

# A framework to study and predict functional trait syndromes using phylogenetic and environmental data

Pablo Sanchez-Martinez<sup>1,2,3</sup>  | David D. Ackerly<sup>4</sup>  | Jordi Martínez-Vilalta<sup>1,2</sup>  |  
Maurizio Mencuccini<sup>2,5</sup>  | Kyle G. Dexter<sup>3,6</sup>  | Todd E. Dawson<sup>4</sup> 

<sup>1</sup>Departament de Biologia Animal, Biologia Vegetal i Ecologia, Univ. Autònoma de Barcelona, Cerdanyola del Vallès, Spain; <sup>2</sup>CREAF, Cerdanyola del Vallès, Barcelona, Spain; <sup>3</sup>School of GeoSciences, University of Edinburgh, Edinburgh, UK; <sup>4</sup>Department of Integrative Biology, University of California, Berkeley, California, USA; <sup>5</sup>ICREA, Pg. Lluís Companys 23, Barcelona, Spain and <sup>6</sup>Royal Botanic Garden Edinburgh, Edinburgh, UK

## Correspondence

Pablo Sanchez-Martinez  
Email: [p.sanchez@creaf.uab.cat](mailto:p.sanchez@creaf.uab.cat)

## Funding information

Institució Catalana de Recerca i Estudis Avançats; Ministerio de Ciencia e Innovación, Grant/Award Number: CGL2017-89149-C2-1-R, PID2021-127452NB-I00 and FPU18/04945

Handling Editor: Karen Bacon

## Abstract

1. Traits do not evolve in isolation but often as part of integrated trait syndromes, yet the relative contributions of environmental effects and evolutionary history on traits and their correlations are not easily resolved.
2. In the present study, we develop a methodological framework to elucidate eco-evolutionary patterns in functional trait syndromes. We do so by separating the amount of variance and covariance related to phylogenetic heritage and environmental variables (*environmental phylogenetic conservatism*), only phylogenetic heritage (*non-attributed phylogenetic conservatism*) and only to environmental variables (*evolutionarily labile environmental effects*). Variance–covariance structures of trait syndromes are displayed as networks. We then use this framework to guide a newly derived imputation method based on machine learning models that predict trait values for unsampled taxa, considering environmental and phylogenetic information as well as trait covariation. *TrEvol* is presented as an R package providing a unified set of methodologies to study and predict multivariate trait patterns and improve our capacity to impute trait values.
3. To illustrate its use, we leverage both simulated data and species-level traits related to hydraulics and the leaf economics spectrum, in relation to an aridity index, demonstrating that most trait correlations can be attributed to *environmental phylogenetic conservatism*.
4. This conceptual framework can be employed to examine issues ranging from the evolution of trait adaptation at different phylogenetic depths to intraspecific trait variation.

## KEYWORDS

environmental effects, evolutionary lability, functional traits, imputation, phylogenetic conservatism, trait syndromes

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

## 1 | INTRODUCTION

Functional traits are defined as morpho-physio-phenological attributes that impact fitness via their effects on individual performance (Violle et al., 2007). As such, they are likely to undergo adaptive evolution in response to environmental drivers (Ackerly et al., 2000). The functional significance of any one trait depends on coordination with other traits, creating functional strategies referred to as trait syndromes that contribute to success under different environmental conditions (Reich et al., 2003; Sanchez-Martinez et al., 2020; Wright et al., 2004). A trait syndrome is a combination of functional trait values that occur in a given taxonomic unit (e.g. a species) and that can be shared with others (e.g. a group of species, which may or may not be closely related). Variability in functional syndromes can lead to the conformation along trait modules (also referred to as trait axes or trait spectra), which are sets of traits that covary, potentially in response to environmental conditions, and that may also be associated with evolutionary legacies (Cavender-Bares et al., 2016; Reich et al., 2003). While the relationships among some functional traits (e.g. Mencuccini et al., 2019; Wright et al., 2004), their relationship with environmental variables (e.g. Bhaskar & Ackerly, 2006; Flo et al., 2021) and trait phylogenetic conservatism (e.g. Ackerly, 2009; Flores et al., 2014) have been widely studied, to our knowledge, a unified framework to study trait syndromes from an eco-evolutionary perspective is still lacking. We posit that a phylogenetically explicit framework describing the multivariate structure of traits and their relationship with the environment will improve our capacity to understand the ecological and evolutionary nature of correlations underlying trait syndromes observed in nature. In turn, this framework may be used to predict plant functional trait values for unsampled taxa, which can help, for example, in determining vegetation responses to environmental changes (Anderegg et al., 2021; Choat et al., 2018; Sanchez-Martinez et al., 2023). The proposed predictive framework can be especially useful in the common situation where sparse trait data undermine the ability to make predictions in understudied locations or vegetation types (García-Valdés et al., 2021; Kattge et al., 2020; Trugman et al., 2020).

Phylogenetic conservatism and evolutionary lability appear in the literature as contrasted concepts describing patterns in functional traits that may arise under different evolutionary scenarios (Blomberg et al., 2003; Blomberg & Garland, 2002). Phylogenetic conservatism also referred as phylogenetic inertia or phylogenetic signal can appear under a scenario of lineage-specific stabilising selection, meaning that values of a given trait will be maintained and transmitted to descendants (Crisp & Cook, 2012). This scenario is compatible with a slow evolution of traits and the existence of strong environmental filtering determining the environmental space that species occupy (Losos, 2008). Under this scenario, phylogenetic conservatism in functional traits is expected to be related to environmental variables describing the species ecological niche. This is translated into a pattern of closely related species occupying similar ecological spaces and responding to the environment by means of similar functional strategies. This pattern of phylogenetic

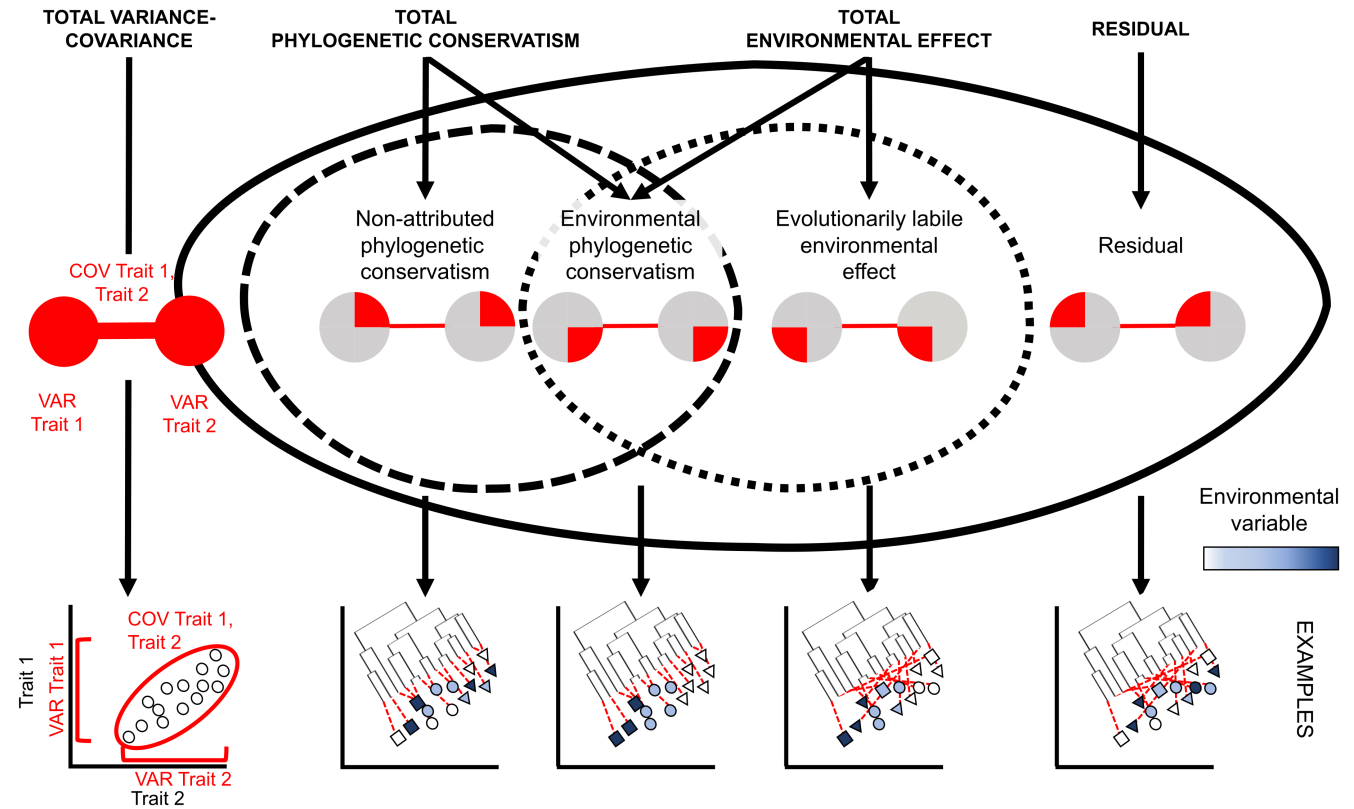
conservatism related to the environmental space that defines the ecological niche is referred to as phylogenetic niche conservatism (Crisp et al., 2009; Crisp & Cook, 2012; Losos, 2008). Because the ecological niche of a given taxonomic entity is usually complex and multivariate (Holt, 2009), quantifying its overall effect on functional syndromes is challenging, and here we take the simpler approach of using individual environmental variables to characterise phylogenetic niche conservatism. We refer to this pattern as *environmental phylogenetic conservatism* for a given trait or trait syndrome in response to a given environmental variable or environmental axis. Then, to characterise phylogenetic niche conservatism, multiple analyses may be required to define the multivariate niche.

Phylogenetic conservatism in trait variation and trait-trait correlations can also be related to evolutionary constraints that are not associated with measured environmental variables. These potentially non-adaptive processes can lead to a slow and constrained evolution as a result of genetic correlation, linkage disequilibrium, pleiotropy, lack of genetic variability or homogenising gene flow, among others (Ackerly, 2009). This conservatism may also be related to unmeasured environmental variables (e.g. soil effects in a study only considering climate). We refer to this conservatism in functional trait syndromes as the *non-attributed phylogenetic conservatism* (Figure 1).

Evolutionary lability is a pattern that may appear in a scenario where evolution is less constrained by ancestral values. This scenario is compatible with rapid evolution, which can be adaptive, leading to patterns such as adaptive radiation (Ackerly, 2009; Ackerly et al., 2000). Under this adaptive scenario, we expect traits to be related to environmental variables exerting selective pressures in a non-phylogenetically structured way, with closely related taxa presenting different strategies that allow them to survive under different conditions. We refer to this pattern as the *evolutionarily labile environmental effect* of a specific environmental variable on trait syndromes.

A better knowledge of functional trait correlations and their environmental and phylogenetic foundations can improve predictive power to infer unmeasured trait values. Previous methodologies allow one to predict missing values by using phylogenetic information (Swenson, 2014) and statistical covariation among traits and environmental variables (Maynard et al., 2022; Poyatos et al., 2018). However, in some cases, traits and their correlations are highly phylogenetically conserved, such that phylogeny is very informative in predicting missing values, while in other cases, traits present a more (evolutionarily) labile correlation in response to environmental variables. In the latter cases, environmental variables may be better predictors of missing functional trait values. In addition, the organisation of traits in modules (i.e. groups of tightly related traits) indicates that not all traits are equally informative for inferring the values of other traits. We propose an imputation method that implements a data-driven procedure to select which predictors to include in each case, based on trait covariation and relationships with phylogeny and environmental data.

The methodology described here uses phylogenetic mixed models to separate the contribution of *environmental phylogenetic*



**FIGURE 1** Conceptual framework. Variance and covariance partitioning into the different components related to *non-attributed phylogenetic conservatism*, *environmental phylogenetic conservatism*, *evolutionarily labile environmental effects* and *residual*. Note how *non-attributed phylogenetic conservatism* and *environmental phylogenetic conservatism* sum to total phylogenetic conservatism and *environmental phylogenetic conservatism* and *evolutionarily labile environmental effect* sum to total environmental effect. Examples representing two traits in each case are presented as networks (showing variance results as pie charts in nodes and correlation as edges). At the bottom, examples for different traits showing extreme cases maximising each one of the components are shown as scatterplots, representing two different hypothetical traits in each case and the phylogeny relating different points, showing the phylogenetic group for each terminal taxon value (different shapes for different major lineages) and an environmental variable as the filling colour. Each point is connected to a terminal taxon in the phylogeny by red-dashed lines. We can see how in the *non-attributed phylogenetic conservatism* scenario, only the phylogeny is related to variance and covariance patterns. In the *environmental phylogenetic conservatism* scenario, both the phylogeny and the environmental variable are related to the variance and covariance patterns. In the *evolutionarily labile environmental effect* scenario, only the environmental gradient is related to the variance and covariance patterns. Finally, in the *residuals*, not the phylogeny nor the environment are related to variances and correlation between traits.

*conservatism*, *non-attributed phylogenetic conservatism* and *evolutionarily labile environmental effects* to trait variances and covariances, helping to elucidate their relative importance in shaping patterns in comparative data. Next, this methodology is used to optimise the use of available data in a newly-derived machine learning algorithm that predicts missing values for functional traits, and which we compare with alternative imputation methods. The reliability of this framework is first tested using simulated data, and it is then applied to a real global dataset of hydraulic (Mencuccini et al., 2019; Sanchez-Martinez et al., 2020) and leaf economics spectrum (LES) traits (Wright et al., 2004) for woody angiosperm species. The methodology presented here is implemented in an accompanying R package named *TrEvol*. A tutorial on how to use the package can be found in [Supporting Information](#) and at the following link: <https://github.com/pablosanchezmart/TrEvol>.

## 2 | METHODS

The proposed method consists of two parts. The first part consists of determining the variance–covariance matrix of pairs of traits by elucidating their association with environmental factors and a phylogeny. We represent these results as trait networks to describe the partition of variance–covariance for a given set of traits, differentiating four components of interest: (1) *environmental phylogenetic conservatism*, (2) *non-attributed phylogenetic conservatism*, (3) *evolutionarily labile environmental effects* and (4) *residual* variance–covariance (not related to the phylogeny, nor to the environmental variable considered). Network metrics can be calculated for each case, which can prove helpful to summarise the multivariate structure of evolutionary patterns in trait syndromes. The second part of the method consists of an imputation algorithm, which uses the previously described variance–covariance structures to make decisions on the use

of available data to produce the most accurate predictions (imputations) of missing values.

## 2.1 | Trait variance–covariance partition

We developed a framework to estimate trait variance–covariance related to an environmental variable or axis and to a phylogeny, so that the individual contribution of these elements (i.e. *evolutionarily labile environmental effect* and *non-attributed phylogenetic conservatism*, respectively) as well as their combined contribution (*environmental phylogenetic conservatism*) can be calculated for pairs of traits. To do so, we used multi-response phylogenetic mixed models implemented in the *MCMCglmm* R package (Hadfield, 2010a). In these models, the phylogenetic effect is set as a random factor on intercepts. To do so, a relatedness matrix is calculated from the phylogeny and it is then used to calculate the effect of the phylogeny on variances for each one of the two traits of interest and their covariances. The phylogenetic effect is analogous to Pagel's lambda, being equal to 1 when the data reflect a pure Brownian Motion model of evolution (Pagel, 1999). As we are interested in evaluating the phylogenetic conservatism of trait relationships with the environment, the environmental variable of interest is also included as a response variable. Then, when an environmental variable is included with the trait pair, we have a model with three response variables for which the phylogenetic and non-phylogenetic variances and covariances are estimated. We understand the environmental variable at the macroevolutionary level, as a component of the ecological niche that can evolve through time.

Because *MCMCglmm* uses a Bayesian framework, it needs prior distributions for the fixed and random effects. We used the default prior for the fixed effects, as suggested in previous implementations (<https://www.mpcm-evolution.com/practice/online-practical-material-chapter-11/chapter-11-1-simple-model-mcmcglmm>). These priors follow an inverse-Gamma distribution, which is canonical and is considered a non-informative prior ([https://devillemereuil.legitux.org/wp-content/uploads/2012/12/tuto\\_en.pdf](https://devillemereuil.legitux.org/wp-content/uploads/2012/12/tuto_en.pdf)) (Hadfield, 2010a, 2010b, 2019). The random effects priors can be manually modified using the *defineModelSpecifications* function of the *TrEvol* package, where users can also modify the number of iterations (default to 1000), the burn-in (default to 100) and thinning (default to 5) (see [Supporting Information](#) for a basic tutorial).

For a specified list of traits, the *computeVarianceCovariancePartition* function selects all pairwise combinations among traits and a single environmental variable (when specified) and builds phylogenetic mixed models including them as response variables (i.e. tri-response models, two traits and one environmental variable). When more than two traits are included, the function iterates and does all possible pairwise combinations among different traits. For each combination of traits, all pairwise complete observations are used to calculate variances and covariances, maximising the use of the data available. For each trait–trait–environment variable combination, the model estimates the amount of phylogenetically dependent (*u*) and

phylogenetically independent (*e*) variances and covariances. Given two traits (T1 and T2) and a continuous environmental variable (E), the model structure is the following:

$$(T1, T2, E) \sim u + e. \quad (1)$$

From this model, the following two variance–covariance matrices are estimated:

$$\begin{array}{l} \text{Matrix } u \text{ (phylogenetically dependent)} \\ \begin{array}{ccc} & T1 & T2 & E \\ T1 & \mathbf{VAR}_u^{T1} & \mathbf{COV}_u^{T1,T2} & \mathbf{COV}_u^{T1,E} \\ T2 & & \mathbf{VAR}_u^{T2} & \mathbf{COV}_u^{T2,E} \\ E & & & \mathbf{VAR}_u^E \end{array} \end{array} \quad (2)$$

$$\begin{array}{l} \text{Matrix } e \text{ (phylogenetically independent)} \\ \begin{array}{ccc} & T1 & T2 & E \\ T1 & \mathbf{VAR}_e^{T1} & \mathbf{COV}_e^{T1,T2} & \mathbf{COV}_e^{T1,E} \\ T2 & & \mathbf{VAR}_e^{T2} & \mathbf{COV}_e^{T2,E} \\ E & & & \mathbf{VAR}_e^E \end{array} \end{array} \quad (3)$$

The elements of these matrices are used to calculate the *environmental phylogenetic variance* ( $\mathbf{VAR}_{env,phylo}^{T1}$ ), the *non-attributed phylogenetic variance* ( $\mathbf{VAR}_{phylo}^{T1}$ ), the *labile environmental variance* ( $\mathbf{VAR}_{env}^{T1}$ ) and the *residual variance* ( $\mathbf{VAR}_{res}^{T1}$ ). Then, these variance components can be aggregated to calculate total phylogenetic variance ( $\mathbf{VAR}_{total,phylo}^{T1}$ ), total environmental variance ( $\mathbf{VAR}_{total,env}^{T1}$ ) and total variance ( $\mathbf{VAR}_{total}^{T1}$ ) for a given trait (e.g. T1), as follows (in bold, aggregation of variance estimates):

$$\mathbf{VAR}_{env,phylo}^{T1} = \left[ \frac{\mathbf{COV}_u^{T1,E}}{\mathbf{VAR}_u^E} \right]^2 \mathbf{VAR}_u^E, \quad (4)$$

$$\mathbf{VAR}_{phylo}^{T1} = \mathbf{VAR}_u^{T1} - \mathbf{VAR}_{env,phylo}^{T1}, \quad (5)$$

$$\mathbf{VAR}_{env}^{T1} = \left[ \frac{\mathbf{COV}_e^{T1,E}}{\mathbf{VAR}_e^E} \right]^2 \mathbf{VAR}_e^E, \quad (6)$$

$$\mathbf{VAR}_{res}^{T1} = \mathbf{VAR}_e^{T1} - \mathbf{VAR}_{env}^{T1}, \quad (7)$$

$$\mathbf{VAR}_{total,phylo}^{T1} = \mathbf{VAR}_u^{T1} = \mathbf{VAR}_{phylo}^{T1} + \mathbf{VAR}_{env,phylo}^{T1}, \quad (8)$$

$$\mathbf{VAR}_{total,env}^{T1} = \mathbf{VAR}_{env}^{T1} + \mathbf{VAR}_{env,phylo}^{T1}, \quad (9)$$

$$\mathbf{VAR}_{total}^{T1} = \mathbf{VAR}_u^{T1} + \mathbf{VAR}_e^{T1} = \mathbf{VAR}_{total,env}^{T1} + \mathbf{VAR}_{pure,phylo}^{T1} + \mathbf{VAR}_{res}^{T1}. \quad (10)$$

The same applies to trait 2. Note that all the variance–covariance estimates used to calculate our metrics are derived from the outputs of the multi-response phylogenetic mixed models (specified in matrices *u* and *e*). When more than two traits are included, different variances can be estimated, as a model is specified for each pairwise combination of traits. In these cases, the mean of all variances' posterior distributions are reported for each trait.

Different variance estimates for the same trait calculated in the current study had a very low variability (lower than the 5% of the mean). However, users can explore this variability by including the traits in the `computeVarianceCovariancePartition` function in pairs, to generate a single estimate of variance for each trait before including all traits together. We recommend this practice to evaluate variability in variance estimates for a given trait before reporting the main results.

Similarly, elements of the two matrices were used to calculate *environmental phylogenetic covariance* ( $COV_{env,phylo}$ ), *non-attributed phylogenetic covariance* ( $COV_{phylo}$ ), *labile environmental covariance* ( $COV_{env}$ ), and *residual covariance* ( $COV_{res}$ ) between all pairwise combinations of traits. Then, these covariance components can be aggregated to calculate total phylogenetic covariance ( $COV_{total,phylo}$ ), total environmental covariance ( $COV_{total,env}$ ) and total covariance ( $COV_{total}$ ) for a given pair of traits (e.g. T1 and T2), as follows (in bold, aggregation of variance estimates):

$$COV_{env,phylo}^{T1,T2} = \frac{COV_u^{T1,E} COV_u^{T2,E}}{VAR_u^E}, \quad (11)$$

$$COV_{phylo}^{T1,T2} = COV_u^{T1,T2} - COV_{env,phylo}^{T1,T2}, \quad (12)$$

$$COV_{env}^{T1,T2} = \frac{COV_e^{T1,E} COV_e^{T2,E}}{VAR_e^E}, \quad (13)$$

$$COV_{res}^{T1,T2} = COV_e^{T1,T2} - COV_{env}^{T1,T2}, \quad (14)$$

$$COV_{total,phylo}^{T1,T2} = COV_u^{T1,T2} = COV_{phylo}^{T1,T2} + COV_{env,phylo}^{T1,T2}, \quad (15)$$

$$COV_{total,env}^{T1,T2} = COV_{env,phylo}^{T1,T2} + COV_{env}^{T1,T2}, \quad (16)$$

$$COV_{total}^{T1,T2} = COV_u^{T1,T2} + COV_e^{T1,T2} = COV_{total,env}^{T1,T2} + COV_{phylo}^{T1,T2} + COV_{res}^{T1,T2}. \quad (17)$$

When no environmental effect is specified, the function calculates total variance and total covariance and phylogenetic variance and phylogenetic covariance, the difference being the total non-phylogenetic variance and covariance. The function allows for representation of variances in absolute or relative terms. When `show_relative_variance = TRUE` (which is the default), the proportion of explained variance from the total variance is reported. Covariance is reported as correlation in all cases (e.g. using phylogenetic variances and covariances to calculate phylogenetic correlation), calculated as follows:

$$Corr = \frac{COV^{T1,T2}}{\sqrt{VAR^{T1}VAR^{T2}}}. \quad (18)$$

The function `computeVarianceCovariancePartition` returns a list containing different elements. The first element includes two results tables, one displaying variance results and the other the covariance results. The second element is the individual model outputs (*MCMCglmm* models from which the estimates to calculate variances and covariance are derived). The third element is the

table of posterior distributions for each variance and covariance component, from which variances and covariances are calculated. These posterior distributions can be used to diagnose models by testing convergence, effective sample sizes and autocorrelation. Tests for the convergence, effective sample sizes and autocorrelation are implemented in *TrEvol* package by means of the `diagnoseModels` function.

The result tables showing variance and covariance (given as correlation) partitions show the mean estimate for each variance and covariance estimate jointly with the credible intervals and a *p*-value representing statistical significance. The interpretation of credible intervals is that there is a 95% probability that the true estimate lies within the interval, given the data. Then, there is a statistically significant effect when the 95% credible interval does not contain zero. This result table also reports *p*-values as an alternative to evaluate statistical significance of estimates. *p*-values are calculated from the probability of direction, *p*-direction, a metric that has shown high correspondence to the frequentist *p*-value (two sided *p*-value = 2 × (1 - *p*-direction)) (Makowski, Ben-Shachar, Chen, et al., 2019). Probability of direction is calculated using the *bayestestR* package (Makowski, Ben-Shachar, & Lüdtcke, 2019) for the distributions obtained for each variance and covariance component reported.

Inputs needed in the `computeVarianceCovariancePartition` function are:

- **traits**: character vector containing the names of the trait or traits to be considered. These variables need to be in the *dataset*. For example `c('trait_1', 'trait_2')`.
- **environmental\_variable**: character vector containing the names of the continuous environmental variable to be considered. These variable need to be in the *dataset*. For example `'aridity_index'`. See [supplementary methods](#) for a version of the methodology that allows to include more than one environmental variable at the same time.
- **dataset**: data frame containing traits, environmental variables and taxon names (e.g. species, genus or other taxon names).
- **phylogeny**: *phylo* object including taxa present in the *dataset*. The function matches internally the *dataset* with the *phylogeny*, so it allows missing values in the data frame and some taxa present or absent in the *phylogeny*. The function will only use complete cases for each pair of traits, so data with incomplete phylogenetic and environmental observations are pruned.
- **Terminal\_taxa**: character identifying which is the column containing the terminal taxa (e.g. `'species'`, which is the default).

## 2.2 | Networks

To represent the multivariate structure, variance-covariance structures can be plotted as networks by the function `plotNetworks`, which shows traits' variance (i.e. variance related to the environment, phylogeny or their combination) as pie charts and correlation

as edges for each one of the pairs of traits included. A given network represents trait variances and correlations related to *environmental phylogenetic conservatism*, *non-attributed phylogenetic conservatism*, *evolutionarily labile environmental effect* and *residuals*. Total correlation can also be plotted.

The same function also calculates and displays network metrics when the *plot\_metrics* argument is set to TRUE, which is the default. These measures are the number of modules (NM) showing how many independent groups of correlated traits exist; edge density (ED) describing the proportion of actual connections among nodes out of all possible connections; maximum absolute correlation coefficient ( $|r|_{\max}$ ) and mean absolute correlation coefficient ( $|r|_{\text{mean}}$ ). In this framework, NM represents the number of trait modules, high ED represents high coordination among all traits and high  $|r|_{\max}$  and  $|r|_{\text{mean}}$  represent networks with a higher dependence among related traits. At the node level, degree (i.e. number of connections per node) can be also calculated and displayed as node size when the argument *degree\_as\_node\_size* is set to TRUE, which is the default. Traits with a high degree value (i.e. higher node size in the visualisation) can be considered hubs.

### 2.3 | Imputation algorithm

The package presented here includes an imputation algorithm performed by the function *imputeTraits*, which uses the previously described variance–covariance structures to guide the trait imputation process. Inputs needed are:

- *dataset*: a data frame containing traits with missing values and optionally environmental variables for a set of terminal taxa. Note that more than one observation per taxon can be included (i.e. the methodology can include within species variability).
- *terminal\_taxa*: character identifying which is the column containing the terminal taxa (e.g. 'species', which is the default).
- *phylogeny*: a 'phylo' object including all terminal taxa present in the dataset.
- *variables\_to\_impute*: character vector indicating the name of the variables to be imputed as they appear in the dataset.
- *predictors*: character vector indicating the list of environmental factors included in the dataset (without missing values) that may be considered as potential predictors (see below).

Phylogenetic information is included to predict by calculating phylogenetic coordinates for each terminal taxon. To calculate phylogenetic coordinates, the algorithm first calculates the phylogenetic distance matrix by means of the *cophenetic.phylo* function and then calculates principal coordinates by means of the *pcoa* function, both from the *ape* package (Paradis & Schliep, 2019). Phylogenetic coordinates are stored internally so the algorithm does not need to calculate them each time that it is run for a given dataset, which can be time consuming specially for big datasets. However, it will recalculate them if the argument *force\_run* is set

to 'TRUE'. Users can also control the number of phylogenetic coordinates to be considered as predictors by using the *number\_of\_phylo\_axis* argument (default set to 10). Note that all terminal taxa present in the input dataset need to be also present in the phylogeny.

The algorithm included in the *imputeTraits* function uses trait relationships as well as the phylogenetic structure and, optionally, environmental factors to predict missing values by performing three chained imputation rounds. In the first round, missing values for individual traits are predicted by using complete data on environmental variables and phylogeny as predictors, when they are informative (i.e. when they present statistically significant relationships with trait variance). In the second round, each missing values of individual traits is sequentially predicted using environmental variables, phylogeny and other traits (including values imputed in the first round) as predictors, when they are informative. In the third round, missing values of individual traits are predicted in turn using environmental variables, phylogeny and other traits (imputed in the second round) as predictors when they are informative (Figure S1). Preliminary analyses showed that imputation performance improved in some cases when a third round of imputation was performed. To elucidate which variables are informative and thus included as predictors in each case, the *computeVarianceCovariancePartition* function is computed internally to calculate phylogenetic and environmental effects for each trait as well as the covariances among the traits to be imputed. Users can also include the results of the *computeVarianceCovariancePartition* manually using the *variance\_results* and the *correlation\_results* arguments of the function *imputeTraits* (see package tutorial).

In each imputation round, imputation of missing values for traits is performed for each one of the traits of interest. Each imputation model is iterated several times (controlled by the *number\_iterations* argument, set to 10 by default). From each imputation round, an imputed matrix is obtained by calculating the mean for all iterations for each predicted value. Standard deviations of predicted values and individual model results are also returned.

The algorithm uses random forest models to predict missing values by means of the *randomForest* function of the *randomForest* R package (Liaw & Wiener, 2002). These models calculate out of bag errors (OOB), which are returned also as a measure of imputation error reported as the normalised root mean square error (NRMSE). Moreover, the developed algorithm calculates cross validation  $R^2$  values by randomly creating NAs in the traits to be imputed, predicting values, and then comparing predictions with observed values. The proportion of NAs that are randomly created to calculate cross validation  $R^2$  is controlled by the *proportion\_NAs* argument of the function (between 0 and 1, set to zero by default). When *proportion\_NAs*=0, cross validation  $R^2$  is not calculated. Parallel processing is enabled in the function when the *number\_clusters* is set to an integer number higher than one, which will determine the number of clusters used. By default, the number of clusters used is set to two. The parallelisation method is based on previous implementations of random forests in R programming (Stekhoven & Bühlmann, 2012).

## 2.4 | Application

To test the methodology, we first applied it to 10 simulated traits, each one with 100 observations (simulated species) with a known variance–covariance structure (Table S1). We also simulated a pure-birth stochastic phylogeny using the *pbtree* function of the *phytools* package (Revell, 2012) for the 100 simulated species. Five traits were simulated to present a phylogenetic component in their variance–covariance structure by means of *simulate\_bm\_model* function of the R package *castor* (Louca & Doebeli, 2018), including the simulated phylogeny and the variance–covariance matrix (Table S1) as the diffusion matrix. The remaining five traits were simulated using the same variance–covariance matrix (Table S1) to present a variance–covariance structure not-explicitly related to the phylogeny by using the function *rnorm\_multi* of the R package *faux* (DeBruine, 2021).

Traits conformed to four trait modules (i.e. traits present correlation with traits in the same module but not with traits in other modules). The first correlation module (M1) was constituted by *Phylo Trait 1*, *Phylo Trait 2*, the variance–covariance structure of which was expected to be related to the phylogeny and to the simulated environmental variable *phylo Env* (i.e. *environmental phylogenetic conservatism* associated with *Phylo Env*). The second correlation module (M2) was constituted by *Phylo Trait 3* and *Phylo Trait 4*, whose variance–covariance was also expected to be related to the phylogeny but not to traits in M1 nor to the environmental variable *Phylo Env* (i.e. *non-attributed phylogenetic conservatism* given *Phylo Env*). The third correlation module (M3) was constituted by *Non Phylo Trait 1*, *Non Phylo Trait 2*, whose variance–covariance was not expected to be related to the phylogeny but was expected to be related to *Non Phylo Env* (i.e. *evolutionarily labile environmental effect* of *Non Phylo Env*). The fourth correlation module (M4) was constituted by *Non Phylo Trait 3* and *Non Phylo Trait 4*, whose variance–covariance was not expected to be related to the phylogeny nor to the previously presented environmental variables (i.e. *residual*) (Figure S2). Note that ‘Env’ variables were considered as environmental variables but the procedure to obtain them was not different from the one implemented for the rest of the simulated variables. The expectations for results were summarised a priori (Table S2). Variable simulations were performed by using the *simulateDataSet* function of the R package presented here, which by default produces the simulated data used in this study.

Next, we applied the methodology to real data on functional traits for angiosperm woody-plant species related to hydraulic function and the leaf economics spectrum (Wright et al., 2004). We used data covering seven leaf traits coming from a previously published dataset (Mencuccini et al., 2019). Functional traits included were specific leaf area ( $\text{cm}^2 \text{g}^{-1}$ , SLA, number of species=1183), leaf nitrogen content ( $\text{mg g}^{-1}$ , number of species=930), maximum photosynthetic rate per area as a measure of photosynthetic capacity ( $\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$ ,  $A_{\text{area}}$ , number of species=206), leaf lifespan (months, LL, Number of species=198), midday leaf xylem tension as a measure of xylem drought exposure (MPa,  $|P_{\text{min}}|$ , number

of species=636), tension at the turgor loss point as a measure of drought tolerance (MPa,  $|P_{\text{tip}}|$ , number of species=556) and leaf-specific maximum hydraulic conductivity as a measure of hydraulic sufficiency ( $\text{Kg m}^{-1} \text{MPa}^{-1} \text{s}^{-1}$ ,  $K_p$ , number of species=878). We considered an aridity index (AI) as an environmental variable related to species niche. The aridity index was obtained from CGIAR (Trabucco & Zomer, 2018) in a previously published analysis (Sanchez-Martinez et al., 2020). Low values of aridity index represent regions with low water availability relative to the atmospheric water demand. All variables were transformed by calculating the natural logarithm of absolute values, as it improved their normality.

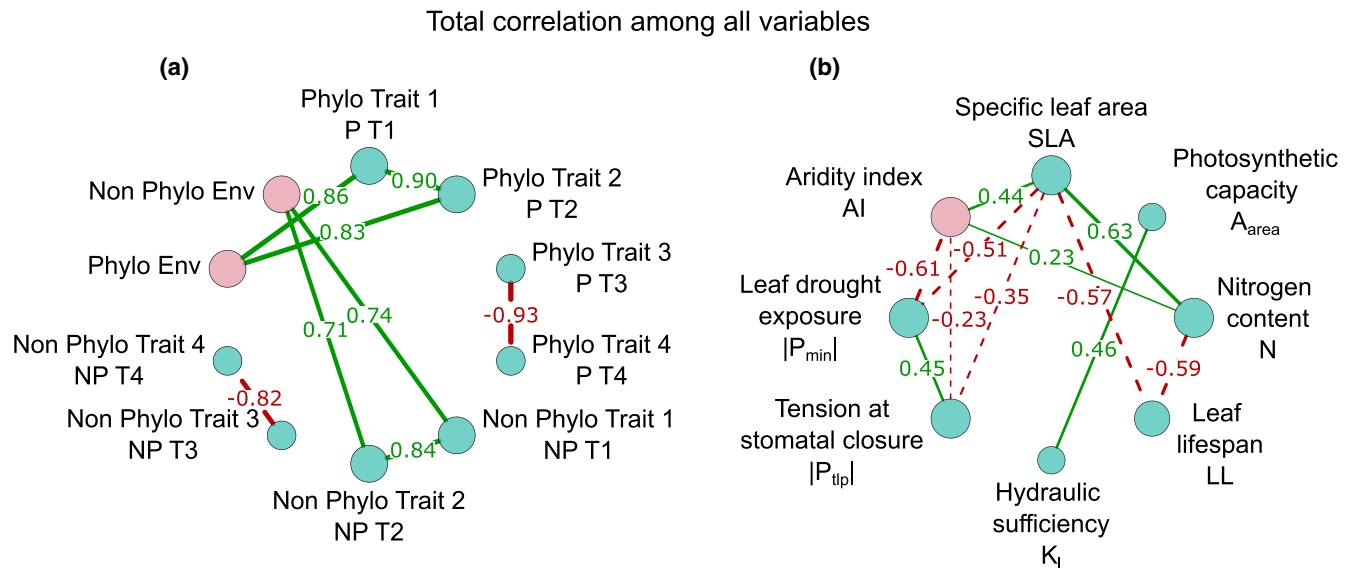
Partitioned trait networks for simulated data and for functional trait data were obtained. Then, the imputation method was implemented both for simulated and for functional traits, considering environmental effects described in each case as potential predictors as well as the phylogeny. In this empirical data case, a species level phylogeny was built using the *V.PhyloMaker* R package (Qian & Jin, 2016). The remaining arguments of the function *imputeTraits* were maintained as default.

To compare results obtained by *imputeTraits* with existing imputation methodologies, we performed a cross validation procedure by using 80%-20%, 50%-50% and 30%-70% of the data to test and train, respectively, for each one of the methodologies, including our function *imputeTraits*. These methodologies include *mean* imputation (i.e. mean for each one of the traits to be taken as imputed values); *MICE* imputation (van Buuren & Groothuis-Oudshoorn, 2011), which uses the covariation of multiple traits to be imputed as well as other complete environmental variables to predict missing values; and a phylogenetic imputation by means of the *phylopars* R package, which uses trait phylogenetic variance–covariance to predict missing values (Goolsby et al., 2017). Thus, predictive performance results reported are not the ones reported internally by the *imputeTraits* function but calculated outside the algorithm in order to use the same cross validation data for all the imputation methods compared. The cross-validation process was iterated 10 times in each case. We acknowledge that a cross-validation using data with an underlying phylogenetic structure may underestimate prediction error and overestimate cross-validation  $R^2$  (Roberts et al., 2017). However, error and  $R^2$  in this implementation are used to compare the predictive capacity of different methodologies, so absolute values are of less interest and we expect the phylogenetic structure to have similar effects on the  $R^2$  and errors for all the methodologies.

## 3 | RESULTS

### 3.1 | Variance–covariance partition results for simulated data

Results obtained by applying the described methodology for a set of simulated traits matched those expected given the variance–covariance matrix used to simulate data, detecting the four hypothesised modules (Table S1; Figure 2a). The methodology correctly detected



**FIGURE 2** Total correlation among variables. (a) Total correlation among simulated variables and (b) total correlation among plant functional traits. Pink nodes represent environmental variables and blue nodes represent traits.

the amount of variance and covariance related to the phylogenetic structure of the data, as is the case for the covariance between *Phylo Trait 1* and *Phylo Trait 2*, between *Phylo Trait 3* and *Phylo Trait 4* and their variances (edges and pie charts, respectively, in [Figures 3](#) and [4](#)). Moreover, it successfully detected the independence between these two modules of phylogenetically correlated traits. The methodology also successfully separated the variance and covariance that is uniquely attributed to shared evolutionary heritage (*non-attributed phylogenetic conservatism*) from *environmental phylogenetic conservatism* associated with *Phylo Env* ([Figures 4b,c](#) and [5b,c](#)). Finally, the methodology successfully detected the *evolutionarily labile environmental effect* (i.e. phylogenetically independent effect of the environmental variable) of *Non Phylo Env* on variances of *Non Phylo Trait 1* and *Non Phylo Trait 2* traits as well as on their covariance ([Figure 4d](#)), while the covariance not related to the phylogenetic structure nor to the given environmental variables were placed within the unexplained or *residual* component ([Figures 3](#) and [4e](#)). Networks presenting a higher contribution to the covariation pattern had higher edge density (ED); those presenting a strong relationship between connected nodes showed a high maximum absolute correlation ( $|r_{\max}|$ ) and mean absolute correlation ( $|r_{\text{mean}}|$ ) and the number of modules related to each one of the components was also successfully detected by the methodology (NM) ([Figures 3](#) and [4](#)).

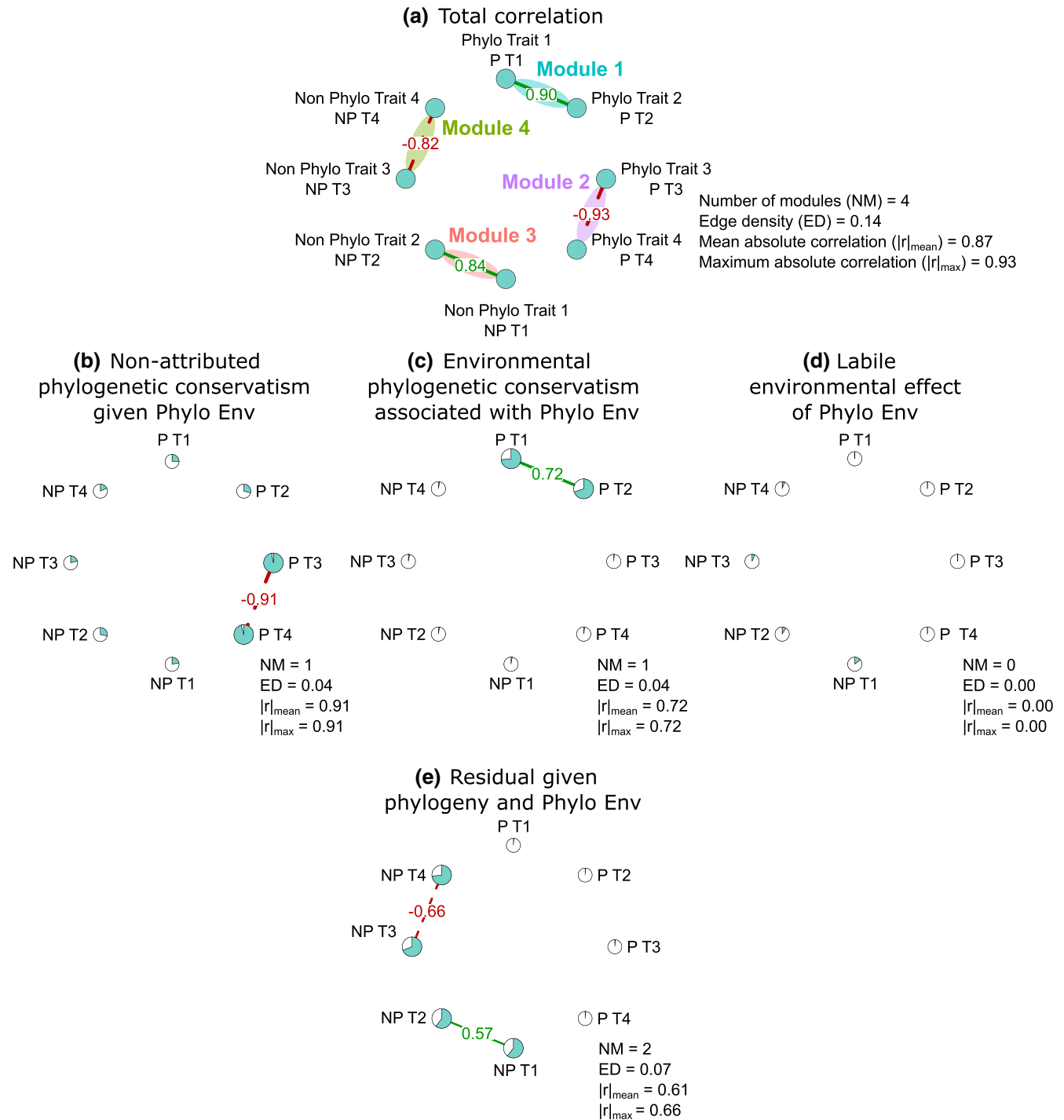
### 3.2 | Variance–covariance partition results for functional trait data

When applied to the dataset of traits of woody angiosperms species, we found that leaf economics traits and leaf hydraulic traits conformed to two main modules ([Figure 5a](#)), that trait variances and covariance were mainly phylogenetically conserved and that

this conservatism was partially associated with aridity ([Figures 2b](#) and [5](#)). The first module represents the correlation between leaf resource uptake and use strategies (i.e. leaf economics) with xylem drought exposure and drought tolerance (quantified as the absolute value of the minimum water potential measured in the xylem,  $|P_{\min}|$  and the absolute value of the water potential at stomatal turgor loss point,  $|P_{\text{tip}}|$ , respectively). This module represents variation in an axis describing acquisitive leaves with low drought exposure-tolerance to leaves with conservative resource uptake and use with high drought exposure and tolerance (specific leaf area is positively related to nitrogen content, negatively related to leaf lifespan and negatively related to drought exposure and tolerance). This module presents phylogenetic conservatism and part of this conservatism is associated with aridity, especially in those correlations involving drought exposure and tolerance ([Figure 5c](#)). Contrarily, most of the phylogenetic conservatism in leaf economic traits and their covariance (specific leaf area, nitrogen content and leaf lifespan) cannot be attributed to aridity ([Figure 5b](#)). The second module relates hydraulic sufficiency with photosynthetic capacity (higher maximum hydraulic conductivity related to higher photosynthetic capacity, both expressed on a per unit leaf area). This module showed phylogenetic conservatism, which cannot be attributed to aridity ([Figure 5b](#)).

The two modules described were shown to be integrated in response to aridity ([Figure 5c](#)). More concretely, specific leaf area was positively related to hydraulic sufficiency and negatively related to photosynthetic capacity, and drought exposure was negatively related to hydraulic sufficiency. Then, the integration between these two modules is mediated by specific leaf area and xylem drought exposure, which show a high degree (number of connections presented by a node) and act as hubs of plant functional trait syndromes in woody angiosperm species. These correlations are related to

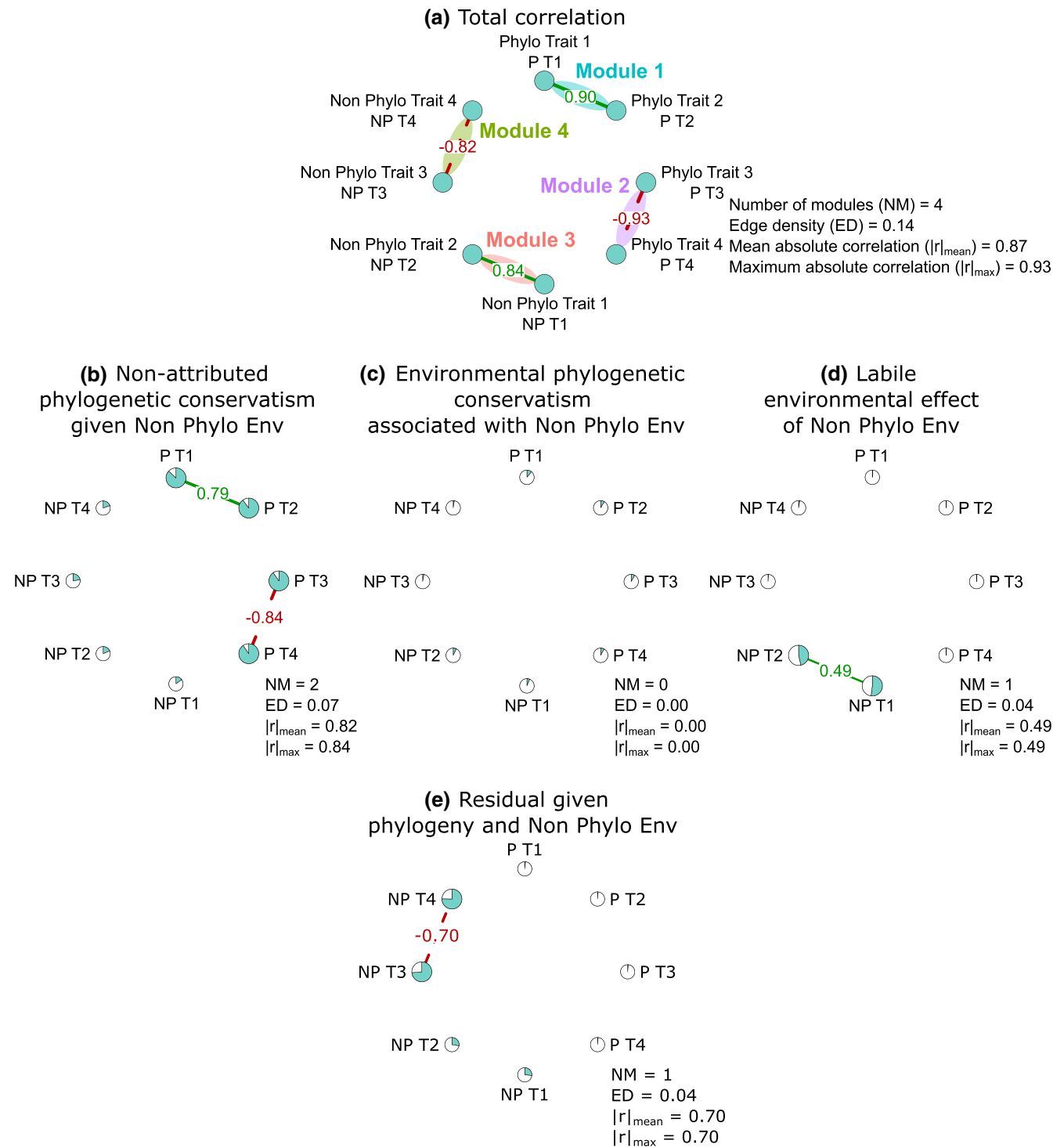




**FIGURE 3** Simulated data networks. Simulated data networks considering a phylogenetically conserved environmental variable (*Phylo Env*). Pie charts represent the variance and edges represent the covariance related to each one of the components (*non-attributed phylogenetic conservatism*, *environmental phylogenetic conservatism*, *evolutionarily labile environmental effect* and *residual*). Green-solid edges represent positive correlations and red-dashed ones negative correlations with a width proportional to the correlation coefficient. Node size is proportional to node degree (number of connections). Network metrics are also shown in each case. (a) Total correlation; (b) *non-attributed phylogenetic conservatism* in variances and correlations given *Phylo Env*; (c) *environmental phylogenetic conservatism* in variances and correlations associated with *Phylo Env*, (d) *labile environmental effect* of *Phylo Env* in variances and correlations and (e) *residual* variance and correlation given the phylogeny and *Phylo Env*.

aridity in a phylogenetically structured manner, describing a pattern of *environmental phylogenetic conservatism* associated with aridity in the integration between plant hydraulics and leaf economics.

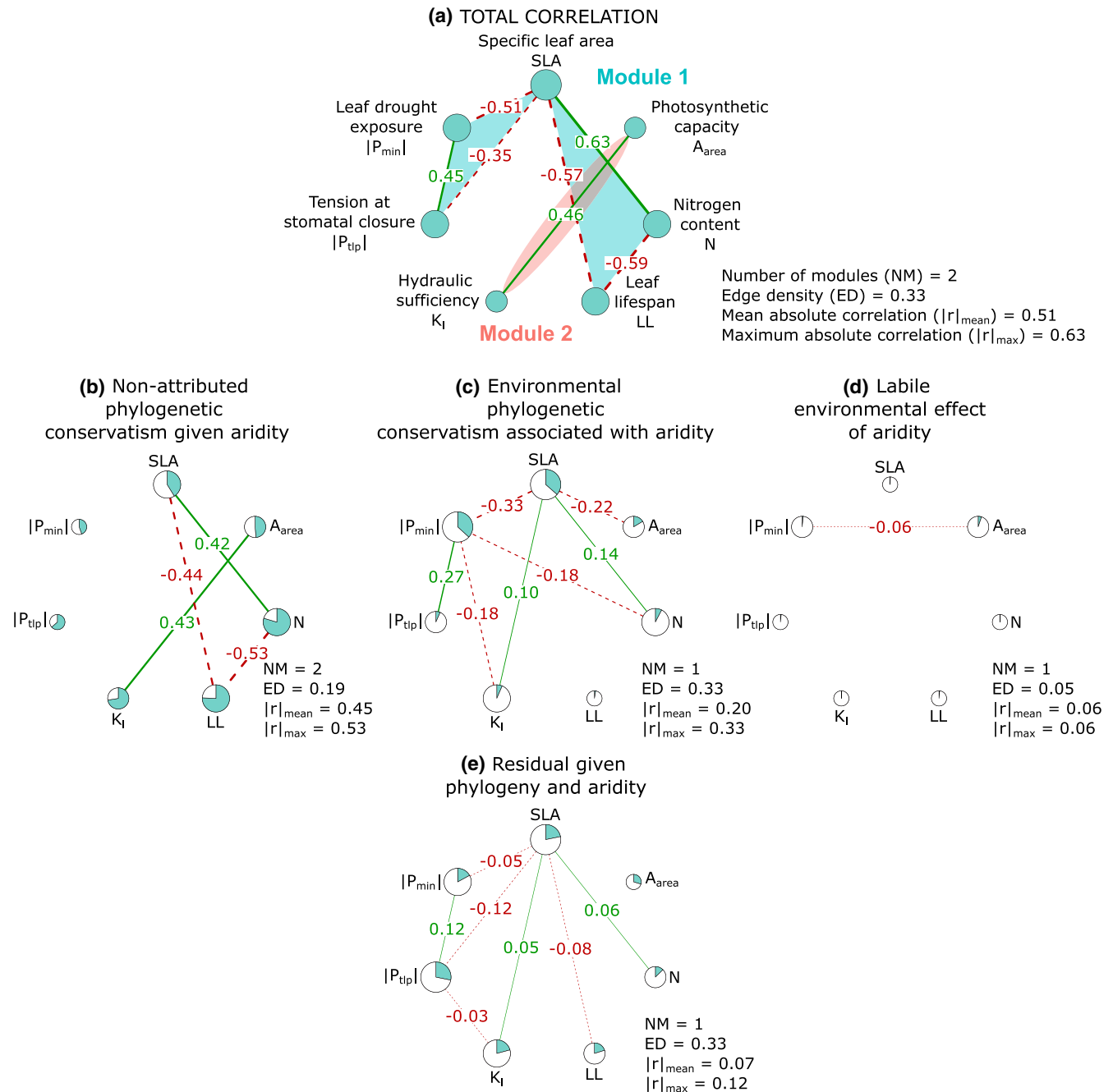
Trait networks supported the predominant pattern of phylogenetic conservatism in functional traits of woody angiosperm species. Networks representing *non-attributed phylogenetic conservatism*



**FIGURE 4** Simulated data networks. Simulated data networks considering a phylogenetically independent environmental variable (*Non Phylo Env*). Pie charts represent the variance and edges represent the covariance related to each one the components (*non-attributed phylogenetic conservatism*, *environmental phylogenetic conservatism*, *evolutionarily labile environmental effect* and *residual*). Green-solid edges represent positive correlations and red-dashed ones negative correlations with a width proportional to the correlation coefficient. Node size is proportional to node degree (number of connections). Network metrics are also shown in each case. (a) Total correlation; (b) *non-attributed phylogenetic conservatism* in variances and correlations given Non Phylo Env; (c) *environmental phylogenetic conservatism* in variances and correlations associated with Non Phylo Env, (d) *labile environmental effect* of Non Phylo Env in variances and correlations and (e) *residual* variance and correlation given the phylogeny and Non Phylo Env.

and *environmental phylogenetic conservatism* associated with aridity showed higher values of edge density (ED), maximum absolute correlation ( $|r|_{\text{max}}$ ) and mean absolute correlation ( $|r|_{\text{mean}}$ ) compared to

the *evolutionarily labile environmental effect* of aridity. The number of modules proved to be useful to detect the number of independent groups of correlated traits, showing the integration of the two



**FIGURE 5** Leaf trait networks considering aridity as the environmental variable of interest. Trait networks using angiosperm woody plant traits data considering aridity (aridity index) as the environmental variable of interest. Pie charts represent the proportion of the variance for each variable related to each one of the components (*non-attributed phylogenetic conservatism*, *environmental phylogenetic conservatism*, *evolutionarily labile environmental effect*, and *residual*) and edges represent the covariance related to each one of these components shown as correlations (green-solid lines represent positive correlations and red-dashed line represent negative correlations with a width proportional to the correlation coefficient). Node size is proportional to node degree (number of connections). Network metrics are also shown, in each case. (a) Total correlation; (b) *non-attributed phylogenetic conservatism* in variances and correlations given aridity; (c) *environmental phylogenetic conservatism* in variances and correlations associated with aridity, (d) *labile environmental effect* of aridity in variances and correlations and (e) *residual* variance and correlation given the phylogeny and aridity.

modules in response to aridity (Figure 5c). The higher ED in the phylogenetic part of trait relationships indicates that most of the trait covariances are phylogenetically conserved. The higher  $|r|_{max}$  and  $|r|_{mean}$  in these networks indicates that phylogenetically conserved relationships are stronger (Figure 5b,c).

### 3.3 | Imputation results

The predictive performance of the presented imputation algorithm improved when informative imputed traits were considered as shown by the higher cross-validation  $R^2$  and lower cross-validation

NRMSE of the second and third imputation rounds compared to results obtained using only phylogenetic and environmental data (first round) (Figure S3), both using simulated and leaf functional trait data. The *imputeTraits* algorithm outperformed *mean* imputation and the *MICE* imputation when using all environmental and traits covariation and performed similarly to *phylopars* imputation using all environmental and traits phylogenetic covariation when traits show phylogenetic structure. The new algorithm outperformed *phylopars* approach when the covariance among traits was not phylogenetically structured (Figure 6). Thus, considering all simulation scenarios, the new algorithm proposed here was the most effective, both in simulated and leaf functional data.

## 4 | DISCUSSION

### 4.1 | A new framework to study functional syndromes evolution

In this study, we developed a new methodological framework to study eco-evolutionary patterns in multiple functional traits. This methodology allows separation of variances and covariances related to phylogenetic heritage from those that are phylogenetically independent, complementing previous methods such as phylogenetically independent contrasts and phylogenetic least squares, which incorporate the phylogenetic structure when analysing trait relationships, identifying correlations that are not associated with a low number of deeply conserved divergences (Blomberg, 2016; Felsenstein, 1988). By using this methodology, evolutionary patterns can be elucidated not only from phylogenetically independent contrast results as done in previous studies (e.g. Ackerly & Reich, 1999; Maherali et al., 2004; Zheng et al., 2019), but also from the phylogenetically conserved component of evolutionary relationships (Sanchez-Martinez et al., 2020). Furthermore, our methodology makes it possible to test whether phylogenetically conserved and phylogenetically independent structures are related to a potential driving variable (here, environmental variable). Therefore, this method quantifies which part of trait variance and covariance is related only to the phylogeny given one or more variable(s) (*non-attributed phylogenetic conservatism*), only to an environmental variable (*evolutionarily labile environmental effect*) and to the phylogenetically conserved environmental effect (*environmental phylogenetic conservatism*). The latter represents (co) variation that is related to both an environmental variable and a phylogeny.

These three components show patterns reflecting different evolutionary processes which could not be distinguished using previously published frameworks, which were not able to consider multiple traits, their phylogenetic structure and their relationship with the environment all at once. Our approach also partitions networks and calculates network metrics such as edge density (ED), number of modules (NM), maximum absolute correlation ( $|r_{\max}|$ ) and mean absolute correlation ( $|r_{\text{mean}}|$ ) for each component, showing how the

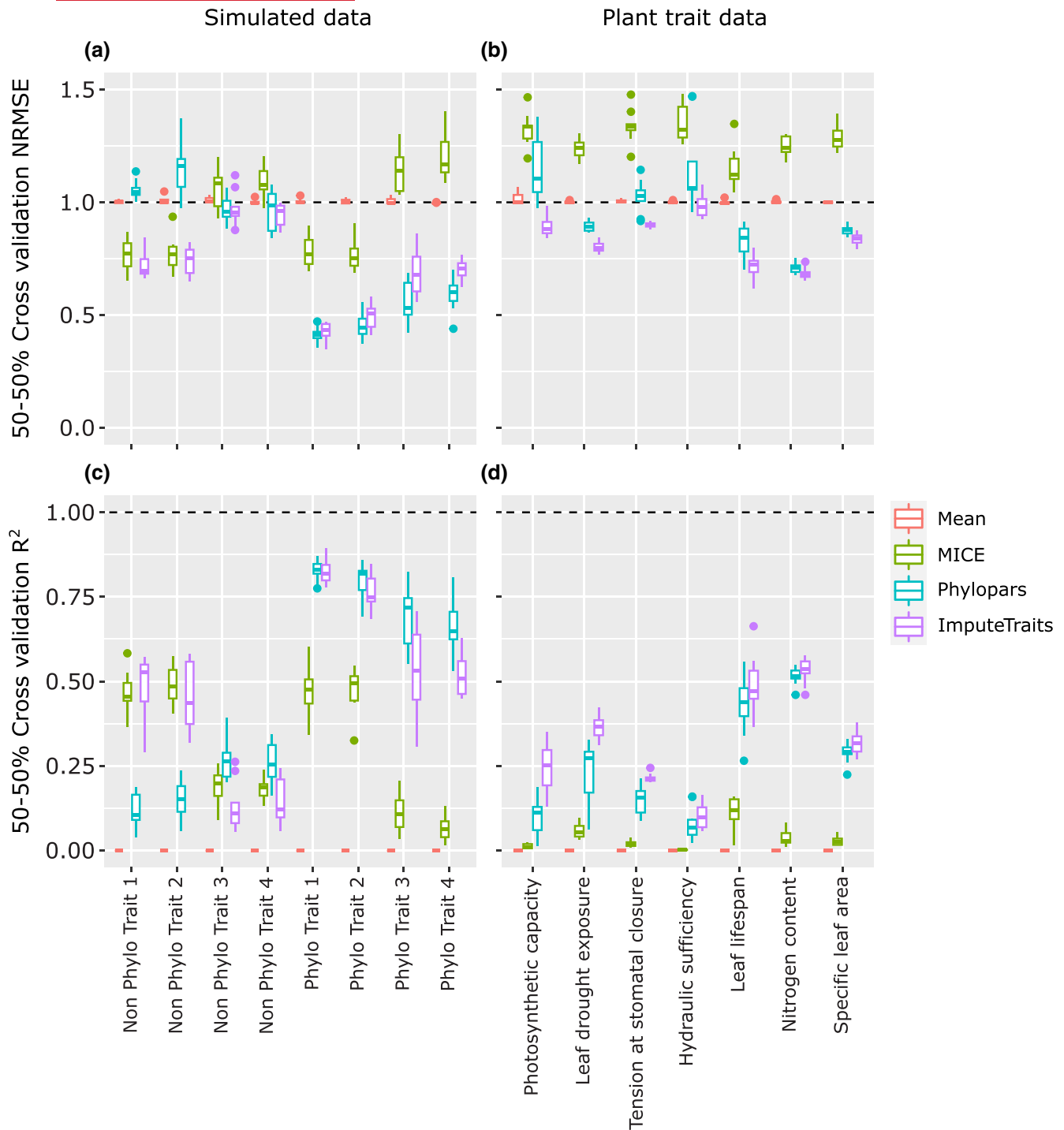
multivariate trait structure is related to the phylogeny and to the environment.

### 4.2 | Leaf economic and hydraulic conform to two phylogenetically conserved modules integrated in response to aridity

Implementing the new methodological approach we demonstrate that the leaf economics spectrum (Wright et al., 2004) and plant hydraulics are evolutionarily integrated, conforming to two evolutionary modules across species at a global scale. The first module describes the coordination between leaf acquisitiveness (Wright et al., 2004) and the coordination between xylem drought exposure and tolerance (Sanchez-Martinez et al., 2020). The second module describes the coordination between leaf hydraulic sufficiency and photosynthetic capacity (Scoffoni et al., 2016). Our results show how these two modules are partially integrated following a pattern of *environmental phylogenetic conservatism* related to aridity. This integration is mainly mediated by specific leaf area and drought exposure, which act as functional hubs relating the two modules previously presented in response to aridity. However, some phylogenetic conservatism in individual trait variability, the correlation among leaf economic traits and the correlation between hydraulic sufficiency and photosynthetic capacity were not attributed to a response to macro-evolutionary patterns of aridity. These functional traits and their integration may be related to other environmental components of species niches (e.g. soil properties, successional status) or may be constrained at the phenotypic or genetic level.

From these results, we can better understand the generalities of woody angiosperm adaptations to aridity. Some angiosperm lineages exposed to low soil water availability and high evaporative demand will tend to present a more negative turgor loss point leading to a higher capability to maintain stomata open and carbon uptake under low water potentials (i.e. anisohydric strategy), presenting a higher drought exposure in the xylem as a consequence (Volaire, 2018). These lineages will tend to present an embolism resistant xylem that will allow maintenance of water transport under low water potentials (Sanchez-Martinez et al., 2020). Drought resistant lineages will tend to have smaller and/or thicker leaves as an adaptation to tolerate water stress (Wright et al., 2017), with lower nitrogen content and higher leaf lifespan, conforming to a conservative strategy of resource acquisition and processing (Reich, 2014; Wright et al., 2004). Importantly, lineages with a lower specific leaf area will also tend to present a relatively higher photosynthetic capacity (per unit area) in response to aridity levels, suggesting that even if there is a phylogenetically conserved coordination between hydraulic conductivity and carbon uptake, this coordination is not associated with the macro-evolutionary response to aridity.

These results mean in turn that under more arid conditions, lineages with a lower specific leaf area may tend to present a higher photosynthetic capacity per unit leaf area independently of their hydraulic sufficiency, probably as an adaptation to maintain



**FIGURE 6** Predictive performance. Predictive performance of the imputation framework. Results for *mean* imputation (red), *MICE* imputation (blue), *phylopars* imputation (green) and *imputeTraits* (*TrEvol* package) imputation (purple) are shown. (a) and (b) show normalised root mean square error (NRMSE) for simulated and leaf traits, respectively. (c) and (d) show  $R^2$  for simulated and leaf traits, respectively. NRMSE and  $R^2$  were calculated from a cross validation procedure using 80% of the data to train models and 20% to test.

carbon uptake under low water availability. Then, a drought resistant functional strategy may compensate for the lower hydraulic sufficiency and the lower leaf area by increasing its photosynthetic capacity per unit area. This result contradicts lineage-specific results reporting that hydraulic sufficiency is evolutionarily coordinated with photosynthetic capacity in response to aridity

(Scoffoni et al., 2016) but supports other evidence showing that photosynthetic capacity may increase under dry conditions (Green et al., 2020; Ramírez-Valiente & Cavender-Bares, 2017). It should be noted, however, that these traits presented the lowest data availability from all the combination of traits used (118 species with both traits), which may limit the ability to extract global

patterns. Therefore, future work on the coordination of photosynthetic capacity with other functional traits will be useful to further elucidate its adaptive role.

The reported results show that leaf economic traits and hydraulic traits are not evolving independently, leading to a pattern of phylogenetic conservatism where closely related species tend to present similar functional strategies. These relationships are partially related to species niches characterisation, specifically to aridity, pointing to a pattern of phylogenetic niche conservatism (Losos, 2008). Based on these results, we expect adaptation in leaf and hydraulic trait syndromes of woody angiosperms to happen in a slow and conserved manner, broadly maintaining their relationships and multivariate structure, with closely related species tending to occupy similar ecological and functional spaces (i.e. having similar trait syndromes). It seems likely that environmental filtering related to variables such as water availability and atmospheric water demand constrains the diversity of trait syndromes that can be present under a given set of conditions and hardwired relationships among traits constrain individual trait variation at the macro-evolutionary scale. Elucidating the directionality of these relationships in future works will help us to better understand which traits are responding to selective pressures directly and which are indirectly related to them as a result of trait integration.

### 4.3 | Using information on trait evolution to perform data imputation

Our statistical framework also enables new approaches to imputation and achieved high predictive performance, especially when trait covariances are considered. The proposed imputation methodology optimises the usage of available data by including phylogenetic, environmental, and additional functional trait data as predictors when they present a significant correlation with the variable to be imputed. This method provides a framework for users that are not familiar with phylogenetic methods, internally testing for the importance of the phylogenetic structure, and including it when informative. Moreover, it allows the consideration of multiple environmental variables, which will only be included when informative. Finally, it allows for the consideration of functional trait data with missing values as predictors by performing imputation and then including them in predictive models when informative.

This methodology outperformed *mean* and *MICE* imputations, while performing comparably to *phylopars* in most cases, and outperforming it for non-phylogenetically conserved traits and observed plant functional traits. However, it slightly underperformed *phylopars* for phylogenetically conserved traits. This may be associated with the fact that we include phylogenetic data by means of a relatively low number of principal components describing most of the phylogenetic structure, which may lead to relatively lower predictive performance in highly phylogenetically structured traits compared to methods that include the whole

phylogeny, such as *phylopars*. Overall, we show how the presented methodology is a good option, especially when using real data, which has a more complex underlying structure than the data simulated in this study.

The fact that the methodology presented here uses widely available information, such as phylogenetic and environmental data, makes it broadly applicable, particularly given the generalised scarcity of functional data at global scales. Moreover, it facilitates the use of phylogenetic information for ecologists without training in phylogenetics.

### 4.4 | Future directions

The new methodology proposed here estimates the phylogenetic effect on variances and covariances based on an underlying Brownian Motion model of evolution. Then, this methodology does not allow, for instance, to include different adaptive optima leading patterns in the data. In future versions of this method, other evolutionary models may be included, such as the Ornstein-Uhlenbeck model of evolution to allow for one or more adaptive optima (Butler & King, 2004).

The proposed methodology also allows further elucidation of patterns at the intraspecific level as well as to use this information to predict missing values. In these cases, the phylogeny is provided at the species level but variability in trait and environmental data can be considered by including more than one observation per species. The application of the proposed framework including within-species variability in future work will allow better quantification of the extent to which variability within species is phylogenetically structured and related to specific environmental variables.

### AUTHOR CONTRIBUTIONS

Pablo Sanchez-Martinez, Maurizio Mencuccini, Jordi Martínez-Vilalta, David D. Ackerly and Todd E. Dawson designed the study; Pablo Sanchez-Martinez created the R Package, analysed the data and wrote the first draft of the manuscript. All authors contributed substantially to revisions.

### ACKNOWLEDGEMENTS

This research was supported by competitive grants CGL2017-89149-C2-1-R and PID2021-127452NB-I00 funded by MCIN/AEI/10.13039/501100011033. Pablo Sanchez-Martinez acknowledges an FPU predoctoral fellowship from the Spanish Ministry of Science, Innovation and Universities (grant FPU18/04945). Jordi Martínez-Vilalta benefited from an ICREA Academia award.

### CONFLICT OF INTEREST STATEMENT

Authors have no conflict of interests to declare.

### PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14304>.

## DATA AVAILABILITY STATEMENT

Data available via the Figshare Digital Repository <https://doi.org/10.6084/m9.figshare.25145615.v1> (Sanchez-Martinez et al., 2024). Code available via the Github Repository <https://github.com/pablo-sanchezmart/Sanchez-Martinez-et-al-2024-A-framework-to-study-and-predict-functional-trait-syndromes/upload/main> (<https://doi.org/10.5281/zenodo.10653998>) (Sanchez-Martinez, 2024).

## ORCID

Pablo Sanchez-Martinez  <https://orcid.org/0000-0002-0157-7800>

David D. Ackerly  <https://orcid.org/0000-0002-1847-7398>

Jordi Martinez-Vilalta  <https://orcid.org/0000-0002-2332-7298>

Maurizio Mencuccini  <https://orcid.org/0000-0003-0840-1477>

Kyle G. Dexter  <https://orcid.org/0000-0001-9232-5221>

Todd E. Dawson  <https://orcid.org/0000-0002-6871-3440>

## REFERENCES

- Ackerly, D. (2009). Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 19699–19706.
- Ackerly, D. D., Dudley, S. A., Sultan, S. E., Schmitt, J., Coleman, J. S., Linder, C. R., Sandquist, D. R., Geber, M. A., Evans, A. S., Dawson, T. E., & Lechowicz, M. J. (2000). The evolution of plant ecophysiological traits: Recent advances and future directions. *BioScience*, *50*, 979.
- Ackerly, D. D., & Reich, P. B. (1999). Convergence and correlations among leaf size and function in seed plants: A comparative test using independent contrasts. *American Journal of Botany*, *86*, 1272–1281.
- Anderegg, L. D. L., Griffith, D. M., Cavender-Bares, J., Riley, W. J., Berry, J. A., Dawson, T. E., & Still, C. J. (2021). Representing plant diversity in land models: An evolutionary approach to make 'functional types' more functional. *Global Change Biology*, *28*, 2541–2554.
- Bhaskar, R., & Ackerly, D. D. (2006). Ecological relevance of minimum seasonal water potentials. *Physiologia Plantarum*, *127*, 353–359.
- Blomberg, (2016). A primer phylogenetic generalized least squares. In L. Z. Garamszegi (Ed.), *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 105–130). Springer Berlin Heidelberg.
- Blomberg, S. P., & Garland, T. (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology*, *15*, 899–910.
- Blomberg, S. P. P., Garland, T., & Ives, A. R. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*, 717–745.
- Butler, M. A., & King, A. A. (2004). Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, *164*, 683–695.
- Cavender-Bares, J., Ackerly, D. D., Hobbie, S. E., & Townsend, P. A. (2016). Evolutionary legacy effects on ecosystems: Biogeographic origins, plant traits, and implications for management in the era of global change. *Annual Review of Ecology, Evolution, and Systematics*, *47*, 433–462.
- Choat, B., Brodribb, T. J., Brodersen, C. R., Duursma, R. A., López, R., & Medlyn, B. E. (2018). Triggers of tree mortality under drought. *Nature*, *558*, 531–539.
- Crisp, M. D., Arroyo, M. T. K., Cook, L. G., Gandolfo, M. A., Jordan, G. J., McGlone, M. S., Weston, P. H., Westoby, M., Wilf, P., & Linder, H. P. (2009). Phylogenetic biome conservatism on a global scale. *Nature*, *458*, 754–756.
- Crisp, M. D., & Cook, L. G. (2012). Phylogenetic niche conservatism: What are the underlying evolutionary and ecological causes? *New Phytologist*, *196*, 681–694.
- DeBruine, L. (2021). Faux: Simulation for factorial designs R package version 1.1.0. *Zenodo*.
- Felsenstein, J. (1988). Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, *19*, 445–471.
- Flo, V., Martínez-Vilalta, J., Mencuccini, M., Granda, V., Anderegg, W. R. L., & Poyatos, R. (2021). Climate and functional traits jointly mediate tree water-use strategies. *New Phytologist*, *231*, 617–630.
- Flores, O., Garnier, E., Wright, I. J., Reich, P. B., Pierce, S., Diaz, S., Pakeman, R. J., Rusch, G. M., Bernard-Verdier, M., Testi, B., Bakker, J. P., Bekker, R. M., Cerabolini, B. E. L., Ceriani, R. M., Cornu, G., Cruz, P., Delcamp, M., Dolezal, J., Eriksson, O., ... Weiher, E. (2014). An evolutionary perspective on leaf economics: Phylogenetics of leaf mass per area in vascular plants. *Ecology and Evolution*, *4*, 2799–2811.
- García-Valdés, R., Vayreda, J., Retana, J., & Martínez-Vilalta, J. (2021). Low forest productivity associated with increasing drought-tolerant species is compensated by an increase in drought-tolerance richness. *Global Change Biology*, *27*, 2113–2127.
- Goolsby, E. W., Bruggeman, J., & Ané, C. (2017). Rphylopars: Fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, *8*, 22–27.
- Green, J. K., Berry, J., Ciais, P., Zhang, Y., & Gentine, P. (2020). Amazon rainforest photosynthesis increases in response to atmospheric dryness. *Science Advances*, *6*(47), 1–10. <https://doi.org/10.1126/sciadv.abb7232>
- Hadfield, J. (2019). *MCMCglmm course notes*.
- Hadfield, J. D. (2010a). MCMCglmm: MCMC methods for multi-response GLMMs in R. *Journal of Statistical Software*, *33*, 1–22.
- Hadfield, J. D. (2010b). MCMCglmm for R. *Journal of Statistical Software*, *33*, 1–22.
- Holt, R. D. (2009). Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 19659–19665.
- Kattge, J., Bönlisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar, C., Aleixo, I., ... Wirth, C. (2020). TRY plant trait database—Enhanced coverage and open access. *Global Change Biology*, *26*, 119–188.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News*, *2*, 18–22.
- Losos, J. B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, *11*, 995–1003.
- Louca, S., & Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. *Bioinformatics*, *34*, 1053–1055.
- Maherali, H., Pockman, W. T., & Jackson, R. B. (2004). Adaptive variation in the vulnerability of woody plants to xylem cavitation. *Ecology*, *85*, 2184–2199.
- Makowski, D., Ben-Shachar, M., & Lüdtke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*, 1541.
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdtke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 1–14.
- Maynard, D. S., Bialic-Murphy, L., Zohner, C. M., Averill, C., van den Hoogen, J., Ma, H., Mo, L., Smith, G. R., Acosta, A. T. R., Aubin, I., Berenguer, E., Boonman, C. C. F., Catford, J. A., Cerabolini, B. E. L., Dias, A. S., González-Melo, A., Hietz, P., Lusk, C. H., Mori, A. S., ... Crowther, T. W. (2022). Global relationships in tree functional traits. *Nature Communications*, *13*, 1–12.

- Mencuccini, M., Rosas, T., Rowland, L., Choat, B., Cornelissen, H., Jansen, S., Kramer, K., Lapenis, A., Manzoni, S., Niinemets, Ü., Reich, P. B., Schrod, F., Soudzilovskaia, N., Wright, I. J., & Martínez-Vilalta, J. (2019). Leaf economics and plant hydraulics drive leaf: Wood area ratios. *New Phytologist*, 224, 1544–1556.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884.
- Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
- Poyatos, R., Sus, O., Badiella, L., Mencuccini, M., & Martínez-Vilalta, J. (2018). Gap-filling a spatially explicit plant trait database: Comparing imputation methods and different levels of environmental information. *Biogeosciences*, 15, 2601–2617.
- Qian, H., & Jin, Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology*, 9, 233–239.
- Ramírez-Valiente, J. A., & Cavender-Bares, J. (2017). Evolutionary trade-offs between drought resistance mechanisms across a precipitation gradient in a seasonally dry tropical oak (*Quercus oleoides*). *Tree Physiology*, 37, 889–901.
- Reich, P. B. (2014). The world-wide 'fast-slow' plant economics spectrum: A traits manifesto. *Journal of Ecology*, 102, 275–301.
- Reich, P. B., Wright, I. J., Cavender-Bares, J., Craine, J. M., Oleksyn, J., Westoby, M., & Walters, M. B. (2003). The evolution of plant functional variation: Traits, spectra, and strategies. *International Journal of Plant Sciences*, 164, S143–S164.
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40, 913–929.
- Sanchez-Martinez, P. (2024). pablosanchezmart/Sanchez-Martinez-et-al-2024-A-framework-to-study-and-predict-functional-trait-syndromes: A framework to study and predict functional trait syndromes using phylogenetic and environmental data (1.0). *Zenodo*, <https://doi.org/10.5281/zenodo.10653998>
- Sanchez-Martinez, P., Ackerly, D. D., Martínez-Vilalta, J., Mencuccini, M., Dexter, K. G., & Dawson, T. E. (2024). Data from: A framework to study and predict functional trait syndromes using phylogenetic and environmental data. *figshare*, <https://doi.org/10.6084/m9.figshare.25145615.v1>
- Sanchez-Martinez, P., Martínez-Vilalta, J., Dexter, K. G., Segovia, R. A., & Mencuccini, M. (2020). Adaptation and coordinated evolution of plant hydraulic traits. *Ecology Letters*, 23, 1599–1610.
- Sanchez-Martinez, P., Mencuccini, M., García-Valdés, R., Hammond, W. M., Serra-Diaz, J. M., Guo, W. Y., Segovia, R. A., Dexter, K. G., Svenning, J. C., Allen, C., & Martínez-Vilalta, J. (2023). Increased hydraulic risk in assemblages of woody plant species predicts spatial patterns of drought-induced mortality. *Nature Ecology & Evolution*, 7, 1620–1632.
- Scoffoni, C., Chatelet, D. S., Pasquet-Kok, J., Rawls, M., Donoghue, M. J., Edwards, E. J., & Sack, L. (2016). Hydraulic basis for the evolution of photosynthetic productivity. *Nature Plants*, 2, 1–8.
- Stekhoven, D. J., & Bühlmann, P. (2012). Missforest-non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28, 112–118.
- Swenson, N. G. (2014). Phylogenetic imputation of plant functional trait databases. *Ecography*, 37, 105–110.
- Trabucco, A., & Zomer, R. J. (2018). Global aridity index and potential evapotranspiration (ETO) climate database v2. *CGIAR Consortium for Spatial Information (CGIAR-CSI)*, 10, 1–10.
- Trugman, A. T., Anderegg, L. D. L., Shaw, J. D., & Anderegg, W. R. L. (2020). Trait velocities reveal that mortality has driven widespread coordinated shifts in forest hydraulic trait composition. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 8532–8538.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45, 1–67.
- Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 116, 882–892.
- Volaire, F. (2018). A unified framework of plant adaptive strategies to drought: Crossing scales and disciplines. *Global Change Biology*, 24, 2929–2938.
- Wright, I. J., Dong, N., Maire, V., Prentice, I. C., Westoby, M., Díaz, S., Gallagher, R. V., Jacobs, B. F., Kooyman, R., Law, E. A., Leishman, M. R., Niinemets, Ü., Reich, P. B., Sack, L., Villar, R., Wang, H., & Wilf, P. (2017). Global climatic drivers of leaf size. *Science*, 357, 917–921.
- Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., ... Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, 428, 821–827.
- Zheng, J., Zhao, X., Morris, H., & Jansen, S. (2019). Phylogeny best explains latitudinal patterns of xylem tissue fractions for Woody angiosperm species across China. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00556>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Figure S1:** Imputation framework.

**Figure S2:** Simulated data scheme.

**Figure S3:** Predictive performance imputeTraits different rounds.

**Figure S4:** Predictive performance cross validation 50–50.

**Table S1:** Variance-covariance matrix used to simulate traits covariation (either phylogenetic or non-phylogenetic).

**Table S2:** Simulated data characteristics and expectations.

**How to cite this article:** Sanchez-Martinez, P., Ackerly, D. D., Martínez-Vilalta, J., Mencuccini, M., Dexter, K. G., & Dawson, T. E. (2024). A framework to study and predict functional trait syndromes using phylogenetic and environmental data. *Methods in Ecology and Evolution*, 15, 666–681. <https://doi.org/10.1111/2041-210X.14304>