ECOGRAPHY

Research

Assessing among-lineage variability in phylogenetic imputation of functional trait datasets

Rafael Molina-Venegas, Juan Carlos Moreno-Saiz, Isabel Castro Parga, T. Jonathan Davies, Pedro R. Peres-Neto and Miguel Á. Rodríguez

R. Molina-Venegas (http://orcid.org/0000-0001-5801-0736) (rmolina@us.es) and M. Á. Rodríguez, Depto de Ciencias de la Vida, Univ. de Alcalá, Madrid, Spain. Present address of RM-V, Inst. of Plant Sciences, Univ. of Bern, Bern, Switzerland. – J. C. Moreno-Saiz, Depto de Biología (Botánica), Univ. Autónoma de Madrid, Madrid, Spain. – I. Castro Parga, Depto de Ecología, Univ. Autónoma de Madrid, Madrid, Spain. – T. J. Davies, Dept of Biology, McGill Univ., Montreal, QC, Canada. – P. R. Peres-Neto, Dept of Biology, Concordia Univ., Montreal, QC, Canada.

Phylogenetic imputation has recently emerged as a potentially powerful tool for predicting missing data in functional traits datasets. As such, understanding the limitations of phylogenetic modelling in predicting trait values is critical if we are to use them in subsequent analyses. Previous studies have focused on the relationship between phylogenetic signal and clade-level prediction accuracy, yet variability in prediction accuracy among individual tips of phylogenies remains largely unexplored. Here, we used simulations of trait evolution along the branches of phylogenetic trees to show how the accuracy of phylogenetic imputations is influenced by the combined effects of 1) the amount of phylogenetic signal in the traits and 2) the branch length of the tips to be imputed. Specifically, we conducted cross-validation trials to estimate the variability in prediction accuracy among individual tips on the phylogenies (hereafter 'tip-level accuracy'). We found that under a Brownian motion model of evolution (BM, Pagel's $\lambda = 1$), tip-level accuracy rapidly decreased with increasing tip branch-lengths, and only tips of approximately 10% or less of the total height of the trees showed consistently accurate predictions (i.e. cross-validation R-squared >0.75). When phylogenetic signal was weak, the effect of tip branch-length was reduced, becoming negligible for traits simulated with $\lambda < 0.7$, where accuracy was in any case low. Our study shows that variability in prediction accuracy among individual tips of the phylogeny should be considered when evaluating the reliability of phylogenetically imputed trait values. To address this challenge, we describe a Monte Carlo-based method that allows one to estimate the expected tip-level accuracy of phylogenetic predictions for continuous traits. Our approach identifies gaps in functional trait datasets for which phylogenetic imputation performs poorly, and will help ecologists to design more efficient trait collection campaigns by focusing resources on lineages whose trait values are more uncertain.

Keywords: branch-lengths, missing data, phylogenetic signal

www.ecography.org

Introduction

Species traits are commonly used to address many ecological and evolutionary questions, such as the processes underlying ecological assemblages (Götzenberger et al. 2012, Kraft et al. 2015), the links between species functional diversity and ecosystem functioning (Cadotte et al. 2011, Valencia-Gómez et al. 2015), the tempo and mode of phenotypic evolution (Felsenstein 1985, Martins 1994, Ackerly 2009) and the distribution of species and ecosystems properties across space and time (Violle et al. 2014). However, gathering trait information is often challenging because the collection of functional trait data is a time- and resource-consuming task. This hurdle is particularly noticeable at the macroecological scale, where trait information for hundreds or thousands of species may be required. Indeed, even one of the largest and most comprehensive functional trait databases compiled to date (i.e. the TRY Plant Trait Database, Kattge et al. 2011) is highly incomplete for many species (Sandel et al. 2015). As such, trait-based studies often deal with missing data.

A common way to proceed when missing values are present is to drop species with missing data and use those values that are known. This procedure would lead to unbiased results if data are missing at random, although statistical power may be reduced due to decreased sample size (Nakagawa and Freckleton 2008). However, data are rarely missing completely at random, since the presence of missing values for a given trait (e.g. seed mass) is often related to the values of another trait (e.g. rarity) (Sandel et al. 2015). In such cases, this can lead to spurious results in comparative studies (González-Suárez et al. 2012, Pakeman 2014) and likely other trait-based analyses. An alternative approach is to impute (via prediction) missing values using other known variables as predictors (Penone et al. 2014, Dray and Josse 2015, Schrodt et al. 2015). Unfortunately, this also poses a challenge in cases where the relationship between missing values and potential predictive variables is weak.

Phylogenetic imputation has emerged as a potentially powerful framework for estimating missing data in functional trait datasets (Swenson 2014). The underlying principle is to model the evolution of a given trait on a phylogeny based on the species trait values available, and then use model parameters to predict the missing values. Phylogenetic imputation has several advantages over non-phylogenetic imputation. First, phylogenetic imputation can be applied to any kind of missing data. Second, traits can be imputed independent of other predictive variables as long as they show some degree of phylogenetic signal. Third, phylogenetic imputation can be conducted even on species for which biological information is scarce (e.g. very rare or even extinct species), assuming their phylogenetic position is known. Given the rapid increase in available molecular data, published phylogenies and major advances in phylogenetic methods, phylogenetically-informed imputation is rapidly becoming a common tool in ecology (Guénard et al. 2014, 2015, Penone et al. 2014, Bland et al. 2015, Goberna and Verdú 2016,

Swenson et al. 2017). Although it has been suggested that phylogenetically-imputed values may introduce high levels of error in local-scale studies (e.g. community ecology), phylogenetic imputation remains a promising tool for exploring macroecological patterns (Swenson et al. 2017).

A common procedure to estimate the reliability of phylogenetically-imputed values is to apply leave-one-out cross-validation (Guénard et al. 2013, 2015). The process begins by pruning all phylogenetic tips (e.g. species) with missing values in respect to the particular trait of interest. Then, each observed value for which trait information exists is dropped one at a time (target species), and re-estimated using information from the remaining species (species used to model trait variation as a function of phylogeny). Finally, a comparison between the observed and predicted values can be performed using the cross-validation R-squared (R^2_{cv}) (or other similar error calculation method, Oba et al. 2003, Penone et al. 2014); the higher the R^2_{cv}, the more reliable the imputed trait values (see Guénard et al. 2015 for details). However, it is important to note that the R^2_{cv} provides an index of the overall accuracy of predictions conducted across the complete set of known values (hereafter 'trait-level accuracy'), but variability in prediction accuracy among individual tips of the phylogeny (in most cases species) may pass unnoticed (which we refer to as 'tip-level accuracy' hereafter). There are reasons why one should expect accuracy to vary considerably across tips. For example, the distribution of terminal branch lengths in phylogenetic trees is often highly skewed (Paradis 2016), and thus some species may be more phylogenetically isolated than others. If predictions made on short terminal branches are more accurate than those on longer terminal branches, then phylogenies showing a disproportionate number of recently-evolved species-rich clades (i.e. many short terminal branches) may return high R^2_{cv} values, though individual predictions conducted on long terminal branches might be poor.

In addition to tip-level variability in imputation accuracy, we might also expect trait-level accuracy to vary among different traits. Phylogenetic signal in trait data is the most obvious factor affecting trait-level accuracy (Goberna and Verdú 2016). Previous studies have shown that traits with little phylogenetic signal are error prone (Swenson et al. 2017); however, the combined effect of variability in prediction accuracy among individual tips and phylogenetic signal in the traits remains largely unexplored. In this paper, we use extensive simulations of trait evolution along simulated phylogenetic trees to show how the accuracy of phylogenetic imputations is influenced by the combined effects of variation in: 1) the amount of phylogenetic signal in the traits and 2) the branch length of the tips to be imputed. Our goal was to determine the set of conditions under which phylogenetic imputation performs well, and those conditions where it performs poorly. We then describe a Monte Carlo-based method that allows one to determine gaps in functional trait datasets that are well-suited to being filled using phylogenetic imputation.

We evaluate the performance of the method using a species-level time-calibrated molecular phylogeny of European and North American tree species together with an empirical trait dataset.

Material and methods

Phylogenetic imputation methods for quantitative traits, an overview

To date, three different families of methods have been put forward for phylogenetic imputation: phylogenetic generalized linear models (pGLM; Swenson 2014, Goolsby et al. 2016a), phylogenetic eigenvector regression models (Diniz-Filho et al. 1998) and phylogenetic eigenvector maps (Guénard et al. 2013).

pGLM estimates the optimal phylogenetic variance–covariance matrix for the observed values by either considering simple transformations of the matrix (e.g. Pagel's lambda) or fitting specific models of evolution (e.g. Ornstein–Uhlenbeck, early-burst) from which missing values can be then estimated (see Swenson 2014 for details). This approach has been implemented in the R package Rphylopars (Goolsby et al. 2016b) and only requires specifying the appropriate phylogenetic variance–covariance transformation or evolutionary model. Although the model that best fits the data is usually unknown, it can be determined through a heuristic search using maximum likelihood (see Freckleton et al. 2002, 2011 for details).

Phylogenetic eigenvector regression models (PVR) (Diniz-Filho et al. 1998, 2011) for imputation begin by extracting the principal coordinates of the pairwise phylogenetic distance matrix between species (generating $n-1$ phylogenetic eigenvectors where n equal the number of terminal tips). The most relevant eigenvectors in explaining the observed trait variation are then used as explanatory variables in a multiple linear regression, with the trait of interest as the response variable (see Diniz-Filho et al. 2012 for a review on model selection methods for phylogenetic eigenvectors). Note that the regression model only includes species for which traits are known. Finally, eigenvector values of the target species are used within the regression model to predict missing values (see Swenson 2014 for details). It is important to note that unlike pGLM, PVR does not assume an explicit model of trait evolution.

Finally, phylogenetic eigenvector mapping (PEM) combines both evolutionary dynamics and information on topology (Guénard et al. 2013). PEM shares some similarities with PVR and, as such, it was conceived to improve over PVR because it additionally considers underlying evolutionary models (Diniz-Filho et al. 2015). In PEM, the topology of the phylogeny is first coded as a binary influence matrix representing ancestor–descendant relationships. This matrix is then transformed according to a given evolutionary model to represent trait change dynamics. Information on trait evolutionary dynamics along the branches of the phylogenetic tree is represented using edge weights. Each branch is assigned a weight $W_{a,\psi}$ proportional to the extent of the change that is expected (according to an assumed evolutionary model) to occur along that branch, where a ($0 \leq a \leq 1$) is the steepness parameter, and ψ ($0 < \psi < \infty$) is the relative evolutionary rate of the trait being modelled. This allows the assignment of different function parameters to different portions of the phylogeny, and thus heterogeneous evolutionary dynamics (Beaulieu et al. 2012, Revell et al. 2012) can be also accommodated. The steepness parameter a is related to Pagel's (1999) κ ($a = 1 - \kappa$), and describes how abrupt the changes in trait values occur through time after any given branching event. When a is set to 0, the expected amount of trait change along the branch is proportional to the square root of the branch length (i.e. Brownian motion; Vienne et al. 2011), whereas when a is set to 1 changes occur at a fixed rate ψ whenever species diverge irrespective of the branch length. Phylogenetic eigenvector maps (PEMs) are obtained by weighting and centring the final influence matrix, and then the most important eigenvectors in explaining the observed trait variation can be used to predict missing values as explained for PVR above. The PEM framework is implemented in the R package MPSEM (Guénard and Legendre 2014).

Because PEM is superior to PVR in that it accommodates different evolutionary models, here we focus on PEM and pGLM (as implemented in the R package Rphylopars; Goolsby et al. 2016b).

The influence of phylogenetic signal and tips branch-lengths on phylogenetic imputation

Simulation of phylogenetic trees
We used the function 'pbtree' in the phytools R package (Revell 2012) to generate a pool of 10 000 simulated, pure-birth ultrametric phylogenies of size n = 100. The height of the trees (i.e. root to tip length) was constrained to 1 in all cases. From this pool, we first randomly selected n = 50 phylogenies to evaluate the different imputation methods. Skewed or discontinuous distributions of terminal branches (i.e. overrepresentation of short terminal branches and a few long branches, Paradis 2016) may mislead the interpretation of the effect of tip branch-lengths on prediction accuracy, because lack of variability and/or continuity in branch-lengths. We thus illustrate our results (i.e. Fig. 2 and Supplementary material Appendix 1 Fig. A1) on the phylogeny in the simulated pool that displayed the maximum evenness in the distribution of tip branch-lengths (i.e. similarity in the number of long versus short terminal branches). We estimated evenness in tip branch-lengths by first grouping the tips of each phylogeny into ten equitable branch-length categories (i.e. terminal branches of length between 0–10% of the tree height, 10–20%, 20–30%, and so on), and then calculated Pielou's evenness index for each phylogeny, retaining the phylogeny having the maximum evenness value. Results from the full sample of 50 phylogenies are provided in Supplementary material Appendix 1 Fig. A2. We also explored the accuracy of predictions using pure-birth ultrametric phylogenies for

n = 50 and n = 200 tips. Finally, we tested for bias in imputation accuracy with tree shape by comparing results on the tree topologies that showed the minimum and maximum values for the gamma statistic (tree steminess; Pybus and Harvey 2000) and the Colless' index (tree imbalance; Mooers and Heard 1997), respectively.

Simulation of quantitative traits

We simulated traits with different levels of phylogenetic signal using a gradient varying from complete lack of phylogenetic signal (Pagel's $\lambda = 0$) to Brownian motion (Pagel's $\lambda = 1$) by rescaling the phylogenetic variance–covariance matrices using Pagel's λ. Rescaling is achieved by multiplying all values off the main diagonal of the covariance matrix by λ. Pagel's λ was varied as follows: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.90, 0.95 and 1. Trait evolution was then simulated along the branches of the rescaled phylogenies following a pure Brownian motion model (BM) of evolution (root value $a = 0$ and instantaneous variance $\sigma^2 = 1$). We only retained traits that demonstrated a phylogenetic signal within $\lambda_x \pm 0.01$, where λ_x is the original λ under which traits were simulated. In the case of $\lambda_x = 0$, we considered all traits that showed a value of the statistic $\lambda < 0.01$. Simulations were conducted iteratively until n = 500 traits per phylogeny were obtained. All trait simulations were conducted using the functions 'fastBM' and 'phylosig' in the phytools R package (Revell 2012) and 'rescale' in the geiger R package (Harmon et al. 2008).

Phylogenetic imputation of simulated data

We conducted leave-one-out cross-validation for each simulated trait separately for both PEM and pGLM. For each tip value i for a given trait, we dropped the value of i (target value) and used the information from the remaining $n-1$ elements (model values) to estimate model parameters and predicted value for i. For PEM, we estimated a single steepness parameter a for the model values in the cross-validation trials using restricted maximum likelihood (function 'PEM.fitSimple' in MPSEM R package). The relative evolution rate ψ was set to 1 in all cases following the suggestion of Guénard et al. (2013). Note that ψ has no effect when its value is assumed to be constant across the phylogeny. We then selected the subset of eigenvectors that minimized information loss using a forward stepwise procedure (function 'lmforwardsequentialAICc' in the MPSEM R package), and retained the regression model with the lowest AICc score (Guénard et al. 2013). Finally, the eigenvector values of the target species were used in the regression model to impute its value.

Phylogenetic imputations using pGLM were generated using the function 'phylopars' as implemented in the R package Rphylopars (Goolsby et al. 2016b). For each tip value i for a given trait, we dropped the value of i and fitted the most appropriate lambda model to the remaining $n-1$ values, and then we used the model parameters to inform the

phylogenetic variance–covariance matrix and estimate the predicted value i.

Estimation of the accuracy of predictions

To assess the influence of phylogenetic signal on phylogenetic imputations, we estimated the overall accuracy of predictions for each cross-validation trial (trait-level accuracy) using the cross-validation R-squared:

$$R_{cv}^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 / n}{s_y^2} \tag{1}$$

where $\hat{y}_i$ is the predicted value for target tip i for any given trait, y_i the simulated 'observed' value for target tip i, n the number of phylogenetic tips, and s_y^2 is the variance of the simulated 'observed' values for the trait. $R_{cv}^2 = 1$ indicates perfect match between predicted and observed values, values $1 > R_{cv}^2 > 0$ indicate imperfect predictions, and $R_{cv}^2 = 0$ are produced when the mean square prediction error equals the mean square deviation from the mean (i.e. the sample variance of the simulated 'observed' values), in which case the model is no better in predicting values than simply assuming the trait mean across all tips (Guénard et al. 2015).

To assess the influence of tip branch-lengths on phylogenetic imputation, we estimated the accuracy of predictions for each individual tip on the phylogenies (tip-level accuracy) using the cross-validation R-squared coefficient as described above (but note the value of n in estimates of tip-level R_{cv}^2 values is equal to 1, Eq. 1). The observed distributions of tip-level R_{cv}^2 values were highly skewed (Supplementary material Appendix 1 Fig. A1). Therefore, we summarized tip-level accuracy as the fraction of tip-level R_{cv}^2 values (n = 500 per tip and phylogenetic signal scenario) that were higher than a given threshold, here 0.75 (hereafter '$P_{0.75}$'). This threshold was set to provide a good contrast among and within phylogenies.

A Monte Carlo-based method to estimate the expected accuracy of phylogenetic imputation

We propose a Monte Carlo-based simulation framework as a pipeline to determine gaps in functional trait datasets that are suited to being filled via phylogenetic imputation. Conversely, the procedure also allows to assess which species are prone to greater levels of uncertainty regarding their imputation. Given a fully-resolved ultrametric phylogeny and a quantitative trait with missing values (Fig. 1), the latter are pruned (step 1 in Fig. 1), and the phylogenetic signal in the observed values of the trait (hereafter λ_x) is estimated using Pagel's λ (Pagel 1999) (step 2). Note that the 'true' phylogenetic signal of the trait (i.e. including missing values) is assumed to be approximately equal to that estimated for the observed values of the trait (i.e. based on the species for which traits are known). To determine the expected tip-level accuracy of missing values, the phylogenetic variance–covariance matrix is rescaled using λ_x (step 3), and traits are simulated
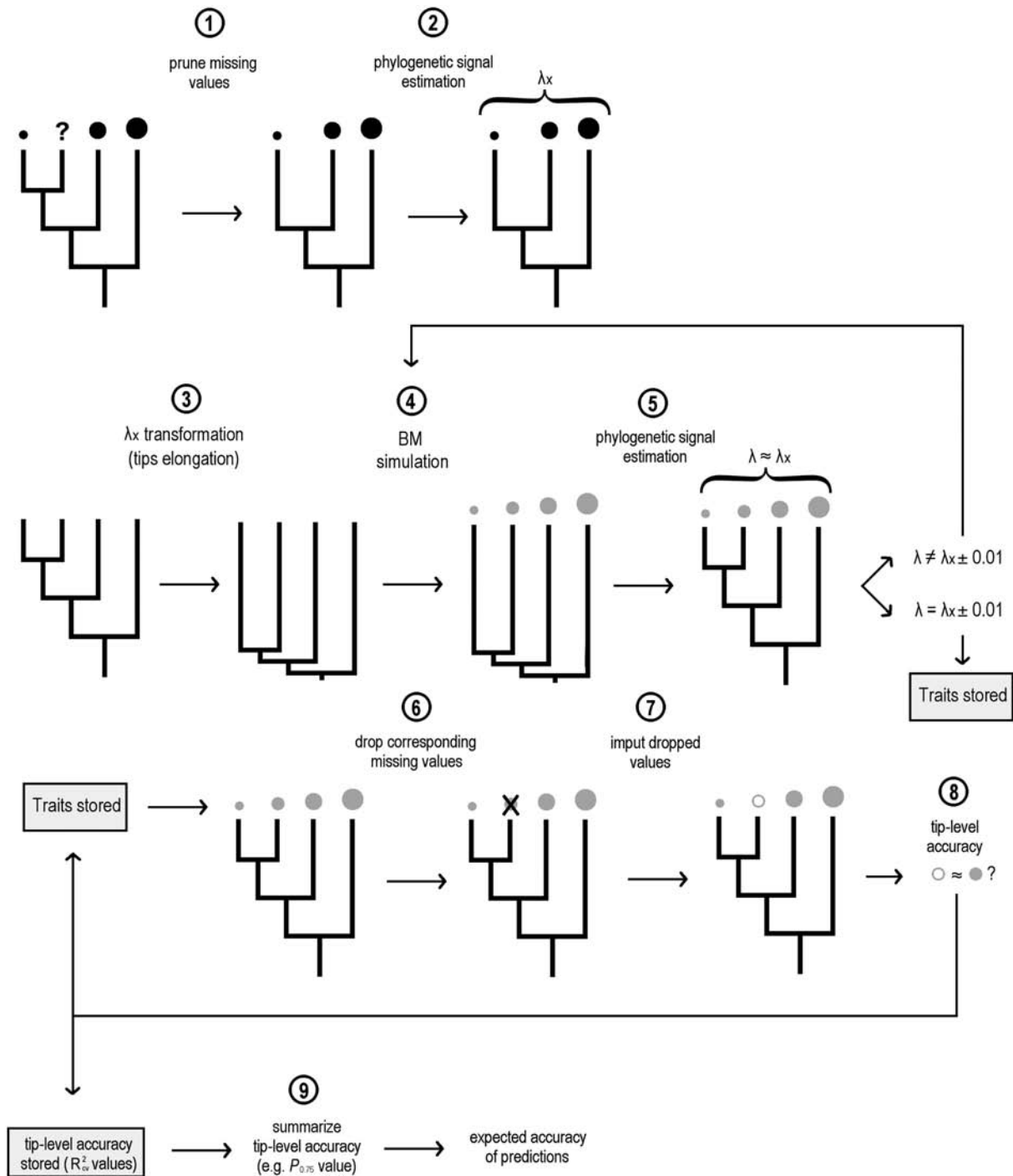
Figure 1. Workflow of the Monte Carlo-based method. Given a phylogeny and a quantitative trait with missing values, the latter are pruned (step 1), and the phylogenetic signal in the observed values of the trait (λ_x) is estimated using Pagel's λ (step 2). The phylogenetic variance–covariance matrix is then rescaled using λ_x (step 3), and a high number of traits are simulated along the branches of the rescaled phylogeny following a pure Brownian motion model (BM) of evolution (step 4). If the value of λ in the simulated trait is approximately equal to λ_x (i.e. $\lambda = \lambda_x \pm 0.01$), the corresponding trait is stored (step 5); otherwise, the workflow goes back to step 4. For each simulated trait, the values corresponding to the phylogenetic placement of missing empirical values are dropped (step 6) and subsequently imputed using the remaining values (step 7). Finally, the accuracy of each individual prediction is measured using the cross-validation R-squared (R^2_{cv}) (step 8), and the distributions of R^2_{cv} values are summarized to obtain an estimate of the expected accuracy of predictions for each tip (step 9).

along the branches of the rescaled phylogeny following a pure Brownian motion model (BM) of evolution (step 4). To ensure that the simulated traits are close to the desired phylogenetic signal (i.e. λ_x), only those that show a phylogenetic signal λ within the bounds of an a priori defined range of values around λ_x are retained (e.g. $\lambda_x \pm 0.01$) (step 5). Traits simulations are conducted iteratively until a high number of traits (e.g. n = 500) are obtained. For each simulated trait, values corresponding to the phylogenetic placement of missing (target) values in the empirical dataset are dropped (step 6) and subsequently imputed using the remaining values (step 7). Finally, the accuracy of each individual prediction (i.e. tip-level accuracy, n = 500 predictions per tip) is measured using the cross-validation R-squared (R^2_{cv}) (step 8), and the distributions of R^2_{cv} values (n = 500 per tip) summarized to obtain an estimate of the expected accuracy of predictions for each tip (step 9). The method is fully implemented in the R environment (R Development Core Team) and an easy-to-use function is provided in Supplementary material Appendix 3 where full details can be found.

Evaluating the performance of the Monte Carlo method using empirical data

To assess the performance of the method, we first constructed a time-calibrated molecular phylogeny for the native trees of Europe and North America following the pipeline of Roquet et al. (2013) (see Supplementary material Appendix 2 for full details on the phylogenetic procedure), and matched this to data on three key functional traits related to persistence, regeneration and dispersal provided by the LEDA Traitbase (Kleyer et al. 2008). Specifically, we compiled information on seed mass (SM), canopy height (CH) and specific leaf area (SLA), and pruned the phylogeny to include only species that had at least one available measurement for each trait (n = 69 species, 61 angiosperms and 8 gymnosperms). When multiple measurements were available for a species, we took the species mean. Trait values were log-transformed to fit assumptions of normality. The phylogenetic signal in the traits was calculated using Pagel's λ (λ = 0.97, 0.82 and 0.65 for SM, CH and SLA respectively; p < 0.001 based on a likelihood ratio test for all traits).

For each empirical trait, we sampled 30, 60 and 90% of the species at random 1000 times, and dropped the trait values corresponding to the placement of sampled (target) species in each random draw. This procedure resulted in three pools of incomplete traits (n = 1000 traits per pool) for each empirical trait. We then estimated the expected accuracy of predictions for the target species using the method described above (Fig. 1). The bounds for the λ statistic of the simulated traits (n = 500 simulations per trait in the sample pool) were set at $\lambda_x \pm 0.01$, where λ_x is the phylogenetic signal in the observed (non-missing) values of the traits. We calculated Pearson's correlation coefficients between the predictions conducted on the target species and their actual values, and regressed these coefficients against the average expected accuracy of predictions (i.e. mean $P_{0.75}$ values) using simple linear

models. Finally, we conducted leave-one-out cross-validations for each empirical trait and estimated the overall accuracy of predictions across all tips (i.e. trait-level accuracy).

Data deposition

Results

Phylogenetic imputation of simulated data

The average trait-level accuracy (i.e. mean trait-level R^2_{cv} value) for λ = 1 ranged between 0.61 and 0.79 (n = 50 phylogenies, n = 500 traits per phylogeny, Supplementary material Appendix 1 Table A1). As expected, trait-level accuracy decreased progressively with lower phylogenetic signal, becoming very poor (mean R^2_{cv} values close to 0) at $\lambda < 0.6$. Although both phylogenetic imputation methods yielded qualitatively similar results for trait-level accuracy, pGLM showed slightly more accurate predictions for traits simulated with $\lambda < 0.9$ (Supplementary material Appendix 1 Table A1).

Tip-level accuracy was strongly and negatively related to tip branch-lengths, with the shape of the relationship being influenced by strength of phylogenetic signal (Fig. 2 and Supplementary material Appendix 1 Fig. A2). Under a scenario of strong phylogenetic signal, λ = 1, $P_{0.75}$ values (i.e. tip-level accuracy) rapidly decreased with increasing tip branch-lengths, and were overall lower than 0.5 for tips longer than 20 to 30% of the total height of the trees (Fig. 2 and Supplementary material Appendix 1 Fig. A2). For weaker phylogenetic signal, the effect of tip branch-length on accuracy was less pronounced, becoming negligible for traits simulated with $\lambda < 0.7$. We did not observe any systematic bias in the relationship between tip-level accuracy ($P_{0.75}$) and tip branch-lengths with the size of the phylogeny or tree shape (Supplementary material Appendix 1 Fig. A3 and Fig. A4), although $P_{0.75}$ values derived from the highly-unbalanced tree seemed to decrease slightly more sharply with increasing tip branch-lengths (Supplementary material Appendix 1 Fig. A4). Both phylogenetic imputation methods yielded similar results for tip-level accuracy (see Supplementary material Appendix 1 Fig. A5 for results obtained with pGLM).

Phylogenetic imputation of empirical traits

As expected, we found a positive relationship between the average expected accuracy of predictions (i.e. mean $P_{0.75}$ values, hereafter referred as to 'expected accuracy') and the correlation between predicted and actual trait values (hereafter referred as to 'actual accuracy') (Fig. 3). This was particularly
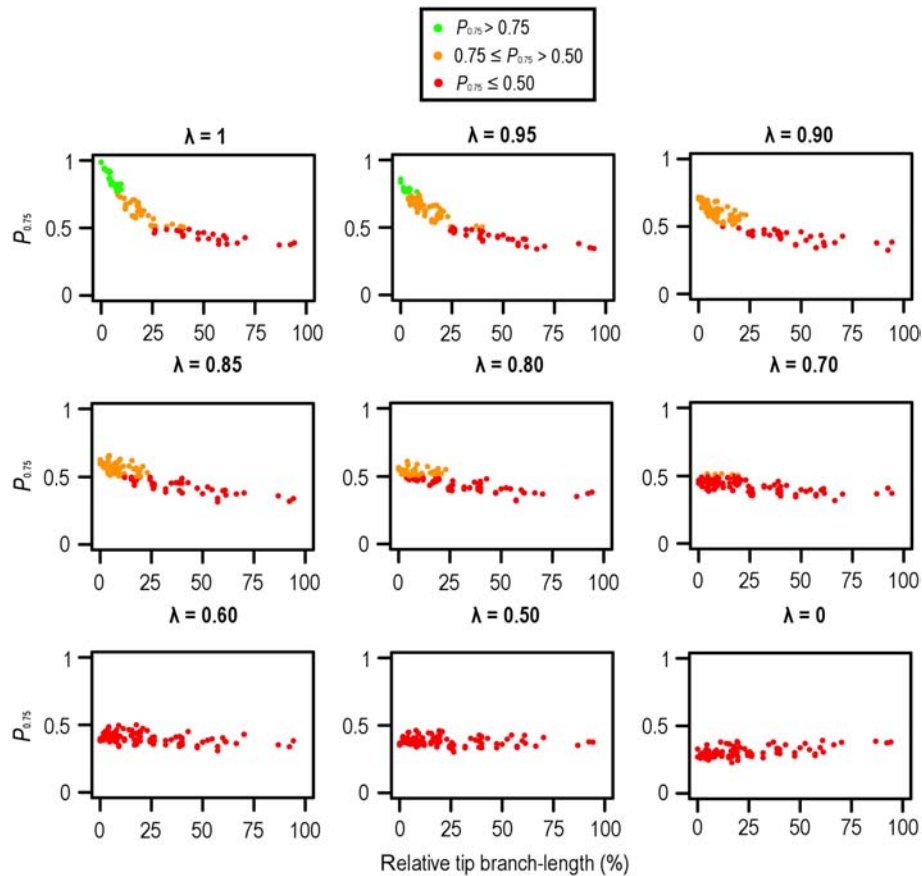
Figure 2. Scatter plots showing the relationship between tip-level accuracy of predictions conducted for simulated traits ($P_{0.75}$, y-axes) and tip branch-lengths (x-axes) across varying scenarios of phylogenetic signal. Trait simulations (n = 500 per phylogenetic signal scenario) were conducted along the branches of the phylogeny that showed maximum evenness in the distribution of tip branch-lengths (see text). Tip-level accuracy was summarized for each tip and phylogenetic signal scenario as the fraction of tip-level R^2_{cv} values that were higher than a certain threshold value set to 0.75 ($P_{0.75}$). Tips are sorted in increasing order of length (relative to the total height of the tree).

noticeable for seed mass (SM) and scenarios where 30% and 60% of trait values were missing ($R^2 = 0.05$; $p < 0.001$, $R^2 = 0.10$; $p < 0.001$ respectively) (Fig. 3a). However, the relationship weakened when 90% of trait values were missing ($R^2 = 0.01$; $p < 0.001$) and prediction intervals were in any case wide. Both the expected and the actual accuracy of predictions decreased as the amount of missing data increased (Table 1). For specific leaf area (SLA), the relationship

between expected and actual accuracy was also significantly positive, though their values were overall lower (Fig. 3b). In contrast, canopy height (CH) showed very weak and non-significant relationships ($R^2 < 0.005$ and $p > 0.1$ in all cases) and the actual accuracy was the lowest among the empirical traits (Table 1). Trait-level accuracy for SM, SLA and CH was 0.73, 0.37 and 0.42, respectively.

Discussion

Phylogenetic imputation is emerging as a promising tool for filling gaps in functional trait datasets by using ancestor–descendant relationships and known trait values for related species. Given the rapid increase in the availability of published phylogenetic trees for species-rich taxonomic groups (Zanne et al. 2014), and large, but gappy, trait databases, there is much interest in the potential application of phylogenetic imputation to explore macroecological patterns (Swenson et al. 2017). It is essential, therefore, to better understand under which circumstances phylogenetic imputation provides reli-

Table 1. Centrality metrics for the distribution of mean $P_{0.75}$ values (average expected accuracy) and Pearson's r coefficients between predicted and actual trait values (actual accuracy) for predictions conducted on empirical traits. The percentages indicate the amount of missing values in the data.

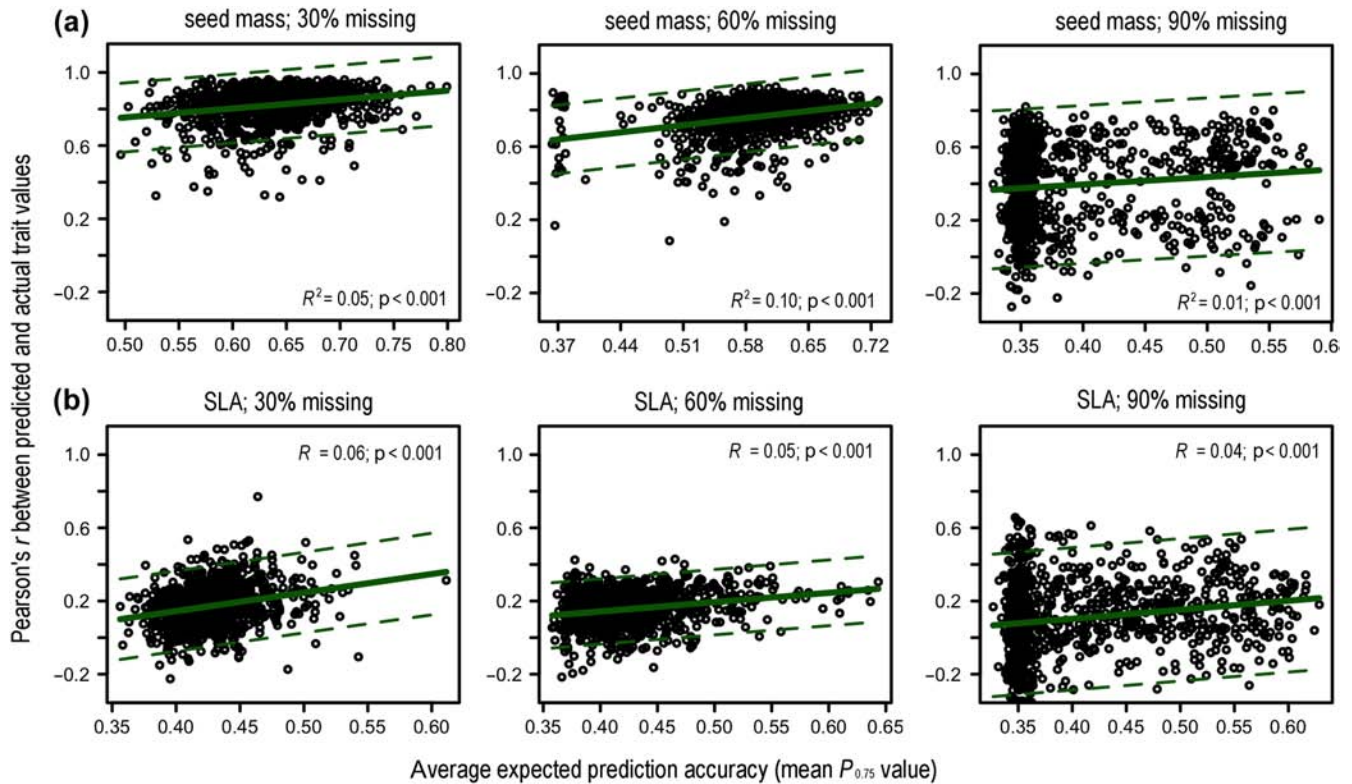|	Mean $P_{0.75}$ values (median)			Pearson's r coefficients (mean)		
Trait	30%	60%	90%	30 %	60%	90%
Seed size	0.63	0.59	0.36	0.82	0.76	0.39
Canopy height	0.47	0.44	0.35	−0.005	−0.02	−0.04
SLA	0.43	0.42	0.37	0.17	0.16	0.11

Figure 3. Scatter plots showing the relationship between average expected accuracy (mean $P_{0.75}$ values) and actual accuracy (i.e. Pearson correlation between predicted and actual trait values) of predictions conducted on empirical traits (i.e. seed mass and specific leaf area) measured for European and North American tree species. Solid and dashed lines represent the regression slope and 95% prediction intervals, respectively.

able estimates of missing species values. In contrast to previous studies (Guénard et al. 2013, 2015, Swenson et al. 2017), we focused on tip-level accuracy of imputed values which show large variation across phylogenetic tips.

Our simulations show how tip-level accuracy of predictions is largely determined by the combined effects of trait phylogenetic signal and the branch length of the tips being imputed. High accuracy is generally achieved for traits with relatively high phylogenetic signal ($\lambda = 1$) and for short tips (those less than approximately 10% of the height of the trees), which subtend species that have had little time for independent evolution. For traits with relatively weak phylogenetic signal (i.e. $\lambda < 0.7$), the accuracy of tip-level predictions was rather low across all tips, irrespective of their lengths. This suggests that phylogenetically imputed values of trait datasets should be interpreted with caution given that 1) most functional traits used in ecological research exhibit a phylogenetic signal $\lambda < 1$ (Liu et al. 2015, Swenson et al. 2017) and 2) the distribution of terminal branch lengths in most phylogenetic trees is seldom uniform (Paradis 2016). For example, although phylogenetic imputation may be useful to predict missing values for highly conserved traits (e.g. seed mass) in 'stemmy' lineages (i.e. clades with short terminal branches), predictions conducted on more 'tippy' lineages (i.e. clades with long terminal branches) may

be unreliable. Importantly, 'tippy' topologies may be particularly prevalent in regional phylogenies where sampling of lineages is often incomplete (i.e. only the species within a given area are included). Therefore, given a trait dataset with missing values (target species) to be phylogenetically-imputed, we suggest that it is best to include as many close relatives of the target species as possible prior to phylogenetic modelling.

We have shown that the expected accuracy of predictions conducted on highly conserved empirical traits (e.g. seed mass) can be reasonably well anticipated by tip branch-length information (Fig. 2 and Fig. 3). Thus, our Monte Carlo-based approach allows one to detect gaps in functional trait databases that are more suitable to be filled by phylogenetic imputation (Supplementary material Appendix 3). Given that predictions conducted on shorter tips are expected to be overall more accurate (Fig. 2), future trait collection campaigns should be oriented to first complete missing values that are subtended from long terminal branches. It is also important to note that functional trait databases may be biased towards certain groups of species (González-Suárez et al. 2012), and thus missing data may be often clumped on the phylogenies (i.e. all values subtended from a given node are missing). In such cases, the relevant information for prediction accuracy will be the distance to the nearest neighbour with a known

trait value, rather than tip branch-length information itself. When missing data show strong phylogenetic clustering, the accuracy of species' trait predictions may be drastically reduced (Goberna and Verdú 2016), which may explain the high incidence of low expected accuracy values for scenarios where 90% of trait values were missing (Fig. 3). Thus, future trait collection campaigns should also aim at sampling across major lineages with high phylogenetic clustering of missing data.

Our trait simulations assumed a single rate of evolution across the phylogeny (i.e. instantaneous variance of the BM process), which is the optimal scenario for phylogenetic imputation. However, this assumption may not hold for most empirical traits, and especially for clades containing large number of species, where a combination of factors may affect evolutionary rates of phenotypic change across lineages. For example, it is well-known that niche conservatism (i.e. the tendency for many ecological traits to remain similar over time; Wiens et al. 2010) leads to strong phenotypic resemblance between closely related species or lineages (Harvey and Pagel 1991). However, adaptive radiations within certain lineages (e.g. trait-linked diversification, Verdú and Pausas 2013) will erode this 'background' phylogenetic signal, because closely-related and recently evolved species rapidly diverge in their fundamental niches to allow coexistence (Abrams 1983). We suggest that in such cases high predicted trait-level accuracy from cross-validation exercises may be misleading if estimated from species outside an adaptive radiation, where imputed values are for species within an adaptive radiation. This may result in inaccurate predictions even in cases where there is high expected accuracy. For example, we found similar trait-level accuracy of predictions (i.e. overall accuracy across all tips) for canopy height and specific leaf area (0.47 and 0.42 respectively), yet the former showed substantially lower correlations between predicted and actual trait values (Table 1). This example illustrates a fundamental limitation of current phylogenetic imputation methods, which assume a phylogenetic model with strong absolute explanatory power in explaining the variance in the observed data informs our predictive accuracy across all tips. A possible solution would be to subset the phylogenetic tree to just close relatives of the target species for which imputed values are required. This approach may be particularly helpful when imputing global datasets that often include hundreds or thousands of species from lineages with distinct evolutionary trajectories.

We conclude that phylogenetic imputation as substitute for directly measured trait data may help to fill the gaps in functional trait datasets, but suggest that information on prediction accuracy should be considered in subsequent analyses. For example, uncertainty in tip-level estimates could be included as measurement error in macroecological analyses. The results of this study and our Monte Carlo-based method to assess predictability error for trait imputation will be useful to future imputation exercises, and help ecologists to design more efficient trait collection campaigns by focusing resources on those lineages whose functional trait values are likely poorly estimated by phylogeny.

References

Abrams, P. 1983. The theory of limiting similarity. – Annu. Rev. Ecol. Syst. 14: 359–376.

Ackerly, D. 2009. Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. – Proc. Natl Acad. Sci. USA 106: 19699–19706.

Beaulieu, J. M. et al. 2012. Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. – Evolution 66: 2369–2383.

Bland, L. M. et al. 2015. Predicting the conservation status of data-deficient species. – Conserv. Biol. 29: 250–259.

Cadotte, M. W. et al. 2011. Beyond species: functional diversity and the maintenance of ecological processes and services. – J. Appl. Ecol. 48: 1079–1087.

Diniz-Filho, J. A. F. et al. 1998. An eigenvector method for estimating phylogenetic inertia. – Evolution 52: 1247–1262.

Diniz-Filho, J. A. F. et al. 2011. Eigenvector estimation of phylogenetic and functional diversity. – Funct. Ecol. 25: 735–744.

Diniz-Filho, J. A. F. et al. 2012. On the selection of phylogenetic eigenvectors for ecological analyses. – Ecography 35: 239–249.

Diniz-Filho, J. A. F. et al. 2015. The best of both worlds: phylogenetic eigenvector regression and mapping. – Genet. Mol. Biol. 38: 396–400.

Dray, S. and Josse, J. 2015. Principal component analysis with missing values: a comparative survey of methods. – Plant. Ecol. 216: 657–667.

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

Freckleton, R. P. et al. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. – Am. Nat. 160: 712–726.

Freckleton, R. P. et al. 2011. Comparative methods as a statistical fix: the dangers of ignoring an evolutionary model. – Am. Nat. 178: E10–E17.

Goberna, M. and Verdú, M. 2016. Predicting microbial traits with phylogenies. – ISME J. 10: 959–967.

González-Suárez, M. et al. 2012. Biases in comparative analyses of extinction risk: mind the gap. – J Anim. Ecol. 81: 1211–1222.

Goolsby, E. W. et al. 2016a. Fast multivariate phylogenetic comparative methods for missing data and within-species variation. – Methods Ecol. Evol. 8: 22–27.

Goolsby, E. W. et al. 2016b Rphylopars: phylogenetic comparative tools for missing data and within-species variation. – R package ver. 0.2.9.

Götzenberger, L. et al. 2012. Ecological assembly rules in plant communities-approaches, patterns and prospects. – Biol. Rev. 87: 111–127.

Guénard, G. and Legendre, P. 2014. MPSEM: modeling phylogenetic signals using eigenvector maps. – R package ver. 0.2-6.

Guénard, G. et al. 2013. Phylogenetic eigenvector maps: a framework to model and predict species traits. – Methods Ecol. Evol. 4: 1120–1131.

Guénard, G. et al. 2014. Using phylogenetic information and chemical properties to predict species tolerances to pesticides. – Proc. R. Soc. B 281: 20133239.

Guénard, G. et al. 2015. Phylogenetics to help predict active metabolism. – Ecosphere 6: 1–11.

Harmon, L. J. et al. 2008. GEIGER: investigating evolutionary radiations. – Bioinformatics 24: 129–131.

Harvey, P. H. and Pagel, M. 1991. The comparative method in evolutionary biology. – Oxford Univ. Press.

Kattge, J. et al. 2011. TRY – a global database of plant traits. – Global Change Biol. 17: 2905–2935.

Kleyer, M. et al. 2008. The Leda Traitbase: a database of life-history traits of northwest European flora. – J. Ecol. 96: 1266–1274.

Kraft, N. J. B. et al. 2015. Plant functional traits and the multidimensional nature of species coexistence. – Proc. Natl Acad. Sci. USA 112: 797–802.

Liu, H. et al. 2015. Strong phylogenetic signals and phylogenetic niche conservatism in ecophysiological traits across divergent lineages of Magnoliaceae. – Sci. Rep. 5: 12246.

Martins, E. P. 1994. Estimating the rate of phenotypic evolution from comparative data. – Am. Nat. 144: 193–209.

Molina-Venegas, R. et al. 2018. Data from: Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. – Dryad Digital Repository, < http://dx.doi. org/10.5061/dryad.12111/1 >.

Mooers, A. O. and Heard, S. B. 1997. Inferring evolutionary process from phylogenetic tree shape. – Q. Rev. Biol. 72: 31–54.

Nakagawa, S. and Freckleton, R. P. 2008. Missing inaction: the dangers of ignoring missing data. – Trends Ecol. Evol. 23: 592–596.

Oba, S. et al. 2003. A Bayesian missing value estimation method for gene expression profile data. – Bioinformatics 19: 2088–2096.

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

Pakeman, R. J. 2014. Functional trait metrics are sensitive to the completeness of the species' trait data? – Methods Ecol. Evol. 5: 9–15.

Paradis, E. 2016. The distribution of branch lengths in phylogenetic trees. – Mol. Phylogenet. Evol. 94: 136–145.

Penone, C. et al. 2014. Imputation of missing data in life-history trait datasets: which approach performs the best? – Methods Ecol. Evol. 5: 961–970.

Pybus, O. G. and Harvey, P. H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. – Proc. R. Soc. B 267: 2267–2272.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). – Methods Ecol. Evol. 3: 217–223.

Revell, L. J. et al. 2012. A new phylogenetic method for identifying exceptional phenotypic diversification. – Evolution 66: 135–146.

Roquet, C. et al. 2013. Building megaphylogenies for macroecology: taking up the challenge. – Ecography 36: 13–26.

Sandel, B. et al. 2015. Estimating the missing species bias in plant trait measurements. – J. Veg. Sci. 26: 828–838.

Schrodt, F. et al. 2015. BHPMF – a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. – Global Ecol. Biogeogr. 24: 1510–1521.

Swenson, N. G. 2014. Phylogenetic imputation of plant functional trait databases. – Ecography 37: 105–110.

Swenson, N. G. et al. 2017. Phylogeny and the prediction of tree functional diversity across novel continental settings. – Global Ecol. Biogeogr. 26: 553–562.

Valencia-Gómez, E. et al. 2015. Functional diversity enhances the resistance of ecosystem multifunctionality to aridity in Mediterranean drylands. – New Phytol. 206: 660–671.

Verdú, M. and Pausas, J. G. 2013. Syndrome-driven diversification in a Mediterranean ecosystem. – Evolution 67: 1756–1766.

Vienne, D. M. et al. 2011. Euclidean nature of phylogenetic distance matrices. – Syst. Biol. 60: 826–832.

Violle, C. et al. 2014. The emergence and promise of functional biogeography. – Proc. Natl Acad. Sci. USA 111: 13690–13696.

Wiens, J. J. et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. – Ecol. Lett. 13: 1310–1324.

Zanne, A. E. et al. 2014. Three keys to the radiation of angiosperms into freezing environments. – Nature 506: 89–92.

Supplementary material (Appendix ECOG-03480 at < www. ecography.org/appendix/ecog-03480 >). Appendix 1–3.