Phylogenetic estimates of species-level phenology improve ecological forecasting

May 5, 2023

Authors:
Ignacio Morales Castilla,[1] T. J. Davies,[2,3] Geoff Legault,[3,4] Daniel M. Buonaiuto,[5,6,7] Catherine J. Chamberlain,[5,6,8] Ailene K. Ettinger,[9] Mira Garner,[3,10] Faith A. M. Jones,[3,11] Deirdre Loughnan,[3] William D. Pearse[12] & E. M. Wolkovich[3,5,6]

Author affiliations:
[1]Edificio Ciencias, Campus Universitario 28805 Alcalá de Henares, Madrid, Spain
[2]Botany, Faculty of Sciences, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada
[3]Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada
[4]Getty
[5]Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA
[6]Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA
[7]Department of Environmental Conservation, University of Massachusetts-Amherst, 160 Holdsworth Way, Amherst, MA, USA
[8]The Nature Conservancy, 334 Blackwell St Ste 300, Durham, NC, USA
[9]The Nature Conservancy of Washington, 74 Wall Street, Seattle, WA USA
[10]Morton Arboretum, USA
[11]Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, 901 83 Umea, Sweden
[12] Imperial College, UK

*Corresponding author: ignacio.moralesc@uah.es

Abstract

Adaptation to ongoing climate change hinges 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, using novel Bayesian hierarchical phylogenetic models, we overcome this challenge and estimate species-level responses to two major environmental cues of spring phenology: temperature and daylength. We find that variation is greater across species than across cues, suggesting the current focus on identifying which cue is the strongest may be less important than predicting how each species responds to a combination of cues. Further, our model provides insights on how evolutionary history has shaped responses to cues, suggesting stronger evolutionary dynamics in responses to temperature—especially cool winter temperatures—than daylength. Our approach provides a major advance in ecological forecasting, with implications for predicting the impacts of climate change and other anthropogenic forces on species interactons and ecosystem functioning.

Introduction

The biological impacts of climate change will have major implications for ecosystem functioning and stability. With rising global temperatures many species have shifted their geographic distributions northward in space and cyclic life-history events—their phenology—earlier in time (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 (Chuine and Regniere, 2017; Ettinger et al., 2020). In the absence of detailed data on individual species, species groupings (e.g., functional groups) have improved ecosystem models (Moorcroft et al., 2001; Griffith et al., 2020), but still capture only a fraction of the important variability (REFS). 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 that data are 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 (Gelman and Hill, 2006), and they thus partially pool ('shrink') towards estimates for species with the most data (and least variable responses), 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 suboptimal under new conditions. More reliable estimates of species-level responses would allow us to better incorporate species differences into models of ecosystem change.

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 (EXAMPLES). In addition, evidence for phylogenetic conservatism appears to depend on method and species, even varying between sites with overlapping species sets (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). A first challenge is thus how to better integrate evolutionary history into multi-species models of plant phenological responses.

Generating robust ecological forecasts requires 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, linking a given phenotypic response to a single cue, for example, time of leafout responding to summed heat during early spring (Davies et al., 2013). These efforts fail to capture the more likely scenario for most phenotypic traits in which multiple cues interacting along evolutionary history have shaped species' responses (Ackerly, 2009). For many plant species, phenolological events are determined by a combination of temperature and light (Chuine and Regniere, 2017). It is likely that additional factors (e.g., other cues or species' physiology) further mediate species responses. Although these mediating factors 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 moves towards integrating evolutionary history in models of phenological responses to environmental change. To understand how evolution has shaped the cues underlying shifting phenology with climate change (Uyeda et al., 2017), we explicitly incorporate phylogenetic structure across model intercepts and slopes (that is, allowing a separate model of evolutionary history for chilling, forcing and photoperiod, see full Methods & Materials for a complete description).

We illustrate our method with an unprecedented dataset on phenological responses to environmental cues (chilling, forcing and photoperiod) determined experimentally for 191 deciduous woody species (by far the most studied group of species in phenology experiments, see Ettinger et al., 2020), in an updated version of the Observed Spring Phenology Responses in Experimental Environments (OSPREE) database (Wolkovich et al., 2019). These data combined with the Smith and Brown (2018) megatree adjusted to our species and our modeling approach allows us to address the common question of which cue has the largest effect on budburst and at the same time provide robust estimates of how cues vary across species. Using spring phenology, we identify historical regime shifts (Uyeda et al., 2017) in phenological responses, and highlight how our approach could advance forecasting of other critical responses to ongoing global change.

Results & Discussion

Most species respond to all three primary cues—forcing, chilling, and photoperiod (Figs. 1, Supporting Table **??**)(see also Laube et al., 2014; Ettinger et al., 2020)—with responses to chilling approximately five-fold greater than to photoperiod (phenological advances of 6.8 days per standardized unit vs 1.2 days, for chilling and photoperiod, respectively; see Table **??**). However, these average estimates 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 the mean 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 current focus on comparing between the three cues may miss the largest variation, and highlights why 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) and 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 stan-

dardized 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 been selected for alongside responses to another cue—e.g. chilling. This could occur if selection operates on responses to both cues for the same reason; for example, if sensitivity to multiple cues provides greater insurance against leafing out before the last frost (Bonamour et al., 2019; Wolkovich and Donahue, 2021). Additionally, linkage or pleiotropism among loci associated with different cues (Nakagawa et al., 2005) could induce these correlations. However, the correlation in species responses across cues was generally weak ($r = 0.31$; between forcing an 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 temperature cues (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 particularly well studied, especially *Fagus sylvatica* (e.g., Basler and Körner, 2012; Zohner et al., 2016; Kramer et al., 2017). 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. 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 (i.e., if local adaptation were 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 **??**).

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 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 traits, 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 indexed 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. However, 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 σ for photoperiod responses (Fig. 2). Here, as in more traditional phylogenetic comparative methods, σ represents the rate of evolution, and thus our results suggest photoperiod responses are also evolving slower than temperature responses (see Fig. **??**).

Phylogenetic conservatism (high λ) and slow evolutionary rates (low σ) 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 where responses are dominated by forcing cues might be more vulnerable to future warming because phylogenetic conservatism (λ) in forcing is higher compared to other cues and its evolutionary rate (σ) is lower than that estimated for chilling. This is misleading, however, as estimates of λ 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 (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 more widely used hierarchical models. This was especially noticeable in temperature responses (for chilling variance across species means, $\beta_{chill,j}$ in eqn. 2, was estimated as 23.45 in the phylogenetic model, versus 17.47 in the non-phylogenetic model; variance across means of $\beta_{force,j}$: 8.74 compared to 5.01) while photoperiod, which had low phylogenetic structure, was more similar across approaches (variance across means of $\beta_{photo,j}$: 0.82 compared to 0.93).

The consequences of including shared evolutionary history in forecasting are most apparent for poorly sampled species nested within more well-sampled clades. For example, forecasts for *Acer campestre*, which has only 6 observations, shift by up to 35% in our phylogenetically informed model compared to one without phylogeny (Fig. 4). In contrast, forecasts for *Betula pendula*, which is one of the most sampled species, are nearly identical across models (Fig. 4). This occurs because cue estimates for *Acer campestre* in the phylogenetically informed model are strongly influenced by other *Acer* species, which diverge from other clades. In the non-phylogentically informed model all species are equally exchangeable and thus *Acer campestre* is pulled strongly towards well-sampled species, such as *Betula pendula*, leading to forecasted shifts that are more similar across all species—especially poorly sampled species (Fig. 4).

The increase in variability across species in our model with phylogenetic structure also decreased the uncertainty in estimates for each individual species temperature responses (Fig. **??**). Thus, traditional (non-phylogentically informed) 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. Notably, photoperiod cues appear weaker, more uniform across species, and less phylogenetically constrained than temperature cues. 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., Betulaceae).

While we focused on spring phenology here, our approach suggests a path forward for more

general forecasting of species-level climate change responses. Our results show how including the phylogenetic relationship of species in a mechanistic model of underlying cues 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. Using this approach improved forecasts of biological responses to climate change could help anticipate impacts on critical ecosystem services from species-level shifts and thus aid mitigation and human adaption to warming.

Methods & Materials

Phenological and Phylogenetic Data

Phenological data: To estimate phenological responses to chilling, forcing and photoperiod we used data from phenological experiments in controlled environments of temperate woody species, 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 dorman*, 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. Ettinger et al. (2020) 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. We focused on angiosperms as gymnosperms are very poorly represented in spring phenology experiments.

Across experiments chilling treatments were often fully or partially applied in the field, thus we estimated field chilling ourselves in Utah units with the `chillR` package combined with daily temperature data (converted to hourly) from both European (E-OBS, version 16, calculating the average of minimum and maximum daily temps, Cornes et al., 2018) and North American (v3, Sheffield et al., 2006) gridded climate datasets. Ettinger et al. (2020) provides additional details on these calculations (however, to have climate data through all study years given the updated OSPREE database, we used a different climate dataset here for North America).

For a phylogenetic tree, we pruned the phylogenetic megatree for seed plants (Smith and Brown, 2018) to extract a subset of the phylogenetic tree containing only the angiosperm species in the OSPREE dataset, then added species that were not present in the megatree as polytomies at the generic level (using the function 'congeneric.merge' in 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 of 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-independece 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 across 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 observation i of species j, we assumed that the timing of phenological events were generated from the following sampling distribution:

$$y_{i,j} \sim \mathcal{N}(\mu_j, \sigma_e^2) \tag{1}$$

where

$$\mu_j = \alpha_j + \beta_{chill,j} X_{chill} + \beta_{force,j} X_{force} + \beta_{photo,j} X_{photo} \tag{2}$$

where σ_e^2 represents random error unrelated to the phylogeny.

Predictors X_{chill}, X_{force}, X_{photo} are standardized chilling, forcing, and photoperiod, and their effects on the phenology of species j are determined by parameters $\beta_{chill,j}$, $\beta_{force,j}$, $\beta_{photo,j}$, representing species' responses (or sensitivities) to each of the cues. These responses, including the species-specific intercept α_j, are elements of the following normal random vectors:

$$\begin{aligned}
\boldsymbol{\alpha} &= [\alpha_1, \ldots, \alpha_n]^T \text{ such that } \boldsymbol{\alpha} \sim \mathcal{N}(\mu_\alpha, \boldsymbol{\Sigma_\alpha}) \\
\boldsymbol{\beta_{chill}} &= [\beta_{1,1}, \ldots, \beta_{1,n}]^T \text{ such that } \boldsymbol{\beta_{chill}} \sim \mathcal{N}(\mu_{\beta_1}, \boldsymbol{\Sigma_{\beta_{chill}}}) \\
\boldsymbol{\beta_{force}} &= [\beta_{2,1}, \ldots, \beta_{2,n}]^T \text{ such that } \boldsymbol{\beta_{force}} \sim \mathcal{N}(\mu_{\beta_2}, \boldsymbol{\Sigma_{\beta_{force}}}) \\
\boldsymbol{\beta_{photo}} &= [\beta_{3,1}, \ldots, \beta_{3,n}]^T \text{ such that } \boldsymbol{\beta_{photo}} \sim \mathcal{N}(\mu_{\beta_3}, \boldsymbol{\Sigma_{\beta_{photo}}})
\end{aligned} \tag{3}$$

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 σ_i^2 is the rate of evolution across a tree for trait i (here assumed to be constant along all branches), λ_i scales branch lengths and therefore is a measure of the "phylogenetic signal" or extent of phylogenetic relatedness on each model parameter (i.e., α_j, $\beta_{force,j}$, $\beta_{force,j}$, $\beta_{photo,j}$), and ρ_{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:

$$\mu_j = \mu_\alpha + \mu_{\beta_{chill}} X_{chill} + \mu_{\beta_{force}} X_{force} + \mu_{\beta_{photo}} X_{photo} + e_{\alpha_j} + e_{\beta_{force,j}} + e_{\beta_{chill,j}} + e_{\beta_{photo,j}} \tag{5}$$

where the residual phylogenetic error terms (e.g., e_{α_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. Model code, including priors used here, are given in the supplement.

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.

Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble version of the E-OBS temperature and precipitation data sets. Journal of Geophysical Research: Atmospheres 123:9391–9409.

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., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

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.

Griffith, D. M., C. P. Osborne, E. J. Edwards, S. Bachle, D. J. Beerling, W. J. Bond, T. J. Gallaher, B. R. Helliker, C. E. R. Lehmann, L. Leatherman, J. B. Nippert, S. Pau, F. Qiu, W. J. Riley, M. D. Smith, C. A. E. Stromberg, L. Taylor, M. Ungerer, and C. J. Still. 2020. Lineage-based functional types: characterising functional diversity to enhance the representation of ecological behaviour in land surface models. New Phytologist 228:15–23.

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.

Moorcroft, P., G. Hurtt, and S. Pacala. 2001. A method for scaling vegetation dynamics: The ecosystem demography model (ed). Ecological Monographs 71:557–585.

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.

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.

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.

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.

Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. Journal of Climate 19:3088–3111.

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 (OSPREE). 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

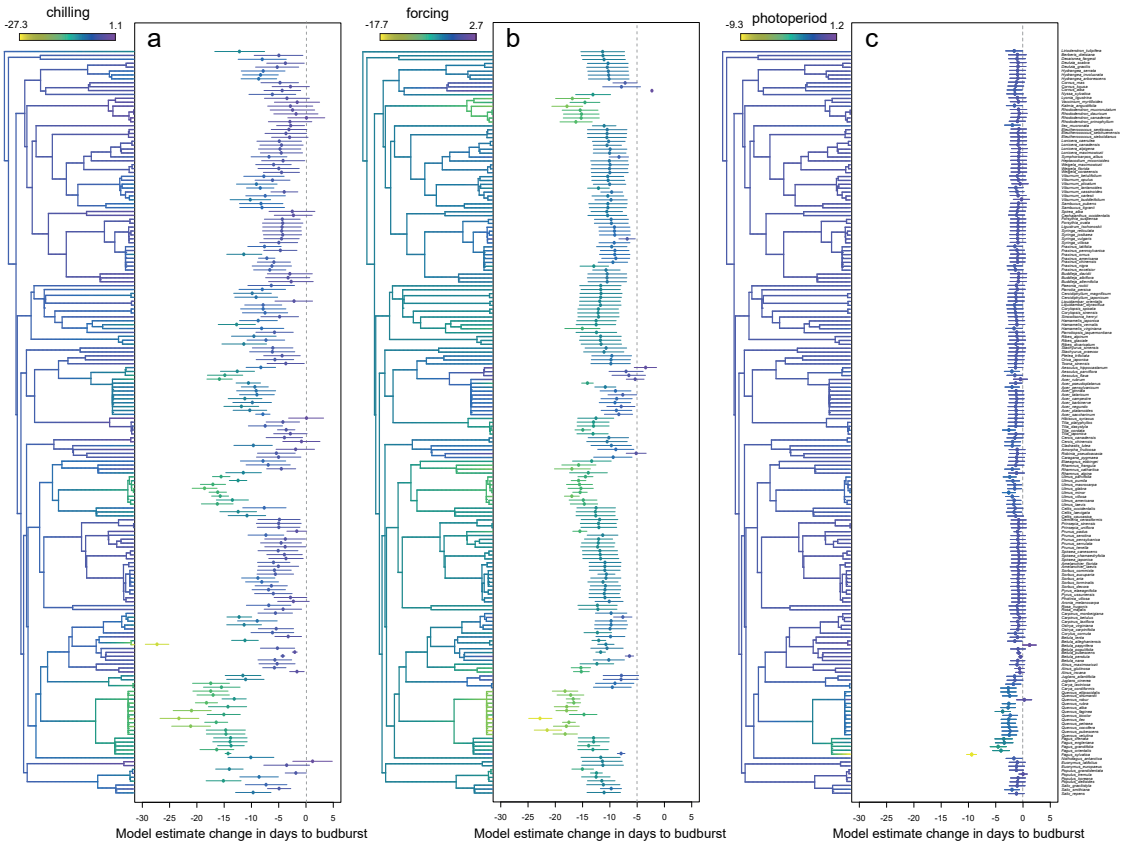

Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured in change in days to budburst per standardized unit (z-transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored acording to an 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 50% Credible Intervals in the diagrams. Note that the color scale varies in each panel. Total tree depth is 81. My.

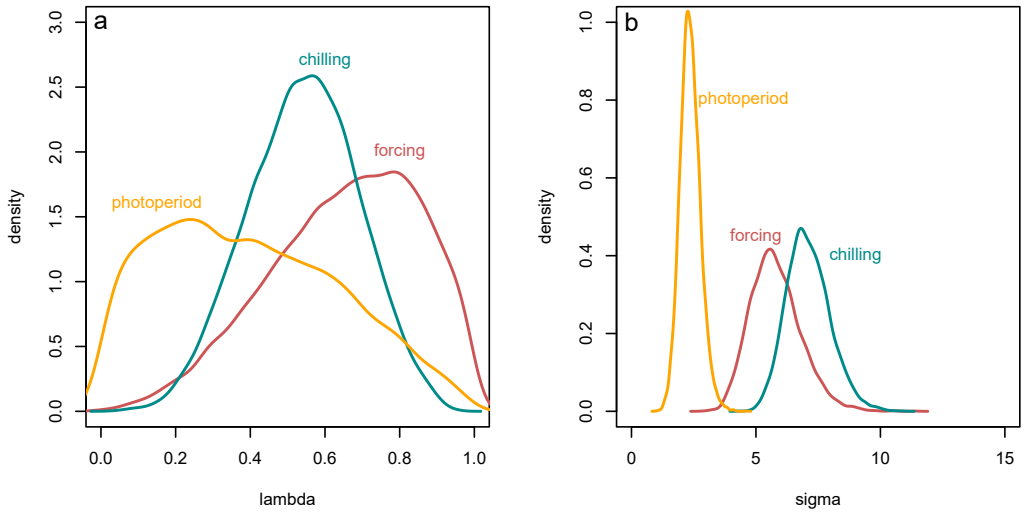

Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters λ and σ estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to λ (a) and σ (b) from the phylogenetic model, and to σ (c) from the non-phylogenetic model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

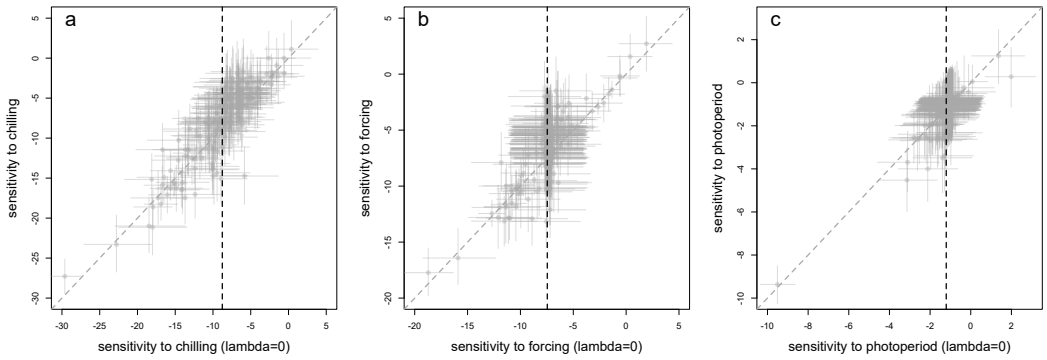

Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue (y-axis), and the more commonly used hierarchical model where species are exchangeable (where λ is constrained to be equal zero, x-axis). While species with large amounts of data may be estimated similarly by both models, in the more commonly used hierarchical model (x-axis) many species are pulled towards the overall average (shown by dashed black lines). The strength of pulling is particularly patent for forcing (b). Panels correspond to sensitivity to chilling (a), to forcing (b), and to photoperiod (c). Dashed grey lines show the one to one relationship (with intercept equal zero and slope equal one).

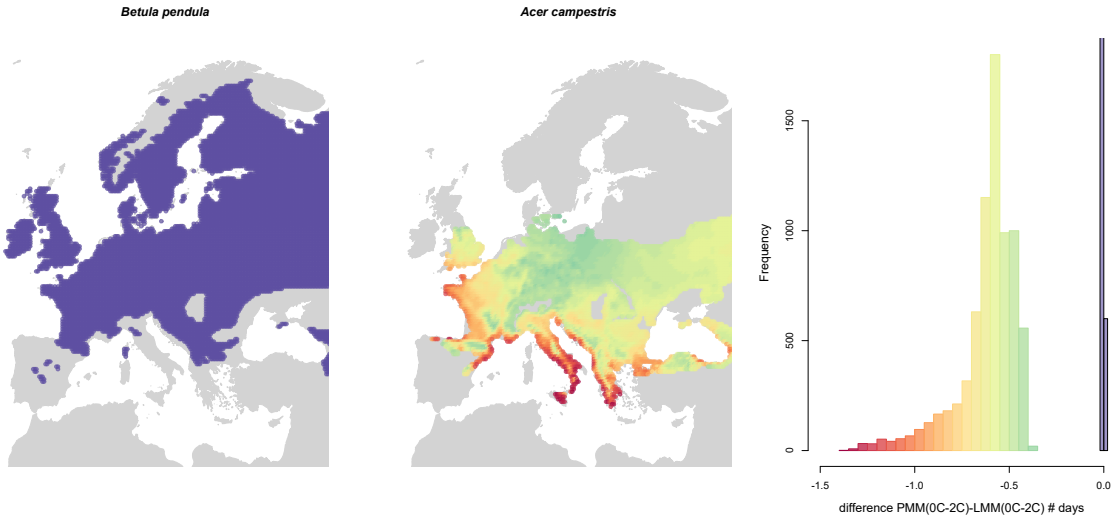

Figure 4: Forecasted shifts can be substantially different for poorly sampled species in well-sampled clades in our phylogenetic mixed model compared to traditional hierarchical approaches. Here we show ...

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_interceptsbf	0.65	0.20	0.22	0.67	0.97	630.96	1.01
lam_interceptsbc	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_interceptsbf	5.80	1.01	4.06	5.70	8.01	1043.34	1.00
sigma_interceptsbc	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_interceptsbf	5.20	0.82	3.76	5.15	6.93	677.74	1.00
sigma_interceptsbc	6.84	0.78	5.40	6.80	8.46	1815.10	1.00
sigma_interceptsbp	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