, 57 rue Cuvier 75 005 Paris, France*

[3] *Department of Plant Taxonomy and Nature Conservation, University of Gdansk, Wita Stwosza 59, 80-308 Gdansk, Poland*

Corresponding author: Benoît Perez-Lamarque (benoit.perez@ens.psl.eu)

Author contributions: BPL, OM, MAS, FM, and HM designed the study. BPL performed the analyses and FM gathered the data. BPL and HM wrote the first draft of the manuscript and all authors contributed to revisions.

Data accessibility statement: All the R functions used to measure phylogenetic signals in bipartite interaction networks, including (simple, partial, and clade-specific) Mantel tests and PBLM, are available in the R-package RPANDA (Morlon *et al.* 2016) (functions *phylosignal_network* and *phylosignal_sub_network*). A tutorial and the simulated networks can be found at https://github.com/BPerezLamarque/Phylosignal_network. Amended functions of *BipartiteEvol* are also included in RPANDA.

Abstract

Whether interactions between species are conserved on evolutionary time-scales is a central question in ecology and evolution. It has spurred the development of both correlative and model-based approaches for testing phylogenetic signal in interspecific interactions: do closely related species interact with similar sets of partners? Here we test the statistical performances of some of these approaches using simulations. We find that one of the most widely used model-based approach, the Phylogenetic Bipartite Linear Model (PBLM), often detects phylogenetic signal when it should not. Simple Mantel tests investigating the correlation between phylogenetic distances and dissimilarities in sets of interacting partners instead have low type-I error rates and satisfactory statistical power, especially when using weighted interactions and phylogenetic dissimilarity metrics; however, they often artifactually detect anti-phylogenetic signals. Partial Mantel tests, which are used to partial out the phylogenetic signal in the number of partners, actually fail at correcting for this confounding effect, and we instead propose the sequential use of simple Mantel tests. We also explore the ability of simple Mantel tests to analyze clade-specific phylogenetic signal, while current methods only measure an overall signal. We provide general guidelines and an empirical application on an interaction network between orchids and mycorrhizal fungi.

Keywords: ecological network, phylogenetic signal, Mantel tests, clade-specific signal, species interactions, mycorrhizal symbiosis.

Introduction

Species in ecological communities engage in numerous types of interspecific interactions, such as pollination, mycorrhizal symbioses, predation, and parasitism (Bascompte *et al.* 2003; Fontaine *et al.* 2011; Martos *et al.* 2012; Bascompte & Jordano 2013). Understanding the processes that shape these interaction networks, including the role of evolutionary history, *i.e.* the phylogeny, is a major focus of ecology and evolution (Rezende *et al.* 2007; Vázquez *et al.* 2009; Krasnov *et al.* 2012; Elias *et al.* 2013; Rohr & Bascompte 2014). One way to assess the role of evolutionary history in shaping contemporary interactions is to test for phylogenetic signal in species interactions, *i.e.* whether closely related species interact with similar sets of partners (Peralta 2016).

Testing for phylogenetic signal in a unidimensional trait (*i.e.* whether a trait is phylogenetically conserved) for a given clade is mainstream (Felsenstein 1985; Blomberg *et al.* 2003; Münkemüller *et al.* 2012). One approach (the 'correlative' approach) is to perform a Mantel test between phylogenetic and trait distances (Mantel 1967); another approach (the 'model-based' approach) relies on trait evolution models such as Pagel's λ (Pagel 1999) or Blomberg's K (Blomberg *et al.* 2003). The model-based approach has a higher ability to detect an existing phylogenetic signal (power) and a lower propensity to infer a phylogenetic signal when it should not (type-I error; Harmon & Glor 2010): The correlative approach should therefore only be used when the model-based approach is not applicable, *e.g.* if the 'trait' data is expressed in terms of pairwise distances.

Testing for phylogenetic signal in species interactions falls in the category of cases where the 'trait' data are pairwise distances, here the between-species dissimilarity in sets of interacting species. Simple Mantel tests have therefore been widely used in this context (*e.g.* Rezende *et al.* 2007; Elias *et al.* 2013; Fontaine & Thébault 2015). Partial Mantel tests have also been used to test whether the

phylogenetic signal reflects more the identity of the interacting partners than their number, *i.e.* the degree, as similarity in the number of partners can increase the value of similarity metrics (Rezende *et al.* 2007; Jacquemyn *et al.* 2011; Aizen *et al.* 2016). Besides such correlative approaches, the Phylogenetic Bipartite Linear Model (PBLM, Ives & Godfray 2006) has been widely used to test for phylogenetic signal in species interactions in a variety of empirical networks, *e.g.* in host-parasite, plant-fungus, and pollination networks (Martos *et al.* 2012; Martín González *et al.* 2015; Xing *et al.* 2020).

Here, using simulations, we perform a comparative analysis of the statistical performances of the correlative (simple or partial) Mantel tests and of the model-based PBLM to test for phylogenetic signal in species interactions (Box 1), considering weighted and unweighted bipartite interaction networks between species from two guilds A and B (Fig. 1). We also investigate the ability of Mantel tests to detect the presence of phylogenetic signal in the different clades of a phylogenetic tree, because current methods only measure the overall signal in the phylogeny. Finally, we provide general guidelines and apply them to an orchid-fungus mycorrhizal network identified across the oceanic island of Réunion (Martos *et al.* 2012).

Methods

Simulating bipartite interaction networks with or without phylogenetic signal in species interactions

We used *BipartiteEvol*, an individual-based eco-evolutionary model (Maliet *et al.* 2020), to generate interaction networks with or without phylogenetic signal between two guilds interacting in a mutualistic, antagonistic, or neutral way. Each individual from guild A (resp. B) is characterized by a multidimensional continuous trait and interacts with one individual from guild B (resp. A). The effect of this interaction on the fitness of each individual from guilds A or B is determined by the distance in trait space of the two interacting individuals, according to a classical trait matching expression parametrized by two parameters α_A and α_B (Supplementary Methods 1, Maliet *et al.* 2020). These parameters determine the nature and specificity of the interaction: positive α_A and α_B correspond to mutualistic interactions, negative α_A and positive α_B to antagonistic interactions (with guild A representing hosts/preys and guild B parasites/predators), high $|\alpha|$ values to scenarios with strong fitness effects (*i.e.* highly specialized interactions), and $|\alpha|$ values close to 0 to more neutral scenarios. *BipartiteEvol* simulates individual's deaths and births (proportional to the individual's fitness) and new individuals have a probability μ to mutate, in which case new traits are drawn independently in a normal distribution centered on the parent traits. Here, we amended the species delineation of the original *BipartiteEvol* model (Maliet *et al.* 2020) and considered that each combination of traits forms a new species. Under this process, closely related species tend to interact with similar sets of partners (*i.e.* there is a phylogenetic signal in species interactions) if (and only if): (1) closely related species have similar traits (*i.e.* there is a phylogenetic signal in species traits) and (2) these traits determine who interacts with whom, *i.e.* $\alpha \neq 0$. Similarly, an anti-phylogenetic signal in species interactions (*i.e.* the tendency for closely related species

127 to associate with dissimilar partners) is expected if there is anti-phylogenetic signal in
128 species traits and $\alpha \neq 0$.

129 Using the R-package RPANDA (Morlon *et al.* 2016; R Core Team 2020), we
130 simulated a total of 2,400 interaction networks with individuals characterized by a six-
131 dimensional trait. To obtain a wide range of network sizes, we considered a total
132 number of 500, 1,000, 2,000, 3,000, 4,000, or 5,000 pairs of interacting individuals per
133 simulation. For each size, we simulated the evolution of 100 neutral networks (α_A=0 ;
134 α_B=0), 120 mutualistic networks (**i**: α_A=1; α_B=1; **ii**: α_A=0.1; α_B=0.1; **iii**: α_A=0.01; α_B=0.01;
135 **iv**: α_A=1; α_B=0.1; **v**: α_A=1; α_B=0.01; and **vi**: α_A=0.1; α_B=0.01) and 180 antagonistic networks
136 (**i**: α_A=-1; α_B=1; **ii**: α_A=-0.1; α_B=0.1; **iii**: α_A=-0.01; α_B=0.01; **iv**: α_A=-1; α_B=0.1; **v**: α_A=-1;
137 α_B=0.01; **vi**: α_A=-0.1; α_B=1; **vii**: α_A=-0.1; α_B=0.01; **viii**: α_A=-0.01; α_B=1; **ix**: α_A=-0.01; α_B=0.1).
138 We used a mutation rate μ=0.01 and followed the interacting individuals during 50^6
139 death events. At the end, we extracted for each guild a species tree from its genealogy
140 by randomly selecting one individual per species (Fig. S1) and reconstructed the
141 corresponding weighted interaction network by counting the number of occurrences
142 of each interspecific interaction.

143 We separated the 2,400 simulated networks between those for which we should
144 expect a phylogenetic signal in species interactions and those for which we should not.
145 We did not expect phylogenetic signal in species interactions in neutral networks and
146 in non-neutral networks with no phylogenetic signal in species traits. Conversely, we
147 expected phylogenetic signal in non-neutral networks with phylogenetic signal in
148 species traits. Phylogenetic signal in species traits was evaluated using Mantel tests
149 (Pearson correlation) between phylogenetic distances and trait distances computed as
150 the Euclidian distances between trait values for each species pair.

151

152 **Computing phylogenetic signal in species interactions**

153

154 <u>Mantel tests:</u> We evaluated the phylogenetic signal in species interactions in guilds A
155 and B separately using simple Mantel tests between phylogenetic and ecological (set

156 of interacting partners) distances. Ecological distances were measured both without
157 accounting for evolutionary relatedness of the interacting partners, using (weighted or
158 unweighted) Jaccard, and accounting for relatedness using (weighted or unweighted)
159 UniFrac distances (Box 1). We used Pearson, Spearman, and Kendall correlations by
160 extending the *mantel* function in the R-package ecodist (Goslee & Urban 2007); we
161 evaluated the significance of each correlation using 10,000 permutations, except for the
162 computationally intensive Kendall correlation (100 permutations only). For each
163 network, we considered that there was a significant phylogenetic signal (resp. anti-
164 phylogenetic signal) if the correlation coefficient was higher (resp. lower) than >95%
165 of the randomized correlations; we computed the p-value of each one-tailed Mantel
166 test as the fraction of the randomized correlations above (resp. below) the original
167 value.
168
169 PBLM: To estimate phylogenetic signal based on PBLM, we modified the function *pblm*
170 from the R-package picante (Kembel *et al.* 2010) to more efficiently perform matrix
171 inversions and handle large interaction networks. We followed Ives & Godfray (2006;
172 Box 1) by considering that the phylogenetic signal is significant when the mean square
173 error (MSE) of the model is smaller than that obtained using star phylogenies (MSE_{star});
174 we also used a more stringent criterion by considering that the signal is significant
175 when the MSE is at least 5% lower than MSE_{star}. Finally, we applied the bootstrapping
176 method of Ives & Godfray (2006; Box 1) for the smallest networks (its computational
177 cost prevented us from applying it on larger networks).
178
179 **Confounding effect of the phylogenetic signal in the number of partners**
180
181 To test the performances of the partial Mantel test at measuring phylogenetic signal in
182 species interactions while controlling for the number of partners (Box 1), we first
183 performed partial Mantel tests between phylogenetic and ecological distances, while
184 controlling for the absolute differences in degrees, on the networks simulated with

185 *BipartiteEvol*. There is no reason to produce a phylogenetic signal in the number of
186 partners in *BipartiteEvol* simulations, and we verified this by performing Mantel tests
187 between phylogenetic distances and degree differences. These analyses were
188 performed to assess whether partial Mantel tests loose power compared to simple
189 Mantel tests.
190
191 Second, we assessed whether partial Mantel tests successfully correct for phylogenetic
192 signal in the number of partners using networks simulated under a process that
193 generate phylogenetic conservatism in the number, but not the identity, of interacting
194 partners (*i.e.* partial Mantel tests should not be significant when applied to such
195 networks). To simulate network with only phylogenetic conservatism in the number
196 of partners in guild A, we first simulated phylogenetic trees for guilds A and B using
197 *pbtree* (R-package phytools; Revell 2012) with a number of species uniformly sampled
198 between 40 and 150 by guild. Next, we simulated the number of partners of the species
199 from guild A using an Ornstein-Uhlenbeck process with an attraction toward 0, a
200 variance of 0.1 (noise of the Brownian motion), and a selection strength (a_A) ranging
201 from 5 (strong stabilizing effect, weak phylogenetic signal) to 0 (Brownian motion,
202 strong phylogenetic signal). We computed the number of partners per species by
203 calibrating the simulated values between 1 and the number of species in guild B and
204 taking the integer part. For each a_A value (5, 1, 0.5, 0.05, or 0), we performed 100
205 simulations using *mvSIM* (R-package mvMORPH; Clavel *et al.* 2015). Finally, for each
206 species in A, we attributed the corresponding number of partners in B at random to
207 obtain binary networks. We checked that our simulations indeed generated a signal in
208 the number of partners by performing simple Mantel tests between phylogenetic and
209 degree difference distances. Finally, we performed on each simulated network a
210 partial Mantel test.
211
212 Given the poor performances of partial Mantel tests (see Results), we tested
213 whether using sequential Mantel tests would provide a good alternative: based on

214 simple Mantel tests, we consider that there is a phylogenetic signal in the identity of
215 the partners if there is a phylogenetic signal in species interactions and no phylogenetic
216 signal in the number of partners. We applied this sequential testing to all our simulated
217 networks.
218
219 **Effect of phylogenetic uncertainty, sampling asymmetry, and network**
220 **heterogeneity on measures of phylogenetic signal in species interactions**
221
222 Unlike simulations (such as those provided by *BipartiteEvol*), empirical bipartite
223 networks suffer from phylogenetic uncertainty (*e.g.* in the microbial partners' tree
224 when studying host-associated microbiota – which often prevents accounting for
225 evolutionary relatedness; *i.e.* using UniFrac distances), sampling asymmetry (*i.e.* one
226 side of the network is more thoroughly sampled than the other), and network
227 heterogeneity (*i.e.* different sub-clades in the network have different levels of
228 phylogenetic signal). We performed additional analyses to investigate the effect of
229 these aspects on phylogenetic signals in species interactions measured using simple
230 Mantel tests.
231
232 First, we tested the effect of phylogenetic uncertainty in the partners' tree on the
233 measure of phylogenetic signal when evolutionary relatedness is accounted for (*i.e.*
234 using UniFrac distances). We performed these analyses to assess whether accounting
235 for the partners' evolutionary relatedness remains advantageous (see Results) when
236 phylogenetic uncertainty is high. To add some variability in the phylogenetic tree of
237 guild B (resp. A) used to compute the UniFrac distances between species pairs from
238 guild A (resp. B), we first simulated, on the original partners tree, the evolution of a
239 short DNA sequence and then reconstructed the tree from the simulated DNA
240 alignment using neighbor-joining (*nj* function, R-package APE (Paradis *et al.* 2004)).
241 We used *simulate_alignment* (R-package HOME; Perez-Lamarque & Morlon 2019) to
242 simulate sequences of length 75, 150, 300, 600, or 1,200 base-pairs, with 30% of variable

243 sites, and a substitution rate of 1.5 (shorter fragments should result in noisier
244 phylogenies).
245
246 Second, we tested the influence of sampling asymmetry on measures of
247 phylogenetic signal. Empirical networks are often an incomplete representation of the
248 actual interactions between two guilds because they are under-sampled, and
249 frequently, in an asymmetrical way. For instance, by sampling targeted species from
250 guild A, observed networks are constituted by few species from guild A which have
251 the complete set of their partners and by often more species from guild B which have
252 an incomplete set of their partners (as they likely interact with unsampled species from
253 guild A). We tested the influence of such sampling asymmetry by selecting only 10%
254 of the most abundant species from guild A in each simulated network (while retaining
255 at least 10 species) and computed phylogenetic signal in these asymmetrically-
256 subsampled networks.
257
258 Third, both Mantel tests and PBLM neglect the heterogeneity within networks.
259 Indeed, a non-significant phylogenetic signal at the level of the entire network can
260 potentially hide a sub-clade presenting significant phylogenetic signal. Alternatively,
261 a phylogenetic signal in the entire network may be driven by only two sub-clades of
262 guilds A and B. To explore the potential heterogeneity of the phylogenetic signal
263 within one guild, one possibility is to apply Mantel tests to the sub-networks formed
264 by a given sub-clade (*e.g.* Song *et al.* 2020). For each node of the tree of guild A having
265 at least 10 descendants, we estimated the clade-specific phylogenetic signal using a
266 Mantel test investigating whether closely related species from this sub-clade of A tend
267 to interact with similar partners (and *vice-versa* for guild B). Using UniFrac distances,
268 we performed the Mantel tests with 100,000 permutations, and introduced a
269 Bonferroni correction for multiple testing to keep a global alpha-risk of 5%. To test this
270 approach, we generated synthetic networks with known sub-clade signal by
271 artificially combining networks simulated under neutrality with networks simulated

with the mutualistic parameters **v** (see Results). We grafted each "mutualistic" phylogenetic tree from guilds A and B within a "neutral" phylogenetic tree by randomly selecting a branch, such that it creates a separate module with strong phylogenetic signal. Such simulations could correspond to the evolution of a different niche, *e.g.* terrestrial *versus* epiphytic plants associating with different mycorrhizal fungi (Martos *et al.* 2012). We then performed our clade-specific analysis of phylogenetic signal and investigated in which nodes we recovered significant phylogenetic signals.

General guidelines and illustration with application on the orchid-fungus mycorrhizal network from La Réunion

We used our results and other empirical considerations to provide general guidelines for testing for phylogenetic signal in interaction networks. We illustrated these guidelines by applying them in a network between orchids and mycorrhizal fungi from La Réunion island (Martos *et al.* 2012). This network encompasses 70 orchid species (either terrestrial or epiphytic species) and 93 molecularly-identified fungal partners (defined according to 97% sequence similarity; Martos *et al.* 2012). We gathered the maximum-likelihood plant and fungal phylogenies on TreeBASE (Study Accession S12721), calibrated the orchid phylogeny using a relaxed clock with *chronos* (Paradis 2013), and arbitrarily added 10 million-years-old polytomies in unresolved genera to obtained a species-level orchid phylogeny.

294 # Results

295

296 **Expected phylogenetic signals in species interactions in *BipartiteEvol* networks**

297

298 The networks simulated using *BipartiteEvol* gave realistic ranges of sizes (from
299 <50 to >250 species; Fig. S2-3).

300 We found a significant phylogenetic signal in species traits for most antagonistic
301 and neutral simulations (Fig. S4). In contrast, for many mutualistic simulations, closely
302 related species often did not tend to have similar traits, except when α_B=0.01 (*i.e.*
303 mutualistic parameters **iii**, **v**, and **vi**; Fig. S4). When α_B were higher (*i.e.* mutualistic
304 parameters **i**, **ii**, and **iv**, we suspect stabilizing selection to occur and erase the
305 phylogenetic signal in the traits (Maliet *et al.* 2020): we therefore do not expect
306 phylogenetic signal in species interactions for these simulations. In addition, we found
307 an anti-phylogenetic signal in species traits in less than 1% of the simulations (likely
308 to be false-positives given the alpha-risk of Mantel tests; Fig. S4): these networks were
309 thus removed when evaluating the performance of the different approaches and we
310 therefore do not expect anti-phylogenetic signal in species interactions for the
311 remaining networks.

312

313 **Tests of phylogenetic signal in species interactions in *BipartiteEvol* networks**

314

315 Using Mantel tests, as expected, we did not find significant phylogenetic signals
316 in species interactions for most neutral networks or for networks with no signal in
317 species traits (Fig. 2, Figs. S5-6-7): the type-I error rate was below 5%, corresponding
318 to the alpha-risk of the test (Table S1), with one notable exception for small networks
319 when using weighted Jaccard distances and Pearson correlations (~8% type-I error).
320 Conversely, we detected a significant unexpected anti-phylogenetic signal in more

321 than 10% of the simulated networks, in particular in the small ones (Fig. 2, Figs. S5-6-
322 7).
323
324 Many mutualistic or antagonistic networks where we expected a phylogenetic
325 signal in species interactions (*i.e.* non-neutral networks with signal in species traits)
326 presented no significant signal with Mantel tests (Fig. 2, Figs. S5-6-7), in particular
327 those simulated with low α_A and α_B values (*e.g.* antagonism **vii**), where non-neutral
328 effects were weak. In mutualistic networks, phylogenetic signals in species
329 interactions were present only when there was a large asymmetry in the effects of trait
330 matching on the fitnesses of the species from guilds A or B (case **v**: α_A=1; α_B=0.01), *i.e.*
331 when only one guild was specialized. Conversely, in antagonistic networks,
332 phylogenetic signals were found mainly when trait matching had a strong impact on
333 the fitness of guild B (the obligate parasites/predators; $\alpha_B \geq 0.1$). Additionally, when
334 phylogenetic signal was significant in one guild, it was generally also significant in the
335 other; in antagonistic networks, the signal was usually higher in guild A compared to
336 guild B (Figs. S5-6-7).
337
338 The statistical power of Mantel tests measuring phylogenetic signal in species
339 interactions seems to be modulated by network size, as phylogenetic signals were less
340 often significant but generally stronger in smaller networks (Figs. S5-6-7). Moreover,
341 Mantel tests based on Pearson correlations had higher power than Spearman and
342 Kendall correlations (Figs. S5-6-7) and weighted UniFrac distances outperformed
343 other ecological distances in terms of power (Figs. S5-6-7; Table S2).
344
345 When using mean square errors to evaluate the significance of PBLM, we found
346 a significant phylogenetic signal in most of the simulated networks including when
347 we did not expect any (Fig. 2e). The propensity of PBLM to detect phylogenetic signal
348 decreased in large unweighted networks, but the type-I errors remained >30%,

349 including when using a more stringent significance cutoff (Figs. S8-9). Similar results
350 were obtained when bootstrapping to evaluate the significance (Fig. S10).
351
352 **Confounding effect of phylogenetic signal in the number of partners**
353
354 As expected, tests of phylogenetic signal in the number of partners were non-
355 significant in the large majority of the *BipartiteEvol* networks, especially the larger ones
356 (Fig. S11). We did however observe significant correlations between ecological
357 distances and degree difference distances (Fig. S12). Partial Mantel tests testing for
358 phylogenetic signal in species interactions while accounting for phylogenetic signal in
359 the number of partners had similar type-I error and power as simple Mantel tests (Figs.
360 S5-13; Table S2). Performing sequential Mantel tests decreased the statistical power by
361 less than 2% (Table S2).
362
363 Networks simulated with phylogenetic conservatism in the number, but not the
364 identity of partners covered a realistic range of sizes (Fig. S14). As expected, Mantel
365 tests revealed significant phylogenetic signals in the number of partners in >60% of
366 these networks, with an increasing percentage of significant tests with decreasing a_A
367 (*i.e.* increasing conservatism in the number of partners; Fig. S18). We found significant
368 correlations between degree differences and ecological distances in most of these
369 simulated networks (Fig. S15). As a result, simple Mantel tests testing for phylogenetic
370 signal in species interactions without accounting for phylogenetic signal in the number
371 of partners were frequently significant (>30%; Fig. S16; Table S3). Partial Mantel tests
372 controlling for degree differences slightly decreased the proportion of false-positives,
373 but it remained high (type-I error >25%; Fig. S17). In addition, partial Mantel tests
374 detected a spurious significant anti-phylogenetic signal in species interactions in >15%
375 of the networks (Fig. S17). Conversely, only few networks with a significant simple
376 Mantel test in species interactions did not produce a significant simple Mantel test in

377 the number of partners, such that sequential Mantel tests had only a ~7% type-I error
378 rate (Table S3).
379
380 **Effect of phylogenetic uncertainty, sampling asymmetry, and network**
381 **heterogeneity**
382
383 The statistical power of Mantel tests using UniFrac distances decreased, as
384 expected, when the length of the simulated DNA sequences decreased (*i.e.* when
385 phylogenetic uncertainty increased; Fig. S19). However, even when the simulated
386 DNA sequences were the shortest (75 base-pairs), resulting in very noisy reconstructed
387 partners' tree (Fig. S20), the statistical power of the Mantel tests using UniFrac
388 distances remained larger than when using Jaccard distances (Fig. S19).
389
390 Our results on the statistical performance of tests of phylogenetic signal were
391 similar when considering sampling asymmetry (Figs. S21-24): PBLM spuriously
392 detected phylogenetic signal when it should not, and Mantel tests had decent
393 statistical performances, especially when using weighted UniFrac distances. In
394 addition, the correlations of the Mantel tests in guild A were generally higher when
395 significant (Fig. S23).
396
397 Our clade-specific tests of phylogenetic signal using Mantel tests while
398 correcting for multiple testing recovered a significant phylogenetic signal in 82% of the
399 nodes where mutualism originated (Fig. S25), as well as in most of the ascending
400 nodes. Conversely, we did not find spurious phylogenetic signals in nodes with only
401 neutrally-evolving lineages (Fig. S25).
402
403 **General guidelines and illustration with application on the orchid-fungus**
404 **mycorrhizal network from La Réunion**
405

Box 2 provides general guidelines based on our results and empirical considerations for accurate tests of phylogenetic signal in interaction networks. We applied these guidelines on the orchid-fungus mycorrhizal network from La Réunion (available in Martos et al. (2012); Fig. 3). First (step 1), simple Mantel tests of phylogenetic signal in species interactions for fungi and orchids revealed a significant but low phylogenetic signal ($R<0.10$) on the orchid side using Jaccard distances; however, the significance disappeared with UniFrac distances (Table S4). Similarly, marginally not-significant and low phylogenetic signals were detected in the mycorrhizal fungi side ($R<0.04$; Table S4). Next (step 2), simple Mantel tests of phylogenetic signal in the number of partners were not significant (p-values>0.05). Our investigation of clade-specific phylogenetic signals in species interactions in orchids (option 1) revealed a significant phylogenetic signal in Angraecinae, a sub-tribe composed of 34 epiphytic species (sequential Mantel test: $R=0.37$; Bonferroni-corrected p-value=0.016; Fig. 4) interacting with 53 fungi, suggesting that closely related Angraecinae tend to interact with more similar mycorrhizal fungi. When we checked the robustness of the significant phylogenetic signal detected in Angraecinae (option 2) by subsampling the Angraecinae clade down to 10 species, we still recovered significant signal in species interactions in both cases (Fig. S26).

Discussion:

We used simulations to perform a comparative analysis of the statistical performances of Mantel tests and the Phylogenetic bipartite linear model (PBLM; Ives & Godfray 2006) for testing for phylogenetic signal in species interactions. Our results highlight the weaknesses of PBLM and partial Mantel tests, and advocate for the use of simple and sequential Mantel tests.

The Phylogenetic bipartite linear model (PBLM) is widely used to test for phylogenetic signal in species interactions, however we found that it has a very high type-I error rate (>30%). PBLM assumes that the interaction strength between two species is determined by the product of two unobserved traits evolving on the phylogenies of guilds A and B respectively, according to two independent Ornstein-Uhlenbeck processes with the selection strengths d_A and d_B (Box 1). PBLM tests the significance of d_A and d_B, which measure the phylogenetic signal of the unobserved traits. A species with a high trait value will have high interaction strengths with many partner species (*i.e.* it is a generalist species), while a species with a low trait value will have low interaction strengths with most partner species, except with the few species with high trait values (*i.e* it is a specialist species). Therefore, we suspect d_A and d_B to measure phylogenetic signals in the number of partners rather than in species interactions. However, we also found significant d_A and d_B in the absence of phylogenetic signal in the number of partners, suggesting that PBLM is sensitive to model misspecification (it relies on strong hypotheses on how the number of partners evolves). In any case, PBLM should not be used as a routine for measuring phylogenetic signal in species interactions. Other model-based approaches inspired from PBLM propose to infer more parameters describing the phylogenetic structure of interactions networks (Rafferty & Ives 2013; Hadfield *et al.* 2014). Unfortunately their computational intensity prevented their incorporation in our comparative analysis.

454 We found that simple Mantel tests have a moderate statistical power and a
455 reasonable type-I error rate when testing for phylogenetic signal in species
456 interactions. Hence, although these tests might fail at detecting low phylogenetic
457 signal, we can trust their results when they are significant. On the contrary, we found
458 a high proportion of false-positives (5-10%) when testing for anti-phylogenetic signal;
459 therefore, detection of such signal in empirical networks should be interpreted with
460 caution. In addition, Pearson correlations performed better than Spearman and
461 Kendall correlations, which is somewhat surprising, as correlations between
462 phylogenetic and ecological distances are not particularly expected to be linear:
463 Spearman and Kendall correlations have less stringent hypotheses, as they only
464 assume monotonicity (Box 1), but they probably loose information. We also reported
465 that using ecological distances that consider interaction abundances and phylogenetic
466 relatedness of the partners, such as weighted UniFrac distances, significantly improves
467 the detection of phylogenetic signal, even when reconstructed partners trees are not
468 robust. Given that species delineation may be somewhat arbitrary, especially for
469 microbial interactors, and that Jaccard distances are directly sensitive to species
470 delineation (Sanders *et al.* 2014), we advocate the use of weighted UniFrac distances.
471 An exception might be if communities of interactors differ mainly in terms of recently
472 diverged species; in this case Jaccard distances may perform better, as UniFrac
473 distances emphasize differences in long branches rather than recent splits (Sanders *et*
474 *al.* 2014). Finally, we found that multiple simple Mantel tests combined with a
475 Bonferroni correction perform rather well to investigate clade-specific phylogenetic
476 signals. Such an approach can therefore be valuable for measuring local phylogenetic
477 signal in large "meta-networks", such as those describing host-microbiota
478 phylosymbiosis (Song *et al.* 2020), which likely have heterogeneous phylogenetic
479 signals across the network.
480
481 While simple Mantel tests have satisfactory statistical performances, these tests do
482 not control for the potential confounding effect of phylogenetic signal in the number

483 of partners. Partial Mantel tests are frequently used for investigating phylogenetic
484 signal in species interactions while controlling for signal in the number of partners;
485 however, we found that they often detected significant signals in species interactions
486 when we simulated signals in only the number of partners. Thus, partial Mantel tests
487 fail at discerning whether evolutionary relatedness strictly affects the identity of
488 partners, independently of the total number of partners associated with each species
489 (Rezende *et al.* 2007). This corroborates the poor statistical performances of partial
490 Mantel tests frequently observed in other contexts (Harmon & Glor 2010; Guillot &
491 Rousset 2013). We therefore suggest to perform sequential simple Mantel tests, testing
492 first for phylogenetic signal in species interactions, and if significant, testing for
493 phylogenetic signal in the number of partners. If there is no signal in the number of
494 partners but a signal in interactions, then we can safely conclude that evolutionary
495 relatedness strictly affects the identity of partners. This approach has a low type-I error
496 rate and a very limited power decrease; however, it does not allow testing if there is a
497 specific signal in species identity when there is a signal in the number of partners. A
498 hint at whether signal in species interactions is entirely due to signal in the number of
499 partners or not can be gained by comparing the correlation coefficients obtained when
500 correlating phylogenetic distance to ecological distance *versus* degree distance.
501
502 By definition, phylogenetic signals in species interactions measure general patterns
503 that are not informative of the processes at play (Losos 2008). A better understanding
504 of the ecological and evolutionary processes playing a role in the assembly of
505 interaction networks (Harmon *et al.* 2019) will require developing integrative process-
506 based approaches, for instance inference machineries for eco-evolutionary models
507 such as *BipartiteEvol*. Classical inferences (generalized least-squares or likelihood-
508 based approaches) might be challenging for such complex models (Hadfield *et al.*
509 2014), but strategies such as machine learning provide promising alternatives.
510

In the mycorrhizal network from La Réunion, we found non-significant or weak phylogenetic signals in species interactions at the level of the entire orchid-fungus network, suggesting these interactions are generally poorly conserved over long evolutionary timescales (Jacquemyn *et al.* 2011; Martos *et al.* 2012). Conversely, clade-specific Mantel tests detected a significant phylogenetic signal in the Angraecinae epiphytic clade that is experiencing a radiation in La Réunion island. This signal is likely produced by the different orchids genera in Angraecinae associating with specific fungal clades (Martos *et al.* 2012). Thus, our results corroborate a trend toward mycorrhizal specialization in epiphytic orchids compared with terrestrial species (Xing *et al.* 2019), as the epiphytic habitats might require particular adaptations and stronger dependences toward specific mycorrhizal fungi.

Interaction networks are increasingly being analyzed to unravel the evolutionary processes shaping their structure and to predict their stability. Currently-used tools for measuring phylogenetic signals are clearly misleading. We propose instead an alternative approach based on sequential Mantel tests. By emphasizing the limits of current tests of phylogenetic signal, we also hope to stimulate new developments in the statistical adjustment to empirical data of process-based models for the evolution of interaction networks.

Acknowledgments:

The authors acknowledge M. Elias, E. Thébault, and D. de Vienne for helpful discussions. They also thank I. Overcast, S. Lambert, I. Quintero, C. Fruciano, J. Clavel, and A. Silva for comments on an early version of the manuscript. This work was supported by a doctoral fellowship from the École Normale Supérieure de Paris attributed to BPL and the École Doctorale FIRE – Programme Bettencourt. Funding of the research of FM was from the Agence Nationale de la Recherche (ANR-19-CE02-0002). HM acknowledges support from the European Research Council (grant CoG-PANDA).

References:

Aizen, M.A., Gleiser, G., Sabatino, M., Gilarranz, L.J., Bascompte, J. & Verdú, M. (2016). The phylogenetic structure of plant-pollinator networks increases with habitat size and isolation. *Ecol. Lett.*, 19, 29–36.

Bascompte, J. & Jordano, P. (2013). *Mutualistic networks*. *Monogr. Popul. Biol.* Princeton. Princeton University Press, Princeton.

Bascompte, J., Jordano, P., Melian, C.J. & Olesen, J.M. (2003). The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci.*, 100, 9383–9387.

Blomberg, S.P., Garland, T. & Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution (N. Y).*, 57, 717–745.

Clavel, J., Escarguel, G. & Merceron, G. (2015). mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.*, 6, 1311–1319.

Elias, M., Fontaine, C. & Frank Van Veen, F.J. (2013). Evolutionary history and ecological processes shape a local multilevel antagonistic network. *Curr. Biol.*, 23, 1355–1359.

Felsenstein, J. (1985). Phylogenies and the comparative method. *Am. Nat.*, 125, 1–15.

Fontaine, C., Guimarães, P.R., Kéfi, S., Loeuille, N., Memmott, J., van der Putten, W.H., *et al.* (2011). The ecological and evolutionary implications of merging different types of networks. *Ecol. Lett.*, 14, 1170–1181.

Fontaine, C. & Thébault, E. (2015). Comparing the conservatism of ecological interactions in plant–pollinator and plant–herbivore networks. *Popul. Ecol.*, 57, 29–36.

Goslee, S.C. & Urban, D.L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.*, 22, 1–19.

Guillot, G. & Rousset, F. (2013). Dismantling the Mantel tests. *Methods Ecol. Evol.*, 4, 336–344.

Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A tale of two

phylogenies: Comparative analyses of ecological interactions. *Am. Nat.*, 183, 174–187.

Hansen, T.F. & Martins, E.P. (1996). Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution (N. Y).*, 50, 1404–1417.

Harmon, L.J., Andreazzi, C.S., Débarre, F., Drury, J., Goldberg, E.E., Martins, A.B., *et al.* (2019). Detecting the macroevolutionary signal of species interactions. *J. Evol. Biol.*, 32, 769–782.

Harmon, L.J. & Glor, R.E. (2010). Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution (N. Y).*, 64, 2173–2178.

Ives, A.R. & Godfray, H.C.J. (2006). Phylogenetic analysis of trophic associations. *Am. Nat.*, 168, E1–E14.

Jacquemyn, H., Merckx, V.S.F.T., Brys, R., Tyteca, D., Cammue, B.P.A.A., Honnay, O., *et al.* (2011). Analysis of network architecture reveals phylogenetic constraints on mycorrhizal specificity in the genus Orchis (Orchidaceae). *New Phytol.*, 192, 518–528.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.

Krasnov, B.R., Fortuna, M.A., Mouillot, D., Khokhlova, I.S., Shenbrot, G.I. & Poulin, R. (2012). Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. *Am. Nat.*, 179, 501–511.

Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.*, 11, 995–1003.

Maliet, O., Loeuille, N. & Morlon, H. (2020). An individual based model for the eco-evolutionary emergence of bipartite interaction networks. *Ecol. Lett.*, ele.13592.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.*, 27, 209–220.

Martín González, A.M., Dalsgaard, B., Nogués-Bravo, D., Graham, C.H., Schleuning, M., Maruyama, P.K., *et al.* (2015). The macroecology of phylogenetically structured hummingbird-plant networks. *Glob. Ecol. Biogeogr.*, 24, 1212–1224.

Martos, F., Munoz, F., Pailler, T., Kottke, I., Gonneau, C. & Selosse, M.-A.A. (2012). The role of epiphytism in architecture and evolutionary constraint within mycorrhizal networks of tropical orchids. *Mol. Ecol.*, 21, 5098–5109.

Morlon, H., Lewitus, E., Condamine, F.L., Manceau, M., Clavel, J. & Drury, J. (2016). RPANDA: An R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol. Evol.*, 7, 589–597.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., *et al.* (2012). How to measure and test phylogenetic signal. *Methods Ecol. Evol.*, 3, 743–756.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*.

Paradis, E. (2013). Molecular dating of phylogenies by likelihood methods: A comparison of models and a new information criterion. *Mol. Phylogenet. Evol.*, 67, 436–444.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Peralta, G. (2016). Merging evolutionary history into species interaction networks. *Funct. Ecol.*, 30, 1917–1925.

Perez-Lamarque, B. & Morlon, H. (2019). Characterizing symbiont inheritance during host–microbiota evolution: Application to the great apes gut microbiota. *Mol. Ecol. Resour.*, 19, 1659–1671.

R Core Team. (2020). R: A language and environment for statistical computing.

Rafferty, N.E. & Ives, A.R. (2013). Phylogenetic trait-based analyses of ecological networks. *Ecology*, 94, 2321–2333.

Revell, L.J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3, 217–223.

Rezende, E.L., Lavabre, J.E., Guimarães, P.R., Jordano, P. & Bascompte, J. (2007).

Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448, 925–928.

Rohr, R.P. & Bascompte, J. (2014). Components of phylogenetic signal in antagonistic and mutualistic networks. *Am. Nat.*, 184, 556–564.

Sanders, J.G., Powell, S., Kronauer, D.J.C., Vasconcelos, H.L., Frederickson, M.E. & Pierce, N.E. (2014). Stability and phylogenetic correlation in gut microbiota: lessons from ants and apes. *Mol. Ecol.*, 23, 1268–1283.

Smouse, P.E., Long, J.C. & Sokal, R.R. (1986). Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Syst. Zool.*, 35, 627–632.

Song, S.J., Sanders, J.G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M.W., *et al.* (2020). Comparative analyses of vertebrate gut microbiomes reveal convergence between birds and bats. *MBio*, 11, 1–14.

Vázquez, D.P., Chacoff, N.P. & Cagnolo, L. (2009). Evaluating multiple determinants of the structure of plant–animal mutualistic networks. *Ecology*, 90, 2039–2046.

Xing, X., Jacquemyn, H., Gai, X., Gao, Y., Liu, Q., Zhao, Z., *et al.* (2019). The impact of life form on the architecture of orchid mycorrhizal networks in tropical forest. *Oikos*, 128, 1254–1264.

Xing, X., Liu, Q., Gao, Y., Shao, S., Guo, L., Jacquemyn, H., *et al.* (2020). The architecture of the network of orchid–fungus interactions in nine co-occurring Dendrobium species. *Front. Ecol. Evol.*, 8, 1–10.

649 ## Text boxes:

650

651 **Box 1: Methods for measuring phylogenetic signal in species interactions:**

652

653 Mantel tests (Mantel 1967) were introduced to examine the correlation between
654 two dissimilarity matrices. They have been used to measure the phylogenetic signal in
655 species interactions by computing the correlation between the matrix of phylogenetic
656 distances (for species pairs from guild A for example, Figure 1) and the matrix of
657 'ecological dissimilarities' between the sets of interacting partners (*i.e.* species from
658 guild B interacting with species pairs from guild A). The correlation (R) with $-1<R<1$ is
659 often evaluated using Pearson correlation, that is the mean of the products of the
660 corresponding elements in the two standardized dissimilarity matrices. Alternatively,
661 R can be evaluated using Spearman correlation (computed from dissimilarities
662 transformed into ranks), or Kendall correlation (computed by counting the pairs of
663 observations that have the same rank). The parametric Pearson correlation is
664 statistically more powerful, but makes stronger hypotheses (it assumes a linear
665 relationship) than the non-parametric Spearman and Kendall correlations (which
666 assume only a monotonic relationship). A positive (resp. negative) correlation
667 indicates a phylogenetic (resp. anti-phylogenetic) signal in species interactions. Its
668 significance is evaluated using randomizations by repeatedly permuting one of the
669 original dissimilarity matrices: one-tailed p-values are obtained by comparing the rank
670 of the original correlation R to the randomized correlations. Ecological dissimilarities
671 of interacting partners between two species from guild A can be measured with
672 various indices. Two classical indices are the Jaccard distance, defined as the number
673 of their unshared partners from guild B divided by their total number of partners, and
674 the UniFrac distance, which incorporates phylogenetic relatedness between partners,
675 computed as the fraction of unshared branch length in the phylogenetic tree of their
676 partners from guild B. Both indices also have a weighted version that accounts for

interaction strength (e.g. weighted Jaccard and generalized/weighted UniFrac; Chen *et al.* 2012).

Partial Mantel tests examine the correlation between two dissimilarities matrices while accounting for a third dissimilarity matrix (Smouse *et al.* 1986). When testing for phylogenetic signal in bipartite interaction networks, these tests are useful for controlling the phylogenetic signal in the number of partners; indeed similarity in the number of partners can decrease the value of ecological dissimilarity metrics independently of the identity of the partners, such that a phylogenetic signal in the number of partners can generate a phylogenetic signal in species interactions that is not linked to an evolutionary conservatism of interacting partners. Partial Mantel tests therefore investigate the correlation between phylogenetic and ecological dissimilarities while controlling for the absolute difference in the degrees (*i.e.* the number of partners) between species pairs (Rezende *et al.* 2007).

The Phylogenetic bipartite linear model (PBLM; Ives & Godfray 2006) assumes that interaction strengths between species from guilds A and B are determined by (unobserved) traits that evolve on the two phylogenies each following a simplified Ornstein-Uhlenbeck process (Blomberg *et al.* 2003). The strength of interaction between two species is assumed to be given by the product of their two traits. Under these assumptions, d_A and d_B can be estimated from the two phylogenies and the matrix of interaction strengths using generalized least squares (Ives & Godfray 2006). d_A and d_B are then interpreted as a measure of phylogenetic signal in species interactions. If $d=1$, the traits evolve as Brownian processes; if $d=0$, there is no effect of the phylogeny (similar as evolution on a star phylogeny); $0<d<1$ generates stabilizing selection and $d>1$ disruptive selection. Ives & Godfray (2006) proposed two approaches to assess the significance of the signal. The simplest consists in comparing the mean square errors (MSE) of the generalized least squares regression to the same MSE obtained using star phylogenies (MSE_{star}). $MSE<MSE_{star}$ is interpreted as a significant phylogenetic signal in species interactions. The second approach uses a bootstraping strategy to build 95% confidence intervals around the estimated d_A and d_B values: the null hypothesis

(absence of phylogenetic signal in guild A, resp. B) is rejected if the confidence interval around d_A (resp. d_B) does not incude 0. While designed primarily for applications to bipartite networks characterized by matrices of interaction strengths (*e.g.* net attack rate of a parasitoid on its hosts), PBLM has been applied to weighted networks characterized by matrices of interaction abundance (*i.e.* the number of times the interaction has been observed) and unweighted (binary) networks, using 1 for the interaction strength when species interact and 0 otherwise (Ives & Godfray 2006; Vázquez *et al.* 2009; Jacquemyn *et al.* 2011; Martos *et al.* 2012; Xing *et al.* 2020).

714 **Box 2: Recommended guidelines to measure phylogenetic signal in species**
715 **interactions within bipartite ecological networks.**
716 This guideline is composed of two fixed steps followed by two optional ones and can
717 be applied as soon as a bipartite interaction network (with or without abundances)
718 and at least the phylogenetic tree of guild A are available. The phylogenetic tree does
719 not need to be binary, rooted, or ultrametric.
720
721 **Step 1:** The first step consists in testing for phylogenetic signal in species interactions
722 for guild A (*i.e.* whether closely related species from guild A tend to interact with
723 similar partners from guild B) using a one-tailed simple Mantel test. This step requires
724 to pick an ecological distance (UniFrac distances are recommended compared to
725 Jaccard distances) and a type of correlation (Pearson correlation by default).
726
727 **Step 2:** Next, to assess whether a phylogenetic signal in species interactions really
728 comes from the identity of species interactions, the second step consists in testing
729 whether there is phylogenetic signal in the number of partners of guild A (*i.e.* whether
730 closely related species from guild A tend to interact with the same number of partners
731 from guild B) using a one-tailed simple Mantel test.
732
733 **Option 1:** Clade-specific phylogenetic signal in guild A can be tested using simple
734 Mantel tests while correcting for multiple testing (*e.g.* Bonferroni correction). It can be
735 used to test whether some clades present different intensities of phylogenetic signal
736 (*e.g.* because of higher specificity).
737
738
739 **Option 2:** The robustness of the findings can be tested by looking at how the
740 conclusions might be affected by phylogenetic uncertainty (e.g. using a Bayesian
741 posterior of tree) or sampling bias. The potential effect of sampling bias can be
742 investigated by subsampling all clades to the same number of species.

If a phylogenetic tree for guild B is available, all these steps can be replicated to test for phylogenetic signal in species interaction in guild B.

Figures:

Figure 1: Illustration of the data used to test for phylogenetic signal in species interactions

Toy example of an interaction network between orchids (in green) and mycorrhizal fungi (in brown) with associated phylogenetic trees. The bipartite interaction network between two guilds A (here the orchids) and B (the fungi) is represented by a matrix which elements indicate either whether or not species interact (*i.e.* 1 if they do and 0 otherwise, 'unweighted' or 'binary' network) or the frequency of the interaction ('weighted' network; for example here we indicated the number of times a given pairwise interaction has been observed using shades of gray from white (no interaction) to dark gray (many interactions)). Each guild is also characterized by a rooted phylogenetic tree, used to compute phylogenetic distances between pairs of species.

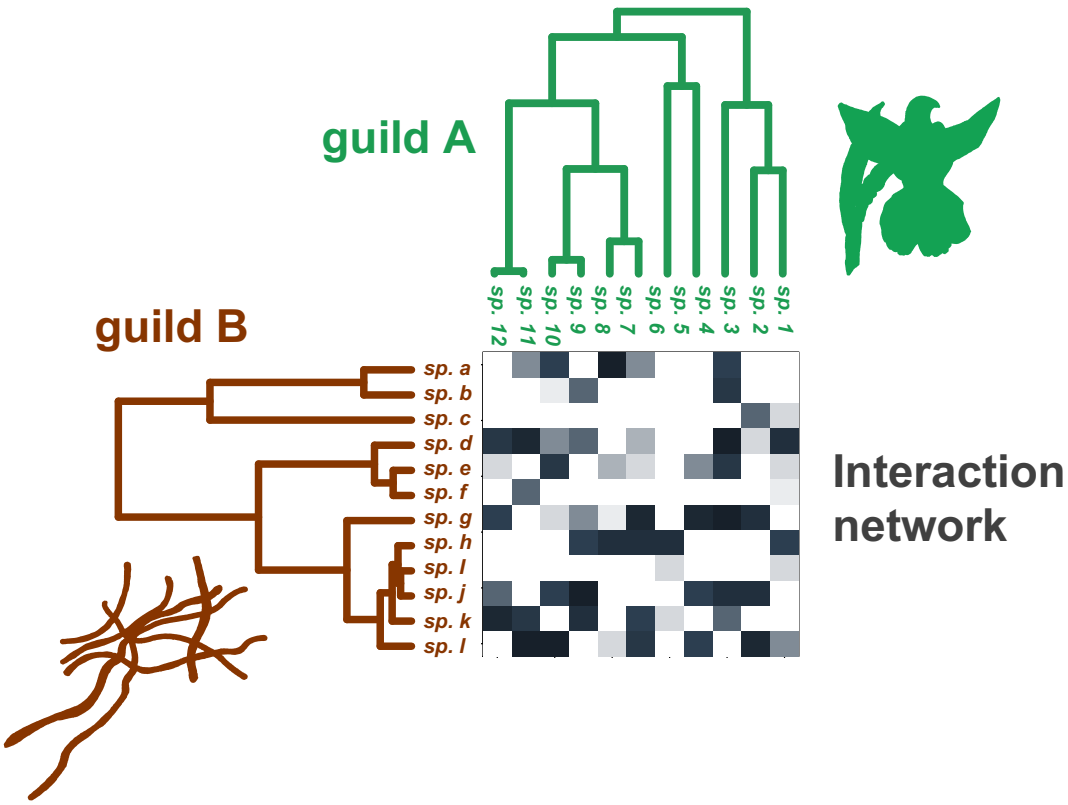


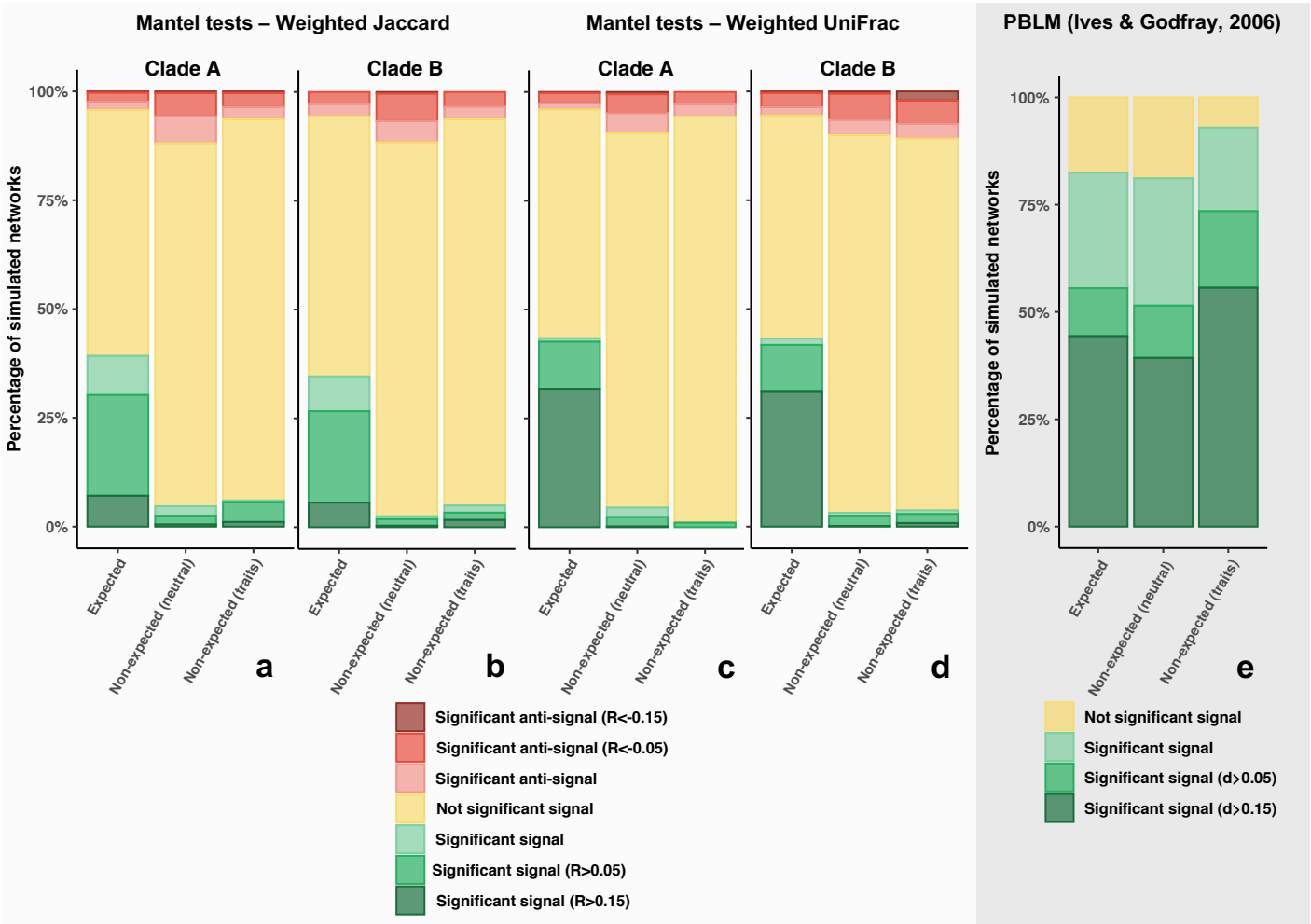
Figure 2: Statistical performances of the simple Mantel tests and the Phylogenetic bipartite linear model (PBLM; Ives & Godfray, 2006)

For each panel, the simulations are divided between networks where phylogenetic signal in species interactions is expected (*i.e.* networks (i) simulated with an effect of the traits on individual fitness - antagonistic and mutualistic simulations - and (ii) presenting traits that are phylogenetically conserved – see Supplementary Figure 2) and networks where phylogenetic signal in species interactions is not expected (*i.e.* neutral simulations ($\alpha = 0$) or simulated networks where we observed no phylogenetic signal in the traits).

a-d: Phylogenetic signals in species interactions estimated using simple Mantel tests with Pearson correlation (R) in the guilds A (a, c) and B (b, d). The different panels in rows correspond to the 2 tested ecological distances: weighted Jaccard (a, b) or weighted UniFrac (c, d) distances. One-tailed Mantel tests between phylogenetic distances and ecological distances were performed using 10,000 permutations. In each panel, the bars indicate the percentage of simulated networks that present a significant positive correlation (p-value>0.05 for the test of phylogenetic signal), a significant negative correlation (in red; p-value>0.05 for the test of anti-phylogenetic signal), or no significant correlation (in yellow; both p-values>0.05).

e: Phylogenetic signals estimated using PBLM. For a given combination of parameters, the bar indicates the percentage of simulated networks that present no significant (in yellow; $MSE \geq MSE_{star}$) or a significant (green; $MSE < MSE_{star}$) phylogenetic signal. Phylogenetic signals are shaded from light green to dark green according to the strength of the signal (e.g. in dark green if $d_A > 0.15$ or $d_B > 0.15$). PBLM were run on the weighted networks.

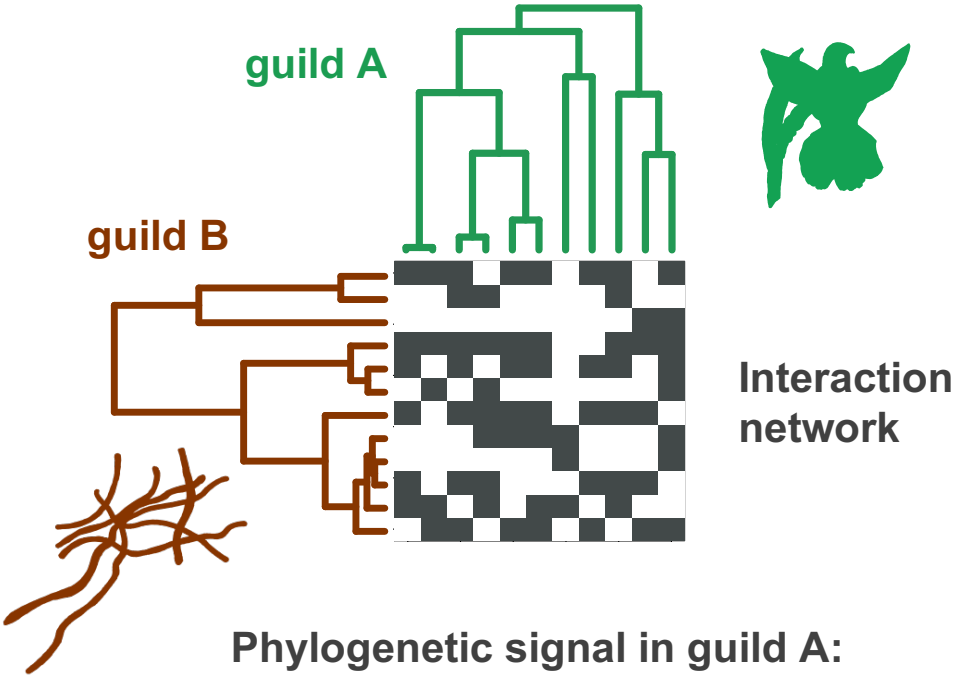
In each panel, the first bar indicates the statistical power of the test, whereas the second and third bar indicate the type-I error rate of the test.


Mantel tests – Weighted Jaccard
Mantel tests – Weighted UniFrac
PBLM (Ives & Godfray, 2006)
Clade A
Clade B
Clade A
Clade B
Percentage of simulated networks
100%
75%
50%
25%
0%
Expected
Non-expected (neutral)
Non-expected (traits)
a
b
c
d
e
Significant anti-signal (R<-0.15)
Significant anti-signal (R<-0.05)
Significant anti-signal
Not significant signal
Significant signal
Significant signal (R>0.05)
Significant signal (R>0.15)
Not significant signal
Significant signal
Significant signal (d>0.05)
Significant signal (d>0.15)

791

Figure 3: Illustration of the recommended guidelines for measuring phylogenetic signal in species interactions within bipartite interaction networks (Box 2).

At the top, a toy example of an interaction network between orchids (in green) and mycorrhizal fungi (in brown) is informed with the phylogenetic trees of each guild. For each step of the guidelines (Box 2), an example of the corresponding function available in the R-package RPANDA is indicated in grey.



Phylogenetic signal in guild A:

Step 1: test the phylogenetic signal in the **species interactions** (simple Mantel test)
(i) choice of ecological distances (Jaccard, UniFrac…)
(ii) with or without interaction abundances

phylosignal_network(network, tree_A, tree_B, method = "GUniFrac", correlation = "Pearson")

Step 2: test the phylogenetic signal in the **degree of generalism** (simple Mantel test)

phylosignal_network(network, tree_A, method = "degree", correlation = "Pearson")

Option 1: investigate clade-specific phylogenetic signals (simple Mantel tests with Bonferroni correction)

phylosignal_sub_network(network, tree_A, tree_B, method = "GUniFrac", correlation = "Pearson")

Option 2: test the robustness of the findings to phylogenetic uncertainty and/or sampling bias

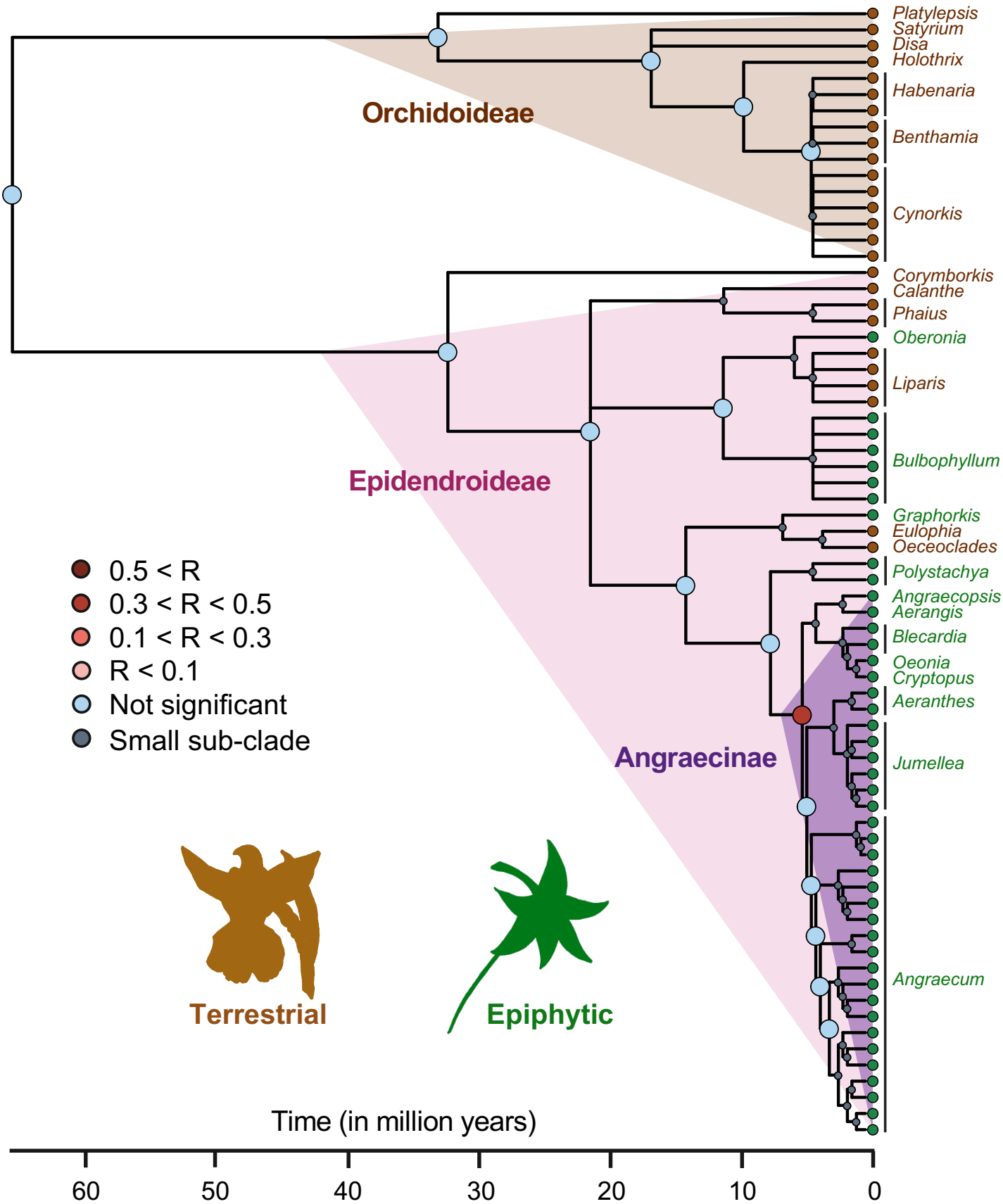
(repeat for guild B)

798

Figure 4: Empirical application on an orchid-fungus interaction network from La Réunion island (Martos *et al.,* 2012): the clade-specific analyses of phylogenetic signal in species interactions revealed a significant phylogenetic signal in the epiphytic subtribe Angraecinae.

The orchid phylogeny (Martos *et al.,* 2012) is represented with its nodes colored according to the results of the Mantel test performed on the corresponding sub-network: in blue if non-significant, in grey when the node has less than 10 descendent species (the Mantel test was not performed), and in red when the phylogenetic signal is significant. Each one-tailed simple Mantel test was performed using the Pearson correlation and 100,000 permutations and its significance was evaluated while correcting for multiple testing (Bonferroni correction).

For each species, its habitat (terrestrial or epiphytic) is indicated at the tips of the tree and the main orchid clades are highlighted in colors. Only the genera are indicated at the tips of the tree (see Supplementary Figure S28 for the species list).


Orchidoideae
Epidendroideae
Angraecinae
Platylepsis
Satyrium
Disa
Holothrix
Habenaria
Benthamia
Cynorkis
Corymborkis
Calanthe
Phaius
Oberonia
Liparis
Bulbophyllum
Graphorkis
Eulophia
Oeceoclades
Polystachya
Angraecopsis
Aerangis
Blecardia
Oeonia
Cryptopus
Aeranthes
Jumellea
Angraecum
0.5 < R
0.3 < R < 0.5
0.1 < R < 0.3
R < 0.1
Not significant
Small sub-clade
Terrestrial
Epiphytic
Time (in million years)
60
50
40
30
20
10
0