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)

Abstract

1. 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?

2. Here we use simulations to test the statistical performances of the two approaches that are the most widely used in the field: Mantel tests and the Phylogenetic Bipartite Linear Model (PBLM). Mantel tests investigate the correlation between phylogenetic distances and dissimilarities in sets of interacting partners, while the PBLM is a model-based approach that relies on strong assumptions on how interactions evolve.

3. We find that PBLM often detects phylogenetic signal when it should not. Simple Mantel tests 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 (*i.e.* closely related species are found to interact with dissimilar partners). 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.

4. 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, Jordano, Melian, & Olesen, 2003; Fontaine et al., 2011; Martos et al., 2012; Bascompte & Jordano, 2013), which are often summarized using bipartite interaction networks (Bascompte & Jordano 2013; Fig. 1). 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, Lavabre, Guimarães, Jordano, & Bascompte, 2007; Vázquez, Chacoff, & Cagnolo, 2009; Krasnov et al., 2012; Elias, Fontaine, & Frank Van Veen, 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, Garland, & Ives, 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.* Cattin *et al.* 2004; 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 ("phylogenetic signal in the number of partners"; Rezende *et al.* 2007; Jacquemyn *et al.* 2011; Aizen *et al.* 2016). Mantel tests, that are easy and fast to run and that do not rely on strong hypotheses, have therefore been vastly used to test for phylogenetic signal in species interactions in empirical networks (Cattin et al., 2004; Rezende et al., 2007; Jacquemyn et al., 2011; Elias et al., 2013; Fontaine & Thébault, 2015). Besides these correlative approaches, several model-based approached have been developed (Ives & Godfray, 2006; Rafferty & Ives, 2013; Hadfield, Krasnov, Poulin, & Nakagawa, 2014; Li, Dinnage, Nell, Helmus, & Ives, 2020), but their complexity and their high computational requirements have largely prevented their general use on empirical networks, with the exception of the Phylogenetic Bipartite Linear Model (PBLM, Ives & Godfray 2006), which has been widely used to test for phylogenetic signal in species interactions in a variety of networks, *e.g.* in host-parasite, plant-fungus, and pollination networks (Ives & Godfray, 2006; Martos et al., 2012; Martín González et al., 2015; Xing et al., 2020). In short, PBLM assumes that interaction strengths between species from the two guilds are determined by (unobserved) traits that evolve on the two phylogenies each following a simplified Ornstein-Uhlenbeck process (Blomberg et al., 2003). PBLM performs a phylogenetic regression to infer the Ornstein-Uhlenbeck parameters, which are then interpreted in terms of phylogenetic signal (Ives & Godfray 2006).

Mantel tests and PBLM sometimes provide contradictory conclusions on empirical data and this is difficult to interpret because the statistical performances of the two approaches have never been compared (Peralta, 2016). Importantly, the statistical performances of PBLM have not been tested. Here, we use simulations, to

perform a comparative analysis of the statistical performances of these two approaches. We consider both weighted and unweighted bipartite interaction networks between species from two guilds A and B (Fig. 1). Our results lead us to propose an alternative approach for measuring phylogenetic signal in interaction networks, the sequential Mantel test. 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 illustrate them on 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 (see Maliet *et al.* 2020 for a complete description of the model), to generate interaction networks with or without phylogenetic signal between two guilds interacting in a mutualistic, antagonistic, or neutral way. In short, 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. Networks simulated using *BipartiteEvol* show typical structural properties observed in empirical networks, including significant nestedness and/or modularity according to the sets of simulated parameters (Maliet et al., 2020): in general, antagonistic networks ($\alpha_A<0$) are modular, while neutral and mutualistic networks ($\alpha_A=0$ or $\alpha_A<0$) tend to be nested. Here, we considered that each combination of traits forms a new species instead of using the species delineation of the original *BipartiteEvol* model (Maliet et

al., 2020). This increased our ability to generate phylogenetic signal but did not affect network structure (results not shown). Therefore, our simulations provided a range of realistic networks.

Under the *BipartiteEvol* model, 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 to associate with dissimilar partners) is expected if there is anti-phylogenetic signal in species traits (*i.e.* closely related species have dissimilar traits) and $\alpha \neq 0$.

Using the R-package RPANDA (Morlon et al., 2016; R Core Team, 2020), we simulated a total of 2,400 interaction networks with individuals characterized by a six-dimensional trait. To obtain a wide range of network sizes, we considered a total number of 500, 1,000, 2,000, 3,000, 4,000, or 5,000 pairs of interacting individuals per simulation. For each size, we simulated the evolution of 100 neutral networks (α_A=0 ; 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; **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 (**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; 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). We used a mutation rate μ=0.01 and followed the interacting individuals during 50^6 death events. At the end, we extracted for each guild a species tree from its genealogy by randomly selecting one individual per species (Fig. S1), we also recorded the number of individuals belonging to each species, and counted the number of occurrences of each interspecific interaction; we then reconstructed the corresponding weighted interaction network.

We separated the 2,400 simulated networks between those for which we should expect a phylogenetic signal in species interactions and those for which we should not. We did not expect phylogenetic signal in species interactions in neutral networks and

in non-neutral networks with no phylogenetic signal in species traits. Conversely, we expected phylogenetic signal in non-neutral networks with phylogenetic signal in species traits. For simplicity and consistency with the rest of the paper, phylogenetic signal in species traits was evaluated using Mantel tests (Pearson correlation) between phylogenetic distances and trait distances computed as the Euclidian distances between trait values for each species pair.

Computing phylogenetic signal in species interactions

We computed phylogenetic signals in species interactions in the simulated networks using Mantel tests and PBLM. Complete descriptions of these methods are available in Supplementary Methods 2. Mantel tests and PBLM rely on different strategies to evaluate the significance of the phylogenetic signal, and it could be argued that results of these significance tests are not directly comparable. Our approach is to follow the methodologies traditionally used in empirical studies, and to compare the conclusions (detection or not of a phylogenetic signal) that an empiricist would make from these analyses.

Mantel tests: We evaluated the phylogenetic signal in species interactions in guilds A and B separately using simple Mantel tests between phylogenetic and ecological (set of interacting partners) distances. Ecological distances were measured both without accounting for evolutionary relatedness of the interacting partners, using (weighted or unweighted) Jaccard, and accounting for relatedness using (weighted or unweighted) UniFrac distances (Supplementary Methods 2). Accounting for evolutionary relatedness of the interacting partners can be particularly relevant for organisms with uncertain species delineations (e.g. microorganisms delineated using only molecular data (Martos et al., 2012; Sanders et al., 2014)). We used Pearson, Spearman, and Kendall correlations (R) by extending the *mantel* function in the R-package ecodist (Goslee & Urban, 2007); the significance of each correlation was evaluated using 10,000

192 permutations, except for the computationally intensive Kendall correlation (100
193 permutations only). For each network, we considered that there was a significant
194 phylogenetic signal (resp. anti-phylogenetic signal) if the correlation coefficient (R)
195 was higher (resp. lower) than >95% of the randomized correlations; we computed the
196 p-value of each one-tailed Mantel test as the fraction of the randomized correlations
197 above (resp. below) the original value.
198
199 PBLM: To estimate phylogenetic signal based on PBLM, we modified the function *pblm*
200 from the R-package picante (Kembel et al., 2010) to more efficiently perform matrix
201 inversions and handle large interaction networks. In short, the parameters d_A and d_B
202 of the Ornstein-Uhlenbeck processes of PBLM were estimated using generalized least
203 squares (Ives & Godfray 2006). d_A and d_B are interpreted as a measure of phylogenetic
204 signal in species interactions: if d=0, there is no effect of the phylogeny (similar as
205 evolution on a star phylogeny, *i.e.* no phylogenetic signal); 0<d<1 generates stabilizing
206 selection (*i.e.* phylogenetic signal) and d>1 disruptive selection (*i.e.* anti-phylogenetic
207 signal). We followed Ives & Godfray (2006; Supplementary Methods 2) by considering
208 that the phylogenetic signal is significant when the mean square error (MSE) of the
209 model is smaller than that obtained using star phylogenies (MSE_{star}); we also used a
210 more stringent criterion by considering that the signal is significant when the MSE is
211 at least 5% lower than MSE_{star}. Finally, we applied the bootstrapping method of Ives &
212 Godfray (2006; Supplementary Methods 2) for the smallest networks. Yet, the
213 computational cost of bootstrapping is very high for large networks. Indeed, even on
214 networks of intermediate sizes (between 50 and 100 species per guild), the PBLM
215 inference can take several days; it thus prevented us from applying it on larger
216 networks.
217
218 **Confounding effect of the phylogenetic signal in the number of partners**
219

To test the performances of the partial Mantel test at measuring phylogenetic signal in species interactions while controlling for the number of partners (Supplementary Methods 2), we first performed partial Mantel tests between phylogenetic and ecological distances, while controlling for the absolute differences in degrees, on the networks simulated with *BipartiteEvol*. There is no reason that *BipartiteEvol* simulations generate a phylogenetic signal in the number of partners, and we verified this by performing Mantel tests between phylogenetic distances and degree differences. Partial Mantel tests were performed to assess whether they lose power compared to simple Mantel tests. If they do not suffer power loss, partial Mantel tests applied to BipartiteEvol simulations should be significant when simple Mantel tests are significant.

Second, we assessed whether partial Mantel tests successfully correct for phylogenetic signal in the number of partners using networks simulated under a process that generate phylogenetic conservatism in the number, but not the identity, of interacting partners (*i.e.* partial Mantel tests should not be significant when applied to such networks). To simulate network with only phylogenetic conservatism in the number of partners in guild A, we first simulated phylogenetic trees for guilds A and B using *pbtree* (R-package phytools; Revell 2012) with a number of species uniformly sampled between 40 and 150 by guild. Next, we simulated the number of partners of the species from guild A using an Ornstein-Uhlenbeck process with an attraction toward 0, a variance of 0.1 (noise of the Brownian motion), and a selection strength (a_A) ranging from 5 (strong stabilizing effect, weak phylogenetic signal) to 0 (Brownian motion, strong phylogenetic signal). We computed the number of partners per species by calibrating the simulated values between 1 and the number of species in guild B and taking the integer part. For each a_A value (5, 1, 0.5, 0.05, or 0), we performed 100 simulations using *mvSIM* (R-package mvMORPH; Clavel *et al.* 2015). Finally, for each species in A, we attributed the corresponding number of partners in B at random to obtain binary networks. We checked that our simulations indeed generated a signal in

the number of partners by performing simple Mantel tests between phylogenetic and degree difference distances. Finally, we performed on each simulated network a partial Mantel test.

Given the poor performances of partial Mantel tests (see Results), we tested whether using sequential Mantel tests would provide a good alternative: based on simple Mantel tests, we consider that there is a phylogenetic signal in the identity of the partners if there is a phylogenetic signal in species interactions and no phylogenetic signal in the number of partners. We applied this sequential testing to all our simulated networks.

Effect of phylogenetic uncertainty, sampling asymmetry, and network heterogeneity on measures of phylogenetic signal in species interactions

Unlike simulations (such as those provided by *BipartiteEvol*), empirical bipartite networks suffer from phylogenetic uncertainty (*e.g.* in the microbial partners' tree when studying host-associated microbiota – which often prevents accounting for evolutionary relatedness; *i.e.* using UniFrac distances), sampling asymmetry (*i.e.* one side of the network is more thoroughly sampled than the other), and network heterogeneity (*i.e.* different sub-clades in the network have different levels of phylogenetic signal). We performed additional analyses to investigate the effect of these aspects on phylogenetic signals in species interactions measured using simple Mantel tests.

First, we tested the effect of phylogenetic uncertainty in the partners' tree on the measure of phylogenetic signal when evolutionary relatedness is accounted for (*i.e.* using UniFrac distances). We performed these analyses to assess whether accounting for the partners' evolutionary relatedness remains advantageous (see Results) when phylogenetic uncertainty is high. To add some variability in the phylogenetic tree of

guild B (resp. A) used to compute the UniFrac distances between species pairs from guild A (resp. B), we first simulated, on the original partners tree, the evolution of a short DNA sequence and then reconstructed the tree from the simulated DNA alignment using neighbor-joining (*nj* function, R-package APE (Paradis, Claude, & Strimmer, 2004)). We used *simulate_alignment* (R-package HOME; Perez-Lamarque & Morlon 2019) to simulate sequences of length 75, 150, 300, 600, or 1,200 base-pairs, with 30% of variable sites, and a substitution rate of 1.5 (shorter fragments should result in noisier phylogenies).

Second, we tested the influence of sampling asymmetry on measures of phylogenetic signal. Empirical networks are often an incomplete representation of the actual interactions between two guilds because they are under-sampled, and frequently, in an asymmetrical way. For instance, by sampling targeted species from guild A, observed networks are constituted by few species from guild A which have the complete set of their partners and by often more species from guild B which have an incomplete set of their partners (as they likely interact with unsampled species from guild A). We tested the influence of such sampling asymmetry by selecting only 10% of the most abundant species from guild A in each simulated network (while retaining at least 10 species) and computed phylogenetic signal in these asymmetrically-subsampled networks.

Third, both Mantel tests and PBLM neglect the heterogeneity within networks. Indeed, a non-significant phylogenetic signal at the level of the entire network can potentially hide a sub-clade of species presenting significant phylogenetic signal. Alternatively, a phylogenetic signal in the entire network may be driven by only two sub-clades of guilds A and B, while the other sub-clades present no significant phylogenetic signal. To explore the potential heterogeneity of the phylogenetic signal within one guild, one possibility is to apply Mantel tests to the sub-networks formed by a given sub-clade (*e.g.* Song *et al.* 2020). For each node of the tree of guild A having

307 at least 10 descendants, we estimated the clade-specific phylogenetic signal using a
308 Mantel test investigating whether closely related species from this sub-clade of A tend
309 to interact with similar partners (and *vice-versa* for guild B). Using UniFrac distances,
310 we performed the Mantel tests with 100,000 permutations, and introduced a
311 Bonferroni correction for multiple testing to keep a global alpha-risk of 5%. To test this
312 approach, we generated synthetic networks with known sub-clade signal by
313 artificially combining networks simulated under neutrality with networks simulated
314 with the mutualistic parameters **v** (see Results). We grafted each "mutualistic"
315 phylogenetic tree from guilds A and B within a "neutral" phylogenetic tree by
316 randomly selecting a branch, such that it creates a separate module with strong
317 phylogenetic signal. Such simulations could correspond to the evolution of a different
318 niche, *e.g.* terrestrial *versus* epiphytic plants associating with different mycorrhizal
319 fungi (Martos et al., 2012). We then performed our clade-specific analysis of
320 phylogenetic signal and investigated in which nodes we recovered significant
321 phylogenetic signals.
322
323 **General guidelines and illustration with application on the orchid-fungus**
324 **mycorrhizal network from La Réunion**
325
326 We used our results and other empirical considerations to provide general
327 guidelines for testing for phylogenetic signal in interaction networks. We illustrated
328 these guidelines by applying them in a network between orchids and mycorrhizal
329 fungi from La Réunion island (Martos et al., 2012). This network encompasses 70
330 orchid species (either terrestrial or epiphytic species) and 93 molecularly-identified
331 fungal partners (defined according to 97% sequence similarity; Martos *et al.* 2012). We
332 gathered the maximum-likelihood plant and fungal phylogenies on TreeBASE (Study
333 Accession S12721), calibrated the orchid phylogeny using a relaxed clock with *chronos*
334 (Paradis, 2013), and arbitrarily added 10 million-years-old polytomies in unresolved
335 genera to obtained a species-level orchid phylogeny.

Results

Expected phylogenetic signals in species interactions in *BipartiteEvol* networks

The networks simulated using *BipartiteEvol* gave realistic ranges of sizes for guilds A and B (from less than 50 to more than 250 species; Fig. S2) and connectance values (*i.e.* ratios of realized interactions, between 5 and 20%; Fig. S3).

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

Computing phylogenetic signal in species interactions in *BipartiteEvol* networks

Using Mantel tests, as expected, we did not find significant phylogenetic signals in species interactions for most neutral networks or for networks with no signal in species traits (Fig. 2, Figs. S5-6-7): the type-I error rate was below 5%, corresponding to the alpha-risk of the test (Table S1), with one notable exception for small networks

364 when using weighted Jaccard distances and Pearson correlations (~8% type-I error).
365 Conversely, we detected a significant unexpected anti-phylogenetic signal in more
366 than 10% of the simulated networks, in particular in the small ones (Fig. 2, Figs. S5-6-
367 7).
368
369 Many mutualistic or antagonistic networks where we expected a phylogenetic
370 signal in species interactions (*i.e.* non-neutral networks with signal in species traits)
371 presented no significant signal with Mantel tests (Fig. 2, Figs. S5-6-7), in particular
372 those simulated with low α_A and α_B values (*e.g.* antagonism **vii**), where non-neutral
373 effects were weak. In mutualistic networks, phylogenetic signals in species
374 interactions were present only when there was a large asymmetry in the effects of trait
375 matching on the fitnesses of the species from guilds A or B (case **v**: α_A=1; α_B=0.01), *i.e.*
376 when only one guild was specialized. Conversely, in antagonistic networks,
377 phylogenetic signals were found mainly when trait matching had a strong impact on
378 the fitness of guild B (the obligate parasites/predators; $\alpha_B \geq 0.1$). Additionally, when
379 phylogenetic signal was significant in one guild, it was generally also significant in the
380 other; in antagonistic networks, the signal was usually higher in guild A compared to
381 guild B (Figs. S5-6-7).
382
383 The statistical power of Mantel tests measuring phylogenetic signal in species
384 interactions seems to be modulated by network size, as phylogenetic signals were less
385 often significant but generally stronger in smaller networks (Figs. S5-6-7). Moreover,
386 Mantel tests based on Pearson correlations had higher power than Spearman and
387 Kendall correlations (Figs. S5-6-7) and weighted UniFrac distances outperformed
388 other ecological distances in terms of power (Figs. S5-6-7; Table S2).
389
390 When using mean square errors to evaluate the significance of PBLM, we found
391 a significant phylogenetic signal in most of the simulated networks including when
392 we did not expect any (Fig. 2e). The propensity of PBLM to detect phylogenetic signal

decreased in large unweighted networks, but the type-I errors remained >30%, including when using a more stringent significance cutoff (Figs. S8-9). Similar results were obtained when bootstrapping to evaluate the significance (Fig. S10).

Confounding effect of the phylogenetic signal in the number of partners

As expected, tests of phylogenetic signal in the number of partners were non-significant in the large majority of the *BipartiteEvol* networks, especially the larger ones (Fig. S11). We did however observe significant correlations between ecological distances and degree difference distances (Fig. S12). Partial Mantel tests testing for phylogenetic signal in species interactions while accounting for phylogenetic signal in the number of partners had similar type-I error and power as simple Mantel tests (Figs. S5-13; Table S2). Performing sequential Mantel tests decreased the statistical power by less than 2% (Table S2).

Networks simulated with phylogenetic conservatism in the number, but not the identity of partners covered a realistic range of sizes (Fig. S14). As expected, Mantel tests revealed significant phylogenetic signals in the number of partners in >60% of these networks, with an increasing percentage of significant tests with decreasing a_A (*i.e.* increasing conservatism in the number of partners; Fig. S18). We found significant correlations between degree differences and ecological distances in most of these simulated networks (Fig. S15). As a result, simple Mantel tests testing for phylogenetic signal in species interactions without accounting for phylogenetic signal in the number of partners were frequently significant (>30%; Fig. S16; Table S3). Partial Mantel tests controlling for degree differences slightly decreased the proportion of false-positives, but it remained high (type-I error >25%; Fig. S17). In addition, partial Mantel tests detected a spurious significant anti-phylogenetic signal in species interactions in >15% of the networks (Fig. S17). Conversely, only few networks with a significant simple Mantel test in species interactions did not produce a significant simple Mantel test in

the number of partners, such that sequential Mantel tests had only a ~7% type-I error rate (Table S3).

Effect of phylogenetic uncertainty, sampling asymmetry, and network heterogeneity on measures of phylogenetic signal in species interactions

The statistical power of Mantel tests using UniFrac distances decreased, as expected, when the length of the simulated DNA sequences decreased (*i.e.* when phylogenetic uncertainty increased; Fig. S19). However, even when the simulated DNA sequences were the shortest (75 base-pairs), resulting in very noisy reconstructed partners' tree (Fig. S20), the statistical power of the Mantel tests using UniFrac distances remained larger than when using Jaccard distances (Fig. S19).

Our results on the statistical performance of tests of phylogenetic signal were similar when considering sampling asymmetry (Figs. S21-24): PBLM spuriously detected phylogenetic signal when it should not, and Mantel tests had decent statistical performances, especially when using weighted UniFrac distances. In addition, the correlations of the Mantel tests in guild A were generally higher when significant (Fig. S23).

Our clade-specific tests of phylogenetic signal using Mantel tests while correcting for multiple testing recovered a significant phylogenetic signal in 82% of the nodes where mutualism originated (Fig. S25), as well as in most of the ascending nodes. Conversely, we did not find spurious phylogenetic signals in nodes with only neutrally-evolving lineages (Fig. S25).

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

Figure 3 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)). 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 (Supplementary Methods 2). 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 that extend PBLM (Rafferty & Ives, 2013; Hadfield et al., 2014; Li et al., 2020) allow to infer parameters thought to reflect the phylogenetic structure of interactions networks, while controlling for phylogenetic

498 signal in the number of parterns as well as heterogeneity in sampling effort (Hadfield
499 *et al.*, 2014). It would have been ideal to include these approaches in our comparative
500 analyses, but this was prohibited by their computational cost. Indeed, preliminary
501 analyses applying the Bayesian approach of Hadfield *et al.* (2014) on a few networks
502 ran several days without reaching convergence. Because of these high computational
503 demands, these methods are never used as a routine to measure phylogenetic signal
504 in species interactions in emprical studies, which is either done using Mantel tests or
505 PBLM. Future model developments of such approaches would thus benefit from faster
506 inferences; our results on PBLM highlight the need to thoroughly test these approaches
507 with simulations before they are applied to empirical systems and biological
508 conclusions are drawn.
509
510 We found that simple Mantel tests have a moderate statistical power (from
511 >90% to <5% depending on the strength of the traits on individuals' fitness) and a
512 reasonable type-I error rate (<5%) when testing for phylogenetic signal in species
513 interactions. Not surprisingly, these tests have a higher power for larger simulated
514 networks. Hence, although simple Mantel tests might fail at detecting low
515 phylogenetic signal, we can trust their results when they are significant. On the
516 contrary, we found a high proportion of simulated networks (5-10%) presenting a
517 significant anti-phylogenetic signal in species interactions, although we did not expect
518 any in our simulations (because we did not observe any anti-phylogenetic signal in
519 species traits). False-positives are therefore frequent when testing for anti-
520 phylogenetic signal using simple Mantel tests and detection of such signal in empirical
521 networks should be interpreted with caution.
522 In addition, Pearson correlations performed better than Spearman and Kendall
523 correlations, which is somewhat surprising, as correlations between phylogenetic and
524 ecological distances are not particularly expected to be linear: Spearman and Kendall
525 correlations have less stringent hypotheses, as they only assume monotonicity
526 (Supplementary Methods 2), but they probably lose information. We also reported that

using ecological distances that consider interaction abundances and phylogenetic relatedness of the partners, such as weighted UniFrac distances, significantly improves the detection of phylogenetic signal, even when reconstructed partners trees are not robust. Given that species delineation may be somewhat arbitrary, especially for microbial interactors, and that Jaccard distances are directly sensitive to species delineation (Sanders et al., 2014), we advocate the use of weighted UniFrac distances. An exception might be if communities of interactors differ mainly in terms of recently diverged species; in this case Jaccard distances may perform better, as UniFrac distances emphasize differences in long branches rather than recent splits (Sanders et al., 2014). Finally, we found that multiple simple Mantel tests combined with a Bonferroni correction perform rather well to investigate clade-specific phylogenetic signals. Such an approach can therefore be valuable for measuring local phylogenetic signal in large "meta-networks", such as those describing host-microbiota phylosymbiosis (Song et al., 2020), which likely have heterogeneous phylogenetic signals across the network.

While simple Mantel tests have satisfactory statistical performances, these tests do not control for the potential confounding effect of phylogenetic signal in the number of partners. Partial Mantel tests are frequently used for investigating phylogenetic signal in species interactions while controlling for signal in the number of partners; however, we found that they often detected significant signals in species interactions when we simulated signals in only the number of partners. Thus, partial Mantel tests fail at discerning whether evolutionary relatedness strictly affects the identity of partners, independently of the total number of partners associated with each species (Rezende et al., 2007). This corroborates the poor statistical performances of partial Mantel tests frequently observed in other contexts (Harmon & Glor, 2010; Guillot & Rousset, 2013). We therefore suggest to perform sequential simple Mantel tests, testing first for phylogenetic signal in species interactions, and if significant, testing for phylogenetic signal in the number of partners. If there is no signal in the number of

partners but a signal in interactions, then we can safely conclude that evolutionary relatedness strictly affects the identity of partners. This approach has a low type-I error rate and a very limited power decrease; however, it does not allow testing if there is a specific signal in species identity when there is a signal in the number of partners. A hint at whether signal in species interactions is entirely due to signal in the number of partners or not can be gained by comparing the correlation coefficients obtained when correlating phylogenetic distance to ecological distance *versus* degree distance.

By definition, phylogenetic signals in species interactions measure general patterns that are not informative of the processes at play (Losos, 2008). A better understanding of the ecological and evolutionary processes playing a role in the assembly of interaction networks (Harmon et al., 2019) will require developing integrative process-based approaches, for instance inference machineries for eco-evolutionary models such as *BipartiteEvol*. Classical inferences (generalized least-squares or likelihood-based approaches) might be challenging for such complex models (Hadfield et al., 2014), but strategies such as machine learning provide promising alternatives.

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).

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:

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

613 networks can be found at https://github.com/BPerezLamarque/Phylosignal_network.

614 Amended functions of *BipartiteEvol* are also included in RPANDA.

615

616 ## Conflict of Interest statement:

617 The authors declare that there is no conflict of interest.

618 **References:**

619

620 Aizen, M. A., Gleiser, G., Sabatino, M., Gilarranz, L. J., Bascompte, J., & Verdú, M.
621 (2016). The phylogenetic structure of plant-pollinator networks increases with
622 habitat size and isolation. *Ecology Letters*, *19*(1), 29–36. doi:10.1111/ele.12539
623 Bascompte, J., & Jordano, P. (2013). *Mutualistic networks*. *Monographs in population*
624 *biology* (Princeton, Vol. 7). Princeton: Princeton University Press.
625 doi:10.1515/9781400848720
626 Bascompte, J., Jordano, P., Melian, C. J., & Olesen, J. M. (2003). The nested assembly
627 of plant-animal mutualistic networks. *Proceedings of the National Academy of*
628 *Sciences*, *100*(16), 9383–9387. doi:10.1073/pnas.1633576100
629 Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in
630 comparative data: Behavioral traits are more labile. *Evolution*, *57*(4), 717–745.
631 doi:10.1111/j.0014-3820.2003.tb00285.x
632 Cattin, M.-F. F., Bersier, L.-F., Banašek-Richter, C., Baltensperger, R., Gabriel, J.-P. P.,
633 Bersler, L. F., … Gabriel, J.-P. P. (2004). Phylogenetic constraints and adaptation
634 explain food-web structure. *Nature*, *427*(6977), 835–839. doi:10.1038/nature02327
635 Clavel, J., Escarguel, G., & Merceron, G. (2015). mvMORPH: An R package for fitting
636 multivariate evolutionary models to morphometric data. *Methods in Ecology and*
637 *Evolution*, *6*(11), 1311–1319. doi:10.1111/2041-210X.12420
638 Elias, M., Fontaine, C., & Frank Van Veen, F. J. (2013). Evolutionary history and
639 ecological processes shape a local multilevel antagonistic network. *Current*
640 *Biology*, *23*(14), 1355–1359. doi:10.1016/j.cub.2013.05.066
641 Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist*,
642 *125*(1), 1–15. doi:10.1086/284325
643 Fontaine, C., Guimarães, P. R., Kéfi, S., Loeuille, N., Memmott, J., van der Putten, W.
644 H., … Thébault, E. (2011). The ecological and evolutionary implications of
645 merging different types of networks. *Ecology Letters*, *14*(11), 1170–1181.

doi:10.1111/j.1461-0248.2011.01688.x

Fontaine, C., & Thébault, E. (2015). Comparing the conservatism of ecological interactions in plant–pollinator and plant–herbivore networks. *Population Ecology*, *57*(1), 29–36. doi:10.1007/s10144-014-0473-y

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19. doi:10.18637/jss.v022.i07

Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, *4*(4), 336–344. doi:10.1111/2041-210x.12018

Hadfield, J. D., Krasnov, B. R., Poulin, R., & Nakagawa, S. (2014). A tale of two phylogenies: Comparative analyses of ecological interactions. *The American Naturalist*, *183*(2), 174–187. doi:10.1086/674445

Harmon, L. J., Andreazzi, C. S., Débarre, F., Drury, J., Goldberg, E. E., Martins, A. B., … Matthews, B. (2019). Detecting the macroevolutionary signal of species interactions. *Journal of Evolutionary Biology*, *32*(8), 769–782. doi:10.1111/jeb.13477

Harmon, L. J., & Glor, R. E. (2010). Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution*, *64*(7), 2173–2178. doi:10.1111/j.1558-5646.2010.00973.x

Ives, A. R., & Godfray, H. C. J. (2006). Phylogenetic analysis of trophic associations. *The American Naturalist*, *168*(1), E1–E14. doi:10.1086/505157

Jacquemyn, H., Merckx, V. S. F. T., Brys, R., Tyteca, D., Cammue, B. P. A. A., Honnay, O., & Lievens, B. (2011). Analysis of network architecture reveals phylogenetic constraints on mycorrhizal specificity in the genus Orchis (Orchidaceae). *New Phytologist*, *192*(2), 518–528. doi:10.1111/j.1469-8137.2011.03796.x

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., … Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. doi:10.1093/bioinformatics/btq166

Krasnov, B. R., Fortuna, M. A., Mouillot, D., Khokhlova, I. S., Shenbrot, G. I., &

675 Poulin, R. (2012). Phylogenetic signal in module composition and species
676 connectivity in compartmentalized host-parasite networks. *American Naturalist*,
677 *179*(4), 501–511. doi:10.1086/664612
678 Li, D., Dinnage, R., Nell, L. A., Helmus, M. R., & Ives, A. R. (2020). phyr: An r
679 package for phylogenetic species-distribution modelling in ecological
680 communities. *Methods in Ecology and Evolution*, *11*(11), 1455–1463.
681 doi:10.1111/2041-210X.13471
682 Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the
683 relationship between phylogenetic relatedness and ecological similarity among
684 species. *Ecology Letters*, *11*(10), 995–1003. doi:10.1111/j.1461-0248.2008.01229.x
685 Maliet, O., Loeuille, N., & Morlon, H. (2020). An individual based model for the eco-
686 evolutionary emergence of bipartite interaction networks. *Ecology Letters*,
687 ele.13592. doi:10.1111/ele.13592
688 Mantel, N. (1967). The detection of disease clustering and a generalized regression
689 approach. *Cancer Research*, *27*(2), 209–220.
690 Martín González, A. M., Dalsgaard, B., Nogués-Bravo, D., Graham, C. H.,
691 Schleuning, M., Maruyama, P. K., … Martinez, N. D. (2015). The macroecology
692 of phylogenetically structured hummingbird-plant networks. *Global Ecology and*
693 *Biogeography*, *24*(11), 1212–1224. doi:10.1111/geb.12355
694 Martos, F., Munoz, F., Pailler, T., Kottke, I., Gonneau, C., & Selosse, M.-A. A. (2012).
695 The role of epiphytism in architecture and evolutionary constraint within
696 mycorrhizal networks of tropical orchids. *Molecular Ecology*, *21*(20), 5098–5109.
697 doi:10.1111/j.1365-294X.2012.05692.x
698 Morlon, H., Lewitus, E., Condamine, F. L., Manceau, M., Clavel, J., & Drury, J. (2016).
699 RPANDA: An R package for macroevolutionary analyses on phylogenetic trees.
700 *Methods in Ecology and Evolution*, *7*(5), 589–597. doi:10.1111/2041-210X.12526
701 Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., &
702 Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in*
703 *Ecology and Evolution*, *3*(4), 743–756. doi:10.1111/j.2041-210X.2012.00196.x

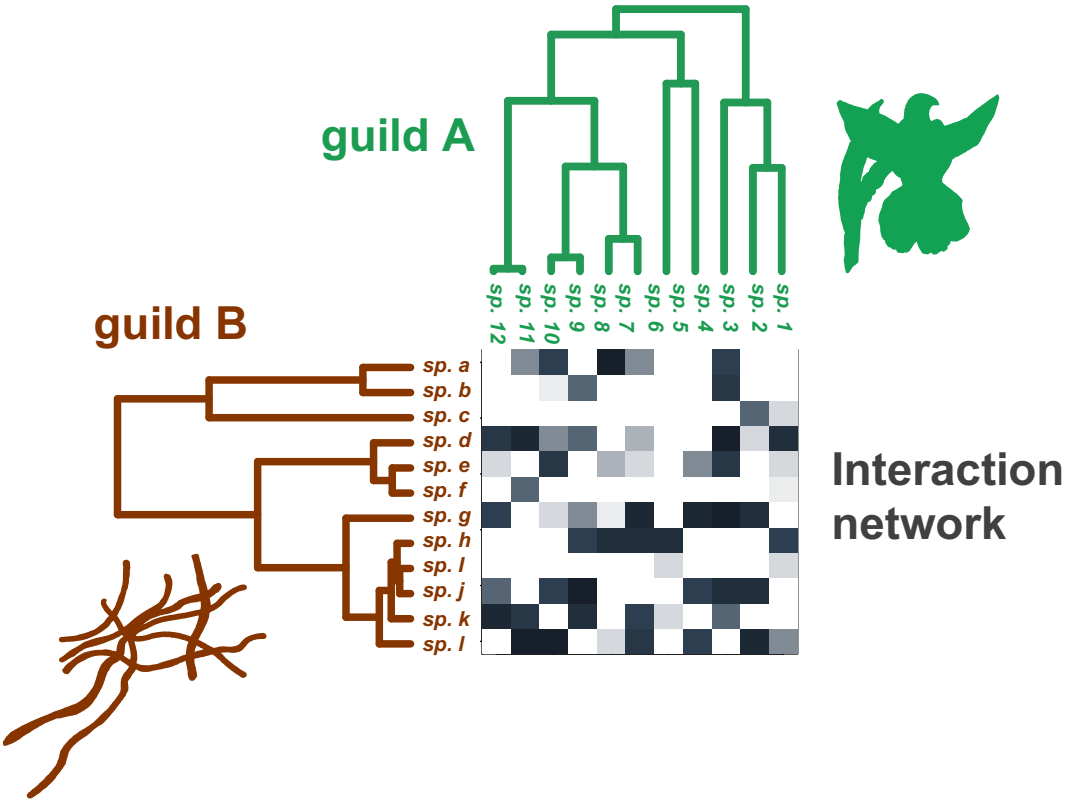
704 Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*,
705 *401*(6756), 877–884. doi:10.1038/44766
706 Paradis, E. (2013). Molecular dating of phylogenies by likelihood methods: A
707 comparison of models and a new information criterion. *Molecular Phylogenetics*
708 *and Evolution*, *67*(2), 436–444. doi:10.1016/j.ympev.2013.02.008
709 Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and
710 evolution in R language. *Bioinformatics*, *20*(2), 289–290.
711 doi:10.1093/bioinformatics/btg412
712 Peralta, G. (2016). Merging evolutionary history into species interaction networks.
713 *Functional Ecology*, *30*(12), 1917–1925. doi:10.1111/1365-2435.12669
714 Perez-Lamarque, B., & Morlon, H. (2019). Characterizing symbiont inheritance
715 during host–microbiota evolution: Application to the great apes gut microbiota.
716 *Molecular Ecology Resources*, *19*(6), 1659–1671. doi:10.1111/1755-0998.13063
717 R Core Team. (2020). R: A language and environment for statistical computing.
718 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
719 https://www.r-project.org/
720 Rafferty, N. E., & Ives, A. R. (2013). Phylogenetic trait-based analyses of ecological
721 networks. *Ecology*, *94*(10), 2321–2333. doi:10.1890/12-1948.1
722 Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology
723 (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223.
724 doi:10.1111/j.2041-210X.2011.00169.x
725 Rezende, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P., & Bascompte, J. (2007).
726 Non-random coextinctions in phylogenetically structured mutualistic networks.
727 *Nature*, *448*(7156), 925–928. doi:10.1038/nature05956
728 Rohr, R. P., & Bascompte, J. (2014). Components of phylogenetic signal in
729 antagonistic and mutualistic networks. *The American Naturalist*, *184*(5), 556–564.
730 doi:10.1086/678234
731 Sanders, J. G., Powell, S., Kronauer, D. J. C., Vasconcelos, H. L., Frederickson, M. E.,
732 & Pierce, N. E. (2014). Stability and phylogenetic correlation in gut microbiota:

733 lessons from ants and apes. *Molecular Ecology*, *23*(6), 1268–1283.
734 doi:10.1111/mec.12611
735 Song, S. J., Sanders, J. G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., … Knight,
736 R. (2020). Comparative analyses of vertebrate gut microbiomes reveal
737 convergence between birds and bats. *MBio*, *11*(1), 1–14. doi:10.1128/mBio.02901-
738 19
739 Vázquez, D. P., Chacoff, N. P., & Cagnolo, L. (2009). Evaluating multiple
740 determinants of the structure of plant–animal mutualistic networks. *Ecology*,
741 *90*(8), 2039–2046. doi:10.1890/08-1837.1
742 Xing, X., Jacquemyn, H., Gai, X., Gao, Y., Liu, Q., Zhao, Z., & Guo, S. (2019). The
743 impact of life form on the architecture of orchid mycorrhizal networks in
744 tropical forest. *Oikos*, *128*(9), 1254–1264. doi:10.1111/oik.06363
745 Xing, X., Liu, Q., Gao, Y., Shao, S., Guo, L., Jacquemyn, H., … Guo, S. (2020). The
746 architecture of the network of orchid–fungus interactions in nine co-occurring
747 Dendrobium species. *Frontiers in Ecology and Evolution*, *8*(May), 1–10.
748 doi:10.3389/fevo.2020.00130
749

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.



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

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

775

776 **a-d:** Phylogenetic signals in species interactions estimated using simple Mantel tests
777 with Pearson correlation (R) in the guilds A (a, c) and B (b, d). The different panels in
778 rows correspond to the 2 tested ecological distances: weighted Jaccard (a, b) or
779 weighted UniFrac (c, d) distances. One-tailed Mantel tests between phylogenetic
780 distances and ecological distances were performed using 10,000 permutations. In each
781 panel, the bars indicate the percentage of simulated networks that present a significant
782 positive correlation (in green; p-value>0.05 for the test of phylogenetic signal), a
783 significant negative correlation (in red; p-value>0.05 for the test of anti-phylogenetic
784 signal), or no significant correlation (in yellow; both p-values>0.05). Significant
785 phylogenetic signals (resp. anti-phylogenetic signals) are shaded from light green to
786 dark green according to the strength of the signal: we arbitrarily considered a "low
787 signal" when R<0.05 (resp. R>-0.05), an "intermediate signal" when 0.05<R<0.15 (resp.
788 -0.05>R>-0.15), and a "strong signal" when R>0.15 (resp. R<-0.15).

789

790 **e:** Phylogenetic signals estimated using PBLM. For a given combination of parameters,
791 the bar indicates the percentage of simulated networks that present no significant (in
792 yellow; $MSE \geq MSE_{star}$) or a significant (green; $MSE < MSE_{star}$) phylogenetic signal.
793 Phylogenetic signals are shaded from light green to dark green according to the
794 strength of the signal: we arbitrarily considered a "low signal" when $d_A < 0.05$ and

795 d_B<0.05, an "intermediate signal" when d_A>0.05 or d_B>0.05, and a "strong signal" when
796 d_A>0.15 or d_B>0.15. PBLM were run on the weighted networks.
797 In each panel, the first bar indicates the statistical power of the test, whereas the second
798 and third bar indicate the type-I error rate of the test. Note that the strength the
799 phylogenetic signals (based on the R and d values) are not directly comparable.
800
801 Results discriminating the simulated networks of different sizes and with different sets
802 of parameters are available in Figures S5 & S8.
803

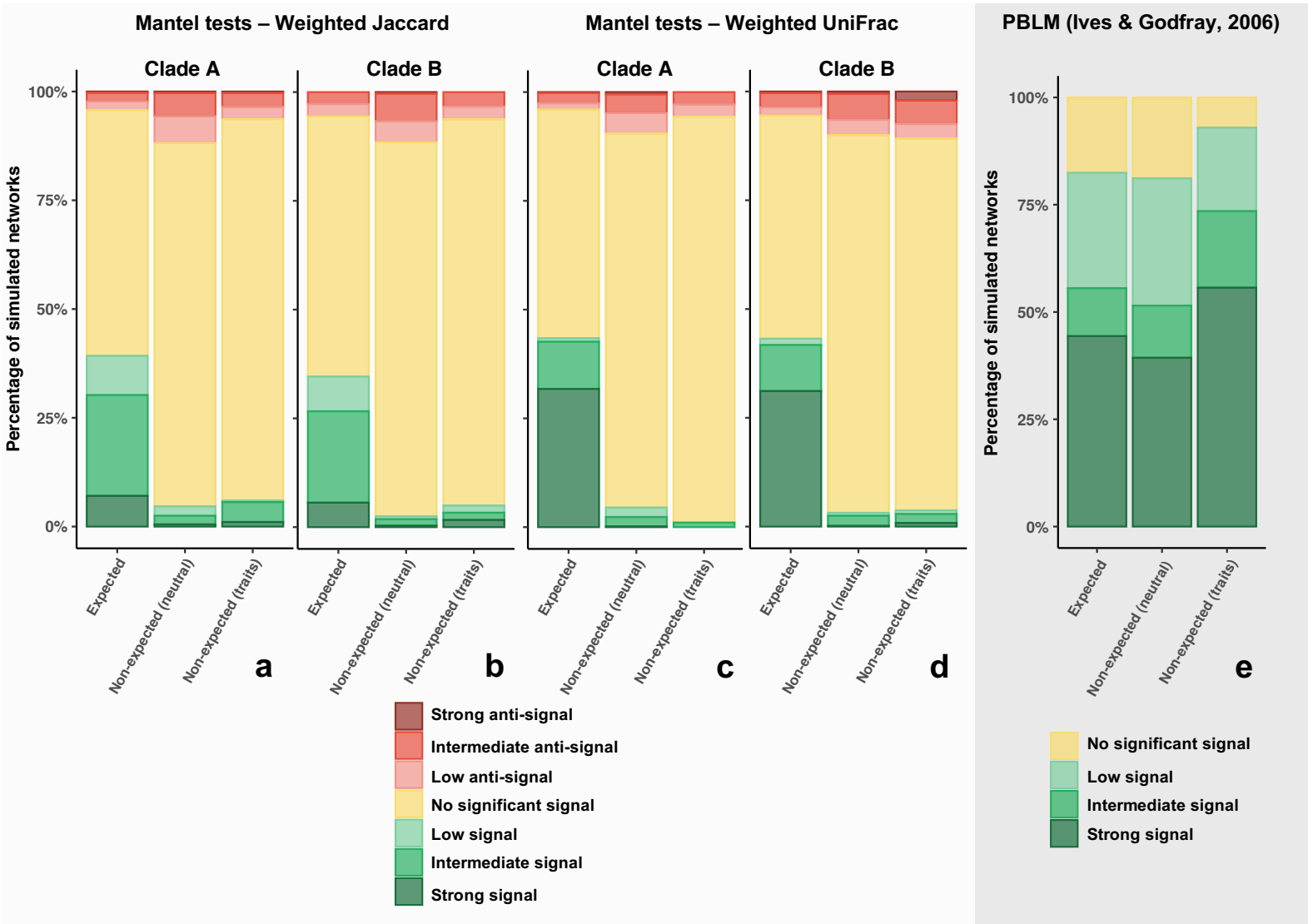


Figure 3: Recommended guidelines to measure phylogenetic signal in species interactions within bipartite ecological networks.

This guideline is composed of two fixed steps followed by two optional ones and can be applied as soon as a bipartite interaction network (with or without abundances) and at least the phylogenetic tree of guild A are available. The phylogenetic tree does not need to be binary, rooted, or ultrametric. For each step, an example of the corresponding function available in the R-package RPANDA is indicated in grey.

Step 1: The first step consists in testing for phylogenetic signal in species interactions for guild A (*i.e.* whether closely related species from guild A tend to interact with similar partners from guild B) using a one-tailed simple Mantel test. This step requires to pick an ecological distance (UniFrac distances are recommended compared to Jaccard distances) and a type of correlation (Pearson correlation by default).

Step 2: Next, to assess whether a phylogenetic signal in species interactions really comes from the identity of species interactions, the second step consists in testing whether there is phylogenetic signal in the number of partners of guild A (*i.e.* whether closely related species from guild A tend to interact with the same number of partners from guild B) using a one-tailed simple Mantel test.

Option 1: Clade-specific phylogenetic signal in guild A can be tested using simple Mantel tests while correcting for multiple testing (*e.g.* Bonferroni correction). It can be used to test whether some clades present different intensities of phylogenetic signal (*e.g.* because of higher specificity).

Option 2: The robustness of the findings can be tested by looking at how the conclusions might be affected by phylogenetic uncertainty (e.g. using a Bayesian posterior of tree) or sampling bias. The potential effect of sampling bias can be 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.

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 **number of partners** (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)

831

832 **Figure 4: Empirical application on an orchid-fungus interaction network from La**
833 **Réunion island (Martos *et al.,* 2012): the clade-specific analyses of phylogenetic**
834 **signal in species interactions revealed a significant phylogenetic signal in the**
835 **epiphytic subtribe Angraecinae.**
836 The orchid phylogeny (Martos *et al.,* 2012) is represented with its nodes colored
837 according to the results of the Mantel test performed on the corresponding sub-
838 network: in blue if non-significant, in grey when the node has less than 10 descendent
839 species (the Mantel test was not performed), and in red when the phylogenetic signal
840 is significant. Each one-tailed simple Mantel test was performed using the Pearson
841 correlation and 100,000 permutations and its significance was evaluated while
842 correcting for multiple testing (Bonferroni correction).
843 For each species, its habitat (terrestrial or epiphytic) is indicated at the tips of the tree
844 and the main orchid clades are highlighted in colors. Only the genera are indicated at
845 the tips of the tree (see Supplementary Figure S28 for the species list).
846

847