New Phytologist

Research

Methods

Modeling trait heterogeneity and inferring causal links in the macroevolution of growth habit in eudicot angiosperms

Suman Neupane[1], **Amy E. Zanne**[2], **Frederic Lens**[3,4] and **Josef C. Uyeda**[5]

[1]Department of Biology, Ball State University, Muncie, IN 47306, USA; [2]Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA; [3]Naturalis Biodiversity Centre, Research Group Functional Traits, Darwinweg 2, Leiden, 2333 CR, the Netherlands; [4]Institute of Biology Leiden, Plant Sciences, Leiden University, Sylviusweg 72, Leiden, 2333 BE, the Netherlands; [5]Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Author for correspondence:
Suman Neupane
Email: suman.neupane@bsu.edu

Received: *2 February 2025*
Accepted: *22 November 2025*

New Phytologist (2026)
doi: 10.1111/nph.70870

Key words: causal modeling, derived woodiness, eudicot angiosperms, growth habit evolution, phylogenetic comparative methods.

Summary

- Phylogenetic comparative methods (PCMs) help researchers understand and predict trait evolutionary relationships. While improvements to PCMs have focused on increasing model complexity, understanding processes remains difficult due to persistent challenges in grounding complex models in biological reality and synthesizing findings across multiple analyses.
- We examined the evolution of growth habit in eudicots (75% of all angiosperms) and tested how variables such as vessel diameter, leaf phenology, and minimum temperature influence macroevolutionary inference. We used a series of PCMs to synthesize our understanding of trait interrelationships, explored plausible causal relationships using phylogenetic path analysis, and employed phylogenetic cross-validation to assess predictive performance among taxa.
- We found that discrete coding of growth form was linked to other measured and unmeasured traits, and that these interrelationships can help overcome limitations arising from incomplete data and simplistic coding of complex traits. Analysis of growth form using phylogenetic path analysis helps reconcile competing views of trait interrelationships from previous studies. Furthermore, including identified covariates improves prediction of growth habit and other traits.
- Our study shows that incorporating causal structure improves macroevolutionary inference, identifies when analyses that omit key causal traits become unreliable, and underscores the importance of integrating phylogenetic models with natural-history knowledge

Introduction

How do we draw meaningful inferences from phylogenetic comparative data? It has been well understood since Felsenstein's (1985) seminal paper that statistical analysis of observational data on interspecific traits can be confounded due to unobserved factors that follow shared ancestry. While phylogenetic comparative methods (PCMs) have developed considerably and now examine a wide variety of macroevolutionary questions, it is widely appreciated that several caveats exist in any analysis of comparative data (and indeed, any observational dataset): correlation does not equal causation; conclusions may be sensitive to model assumptions; and models are grossly simplified versions of biological reality (Freckleton, 2009; Cooper *et al.*, 2016). While these points are largely recognized by researchers, the goal is that these models are nonetheless useful for elucidating meaningful information about biological processes. However, it remains critical to ask how far our current methods are from the known underlying biology of a complex set of interacting traits, and how sustained progress toward extracting meaning from the limited and finite comparative datasets can be made.

An overarching goal of PCMs has been to understand the tempo and mode of evolution of traits of interest and to link patterns among multiple traits and environmental variables (O'Meara, 2012). To account for the unobserved phylogenetic confounding, a variety of trait evolution models exist for discrete (e.g. MK model: continuous-time Markov chains with k states) and continuous (multivariate Brownian motion and Ornstein-Uhlenbeck models: multivariate Gaussian processes) traits. Variations of these models allow researchers to understand how these different traits might influence each other and are influenced by the environment using the comparative method (Felsenstein, 1985; O'Meara, 2012; Uyeda *et al.*, 2018). Although the evolutionary linear models such as phylogenetic independent contrasts (Felsenstein, 1985) and phylogenetic generalized least squares (PGLS; Grafen, 1989) are often described as removing phylogenetic nonindependence by accounting for expected phylogenetic covariance between species, these methods are better thought of as modeling the evolution of unobserved confounding factors (Hansen *et al.*, 2008; Uyeda *et al.*, 2018; Blomberg *et al.*, 2020). Quite understandably, researchers traditionally examine only a few factors and relationships at a time; and therefore, accounting for these background factors with stochastic modeling is a reasonable solution to the limitations of comparative analyses. However, it also illustrates a significant challenge: how do we build an evolutionary story for a particular group, when each analysis 'starts over' from scratch? In essence, each successive study may iteratively 'background' different traits to be subsumed into the non-focal component of the model, often Brownian motion or an alternative model for residual variation. As a result, each subsequent analysis of a clade seems to 'forget' previously learned information about the important factors driving evolution in the group.

As an example, this manuscript examines the evolution of plant traits, particularly growth form (plant habit and height) in eudicot angiosperms to understand the causal processes underlying these trait correlations. Using phylogenetic regression models, Zanne *et al.* (2018) used data from growth habits (woody vs herbaceous), vessel size (cross-sectional area calculated from diameter assuming circular lumina), leaf phenologies (deciduous vs evergreen), and environmental limits (minimum temperature and precipitation at the range extremes) across the angiosperm phylogeny. They showed that conduit size varies in concert with leaf phenology and minimum temperature, with linear and nonlinear relationships between vessel sizes and minimum temperature limits depending on whether lineages are deciduous and evergreen, respectively. However, other studies using similar methods and data (e.g. Olson & Rosell, 2013; Olson *et al.*, 2014, 2018, 2021) show that plant height is the strongest predictor of conduit diameter variation across temperature gradients and argue that if selection from freezing or drought acts against wider conduits, it does so by reducing plant size as vessel diameter and plant height are strongly correlated traits. Such contrasting papers might seem to suggest opposing conclusions that nevertheless make similar predictions. However, it is likely that specific organismal traits and environmental variables are embedded as parts of a larger set of interrelated variables that are causally linked to one another, and therefore may or may not be properly considered confounding variables. Understanding when background variables should be treated as confounders in a phylogenetic analysis requires that phylogenetic comparative analyses meaningfully confront the theory behind causal inference (Hardenberg & Gonzalez-Voyer, 2013; Pearl, 2014; Uyeda *et al.*, 2018; Keele *et al.*, 2020).

To illustrate, consider this regression model: $Y = \alpha X + \beta Z + b$. In standard practice, the partial regression coefficient α in the regression model can be interpreted as the average causal effect of X on Y after controlling for the known confounding variable Z and can be extended to any number of confounders. This multiple regression approach is commonly applied to 'control' for any number of variables that are not of primary interest and, in PGLS, are not explicitly treated, but modeled as latent variables whose effects evolve under Brownian motion (or another model). However, this approach is valid for only a limited set of causal structures. Specifically, variable Z is a confounder if it is a common cause of both X and Y (i.e. $X \leftarrow Z \rightarrow Y$), controlling for Z renders X and Y conditionally independent unless X has a direct effect on Y. However, if Z is the common effect of both X and Y (i.e. $X \rightarrow Z \leftarrow Y$), it is known as a collider. Here, conditioning on Z can induce spurious association between X and Y, and hence adding such collider variables to multiple regression to 'control' for possible confounders can be positively misleading and introduce a spurious effect of X on Y (Pearl, 2009). Other relationships for variables, such as Z mediating the effect of X on Y, will likewise invalidate the interpretation of regression coefficients as causal effects. Indeed, only through analyzing a graph of the causal path structure, can we identify the proper way to measure the total effect of a variable, for example, the effect of minimum temperature on the vessel diameter in angiosperms. It is possible, for example, that this effect is both indirect (mediated through the plant's height, Olson *et al.*, 2018), as well as direct (not mediated through the plant's height). When the mediator (e.g. leaf phenology) involves interactions with the predictor variable (Zanne *et al.*, 2018), the total effect is nonlinear and effects along the direct and indirect paths are not additive. Hence, any attempt to remove the confounding effect must consider how these variables are connected in a logical path structure (Keele *et al.*, 2020). Such considerations are rarely explicitly included in comparative analyses, especially when known relationships are treated as latent background in a phylogenetic model. While no observational study can definitively demonstrate causation, correlations can disprove many possible causal structures – providing a means for ranking causal hypotheses by their plausibility (Shipley, 2000). Therefore, when considering the evolution of a trait and its underlying causes, comparative studies should be working toward understanding this underlying causal path structure and building upon existing knowledge when new hypotheses are tested.

Here, we explore the evolution of plant growth habit by studying the underlying causal structure of the evolution of this trait in eudicots in relation to its vessel diameter, leaf phenological traits, and environmental variables (specifically minimum temperature of the coldest month). This manuscript is divided into three sections of methods and their corresponding results, as each of these sections build on the previous one. We begin in Section I with individual analyses of traits, particularly growth habit and cross-sectional vessel diameter, which pose interrelated challenges for the assumptions of comparative models. In this section, we demonstrate how these violations affect trait reconstructions and phylogenetic comparative analyses. In Section II, we then examine the interrelationships among variables using causal inference via Phylogenetic Path Analysis. We identify the causal graph that is consistent with the observed data and discuss its implications for our third aim, phylogenetic prediction. In Section III, we use hypothesized causal predictors to cross-validate our dataset and compare it to the natural history of specific taxa for which the prediction fails, as well as identify what variables cause misprediction of these taxa. Finally, we summarize our analyses and highlight remaining gaps – demonstrating one way for how analyses of plant growth form evolution can connect relevant predictors and prior knowledge, suggest guidance on how the impact of other variables can be fit into future analyses, and how these future analyses can build upon existing research to obtain a synthetic understanding of macroevolutionary processes.

I. Woodiness in continuum: modeling angiosperm wood evolution and the concept of derived woodiness

Modeling discrete trait heterogeneity

Flowering plants (angiosperms) have evolved various life-history strategies to adapt to the Earth's diverse climatic conditions. Herbaceous (i.e. nonwoody) annual plants die every year after shedding their seeds to escape the harshest periods in the year as seeds in the soil. By contrast, longer-lived perennials that possess an active vascular cambium must withstand daily and seasonal variation in environmental conditions. One way for woody perennials to cope with stress-induced damage is to produce a new layer of secondary xylem (i.e. wood) and secondary phloem by the vascular cambium (Gauthey *et al*., 2022). Consequently, annual plants exhaust their shoot meristems during the development of their reproductive system, while perennials maintain their meristems as active during consecutive growing seasons. Although all perennial nonmonocot angiosperms are able to produce a negligible amount of wood at the base of their aboveground stems, this wood development is often so limited in many perennials that they are considered to be herbaceous (Kidner *et al*., 2016).

Wood is a complex tissue consisting of multiple cell types, including different conduits, fibers and parenchyma cells. This diversity leads to significant variation in microscopic wood anatomy patterns across lineages that can be used to identify woody plants (e.g. Gasson, 2011; Lens *et al*., 2020). In this study, we acknowledge that wood can be defined through both anatomical (see previous paragraph) and functional perspectives. From a functional point of view, woody species have an aboveground stem that persists over time under changing environmental conditions (see Zanne *et al*., 2014), regardless of whether secondary xylem is formed. Since our dataset is derived from Zanne *et al*. (2014), which employs a functional interpretation, we adopt this approach to maintain consistency with the data source. This functional definition allows for the classification of plants, such as palms (monocots) as woody due to their structural persistence, even though they lack a vascular cambium and consequently never produce wood.

Despite structural and morphological variability, macroevolutionary studies often treat growth form as a single, homogeneous trait with binary states (woody vs herbaceous). These studies typically model the evolution of these states (e.g. for ancestral state reconstructions) using continuous-time Markov models (O'Meara, 2012). For such morphological discrete traits, Lewis (2001) proposed the Mk model of trait evolution, which – like all Markov models – exhibits a memoryless property. Under this framework, the probability of transitioning between states (defined by a fixed instantaneous transition rate matrix) depends solely on the current state, regardless of how long the trait has persisted in that state. However, the remarkable diversity of wood anatomical types and woody forms across angiosperm lineages arises both from the legacy of their evolutionary history (Luo *et al*., 2023) and ecological adaptations that drive the complexity of xylem anatomy (163 wood anatomical traits in angiosperms defined and illustrated in IAWA Committee, 1989). Therefore, it is likely that woody and herbaceous clades evolve across the angiosperm phylogeny at different transition rates (Smith & Donoghue, 2008; Smith & Beaulieu, 2009; Beaulieu *et al*., 2013), due to both intrinsic and extrinsic (ecological) factors. Additionally, closely related species in many angiosperm families often have a different growth form suggesting many transitions toward woodiness (Zizka *et al*., 2022); in other lineages the woody-herbaceous growth form differences among close relatives are subtle, making it challenging to clearly score growth form as a binary trait (Lens *et al*., 2012).

Recognizing the unrealistic nature of treating woodiness as a simple binary trait in angiosperm growth habit, Beaulieu *et al*. (2013; Beaulieu & O'Meara, 2014) developed a hidden-state Markov model that captures the possibility of underlying differences in states and their likelihood to transition between observed states. The inclusion of additional 'hidden states' in a Markov model allows for differences in transition rates to account for some of the underlying heterogeneity in real biological processes, such as identifying 'fast woody' and 'slow woody' clades. These terms refer to the transition rates, where taxa with fast rates switch between trait states more frequently than those with slow rates. While quite flexible, such models make limited assumptions about what the hidden states actually represent. Furthermore, hidden states may not be entirely hidden (Tarasov, 2019) and may instead be tightly related to observed and measured traits in a particular dataset. In practice, hidden-state models applied to complex characters may often be preferred because they implicitly allow for underlying continuity in the state space of the character. An alternative to discrete hidden-state modeling, the threshold model from quantitative genetics (Wright, 1934) has been adapted to macroevolution (Felsenstein, 2005; Hadfield & Nakagawa, 2010; Revell, 2014). It assumes an underlying quantitative latent variable known as a liability (a continuous form of the trait often modeled to evolve under Brownian motion along the phylogeny) that determines an observable categorical state of a trait. Unlike the Mk model, the threshold model predicts that lineages near the threshold will transition frequently between discrete states, whereas lineages far from the threshold will be largely fixed, producing fast and slow states naturally from an underlying continuous state space. Furthermore, the threshold model allows evolutionary covariances to be estimated between the latent liability and continuous observable traits – which in some cases, may directly measure the liability itself (Felsenstein, 2005). Establishing such links between different traits is especially important, as a major challenge to any comparative analysis that synthesizes relationships across multiple variables is the fragmentary nature of most trait databases. For example, plant trait databases are biased in reporting conduit (in angiosperms largely vessel) diameter for large woody plants and omitting herbaceous ones. Here, we evaluate whether phylogenetic signal of herbaceous ancestry is retained in the phylogeny and how this may affect inferences of vessel diameter based solely on data from woody plants.

Ancestral and phylogenetically derived woodiness

Earlier ancestral state reconstructions of the growth form trait using parsimony suggested the presence of both 'ancestral' and 'derived' woodiness, with early angiosperms inferred to be ancestrally woody (Zizka *et al.*, 2022). Lineages that maintain this woody trait without interruption throughout evolutionary history are also considered ancestrally woody (Feild *et al.*, 2004; Doyle, 2012). While transitions from ancestral woodiness to herbaceousness are frequent across angiosperm lineages, the reverse transitions, from herbaceous to derived woody forms, are less frequent and are often associated with specific environmental conditions on (sub)tropical islands – a phenomenon known as insular woodiness. This evolutionary shift has occurred at least 175 times on (sub)tropical islands, with most transitions concentrated in a small number of plant families (Zizka *et al.*, 2022). There are several proposed hypotheses for woodiness evolution, including competition for sunlight among herbaceous colonizers that settle in open habitats on islands leading to the selection of taller-stemmed plants (Darwin, 1859; Givnish, 1995), the increased lifespan of plants due to the frost-free, aseasonal climate and lack of native herbivores on the island (Carlquist, 1974), and the adaptation of plants to drought through increased amount of wood tissue as a means of delaying drought-induced root-to-shoot water blockage in the plant's vasculature (Lens *et al.*, 2013; Choat *et al.*, 2018; Dória *et al.*, 2019). In the recent comprehensive analysis of insular woody plants using the structural equation model, Zizka *et al.* (2022) found that the prevalence of insular woodiness is primarily influenced by a series of variables related to drought, but also to reduced herbivory and longer growing seasons (absence of frost). The classification of taxa into ancestral and derived woodiness categories is mainly inferred through phylogenetic trait mapping. Furthermore, genetic studies on herbaceous *Arabidopsis thaliana* (L.) Heynh. show that a shift to woodiness could be easily modulated by a simple genetic switch where two genes involved in flowering time control are knocked out resulting in considerable wood development in double mutant types (Melzer *et al.*, 2008). Likewise, overexpression of the rejuvenator gene AHL15 (which is involved in the same wood pathway downstream of SOC1 and FUL) delays the onset of the reproductive stage and leads to a similar woody phenotype (Rahimi *et al.*, 2022). Thus, this labile nature of plant woodiness is due to the variable activity of the vascular cambium under specific environmental conditions and the widely scattered distribution of woody and herbaceous lineages across the angiosperm Tree of Life warrants further exploration of the impact of modeling assumptions on these interpretations of derived woodiness (Zizka *et al.*, 2022).

Specifically, in this section, we ask three questions: (1) given that traits like plant growth habit are relatively labile in some groups, how does the growth form threshold compare with Markov models in modeling woodiness evolution across macroevolutionary scales in eudicot angiosperms? (2) Does macroevolutionary modeling and ancestral state reconstruction of woody and herbaceous plant evolution in angiosperms argue for a distinct designation of derived woodiness? (3) Given biases in the acquisition of certain traits that covary with growth habit, such as vessel diameter, can we use model-based approaches to impute missing values?

Methods and results: ancestral and derived woodiness under alternative models of evolution

To infer derived woodiness based solely on phylogeny and character mapping, we used a eudicot phylogeny (extracted from the seed plant phylogeny; Smith & Brown, 2018) as a test case. Growth habit data (woody/herbaceous) were obtained from the Global Woodiness Database (Cornwell *et al.*, 2013), which employs a functional definition of woodiness (see Supporting Information Dataset S1 for complete data sources and standardization procedures across all analyses). Using the corHMM function from the *R* package corHMM (Beaulieu & O'Meara, 2016), we generated a matrix of maximum likelihood estimates (MLE) for transition rates between herbaceous and woody states under a hidden rates model. This model also allowed us to infer the most likely state at each tip. The model assumes that each of the four possible states – slow woody, fast woody, slow herbaceous, and fast herbaceous – had distinct transition rates across different portions of the eudicot phylogeny.

Using the MLEs of transition rates and the most likely states at the tips, we generated 100 stochastic character maps (Huelsenbeck *et al.*, 2003; Bollback, 2006) with the R package make.simmap (Revell, 2012). These maps reconstructed ancestral trait states along the branches of the phylogeny, with the root fixed as a woody ancestor based on prior assumptions. We used these stochastic maps to determine whether each tip taxon retained its primary state (i.e. the ancestrally woody state) throughout its evolutionary history. Specifically, if a transition from woody at the tip to herbaceous along the branches was observed, followed by a reversion to woody (i.e. Woody → Herbaceous → Woody), the tip was classified as having a derived woody state. We quantified the ancestral or derived woodiness of each tip as the percentage of instances it was classified as ancestral or derived across the 100 stochastic maps.

We also compared our simmap-based list of derived woody taxa with the derived woody taxa on islands identified by Zizka *et al.* (2022). Note, our goal was not to establish a particular taxon as ancestrally or derived woody, as such designations are made based on more lines of evidence than simply macroevolutionary models. Rather, our goal was to evaluate to what extent the inference of derived woodiness from macroevolutionary models is dependent on macroevolutionary assumptions about the underlying causes of trait variation.

The apparently low prevalence of ancestrally woody states in reconstruction (or at least, high uncertainty, Figs 1, S1) despite widespread clades in the 'slow woody' category appears to result primarily from rapid radiations of herbaceous and woody taxa near the base of the eudicot phylogeny. Under this scenario, the models reconstruct a fast woody ancestor for most major lineages of eudicots. The model predicts that ancestral 'fast woody' states (or, in the threshold model, liabilities near the threshold, see the threshold model results in the subsequent section) in these basal lineages likely spent some time in an herbaceous state, even if they are now classified as 'slow woody' or far from the threshold. Thus, macroevolutionary modeling alone reconstructs low confidence in ancestral woodiness across eudicot taxa.

The simmap results show that very few lineages (972 out of 10225 woody taxa in our dataset) retain their ancestrally woody status with high probability (Fig. 1, also see Fig. S1). Among the 10225 woody taxa, 272 woody species overlap with the derived woody taxa on islands identified by Zizka *et al.* (2022). Of these 272 shared taxa, 259 taxa show more than 90% support for derived woodiness in our SIMMAP-based derived wood designation, defined as fewer than 10 of 100 stochastic reconstructions inferring ancestral woodiness. Furthermore, 180 out of 272 (66%) of the derived woody taxa identified by Zizka *et al.* (2022) fall under the 'fast woody' category in the corHMM model. Our results indicate that these taxa are not inferred to have maintained a persistent woody state but instead remain in a transitional phase, characterized by high transition rates between herbaceous and woody states.

Methods and results: growth habit covaries with many other traits, and they jointly carry information about the evolutionary history across eudicot lineages

Most of the derived woody taxa (66%) designated by Zizka *et al.* (2022) had fast transition rates which suggests that these lineages are potentially near the threshold zone where the growth habit transitions are quite labile. To explore this, we fitted the threshold model with covariates using the efficient phylogenetic multivariate mixed model framework implemented in a custom version of the *R* package MCMCglmm (Hadfield, 2010) referred to as MCMCglmmRAM (Hadfield, 2015). Unlike the original MCMCglmm (Hadfield, 2010), this implementation uses the reduced animal model (RAM) to integrate out tip-level phylogenetic effects deterministically, which provides much faster and more stable MCMC sampling for discrete traits, especially threshold models. The MCMC algorithm approximates the distribution of the parameters (covariance matrix of continuous traits and liabilities of the discrete traits) by sampling the joint posterior distribution of covariances and liabilities (see Notes S1 for detailed model specifications, variance structures, and implementation code).

The MCMCglmmRAM was performed on the eudicot phylogeny (extracted from seed plant phylogeny; Smith & Brown, 2018). Climatic data (temperature) were obtained by querying specimens' georeferenced locations against WorldClim variables (Hijmans *et al.*, 2004). Following Zanne *et al.* (2018), we specifically used BIO6 (minimum temperature of the coldest month), taking the 2.5th percentile across georeferenced locations for each species as the limiting value affecting plant hydraulic conductivity.

Conduit size was expressed as mean vessel diameters. For woody taxa, vessel diameter data were obtained from the Angiosperm Global Vessel Anatomy Database (Zanne *et al.*, 2010), and

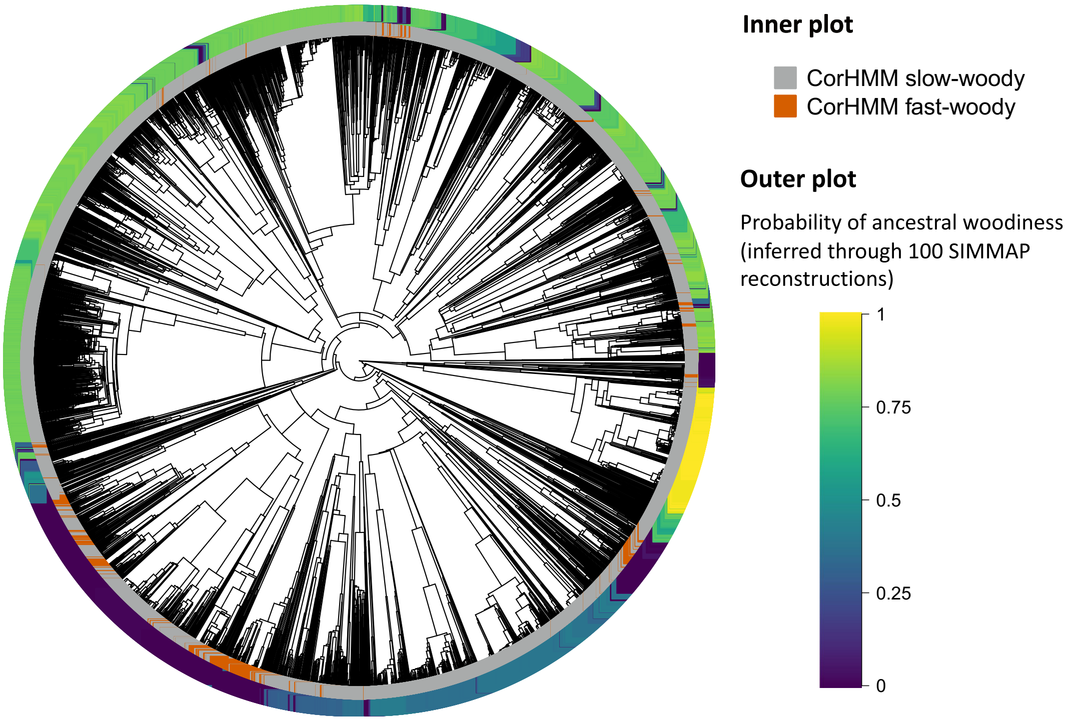


Fig. 1 Distribution of corHMM slow-woody (gray) and fast-woody (orange) lineages (inner ring), and the SIMMAP reconstruction of ancestral (and derived) woodiness (outer ring) across the eudicot phylogeny including only woody species. The probability of ancestral woodiness was inferred through 100 stochastic character maps as the percentage of maps (out of 100 maps) each woody lineage maintained its woody state throughout the history of lineage (i.e. from root to the tip) under Mk2 states (herbaceous, woody) model.

for herbaceous taxa, from The Xylem Database (Schweingruber & Landolt, 2005). Leaf phenology data (deciduous, evergreen, variable) were obtained from Zanne *et al.* (2018), originally sourced from the Global Leaf Phenology Database (Wright *et al.*, 2014). Growth habit (woody/herbaceous) data were obtained from the Global Woodiness Database (Cornwell *et al.*, 2013). Vegetative plant height (maximum vegetative height) data were obtained from the TRY Plant Trait Database – Trait ID 3106 (plant height vegetative; Kattge *et al.*, 2011; https://www.try-db.org, accessed March 2019). All variables – vessel diameter, minimum temperature, and plant height – were log-transformed for analysis (see Dataset S1 for data sources and filtering details for all trait data).

Mapping threshold liabilities of woody taxa with their probability of being derived woody (i.e. 1 minus the probability of ancestral woodiness) revealed strong correlations (Fig. 2). Taxa with high probability of derived woodiness lie near the threshold based on posterior distributions of liabilities. Next, we converted the covariances of a set of traits and environmental variables – growth habit, plant height, vessel diameter, minimum temperature, and leaf phenology – estimated under a phylogenetic mixed effects model for threshold traits, into pairwise correlations (posterior distributions shown in Fig. 3). All variables, except vessel diameter and plant height with leaf phenology, were significantly correlated, with posterior distributions bounded away from zero.

We estimated vessel size optima (vessel diameter) for herbaceous and woody regimes by allowing adaptive shifts under a multi-optima Ornstein–Uhlenbeck (OU) model of trait evolution. We used the same eudicot simmap phylogenies, with woody (slow and fast) and herbaceous regimes ‘painted’ as fixed locations of shifts on the internal branches of the phylogeny. For the purpose of imputation, we removed all herbaceous tips from the simmap phylogeny (e.g. Fig. 4a) and treated their vessel diameter data as missing, retaining only woody taxa and their corresponding vessel diameter for analysis. Using these painted regimes with woodiness-only data, we estimated vessel size optima for both herbaceous (absent in the tips) and woody regimes (present in the tips with their vessel sizes) under Bayesian rjMCMC multi-optima OU process using the *R* package bayou (Uyeda & Harmon, 2014) and summarized these optima over 100 stochastic maps (see Fig. S2 for detailed workflow). The aggregate mean vessel size (cross-sectional diameter in mm) optima for corHMM designated regimes, herbaceous (fast and slow herbaceous combined), fast woody, and slow woody growth habits estimated by BAYOU (using from 100 stochastic maps) are 0.048 (95% credible interval 0.03–0.07), 0.066 (95% credible interval 0.03–0.12), and 0.07 (95% credible interval 0.069–0.08), respectively. Despite the dataset having no observations of vessel sizes in herbaceous plants, we reconstructed a vessel size distribution for herbaceous plants from the model that was near identical to its true distribution from literature-reported values (Fig. 4b), suggesting that the covarying traits of vessel size and plant growth habits can be used to estimate missing value trait (e.g. herbaceous vessel size) using the model-based trait imputation process. Particularly, information on unobserved vessel sizes can be leveraged from recent derived woody plants, which maintain signal of trait values in the herbaceous state.

Summary: growth habit is a complex trait

In this subsection, we present three key findings from our analysis of eudicot growth habit: (1) single trait reconstructions using phylogenetic comparative methods fail to validate designations of ancestral woodiness (Fig. 1); (2) a threshold model best captures the data by modeling growth habit as a latent continuum, supported by strong correlations between threshold liabilities and derived woodiness probabilities (Fig. 2) and significant pairwise trait covariances (Fig. 3); and (3) these trait interrelationships can

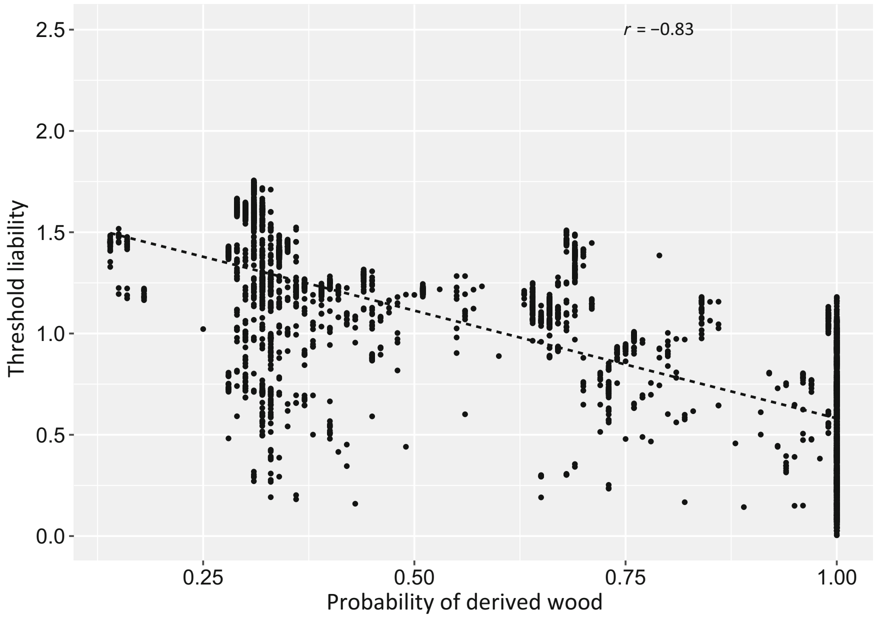


Fig. 2 Scatter plot showing the relationship between each woody taxon’s probability of derived woodiness (inferred under a Mk model) and its average threshold liability. Taxa closer to the threshold zone (i.e. threshold liability near 0) exhibit a higher probability of being derived woody. The fitted dotted regression line indicates a strong negative association between the two variables ($r = -0.83$). Probabilities were estimated from 100 stochastic character maps as the proportion of maps in which a lineage did not remain woody throughout its evolutionary history (root to tip).

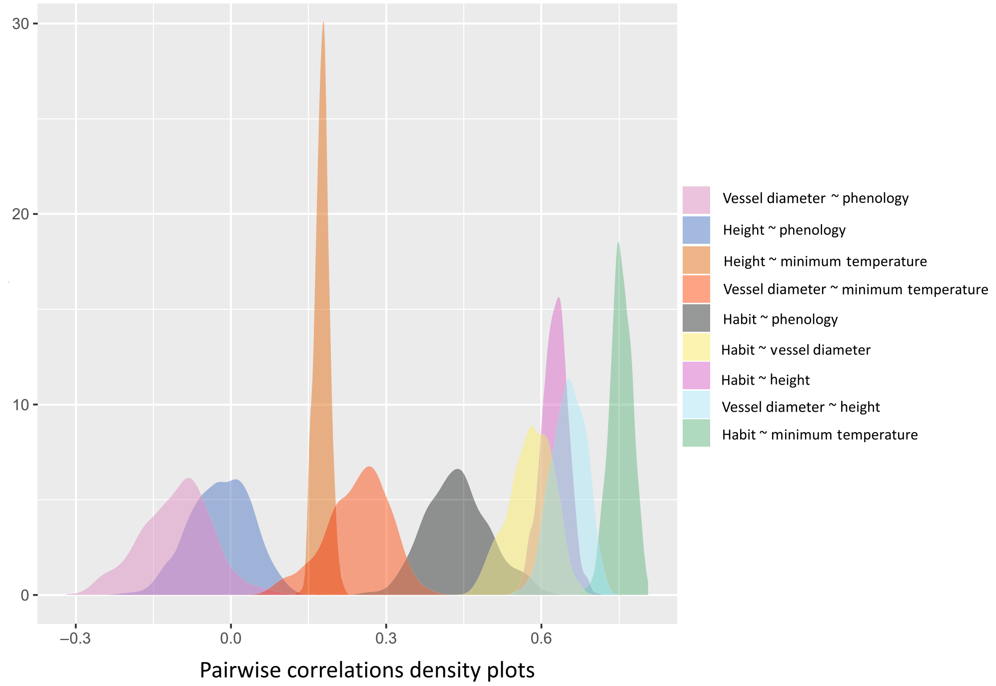


Fig. 3 Testing correlations among different variables affecting plant growth habit, height, and conduit size. Each colored density curve represents the posterior distribution of the correlation coefficient for a pair of variables (growth habit, log-plant height, log-cross-sectional vessel diameter, log-minimum temperature, and leaf phenology), estimated using MCMCglmmRAM. Correlations were obtained by converting the posterior covariance matrices from the phylogenetic mixed-effect models into pairwise correlation coefficients.

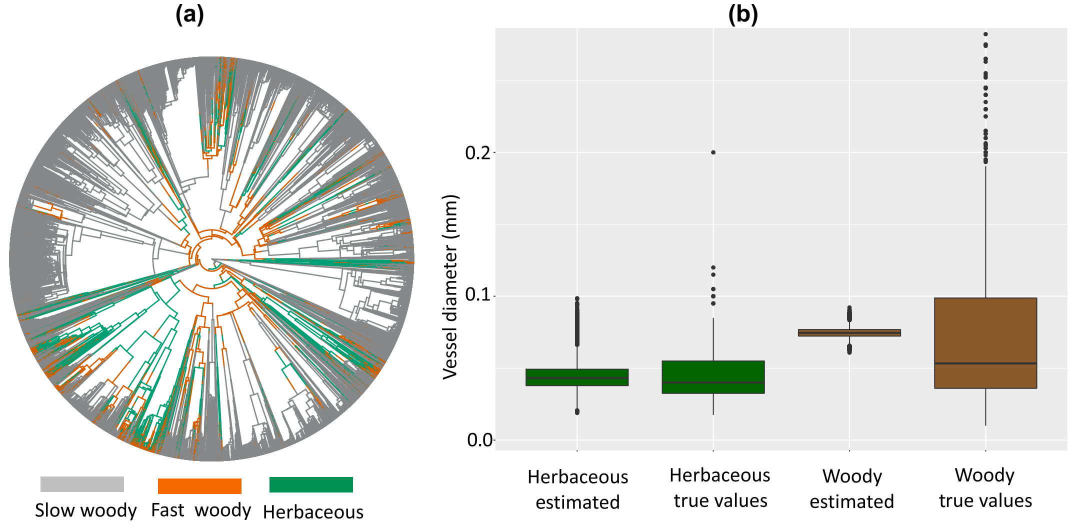


Fig. 4 Imputation of missing vessel diameter data for herbaceous lineages using a multi-optima Ornstein–Uhlenbeck (OU) model. (a) Example SIMMAP reconstruction (one of 100) showing internal branches colored by regime: slow woody (gray), fast woody (orange), and herbaceous (green). For the purpose of imputation, we dropped all herbaceous tips and treated their vessel diameter as missing. Only vessel diameters for woody species were retained. (b) Under a Bayesian rjMCMC multi-optima OU process (implemented in BAYOU), we used the woody-only data (vessel sizes) alongside the painted regimes to estimate vessel size (diameter) optima for both herbaceous and woody regimes. The box plots compare predicted optimal vessel sizes for missing herbaceous taxa (inferred from the woody data and regime assignments) against their true observed values. Estimates shown here are summarized across 100 stochastic maps. Boxplots show the median (horizontal line), the 25th–75th percentiles (box), whiskers extending to 1.5× the interquartile range, and outliers as points.

aid in trait prediction of missing data, even when data are heavily biased in their collection, such as the case of vessel diameter that is predominantly reported for woody plants (Fig. 4).

More specifically, our analysis further demonstrates that phylogenetic comparative data alone cannot validate the existence of ancestrally woody states, defined as woody lineages that have always remained woody during evolutionary history of angiosperms. Although the idea of derived woodiness originated with Darwin and Wallace (Darwin, 1859; Wallace, 1878) and was elaborated by Carlquist's (1974, 2009, 2012) observation on many (sub-tropical) islands, these results were later supported in studies across plant groups using phylogenetic analysis through trait mapping (e.g. Böhle *et al.*, 1996; Givnish, 1998; Kim *et al.*, 2004; Lee *et al.*, 2005; Lens *et al.*, 2012, 2013; Neupane *et al.*, 2017; Frankiewicz *et al.*, 2020; Zizka *et al.*, 2022). These phylogenetic trait-mapping studies were initially based on the parsimony method, which minimizes the number of steps, and therefore often implied no more than a single change per branch. Later work increasingly adopted stochastic models (maximum likelihood, Bayesian) that allowed multiple transitions along a branch and provided probabilistic ancestral-state estimates. Using these phylogenetic methods, many clades that we found to be derived woody have been considered ancestrally woody since the origin of angiosperms.

However, using models that capture more of the heterogeneity underlying growth habit by allowing clades specific rates of state changes (corHMM, or the threshold model) and the possibility of states changes along branches, we found that the phylogenetic evidence alone using traditional trait mapping is not sufficient in confirming whether a plant species is ancestrally woody (Fig. 1). Widespread ancestral woodiness may certainly occur in eudicots (and it is often based on more than phylogenetic evidence), but phylogenetic evidence alone provides equivocal support for a vast majority of eudicots. For example, Klimeš *et al.* (2022) estimated that only 9% of Lamiidae species retain primary woodiness, compared to much higher proportions in Fabidae (71%) and Malvidae (78%), highlighting how unevenly ancestral woodiness is distributed across major eudicot clades. This equivocal support in our analyses emerges from the observation that the base of the eudicot radiation has a number of short branches from which numerous clades radiate (see Zuntini *et al.*, 2024) with a large diversity of growth habit that remain relatively static thereafter. In other words, several major clades branch off at the base of the phylogeny that are either ancestrally woody or herbaceous. To generate such clades, the model confidently reconstructs the ancestral branches of these groups to be 'fast woody', so that both herbaceous and woody lineages can descend from them. However, if these lineages indeed were 'fast woody', then the probability that they briefly transitioned to 'fast herbaceous' is high. This scenario corresponds well with the concept of a threshold model, in which lineages near the threshold transition at high rates, but that as a lineage maintains a woody state, it drifts further and further from the threshold (Fig. 2).

Auxiliary traits associated with ancestral woodiness may provide independent evidence that these lineages maintained woodiness throughout their evolutionary history. However, incorporating rate heterogeneity into comparative methods revealed that absent this evidence, many of the woody states in comparative analyses are likely representative of the underlying continuous nature of the woody–herbaceous continuum with thresholds in state change. These results highlight the need for caution when interpreting results based solely on parsimony reconstructions of complex traits.

Revealing the underlying heterogeneity of rates in growth habit has the additional effect of revealing a path toward clarifying its history – many familiar traits are informative about where a plant lies on the continuum of herbaceous to woody. Macroevolutionary studies on plant growth habit evolution do show that the switch from herbaceous to woody can occur in a range of environmental conditions such as in arid areas of Canary Islands (e.g. Lens *et al.*, 2013; Dória *et al.*, 2018), southern Africa (Frankiewicz *et al.*, 2021) and other insular environments with moderate climate of oceanic island or island-like tropical mainland mountains (Neupane *et al.*, 2017; Nürk *et al.*, 2019; Zizka *et al.*, 2022). Indeed, our threshold models reveal widespread significant covariances between habit and leaf phenology, stem length, environmental temperature and cross-sectional vessel diameter (Fig. 3). Considering that plant height is a strong covariate of woodiness, it becomes clear why our reconstructions failed to recover ancestral woodiness, since the models treated woodiness as a binary trait and ignored its correlation with height. Furthermore, it suggests that ancestral state reconstructions and the inferences that depend on them can benefit from the inclusion of additional covariates underlying discretely coded characters.

Previous work (e.g. Westoby *et al.*, 2002; Díaz *et al.*, 2016) supports the strong correlations among plant functional traits. However, to build a cohesive understanding of how these factors are interrelated, simply demonstrating correlations is not sufficient. Does a phylogenetic regression of growth habit against a given predictor require that all other significantly correlated variables be included as covariates? To understand the possible causal structures underlying these relationships, we turn toward phylogenetic path analysis (Hardenberg & Gonzalez-Voyer, 2013) in the next section as a guide for determining which variables are confounders, mediators, or colliders (see Introduction for detailed explanations of these concepts) and how such questions can be pursued in individual analyses, despite these complexities.

II. Causal modeling: phylogenetic path analysis of temperature effects on growth form, leaf phenology, and vessel size

Our previous section highlights the need to make sense of a complex set of interrelated intrinsic and extrinsic traits when most pairwise comparisons have significant correlations. The outstanding questions are how to make sense of such a wide array of correlated traits to understand cause, effect and make predictions? How do we interpret results from standard comparative methods (Markov, BM, etc.) of individual trait analyses when the real underlying processes are a complex interrelated web of cause and effect? Here we revisit Olson *et al.* (2018), who found maximum plant height to be the main predictor of vessel diameter across temperature gradients and Zanne *et al.* (2018), who linked conduit size to leaf phenology and minimum temperature. Using causal-inference and phylogenetic path analysis (Hardenberg & Gonzalez-Voyer, 2013), we integrate these perspectives into a more holistic evolutionary framework of growth-habit and conduit diameter evolution.

Methods and results: phylogenetic path analysis

We tested for direct and indirect effects of an environmental variable (minimum temperature) on plant traits (cross-sectional conduit diameter, plant height, leaf phenology, woodiness) and

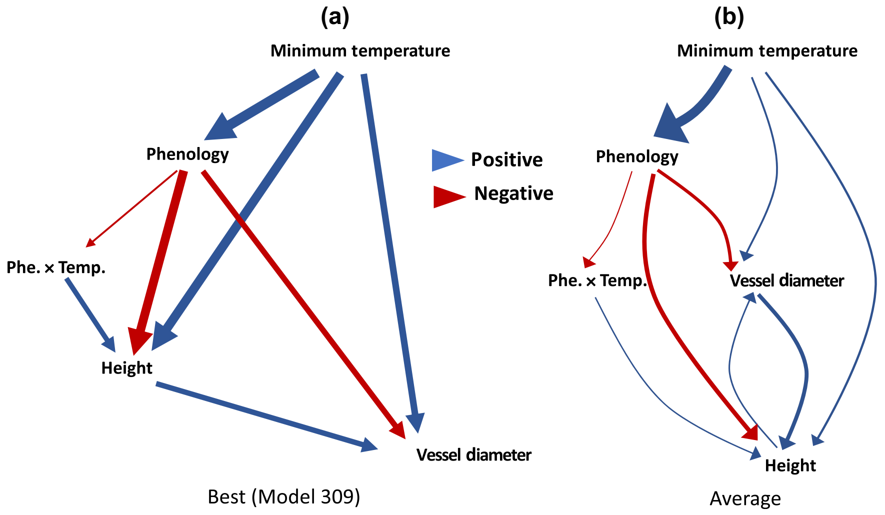


Fig. 5 Phylogenetic path analysis results depicting causal relationships among plant traits (log-cross-sectional vessel diameter, log-plant height, and leaf phenology), an environmental variable (log-minimum temperature), and their interaction (log-minimum temperature × phenology). Arrows represent significant causal links; line width indicates the magnitude of the standardized regression coefficient, and color denotes the direction of the relationship (blue, positive; red, negative). (a) One of the best-fitting causal models (model 309) based on CICc criteria. (b) The averaged path model across the five best-fitting models (ΔCICc < 2).

among plant traits using confirmatory path analysis. Phylogenetic path analysis (Hardenberg & Gonzalez-Voyer, 2013) allows inference of plausible underlying causal structures of these variables by the method of d-separation. In this approach, all pairs of variables that are predicted to be conditionally independent (called the d-separated variables) in the causal model (depicted as Directed Acyclic Graph or DAG) are listed. These d-separation statements are translated to statistical linear models (also accounting for phylogenetic non-independence using phylogenetic least squares) and tested for the statistical independence conditioning on other variables included in the causal path (the causal parents). The model fit for the particular causal model was evaluated by combining *P*-values of individual d-separation statements using Fischer's *C* statistic (Shipley, 2000). The phylogenetic path analysis was performed using the R package PHYLOPATH (van der Bijl, 2018) using the eudicot phylogeny extracted from the seed plant phylogeny from Smith & Brown (2018).

First, we combined the variables included in models in Olson *et al.* (2018) and Zanne *et al.* (2018) and tested their relationships in our causal path model. We also added an interaction term between minimum temperature and leaf phenology in the causal path structure as it was shown by Zanne *et al.* (2018) that conduit size-minimum temperature relationships varied between deciduous (linear relationship) and evergreen species (sigmoidal relationship). A total of 607 plausible path models were compared. Of the 607 path models that were tested to infer the causal structure among log-cross-sectional conduit diameter, log-minimum temperature, log-plant height, leaf phenology, and interaction variable (log-minimum temperature × leaf phenology), five models were favored (Fig. S3) based on CICc criteria (C-statistic information criterion corrected for small sample sizes) with Δ CICc values < 2. Minimum temperature had a direct positive effect on cross-sectional conduit diameter and plant height (Figs 5, S3). However, the causal direction between cross-sectional conduit diameter and plant height was unresolved: three models favor a direct effect of height on cross-sectional conduit diameter and two models favor the reverse. In all five models, evergreens had higher minimum temperature than deciduous and when phenology had direct causal path on the conduit size and plant height, evergreens favored smaller conduit size and plant height vs deciduous (Fig. 5). Further, there was also a significant interaction effect of minimum temperature and phenology on height (model 309, model 325, model 60) or on conduit size and height both (model 52, model 301). The significant positive interaction between minimum temperature and evergreen phenology indicates that minimum temperature has a stronger effect on height or conduit size in evergreen species than in deciduous species.

Second, we expanded the model to include the plant growth habit (woodiness) as a trait that could be influenced by other variables (minimum temperature, plant height, leaf phenology, and cross-sectional conduit diameter). The growth habit was treated as a continuous trait in the form of posterior tip liabilities of plant growth habit obtained using MCMCglmmRAM under a simple model with no covariance with other traits (i.e. random effects only). The phylogenetic path analysis was performed over all posterior tip liabilities (1800 samples) and the results were averaged over the best selected models (with Δ CICc values < 2) for each sample of posterior tip liabilities. Two types of path structures (i.e. variables combinations) were explored: (1) four variables path models with the selected variables except for leaf phenology (138 possible path models); and (2) five variables path models (7338 possible path models) including leaf phenology (categorical). In both averaged models (Fig. S4A,B), as expected plant growth habit exhibited as a threshold property where plant height and growth habit threshold liabilities were strongly linked but their causal direction was unresolved. Also, we found that plant growth habit was influenced by conduit size and to some extent by minimum temperature (Fig. S4A).

Summary: path analyses unite findings

The results from the path analyses are consistent with both the findings of Olson *et al.* (2018) and Zanne *et al.* (2018) and suggest that both studies examined a subset of the underlying causal structure relating phenology, vessel diameter, and height. The best-fitting single model measured a direct effect of plant height on vessel cross-sectional diameter (Olson *et al*., 2018), along with direct effects of temperature and leaf phenology (Zanne *et al*., 2018), while model-averaged paths showed stronger effects of vessel cross-sectional diameter on height. Furthermore, the interaction of phenology and temperature observed to affect vessel cross-sectional diameter in Zanne *et al*. (2018) appears mediated through vessel diameter's correlation with height.

While we attempted to include growth habit in the analyses using posterior probabilities of liabilities from the threshold model, these discrete traits did not have enough information to strongly support alternative causal relationships, although the best fitting models included direct effects of height (Fig. S4). In fact, all variables except leaf phenology appeared to impact growth habit across the best and averaged models. In the next section, we examine how inclusion of these variables impacts prediction and misprediction of growth habit across taxa.

III. Phylogenetic prediction using cross-validation

The causal graphs obtained through phylogenetic path analysis highlight one approach for making sense of the widespread correlations among plant traits and environmental variables and allowing piecewise study of trait evolution at the comparative scale. While the use of discrete traits and/or liabilities are limited in their abilities to distinguish among causal hypotheses for growth habit, we sought to determine how inclusion of variables impacted the predictability of growth habit across our dataset using phylogenetic cross-validation.

Methods and results: identifying mispredicted species across models

To evaluate whether including covariates improves prediction of growth habit, we compared two types of models: (1) a simple Markov model (Mk, Lewis, 2001) for discrete traits with homogenous transition rates for the entire phylogeny lacking any covarying traits and variables; and (2) a multivariate phylogenetic threshold model framework. We then compared cross-validation accuracy before (Mk model) and after (multivariate threshold model) the inclusion of covariates associated with growth habit from our causal modeling framework. To test the predictive ability of plant growth habit under both a fixed rates continuous-time Markov Model ('Mk model') and a multivariate threshold model, we iteratively cross-validated each tip in our phylogenetic dataset (5190 species) by replacing its known value of growth habit with a null value. We then predicted its state under both types of models and compared this prediction against its known state.

For Mk model predictions, we used hsp_mk_model function of the *R* package Castor (Louca & Doebeli, 2018). This package allows for missing tip states to be reconstructed using the maximum likelihood approach. For the threshold model, we used the *R* package MCMCglmmRAM (Hadfield, 2015), where the predicted growth habit of missing state is calculated as the percentage of imputed posterior tip liabilities above (woody) or below (herbaceous) the threshold value. The liabilities were obtained under a multivariate phylogenetic mixed-effects model. Initially, we included only growth habit as a single variable, subsequently including different sets of covariates in the causal path (minimum temperature, height, leaf phenology, and vessel diameter; see Fig. S4). We also tested minimum annual precipitation (BIO14; log-transformed after adding Euler's constant to accommodate zeros) as an additional covariate, using 2.5th percentile values across georeferenced localities from WorldClim (Hijmans *et al*., 2004), following Zanne *et al*. (2018) (see Dataset S1). However, precipitation did not have a significant effect in our predictions.

The performance of the Mk model and the threshold model were comparable, with *c*. 95% correct predictions of tip states. The threshold model slightly outperformed the Mk model (Mk mispredictions = 258 of the 4132 tips in the phylogeny; threshold mispredictions = 244 of 4132). There was a total of 307 mispredictions of habit state across both models, with 195 mispredictions shared by both models. We rigorously evaluated whether these mispredictions were indeed mispredictions rather than miscodings of the habit data by manually reevaluating each species. We found that 35 of the taxa (of the original 244 mispredicted habit trait states under threshold model) had erroneous assignment of their growth habit in the original trait data. While not the focus of our study, this indicates that phylogenetic cross-validation can help detect the incorrect coding of comparative data.

Methods and results: predictive modeling and cross-validation

To understand the effect of covariates in trait prediction (vessel size, plant height, minimum temperature, leaf phenology, Fig. S4) in explaining growth habit for 'challenging taxa', we focused our prediction on the 307 mispredicted taxa using cross-validation with the threshold model by estimating tip liabilities under various multivariate scenarios using MCMCglmmRAM. We chose to keep the 35 miscoded taxa in this set under the assumption that these taxa may be challenging to code as a binary trait. The set of models used to obtain the tip liabilities were:

- Null model: no causal effect of other variables on tip liabilities (phylogenetically covarying random effects only)
- Model 2: liabilities covarying with plant height only
- Model 3: liabilities covarying with vessel diameter only
- Model 4: liabilities covarying with vessel diameter and plant height.
- Model 5: liabilities covarying with vessel diameter, plant height, and minimum temperature.

While other combinations of models are possible, we limited the number of models to those that appeared to have support under the causal path analyses of Section II. To assess the performance of each model in predicting the plant habit, we compared the distribution of posterior probabilities of the actual states of the 307 test taxa across models 1–5 (Fig. 6). Additionally, we randomly chose multiple sets of 307 taxa regardless of whether

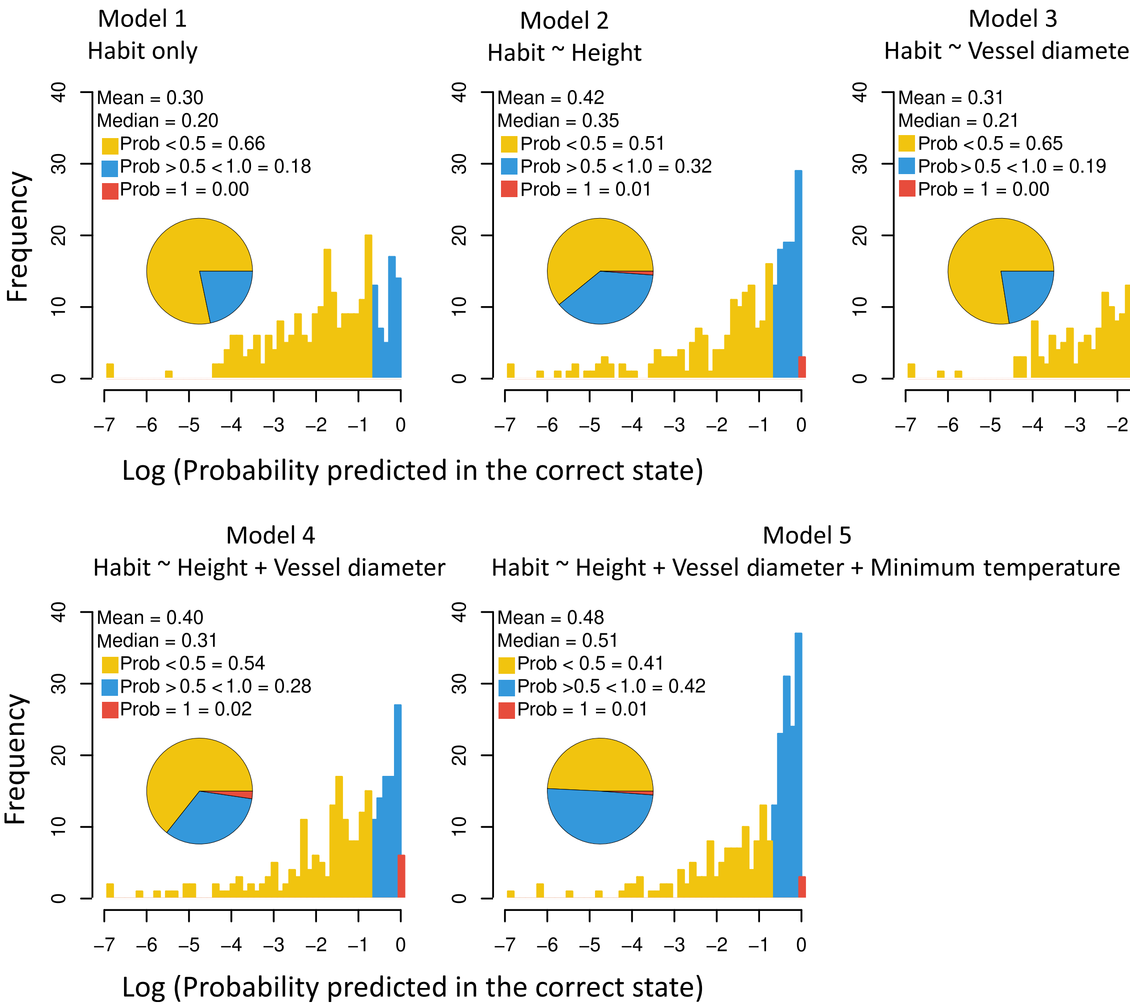


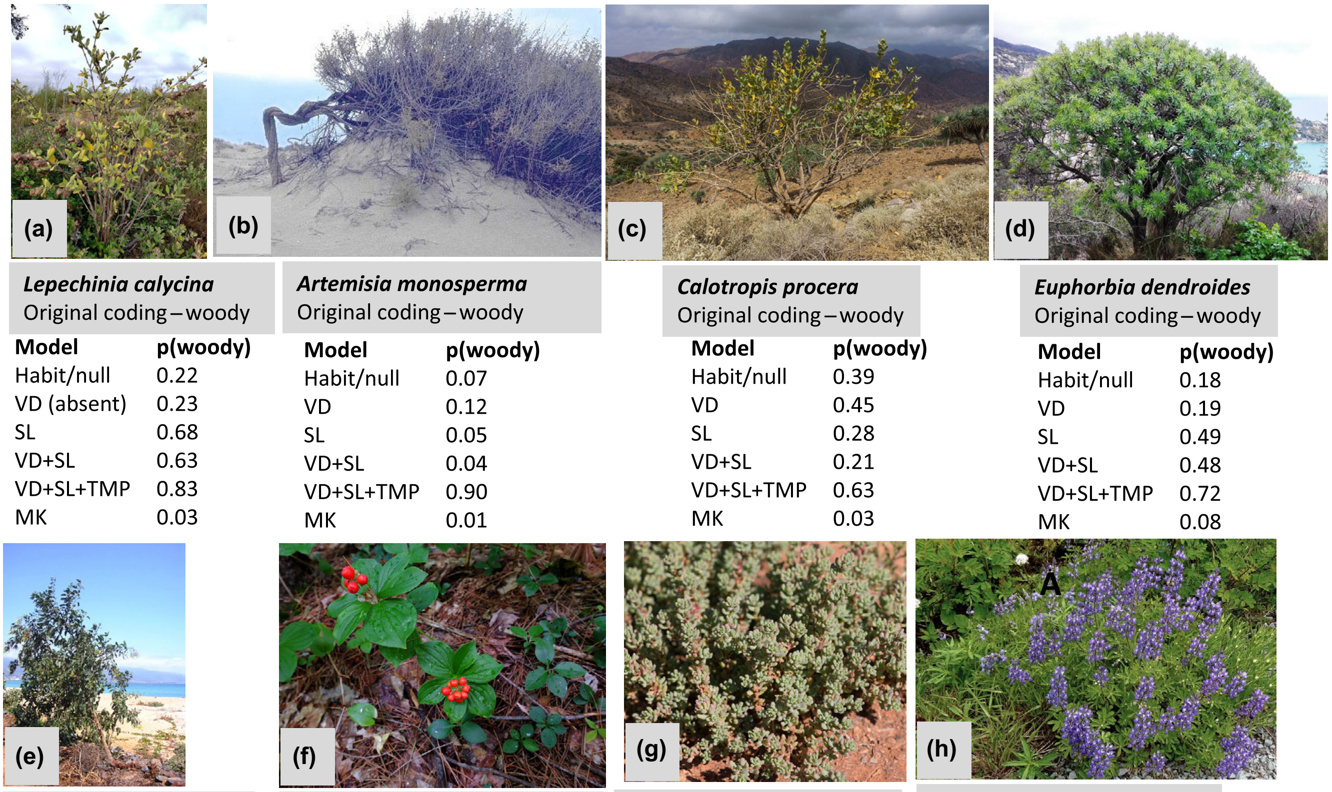
Fig. 6 Distribution of log-transformed proportions of posterior tip liabilities (*x*-axis) correctly matching the actual growth habit (herbaceous vs woody) for 307 taxa (initially misclassified and treated as missing) across five threshold models estimated using MCMCglmmRAM. Each bar (*y*-axis) represents the frequency of species with a given log-transformed proportion of correct liabilities. Species in categories farther to the right indicate higher proportions of posterior tip liabilities correctly predicting their actual state, with 100% (i.e. 1 or 0 on the log scale) appearing at the extreme right. Models showing more bars shifted toward the right indicate better overall prediction accuracy of growth habits. The inset pie charts and summary statistics illustrate the overall prediction accuracy for each model.

they were correctly predicted in the original analyses and repeated the cross-validation analysis to generate a comparison distribution of posterior probabilities to determine how the inclusion of covariates impacted overall trait reconstruction across all taxa (Fig. S5).

Comparisons show that the predictability of the threshold model improves substantially as we add covariates. We found that including covarying factors (Models 2–5) increased the probability of support for either woody or herbaceous predictions compared to the Null model (Figs 6, S5). Random subsets were overall predicted with extremely high accuracy as expected given the conserved nature of growth habit. In particular, adding height increased prediction habit state certainty for random subsets by increasing the proportion of taxa with posterior probabilities = 1 (Fig. S5), while the combined effect of height, vessel diameter, and minimum temperature have the highest prediction accuracy across both mispredicted and random subsets (Figs 6, S5).

We further inspected the 35 erroneous assignments and 209 correct assignments of their growth habits under models 1–5 and assessed how our predictions were influenced by the set of covarying factors (determined by the causal link) included in the prediction model. We particularly focused on taxa where the results were altered by the inclusion/exclusion of predictors in the model. Most of these mispredictions occurred in plants that lie in the boundary zone between herbaceous and woody states, which are often characterized as suffrutescent herbs, subshrubs, and perennials, and some of which are characterized as derived woody.

For most taxa, unsurprisingly, plant height had the largest effect (Model 2) on plant growth habit predictions. For example, height greatly improved prediction for *Lepechinia calycina* (Benth.) Epling (Fig. 7a), an evergreen perennial shrub found in the rocky slopes in California that has an extremely low probability of being woody under both null threshold (0.22) and Mk (0.03) models, but switches to being predicted to be woody when



Lepechinia calycina
Original coding – woody

Model	p(woody)
Habit/null	0.22
VD (absent)	0.23
SL	0.68
VD+SL	0.63
VD+SL+TMP	0.83
MK	0.03

Artemisia monosperma
Original coding – woody

Model	p(woody)
Habit/null	0.07
VD	0.12
SL	0.05
VD+SL	0.04
VD+SL+TMP	0.90
MK	0.01

Calotropis procera
Original coding – woody

Model	p(woody)
Habit/null	0.39
VD	0.45
SL	0.28
VD+SL	0.21
VD+SL+TMP	0.63
MK	0.03

Euphorbia dendroides
Original coding – woody

Model	p(woody)
Habit/null	0.18
VD	0.19
SL	0.49
VD+SL	0.48
VD+SL+TMP	0.72
MK	0.08

Nicotiana glauca
Original coding–woody

Model	p(woody)
Habit/null	0.16
VD	0.20
SL	0.47
VD+SL	0.42
VD+SL+TMP	0.63
MK	0.12

Cornus canadensis
Original coding – herbaceous

Model	p(woody)
Habit/null	0.96
VD (absent)	0.96
SL	0.69
VD+SL	0.78
VD+SL+TMP	0.36
MK	0.98

Osteocarpum_acropterum
Original coding – woody

Model	p(woody)
Habit/null	0.96
VD (absent)	0.96
SL	0.49
VD+SL	0.62
VD+SL+TMP	0.62
MK	0.98

Lupinus latifolius
Original coding – herbaceous

Model	p(woody)
Habit/null	0.03
VD (absent)	0.03
SL	0.84
VD+SL	0.79
VD+SL+TMP	0.93
MK	0.003

Fig. 7 Examples of mispredicted taxa and their estimated probabilities of woodiness estimated under the MK and different threshold models estimated using MCMCglmmRAM. Probabilities of woodiness were calculated as the percentage of posterior tip liabilities exceeding the threshold value. The threshold models include: (1) null (no covarying traits); (2) habit ~ cross-sectional vessel diameter (VD); (3) habit ~ stem length/plant height (SL); (4) habit ~ VD + SL; (5) habit ~ VD + SL + minimum temperature (TMP). Panels (a–f) show cases where prediction accuracy improved after including covarying traits; (g) illustrates a case where covarying traits helped detect an incorrectly assigned habit; and (h) shows a case where including covarying traits reduced prediction accuracy. Photo use permissions: (a) *Lepechinia calycina*, photo by Bill Hubick https://www.inaturalist.org/observations/53324917) used with permission; (b) *Artemisia monosperma*, photo by Dr Avinoam Danin (https://flora.org.il/en/plants/artmon/) used with permission; (c) *Calotropis procera*, photo by A. Savin (Free Art License, Wikimedia Commons, https://es.wikipedia.org/wiki/Calotropis_procera#/media/Archivo:ET_Afar_asv2018-01_img87_Abala_surroundings.jpg); (d) *Euphorbia dendroides*, photo by Anne-Hélène Paradis (https://www.inaturalist.org/observations/108786878) used with permission; (e) *Nicotiana glauca*, photo by Forest & Kim Starr (Starr Environmental, licensed under Creative Commons Attribution (CC BY) http://www.starrenvironmental.com/images/image/?q=24526521762); (f) *Cornus canadensis* by M. Costea and G. Costea (http://phytoimages.siu.edu/imgs/Cusman1/r/Cornaceae_Cornus_canadensis_75731.html) used with permission; (g) *Osteocarpum acropterum* by Kym Nicolson (https://www.inaturalist.org/observations/61635424) used with permission; (h) *Lupinus latifolius* by Jean Pawek (https://calphotos.berkeley.edu/cgi/img_query?seq_num=385101&one=T) used with permission.

height (0.68) and other covariates are included in the threshold model. Similarly, for some taxa that are particularly adapted to arid climates, including temperature in the predictive model altered results drastically. *Artemisia monosperma* Delile (Fig. 7b) is a dominant perennial shrub of Arabian deserts that was originally coded as woody. The models that exclude temperature, including the MK model, predict it to be herbaceous with high certainty (> 0.80), implying that including minimum temperature reverses this prediction to the correct state.

Similarly, woody plants that are small trees/large shrubs (e.g. *Calotropis procera* (Aiton) W.T.Aiton, *Euphorbia dendroides* L., *Nicotiana glauca* Graham; Fig. 7c–e) showed substantial model prediction improvement when minimum temperature was included in the model (Model 5). Likewise, some herbaceous taxa originally predicted to be woody had improved prediction of correct growth habit states when covariates were included. The herbaceous species, *Cornus canadensis* L., a north Eurasian-North American member of a genus that otherwise contains trees and large shrubs, was predicted by the Mk model and the univariate threshold model to be woody (> 0.96). However, inclusion of plant height drops these probabilities considerably, and support for the majority of the posterior probability becomes herbaceous once temperature was included (Fig. 7f).

Ambiguity in character coding resulting in erroneous trait assignments in our original dataset was also apparent across the 35 miscoded taxa. For example, the perennial herb *Osteocarpum acropterum* (F.Muell. & Tate) Volkens, a succulent of *c.* 20 cm high, grows in saline areas in temperate Australia. It was originally coded as woody in our database. However, when height was included in the prediction, support for woodiness decreased, favoring the herbaceous state. Without height, the threshold model results were comparable to the Mk model predictions (Fig. 7g). Similarly, the New Zealand natives *Muehlenbeckia australis* (G.Forst.) Meisn. and *Muehlenbeckia complexa* (A.Cunn.) Meisn. are robust vines with the main stem up to 10 cm in diameter. These perennial woody climbers were coded as herbaceous and predicted to be herbaceous under the MK and threshold models when height was not included (Null model).

However, for a few taxa, inclusion of covariates actually decreased support for predictions of the true growth habit (contrary to the pattern seen overall, see Figs 6, S5). For example, including height reversed the predictions for 10 species: *Thymus mongolicus* (Ronniger) Ronniger, *Artemisia maritima* L., *Cephalaralia cephalobotrys* (F.Muell.) Harms, *Batis maritima* L., *Anabasis salsa* (Ledeb.) Benth. ex Volkens, *Euphorbia hierosolymitana* Boiss, *Chenopodium album* L., *Artemisia japonica* Thunb., and *Lupinus latifolius* Lindl. ex J.Agardh (Fig. 7h).

Summary: phylogenetic prediction improves by inclusion of causal paths

Overall, we found that height and minimum temperature improve the prediction of growth habit in the mispredicted taxa, the vast majority of which appear on the 'threshold' between woody and herbaceous. These taxa represent many cases of derived woodiness, or recent transitions to herbaceousness. Manual coding of these taxa by experts can be challenging (Lens *et al.*, 2012), let alone by phylogenetic comparative modeling. In our case, it was also challenging to unambiguously identify the causal structure of growth form evolution, owing to it being a discrete trait with limited signal to evaluate conditional independence.

Interestingly, consistent with the results of the phylogenetic path analyses in Section II (Fig. S4), we found improved prediction with the inclusion of covariates, particularly height and minimum temperature. Cross-validation is particularly useful in identifying the 'tipping point' variables that underly improved predictive power as well as demonstrating exceptional cases where their inclusion runs counter to expectations. For example, we show exemplar taxa with improved assignation to the correct plant habit state (herbaceous, Fig. 7a–e; woody, Fig. 7f,g) when height and minimum temperature are added to the underlying path causal model. Taxa for whom predictive power decreases from true values have particularly odd values in our dataset. For example, *Lupinus latifolius* is typically a medium-sized perennial herb, but in our dataset carries a particularly large value for height, of *c.* 3 m. While this is potentially possible for large herbs (e.g. giant hogweed, *Heracleum mantegazzianum* Sommier & Levier), if accurate, it likely represents an extreme maximum in ideal conditions that is not typical for all the populations in the species. Examination of biological exemplars can potentially identify what is missing from the model as well – highlighting taxa with natural history features that potentially elucidate missing components of the model.

Synthesis and discussion

In this manuscript, our primary goal was to demonstrate how phylogenetic comparative methods can be used to build a growing understanding of interrelated trait relationships. In general, the main justification for using phylogenetic comparative methods is that there are phylogenetic effects in the residual variation (Revell, 2010). Such variation can potentially lead to confounding estimation of the relationships among focal variables, as well illustrated by Felsenstein's famed worst-case scenario (Felsenstein, 1985). It is common for different comparative studies to focus on different sets of focal variables, while sequentially relegating other factors – even those previously established to be important and potentially causal – to mere nuisance or background variation modeled under evolutionary processes such as Brownian Motion (Grafen, 1989). However, this practice can lead to analytical amnesia that is antithetical to the goal of building a comprehensive understanding of the biology underlying macroevolutionary patterns. Too often we find that large macroevolutionary datasets have many sets of traits with strong correlations. Thus, the challenge of macroevolutionary analyses is not simply finding significant relationships (e.g. Fig. 3), but in interpreting the complex web of interrelationships among variables. The real task lies in making sense of these correlations and linking them to underlying biological processes. Importantly, such interrelationships can also be leveraged to reconstruct unobserved variables.

In particular, little guidance has been established in the field for when a significantly covarying variable should be excluded from a comparative analysis (Farallo *et al.*, 2020; Keele *et al.*, 2020). Inappropriate inclusion of variables can create overly complex models, introduce uncertainty due to partial or incomplete data, and worst of all, potentially create bias or spurious correlations by improperly classifying variables as confounding. Here we argue that phylogenetic path analysis and the logic of causal inference is particularly valuable for guiding researchers toward meaningful macroevolutionary inferences while building a comprehensive understanding about the relationships among variables (Hardenberg & Gonzalez-Voyer, 2013). Establishing plausible Directed Acyclic Graphs (DAGs) of trait relationships can facilitate piece-wise study of trait relationships. The strength of this approach is because path analyses can identify when several variables exist along the same common causal path. Consequently, potentially alternative causal explanations for macroevolutionary patterns can be identified as pieces of the full DAG – such as we identified for the results of Zanne *et al.* (2018) and Olson *et al.* (2018), which are both consistent as partial descriptors of Fig. 5. Such causal DAGs can facilitate the identification of variables that are confounders (influence both predictor and outcome, creating spurious associations), colliders (are influenced by both the predictor and outcome, introducing bias if controlled for), and mediating variables (lie on the causal pathway between predictor and outcome) along the causal path of interest – all of which will impact how additional variables are treated in regression analyses to estimate the causal effect of interest (Pearl, 2009; Shipley, 2016; Keele *et al.*, 2020).

As an example, the previously identified relationships between vessel diameter and climate (Zanne *et al.*, 2018) could result simply as a by-product of temperature effects on plant height since vessel diameter is known to covary with plant height (Olson *et al.*, 2018). However, our phylogenetic path analysis identified a direct effect of low temperature on cross-sectional vessel area independent of its effect on plant height, as mediated with its interaction with leaf phenology (evergreen vs deciduous) and temperature. This explains well the findings of Zanne *et al.* (2018), which showed an overall negative relationship between minimum temperature at the extremes of a species' range with cross-sectional vessel area – an effect likely driven primarily by the covariance of vessel diameter and height, as in Olson *et al.* (2018). However, the difference in the shape of the relationship (sigmoidal vs linear) between evergreen and deciduous woody plants, respectively, indicates that temperature has effects on cross-sectional vessel area independent of its effect on height. One simplifying assumption in these previous studies was that all lineages remained entirely woody throughout their evolutionary history – an assumption that is known to be violated (Zanne *et al.*, 2018). Traits such as cross-sectional vessel area can be biased in their collection; vessel size is of great interest for the physiological properties of woody species, but has received far less interest in herbaceous plants, meaning there is a paucity of vessel diameter data in herbs. Such biases are common and can often limit comprehensive and realistic phylogenetic comparative methods. Nevertheless, comparative methods can potentially overcome these challenges by integrating knowledge of interacting traits within the causal model framework.

Another plant trait that is a good example showcasing a common problem in phylogenetic comparative methods is growth habit. This trait is often discretely coded (e.g. woody or herbaceous), but the underlying biological reality is one of a complex, polygenic trait with continuous variation. How should we model such traits in a comparative framework? One option is to model this unseen variation as a hidden-state model in software such as corrHMM and HiSSE (Beaulieu *et al.*, 2013; Beaulieu & O'Meara, 2016). These tools are essential components of the comparative toolkit. However, as pointed out by Tarasov (2019, see the 'Two-Scientist Paradox') – hidden states are often not particularly hidden. Indeed, trees are generally well-known to be tall plants. Therefore, the likelihood of transitions from woody to herbaceous plants are going to be impacted by this quite visible trait. While such facts are obvious, the distance between this reality and standard Mk models, or even hidden state models, becomes strikingly apparent.

Clearly, inclusion of covariates improves trait prediction for plant growth habit, and the phylogenetic distribution of discrete characters can facilitate the joint reconstruction of growth habit, plant height, vessel diameter, and minimum temperature (Figs 4, 6). Likewise, while we show improved cross-validation at the tips, similar performance should be expected for internal nodes in the phylogeny. Almost all phylogenetic comparative models account for node-based estimates of ancestral states, explicitly or implicitly – for example, integrating over uncertainty in the ancestral states of the residual variation in the case of PGLS with Brownian Motion to estimate the regression coefficients of variables. While PGLS assumes phylogenetically covarying residuals, it is an inappropriate model when extrinsic factors influence a trait with adaptation. Instead, such relationships must account also for the ancestral states of the predictor and response, and model the lag time between variables (Hansen, 2008; Uyeda *et al.*, 2018). Thus, it becomes important to distinguish between efforts to make predictions (e.g. accurately infer ancestral states) and efforts to measure hypothesized causal effects in phylogenetic comparative analyses. For one, the introduction of covariates can potentially greatly improve inference of past ancestral states and subsequently inform and reduce the uncertainty of the phylogenetic model. However, only confounding variables should be included in phylogenetic regression models intended to estimate causal effects from regression coefficients (Keele *et al.*, 2020). There are currently limited methods available for integrating a fully phylogenetic framework with the phylogeny being part of the DAG (Uyeda *et al.*, 2018).

While such methods are poorly developed, we think that progress can be made by separating out the goals of trait reconstruction, reducing the uncertainty of ancestral states, and establishing estimates of causal effects. We have shown, for example, how treating growth habit as a threshold trait and combining it with covariates can impact the inference of woodiness at the base of the eudicot phylogeny (Figs 6, S5) and incorporating this additional information can potentially eliminate uncertainty for estimation of the causal effects of, for example, climate on growth habit. While accounting for these covariances with multivariate responses can be challenging, reconstruction of the independent variable from a set of covarying traits is more straightforward. For example, the common task of visualizing evolutionary regimes ('regime painting') can be facilitated by jointly modeling plant height, vessel diameter, and growth habit using a multivariate threshold model. Stochastic mapping of the underlying liability can then be painted on the phylogeny with uncertainty and used as a subsequent cause of evolution in other traits. Ideally, such analyses could be conducted as a joint inference (Höhna *et al.*, 2016, 2017; Landis *et al.*, 2018).

Conclusion

The distance between 'out-of-the-box' methods in the standard comparative toolkit, and the biological reality of traits is often immense and can limit inferences in the face of large multi-trait datasets. In particular, the solution appears not to be more complex models, but more careful thought about the biology and the interrelationships among traits, as well as understanding the nature of cause and effect in the system. Here we elaborate on one approach to synthesize and build understanding of plant growth habit evolution and its underlying causes, better connecting the traits to their biological underpinnings. We highlight future directions for phylogenetic comparative methods and the need to provide guidance and resources for making sense of macroevolutionary patterns.

Acknowledgements

We thank Anna Soffin for assistance in verifying growth forms for several mispredicted species. We are also grateful to Sergei Tarasov, Nic Bone, and Sean McHugh for their helpful comments during the conception of this project. JCU was supported by NSF-DEB-1942717, and SN was supported by startup funds provided to JCU during his time at Virginia Tech.

Competing interests

None declared.

Author contributions

JCU and SN conceived the project. JCU, SN, AEZ, and FL planned and conceptualized the research. FL and AEZ provided the data. SN and JCU performed the analyses and wrote the manuscript, and AEZ and FL reviewed and edited multiple versions.

ORCID

Frederic Lens https://orcid.org/0000-0002-5001-0149
Suman Neupane https://orcid.org/0000-0002-8094-3775
Josef C. Uyeda https://orcid.org/0000-0003-4624-9680
Amy E. Zanne https://orcid.org/0000-0001-6379-9452

Data availability

All data files, analysis scripts, and model outputs supporting this study are archived in the Dryad Digital Repository (https://doi.org/10.5061/dryad.dfn2z35cg). Additional Supporting Information includes Dataset S1 (data sources and filtering steps), Notes S1 (methodological details of MCMCglmmRAM analyses), and Figs S1–S5.

References

Beaulieu JM, O'Meara BC. 2014. Hidden Markov models for studying the evolution of binary morphological characters. In: Garamszegi LZ, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology.* Berlin, Heidelberg, Germany: Springer, 395–408.

Beaulieu JM, O'Meara BC, Donoghue MJ. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic Biology* **62**: 725–737.

Beaulieu JM, O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* **65**: 583–601.

van der Bijl W. 2018. PHYLOPATH: easy phylogenetic path analysis in r. *PeerJ* **6**: e4718.

Blomberg SP, Rathnayake SI, Moreau CM. 2020. Beyond Brownian motion and the Ornstein–Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. *The American Naturalist* **195**: 145–165.

Böhle U-R, Hilger HH, Martin WF. 1996. Island colonization and evolution of the insular woody habit in Echium l.(Boraginaceae). *Proceedings of the National Academy of Sciences, USA* **93**: 11740–11745.

Bollback JP. 2006. Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**: 88.

Carlquist. 2009. Xylem heterochrony: an unappreciated key to angiosperm origin and diversifications. *Botanical Journal of the Linnean Society* **161**: 26–65.

Carlquist. 2012. How wood evolves: a new synthesis. *Botany* **90**: 901–940.

Carlquist S. 1974. *Island biology, chap. 10.* New York, NY, USA: Columbia University Press, 350–428.

Choat B, Brodribb TJ, Brodersen CR, Duursma RA, López R, Medlyn BE. 2018. Triggers of tree mortality under drought. *Nature* **558**: 531–539.

Cooper N, Thomas GH, FitzJohn RG. 2016. Shedding light on the 'dark side' of phylogenetic comparative methods. *Methods in Ecology and Evolution* **7**: 693–699.

Cornwell WK, FitzJohn RG, Stevens PF, Calaminus A, Soltis DE, Soltis PS, Govaerts R, Wright IJ, Oleksyn J, Reich PB *et al.* 2013. Global woodiness database. Data from: Three keys to the radiation of angiosperms into freezing environments. doi: 10.5061/dryad.63q27/2.

Darwin C. 1859. *The origin of species by means of natural selection.* London, UK: John Murray.

Díaz S, Kattge J, Cornelissen JH, Wright IJ, Lavorel S, Dray S, Reu B, Kleyer M, Wirth C, Prentice IC *et al.* 2016. The global spectrum of plant form and function. *Nature* **529**: 167–171.

Dória LC, Meijs C, Podadera DS, Del Arco M, Smets E, Delzon S, Lens F. 2019. Embolism resistance in stems of herbaceous Brassicaceae and Asteraceae is linked to differences in woodiness and precipitation. *Annals of Botany* **124**: 1–14.

Dória LC, Podadera DS, del Arco M, Chauvin T, Smets E, Delzon S, Lens F. 2018. Insular woody daisies (Argyranthemum, Asteraceae) are more resistant to drought-induced hydraulic failure than their herbaceous relatives. *Functional Ecology* **32**: 1467–1478.

Doyle JA. 2012. Molecular and fossil evidence on the origin of angiosperms. *Annual Review of Earth and Planetary Sciences* **40**: 301–326.

Farallo VR, Muñoz MM, Uyeda JC, Miles DB. 2020. Scaling between macro-to microscale climatic data reveals strong phylogenetic inertia in niche evolution in plethodontid salamanders. *Evolution* **74**: 979–991.

Feild TS, Arens NC, Doyle JA, Dawson TE, Donoghue MJ. 2004. Dark and disturbed: a new image of early angiosperm ecology. *Paleobiology* **30**: 82–107.

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

Felsenstein J. 2005. Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **360**: 1427–1434.

Frankiewicz KE, Banasiak Ł, Oskolski AA, Magee AR, Alsarraf M, Trzeciak P, Spalik K. 2021. Derived woodiness and annual habit evolved in African umbellifers as alternative solutions for coping with drought. *BMC Plant Biology* **21**: 1–18.

Frankiewicz KE, Oskolski AA, Banasiak Ł, Fernandes F, Reduron JP, Reyes-Betancort JA, Szczeparska L, Alsarraf M, Baczyński J, Spalik K. 2020. Parallel evolution of arborescent carrots (Daucus) in Macaronesia. *American Journal of Botany* **107**: 394–412.

Freckleton RP. 2009. The seven deadly sins of comparative analysis. *Journal of Evolutionary Biology* **22**: 1367–1375.

Gasson P. 2011. How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES. *IAWA Journal* **32**: 137–154.

Gauthey A, Peters JMR, López R, Carins-Murphy MR, Rodriguez-Domínguez CM, Tissue DT, Medlyn BE, Brodribb TJ, Choat B. 2022. Mechanisms of xylem hydraulic recovery after drought in *Eucalyptus saligna*. *Plant, Cell & Environment* **45**: 1216–1228.

Givnish TJ. 1995. Plant stems: Biomechanical adaptation for energy capture and influence on species distributions. In: Gartner BL, ed. *Plant stems: physiology and functional morphology*. New York, NY, USA: Chapman & Hall, 3–49.

Givnish TJ. 1998. Adaptive plant evolution on islands: classical patterns, molecular data, new insights. In: Grant PR, ed. *Evolution on islands*. Oxford, UK: Oxford University Press, 281–304.

Grafen A. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **326**: 119–157.

Hadfield J, Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology* **23**: 494–508.

Hadfield JD. 2010. Mcmc methods for multi-response generalized linear mixed models: the MCMCGLMM r package. *Journal of Statistical Software* **33**: 1–22.

Hadfield JD. 2015. Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models. *Methods in Ecology and Evolution* **6**: 706–714.

Hansen TF. 2008. Macroevolutionary quantitative genetics? a comment on Polly (2008). *Evolutionary Biology* **35**: 182–185.

Hansen TF, Pienaar J, Orzack SH. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* **62**: 1965–1977.

Hardenberg A v, Gonzalez-Voyer A. 2013. Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution* **67**: 378–387.

Hijmans R, Cameron S, Parra J, Jones P, Jarvis A. 2004. *The worldclim interpolated global terrestrial climate surfaces. v.1.3*. [WWW document] URL http://www.worldclim.org/ [accessed 13 February 2018].

Höhna S, Landis MJ, Heath TA. 2017. Phylogenetic inference using RevBayes. *Current Protocols in Bioinformatics* **57**: 6–16.

Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* **65**: 726–736.

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. *Systematic Biology* **52**: 131–138.

IAWA Committee. 1989. IAWA list of microscopic features for hardwood identification. *IAWA Bulletin, New Series* **10**: 219–332.

Kattge J, Diaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ *et al.* 2011. TRY–a global database of plant traits. *Global Change Biology* **17**: 2905–2935.

Keele L, Stevenson RT, Elwert F. 2020. The causal interpretation of estimated associations in regression models. *Political Science Research and Methods* **8**: 1–13.

Kidner C, Groover A, Thomas DC, Emelianova K, Soliz-Gamboa C, Lens F. 2016. First steps in studying the origins of secondary woodiness in Begonia (Begoniaceae): combining anatomy, phylogenetics, and stem transcriptomics. *Biological Journal of the Linnean Society* **117**: 121–138.

Kim S, Soltis DE, Soltis PS, Zanis MJ, Suh Y. 2004. Phylogenetic relationships among early-diverging eudicots based on four genes: were the eudicots ancestrally woody? *Molecular Phylogenetics and Evolution* **31**: 16–30.

Klimeš A, Šímová I, Zizka A, Antonelli A, Herben T. 2022. The ecological drivers of growth form evolution in flowering plants. *Journal of Ecology* **110**: 1525–1536.

Landis MJ, Freyman WA, Baldwin BG. 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution* **72**: 2343–2359.

Lee C, Kim S-C, Lundy K, Santos-Guerra A. 2005. Chloroplast DNA Phylogeny of the woody *Sonchus* alliance (Asteraceae: Sonchinae) in the Macaronesian islands. *American Journal of Botany* **92**: 2072–2085.

Lens F, Davin N, Smets E, del Arco M. 2013. Insular woodiness on the Canary Islands: a remarkable case of convergent evolution. *International Journal of Plant Sciences* **174**: 992–1013.

Lens F, Eeckhout S, Zwartjes R, Smets E, Janssens SB. 2012. The multiple fuzzy origins of woodiness within Balsaminaceae using an integrated approach: where do we draw the line? *Annals of Botany* **109**: 783–799.

Lens F, Liang C, Guo Y, Tang X, Jahanbanifard M, da Soares Correa Silva F, Ceccantini G, Verbeek F. 2020. Computer-assisted timber identification based on features extracted from microscopic wood sections. *IAWA Journal* **41**: 660–680.

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913–925.

Louca S, Doebeli M. 2018. Efficient comparative phylogenetics on large trees. *Bioinformatics* **34**: 1053–1055.

Luo A, Xu X, Liu Y, Li Y, Su X, Li Y, Lyu T, Dimitrov D, Larjavaara M, Peng S *et al.* 2023. Spatio-temporal patterns in the woodiness of flowering plants. *Global Ecology and Biogeography* **32**: 384–396.

Melzer S, Lens F, Gennen J, Vanneste S, Rohde A, Beeckman T. 2008. Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nature Genetics* **40**: 1489–1492.

Neupane S, Lewis PO, Dessein S, Shanks H, Paudyal S, Lens F. 2017. Evolution of woody life form on tropical mountains in the tribe Spermacoceae (Rubiaceae). *American Journal of Botany* **104**: 419–438.

Nürk NM, Atchison GW, Hughes CE. 2019. Island woodiness underpins accelerated disparification in plant radiations. *New Phytologist* **224**: 518–531.

O'Meara BC. 2012. Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics* **43**: 267–285.

Olson ME, Anfodillo T, Gleason SM, McCulloh KA. 2021. Tip-to-base xylem conduit widening as an adaptation: causes, consequences, and empirical priorities. *New Phytologist* **229**: 1877–1893.

Olson ME, Anfodillo T, Rosell JA, Petit G, Crivellaro A, Isnard S, León-Gómez C, Alvarado-Cárdenas LO, Castorena M. 2014. Universal hydraulics of the flowering plants: vessel diameter scales with stem length across angiosperm lineages, habits and climates. *Ecology Letters* **17**: 988–997.

Olson ME, Rosell JA. 2013. Vessel diameter–stem diameter scaling across woody angiosperms and the ecological causes of xylem vessel diameter variation. *New Phytologist* **197**: 1204–1213.

Olson ME, Soriano D, Rosell JA, Anfodillo T, Donoghue MJ, Edwards EJ, León-Gómez C, Dawson T, Camarero Martínez JJ, Castorena M *et al.* 2018. Plant height and hydraulic vulnerability to drought and cold. *Proceedings of the National Academy of Sciences, USA* **115**: 7551–7556.

Pearl J. 2009. *Causality*. Cambridge, UK: Cambridge University Press.

Pearl J. 2014. Comment: understanding Simpson's paradox. *The American Statistician* **68**: 8–13.

Rahimi A, Karami O, Lestari AD, de Werk T, Amakorová P, Shi D, Novák O, Greb T, Offringa R. 2022. Control of cambium initiation and activity in Arabidopsis by the transcriptional regulator ahl15. *Current Biology* **32**: 1764–1775.

Revell. 2014. Ancestral character estimation under the threshold model from quantitative genetics. *Evolution* **68**: 743–759.

Revell LJ. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* **1**: 319–329.

Revell LJ. 2012. PHYTOOLS: an r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Schweingruber F, Landolt W. 2005. *The xylem database*. [WWW document] URL http://www.wsl.ch/dendropro/xylemdb/ [accessed 13 March 2019].

Shipley B. 2016. *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference with R*. Cambridge, UK: Cambridge University Press.

Shipley B. 2000. A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling* **7**: 206–218.

Smith SA, Beaulieu JM. 2009. Life history influences rates of climatic niche evolution in flowering plants. *Proceedings of the Royal Society B: Biological Sciences* **276**: 4345–4352.

Smith SA, Brown JW. 2018. Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* **105**: 302–314.

Smith SA, Donoghue MJ. 2008. Rates of molecular evolution are linked to life history in flowering plants. *Science* **322**: 86–89.

Tarasov S. 2019. Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. *Systematic Biology* **68**: 698–716.

Uyeda JC, Harmon LJ. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Systematic Biology* **63**: 902–918.

Uyeda JC, Zenil-Ferguson R, Pennell MW. 2018. Rethinking phylogenetic comparative methods. *Systematic Biology* **67**: 1091–1109.

Wallace AR. 1878. *Tropical nature and other essays*. London, UK: Macmillan.

Westoby M, Falster DS, Moles AT, Vesk PA, Wright IJ. 2002. Plant ecological strategies: some leading dimensions of variation between species. *Annual Review of Ecology and Systematics* **33**: 125–159.

Wright IJ, Royer DL, Ordonez A, Oleksyn J, Reich PB. 2014. Global Leaf Phenology Database. Data from: Three keys to the radiation of angiosperms into freezing environments. *Dryad Digital Repository*. doi: 10.5061/dryad.63q27/3.

Wright S. 1934. An analysis of variability in number of digits in an inbred strain of Guinea pigs. *Genetics* **19**: 506–536.

Zanne AE, Pearse WD, Cornwell WK, McGlinn DJ, Wright IJ, Uyeda JC. 2018. Functional biogeography of angiosperms: life at the extremes. *New Phytologist* **218**: 1697–1709.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* **506**: 89–92.

Zanne AE, Westoby M, Falster DS, Ackerly DD, Loarie SR, Arnold SE, Coomes DA. 2010. Angiosperm wood structure: global patterns in vessel anatomy and their relation to wood density and potential conductivity. *American Journal of Botany* **97**: 207–215.

Zizka A, Onstein RE, Rozzi R, Weigelt P, Kreft H, Steinbauer MJ, Bruelheide H, Lens F. 2022. The evolution of insular woodiness. *Proceedings of the National Academy of Sciences, USA* **119**(37): e220862911.

Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C *et al.* 2024. Phylogenomics and the rise of the angiosperms. *Nature* **629**: 843–850.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Trait datasets, filtering steps.

Fig. S1 SIMMAP reconstruction of ancestral woodiness across the eudicot phylogeny.

Fig. S2 SIMMAP reconstructions and BAYOU analyses of vessel size evolution in eudicots.

Fig. S3 Directed acyclic graphs of the top five phylogenetic path models showing significant causal relationships among traits and temperature.

Fig. S4 Averaged phylogenetic path models showing key causal links among plant habit, other traits, and minimum temperature.

Fig. S5 Prediction accuracy of growth habit for 307 eudicot taxa across five threshold-liability models estimated using MCMCglmmRAM.

Notes S1 Phylogenetic threshold and multivariate MCMCglmmRAM analyses.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.