A PHYLOGENETIC COMPARATIVE METHOD FOR EVALUATING TRAIT COEVOLUTION ACROSS TWO PHYLOGENIES FOR SETS OF INTERACTING SPECIES

Dean C. Adams[1,2,3] and John D. Nason[1]

[1]*Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames IA, 50011*

[2]*Department of Statistics, Iowa State University, Ames IA, 50011*

[3]*Email: dcadams@iastate.edu*

Short title: Co-phylogenetic regression

Key words - phylogenetic comparative method; macroevolution; coevolution; Ficus; Agaonidae

AUTHOR CONTRIBUTIONS: DCA and JDN conceived of the project, and wrote the manuscript; DCA wrote computer code and performed all statistical simulations and data analyses

ACKNOWLEDGMENTS

We thank G. Weiblen for sharing his fig and fig wasp data with us and N. Davis for assistance with the data. This work was sponsored in part by US National Science Foundation grants DEB-1257287 and DEB-1556379 (to DCA), and DEB-1146312 and DEB-1556853 (to JDN).

Data archived in DRYAD: http://dx.doi.org/10.5061/dryad.r3765

ABSTRACT

Evaluating trait correlations across species within a lineage via phylogenetic regression is fundamental to comparative evolutionary biology, but when traits of interest are derived from two sets of lineages that co-evolve with one another, methods for evaluating such patterns in a dual-phylogenetic context remain underdeveloped. Here we extend multivariate permutation-based phylogenetic regression to evaluate trait correlations in two sets of interacting species while accounting for their respective phylogenies. This extension is appropriate for both univariate and multivariate response data, and may utilize one or more independent variables, including environmental covariates. Imperfect correspondence between species in the interacting lineages can also be accommodated, such as when species in one lineage associate with multiple species in the other, or when there are unmatched taxa in one or both lineages. For both univariate and multivariate data, the method displays appropriate type I error, and statistical power increases with the strength of the trait covariation and the number of species in the phylogeny. These properties are retained even when there is not a 1:1 correspondence between lineages. Finally, we demonstrate the approach by evaluating the evolutionary correlation between traits in fig species and traits in their agaonid wasp pollinators. R computer code is provided.

INTRODUCTION

Interactions between species, such as predator-prey, host-parasite, competition, and mutualism, have long fascinated ecologists and evolutionary biologists (Rosenzweig and MacArthur 1963; Anderson and May 1978; Tilman 1982; West et al. 2007). These interactions underlie processes responsible for adaptive phenotypic evolution in space and time and trait coevolution between species (Thompson 2005), as well as biological diversification, including patterns of co-diversification (Page 2003). While many studies have focused on interactions between species pairs, recent studies have emphasized that to gain insight into the ecological and evolutionary consequences of species interactions, we must account for the network of interactions that characterize communities of co-occurring species (e.g., Polis 1991; Strauss 1991; Bascompte et al. 2003). For example, species coexistence and the resilience of ecological networks are emergent properties of natural communities that may be poorly predicted from the independent analysis of component species pairs (Montoya et al. 2006; Amarasekare 2008; Thébault and Fontaine 2010).

While the network variables of interest are often ecological, evaluating such variables requires a phylogenetic perspective since the species interacting within communities possess shared evolutionary histories and are not statistically independent. Indeed, several procedures have been developed to model the evolution of community structure and the strength of associations among species that explicitly account for phylogenetic non-independence in a pair of associated lineages (Legendre et al. 2002; Henry et al. 2013; Rafferty and Ives 2013; Hadfield et al. 2014). In certain circumstances, the effect of trait values from one phylogenetically related set of species on the strength of ecological interaction with another set of species can also be incorporated (Rafferty and Ives 2013; see also Hadfield et al. 2014). Utilizing such approaches, researchers have been able to elucidate the extent to which species associations, incidence rates, and community

structure have been affected by evolutionary history (Longdon et al. 2014; Calatayud et al. 2016). Likewise, some methods model phenotypic evolution of continuous characters in one lineage while taking into consideration changes in trait values in a second set of co-evolving taxa that interact with the focal clade (Manceau et al. 2017). However, the latter approach does not directly evaluate the correlation between traits across the two co-evolving lineages.

As with the evaluation of ecological variables across a community, phylogenetic non-independence of taxa must be accounted for in evaluating the correlation in two sets of trait values across species, a problem that is addressed using phylogenetic comparative methods such as phylogenetic independent contrasts (Felsenstein 1985) or phylogenetic generalized least squares (Martins and Hansen 1997). An important and related area of study is coevolution and the extent to which phenotypic traits covary between interacting sets of taxa. Biologically, it may be expected that traits will be correlated between sets of species, as changes in the traits of one species may exert selective pressures that result in changes in the traits of their interacting species, and vice versa (Heinrich and Raven 1972; but see Janzen 1980; Nuismer et al. 2010). Indeed, such patterns of coadaptation may be predicted to occur in the traits of plants and their pollinators (e.g., Weiblen 2004; Smith et al. 2008), predators and their prey (e.g., West et al. 1991; Abrams 2000), hosts and their parasites (e.g., Morand et al. 2000; Morand and Poulin 2003; Johnson et al. 2005), and other species interactions. However, deciphering the ecological and evolutionary significance of trait correlations in such systems is complicated by the fact that the species within each interacting lineage are not independent. As such, the phylogenetic history of each interacting lineage can influence patterns of trait variation among species within each group. Thus, identifying trait correlations between sets of interacting species requires a phylogenetic perspective (for discussion see Morand and Poulin 2003; Weiblen 2004).

Methods such as phylogenetic independent contrasts (PIC) and phylogenetic generalized

least squares (PGLS) condition the data on a single phylogeny to account for the lack of independence in trait values within that lineage. However, in the case of interacting lineages, one must account for the separate phylogenetic histories of each interacting lineage. A few approaches have attempted to account for multiple phylogenies while assessing trait correlations, either by obtaining predicted values for both traits using phylogenetic eigenvector regression (Morand and Poulin 2003), or by obtaining residuals for each trait using phylogenetic autocorrelation approaches (Weiblen 2004). Unfortunately, the methods upon which these two approaches are based have numerous analytical challenges. For instance, phylogenetic eigenvector regression displays unacceptably high type I error rates (Freckleton et al. 2011), has no underlying evolutionary model, and suffers other procedural difficulties (e.g., if all eigenvectors are used, the data predicted by the phylogeny is identical to the original data where phylogenetic relationships were ignored: see Rohlf 2001; Adams and Church 2011). Likewise, phylogenetic autocorrelation can result in multiple values that produce identical optimal fits of the data to the phylogeny, limiting the utility of the method and complicating biological interpretation (see Rohlf 2001). Because of these issues, an alternative and more general method for studying trait correlations in interacting species while accounting for their respective phylogenies is desired.

In this article, we propose a new statistical procedure for evaluating the covariation between traits for two sets of interacting species, while accounting for their respective phylogenies. The approach is based on a Brownian motion model of evolution, and is extended from one implementation of phylogenetic regression: multivariate permutation-based PGLS (Adams 2014). The method is general, may be used with either univariate or multivariate response data, and may utilize one or more independent variables, including environmental covariates. The method is also appropriate for datasets where there is not a 1:1 correspondence between species in the interacting lineages, such as when species in one lineage associate with multiple species in the other, or when

there are unmatched taxa in one lineage relative to another. Using computer simulations we show that the method displays appropriate type I error rates and high statistical power for both the univariate and multivariate cases, and retains these desirable properties even when there is not a 1:1 correspondence between lineages. We then provide an empirical example demonstrating the utility of the approach by examining the evolutionary correlation between phenotypic traits in fig species and phenotypic traits their agaonid wasp pollinators. Computer code written in R for implementing the procedure is also provided.

PHYLOGENETIC REGRESSION

Phylogenetic comparative methods evaluate trends of evolutionary covariation among traits by conditioning the data on the phylogeny under a particular model of evolution. Typically, a Brownian motion model is utilized (e.g., Felsenstein 1973, 1981), though other models, such as Ornstein-Uhlenbeck (Hansen 1997; Butler and King 2004) or early-burst (Blomberg et al. 2003) models may be used in some applications. Most implementations of phylogenetic regression are based on a generalized least squares (GLS) model of the form:

$$\mathbf{Y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \varepsilon \qquad 1$$

where $\mathbf{Y}$ is a $N \times p$ matrix of trait values for the N species across p dependent trait dimensions, and $\mathbf{X}$ is a $N \times k$ design matrix containing one or more independent variables. In this model, the residuals (ε) are not independent, but are normally distributed as $\mathcal{N}(0, \mathbf{C})$. Here, $\mathbf{C}$ is an $N \times N$ phylogenetic covariance matrix describing the expected covariance between species resulting from shared evolutionary history as characterized by the phylogeny (Rohlf 2001; Blomberg et al. 2003;

O'Meara et al. 2006).

To obtain parameter estimates for the model while conditioning the data on the phylogeny, several algebraic approaches are commonly utilized, and all yield identical parameter estimates when implemented properly (Garland and Ives 2000; Rohlf 2001; Blomberg et al. 2012). The first is based on phylogenetically independent contrasts (Felsenstein 1985). Here, *N-1* contrast scores are obtained for each trait in **X** and **Y**, found as the difference in trait values between the two descendent taxa for each node on the phylogeny divided by the square root of the sum of their branch lengths. Model parameters, minus the intercept, are then found using the contrast scores $\mathbf{X}_{pic}$ and $\mathbf{Y}_{pic}$ as:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}_{\boldsymbol{pic}}^{\boldsymbol{t}}\mathbf{X}_{\boldsymbol{pic}}\right)^{-\mathbf{1}}\mathbf{X}_{\boldsymbol{pic}}^{\boldsymbol{t}}\mathbf{Y}_{\boldsymbol{pic}} \qquad 2$$

Likelihood ratio tests or their derived parametric summary measures (e.g., F-ratios) are typically used to statistically evaluate the fit of the regression model while accounting for phylogeny (see discussion in Adams and Collyer 2017).

The second implementation, often called phylogenetic generalized least squares (PGLS), uses the standard algebraic solution for GLS models (Grafen 1989; Martins and Hansen 1997). This approach is generally more flexible than that of independent contrasts because non-Brownian motion models, as well as complex statistical designs that include multiple independent variables, environmental covariates, and categorical factors, are more easily accommodated using this implementation (Pennell and Harmon 2013). First, a column of ones is appended to **X** to include the intercept. Model parameters are then obtained using the GLS solution:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^{t}\ \mathbf{C}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{t}\ \mathbf{C}^{-1}\mathbf{Y} \quad 3$$

As above, likelihood ratio tests and parametric summary measures are typically used to statistically evaluate the fit of the regression model conditioned on the phylogeny.

A third implementation uses phylogenetic transformation (Garland and Ives 2000; Adams 2014). Here, a phylogenetic transformation matrix is obtained as $\mathbf{P} = \left(\mathbf{U}\mathbf{W}^{-1/2}\mathbf{U}^{t}\right)^{-1}$, where **U** and **W** are the eigenvalues and eigenvectors of the phylogenetic covariance matrix **C** (Garland and Ives 2000). Next, the independent variable, **X** (including a column of ones), and the dependent variables, **Y**, are transformed by **P** via projection: $\mathbf{X}_{phy} = \mathbf{PX}$ and $\mathbf{Y}_{phy} = \mathbf{PY}$. This results in data conditioned on the phylogeny and whose error is now independent of the phylogeny as expressed in **C**. Model parameters are then estimated as:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}_{\boldsymbol{phy}}^{\boldsymbol{t}}\mathbf{X}_{\boldsymbol{phy}}\right)^{-1}\mathbf{X}_{\boldsymbol{phy}}^{\boldsymbol{t}}\mathbf{Y}_{\boldsymbol{phy}} \quad 4$$

Statistical evaluation of the fit of these models is accomplished using likelihood ratio tests (as above), or by generating empirical sampling distributions of test statistics through permutation procedures (e.g., Adams 2014; Adams and Collyer 2015).

Finally, a phylogenetic generalized linear mixed model (P-GLMM) has been proposed as an alternative approach to model the co-evolutionary relationship between traits while accounting for non-independence due to phylogeny (Lynch 1991; see also Housworth et al. 2004; Hadfield and Nakagawa 2010; de Villemereuil and Nakagawa 2014). This method is derived from the 'animal model' in quantitative genetics, but instead of using an animal pedigree to account for the relationships among individuals, the species phylogeny is used to account for the relationships

among taxa. The model is of the form:

$$\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} + \varepsilon \qquad 5$$

With this model, the set of independent variables and their parameters ($\mathbf{X}\hat{\boldsymbol{\beta}}$) are treated as fixed effects, while a phylogenetic random effect, described by the phylogenetic relatedness matrix and its effect parameters ($\mathbf{Z}\hat{\mathbf{u}}$) is also included. Note that this formulation differs somewhat from that of the GLS model, where phylogenetic non-independence is incorporated in the error structure using the phylogenetic covariance matrix **C**. With P-GLMM, model parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ are estimated using restricted maximum likelihood or MCMC approaches, and likelihood ratio tests, posterior probability distributions, or related measures are used to statistically evaluate the model. For bivariate regression, P-GLMM yields results similar to those obtained from GLS applications (see e.g., Housworth et al. 2004).

PHYLOGENETIC REGRESSION ON TWO PHYLOGENIES

Phylogenetic regression estimates the covariation between traits while conditioning the data on a single phylogeny under a specified model of evolution. This statistical approach is appropriate for most applications, as both the independent (**X**) and dependent (**Y**) variables are typically derived from the same set of species. However, when the traits are from two sets of species that co-evolve with one another, this is not the case, as one set of traits (**X**) is derived from one lineage, while the other set of traits (**Y**) is derived from a second lineage. Thus, in this scenario, properly accounting for species non-independence requires conditioning the data on two phylogenies in some manner. To accomplish this one might consider a P-GLMM approach, where

a second random effect is added to the model to incorporate the second phylogeny. However, this conditions the response (**Y**) data on *both* phylogenies simultaneously, which is the incorrect statistical model for the present application. Instead, the independent (**X**) variables should be conditioned on the phylogeny for their species, while the dependent (**Y**) variables are conditioned on the phylogeny for their species. Only by conditioning each data set on its respective phylogeny can one properly account for the expected covaration among species trait values for both the **X** and **Y** variables. Accomplishing this requires a different implementation.

As a solution to this problem, we propose an extension of the multivariate permutation-based PGLS procedure (Adams 2014; Adams and Collyer 2015). In the description that follows and for ease of presentation, we will refer to one set of interacting species as the plants and the second set of interacting species as the pollinators, though we recognize that other types of interacting species could be evaluated. First, trait values for one or more independent variables are obtained from the plant species, which may include measures of plant phenotypes, environmental covariates, or factors describing categories to which the plant species belong (e.g., upland versus lowland). All independent variables are assembled column-wise in a design matrix **X** containing N_X rows (categorical variables are represented as a series of dummy variables). A column of ones is included in **X** so the model intercept may also be estimated. Likewise, the continuous trait values for one or more dependent variables are obtained from the pollinator species, which are assembled column-wise in a response matrix **Y** containing N_Y rows. Finally, for both the plant and pollinator lineages, estimates of the phylogeny for the species must be available.

In addition to the data described above a species pairs matrix must be assembled, which specifies which pollinator is associated with which plant. For cases of a perfect 1:1 correspondence between plants and pollinators, each pollinator species will be associated with only one plant species, and these associations will be found on the $N_X = N_Y$ rows of the species pairs matrix.

However, a more realistic scenario is when there is an imperfect match between pollinators and plants. Indeed, having multiple connections between species is common in nature (Bascompte et al. 2003; Jordano et al. 2003), for instance when there is or more than one plant per pollinator, or when there is more than one pollinator per plant. Other mechanisms generating imperfect pairings include horizontal transmission of symbionts (e.g., parasites) to new hosts, which increases host breadth (Henry et al. 2013) or leads to the accumulation of symbionts on hosts (Stireman III et al. 2005). To accommodate such scenarios, species that have multiple associations with interacting species will be represented on multiple rows of the species pairs matrix, corresponding to those species with which they associate in nature. In these cases, the number of rows in the species pairs matrix depends on the total number of associations between plants and pollinators, which may differ from both N_X and N_Y. Finally, it is biologically possible that there are unmatched species in the dataset, for instance when a plant species in the phylogeny does not have an associated pollinator. In such cases, the unmatched species are not paired with a species in the other lineage in the species pairs matrix.

With the data components above, the following operations are performed. First, phylogenetic transformation matrices are calculated from both the plant and pollinator phylogenies, termed $\mathbf{P}_X$ and $\mathbf{P}_Y$ respectively. Next, the dependent and independent variable matrices are transformed by their respective phylogenies as: $\mathbf{X}_{phy.t} = \mathbf{P}_X\mathbf{X}$ and $\mathbf{Y}_{phy.t} = \mathbf{P}_Y\mathbf{Y}$. Using the associations described in the species pairs matrix, $\mathbf{X}_{phy}$ and $\mathbf{Y}_{phy}$ are then assembled from the rows of $\mathbf{X}_{phy.t}$ and $\mathbf{Y}_{phy.t}$. These are identical to $\mathbf{X}_{phy.t}$ and $\mathbf{Y}_{phy.t}$ when there is a 1:1 correspondence between plants and pollinators, but for scenarios where there is more than one plant per pollinator or vice versa, $\mathbf{X}_{phy}$ and $\mathbf{Y}_{phy}$ will contain replicated values from $\mathbf{X}_{phy.t}$ and/or $\mathbf{Y}_{phy.t}$ in the rows corresponding to their multiple species associations. Then, parameter estimates for the model are calculated as:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}_{\boldsymbol{phy}}^{\boldsymbol{t}}\mathbf{X}_{\boldsymbol{phy}}\right)^{-\mathbf{1}}\mathbf{X}_{\boldsymbol{phy}}^{\boldsymbol{t}}\mathbf{Y}_{\boldsymbol{phy}} \quad 6$$

and predicted values are obtained as: $\widehat{\boldsymbol{Y}}_{\boldsymbol{X}} = \mathbf{X}_{phy}\widehat{\boldsymbol{\beta}}$.

Next, if **X** contains only a single variable, a reduced model ($\mathbf{X}_r$) containing only a column of ones is transformed by $\mathbf{P}_X$ and the above calculations repeated to obtain predicted values from this reduced model ($\widehat{\boldsymbol{Y}}_{\mathbf{1}}$). Variation explained by the model is then found from the trace of the outer-product of the difference in predicted values (Adams 2014; Adams and Collyer 2015):

$$\mathbf{SS}_{\boldsymbol{X}} = tr\left[\left(\widehat{\boldsymbol{Y}}_{\boldsymbol{X}} - \widehat{\boldsymbol{Y}}_{\mathbf{1}}\right)\left(\widehat{\boldsymbol{Y}}_{\boldsymbol{X}} - \widehat{\boldsymbol{Y}}_{\mathbf{1}}\right)^{\boldsymbol{t}}\right] \quad 7$$

On the other hand, if the design matrix **X** contains more than one factor, a series of reduced model matrices are generated by sequentially removing terms from the original design matrix. These matrices are then transformed by $\mathbf{P}_\mathbf{X}$, and the above procedure is used in sequential fashion to obtain the sums-of-squares for each term in the model, from which the corresponding F-ratio and R^2 is obtained (described in Collyer et al. 2015; see also Anderson 2001). Finally, the significance of each model term is evaluated by generating an empirical sampling distribution of F-statistics using permutation. Here the phylogenetic transformation by $\mathbf{P}_\mathbf{Y}$ is performed, the reduced model is fit and residuals from this reduced model are permuted, and SS, MS, and F-values are obtained. The proportion of permuted values greater than the observed is used as an estimate of significance (see Adams 2014; Adams and Collyer 2015). It should be noted that when the phylogeny for the plants and pollinators is the same, the procedure above yields identical parameter estimates and summary test measures to the original permutation-based PGLS procedure (Supplemental

Material). Computer code in R for implementing the entire procedure is found in the Appendix.

STATISTICAL PERFORMANCE

To evaluate the statistical performance of the approach described above, a series of computer simulations were conducted. Simulations were performed under a Brownian motion model of evolution, and evaluated four different scenarios: 1) a bivariate regression with a single **Y** variable and a 1:1 correspondence among taxa, 2) a multivariate regression with a **Y** matrix (p = 9) and a 1:1 correspondence among taxa, 3) a bivariate regression with a single **Y** variable and an imperfect correspondence among taxa, and 4) a multivariate regression with a **Y** matrix (p = 9) and an imperfect correspondence among taxa. The simulations utilized random phylogenies for both the plants and their pollinators, which were generated using a random-splits model. Simulations with a 1:1 perfect plant-pollinator correspondence were conducted at four different levels of species richness (N = 16, 32, 64, 128), and each pollinator species was associated with exactly one plant species. For scenarios with an imperfect correspondence, plant phylogenies were simulated with species richness of (N_X = 16, 32, 64, 128) while pollinator phylogenies contained 25% fewer species (N_X = 12, 24, 48, 96). For species matching, each plant was associated with one pollinator species, however, because there were fewer pollinator than plant species, a third of the pollinators were associated with two plant species. Thus, these simulations evaluated scenarios of imperfect matching with multiple associations for some of the species.

For each simulation scenario above we evaluated both type I error rates and statistical power. Simulations evaluating type I error rates assumed no relationship between **X** and **Y**, and thus used no initial covariation between traits (σ_{XY} = 0.0). By contrast, simulations evaluating statistical power assumed a positive relationship between **X** and **Y**, and thus used positive initial

levels of covariation between traits. The degree of covariation between **X** and **Y** varied depending upon the desired strength of the **Y~X** relationship (σ_{XY} = 0.2, 0.4, 0.6, 0.8, 0.99).

To simulate the data, an initial input covariance matrix (**S**) was constructed, with the variance of each trait as (σ = 1.0) and the covariance between traits chosen from the values above (depending on simulation condition). For bivariate regressions **S** was a 2 × 2 matrix while for multivariate regressions **S** was a 10 × 10 matrix. Data were drawn from a normal distribution, $N(0,\mathbf{S})$. Random phylogenies were generated, and the data were back-transformed to each using the inverse of their respective phylogenetic transformation matrices, $\mathbf{P}^{-1}$. This procedure was utilized, rather than simulating along the phylogeny directly, because the data were meant to come from two distinct phylogenies. However, it can be shown that simulating data in this manner yields regression parameters identical to those obtained from the data back-transformed by the phylogeny, and then evaluated using PGLS (results not presented). For each simulation condition, 1000 datasets were generated in this manner, and for all datasets, the relationship between **Y** and **X** was evaluated using the new phylogenetic regression approach for two phylogenies described above. The proportion of significant results (out of 1000) was then treated as an estimate of the Type I error (when $\sigma_{XY} = 0.0$) or statistical power (when $\sigma_{XY} > 0.0$) of both approaches.

Results: When there was a 1:1 correspondence between taxa, the approach displayed appropriate Type I error rates, near the nominal α = 0.05 (Fig. 1). This was the case for both univariate and multivariate response data. Additionally, the power of the test increased as the degree of covariation between **X** and **Y** increased, and as the dimensionality of **Y** increased. As expected, as the number of species in the phylogeny increased, statistical power also increased (Fig. 1A & B); a pattern frequently found with phylogenetic comparative methods (e.g., O'Meara et al. 2006; Adams 2014). Importantly, when non 1:1 scenarios were examined, all of these

statistical properties were maintained (Fig. 1C & D), revealing that the approach was robust to imperfect species associations between interacting lineages, such as when there is more than one plant species associated with a pollinator species. Overall, these simulations revealed that the approach is capable of detecting significant evolutionary associations between variables, while accounting for dual-species phylogenies, when such patterns are present.

A Biological Example

To illustrate the utility of the approach described above, we evaluated the degree of evolutionary association between phenotypic traits in species of fig and their wasp pollinators. Figs (genus *Ficus*, family Moraceae) are highly diverse (750-plus species) and widely distributed across tropical and subtropical habitats. Fig wasp pollinators (family Agaonidae, superfamily Chalcidoidea) are obligately associated with figs and typically have a one-to-one species specificity with their hosts (Weiblen 2002; Marussich and Machado 2007). A textbook example of a highly coevolved mutualism (Janzen 1979; Herre et al. 2008), previously inseminated pollinators are attracted to enclosed fig inflorescences (syconia) by species-specific volatile attractants. A wasp then enters the syconium through a small ostiole and provides pollination services to the plant while laying eggs into a subset of its many ovules, with each ovule supporting one developing wasp larva.

The data examined here are found in Weiblen (2004) and contain the mean style length for individual plant species and the mean ovipositor length for its associated pollinator species; pollinators oviposit through the style into ovules and so a correlation between style and ovipositor length across species would be indicative of co-adaptation. The phylogenetic relationships for the plants are from Weiblen (2000), and are represented by a strict consensus tree of 208 trees obtained from a parsimony analysis of the ITS region of ribosomal DNA (see Weiblen 2000). The

phylogenetic relationships for the pollinators are represented by the single most parsimonious tree obtained from an analysis of cytochrome oxidase I (Weiblen 2001). Phenotypic traits from Weiblen 2004) were associated with their respective taxa in these phylogenies, resulting in a dataset of 39 plant-pollinator species pairs (Table 2 of Weiblen 2004). Plant-pollinator associations, along with their respective phylogenies are found in Fig. 2A. The evolutionary correlation of style and ovipositor length across species while accounting for both the phylogeneis of the plants and the pollinators was accomplished using the multivariate permutation-based PGLS procedure described above. All analyses were performed in R 3.4.1 (R Development Core Team 2017) using routines written by the senior author and found in the Appendix (all data and R scripts used for the analysis may be obtained from DRYAD: http://dx.doi.org/10.5061/dryad.r3765).

Results: Using analyses that did not account for phylogenetic associations, there was a significant relationship between species means for style length and ovipositor length ($\beta = 0.735$, $P < 0.0001$, $r = 0.899$: Fig. 2B). When the phylogenies of both plants and pollinators were taken into consideration, this relationship was still highly significant ($\beta = 0.577$, $P < 0.0001$, $r = 0.548$: Fig. 2C), though it was revealed that the original analysis overestimated the biological association (r) between style and ovipositor length by approximately 60%. Because the latter analysis properly accounted for both the plant and pollinator phylogeny in the analysis, we conclude that the covariation between style length and ovipositor length cannot be explained by the lack of independence of species due to their evolutionary histories alone. This suggests that the co-evolution between ovipositor length and style length displays an adaptive signature in this system.

DISCUSSION

The development of robust methods accounting for the phylogenetic non-independence of taxa in evaluating the correlation in two sets of trait values across species within a lineage has been

an active area of evolutionary research. In contrast, methods accounting for phylogeny in evaluating the correlation in trait values between two interacting lineages are underdeveloped, and the few approaches proposed have recognized analytical challenges. To address these issues we extended multivariate permutation-based phylogenetic regression approaches to allow for the assessment of trait correlations across two sets of interacting species while accounting for their respective phylogenies. The approach is general and may be used to evaluate either univariate or multivariate response data, and utilize one or more independent variables, including environmental covariates. Additionally, the method may be used with datasets where there is a 1:1 correspondence between species in the interacting lineages, as well as for scenarios with imperfect correspondence between species in the interacting lineages, such as when there is more than one parasite per host (or vice versa), or when there are unmatched taxa in one lineage relative to another. Using computer simulations we demonstrated that the approach displays appropriate type I error rates and high statistical power for both the univariate and multivariate cases, and retains these desirable properties even when there is not a 1:1 correspondence between lineages (Fig. 1). A particularly relevant application of this method is to evaluate the correlation between traits modulating species interactions between obligately-associated lineages. We provide such an example, identifying significant positive covariation between style length in fig flowers and the ovipositor length of pollinating fig wasps that insert their ovipositors the length of fig styles in the egg-laying process. This example thus demonstrates the utility of our procedure for identifying evolutionary correlations between traits in species from two co-evolving lineages.

The interactions between species, such as predator-prey relationships, interspecific competition, and mutualism, have long attracted the attention of ecologists and evolutionary biologists, as such interactions underlie processes responsible for adaptive phenotypic evolution in space and time. Indeed, interactions between species are considered to result in the co-evolution

of phenotypic traits, and considerable effort has been devoted to documenting the consequences of such interactions (Ehrlich and Raven 1964; Anderson and May 1978; Weiblen 2004). Recent phylogenetic approaches have furthered our understanding of how such traits evolve, by providing models that describe evolutionary changes in traits of a focal clade while considering interactions between those species (Nuismer and Harmon 2015) or conditioning on the trait values in a second set of co-evolving taxa that interact with them (Drury et al. 2016; Manceau et al. 2017). Our method provides a useful complement to these approaches, by affording a means by which the co-evolution of traits associated with such interactions may be evaluated using a dual-phylogenetic perspective. Through an examination of traits in fig species and traits in their agaonid wasp pollinators, we demonstrated the method's utility for illuminating co-evolutionary trends in mutualistic interactions (Fig. 2). However, we anticipate that our approach can provide equally useful insights into the evolution of trait correlations in other systems. For example, our procedure may shed light on trait changes observed in evolutionary arms races as characterized by some predator-prey interactions (e.g., Brodie and Jr 1991; Brodie et al. 2003), or the phenotypic responses of competition between clades (Silvestro et al. 2015). Characterizing evolutionary patterns of trait covariation in such systems would provide further quantitative evidence on the patterns of trait divergence, directional (runaway) trends, and phenotypic matching across species (sensu Nuismer and Harmon 2015), thereby enabling deeper insight into the macroevolutionary patterns that result from processes related to species interactions.

One possible extension of our approach is to consider the evolution of trait correlations at the population level. As with interspecific interactions, pairs of species interact locally within communities across the landscape. Therefore, how these interactions affect trait values among populations may be of interest, and it logically follows that accounting for the non-independence among populations is required when evaluating such trait correlations. However, in these cases, a

phylogeny is not the appropriate representation of the lack of independence among populations, as gene flow occurs across multiple populations in a network (Felsenstein 2002; Dyer and Nason 2004). Rather, a migration matrix among populations (**M**) describes the lack of independence across populations (see Felsenstein 2002; Stone et al. 2011). Thus, obtaining the migration matrices among populations (**M**) and using them in place of phylogenetic covariance matrices **C** in the procedure above, one may evaluate the correlation among traits between two sets of interacting species at the population level.

In conclusion, the method presented here of phylogenetic regression across two phylogenies adds to the suite of available approaches that explicitly account for phylogenetic non-independence in a pair of associated lineages. Specifically, it provides a robust procedure for evaluating the degree of trait covariation between lineages while accounting for their respective phylogenies. This new approach complements existing procedures that evaluate patterns of co-diversification of communities of co-evolving taxa (Rafferty and Ives 2013; Hadfield et al. 2014), as well as recent approaches of phenotypic matching across such lineages (Manceau et al. 2017). Through extension, we show that our approach also benefits developing methods for understanding how patterns of trait variation accumulate over macro- and microevolutionary timescales.

REFERENCES

Abrams, P. A. 2000. The evolution of predator-prey interactions: theory and practice. Ann. Rev. Ecol. Evol. Syst. 31:79-105.

Adams, D. C. 2014. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675-2688.

Adams, D. C. and J. O. Church. 2011. The evolution of large-scale body size clines in Plethodon: evidence of heat-balance or species-specific artifact? Ecography 34:1067-1075.

Adams, D. C. and M. L. Collyer. 2015. Permutation tests for phylogenetic comparative analyses of high-dimensional shape data: what you shuffle matters. Evolution 69:823-829.

Adams, D. C. and M. L. Collyer. 2017. Multivariate phylogenetic comparative methods: Evalutions, comparisons, and recommendations. Syst. Biol. In Press: DOI:10.1093/sysbio/syx1055.

Amarasekare, P. 2008. Spatial dynamics of foodwebs. Ann. Rev. Ecol. Syst. 39:479-500.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26:32-46.

Anderson, R. M. and R. M. May. 1978. Regulation and stability of host-parasite population interactions: I. Regulatory processes. J. Anim. Ecol. 47:219-247.

Bascompte, J., P. Jordano, C. J. Melián, and L. M. Olesen. 2003. The nested 348 assembly of plant–animal mutualistic networks. Proc. Natl. Acad. Sci. U.S.A. 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.

Brodie, E. D. and E. D. B. Jr. 1991. Evolutionary response of predators to dangerous prey: Reduction of toxicity of newts and resistance of garter snakes in island populations. Evolution 45:221-224.

Brodie, E. D., B. Ridenhour, and E. Brodie. 2003. The evolutionary response of predators to dangerous prey: Hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. Evolution 56:2067–2082.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164:683-695.

Calatayud, J., J. L. Hórreo, J. Madrigal-González, A. Migeon, M. Rodríguez, S. Magalhaes, and J. Hortal. 2016. Geographay and major host evolutionary transitions shape the resource use of plant parasites. Proc. Natl. Acad. Sci. U.S.A. 113:9840-9845.

Collyer, M. L., D. J. Sekora, and D. C. Adams. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. Heredity 115:357-365.

de Villemereuil, P. and S. Nakagawa. 2014. General quantitative genetic methods for comparative biology. Pp. 287-304 *in* L. Z. Garamszegi, ed. Modern phylogenetic comparative methods and their application in evolutionary biology. Springer-Verlag, Berlin.

Drury, J., J. Clavel, M. Manceau, and H. Morlon. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. Syst. Biol. 65:700–710.

Dyer, R. J. and J. D. Nason. 2004. Population graphs: the graph theoretic shape of genetic structure. Molec. Ecol. 13:1713-1727.

Ehrlich, P. R. and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586–608.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Americal Journal of Human Genetics 25:471-492.

Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35:1229-1242.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1-15.

Felsenstein, J. 2002. Contrasts for a within-species comparative method. Pp. 118–129 *in* M. Slatkin, and M. Veuille, eds. Modern developments in theoretical population genetics Oxford University Press, Oxford, UK.

Freckleton, R. P., N. Cooper, and W. Jetz. 2011. Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. Am. Nat. 178:E10-E17.

Garland, T. J. and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346-364.

Grafen, A. 1989. The phylogenetic regression. Phil. Trans. Roy. Soc. London B. 326:119-157.

Hadfield, J. D., B. R. Krasnov, R. Poulin, and S. Nakagawa. 2014. A tale of two phylogenies: Comparative analysis of ecological interactions. Am. Nat. 183:174-187.

Hadfield, J. D. and S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J. Evol. Biol. 23:494-508.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341-1351.

Heinrich, B. and P. H. Raven. 1972. Energetics and pollination ecology. Science 176:597-602.

Henry, L. M., J. Peccoud, J. C. Simon, J. Hadfield, M. J. C. Maiden, J. H. Ferrari, and H. C. J. Godfray. 2013. Horizontally transmitted symbionts and host colonization of ecological niches. Curr. Biol. 23:1713–1717.

Herre, E. A., K. C. Jandér, and C. A. Machado. 2008. Evolutionary ecology of figs and their associates: Recent progress and outstanding puzzles. Ann. Rev. Ecol. Evol. Syst. 39:439–458.

Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. Am. Nat. 163:84-96.

Janzen, D. H. 1979. How to be a fig. Ann. Rev. Ecol. Syst. 10:13-51.

Janzen, D. H. 1980. When is it coevolution? Evolution 34:611-612.

Johnson, K. P., S. E. Bush, and D. H. Clayton. 2005. Correlated evolution of host and parasite body size: tests of Harrison's fule using birds and lice. Evolution 59:1744-1753.

Jordano, P., J. Bascompte, and J. M. Olesen. 2003. Invariant properties in coevolutionary networks of plant–animal interactions. Ecol. Lett. 6:69-81.

Legendre, P., Y. Desdevises, and E. Bazin. 2002. A statistical test for host-parasite coevolution. Syst. Biol. 51:217–234.

Longdon, B., M. A. Brockhurst, C. A. Russell, J. J. Welch, and F. M. Jiggins. 2014. The evolution and genetics of virus host shifts. PLoS Pathog. 10:e1004395.

Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065-1080.

Manceau, M., A. Lambert, and H. Morlon. 2017. A unifying comparative phylogenetic framework including traits co-evolving acxross interacting lineages. Syst. Biol. 66:551–568.

Martins, E. P. and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646-667.

Marussich, W. A. and C. A. Machado. 2007. Host-specificity and coevolution among pollinating and nonpollinating New World fig wasps. Molec. Ecol. 16:1925-1946.

Montoya, J. M., S. L. Pimm, and R. V. Solé. 2006. Ecological networks and their fragility. Nature 442:259-264.

Morand, S., M. S. Hafner, R. D. M. Page, and D. L. Reed. 2000. Comparative body size relationships in pocket gophers and their chewing lice. Biol. J. Linn. Soc. 70:239-249.

Morand, S. and R. Poulin. 2003. Phylogenies, the comparative metehod and parasite evolutionary ecology. Adv. Parasitol. 54:281-302.

Nuismer, S. L., R. Gomulkiewicz, and B. J. Ridenhour. 2010. When is correlation coevolution? Am. Nat. 175:525-537.

Nuismer, S. L. and L. J. Harmon. 2015. Predicting rates of interspecific interaction from phylogenetic trees

Ecol. Lett. 18:17–27.

O'Meara, B. C., C.Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922-933.

Page, R. D. M. 2003. Tangled trees: phylogeny, cospeciation, and coevolution. The University of Chicago Press, Chicago, IL.

Pennell, M. W. and L. J. Harmon. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. Ann. New York Acad. Sci. 1289:90-105.

Polis, G. A. 1991. Complex desert food webs: an empirical critique of food web theory. Am. Nat. 138:123-155.

R Core Team. 2017. R: a language and environment for statistical computing. Version 3.4.1. http://cran.R-project.org. R Foundation for Statistical Computing, Vienna.

Rafferty, N. E. and A. R. Ives. 2013. Phylogenetic trait-based analyses of ecological networks. Ecology 94:2321–2333.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55:2143-2160.

Rosenzweig, M. L. and R. H. MacArthur. 1963. Graphical representation and stability conditions of predator-prey interactions. Am. Nat. 97:209-223.

Silvestro, D., A. Antonelli, N. Salamin, and T. B. Quental. 2015. The role of clade competition in the diversification of North American canads. Proc. Natl. Acad. Sci. U.S.A. 112:8684-8689.

Smith, S. D., C. Ane, and D. A. Baum. 2008. The role of pollinator shifts in the floral diversification of *lochroma* (Solanaceae). Evolution 62:793-806.

Stireman III, J. O., J. D. Nason, and S. B. Heard. 2005. Host-associated genetic differentiation in phytophagous insects: general phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. Evolution 59:2573-2587.

Stone, G. N., S. Nee, and J. Felsenstein. 2011. Controlling for non-independence in comparative analysis of patterns across populations within species. Phil. Trans. Roy. Soc. 366:1410-1424.

Strauss, S. Y. 1991. Indirect effects in community ecology: their definition, study and importance. Trends Ecol. Evol. 6:206-210.

Thébault, E. and C. Fontaine. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853-856.

Thompson, J. N. 2005. The geographic mosaic of coevolution. The University of Chicago Press, Chicago, IL.

Tilman, D. 1982. Resource competition and community structure. Princeton University Press, Princeton, NJ.

Weiblen, G. D. 2000. Phylogenetic relationships of functionally dioecious Ficus (Moraceae) based on ribosomal DNA sequences and morphology. Am. J. Bot. 87:1342-1357.

Weiblen, G. D. 2001. Phylogenetic relationships of fig wasps pollinating functionally dioecious figs based on mitochondrial DNA sequences and morphology. Syst. Biol. 50:243-267.

Weiblen, G. D. 2002. How to be a fig wasp. Ann. Rev. Entomol. 47:299–330.

Weiblen, G. D. 2004. Correlated evolution in fig pollination. Syst. Biol. 53:128-139.

West, K., A. Cohen, and M. Brown. 1991. Morphology and behavior of crabs and gastropods from Lake Tanganyika, Africa: implications for lacustrine predator-prey coevolution. Evolution 45:589-607.

West, S. A., A. S. Griffin, and A. Gardner. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20:415-432.

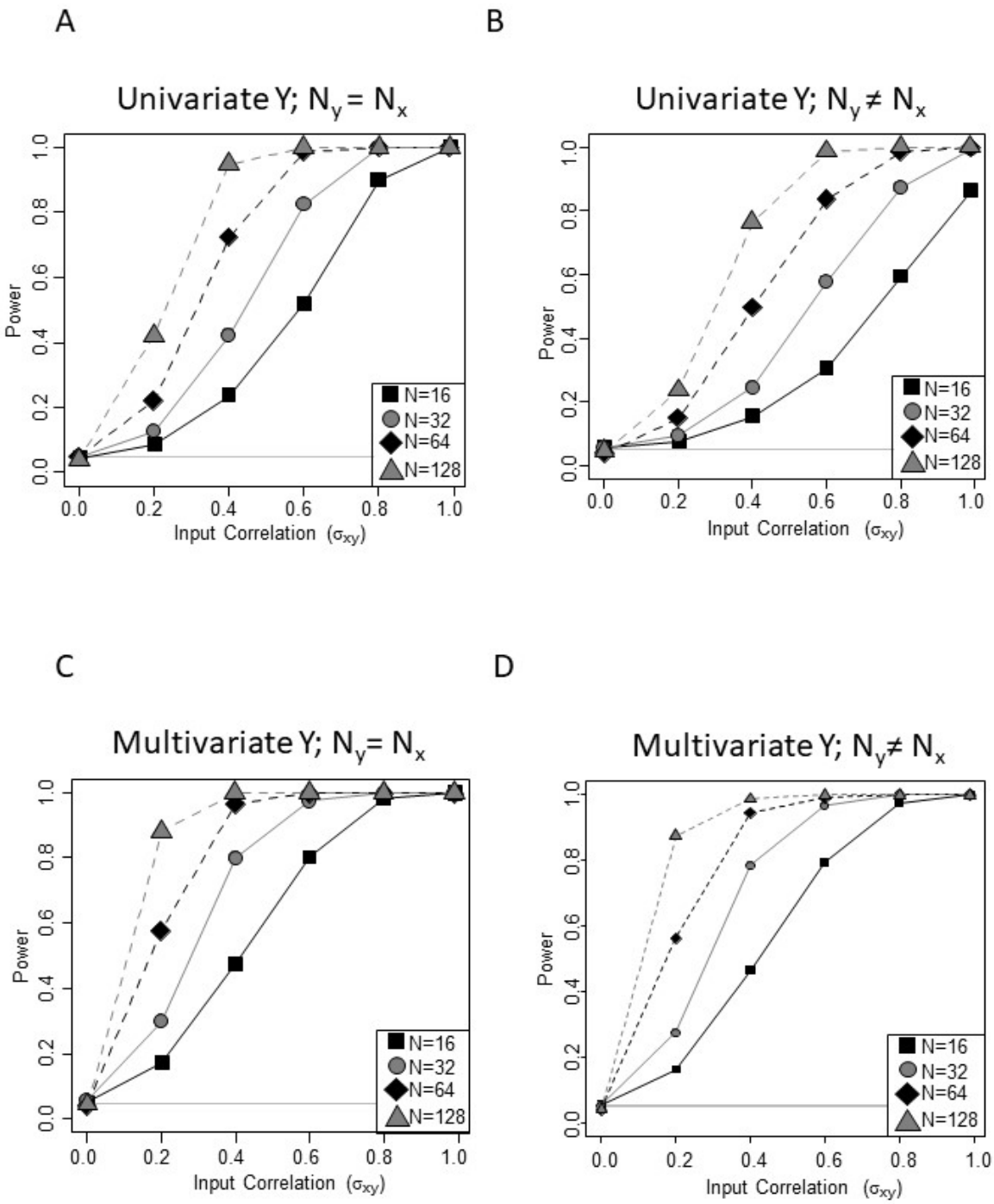
A
Univariate Y; $N_y = N_x$
B
Univariate Y; $N_y \neq N_x$
C
Multivariate Y; $N_y = N_x$
D
Multivariate Y; $N_y \neq N_x$
Power
Input Correlation (σ_{xy})
N=16
N=32
N=64
N=128
0.0
0.2
0.4
0.6
0.8
1.0

FIGURE. 1. Simulation results evaluating the Type I error (when $\sigma_{XY} = 0.0$) and statistical power (when $\sigma_{XY} > 0.0$) of the phylogenetic regression approach developed here for evaluating trait correlation between two lineages of interacting species under the following conditions: A) univariate response data with a 1:1 correspondence among taxa, B) multivariate response data with a 1:1 correspondence among taxa, C) univariate response data with an imperfect correspondence among taxa, and D) multivariate response data with an imperfect correspondence among taxa. For all scenarios, curves for increasing numbers of species in the phylogeny are shown.

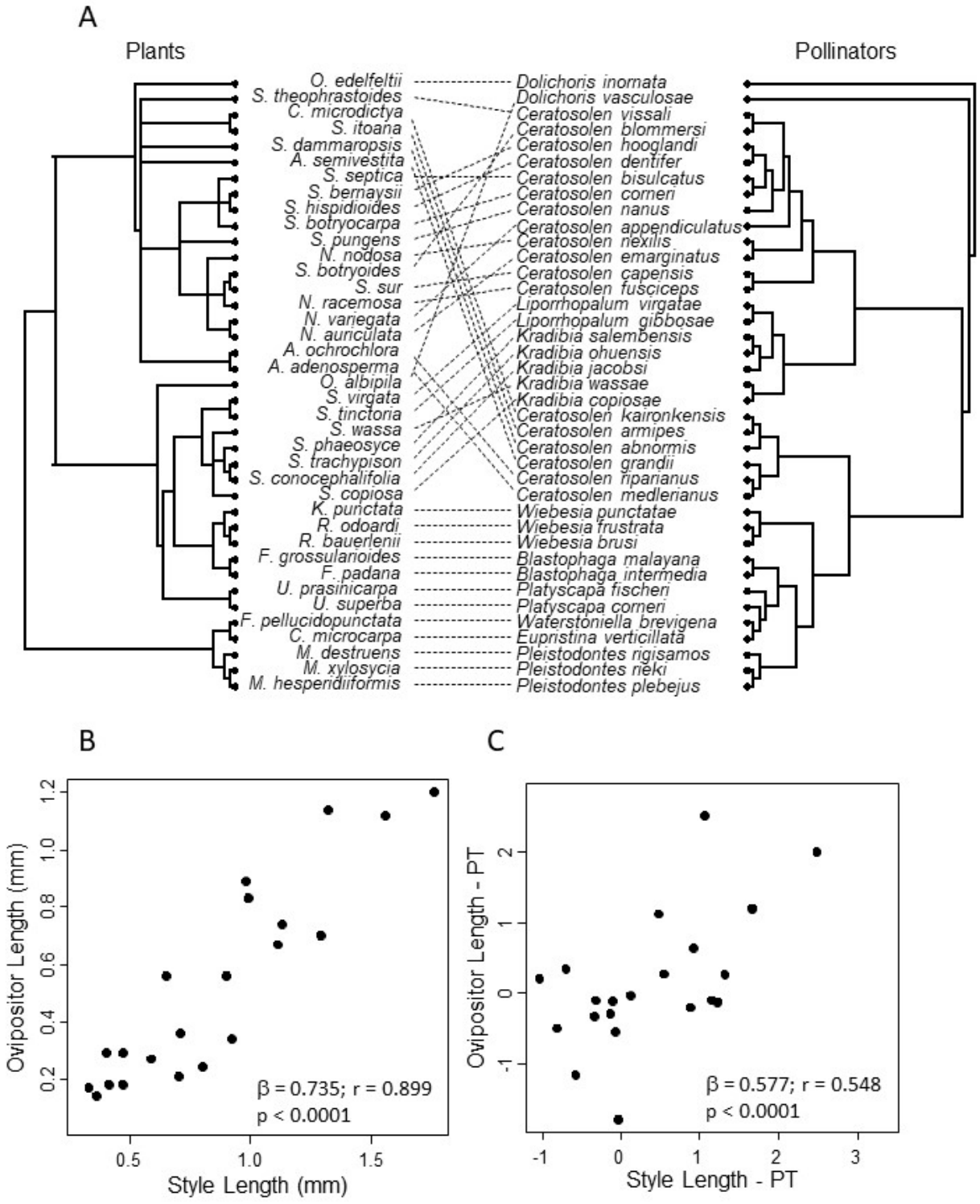

FIGURE 2. A) Phylogenetic relationships of figs and their pollinators, with the plant-pollinator

associations indicated (adapted from Weiblen 2004). B) Relationship between style length and ovipositor length for original data from 39 pairs of associated plant and pollinator species, and C) relationship between the same traits for the same 39 species pairs as viewed using the phylogenetically-transformed (PT) variables. In C), the association between traits was evaluated using the procedure developed here (see text).

APPENDIX

COMPUTER CODE FOR R

```
#Co-evolution of traits for two interacting lineages

#CoPhy.pgls: Function performs a phylogenetic regression for two
traits (x & Y), obtained from two co-evolving lineages (e.g., hosts
#and their parasites). The user provides a formula describing the
#linear model to be examined (e.g., y~x1+x2), where the dependent
#variables (Y) are derived from one set of interacting species (e.g.,
#parasites) while the independent variables are derived from a second
#set interacting species (e.g., hosts). Phylogenetic transformation of
#both X and Y is performed and the model is statistically evaluated
#using multivariate permutation PGLS procedures (see Adams 2014;
#Evolution; Adams and Collyer 2015; Evolution).

#The method is capable of accommodating perfect 1:1 correspondence
#between X and Y taxa, in which case each species in X is matched to
#one (and only one) species in Y.  Additionally, patterns in data with
#imperfect species correspondence may also be evaluated: where a
#parasite inhabits more than one host; where more than one parasite is
#found on a single host, or cases where some host or parasites are not
#matched to a species in the other lineage (see below for details).

#Parameters
# @param f1 - A formula for the linear model (e.g., y ~ x1 + x2)

# @param dataX - A list containing at least two elements: 1: the
  #phylogeny (of class 'phylo') for the species represented by the
  #independent (X) variables, and 2: one or more vectors containing
  #the values for the independent (X) variables, whose element names
  #correspond to the species names in the X-phylogeny.  The name of
  #each vector must correspond to the name of one of the independent
  #variables as listed in f1 (e.g., x1, x2, etc.).

# @param dataY - A list containing two elements: 1: the phylogeny (of
  #class 'phylo') for the species represented by the dependent (Y)
  #variables, and 2: a vector (or matrix) containing the dependent (Y)
  #variables, whose names/rownames correspond to the species names in
  #the Y-phylogeny. The name of this matrix must correspond to the
  #name of the independent variable as listed in f1 (e.g., Y).

# @param matchlist - A matrix describing which species for X
  #correspond to species for Y, in X-Y order, i.e., the dependent (X)
  #variable species are first, followed by the independent (Y)
  #species. If more than one 'host' (X-species) contains more than one
  #'parasite' (Y-species), that species' name is repeated in the
  #matchlist at the corresponding locations for its multiple
  #associations.  For the converse scenario (where a 'parasite' is
  #found on more than one host, its name is repeated in the matchlist
```

```
  #in the appropriate locations. Finally, for any species that is not
  #matched with a species in the other lineage, 'NA' is used to
  #designate its 'match' in the second lineage.

# @param iter - The number of iterations for significance testing

###############
CoPhy.pgls<-function(f1,dataX=NULL,dataY=NULL, matchlist, iter=999){
  library(ape)
  Pmat<-function(phy,x){
    C <- vcv.phylo(phy, anc.nodes = FALSE)
    C <- C[rownames(x), rownames(x)]
    eigC <- eigen(C)
    lambda <- zapsmall(eigC$values)
    if (any(lambda == 0)) {   lambda = lambda[lambda > 0]  }
    eigC.vect = eigC$vectors[, 1:(length(lambda))]
    Pmat<-eigC.vect %*% diag(sqrt(lambda)) %*% t(eigC.vect)
    Px<-if (det(Pmat) > 1e-08) qr.solve(Pmat) else fast.ginv(Pmat)
    rownames(Px) <- colnames(Px) <- colnames(C)
    Px
  }
  fast.ginv<-function (X, tol = sqrt(.Machine$double.eps)) {
    k <- ncol(X)
    Xsvd <- La.svd(X, k, k)
    Positive <- Xsvd$d > max(tol * Xsvd$d[1L], 0)
    rtu <- ((1/Xsvd$d[Positive]) * t(Xsvd$u[, Positive, drop =
FALSE]))
    v <- t(Xsvd$vt)[, Positive, drop = FALSE]
    v %*% rtu
  }
  pval<-function(s){
    p = length(s)
    r = rank(s)[1] - 1
    pv = 1 - r/p
    pv
  }
  if(is.null(dataX))stop("No dataX provided.")
  if(is.null(dataY))stop("No dataY provided.")
  NA.vals<-length(which(is.na(matchlist)==TRUE))
  if(NA.vals>0){
    NA.rows<-unlist(apply(matchlist,2,function (x) which(is.na(x))))
    matchlist<-matchlist[-NA.rows,]
  }
  for(i in 1:length(dataX)) assign(names(dataX)[i], dataX[[i]])
  checkX <- sapply(dataX, class)
  if(is.na(match("phylo",checkX)))stop("No phylogeny of class Phylo in
list 'dataX'.") else{
    phyX<-dataX[[match("phylo",checkX)]]
  }
  y.tmp<-rep(1,length(phyX$tip.label))
  tmp.f<-as.formula(paste("y.tmp", formula(f1)[3],sep="~"))
  x.mod<-model.matrix(terms(formula(tmp.f)))
```

```
  Terms<-terms(tmp.f)
  term.labels <- attr(Terms, "term.labels")
  X.k <- attr(x.mod, "assign")
  QRx <- qr(x.mod)
  x.mod <- x.mod[, QRx$pivot, drop = FALSE]
  x.mod <- x.mod[, 1:QRx$rank, drop = FALSE]
  X.k <- X.k[QRx$pivot][1:QRx$rank]
  uk <- unique(c(0,X.k))
  k <- length(attr(Terms, "term.labels"))
  Xs <- lapply(1:length(uk), function(j)  Xj <- x.mod[, X.k %in%
uk[1:j]])
  Xrs <- Xs[1:k]
  Xfs <- Xs[2:(k+1)]
  Px<-Pmat(phy = phyX,x = as.matrix(dataX[[which(checkX
!="phylo")[1]]]))
  nX<-nrow(Px)
  Xr <- lapply(Xrs, function(x) crossprod(Px, as.matrix(x)))
  Xf <- lapply(Xfs, function(x) crossprod(Px, as.matrix(x)))
  Xr.m<-lapply(Xr, function(x)
as.matrix(x[match(matchlist[,1],rownames(x)),]))
  Xf.m<-lapply(Xf, function(x)
as.matrix(x[match(matchlist[,1],rownames(x)),]))
  for(i in 1:length(dataY)) assign(names(dataY)[i], dataY[[i]])
  checkY <- sapply(dataY, class)
  if(is.na(match("phylo",checkY)))stop("No phylogeny of class Phylo in
list 'dataY'.") else{
    phyY<-dataY[[match("phylo",checkY)]]
  }
  Py<-Pmat(x = as.matrix(dataY[[which(checkY !="phylo")[1]]]),phy =
phyY)
  Y<-as.matrix(get(names(dataY)[which(checkY!="phylo")]))
  ind <- c(list(1:nrow(Y)), (Map(function(x) sample.int(nrow(Y),
nrow(Y)), 1:iter)))
  perms <- length(ind)
  Y.pr<-crossprod(Py,Y)
  Y.pr.m<-as.matrix(Y.pr[match(matchlist[,2],rownames(Y.pr)),])
  N.xy<-n<-nrow(Y.pr.m)
  p<-ncol(Y.pr.m[[1]])
  gls.init<-lapply(1:length(Xr.m), function(j)
lm.fit(Xr.m[[j]],Y.pr.m) )
  fitted<-lapply(1:length(Xr.m), function(j)
as.matrix(gls.init[[j]]$fitted.values))
  res<-lapply(1:length(Xr.m), function(j)
as.matrix(gls.init[[j]]$residuals))
  SS.p <- lapply(1: perms, function(j){
    Yi<-Map(function(f, r) f + r[ind[[j]], ], fitted, res)
    c(Map(function(y, ur, uf) sum(.lm.fit(ur,y)$residuals^2  -
.lm.fit(uf,y)$residuals^2 ),
          Yi, Xr.m, Xf.m),
      sum(.lm.fit(Xr.m[[k]],Yi[[k]])$residuals^2)-
sum(.lm.fit(Xr.m[[k]],Yi[[k]])$residuals^2  -
.lm.fit(Xf.m[[k]],Yi[[k]])$residuals^2 ),
```

```
      sum(.lm.fit(Xr.m[[1]],Yi[[1]])$residuals^2) )
  })
  SS.p <- matrix(unlist(SS.p), k+2, perms)
  SS <- SS.p[1:k,]
  SSE <- SS.p[k+1,]
  SSY <- SS.p[k+2,]
  df<-sapply(1:k, function(j) qr(Xf.m[[j]])$rank - qr(Xr.m[[j]])$rank)
  dfE <- n- sum(df) -1
  df <- c(df, dfE, n-1)
  MS <- SS/df[1:k]
  MSE <- SSE/df[k+1]
  SSE.mat <- matrix(SSE, k, length(SSE), byrow = TRUE)
  MSE.mat <- matrix(MSE, k, length(MSE), byrow = TRUE)
  if(is.matrix(SS)){
    Fs <- (SS/df[1:k])/MSE.mat
    P.val <- apply(Fs, 1, pval)
  } else {
    MSE <- SSE/df[2]
    Fs <- (SS/df[1])/MSE
    P.val <- pval(Fs)
  }
  SS.obs<-SS.p[,1]
  if(k==1){MS.obs <- c(MS[1],MSE[[1]],NA)} else MS.obs <-
c(MS[,1],MSE[[1]],NA)
  R2.obs <- c((SS.obs/SSY[[1]])[1:k],NA,NA)
  if(k==1){Fs <- c(Fs[1],NA,NA)} else Fs <- c(Fs[,1],NA,NA)
  Pr<-c(P.val,NA,NA)
  anova.tab <- data.frame(df,SS=SS.obs,MS=MS.obs,Rsq=R2.obs,F=Fs,
Pr=Pr)
  rownames(anova.tab) <- c(term.labels, "Residuals", "Total")
  b <- lm(Y.pr.m~Xf.m[[k]]-1)$coefficients
  out <- list(anova.table = anova.tab, reg.coef = b)
  out
}
```