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

Running title: Measuring phylogenetic signal in interactions

Benoît Perez-Lamarque [1,2], Odile Maliet [1], Benoît Pichon [3,1], Marc-André Selosse [2,4], 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] *Institut d'écologie et des sciences de l'environnement (iEES), Sorbonne Université, CNRS, UPEC, CNRS, IRD, INRA, 75 005 Paris, France*

[4] *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)

Supplementary data: https://github.com/BPerezLamarque/Phylosignal_network/blob/master/Supplementary_figures_phylo_signal_network.pdf

Abstract

Whether interactions between species are conserved on evolutionary time-scales 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 partners? 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 PBLM is a model-based approach that relies on strong assumptions on how interactions evolve. We find that PBLM often detects phylogenetic signal when it should not. Simple Mantel tests instead have low type-I error rates and moderate statistical power; however, they often artifactually detect that closely related species 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. We provide general guidelines and an 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, herbivory, and parasitism (Bascompte et al. 2003; Fontaine et al. 2011; Martos et al. 2012; Bascompte and 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, 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 and 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.

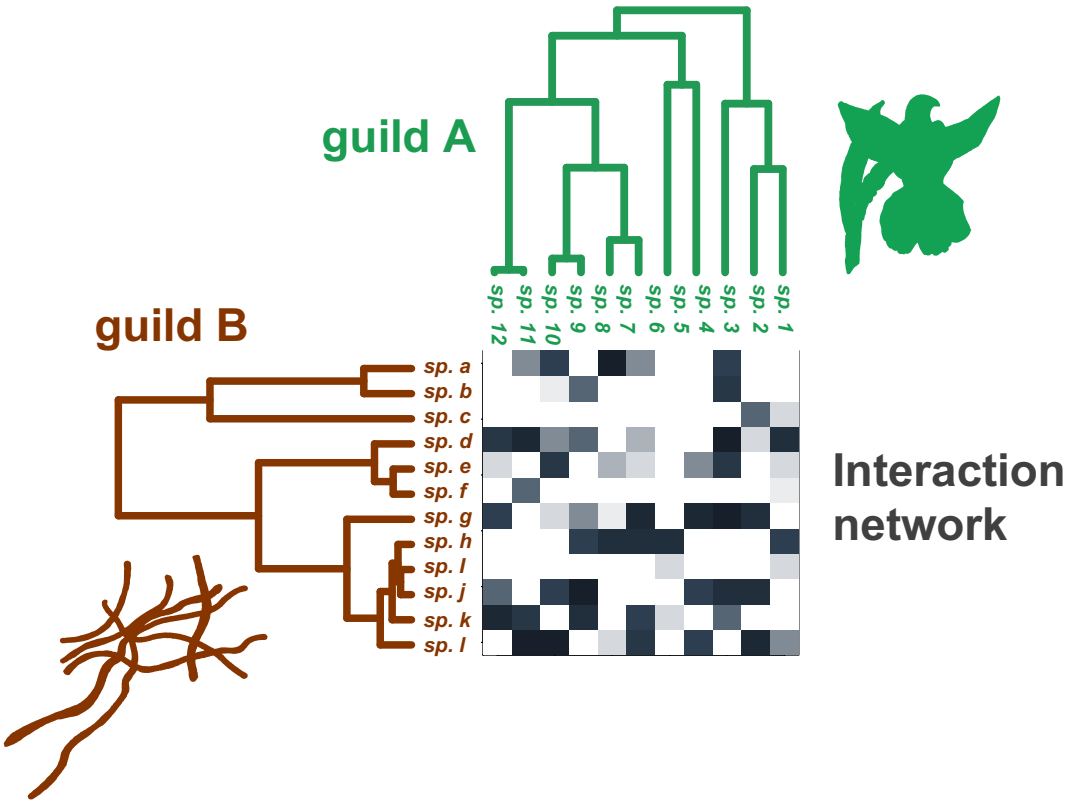
Testing for phylogenetic signal in species interactions falls in the category of cases where the 'trait' data are pairwise distances, here the between-species dissimilarity in sets of interacting species. Simple Mantel tests have therefore been

widely used in this context (*e.g.* 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 and Thébault 2015). Besides these correlative approaches, several model-based approaches have been developed (Ives and Godfray 2006; Rafferty and Ives 2013; Hadfield et al. 2014; Li et al. 2020). The first of this model, 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 networks, *e.g.* in host-parasite, plant-fungus, and pollination networks (Ives and 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). Other models have been developed more recently (Rafferty and Ives 2013; Hadfield et al. 2014; Li et al. 2020), including the phylogenetic generalized linear mixed model (PGLMM; Rafferty and Ives 2013) that uses linear mixed models to infer phylogenetic signals in both the number of partners and species interactions. Yet, the higher computational requirements of these methods have prevented their widespread use on empirical networks. PBLM thus remains the method frequently used in empirical studies (*e.g.* Xing et al. 2020; Corro et al. 2021).

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 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, as phylogenetic signal may be localized. 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).

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.



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 in the simulated networks without affecting their overall structure.

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 5.10^7 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 evaluated whether these simulations generated realistic networks by comparing their structure with that of empirical networks (Supplementary Methods 2).

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. We evaluated phylogenetic signal in species traits using two approaches. First, for simplicity and consistency with the rest of the paper, we used Mantel tests (Pearson correlation) between phylogenetic distances and trait distances computed as the Euclidian distances between trait values for each species pair. Second, given that model-based approaches usually perform better (Harmon and Glor 2010), we used a multivariate extension of Pagel's λ (Pagel 1999) implemented in R (Goolsby 2015); we assessed the significance of the phylogenetic signal in species traits with likelihood ratio tests comparing the inferred Pagel's λ model to a null model where λ=0 (*i.e.* no phylogenetic signal).

Computing phylogenetic signal in species interactions

We computed phylogenetic signal in species interactions in the simulated networks using Mantel tests and PBLM, as well as the computationally-intensive PGLMM for the smallest networks. Complete descriptions of these methods are available in Supplementary Methods 3. Mantel tests, PBLM, and PGLMM rely on different strategies to evaluate the significance of the phylogenetic signal, and it could be argued that results of these tests are not directly comparable. Our approach is to follow the methodologies traditionally used in empirical studies and compare their conclusions (detection or not of a phylogenetic signal).

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 3 (Lozupone et al. 2011)). 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 and Urban 2007); the significance of each correlation was evaluated using 10,000 permutations, except for the computationally intensive Kendall correlation (100 permutations only). For each network, we considered that there was a significant phylogenetic signal (resp. anti-phylogenetic signal) if the correlation coefficient (R) was higher (resp. lower) than >95% of the randomized correlations; we computed the p-value of each one-tailed Mantel test as the fraction of the randomized correlations above (resp. below) the original value.

PBLM: To estimate phylogenetic signal based on PBLM, we modified the function *pblm* from the R-package picante (Kembel et al. 2010) to more efficiently perform matrix inversions and handle large interaction networks. In short, the parameters d_A and d_B of the Ornstein-Uhlenbeck processes of PBLM were estimated using generalized least squares (Ives & Godfray 2006). d_A and d_B are interpreted as a measure of phylogenetic signal in species interactions: if d=0, there is no effect of the phylogeny (similar as evolution on a star phylogeny, *i.e.* no phylogenetic signal); 0<d<1 generates stabilizing selection (*i.e.* phylogenetic signal) and d>1 disruptive selection (*i.e.* anti-phylogenetic signal). We followed Ives & Godfray (2006; Supplementary Methods 3) by considering that the phylogenetic signal is significant when the mean square error (MSE) of the model is smaller than that obtained using star phylogenies (MSE_{star}); we also used a more stringent criterion by considering that the signal is significant when the MSE is at least 5% lower than MSE_{star}. Finally, we applied the bootstrapping method of Ives & Godfray (2006; Supplementary Methods 3) to the smallest networks. A single PBLM inference can take several days to run (time measured on an Intel 2.8 GHz MacOSX laptop) on networks of intermediate sizes (between 50 and 100 species per guild),

which prevented us from applying the bootstrap approach to large networks; we therefore tested this approach on networks simulated with 500 individuals (*i.e.* a total of 400 networks).

PGLMM: We performed analyses of the statistical performances of PGLMM (Rafferty and Ives 2013) using the function *pglmm* in the R-package phyr (Li et al. 2020). Following the procedure used in Lajoie and Kembel (2021), we fitted for each network different models accounting or not for phylogenetic signals in both the number of partners and in the species interactions in both clades, using restricted maximum likelihood and evaluating significance with likelihood ratio tests. Because fitting these models can require large amount of memory (e.g. >80 Go for some networks with >50 species per guild), we tested this approach on networks simulated with 500 individuals. We fitted the PGLMM using either a Gaussian or a Poisson distribution of abundances for weighted networks , and a binomial distribution (presence/absence data) for unweighted networks (Li et al. 2020).

Confounding effect of the phylogenetic signal in the number of partners

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 3), 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 between phylogenetic and ecological distances, while controlling for the absolute differences in degrees.

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 uncertainty in the phylogenetic reconstructions (*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 signal 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 et al. 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 at least 10 descendants, we estimated the clade-specific phylogenetic signal using a Mantel test investigating whether closely related species from this sub-clade of A tend to interact with similar partners (and *vice-versa* for guild B). Using UniFrac distances, we performed the Mantel tests with 100,000 permutations, and introduced a Bonferroni correction for multiple testing to keep a global alpha-risk of 5%. To test this approach, we generated synthetic networks with known sub-clade signal by 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

359 randomly selecting a branch, such that it creates a separate module with strong
360 phylogenetic signal. Such simulations could correspond to the evolution of a different
361 niche, *e.g.* terrestrial *versus* epiphytic plants associating with different mycorrhizal
362 fungi (Martos et al. 2012). We then performed our clade-specific analysis of
363 phylogenetic signal and investigated in which nodes we recovered a significant
364 phylogenetic signal.

365

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

368

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

Results

Expected phylogenetic signal in species interactions in *BipartiteEvol* networks

The networks simulated using *BipartiteEvol* gave large ranges of sizes for guilds A and B (from less than 50 to more than 250 species; Fig. S2) and had structural properties comparable to those of empirical networks (Fig. S3), meaning that these simulated networks are realistic.

Using Mantel tests, we found a significant phylogenetic signal in species traits for most antagonistic and neutral simulations (Fig. S4A). 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. S4A). 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, which represent ~40% of the mutualistic simulations. In addition, we found an anti-phylogenetic signal in species traits in less than 1% of the simulations (Fig. S4A). Given that we do not expect *BipartiteEvol* to generate anti-phylogenetic signal in species traits and given that the alpha-risk of Mantel tests is 5%, these 1% of networks with an anti-phylogenetic signal in species traits are likely false-positives. We removed them when evaluating the performance of the different approaches and we therefore do not expect anti-phylogenetic signal in species interactions for the networks we tested. Results were similar with Pagel's λ, with a significant phylogenetic signal in species traits for almost all antagonistic and neutral simulations, and in ~65% of the mutualistic simulations (Fig. S4B). Mantel tests and Pagel's λ lead to identical conclusions for >95% of the simulated networks.

Computing phylogenetic signal in species interactions in *BipartiteEvol* networks

Using Mantel tests, as expected, we did not find a significant phylogenetic signal in species interactions for most neutral networks or for networks with no signal in species traits (Figs. 2 & S5): 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 when using weighted Jaccard distances and Pearson correlations (~8% type-I error). Conversely, we detected a significant unexpected anti-phylogenetic signal in more than 10% of the simulated networks, in particular in the small ones (Figs. 2 & S5).

Many mutualistic or antagonistic networks where we expected a phylogenetic signal in species interactions (*i.e.* non-neutral networks with signal in species traits) presented no significant signal with Mantel tests (Figs. 2 & S5), in particular those simulated with low α_A and α_B values (*e.g.* antagonism **vii**), where non-neutral effects were weak. Mantel tests measuring phylogenetic signal in species interactions were most often not significant unless the phylogenetic signal in species traits was strong ($R>0.6$; Fig. S6). Even when the phylogenetic signal in species traits was very strong ($R>0.9$), the phylogenetic signal in species interactions was not significant in many networks. In mutualistic networks, phylogenetic signals in species interactions were present only when there was a large asymmetry in the effects of trait matching on the fitnesses of the species from guilds A or B (case **v**: $\alpha_A=1$; $\alpha_B=0.01$), *i.e.* when only one guild was specialized. Conversely, in antagonistic networks, phylogenetic signals were found mainly when trait matching had a strong impact on the fitness of guild B (the obligate parasites/predators; $\alpha_B \geq 0.1$). Additionally, when phylogenetic signal was significant in one guild, it was generally also significant in the other; in antagonistic networks, the signal was usually higher in guild A compared to guild B (Fig. S5).

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 according to a Mantel test – see Supplementary Figure 4A) 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). Results are similar when the expectations are based on Pagel's λ to measure the phylogenetic signals in species traits (Supplementary Figure 10).

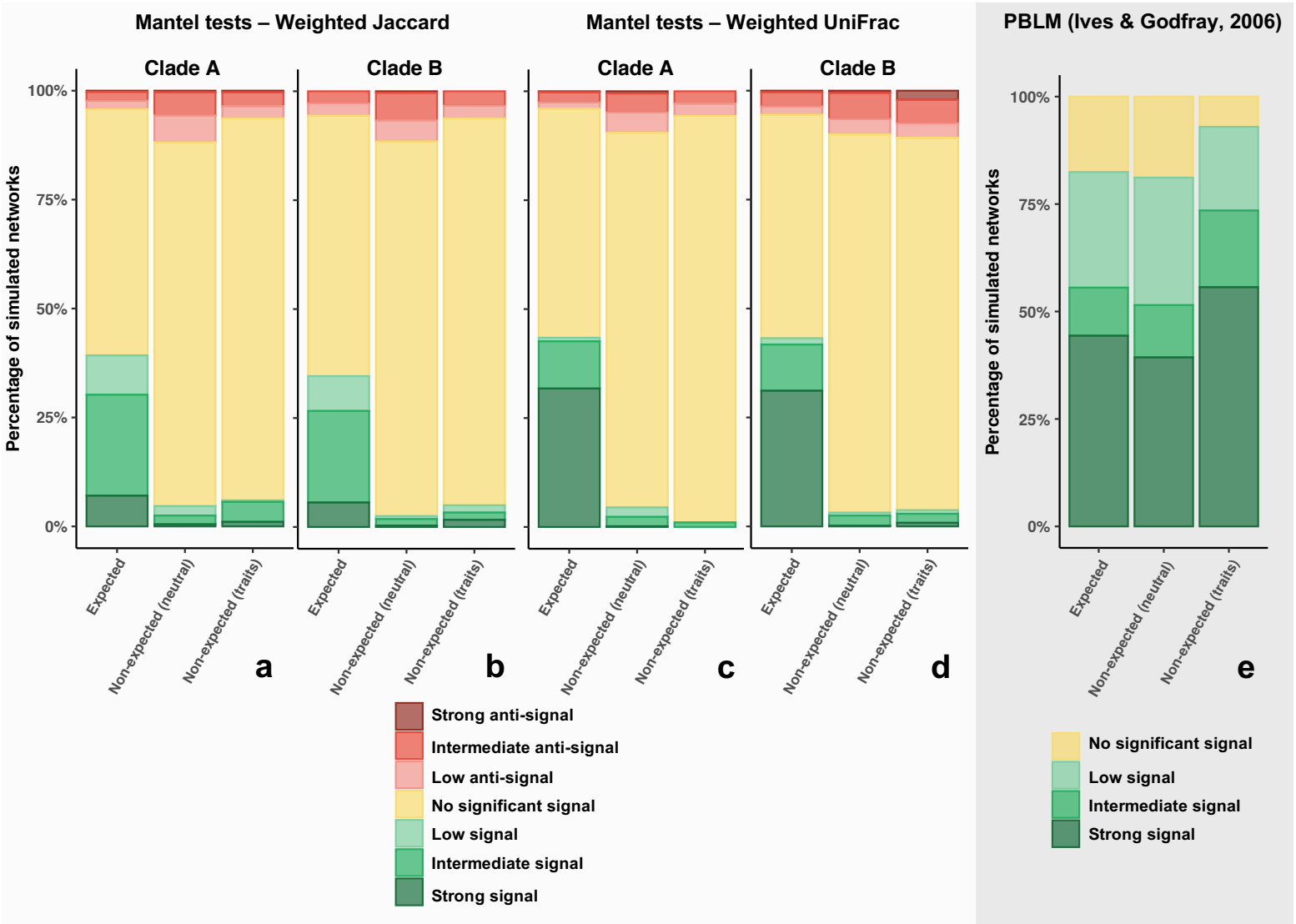
a-d: Phylogenetic signals in species interactions estimated using simple Mantel tests with Pearson correlation (R) in the guilds A (a, c) and B (b, d). The different panels in rows correspond to the 2 tested ecological distances: weighted Jaccard (a, b) or weighted UniFrac (c, d) distances. One-tailed 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 (in green; 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). Significant phylogenetic signals (resp. anti-phylogenetic signals) are shaded from light green to dark green according to the strength of the signal: we arbitrarily considered a "low signal" when R<0.05 (resp. R>-0.05), an "intermediate signal" when 0.05<R<0.15 (resp. -0.05>R>-0.15), and a "strong signal" when R>0.15 (resp. R<-0.15).

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: we arbitrarily considered a "low signal" when d_A<0.05 and d_B<0.05, an "intermediate signal" when d_A>0.05 or d_B>0.05, and a "strong signal" when 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. Note that the strength the phylogenetic signals (based on the R and d values) are not directly comparable.

Results discriminating the simulated networks of different sizes and with different sets of parameters are available in Figures S5 & S7.



476 The statistical power of Mantel tests measuring phylogenetic signal in species
477 interactions seems to be modulated by network size, as phylogenetic signals were less
478 often significant but generally stronger in smaller networks (Fig. S5). Moreover,
479 Mantel tests based on Pearson correlations had higher power than Spearman and
480 Kendall correlations (Fig. S5) and weighted UniFrac distances outperformed other
481 ecological distances in terms of power (Fig. S5; Table S2).
482
483 When using mean square errors to evaluate the significance of PBLM, we found
484 a significant phylogenetic signal in species interactions in most of the simulated
485 networks including when we did not expect any (Fig. 2e). The strength and the
486 significance of the inferred phylogenetic signals were independent of the strength of
487 the phylogenetic signal in species traits (Fig. S6). The propensity of PBLM to detect
488 phylogenetic signal decreased in large unweighted networks, but the type-I errors
489 remained >30%, including when using a more stringent significance cutoff (Figs. S7).
490 Similar results were obtained when bootstrapping to evaluate the significance (Fig.
491 S8). PGLMM on weighted networks with a Gaussian or Poisson distribution had
492 slightly lower but still high type-I error rates (>25% or 20%, respectively) and
493 intermediate statistical power (<50%) when measuring phylogenetic signals in species
494 interactions (Fig. S9). PGLMM also often artifactually detected phylogenetic signals in
495 the number of partners (Fig. S9). Conversely, PGLMM on unweighted networks never
496 detected any significant signal (Fig. S9).
497
498 We inferred similar statistical performances of both Mantel tests and PBLM
499 when we used Pagel's λ to evaluate phylogenetic signal in species traits (Figs. S6 and
500 S10).
501

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 & S13; 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. S15). We found significant correlations between degree differences and ecological distances in most of these simulated networks (Fig. S16). 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. S17; 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. S18). In addition, partial Mantel tests detected a spurious significant anti-phylogenetic signal in species interactions in >15% of the networks (Fig. S18). 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

559 (available in Martos et al. (2012)). First (step 1), simple Mantel tests of phylogenetic
560 signal in species interactions for fungi and orchids revealed a significant but low
561 phylogenetic signal (R<0.10) on the orchid side using Jaccard distances; however, the
562 significance disappeared with UniFrac distances (Table S4). Similarly, marginally not-
563 significant and low phylogenetic signals were detected in the mycorrhizal fungi side
564 (R<0.04; Table S4). Next (step 2), simple Mantel tests of phylogenetic signal in the
565 number of partners were not significant (p-values>0.05). Our investigation of clade-
566 specific phylogenetic signals in species interactions in orchids (option 1) revealed a
567 significant phylogenetic signal in Angraecinae, a sub-tribe composed of 34 epiphytic
568 species (sequential Mantel test: R=0.37; Bonferroni-corrected p-value=0.016; Fig. 4)
569 interacting with 53 fungi, suggesting that closely related Angraecinae tend to interact
570 with more similar mycorrhizal fungi. When we checked the robustness of the
571 significant phylogenetic signal detected in Angraecinae (option 2) by subsampling the
572 Angraecinae clade down to 10 species, we still recovered significant signal in species
573 interactions in both cases (Fig. S26).
574

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)

602

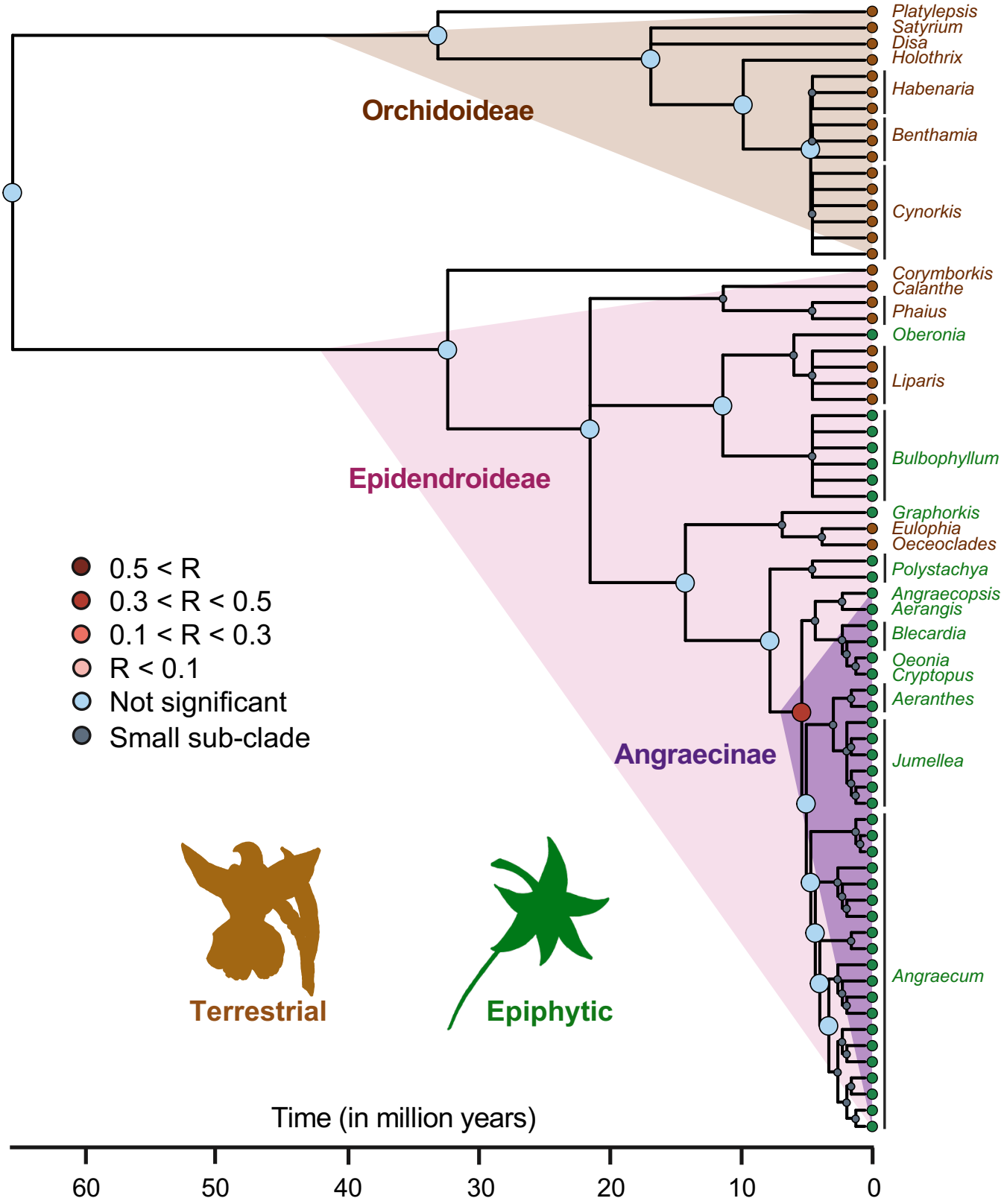
603

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 S26 for the species list).

619



620 ## Discussion:

621

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

627

628 The Phylogenetic bipartite linear model (PBLM) is widely used to test for
629 phylogenetic signal in species interactions, however we found that it has a very high
630 type-I error rate (>30%). PBLM assumes that the interaction strength between two
631 species is determined by the product of two unobserved traits evolving on the
632 phylogenies of guilds A and B respectively, according to two independent Ornstein-
633 Uhlenbeck processes with the selection strengths d_A and d_B (Supplementary Methods
634 3). PBLM tests the significance of d_A and d_B, which measure the phylogenetic signal of
635 the unobserved traits. A species with a high trait value will have high interaction
636 strengths with many partner species (*i.e.* it is a generalist species), while a species with
637 a low trait value will have low interaction strengths with most partner species, except
638 with the few species with high trait values (*i.e* it is a specialist species). Therefore, we
639 suspect d_A and d_B to measure phylogenetic signals in the number of partners rather
640 than in species interactions. However, we also found significant d_A and d_B in the
641 absence of phylogenetic signal in the number of partners, suggesting that PBLM is
642 sensitive to model misspecification (it relies on strong hypotheses on how the number
643 of partners evolves). In any case, our results suggest that PBLM should not be used as
644 a routine for measuring phylogenetic signal in species interactions.

645

646 Other model-based approaches that extend PBLM (Rafferty and Ives 2013;
647 Hadfield et al. 2014; Li et al. 2020) allow to infer parameters thought to reflect the
648 phylogenetic structure of interactions networks, while controlling for phylogenetic

649 signal in the number of parterns as well as heterogeneity in sampling effort (Hadfield
650 *et al.*, 2014). Our analyses using the PGLMM approach (Rafferty and Ives 2013) on the
651 smallest simulated networks suggested that it also has high type-I error rates and
652 intermediate statistical power when using weighted interactions. It would have been
653 ideal to also test this approach on larger networks, but this was prohibited by their
654 computational cost. Indeed, fitting PGLMM can require >80 Go of memory for some
655 networks and our application of the Bayesian approach of Hadfield *et al.* (2014) ran
656 several days (on an Intel 2.8 GHz, MacOSX laptop) without reaching convergence.
657 Because of these high computational demands, these methods are typically not used
658 to measure phylogenetic signal in species interactions in empirical studies, which is
659 either done using Mantel tests or PBLM (see Fontaine and Thébault 2015; Xing et al.
660 2020; Corro et al. 2021 for recent examples). Future model developments of such
661 approaches would thus benefit from faster inferences and our results highlight the
662 need to thoroughly test these approaches with simulations before they are applied to
663 empirical systems and biological conclusions are drawn.
664
665 We found that simple Mantel tests have a moderate statistical power and a
666 reasonable type-I error rate (<5%) when testing for phylogenetic signal in species
667 interactions. Not surprisingly, these tests have a higher power for larger simulated
668 networks. The fact that Mantel tests have a moderate power for measuring
669 phylogenetic signals in species interactions corroborates the findings about Mantel
670 tests in other contexts (Harmon and Glor 2010; Guillot and Rousset 2013). Hence,
671 although simple Mantel tests might fail at detecting low phylogenetic signal, we can
672 trust their results when they are significant. On the contrary, we found a high
673 proportion of simulated networks (5-10%) presenting a significant anti-phylogenetic
674 signal in species interactions, although we did not expect any in our simulations
675 (because we did not observe any anti-phylogenetic signal in species traits). False-
676 positives are therefore frequent when testing for anti-phylogenetic signal using simple

677 Mantel tests and detection of such signal in empirical networks should be interpreted
678 with caution.
679 In addition, Pearson correlations performed better than Spearman and Kendall
680 correlations, which is somewhat surprising, as correlations between phylogenetic and
681 ecological distances are not particularly expected to be linear: Spearman and Kendall
682 correlations have less stringent hypotheses, as they only assume monotonicity
683 (Supplementary Methods 3), but they probably lose information. We also reported that
684 using ecological distances that consider interaction abundances and phylogenetic
685 relatedness of the partners, such as weighted UniFrac distances, significantly improves
686 the detection of phylogenetic signal, even when reconstructed partners trees are not
687 robust. Given that species delineation may be somewhat arbitrary, especially for
688 microbial interactors, and that Jaccard distances are directly sensitive to species
689 delineation (Sanders et al. 2014), we advocate the use of weighted UniFrac distances.
690 An exception might be if communities of interactors differ mainly in terms of recently
691 diverged species; in this case Jaccard distances may perform better, as UniFrac
692 distances emphasize differences in long branches rather than recent splits (Sanders et
693 al. 2014). Finally, we found that multiple simple Mantel tests combined with a
694 Bonferroni correction perform rather well to investigate clade-specific phylogenetic
695 signals. Such an approach can therefore be valuable for measuring local phylogenetic
696 signal in large "meta-networks", such as those describing host-microbiota
697 phylosymbiosis (Song et al. 2020), which likely have heterogeneous phylogenetic
698 signals across the network.
699
700 While simple Mantel tests have satisfactory statistical performances, these tests do
701 not control for the potential confounding effect of phylogenetic signal in the number
702 of partners. Partial Mantel tests are frequently used for investigating phylogenetic
703 signal in species interactions while controlling for signal in the number of partners;
704 however, we found that they often detected significant signals in species interactions
705 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 and Glor 2010; Guillot and Rousset 2013). An alternative possibility is 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

735 epiphytic clade that is experiencing a radiation in La Réunion island. This signal is
736 likely produced by the different orchids genera in Angraecinae associating with
737 specific fungal clades (Martos et al. 2012). Thus, our results corroborate a trend toward
738 mycorrhizal specialization in epiphytic orchids compared with terrestrial species
739 (Xing et al. 2019), as the epiphytic habitats might require particular adaptations and
740 stronger dependences toward specific mycorrhizal fungi.
741
742 Interaction networks are increasingly being analyzed to unravel the
743 evolutionary processes shaping their structure and to predict their stability. Currently-
744 used tools for measuring phylogenetic signals are clearly misleading. The alternative
745 approach we propose based on sequential Mantel tests avoids false positive, but its
746 statistical power is limited. By emphasizing the limits of current tests of phylogenetic
747 signal, we hope to stimulate new developments in the statistical adjustment to
748 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:

All authors designed the study. BPL performed the analyses, BPi performed the analyses on the network structures, and FM gathered the empirical data. BPL and HM wrote the first draft of the manuscript and all authors contributed to revisions.

Data accessibility:

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.

The scripts for simulating the networks and for measuring the phylogenetic signals in species interactions are available at:

https://github.com/BPerezLamarque/Phylosignal_network/tree/master/simulations

776 Supplementary data (including Supplementary Methods, Tables, and Figures) are
777 available at:
778
779 https://github.com/BPerezLamarque/Phylosignal_network/blob/master/Supplementa
780 ry_figures_phylo_signal_network.pdf
781

782 ## Conflict of Interest statement:

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

References:

Aizen, M. A., G. Gleiser, M. Sabatino, L. J. Gilarranz, J. Bascompte, and M. Verdú. 2016. The phylogenetic structure of plant-pollinator networks increases with habitat size and isolation. Ecology Letters 19:29–36.

Bascompte, J., and P. Jordano. 2013. Mutualistic networks. Monographs in population biology (Princeton., Vol. 7). Princeton University Press, Princeton.

Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. The nested assembly of plant-animal mutualistic networks. Proceedings of the National Academy of Sciences 100:9383–9387.

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

Cattin, M.-F. F., L.-F. Bersier, C. Banašek-Richter, R. Baltensperger, J.-P. P. Gabriel, L. F. Bersler, C. Banašek-Richter, et al. 2004. Phylogenetic constraints and adaptation explain food-web structure. Nature 427:835–839.

Clavel, J., G. Escarguel, and G. Merceron. 2015. mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. (T. Poisot, ed.)Methods in Ecology and Evolution 6:1311–1319.

Corro, E. J., F. Villalobos, A. Lira-Noriega, R. Guevara, P. R. Guimarães, and W. Dáttilo. 2021. Annual precipitation predicts the phylogenetic signal in bat–fruit interaction networks across the Neotropics. Biology Letters 17.

Elias, M., C. Fontaine, and F. J. Frank Van Veen. 2013. Evolutionary history and ecological processes shape a local multilevel antagonistic network. Current Biology 23:1355–1359.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1–15.

Fontaine, C., P. R. Guimarães, S. Kéfi, N. Loeuille, J. Memmott, W. H. van der Putten, F. J. F. F. van Veen, et al. 2011. The ecological and evolutionary implications of merging different types of networks. Ecology Letters 14:1170–1181.

Fontaine, C., and E. Thébault. 2015. Comparing the conservatism of ecological interactions in plant–pollinator and plant–herbivore networks. Population Ecology 57:29–36.

Goolsby, E. W. 2015. Phylogenetic comparative methods for evaluating the evolutionary history of function-valued traits. Systematic Biology 64:568–578.

Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22:1–19.

Guillot, G., and F. Rousset. 2013. Dismantling the Mantel tests. Methods in Ecology and Evolution 4:336–344.

Hadfield, J. D., B. R. Krasnov, R. Poulin, and S. Nakagawa. 2014. A tale of two phylogenies: Comparative analyses of ecological interactions. The American Naturalist 183:174–187.

Harmon, L. J., C. S. Andreazzi, F. Débarre, J. Drury, E. E. Goldberg, A. B. Martins, C. J. Melián, et al. 2019. Detecting the macroevolutionary signal of species interactions. Journal of Evolutionary Biology 32:769–782.

Harmon, L. J., and R. E. Glor. 2010. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. Evolution 64:2173–2178.

Ives, A. R., and H. C. J. Godfray. 2006. Phylogenetic analysis of trophic associations. The American Naturalist 168:E1–E14.

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

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

Krasnov, B. R., M. A. Fortuna, D. Mouillot, I. S. Khokhlova, G. I. Shenbrot, and R. Poulin. 2012. Phylogenetic signal in module composition and species connectivity in compartmentalized host-parasite networks. American Naturalist 179:501–511.

842 Lajoie, G., and S. W. Kembel. 2021. Plant-bacteria associations are phylogenetically
843 structured in the phyllosphere. Molecular Ecology 30:5572–5587.
844 Li, D., R. Dinnage, L. A. Nell, M. R. Helmus, and A. R. Ives. 2020. phyr: An r package
845 for phylogenetic species-distribution modelling in ecological communities. (S. Price,
846 ed.)Methods in Ecology and Evolution 11:1455–1463.
847 Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the
848 relationship between phylogenetic relatedness and ecological similarity among
849 species. Ecology Letters 11:995–1003.
850 Lozupone, C., M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight. 2011. UniFrac:
851 an effective distance metric for microbial community comparison. The ISME Journal
852 5:169–172.
853 Maliet, O., N. Loeuille, and H. Morlon. 2020. An individual based model for the eco-
854 evolutionary emergence of bipartite interaction networks. (T. Poisot, ed.)Ecology
855 Letters ele.13592.
856 Mantel, N. 1967. The detection of disease clustering and a generalized regression
857 approach. Cancer Research 27:209–220.
858 Martín González, A. M., B. Dalsgaard, D. Nogués-Bravo, C. H. Graham, M.
859 Schleuning, P. K. Maruyama, S. Abrahamczyk, et al. 2015. The macroecology of
860 phylogenetically structured hummingbird-plant networks. Global Ecology and
861 Biogeography 24:1212–1224.
862 Martos, F., F. Munoz, T. Pailler, I. Kottke, C. Gonneau, and M.-A. A. Selosse. 2012.
863 The role of epiphytism in architecture and evolutionary constraint within
864 mycorrhizal networks of tropical orchids. Molecular Ecology 21:5098–5109.
865 Morlon, H., E. Lewitus, F. L. Condamine, M. Manceau, J. Clavel, and J. Drury. 2016.
866 RPANDA: An R package for macroevolutionary analyses on phylogenetic trees.
867 Methods in Ecology and Evolution 7:589–597.
868 Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W.
869 Thuiller. 2012. How to measure and test phylogenetic signal. Methods in Ecology
870 and Evolution 3:743–756.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Paradis, E. 2013. Molecular dating of phylogenies by likelihood methods: A comparison of models and a new information criterion. Molecular Phylogenetics and Evolution 67:436–444.

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

Peralta, G. 2016. Merging evolutionary history into species interaction networks. Functional Ecology 30:1917–1925.

Perez-Lamarque, B., and H. Morlon. 2019. Characterizing symbiont inheritance during host–microbiota evolution: Application to the great apes gut microbiota. Molecular Ecology Resources 19:1659–1671.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafferty, N. E., and A. R. Ives. 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 in Ecology and Evolution 3:217–223.

Rezende, E. L., J. E. Lavabre, P. R. Guimarães, P. Jordano, and J. Bascompte. 2007. Non-random coextinctions in phylogenetically structured mutualistic networks. Nature 448:925–928.

Rohr, R. P., and J. Bascompte. 2014. Components of phylogenetic signal in antagonistic and mutualistic networks. The American Naturalist 184:556–564.

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

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

900 Vázquez, D. P., N. P. Chacoff, and L. Cagnolo. 2009. Evaluating multiple
901 determinants of the structure of plant–animal mutualistic networks. Ecology
902 90:2039–2046.
903 Xing, X., H. Jacquemyn, X. Gai, Y. Gao, Q. Liu, Z. Zhao, and S. Guo. 2019. The impact
904 of life form on the architecture of orchid mycorrhizal networks in tropical forest.
905 Oikos 128:1254–1264.
906 Xing, X., Q. Liu, Y. Gao, S. Shao, L. Guo, H. Jacquemyn, Z. Zhao, et al. 2020. The
907 architecture of the network of orchid–fungus interactions in nine co-occurring
908 Dendrobium species. Frontiers in Ecology and Evolution 8:1–10.
909