

# Unravelling the phenology-phylogeny tangle.

December 19, 2022

Authors:

The Wolkovich Lab in 2019 & collaborators <sup>1,2,3,4</sup>

*Author affiliations:*

<sup>1</sup>Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4;

<sup>2</sup>Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA;

<sup>3</sup>Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA;

<sup>4</sup>Edificio Ciencias, Campus Universitario 28805 Alcalá de Henares, Madrid, Spain

\*Corresponding author: ignacio.moralesc@uah.es

## Abstract

Mitigation and adaptation to ongoing climate change hinge on accurate ecological forecasting to predict shifts in key ecosystem services, such as carbon storage and biodiversity maintenance. Forecasts to date, however, have generally failed to capture the important variability in biological responses, especially observed across species. Here, we use a Bayesian hierarchical phylogenetic model to overcome this challenge and estimate species-level phenological responses to two major environmental cues: temperature and daylength. We find that species-specific variability is higher than variation across cues, reinforcing the idea that understanding how each species responds to a combination of cues may be more critical than knowing what cue is stronger. Further, our model provides insights on how evolutionary history has shaped responses to cues, suggesting greater evolution of responses to temperature—especially cool winter temperatures—than daylength. Our approach provides a major advance in ecological forecasting, with implications for predicting impacts on key ecosystem services from climate change, biological invasions and related threats to biodiversity.

## Introduction

The biological impacts of climate change will have major implications for the future sustainability of ecosystems. With rising global temperatures species have shifted northward in space and earlier in time on average (IPCC, 2014; Parmesan and Yohe, 2003), against a background of high variability. These shifts have cascading consequences on many ecosystem services including carbon storage, making both mitigation and human adaptation to future warming dependent on accurate ecological forecasts.

While ecological forecasting has improved over recent years (Dietze, 2017; Lewis et al., 2022), it remains a challenge to reproduce the high variability observed in responses to date (IPCC, 2014). Some of this variability results from the complexity of climate change itself, including regional and seasonal variation in warming that underlies average trends alongside shifts in other climate axes (e.g. precipitation). Much of it, however, is driven by species-specific variation, reflecting evolved differences in species’ sensitivities to underlying environmental cues and their interactions, which we know well for only a few well-studied species. In the absence of detailed data on individual species, species groupings (e.g., functional groups) have improved ecosystem models, but capture only a fraction of the important variability. Improving forecasts, thus, will require models that accurately predict species-level differences in responses to complex environmental change.

Recent efforts that have attempted to model species-specific responses to the environment (Diez et al., 2012) are often confounded by data availability—especially the common problem of data highly biased to some species and sparse across others. The rise of Bayesian hierarchical models can allow inference across species in such cases. However, underlying most hierarchical models is an implicit assumption that species are exchangeable, and they thus partially pool (‘shrink’) towards estimates for species with the most data (and least variable responses) (CITES), making inference at the species-level unreliable (Ettinger et al., 2020).

Including the evolutionary history of species relationships in models of species responses could at once provide more robust species-level estimates than current approaches and a better understanding of the evolutionary constraints that might limit future adaptation to change. For example, strong phylogenetic niche conservatism (Wiens et al., 2010) could potentially inhibit adaptive responses by drawing species back to an evolutionary conserved optimum, which is sub-optimal under new conditions.

Research using long-term observational data has especially highlighted the role that evolutionary history may play in structuring plant phenological responses—which are critical to accurate forecasts of carbon storage. Phylogenetic signal in plant phenology (Kochmer and Handel, 1986; Willis et al., 2008; Davies et al., 2013)) suggests species responses to cues have diverged over macro-evolutionary timescales, helping explain species present day differences. However, approaches using traditional phylogenetic comparative methods (e.g., Willis et al.,

2008; CaraDonna and Inouye, 2014; Yang et al., 2021), have produced conflicting results. Evidence for phylogenetic conservatism appears to depend on method and species, even varying from one site to the next for the same clade (e.g., Rafferty and Nabity, 2017), which violates the fundamental idea of shared evolutionary history—the common ancestor of two sets of species cannot possess two separate evolutionary histories for the same trait.

To generate robust ecological forecasts will require addressing a second major hurdle—underlying environmental cues that are complex and interacting. Decades of research have informed our understanding of how species use environmental cues to time their phenotypic responses with the temporal distribution of key resources and to avoid periods of high abiotic or biotic stress (Larcher, 1980; Bonamour et al., 2019). Commonly, however, responses to environmental cues, and their evolution, are studied individually, for example, linking a given phenotypic response to a single cue, such as time of leafout responding to summed heat during early spring (e.g., Davies et al., 2013). These efforts ignore a more likely scenario for most phenotypic traits where multiple cues interacting along evolutionary history have shaped species’ responses (Ackerly, 2009). For many plant species, phenological events are determined by a combination of temperature and light (Chuine and Regniere, 2017). It is likely that additional cues (such as soil moisture) and species’ physiology, further mediate species responses and, although they are often less well understood (Chuine and Regniere, 2017). They can be accounted for in models either as latent processes or by allowing non-stationarity in responses across species (Davies et al., 2019).

Spring plant phenology may represent our best opportunity to improve forecasts of species’ responses to interacting environmental cues. Beyond being the most studied biological impact of climate change, the primary cue system is well established (Chuine and Regniere, 2017), especially for temperate woody species where phenology is generally thought to be determined by two components of temperature—chilling (cool temperatures during dormancy period over winter) and forcing (warm temperatures, generally in the spring)—and photoperiod (Ettinger et al., 2021). Plant phenology is also one of few phenotypic traits with extensive experimental data on responses to multiple environmental cues across species. Recent multi-species analyses considering forcing, chilling and photoperiod have shown that chilling and forcing together often determine complex non-linear responses to warming, but cannot forecast beyond several well-studied species (Ettinger et al., 2020).

Here we present a novel Bayesian framework that extends upon phylogenetic mixed models (Housworth et al., 2004) to examine how chilling, forcing (metrics of temperature) and photoperiod together determine plant phenology. By allowing non-stationarity in species responses across phylogeny (Davies et al., 2019), our model departs from most previous work and assumptions of traditional phylogenetic comparative methods (e.g. Freckleton et al., 2002; Ives and Helmus, 2011; Hadfield, 2010), and move towards integrating evolutionary history in models of phenological responses to environmental change. We illustrate our method with an unprecedented dataset on phenological responses to environmental cues determined experimentally for 191 deciduous woody species (by far the most studied group of species in phenology experi-

ments, see Ettinger et al., 2020). Our method allows us to ask: which cue has the largest effect on budburst, how do cues vary across species, and how evolutionary history has shaped species responses to individual cues? Our results allow us to identify historical regime shifts (Uyeda et al., 2017) in phenological responses across the plant phylogeny, and have relevance for forecasting under ongoing change.

## Methods

*Working on this below for now ...*

## Results & Discussion

Most species are sensitive to all three primary cues—forcing, chilling, and photoperiod (Figs. 1, Supporting Table ??)(see also Laube et al., 2014; Ettinger et al., 2020)—with sensitivity to chilling approximately five-fold greater than sensitivity to photoperiod, on average (phenological advances of 6.8 days per standardized unit vs 1.2 days, for chilling and photoperiod, respectively; see Table ??). However, these average sensitivities estimated across species, fail to capture the large differences in species’ responses to both chilling and forcing (Figs. 1, Supporting Table ??). By allowing species responses to vary, based on a model including their shared evolutionary history, we found species differences dwarfed differences between cues, especially temperature cues (Fig. 1). The largest in magnitude cue—chilling—varied 25-fold between species, while variation to forcing varied 6-fold. These results suggest the strong focus on comparing cues may miss the largest variation. Robust phenological forecasts must account for both the complexity of multiple cues and species-level variation in responses to them.

### *Differences across clades & cues*

The large differences across species produced striking differences between clades. Oaks and beeches (Fagaceae), elms (Ulmaceae) and buckthorns (Rhamnaceae) are highly sensitive to chilling while rhododendrons (Ericaceae), butterfly bushes (Scrophulariaceae) or spindles (Celastraceae) show little to no response to chilling (Fig. 1 a). Similar clade-level variation was observed for forcing, where some of these clades—e.g., Ericaceae, Rhamnaceae, Ulmaceae, or Fagaceae—were particularly sensitive (advancing their budburst more than 10 days per standardized unit of forcing) and others such as the Sapindaceae, Cornaceae or Juglandaceae show little response (Fig. 1 b).

Some species responded strongly to both temperature cues, which could suggest the existence of syndromes where the genetic basis for responses to one cue—e.g., forcing—has been selected for alongside responses to another cue—e.g. chilling. We might expect, for example, early-season species to be sensitive to multiple cues as mistiming phenology could result in exposure to harsh conditions, leading to tissue loss or death (Sakai and Larcher, 1987), and sensitivity to multiple cues likely provides greater insurance against mistiming (Wolkovich and Donahue, 2021). Alternatively or additionally, linkage or pleiotropism among loci associated with sensitivity to different cues (Nakagawa et al., 2005) could induce correlated cues. However, the correlation in species sensitivities across cues was weak ( $r = 0.31$ ; between forcing and chilling) and some genera, such as *Tilia*, and *Rhododendron* (Ericaceae), displayed strong responses to forcing but weak responses to chilling, while others, such as *Acer* (Sapindaceae), show moderately strong responses to chilling but weak responses to forcing (Fig. 1). Species sensitivity to one cue, thus, does not constrain sensitivity to another cue, and it seems selection can operate independently on responses to different cues (Bonamour et al., 2019).

In contrast to both chilling and forcing, species-level responses to photoperiod were almost

uniform across species. This provides novel insight on a large debate over the prevalence of photoperiod cues in temperate trees, where previous experiments (Basler and Körner, 2012; Zohner et al., 2016) and models (e.g., Hunter and Lechowicz, 1992; Schaber and Badeck, 2003) suggested important variability across species that may constrain the responses of certain species to warming (Way and Montgomery, 2015). Our results suggest variability is limited to a handful of species in Fagaceae, which have been heavily studied, especially *Fagus sylvatica* (e.g., Basler and Körner, 2012; Zohner et al., 2016; Kramer et al., 2017). These same few species are also where most evidence of local adaptation in photoperiod cues for spring phenology comes from (e.g., Kramer et al., 2017), in contrast with common garden studies of other species, which found little evidence of local adaptation in spring (but not fall) phenology (Aitken and Bemmels, 2016). The uniformity of response to photoperiod that we found supports this latter view of generally low local adaptation in photoperiod cues for spring phenology—if local adaptation was high in photoperiod cues we would have expected more variability across species.

#### *Phylogenetic structure of phenological cues*

Variation—or lack thereof—in cues across species and clades provides possible insights into the evolution of cues across the phylogeny. While responses to each cue were phylogenetically structured, with closely related species exhibiting more similar sensitivities than distantly related species, the strength of phylogenetic conservatism in response differs between cues (Fig. 2). Responses to temperature (forcing and chilling) were moderately structured ( $\lambda = 0.65$  and  $\lambda = 0.54$ , for forcing and chilling, respectively). Phylogenetic structure in species responses to photoperiod was comparatively weak ( $\lambda = 0.39$ ) (see Fig. ??, Table ??), emphasizing again that photoperiod diverges from other climate cues. In addition to being weaker and more uniform across species, sensitivity to photoperiod appears to also evolutionarily constrained than sensitivities to temperature cues.

Differences between species in their temperature responses represent shifts in the slope of the relationship between the observed phenology and the cue. The observed phylogenetic structure in temperature responses (forcing and chilling) would be consistent with an interaction with a non-measured (i.e., latent) trait that moderates responses, and which also covaries with phylogeny (Davies et al., 2019). This fits fundamentally with the idea that early-season phenology plays a critical role in shaping species temporal niches (Gotelli and Graves, 1996) and thus should covary with a suite of life-history traits, including whether species are early-active with rapid return on investment trait, or start later in the season and have traits associated with higher competitive abilities (e.g., Grime, 1977; Wolkovich and Cleland, 2014).

Weak phylogenetic signal in photoperiod sensitivity (Fig. 2) might seem at odds with the uniformity of species response. However, somewhat counterintuitively, both uniform and random responses can manifest as low phylogenetic signal when estimated by Brownian motion expectations (Wiens et al., 2010). While rapid local adaptation within species might erase the phylogenetic structure in photoperiod responses, it does not agree with the uniformity we find in species'



responses. If responses to photoperiod evolved early in plants, as seems likely (Serrano-Bueno et al., 2017), and subsequent selection on photoperiod sensitivity was constrained by stabilizing selection operating on other life-history attributes sensitive to photoperiod (e.g., Rinne et al., 1994; Wilczek et al., 2014; Azeez and Sane, 2015), we would predict both low interspecific variation and weak phylogenetic signal in responses, matching observations. This latter interpretation is also consistent with our estimates of lower  $\sigma$  for photoperiod responses (Fig. 2). Here, as in more traditional phylogenetic comparative methods,  $\sigma$  represents the rate of evolution, and thus our results suggest photoperiod responses are also evolving slower than temperature responses (and we can add in ref to supp)

Phylogenetic conservatism (high  $\lambda$ ) and slow evolutionary rates (low  $\sigma$ ) in traits has sometimes been interpreted as indicative of evolutionary constraints to adaptive change (Wiens et al., 2010; Bennett et al., 2021). If this were the case, we might then conclude that species with strong forcing response might be more vulnerable to future warming because phylogenetic conservatism ( $\lambda$ ) in responses to forcing is higher compared to other cues, and its evolutionary rate ( $\sigma$ ) is lower. This is misleading, however, as estimates of  $\lambda$  are independent from the rate of evolution, and macroevolutionary rates are estimated on phylogenetic trees that integrate across millions of years of evolutionary history, and thus do not necessarily inform us of maximum possible rates of evolution over much shorter timescales. Indeed, there is accumulating evidence for rapid evolution to shifting climates (Bradshaw and Holzapfel, 2006; Franks et al., 2014). Our estimates are thus more useful in providing unique insights into the evolutionary history of phenological cues, and emphasize the critical importance of incorporating species-level differences in ecological forecasts

#### *Forecasting species-level responses*

Our results highlight that species-level variability can be extremely high—when properly estimated. Our approach, which partially pooled species responses based on their shared evolutionary history, did not have major effects on the overall average across species (model slopes for forcing and chilling shifted by 7.2% and 3.7%, respectively; Fig. 3), but estimated substantially higher variation across species compared with currently more widely used hierarchical models. This was especially noticeable in temperature responses (for chilling variance across species was  $\beta_{phylo} = 23.45$  compared to  $\beta_{non-phylo} = 17.47$ , for forcing it was  $\beta_{phylo} = 8.74$  compared to  $\beta_{non-phylo} = 5.01$ ) while photoperiod, which had low phylogenetic structure was more similar across approaches (variance of  $\beta_{phylo} = 0.82$  compared to  $\beta_{non-phylo} = 0.93$ ). The increase in variability across species in our model with phylogenetic structure decreased the uncertainty in estimates for each individual species temperature responses (by 6% for chilling, 6% for forcing and 2% for photoperiod). Thus, traditional approaches that partially pool across species (most hierarchical models in ecology) may also lead to less precise predictions and forecasts of phenology for individual species, although overall model accuracy might still appear reasonable (see Appendix XX in Supporting Information).

The contrasts between temperature sensitivities and photoperiod sensitivities—in both their variability across species and phylogenetic structure—have important implications for generating multi-species forecasts. For temperature responses, the large variability among species makes predicting species individual responses challenging, but the phylogenetic structure in responses lets us borrow information from close relatives to improve our predictions. While this must be done with care (Molina-Venegas et al., 2018), our results could provide reasonable predictions for many temperate woody plant species from clades for which we have multiple species and experimental observations at varying treatment levels (e.g., ANY EXAMPLES HERE?). For photoperiod, the weak phylogenetic structure in responses indicates that phylogeny retains little information, but because species responses are generally more uniform, we can be more confident in assuming the mean species response across species, comforted in the knowledge that small errors will likely not have large impact given the relatively weak overall contribution of this cue, except for *Fagus sylvatica*. As *Fagus sylvatica* is nearly five times more sensitive to photoperiod than most other measured tree species, our results caution against using it to draw inferences of photoperiod responses more widely.

While we focused on spring phenology here, our modeling approach suggests a path towards more robust forecasting via species-level estimates of climate change responses. Our results show how including the phylogenetic relationship of species can overcome major limitations of most current hierarchical models (producing biased model estimates, underestimating the full variability across species, and increasing uncertainty around individual species’ estimates) while at once providing insight into the evolutionary history of biological responses. Improving forecasting of biological responses to climate allows anticipating climate change impacts on key ecosystem services (e.g., carbon storage, others?) and planning adaptive measures accordingly.

## Methods in short form

*Data sources:* To estimate phenological responses to chilling, forcing and photoperiod we used data from phenological experiments of temperate woody species conducted in controlled environments, brought together in the Observed Spring Phenology Responses in Experimental Environments (OSPREE) database (Wolkovich et al., 2019), which we updated in July 2019 following previous methods (see full Methods & Materials below). We focused on angiosperms as gymnosperms are very poorly represented in spring phenology experiments and pruned Smith and Brown (2018) to our species, adding species that were not present in the megatree as polytomies at the generic level (using the function *congeneric.merge*; (Pearse et al., 2015)).

## Methods & Materials (Full)

Chunks to maybe work in...

- Common phylogenetic regression accounts for phylogenetic relationships as a grouping factor either explicitly (PMM) or implicitly (PGLS). Here we present one possible approach that accounts for more complex interactions going on among predictors, which would be reflected in the species-level slopes being allowed to vary as a function of the phylogeny, rather than keeping slopes constant and only allowing the intercepts (or residuals) to vary.
- In a first attempt at establishing whether or not it is important, we compare results from a common hierarchical model with partial pooling on the slopes that does not allow for phylogenetic constraints to affect slope estimates against results from a phylogenetic hierarchical model allowing phylogeny to constrain partially pooled slopes.

## Phenological and Phylogenetic Data

*Phenological data:* To estimate phenological responses to chilling, forcing and photoperiod we used data from phenological experiments of temperate woody species conducted in controlled environments, brought together in the Observed Spring Phenology Responses in Experimental Environments (OSPREE) database. In July 2019, we updated an earlier version of this database (Wolkovich et al., 2019) by reviewing all papers found through searching ISI Web of Science and Google Scholar with the following terms:

1. TOPIC = (budburst OR leaf-out) AND (photoperiod OR daylength) AND temperature\*, which yielded 623 publications
2. TOPIC = (budburst OR leaf-out) AND dormant\*, which yielded 270 publications

We scraped data from all papers of woody species that tested for photoperiod and/or temperature effects on budburst, leafout, or flowering, resulting in 56 papers. ? used a portion (72 experiments across 49 papers) of the earlier OSPREE database and provides extensive methods on the database creation and cleaning. For our analysis here, we included all budburst experiments where we could quantify chilling, forcing and photoperiod levels, resulting in 44 studies from 33 papers. Across experiments chilling treatments were often fully or partially applied in the field, thus we estimated field chilling ourselves using daily temperature data from ... [Cat and Nacho – add here: Be sure to include updated info on our datasets and which chilling metric we used]. ? provides additional details on these calculations (however, to have climate data through all our study years, we used a different climate dataset here for North America).

We analyze 2 different subsets of species in the OSPREE database to explore differences across two major groups of taxa, angiosperms and gymnosperms, for which there are markedly different number of species (194 angiosperms vs. 19 gymnosperms), and whose deep evolutionary divergence advises for separate analyses ().

We used the phylogenetic megatree for seed plant from Smith and Brown (2018) to extract a subset of the phylogenetic tree containing only the angiosperm species in the OSPREE dataset (Wolkovich et al., 2019). The species that were not present in the megatree were added as polytomies at the generic level (using the function *congeneric.merge*; (Pearse et al., 2015)) with a branch length of zero. Polytomies represent 26.8% of the full angiosperm dataset. To test for the ability of polytomies to bias our results we run sensitivity analyses excluding these species from models (which lead to 142 angiosperms; see Supporting Information).

## Bayesian hierarchical phylogenetic model

Commonly used phylogenetic regression methods today (e.g., PGLS and PMM) were originally conceived as statistical corrections for phylogenetic non-independence across observations—generally species—thus allowing multi-species studies to meet the assumptions linear regression (Freckleton et al., 2002). These corrections incorporated phylogenetic structure in the regression by modifying the residual variance-covariance matrix to substitute off-diagonal elements of zero (the value given the assumption of independence across observations) for shared phylogenetic branch lengths representing pairwise covariances (under phylogenetic non-independence among observations). Off-diagonals were also allowed to include a multiplying parameter—generally referred to as lambda—which is a transformation indicating the amount of phylogenetic relatedness among species (see below). Because the original aim of these methods was to correct for statistical nuance, the underlying assumption of phylogenetic regressions is that phylogenetic relatedness would only affect either model residuals (in PGLS approaches, Freckleton et al., 2002), or the model intercepts (e.g., in many PMM approaches, Housworth et al., 2004).

Because our aim is to understand how evolution may have imprinted biological responses to multiple interactive cues, our approach expands the above methods by explicitly incorporating phylogenetic structure accross model intercepts and slopes. Doing so allows explicitly estimating the amount of phylogenetic relatedness in species’ sensitivities to each cue, when these sensitivities are modelled in a multi-predictor regression setting.

For each  $j$  species, we assumed that data were generated from the following sampling distribution:

$$y_j \sim \mathcal{N}(\mu_j, \sigma_e^2) \quad (1)$$

where

$$\mu_j = \alpha_j + \beta_{1,j}X_2 + \beta_{2,j}X_2 + \beta_{3,j}X_3 \quad (2)$$

Predictors  $X_1$ ,  $X_2$ ,  $X_3$  are standardized forcing, chilling, and photoperiod, and their effects on the phenology of species  $j$  are determined by parameters  $\beta_{1,j}$ ,  $\beta_{2,j}$ ,  $\beta_{3,j}$ , representing species' responses (or sensitivities) to each of the cues. These responses, including the species-specific intercept  $\alpha_j$ , are elements of the following normal random vectors:

$$\begin{aligned}\boldsymbol{\alpha} &= \{\alpha_1, \dots, \alpha_n\}^T \text{ such that } \boldsymbol{\alpha} \sim \mathcal{N}(\mu_{\alpha}, \boldsymbol{\Sigma}_{\alpha}) \\ \boldsymbol{\beta}_1 &= \{\beta_{1,1}, \dots, \beta_{1,n}\}^T \text{ such that } \boldsymbol{\beta}_1 \sim \mathcal{N}(\mu_{\beta_1}, \boldsymbol{\Sigma}_{\beta_1}) \\ \boldsymbol{\beta}_2 &= \{\beta_{2,1}, \dots, \beta_{2,n}\}^T \text{ such that } \boldsymbol{\beta}_2 \sim \mathcal{N}(\mu_{\beta_2}, \boldsymbol{\Sigma}_{\beta_2}) \\ \boldsymbol{\beta}_3 &= \{\beta_{3,1}, \dots, \beta_{3,n}\}^T \text{ such that } \boldsymbol{\beta}_3 \sim \mathcal{N}(\mu_{\beta_3}, \boldsymbol{\Sigma}_{\beta_3})\end{aligned}\tag{3}$$

where the means of the multivariate normal distributions are root trait values (i.e., values of cue responses prior to evolving across a phylogenetic tree) and  $\boldsymbol{\Sigma}_i$  are  $n \times n$  phylogenetic variance-covariance matrices of the form:

$$\begin{bmatrix} \sigma_i^2 & \lambda_i \times \sigma_i \times \rho_{12} & \dots & \lambda_i \times \sigma_i \times \rho_{1n} \\ \lambda_i \times \sigma_i \times \rho_{21} & \sigma_i^2 & \dots & \lambda_i \times \sigma_i \times \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_i \times \sigma_i \times \rho_{n1} & \lambda_i \times \sigma_i \times \rho_{n2} & \dots & \sigma_i^2 \end{bmatrix}\tag{4}$$

where  $\sigma_i^2$  is the rate of evolution across a tree for trait  $i$  (here assumed to be constant along all branches),  $\lambda_i$  scales branch lengths and therefore is a measure of the “phylogenetic signal” or extent of phylogenetic relatedness on each model parameter (i.e.,  $\alpha_j$ ,  $\beta_{1,j}$ ,  $\beta_{2,j}$ ,  $\beta_{3,j}$ ), and  $\rho_{xy}$  is the phylogenetic correlation between species  $x$  and  $y$ , or the fraction of the tree shared by the two species.

The above specification is equivalent to writing equation 2 in terms of root trait values and residuals, such that:

$$\mu_j = \mu_{\alpha} + \mu_{\beta_1} X_1 + \mu_{\beta_2} X_2 + \mu_{\beta_3} X_3 + e_{\alpha_j} + e_{\beta_{1,j}} + e_{\beta_{2,j}} + e_{\beta_{3,j}}\tag{5}$$

where the residual error terms (e.g.,  $e_{\alpha_j}$ ) are elements of normal random vectors from multivariate normal distributions centered on 0 with the same phylogenetic variance-covariance matrices as in equation 4.

### Interpretation of $\lambda_j$ and $\sigma_j^2$ on slopes and intercepts

Most current phylogenetic regression approaches aimed at controlling for phylogenetic non-independence of analysis units (i.e. species, see Revell, 2010) assume the  $\lambda$  scaling parameter is constant across the full set of predictors in the model. Thus,  $\lambda$  is estimated as a single parameter based on one single residual term VCV matrix. While useful for correcting for phylogenetic

non-independence this approach does not allow the phylogeny to differentially affect different predictors (i.e. environmental cues in our example, which refer to simply as cues hereafter). In models with multiple cues, species responses to all cues are estimated as similarly phylogenetically structured, but this may not be the case. For example, in a PGLS model with three cues, it would be possible to have a high (i.e. close to 1) value of  $\lambda$ , due to either a strong phylogenetic signal in the response, but no phylogenetic structuring in the cues, or one or more predictors being strongly phylogenetically structured. In the latter case, phylogenetic structuring of responses to cues could be correlated (i.e., responses to cues evolving in a correlated fashion) or uncorrelated (i.e., independent evolution of responses to cues). Discerning these different situations is not trivial as they would inform whether responses to predictors configure in a structured fashion along the evolutionary process. However, most current approaches act as a black box regarding this information; they simply inform whether or not model residuals are phylogenetically structured (i.e. in PGLS) or the amount of model variance attributable to the phylogeny and independent from other sources of variation (i.e., in PMM, see Housworth et al. (2004)).

Because we are specifically interested in estimating the phylogenetic structure of each cues, our approach explicitly partitions variance into specific components relative to the model intercept and predictor (cue) slopes (see equation 5). The multivariate normal distributions of the intercept and slope terms include each a variance term (see equation 3), modelled with a  $\lambda$  scaling parameter. The interpretation of  $\lambda$ s in our models are analogous to Pagel’s Pagel (1999)  $\lambda$  parameter (Housworth et al., 2004), constrained to range from 0 to 1, with values of 0 indicating absence of phylogenetic relatedness, and values of 1 indicating *Brownian Motion* evolution (BM). Estimated  $\lambda$ s are not fully equivalent to computing phylogenetic signal of the slopes of each cue separately (i.e., fitting a multilevel regression model with species as a grouping factor on intercepts, and subsequently estimating phylogenetic signal for model slopes). Instead, they are a relative metric of phylogenetic relatedness allowing us to compare among responses known to interact with each other and estimated simultaneously. This approach has the further benefit of adjusting our partial pooling (‘random effect’ of species) based on evolutionary distance, more strongly pooling closely related species, and only weakly pooling distantly related species (see Gaussian process models in Gelman et al., 2014).

A traditional interpretation of  $\sigma^2$ s under Brownian Motion evolution, is an ‘evolutionary rate’ (Revell et al., 2008). In PGLS,  $\sigma_\epsilon^2$  is estimated for the model error term, which is distributed as a multivariate normal with VCV matrix given by  $\sigma_\epsilon^2 \mathbf{\Sigma}_i$ . Here, similar to our approach to  $\lambda$ , we estimate four  $\sigma^2$  values, corresponding to each model parameter. In our particular case (i.e., modelling a phenological response to three environmental cues),  $\sigma_\alpha^2$  for the intercept could be interpreted as the phenological variation across species accumulated along evolution independently from the cues. The  $\sigma_{\beta_1}^2$ ,  $\sigma_{\beta_2}^2$ , and  $\sigma_{\beta_3}^2$ , corresponding

to model slopes, would represent the phylogenetic variance linked to species responses to each of the modelled cues (i.e., forcing, chilling, and photoperiod, respectively). This is, the variability in how species shift their phenology responding to temperature and light, accumulated along the evolutionary process and considered in concert.

## References

- Ackerly, D. 2009. Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences of the United States of America* 106:19699–19706.
- Aitken, S. N., and J. B. Bemmels. 2016. Time to get moving: assisted gene flow of forest trees. *Evol Appl* 9:271–90.
- Azeez, A., and A. P. Sane. 2015. Photoperiodic growth control in perennial trees. *Plant signaling & behavior* 10:e1087631.
- Basler, D., and C. Körner. 2012. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology* 165:73–81.
- Bennett, J. M., J. Sunday, P. Calosi, F. Villalobos, B. Martínez, R. Molina-Venegas, M. B. Araújo, A. C. Algar, S. Clusella-Trullas, B. A. Hawkins, et al. 2021. The evolution of critical thermal limits of life on earth. *Nature communications* 12:1–9.
- Bonamour, S., L. M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B-Biological Sciences* 374.
- Bradshaw, W. E., and C. M. Holzapfel. 2006. Evolutionary response to rapid climate change. *Science* 312:1477–1478.
- CaraDonna, P. J., and D. W. Inouye. 2014. Phenological responses to climate change do not exhibit phylogenetic signal in a subalpine plant community. *Ecology* 96:355–361.
- Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics* 48:159–182.
- Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism in plant phenology. *Journal of Ecology* 101:1520–1530.
- Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology and biogeography* 28:275–285.

- Dietze, M. 2017. Ecological forecasting. *In* Ecological Forecasting. Princeton University Press.
- Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins, C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to community patterns. *Ecology letters* 15:545–553.
- Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha, and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses to warming. *Nature Climate Change* pages 1–6.
- Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich. 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist* 230:462–474.
- Franks, S. J., J. J. Weber, and S. N. Aitken. 2014. Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications* 7:123–139.
- Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* 160:712–726.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. 3rd ed. CRC Press, New York.
- Gotelli, N. J., and G. R. Graves. 1996. The temporal niche. Pages 95–111 *in* N. J. Gotelli and G. R. Graves, eds. *Null Models in Ecology*. Smithsonian Institution, Washington, DC.
- Grime, J. P. 1977. Evidence for existence of 3 primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist* 111:1169–1194.
- Hadfield, J. D. 2010. Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of statistical software* 33:1–22.
- Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The American Naturalist* 163:84–96.
- Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of budburst in temperate trees. *Journal of Applied Ecology* 29:597–604.
- IPCC. 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Ives, A. R., and M. R. Helmus. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. *Ecological Monographs* 81:511–525.
- Kochmer, J. P., and S. N. Handel. 1986. Constraints and competition in the evolution of flowering phenology. *Ecological monographs* 56:303–325.



- Kramer, K., A. Ducousso, D. Gomory, J. K. Hansen, L. Ionita, M. Liesebach, A. Lorent, S. Schueler, M. Sulkowska, S. de Vries, and G. von Wuehlisch. 2017. Chilling and forcing requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between provenances and are phenotypically plastic. *Agricultural and Forest Meteorology* 234:172–181.
- Larcher, W. 1980. *Plant Physiological Ecology*. Springer-Verlag.
- Laube, J., T. H. Sparks, N. Estrella, J. Höfler, D. P. Ankerst, and A. Menzel. 2014. Chilling outweighs photoperiod in preventing precocious spring development. *Global Change Biology* 20:170–182.
- Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins, M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, et al. 2022. The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution* .
- Molina-Venegas, R., J. C. Moreno-Saiz, I. Castro Parga, T. J. Davies, P. R. Peres-Neto, and M. Á. Rodríguez. 2018. Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. *Ecography* 41:1740–1749.
- Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005. Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological model. *Theoretical and Applied Genetics* 110:778–786.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421:37.
- Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker, and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. *Bioinformatics* 31:2888–2890.
- Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering time under climate change. *Journal of Ecology* 105:627–633.
- Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1:319–329.
- Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. *Systematic biology* 57:591–601.
- Rinne, P., A. Saarelainen, and O. Junttila. 1994. Growth cessation and bud dormancy in relation to ABA level in seedlings and coppice shoots of *Betula pubescens* as affected by a short photoperiod, water stress and chilling. *Physiologia Plantarum* 90:451–458.
- Sakai, A. K., and W. Larcher. 1987. *Frost Survival of Plants*. Ecological Studies. Springer-Verlag.

- Schaber, J., and F. Badeck. 2003. Physiology-based phenology models for forest tree species in germany. *International Journal of Biometeorology* 47:193–201. International Conference on the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- Serrano-Bueno, G., F. J. Romero-Campero, E. Lucas-Reina, J. M. Romero, and F. Valverde. 2017. Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology* 37:10–17.
- Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny. *American journal of botany* 105:302–314.
- Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of energetic scaling across the vertebrate tree of life. *The American Naturalist* 190:185–199.
- Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I. Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology letters* 13:1310–1324.
- Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 111:7906–7913.
- Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phylogenetic patterns of species loss in thoreau’s woods are driven by climate change. *Proceedings of the National Academy of Sciences* 105:17029–17033.
- Wolkovich, E. M., and E. E. Cleland. 2014. Phenological niches and the future of invaded ecosystems with climate change. *AoB Plants* 6:plu013.
- Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and communities in non-stationary environments. *Biological Reviews* .
- Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-PREE). doi:10.5063/F1CZ35KB.
- Yang, Z., Y. Du, M. Shen, N. Jiang, E. Liang, W. Zhu, Y. Wang, and W. Zhao. 2021. Phylogenetic conservatism in heat requirement of leaf-out phenology, rather than temperature sensitivity, in tibetan plateau. *Agricultural and Forest Meteorology* 304.
- Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate Change* 6:1120–1123.

## Tables and Figures

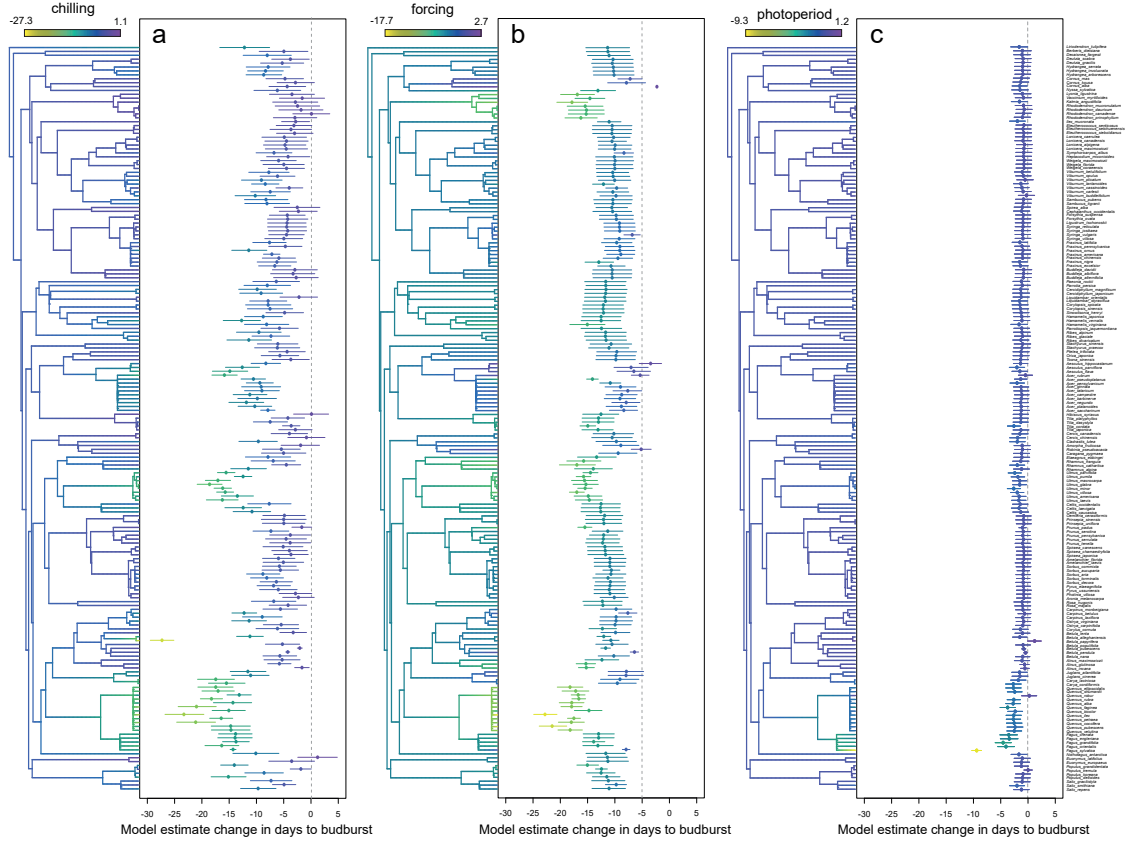


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured in change in days to budburst per standardized unit (z-transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored according to an estimation of ancestral character states, being the states at the tips the species' sensitivities to a cue, as estimated by our hierarchical phylogenetic model. Species sensitivities are shown along with 95% Credible Intervals in the diagrams. Note that the color scale varies in each panel. Total tree depth is 81. My.

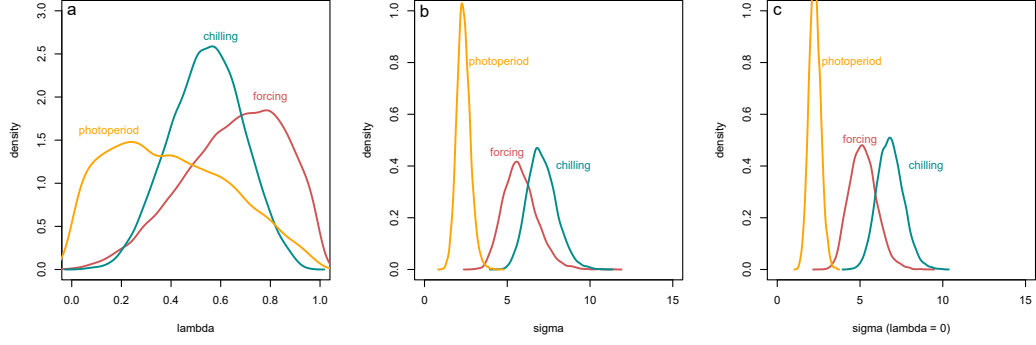


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters  $\lambda$  and  $\sigma$  estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to  $\lambda$  (a) and  $\sigma$  (b) from the phylogenetic model, and to  $\sigma$  (c) from the non-phylogenetic model (where  $\lambda$  is constrained to be equal zero).

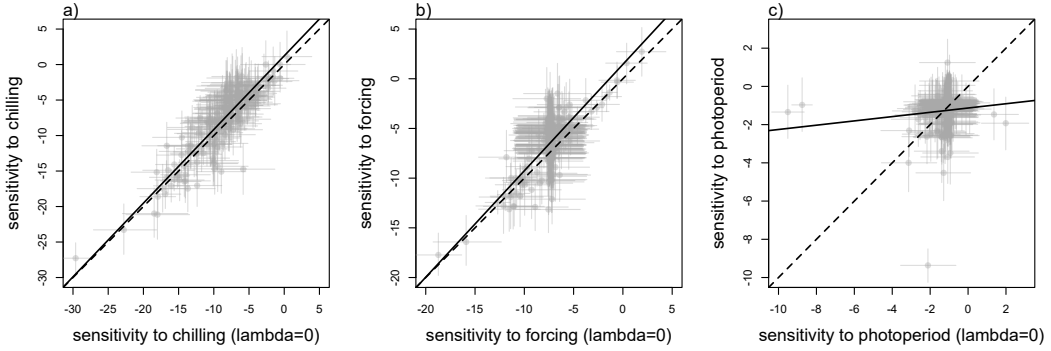


Figure 3: Correlations between model parameters as estimated by the full model and the models where  $\lambda$  is constrained to be equal zero. Panels correspond to sensitivity to chilling (a), to forcing (b), and to photoperiod (d).

X	mean	sd	X2.5.	X50.	X97.5.	n_eff	Rhat
a_z	30.63	3.41	23.94	30.66	37.26	12315.84	1.00
b_zf	-6.12	2.11	-10.24	-6.15	-1.85	3989.87	1.00
b_zc	-6.86	2.18	-10.98	-6.91	-2.39	7444.80	1.00
b_zp	-1.22	0.77	-2.73	-1.22	0.36	2482.96	1.00
lam_interceptsa	0.34	0.10	0.16	0.34	0.55	7668.82	1.00
lam_interceptsb	0.65	0.20	0.22	0.67	0.97	630.96	1.01
lam_interceptsc	0.54	0.15	0.25	0.55	0.82	1834.14	1.00
lam_interceptsbp	0.40	0.24	0.03	0.38	0.88	672.39	1.00
sigma_interceptsa	15.99	1.15	13.98	15.91	18.47	6970.37	1.00
sigma_interceptsb	5.80	1.01	4.06	5.70	8.01	1043.34	1.00
sigma_interceptsc	7.10	0.88	5.53	7.04	8.99	1767.13	1.00
sigma_interceptsbp	2.36	0.41	1.61	2.34	3.23	636.82	1.01
sigma_y	12.58	0.18	12.24	12.58	12.93	10904.90	1.00

X	mean	sd	X2.5.	X50.	X97.5.	n_eff	Rhat
a_z	31.79	1.28	29.29	31.77	34.35	13779.62	1.00
b_zf	-7.46	0.89	-9.19	-7.46	-5.71	2960.28	1.00
b_zc	-8.75	0.81	-10.29	-8.76	-7.11	6051.59	1.00
b_zp	-1.21	0.46	-2.10	-1.20	-0.29	2175.88	1.00
sigma_interceptsa	16.35	1.00	14.46	16.31	18.41	10178.43	1.00
sigma_interceptsb	5.20	0.82	3.76	5.15	6.93	677.74	1.00
sigma_interceptsb	6.84	0.78	5.40	6.80	8.46	1815.10	1.00
sigma_interceptsb	2.27	0.35	1.61	2.25	2.99	649.15	1.00
sigma_y	12.57	0.18	12.23	12.57	12.94	12887.31	1.00