

APPLICATION

paco: implementing Procrustean Approach to Cophylogeny in R

Matthew C. Hutchinson^{*,1,2}, E. Fernando Cagua², Juan A. Balbuena³, Daniel B. Stouffer² and Timothée Poisot⁴

¹Department of Ecology and Evolutionary Biology, Princeton University, 106A Guyot Hall, Princeton, NJ 08544, USA; ²Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand;

³Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, 2 Professor José Beltrán Martínez Street, Paterna, Valencia 46980, Spain; and ⁴Department of Biological Sciences, University of Montréal, Pavillon Marie-Victorin, 90 Vincent-d'Indy Avenue, Montréal, QC H2V 2S9, Canada

Summary

1. The concordance of evolutionary histories and extant species interactions provides a useful metric for addressing questions of how the structure of ecological communities is influenced by macro-evolutionary processes.
2. We introduce `paco` (v0.3.1), an R package to perform Procrustean Approach to Cophylogeny. This method assesses the phylogenetic congruence, or evolutionary dependence, of two groups of interacting species using both ecological interaction networks and their phylogenetic history.
3. We demonstrate the functionality of `paco` through its application to empirical host-parasite and plant-pollinator communities.
4. Although the package is intended to assess the phylogenetic congruence between groups of interacting species, the method is also directly applicable to other scenarios that may show phylogenetic congruence including historical biogeography, molecular systematics, and cultural evolution.

Key-words: codivergence, cophylogeny, cospeciation, covariance, evolutionary ecology, phylogenetic community structure, phylogenetic congruence, phylogenetic structure

Introduction

Inquiry at the interface of ecological and evolutionary biology has rapidly expanded with the emergence of robust phylogenetic comparative methods (Webb *et al.* 2002; Cavender-Bares *et al.* 2009; Mouquet, Devictor & Meynard 2012). The relationships between extant ecological interactions and their evolutionary history provide a novel perspective on the study of ecological communities (Cavender-Bares *et al.* 2009; Vamosi *et al.* 2009). For instance, work in this area has already highlighted the existence of phylogenetic structure in ecological communities (Emerson & Gillespie 2008; Nuismer, Jordano & Bascompte 2013; Eklöf & Stouffer 2016), the role of evolution in community assembly (Kraft *et al.* 2007), and the feedback between ecological interactions and evolutionary processes (Galletti, Guevara & Côrtes 2013). Here, we contribute a new item to the toolbox – `paco` – the implementation in R of the PACo method (Balbuena, Míguez-Lozano & Blasco-Costa 2013) to explore the phylogenetic congruence of interacting species.

Cophylogeny is the study of concordance between the phylogenies and interactions of two groups of species (Page 2003; Balbuena, Míguez-Lozano & Blasco-Costa 2013). Phylogenetic congruence of interactions between two clades – interactions between species of similar phylogenetic position – indicates that there is shared evolutionary history between the groups (Hafner *et al.* 2003; but see Herrera, Hirooka & Chaverri 2016). Although phylogenetic congruence suggests an evolutionary link between ecologically associated species, the processes that result in phylogenetic congruence remain unclear (Poisot 2015). Phylogenetic congruence between species has been causally attributed to both coevolutionary (Smith *et al.* 2008; Godsoe *et al.* 2009) and biogeographical (e.g. co-vicariance; Weckstein 2004) processes. Likewise a cophylogenetic pattern may also arise from evolutionary tracking of one group by another – such as the evolution of *Polystoma* in response to habitat acquisition by their amphibian hosts (Bentz *et al.* 2006). Nevertheless, phylogenetic congruence between clades provides additional means with which to study how the ecological associations of species and clades are reflected in their evolutionary history.

There are several tools (e.g. TREEMAP, ParaFit, Jane, PACo, RASCAL) currently available to assess phylogenetic

*Correspondence author. E-mail: matthewhutchinson15@gmail.com

congruence between interacting species (e.g. Charleston & Page 2002; Legendre, Desdevises & Bazin 2002; Hommola *et al.* 2009; Conow *et al.* 2010; Balbuena, Míguez-Lozano & Blasco-Costa 2013; Drinkwater & Charleston 2016). These tools can be typically classified as either event-based or global-fit methods (Desdevises 2007; Balbuena, Míguez-Lozano & Blasco-Costa 2013). Event-based methods map one evolutionary tree to the other, based on either edges or nodes, with the aim of understanding the underlying evolutionary processes, particularly coevolutionary associations (Conow *et al.* 2010; Drinkwater & Charleston 2014). It is now feasible to use event-based approaches for large datasets thanks to recent advances that significantly reduce their computational cost (Drinkwater & Charleston 2014; Libeskind-Hadas *et al.* 2014; Drinkwater & Charleston 2016). Conversely, global-fit approaches assess the degree of congruence between two phylogenies and some can also identify the specific interactions (or links) that contribute most to their concordance (Legendre, Desdevises & Bazin 2002; Desdevises 2007; Balbuena, Míguez-Lozano & Blasco-Costa 2013). Although global-fit methods cannot make the evolutionary inferences of event-based methods, it has been suggested that the role of coevolution in particular can be inferred from the extent to which phylogenetic congruence can be identified by global-fit methods (Desdevises 2007). Global-fit methods are also attractive in that, unlike some event-based methods (although see the computational advances described above), they can handle large-scale ecological data making it possible to examine phylogenetic congruence across even exceptionally rich clades of interacting species (Meier-Kolthoff *et al.* 2007; Balbuena, Míguez-Lozano & Blasco-Costa 2013).

Recently, Balbuena, Míguez-Lozano & Blasco-Costa (2013) introduced a novel global-fit method, PACo. Their method – Procrustean Approach to Cophylogeny – assesses phylogenetic congruence with the explicit aim to test the dependency of one phylogeny on the other (Balbuena, Míguez-Lozano & Blasco-Costa 2013). Going further than other methods, consideration of evolutionary dependency by PACo also allows coevolutionary inferences to be made at the level of species' interactions (Balbuena, Míguez-Lozano & Blasco-Costa 2013; Poisot & Stouffer 2015). Furthermore, PACo has been shown to perform as well as, or better than, other global-fit methods such as ParaFit and Hommola *et al.*'s correlation-based test (Legendre, Desdevises & Bazin 2002; Hommola *et al.* 2009; Balbuena, Míguez-Lozano & Blasco-Costa 2013). Therefore, PACo provides a robust, versatile tool capable of not only exploring cophylogenetic hypotheses in interacting clades but also the potential coevolutionary dynamics of those clades (Balbuena, Míguez-Lozano & Blasco-Costa 2013).

Here, we introduce the R package, `paco` (v0.3.1; doi:10.5281/zenodo.192508): a tool for undertaking cophylogenetic analysis in R based on the Procrustean approach of Balbuena, Míguez-Lozano & Blasco-Costa (2013). This package formalises and extends the original R code provided by Balbuena, Míguez-Lozano & Blasco-Costa (2013) in several key areas to increase the utility of the method (Table 1). As a package on the CRAN database, an immediate advantage of `paco` is increased accessibility for both installation and use.

Table 1. The functions of `paco` and their description

| Function | Description |
|----------------------------------|--|
| <code>prepare_paco_data()</code> | Assigns the order of superimposition where P is superimposed on H. Also checks that the interaction matrix and respective phylogenetic matrices are adequate for PACo analysis |
| <code>add_pcoord()</code> | Performs a Principal Coordinates analysis of the phylogenetic distance matrices |
| <code>PACo()</code> | Implements the Procrustean superimposition of the phylogenetic objects, performing their cophylogenetic analysis |
| <code>paco_links()</code> | Returns the contribution of individual interactions to the overall cophylogenetic signal based on a jackknifing procedure |
| <code>residuals_paco()</code> | Returns the contribution of individual interactions to the overall cophylogenetic signal in terms of the raw residuals of the Procrustean superimposition |

Furthermore, `paco` has the advantage that code for the analysis of phylogenetic congruence is broken into distinct functional steps providing a more user-friendly approach (Table 1). Finally, `paco` extends the analysis of individual links with parallel implementation of the `paco_links` function and a new function (`residuals_paco`) to obtain the raw Procrustean residuals of each link.

Overview of `paco`

The most widely-used global-fit approach to cophylogeny to date – ParaFit (Legendre, Desdevises & Bazin 2002) – uses the 4th-corner method to examine non-independence between phylogenies based on their interactions. The approach implemented here instead addresses phylogenetic congruence with an analogous tool, Procrustean superimposition. Both approaches can provide a statistic for the global congruence of phylogenies and the contribution of individual links to the global congruence (Legendre, Desdevises & Bazin 2002; Balbuena, Míguez-Lozano & Blasco-Costa 2013). However the Procrustean method of PACo – which assesses the degree to which the parasite phylogeny tracks that of the host – performs better than ParaFit in terms of Type I error rates and statistical power while also allowing a more direct inference of a coevolutionary process (Balbuena, Míguez-Lozano & Blasco-Costa 2013).

To undertake cophylogenetic analysis with `paco`, three items of data are required (Fig. 1). First, two dissimilarity matrices (i.e. those based on the respective phylogenies) are needed, one for each group of interest (e.g. one for the hosts and one for the parasites). Second, a binary matrix describing the associations between the two groups is needed for the appropriate superimposition. PACo analysis is then a straightforward procedure (Fig. 2). The data are initially projected into multivariate space via Principal Coordinates Analysis in `paco::add_pcoord` – where points (species) are associated based on observed interactions between them. These matrices

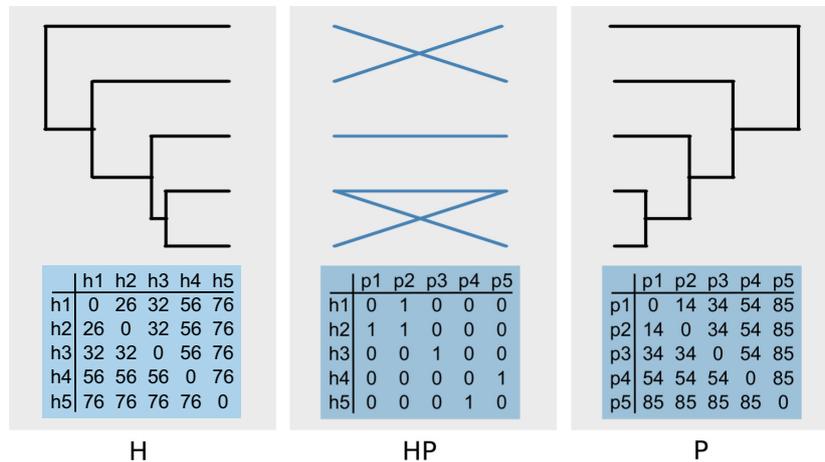


Fig. 1. The three items of eco-evolutionary data required for cophylogenetic analysis of species' interactions in *paco* (simplified from Balbuena, Míguez-Lozano & Blasco-Costa 2013). H and P show the distance matrices of each group that are required. These are generated from the phylogeny of each group. Alongside the distance matrices, *paco* takes an association matrix describing interactions between groups (HP). As shown, the associations between groups should be an adjacency matrix with the group that will be superimposed *onto* (H) in the rows and the group that will be superimposed (P) in the columns.

```
## first load the required packages and data
library(paco)
library(ape)
data(gopherlice)
## next transform the phylogenies to distance matrices
gtree <- cophenetic(gophertree)
ltree <- cophenetic(licetree)
## bundle the data together and add Principal Coordinates
D <- prepare_paco_data(H = gtree, P = ltree, HP = gl.links)
D <- add_pcoord(D, correction = 'cailliez')
## perform cophylogenetic analysis
D <- PACo(D, nperm = 1000, seed = 12, method = 'r0',
symmetric = FALSE)
## return interaction-specific cophylogenetic contributions
based on a jackknifing procedure
D <- paco.links(D)
## return interaction-specific cophylogenetic contributions in
terms of individual superimposition residuals
res <- residuals_paco(D$proc)
```

Fig. 2. The step-wise methodology of conducting cophylogenetic analysis in *paco*. Initialising the data requires constructing distance matrices (i.e. finding the patristic distances of each group based on their phylogeny with the *ape::cophenetic* function). After assigning the order of superimposition and grouping the data ($2 \times$ phylogenetic distance matrices and association matrix) in *paco::prepare_paco_data*, Principal Coordinates Analysis of these three matrices are done in *paco::add_pcoord* with appropriate correction of negative eigenvalues. A Procrustean Approach to Cophylogeny is done in *paco::PACo*, and finally the contribution of individual links can be returned with *paco::paco_links* or *paco::residuals_paco*. For each step in the process, the results are concatenated into object 'D', a list of class 'paco'.

then undergo a Procrustean superimposition in *paco::PACo* where the level of cophylogenetic signal is taken as the global sum of squared residuals (m_{XY}^2) in the best-fit superimposition of the two phylogenies (Balbuena, Míguez-Lozano & Blasco-Costa 2013). Finally, the contribution of individual interactions to the overall cophylogenetic signal can be explored with *paco::residuals_paco* or *paco::paco_links*. The former returns the observed residuals of Procrustean

superimposition where a small contribution (residual) to m_{XY}^2 indicates a link more congruent with a cophylogenetic hypothesis. Similarly, the contributions of individual links, as quantified by *paco::paco_links*, are estimated with a jackknife procedure, whereby one link is removed and the fit reassessed. The contribution of that link is therefore the change in fit when said link is removed. Again, a link that shows stronger support to a cophylogenetic hypothesis is one that makes a small contribution to m_{XY}^2 .

Although the global fit of the phylogenies can provide some indication of phylogenetic congruence between the two groups, the statistical and eco-evolutionary significance of this value is dependent on an appropriate null model. Thus, the significance level returned by *paco::PACo* indicates whether or not the observed interaction network is significantly more congruent with both group's evolutionary history than an ensemble of random instances of itself. We have written *paco* in such a way that it provides the user with a large suite of possible null models by employing the swap algorithms of *vegan::nullmodel* (Table 2; Oksanen *et al.* 2016). Each method – as an argument to *paco::PACo* – is described in depth in the documentation of the R function *vegan::commsim* (Oksanen *et al.* 2016) and in Wright *et al.* (1997). Briefly, the potential algorithms allow binary null models that conserve the total number of interactions (*r00*), degree-distribution (number of interactions per species) of either row group (*r0*), column group (*c0*), or both groups (*swap*, *tswap*, *quasiswap*, and *backtracking*; Table 2). Additionally, null models *r1* and *r2* rewire the network based on species-specific interaction information while maintaining the number of interactions in each row (Table 2; Wright *et al.* 1997). Note that quantitative randomisation algorithms are also available but not explored here since *paco* does not make use of this information. The user can also retrieve the measure of phylogenetic congruence, m_{XY}^2 , for these randomisations with the argument 'shuffled=TRUE' to the *paco::PACo* function.

Table 2. The randomisation algorithms available in paco from the R package *vegan* (Oksanen *et al.* 2016)

| Algorithm | Constraints | Description |
|------------------------|---|---|
| r00 | Overall fill (the total number of interactions in the network) is maintained | Should be used when <i>only</i> the number of interactions is assumed to influence cophylogenetic signal |
| r0 | Overall fill and row degree are maintained | Should be used under the assumption that the column group tracks the evolution of the row group, therefore, maintaining the degree of the species in the row group |
| c0 | Overall fill and column degree are maintained | Should be used under the assumption that the row group tracks the evolution of the column group, therefore, maintaining the degree of the species in the column group |
| backtracking and swaps | Overall fill, column degree, and row degree are maintained | Should be used when it is unclear which group is tracking the other, therefore, conserving the degree of both groups |
| r1 and r2 | Overall fill and row degree are maintained. Randomisation of network is probabilistic and based on the degree of column species | Should be used when the column group is assumed to track the row group and when the specialisation/generalism of the column group also determines cophylogenetic signal |

Each null model algorithm is listed alongside its constraints and the situation in which it could be used. 'swaps' refers to the `quasiswap`, `swap`, and `tswap` methods. Note that 'r00' is the least conservative of the algorithms listed as each interaction between species contributes to the Procrustes global fit statistic. Therefore, if the number of interactions is not conserved between randomisations you are more likely to see a statistically significant, but not necessarily meaningful, difference between observed and permuted matrices. We include it here for completeness despite the fact that it may not be applicable to some situations in which phylogenetic congruence may be studied.

Our implementation of PACo makes one additional and major improvement upon the original method. Retaining the ability to test the dependence of one phylogeny on the other (Fig. S1, Supporting Information), as in Balbuena, Míguez-Lozano & Blasco-Costa (2013) and most appropriate for systems where one clade is thought to drive the evolution of the other, it is now also possible to test the dependence of both phylogenies on each other (Fig. S1). This variation in implementation is achieved with the `vegan::procrustes` function in the R package *vegan* (Peres-Neto & Jackson 2001; Oksanen *et al.* 2016) and can be specified with the `symmetric` argument of the `PACo` function. This adaptation may be particularly important when considering more diffuse interaction types such as pollination or seed dispersal where it is not clear whether one clade depends on the evolution of the other. It should also be noted that the `vegan::procrustes` function can return warnings when the phylogenies have differing numbers of tips which, in turn, leads to differing numbers of axes in the Principal Coordinates analysis. These warnings are trivial, as noted by Balbuena, Míguez-Lozano & Blasco-Costa (2013), because the smaller of two is complemented with columns of zeros to make both matrices the same size (Legendre & Legendre 2012; Oksanen *et al.* 2016). These warnings can be suppressed when calling `paco::PACo` or `paco::paco_links` with the `'proc.warnings'` argument.

The final important note on implementation addresses the typically non-Euclidean nature of phylogenetic distances. Non-Euclidean distances pose a problem when they are transformed to Principal Coordinates (in `paco::add_pcoord`) as this can lead to negative eigenvalues being produced that cannot be represented in the real space that `paco` operates in. To avoid this problem the user can take several approaches. First, an eigenvalue correction may be applied as the `correction` argument to `paco::add_pcoord`. Possible corrections include 'cailliez' (Cailliez 1983) and 'lingoes' (Lingoes 1971), both of which correct negative eigenvalues with the

addition of a constant (Paradis, Claude & Strimmer 2004). Alternatively, de Vienne, Aguileta & Ollier (2011) showed that the matrix of patristic distances produced by the `ape::cophenetic` function can be made Euclidean by taking the element-wise square-root of the distances. To implement this correction the user would apply the `sqrt` function to the patristic distance matrices produced by `ape::cophenetic` and feed 'none' (the default) to the correction argument in `paco::add_pcoord`.

Example analysis

The `paco` package for R we present here follows a recipe-like structure for undertaking a Procrustean Approach to Cophylogeny (Fig. 2). To demonstrate the method and utility of `paco`, we develop two specific examples. First, we recreate the classic cophylogenetic analysis of gophers and their lice ectoparasites (Hafner & Nadler 1988) featured in Balbuena, Míguez-Lozano & Blasco-Costa (2013) and provided as example data in `paco`. Second, we demonstrate the increased utility of the PACo method to analyse the phylogenetic congruence of a pollination community, using an interaction network from Arroyo, Primack & Armesto (1982).

PHYLOGENETIC CONGRUENCE OF GOPHERS AND CHEWING LICE

Hosts and their parasites represent the classic study system of cophylogenetic analysis due to their intimate and specific associations (Hafner & Nadler 1988; Weckstein 2004; Bentz *et al.* 2006). Moreover, the general dependence of parasites on their hosts means that when exploring host-parasite phylogenetic congruence, we can assume that the hosts drive the evolution of their parasites (Light & Hafner 2007). This assumption informs our implementation of `paco` in three ways.

1. We superimpose the parasite phylogeny on the host phylogeny in order to test the specific hypothesis that chewing lice speciation tracks gopher speciation. This is achieved in `paco::prepare_paco_data` by giving the argument `H` the gopher distance matrix and argument `P` the chewing lice distance matrix.

2. The symmetric argument of `paco::PACo` is set to `FALSE` (the default). In this case, the lice phylogeny will be scaled and transformed to fit gopher phylogeny and thus the best fit of the superimposition is returned relative to the size of the gopher phylogeny.

3. The null model algorithm may be less conservative under this evolutionary assumption as we test the hypothesis that the evolution of chewing lice tracks the gopher phylogeny. In this case, we will set the `method` argument of `paco::PACo` to `r0` to maintain the number of interactions of the gopher species only (Table 2).

The results of this cophylogenetic analysis with `paco` conform to the original results of Hafner & Nadler (1988). Hafner & Nadler (1988) saw a distinct cophylogenetic signal between gophers and their chewing lice and `paco` returns a significant cophylogenetic signal between the groups ($m_{XY}^2 = 0.073$, $P < 0.001$, $n = 1000$) where none of the 1000 random instances of the interaction network show more phylogenetic congruence than the observed network (Fig. 3b). This supports the hypothesis that the chewing lice phylogeny tracks the gopher phylogeny, and that the two groups have

undergone coupled evolutionary change. The results of `paco` can be visually confirmed when examining a diagrammatic representation of the interaction network, where only a handful of interactions are not concordant with speciation events (Fig. 3a).

PHYLOGENETIC CONGRUENCE OF AN ANDEAN POLLINATION NETWORK

Second, we examine the phylogenetic congruence between a plant and pollinator community of the Chilean Andes. This network, originally collected and presented by Arroyo, Primack & Armesto (1982) and collated by Rezende *et al.* (2007), is freely available from the Web of Life database (<http://www.web-of-life.es/>). In order to perform these analyses, we removed unidentified and unconnected species from the original network. The plant and pollinator phylogenies used in this analysis were built with the R package `ape` (Paradis, Claude & Strimmer 2004) from taxonomic tables and dated with the `bladj` function of `phylocom`. The ages used to date these phylogenies were taken from the Wikström, Savolainen & Chase (2001) angiosperm phylogeny and Misof, Liu & Meusemann (2014) insect phylogeny respectively.

Cophylogenetic analysis of this plant-pollinator community differs from the previous example in several important ways. First, we investigate phylogenetic congruence here at the scale of a community of co-occurring, but not necessarily closely

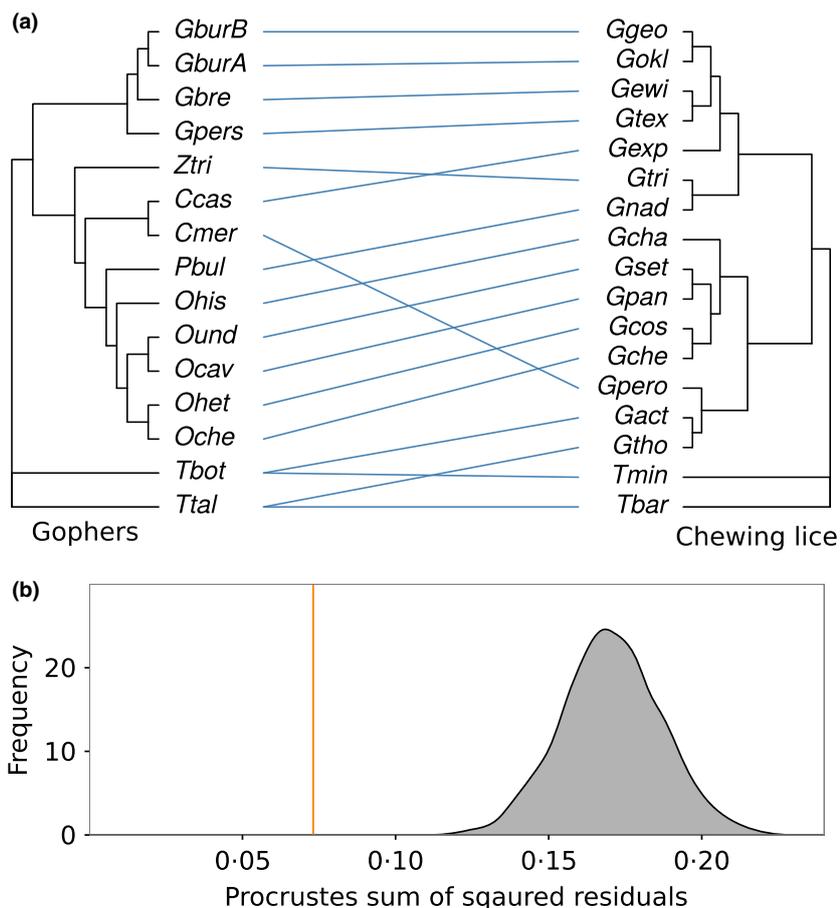


Fig. 3. (a) The interaction network and phylogenies of gophers and chewing lice (Hafner & Nadler 1988). The paucity of interactions makes the phylogenetic congruence visually apparent between these two groups as interactions tend to occur between species of coupled evolutionary history. A table of species' full taxonomic names and their identifier used here (first letter of genus and first 3 letters of species) can be found in Table S1. (b) Analysis with `paco` supports the above conclusion with an observed best-fit Procrustes superimposition (orange line) better than the same for any of the ensemble of network randomisations in the null model.

related species, rather than the clade-level of the above example and characteristic of most of the cophylogeny literature (Weckstein 2004; Desdésives 2007; Smith *et al.* 2008; Godsoe *et al.* 2009). Second, pollination interactions are less intimate and more abundant than the specific parasitic relationships of gophers and chewing lice. Investigation of phylogenetic congruence between flowering plants and pollinators therefore requires an alternative approach. Note that analysis code for this example and data used can be found online (code as Appendix S2 and data at Data Dryad).

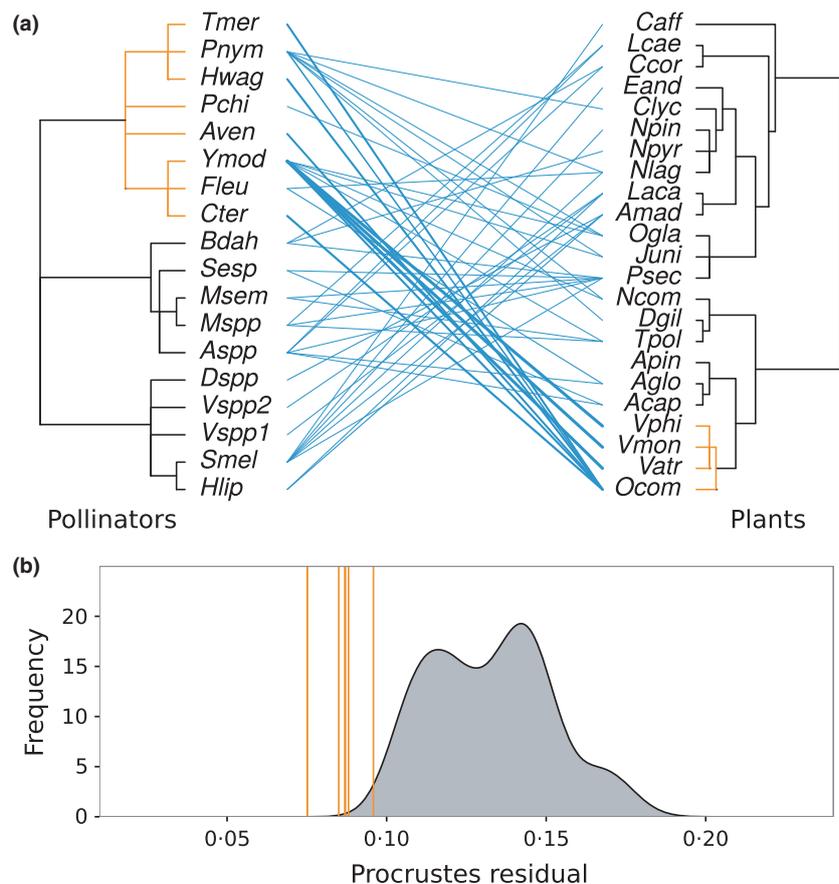
In ecological systems such as pollination interactions, where it is not clear which group may be driving the evolution of the other, the implementation of `paco` should be more conservative. First, the designation of H (the phylogeny superimposed on) and P (the phylogeny that is superimposed on the other) in `paco::prepare_paco_data` is unimportant because in `paco::PACo` the `symmetric` argument is set to `TRUE`. This means that both phylogenies are standardised prior to superimposition resulting in the best-fit of the superimposition being independent of both phylogenies. Second, in a more diffuse ecological system such as pollination, we implement a more constrained null model (`method='quasiswap'`) in `paco::PACo` where the number of interactions is conserved for each species and accordingly in the network as a whole (Table 2).

Cophylogenetic analysis with `paco` of this pollination community shows that, despite the weaker taxonomic relationships and interaction intimacy, a significant cophylogenetic signal is present ($m_{XY}^2 = 0.83$, $P < 0.001$, $n = 1000$). Phylogenetic

congruence at this scale – given that the pollinators and flowering plants in the community are depauperate representations of their respective clades – more likely indicates that lineages within those clades have evolutionarily congruent relationships rather than describing phylogenetic congruence between these specific species. However, digging deeper into this relationship, we can observe that specific pollination interactions show more support for a hypothesis of phylogenetic congruence than others with `paco::residuals_paco` (note, `paco::paco_links` may also be used for this assessment).

In this network, it is clear to see that nine interactions appear to have a much stronger cophylogenetic signal (smaller residual distance) than the rest (Fig. 4). The interactions that show the most support for phylogenetic congruence occur between species from two rosoid sister orders (Malpighiales and Oxalidales; hereafter M/O) and the lepidopteran species of the network (Fig. 4a). Indeed, if we focus on just the interactions between these two groups, we see that they show significantly stronger support to a hypothesis of phylogenetic congruence than the other interactions in the network based on comparing residuals of these interactions to those of the remaining interactions (Welch's *t*-test; $t = -13.07$, d.f. = 33.2, $P < 0.001$, Figs 4b and 5). This makes intuitive sense from the plant's perspective given that all but one of the interactions these plants participate in (in this network) occur with lepidoptera, exemplified by the three members of *Viola* only interacting with the nymphalid *Yramea modesta* (Fig. 4a). On the other hand, the lepidoptera-M/O interactions only account

Fig. 4. The interaction network and phylogenies of flowering plants and pollinators from the Chilean Andes (Arroyo, Primack & Armesto 1982). The community as a whole shows significant phylogenetic congruence, as analysed with `paco`. Moreover, in (a) the interactions between plants and pollinators are weighted by their contribution to the overall phylogenetic congruence – where thicker lines indicate a smaller residual distance or an interaction that shows stronger support to a hypothesis of phylogenetic congruence. With these weights, it can be seen that the interactions in the network that show the strongest support for phylogenetic congruence occur between a clade of eight lepidopteran species (Tmer, Pnym, Hwag, Pchi, Aven, Ymod, Fleu, Cter) and a clade of plants – three *Viola* species (Vphi, Vmon, Vatr) and *O. compacta* (Ocom) – present in this network (orange clades in each phylogeny). A table of species' full taxonomic names and their identifier used here (first letter of genus and first 3 letters of species) can be found in Table S2. In (b) this result is presented more explicitly where the orange lines are the interactions that occur between the two orange clades in (a), and the distribution gives the values of the rest of the interactions in the network.



for 41% of interactions (9/22) that the lepidopteran species participate in (Fig. 4a). This would suggest that it is the plant species (3 *Viola* spp. and *Oxalis compacta*) that drive the cophylogenetic signal of these interactions. In addition, we can examine the role of species' degree in predicting residual distance with a linear regression. Specialisation or generalism of the plant species (quantified by the number of interaction partners a plant species has) in this network does not predict the cophylogenetic signal of their interactions despite the perhaps intuitive relationship between phylogenetic congruence and specialisation ($t = -0.84$, $P = 0.40$). While each of the *Viola* species only engages in one interaction, *O. compacta* is the most generalist plant species in the network with seven interactions, albeit six of them with the lepidopterans.

Given that ecological communities are labile assemblages, it is to be expected that certain interactions are more consistent with a hypothesis of phylogenetic congruence than others. In this network of plant–pollinator interactions from the Chilean Andes, those interactions are the ones between the lepidopterans species and the species of the rosid clades containing the Malpighiales and Oxalidales. The disparate strength at which interactions across the network support a hypothesis of phylogenetic congruence likely has differing causes and impacts on both interactions and species. Therefore, further work on patterns of phylogenetic congruence should be aimed at teasing apart the ecological and evolutionary consequences that these differences in phylogenetic congruence have on the participant species.

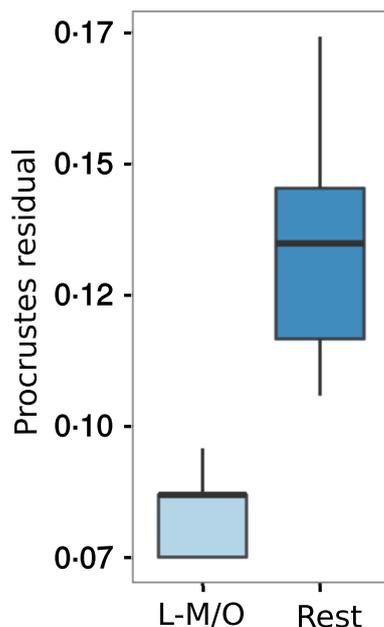


Fig. 5. The comparison of the interactions that are the most phylogenetically congruent (as indicated by `residuals_paco`) between lepidopterans and the rosid clade (L–M/O) compared to the remainder of the network (Rest). The residuals of the L–M/O interactions are significantly smaller than the remainder of the interactions in the network, indicating that these interactions show significantly greater phylogenetic congruence than the rest. This suggests that significant phylogenetic congruence at the whole network scale may be driven by the interactions between these taxa.

Wider applicability and future development

The current utility of `paco` allows the application of the PACo method to a variety of cophylogenetic problems, both ecological and otherwise. As outlined above, classical cophylogeny is the assessment of congruence between the phylogenies of two interacting clades (Hafner & Nadler 1988; Weckstein 2004; Balbuena, Míguez-Lozano & Blasco-Costa 2013). However, the Procrustean method performed in `paco` is also amenable to exploring, among other questions, the concordance of phylogenies derived from different sources (Pérez-Escobar, Balbuena & Gottschling 2016), the geographic variation of ecological traits, and the parallel evolution of genes and culture.

Aside from the phylogenetic congruence of interacting clades, the PACo method has also been directly applied to the study of the concordance of gene-derived phylogenies. Pérez-Escobar, Balbuena & Gottschling (2016) developed a pipeline based on PACo to examine the congruence between phylogenies inferred from nuclear and organelle markers in the Neotropical orchids *Catasetinae*. This was achieved using a nuclear gene phylogeny, an organelle gene phylogeny, and specific organelle–host associations as the input data to the PACo method. Doing so revealed phylogenetically incongruent associations between the trees derived from chloroplast and nuclear sequences in three genera (Pérez-Escobar, Balbuena & Gottschling 2016). Testing the congruence between phylogenies such as this can be easily performed with `paco` and has the potential to identify important evolutionary events in the history of a clade (Pérez-Escobar, Balbuena & Gottschling 2016). Therefore, a useful development – from the perspective of comparing phylogenetic trees – would be the adaptation of `paco_links` to identify the outlier associations between phylogenies in the same way achieved by Pérez-Escobar, Balbuena & Gottschling (2016). In addition, a similar approach can be adopted for comparison of multiple gene trees by means of a Generalised Procrustes Analysis (Gower 1975) to identify consensus phylogenies, detect phylogenetic outliers, and comparing prospective trees to a reference tree (as previously suggested by Choi & Gomez 2009).

In addition to the areas in which PACo has already been applied and where `paco` can immediately be implemented, there are new possibilities for ecological analysis that the future development of `paco` could cater to. In such cases, distance measures based on phylogenies may not be appropriate to use and, therefore, other dissimilarity measures may be used in their stead. For example, Nieberding, Joussein & Desdevises (2010) proposed a cophylogenetic method based on ParaFit that can be applied to geographic trait variation at the phylogeographic rather than phylogenetic scale. This approach takes an association matrix – weighted by ecological traits – describing coexistence relationships between individuals of two species along with the phylogeographic trees of those species (Nieberding, Joussein & Desdevises 2010). With this method, it is possible to test how traits and geographic areas contribute to the congruence of the ecological association (Nieberding, Joussein & Desdevises 2010). The innate similarities between

ParaFit and PACo indicate that this method could be also be addressed with a Procrustean approach. Development of paco to tackle these questions would, with the right data, facilitate study of the phylogenetic congruence between taxa at both the macro- and micro-evolutionary scale.

Finally, potentially phylogenetically congruent associations exist outside eco-evolutionary systems. In the social sciences, the concept of phylogenetic congruence is readily applicable (as demonstrated by the analysis of population and cultural divergence in Iranian tribes; Tehrani, Collard & Shennan 2010). Likewise, paco may be ideal to investigate synchronous cultural developments such as the analysis of long-house design and language in the Pacific North-West (Jordan & O'Neill 2010). The extension of cophylogenetic methods – and indeed the phylogenetic comparative method as a whole – into the social sciences promises to extend these methods to the benefit of both eco-evolutionary inquiry and the humanities.

Authors' contributions

T.P., M.C.H. and E.F.C. were responsible for package development; M.C.H. led the writing of the manuscript and performed analyses. All authors contributed substantially to writing the code for the package, contributed critically to manuscript drafts, and gave final approval for publication.

Acknowledgements

We thank the original co-developers of PACo – R. Míguez-Lozano and I. Blasco-Costa – for helping to introduce the method and cutting the path for its implementation as a package in R. We acknowledge the original authors of each dataset, as well as E. Rezende and J. Bascompte for providing the ecological network database from which data was taken. We also acknowledge the work of the artists (Adrian Reich and Alexander Schmidt-Lebuhn) whose pictures are used in Fig. S1. M.C.H. acknowledges a University of Canterbury Summer Scholarship. E.F.C. acknowledges the support of a University of Canterbury Doctoral Scholarship, a New Zealand International Doctoral Research Scholarship, and a University of Canterbury Meadow Mushrooms Postgraduate Scholarship. J.A.B. acknowledges the MINECO/FEDER (CGL2015-71146-P) and the Generalitat Valenciana (PROMETEOII/2015/018). D.B.S. acknowledges the support of a Rutherford Discovery Fellowship, administered by the Royal Society of New Zealand. T.P. thanks Compute Québec for providing computational resources during development.

Data accessibility

paco MIT open source license held by the Poisot Lab @ U. Montréal/the Stouffer Lab @ U. Canterbury, 2015. PACo is tested on previous, current, and future versions of R. The gopher and lice data used, source code, and documentation are freely available on CRAN (<http://cran.r-project.org/web/packages/paco>) and GITHUB (<http://github.com/PoisotLab/paco>). Furthermore, the plant-pollinator interaction web, phylogenies, and analysis code are available on Data Dryad <https://doi.org/10.5061/dryad.j555d> (Hutchinson *et al.* 2017)

References

Arroyo, M.T.K., Primack, R.B. & Armesto, J.J. (1982) Community studies in pollination ecology in the high temperate Andes of central Chile. I. Pollination mechanisms and altitudinal variation. *American Journal of Botany*, **69**, 82–97. ISSN 0378-2697.

Balbuena, J.A., Míguez-Lozano, R. & Blasco-Costa, I. (2013) PACo: a novel Procrustes application to cophylogenetic analysis. *PLoS ONE*, **8**, e61048.

Bentz, S., Sinnappah-Kang, N.D., Lim, L.H.S., Lebedev, B., Combes, C. & Verneau, O. (2006) Historical biogeography of amphibian parasites, genus *Polystoma* (Monogenea: Polystomatidae). *Journal of Biogeography*, **33**, 742–749.

Cailliez, F. (1983) The analytical solution of the additive constant problem. *Psychometrika*, **48**, 305–308.

Cavender-Bares, J., Kozak, K.H., Fine, P.V. & Kembel, S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.

Charleston, M. & Page, R. (2002) TREEMAP 2.0 B: A Macintosh program for the analysis of how dependent phylogenies are related, by cophylogeny mapping.

Choi, K. & Gomez, S.M. (2009) Comparison of phylogenetic trees through alignment of embedded evolutionary distances. *BMC Bioinformatics*, **10**, 1.

Conow, C., Fielder, D., Ovadia, Y. & Libeskind-Hadas, R. (2010) Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms for Molecular Biology*, **5**, 1.

Desdevises, Y. (2007) Cophylogeny: insights from fish-parasite systems. *Parasitologia*, **49**, 125.

Drinkwater, B. & Charleston, M.A. (2014) An improved node mapping algorithm for the cophylogeny reconstruction problem. *Coevolution*, **2**, 1–17.

Drinkwater, B. & Charleston, M.A. (2016) RASCAL: a randomized approach for coevolutionary analysis. *Journal of Computational Biology*, **23**, 218–227.

Eklöf, A. & Stouffer, D.B. (2016) The phylogenetic component of food web structure and intervality. *Theoretical Ecology*, **9**, 107–115.

Emerson, B.C. & Gillespie, R.G. (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, **23**, 619–630.

Galetti, M., Guevara, R., Cortés, M.C. *et al.* (2013) Functional extinction of birds drives rapid evolutionary changes in seed size. *Science*, **340**, 1086–1090.

Godsoe, W., Strand, E., Smith, C.I., Yoder, J.B., Esque, T.C. & Pellmyr, O. (2009) Divergence in an obligate mutualism is not explained by divergent climatic factors. *New Phytologist*, **183**, 589–599.

Gower, J.C. (1975) Generalized procrustes analysis. *Psychometrika*, **40**, 33–51.

Hafner, M.S. & Nadler, S.A. (1988) Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*, **332**, 258–259.

Hafner, M.S., Demastes, J.W., Spradling, T.A. & Reed, D.L. (2003) Cophylogeny between pocket gophers and chewing lice. *Tangled Trees: Phylogeny, Cospeciation, and Coevolution* (ed R.D.M. Page), pp. 195–218. University of Chicago Press, Chicago, IL, USA.

Herrera, C.S., Hirooka, Y. & Chaverri, P. (2016) Pseudocospiculation of the mycoparasite *Cosmospora* with their fungal hosts. *Ecology and Evolution*, **6**, 1504–1514.

Hommola, K., Smith, J.E., Qiu, Y. & Gilks, W.R. (2009) A permutation test of host–parasite cospeciation. *Molecular Biology and Evolution*, **26**, 1457–1468.

Hutchinson, M.C., Cagua E.F., Balbuena, J.A., Stouffer, D.B., & Poisot, T. (2017) Data from: paco: implementing Procrustean Approach to Cophylogeny in R. *Dryad Digital Repository*, <https://doi.org/10.5061/dryad.j555d>.

Jordan, P. & O'Neill, S. (2010) Untangling cultural inheritance: language diversity and long-house architecture on the Pacific northwest coast. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365**, 3875–3888.

Kraft, N.J., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007) Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *The American Naturalist*, **170**, 271–283.

Legendre, P. & Legendre, L.F. (2012) *Numerical Ecology*, vol. 24. Elsevier, Amsterdam, the Netherlands.

Legendre, P., Desdevises, Y. & Bazin, E. (2002) A statistical test for host–parasite coevolution. *Systematic Biology*, **51**, 217–234.

Libeskind-Hadas, R., Wu, Y.C., Bansal, M.S. & Kellis, M. (2014) Pareto-optimal phylogenetic tree reconciliation. *Bioinformatics*, **30**, i87–i95.

Light, J.E. & Hafner, M.S. (2007) Cophylogeny and disparate rates of evolution in sympatric lineages of chewing lice on pocket gophers. *Molecular Phylogenetics and Evolution*, **45**, 997–1013.

Lingoes, J.C. (1971) Some boundary conditions for a monotone analysis of symmetric matrices. *Psychometrika*, **36**, 195–203.

Meier-Kolthoff, J.P., Auch, A.F., Huson, D.H. & Göker, M. (2007) CopyCat: cophylogenetic analysis tool. *Bioinformatics*, **23**, 898–900.

Misof, B., Liu, S., Meusemann, K. *et al.* (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**, 763–767.

Mouquet, N., Devictor, V., Meynard, C.N. *et al.* (2012) Ecophylogenetics: advances and perspectives. *Biological Reviews*, **87**, 769–785.

Nieberding, C., Jousset, E. & Desdevises, Y. (2010) The use of co-phylogeographic patterns to predict the nature of host-parasite interactions, and vice versa. *The Biogeography of Host-Parasite Interactions*. (eds B.R. Krasnov & S. Morand) pp. 59–71. Oxford University Press, New York, NY, USA.

Nuismer, S.L., Jordano, P. & Bascompte, J. (2013) Coevolution and the architecture of mutualistic networks. *Evolution*, **67**, 338–354.

Oksanen, J., Blanchet, F.G., Kindt, R. *et al.* (2016) *vegan: Community Ecology Package*. R package version 2.3-3. Available at: <http://CRAN.R-project.org/package=vegan> (accessed 20 January 2016).

- Page, R.M. (2003) Introduction. *Tangled Trees: Phylogeny, Cospeciation, and Coevolution* (ed R.M. Page), pp. 1–22. University of Chicago Press, Chicago, IL, USA.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Peres-Neto, P.R. & Jackson, D.A. (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, **129**, 169–178.
- Pérez-Escobar, O.A., Balbuena, J.A. & Gottschling, M. (2016) Rumbling orchids: how to assess divergent evolution between chloroplast endosymbionts and the nuclear host. *Systematic Biology*, **65**, 51–65.
- Poisot, T. (2015) When is cophylogeny evidence of coevolution? *Parasite Diversity and Diversification* (eds B.R. Morand, S. Krasnov & D.T.J. Littlewood), pp. 420–434. Cambridge University Press, Cambridge, UK.
- Poisot, T. & Stouffer, D. (2015) Coevolution leaves a stronger imprint on interactions than on community structure. *bioRxiv*, p. 033050.
- Rezende, E.L., Lavabre, J.E., Guimaraes, Jr., P.R., Jordano, P. & Bascompte, J. (2007) Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, **448**, 925–9U6.
- Smith, C.I., Godsoe, W.K., Tank, S., Yoder, J.B. & Pellmyr, O. (2008) Distinguishing coevolution from covariation in an obligate pollination mutualism: asynchronous divergence in Joshua tree and its pollinators. *Evolution*, **62**, 2676–2687.
- Tehrani, J.J., Collard, M. & Shennan, S.J. (2010) The cophylogeny of populations and cultures: reconstructing the evolution of Iranian tribal craft traditions using trees and jungles. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **365**, 3865–3874.
- Vamosi, S., Heard, S., Vamosi, J. & Webb, C. (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, **18**, 572–592.
- de Vienne, D.M., Aguilera, G. & Ollier, S. (2011) Euclidean nature of phylogenetic distance matrices. *Systematic Biology*, **60**, 826–832.
- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **11**, 475–505.
- Weckstein, J.D. (2004) Biogeography explains cophylogenetic patterns in toucan chewing lice. *Systematic Biology*, **53**, 154–164.
- Wikström, N., Savolainen, V. & Chase, M.W. (2001) Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **268**, 2211–2220.
- Wright, D.H., Patterson, B.D., Mikkelsen, G.M., Cutler, A. & Atmar, W. (1997) A comparative analysis of nested subset patterns of species composition. *Oecologia*, **113**, 1–20.

Received 6 July 2016; accepted 5 January 2017

Handling Editor: Richard Fitzjohn

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Visual representation of the PACo procedure and tables were abbreviations used in the main text figures are matched to their binomial names.

Appendix S2. R code for undertaking cophylogenetic analysis on the Mendoza pollination network that was used in the main text and visualising these results.

Fig. S1. Diagrammatical representation of the alternative approaches to Procrustean superimposition in paco. In each panel (a–c), the shapes represent a phylogeny of pollinators (blue) and plants (orange) projected in multidimensional space. Dashed lines between shapes represent the observed interactions between the two groups. Procrustean superimposition aims to find the best fit of these phylogenetic objects based on the interactions observed between them. (a) The original projection of the two phylogenies and interactions between them before any superimposition is attempted. This projection is the starting point of both of the approaches to Procrustean superimposition that are shown in (b) and (c). (b) Procrustean superimposition following the ‘asymmetric’ method whereby, if we start with the original projection shown in (a), one phylogenetic object (orange) is fitted to the other (blue) through scaling, rotation and transformation. In this case, the fitting between phylogenies, and thus phylogenetic congruence, is in terms of the phylogeny being fitted to. (c) Procrustean superimposition following the ‘symmetric’ method. In this case, if we start with the original projection shown in (a), both phylogenies are scaled and transformed to get the best fitting. With the ‘symmetric’ method, the cophylogenetic signal is not in terms of either phylogeny but instead a standardised value.

Table S1. Taxonomic species names of pocket gophers, their chewing lice, and the IDs used for the tanglegram of their phylogenies and interactions (Fig. 3). G/L refers to guild. ‘G’ indicates a gopher species while ‘L’ indicates a louse species.

Table S2. Taxonomic species names of flowering plants, their pollinators, and the IDs used for the tanglegram of their phylogenies and interactions (Fig. 4). Pol/Pla refers to guild. ‘Pol’ indicates a pollinator species while ‘Pla’ indicates a plant species.