

Phylogenetic estimates of species-level phenology improve ecological forecasting

immediate

June 15, 2024

Author list

Ignacio Morales-Castilla,^{1*} T. J. Davies,^{2,3} Geoffrey Legault,³ D. M. Buonaiuto,^{4,5,6} Catherine J. Chamberlain,^{4,5,7} Ailene K. Ettinger,^{5,8} Mira Garner,³ Faith A. M. Jones,^{3,10} Deirdre Loughnan,³ William D. Pearse,¹¹ Darwin S. Sodhi³ & E. M. Wolkovich^{3,4,5}

Affiliations

¹GloCEE - Global Change Ecology and Evolution Group, Department of Life Sciences, University of Alcalá, Alcalá de Henares, Spain

²Botany, Faculty of Sciences, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

³Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

⁴Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA

⁵Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA

⁶Department of Environmental Conservation, University of Massachusetts-Amherst, 160 Holdsworth Way, Amherst, MA, USA

⁷The Nature Conservancy, 334 Blackwell St Ste 300, Durham, NC, USA

⁸The Nature Conservancy of Washington, 74 Wall Street, Seattle, WA USA

¹⁰Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, 901 83 Umea, Sweden

¹¹Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, SL5 7PY, UK

*Corresponding author: Ignacio Morales-Castilla; e-mail: ignacio.moralesc@uah.es

Abstract

The ability to adapt to climate change requires accurate ecological forecasting. Current forecasts, however, have failed to capture important variability in biological responses, especially across species. Here, we present a novel method using Bayesian hierarchical phylogenetic models to overcome this challenge and estimate species-level responses. We apply our method to phenological experiments manipulating temperature and daylength and show species differences are larger than the average differences between cues. We illustrate how a focus on certain clades can bias prediction, but that predictions may be improved by integrating information on phylogeny. Our model demonstrates an underlying phylogenetic structure in plant phenological responses to temperature cues, whereas responses to photoperiod appear weaker, more uniform across species, and less phylogenetically constrained. Our approach provides an advance in ecological forecasting, with implications for predicting the impacts of climate change and other anthropogenic forces on ecosystems.

Main text

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 poleward in space and recurring life-history events—their phenology—earlier in time (1; 2), 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 (3).

While ecological forecasting has improved over recent years (4; 5), it remains a challenge to reproduce the high variability observed in biological responses such as phenology, physiology or demography to environmental cues (1). 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, could be driven by species-specific variation, reflecting evolved differences in species sensitivities to underlying environmental cues and their interactions. Unfortunately, we can only estimate the sensitivities to cues for a few well-studied species (6; 7). In the absence of detailed data on individual species, species groupings (e.g., functional groups) have improved ecosystem models (8; 9), but still capture only a fraction of the important variability (10).

Recent efforts that have attempted to model species-specific responses to the environment (11) are often restricted by data availability—especially the common problem that data are often prevalent for 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 (all species represent samples drawn from the same underlying distribution, 12). As such, they partially pool (‘shrink’) towards estimates for species with the most data and least variable responses, making inference at the species-level unreliable (7). More reliable estimates of species-level responses would allow us to better incorporate species differences into models of ecosystem change.

Including the evolutionary history of species relationships in models of species responses could provide more robust species-level estimates than current approaches and a better understanding of the evolutionary constraints that might limit adaptation to change. For example, strong phylogenetic niche conservatism (13) could potentially inhibit adaptive responses by drawing species back to an evolutionary conserved optimum, which is sub-optimal under new conditions. While incorporating such evolutionary history is traditionally seen as necessary, either as a statistical correction or to better understand species evolutionary history, the use of such phylogenetic information should also improve model fitting and forecasts (14).

Research using long-term observational data has 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, including dates of budburst, leafout and first flowering (15; 16; 17), suggests that more closely related species share more similar phenologies, likely reflecting evolutionary conservatism in responses to common cues. There are two broad explanations for why we might expect phylogenetic conservatism in phenological traits. First, close relatives will tend to share similar ecologies and physiologies, and thus be sensitive to similar environmental pressures. Second, close relatives derive from common geographic centers of origin, and thus their ancestors will have been exposed to—and have adapted to—similar environmental cues (17).

However, approaches using traditional phylogenetic comparative methods, have produced conflicting results, with some studies reporting evidence of phylogenetic structure in phenology-linked species declines (e.g., 16) and in some phenophases, but not others (e.g., 18), and in responses to some cues, but not others (e.g., 19). In

addition, evidence for phylogenetic conservatism of phenological responses appears to depend on method and species, even varying between sites with overlapping species sets (e.g., 20), 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). Thus, a first challenge is 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 understanding of how species use environmental cues to time their phenotypic responses with the temporal distribution of key resources while avoiding periods of high stress (21; 22). 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 (17). 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 (23). For many plant species, phenological events are determined by a combination of temperature and light (6), with additional factors (e.g., other cues—like humidity, or species physiology—vasculature or leaf structure) likely further mediating species responses. Although these mediating factors are not well understood (6), they can be accounted for in models either as latent processes or by allowing non-stationarity in responses across species (24).

Spring plant phenology may represent the 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 (6), 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 (25). 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 (7).

Here we present a Bayesian framework that extends upon phylogenetic mixed models (26) to examine how chilling, forcing (both metrics of temperature) and photoperiod together determine spring plant phenology. By allowing non-stationarity in species responses across the phylogeny (24), our model departs from previous work and assumptions of traditional phylogenetic comparative methods concerned with phylogenetic correction (e.g., 14), 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 (27), 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 Methods 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 7), in an updated version of the Observed Spring Phenology Responses in Experimental Environments (OSPREE) database (28). These data combined with the (author?) (29) megatree (i.e., hypothesized phylogenetic relationships for a large number of species and built from multiple smaller clade-level trees) adjusted to our species, and 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 (27) in phenological responses, and highlight how our approach could advance forecasting of other critical responses to ongoing global change.

Most species respond to all three primary cues—forcing (warm temperature experimental treatments generally starting in late winter), chilling (cool temperature treatments starting in autumn), and photoperiod (Fig. 1, Supplementary Table S2)—with responses to chilling approximately five-fold greater than to photoperiod (phenological advances of 6.9 days per standardized unit vs 1.2 days, for chilling and photoperiod, respectively; see Supplementary Table S2). We estimated lower average responses to temperature compared to a model without phylogeny (model slopes for forcing and chilling decreased by 18% and 22%, respectively; see Extended Data Fig. Extended Data 1); responses to chilling and forcing were also more similar when including phylogeny (though chilling was still greater: 6.9 vs. 6.1 per standard unit), which contrasts with previous results suggesting chilling responses are much greater than forcing (30; 7).

These average estimates, however, fail to capture the large differences in species responses to both chilling and forcing (Fig. 1, Supplementary Table S6). 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 cue in magnitude—chilling—varied 24-fold between species, while variation to forcing varied 7-fold. This variation indicates large differences between chilling and forcing occur due to differences found at the species-level rather than due to differences across species (e.g., the average effect across species as previously suggested, 30; 7). These results highlight 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. For example, several groups—oaks and beeches (Fagaceae), elms (Ulmaceae) and buckthorns (Rhamnaceae)—are highly sensitive to chilling while others—rhododendrons (Ericaceae), butterfly bushes (Scrophulariaceae) and spindles (Celastraceae)—show little to no response to chilling (Fig. 1a). Similar clade-level variation was observed for forcing, where some of these clades—e.g., Ericaceae, Rhamnaceae, Ulmaceae, or Fagaceae—were particularly sensitive and others, such as the Sapindaceae, Cornaceae or Juglandaceae, show little response (Fig. 1b).

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. This could occur if selection operates jointly on responses to both cues; for example, if sensitivity to multiple cues provides greater insurance against leafing out before the last frost (22; 31). Additionally, linkage or pleiotropism among loci associated with different cues (32) could induce across-cue correlations. However, the correlation in species responses across cues was generally weak ($r = 0.31$; between forcing and chilling; Supplementary Fig. S4) 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). whyuse1Species sensitivity to one cue, thus, does not constrain sensitivity to another cue, and it seems selection can operate independently on responses to different cues (22).whyuse1E

In contrast to temperature cues (chilling and forcing), species-level responses to photoperiod were almost uniform across species. whyuse2This consistency provides insight on a large debate over the prevalence of photoperiod cues in temperate trees, where previous experiments (33; 34) and models (e.g., 35; 36) suggested important variability across species that may constrain the responses of certain species to warming (37). Our results indicate variability is limited to a handful of species in Fagaceae, which have been particularly well studied, especially *Fagus sylvatica* (e.g., 33; 34; 38). 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., 38), in contrast with common garden studies of other species, which find little evidence of local adaptation in spring (but not fall) phenology (39). The uniformity of response to photoperiod in our results 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).

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 differed 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 ($\lambda = 0.4$; see Fig. 2, Supplementary Table S2).

Differences among 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 (24). This fits fundamentally with the idea that early-season phenology plays a critical role in shaping species temporal niches (40) 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., 41; 31).

Weak phylogenetic signal in photoperiod sensitivity (Fig. 2) might seem at odds with the uniformity of species response—i.e., there is very little variation in the responses to photoperiod across species. However, somewhat counterintuitively, both uniform and random responses can manifest as low phylogenetic signal when indexed by Brownian motion expectations (13). 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 (42), and subsequent selection on photoperiod sensitivity was constrained by stabilizing selection operating on other life-history attributes sensitive to photoperiod (e.g., 43; 44; 45), 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 Supplementary Fig. S6).

Phylogenetic conservatism (high λ) and slow evolutionary rates (low σ) in traits has sometimes been interpreted as indicative of evolutionary constraints to adaptive change (13; 46). 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. 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, 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, $var(\beta_{chill,j})$ from eqn. 2), was estimated as 23.55 in the phylogenetic model, versus 17.47 in the non-phylogenetic model; variance across means, $var(\beta_{force,j})$: 8.75 compared to 5.01) while photoperiod, which had low phylogenetic structure, was more similar across approaches (variance across means, $var(\beta_{photo,j})$: 0.83 compared to 0.64).

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 the number of days until budburst after forcing starts, when comparing our phylogenetically informed model to one without phylogeny (see Fig. 4, Extended Data Fig. Extended Data 2). In contrast, forecasts for *Betula pendula*, which is one of the most sampled species, are nearly identical across models (Fig. 4, Extended Data Fig. Extended Data 2). 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-phylogenetically informed model all species are equally exchangeable and thus *Acer campestre* is pulled strongly towards well-sampled species, such as *Betula pendula* ($n = 311$), leading to forecasted shifts that are more similar across all 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 (Extended Data 3). Thus, traditional (non-phylogenetically informed) approaches that partially pool across species (most hierarchical models in ecology, e.g., 47; 7) may also lead to less precise predictions and forecasts of phenology for individual species, although overall model accuracy might still appear reasonable (see Extended Data Fig. Extended Data 4). Another advantage of our Bayesian approach is that we are also better able to accommodate imprecision in the data that informs our model, which might arise from multiple sources, including measurement or experimental error, and the general stochasticity associated with limited sample sizes and unbalanced species representation. Critically, by partially pooling across species and weighting by phylogeny, we gain strength from species estimates that are informed by more data, such as within *Betula* and *Fagaceae*, but avoid skewing estimates for phylogenetically distant clade that may have been exposed to different selective regimes. We found species estimates were robust through cross-validation, the phylogenetic model better predicted observed values for held-out data, and yielded more stable species coefficients compared to a hierarchical model (Extended Data Figs. Extended Data 5-Extended Data 6; see “Leave-One-Clade-Out model cross validation” section in Supplementary Information).

The contrasts between temperature and photoperiod responses—in both their variability across species and phylogenetic structure—have important implications for generating multi-species forecasts. Notably, responses to photoperiod appear weaker, more uniform across species, and less phylogenetically constrained compared to temperature. For temperature responses, the large variability among species makes predicting species-level responses challenging, but the phylogenetic structure in responses lets us borrow information from close relatives to improve our predictions. However, given that Brownian motion (our assumed model of evolution) is an extremely noisy process, we recommend imputation only for missing taxa that are closely related to other well sampled species or clades (48; 49).

While we focused on spring phenology here, our new 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—correcting biased model estimates (Extended Data Fig. Extended Data 7), estimating the full variability across species and reducing uncertainty around individual species estimates—while at once providing insight into the evolutionary history of biological responses. Using this

approach improved forecasts of phenological responses to climate change and could help anticipate impacts on critical ecosystem services from species-level shifts and thus aid mitigation and human adaption to warming.

Acknowledgements

We thank the many researchers who conducted the experiments and built the phylogeny used in this manuscript. I.M.-C. acknowledges funding from the Spanish Ministry for Science and Innovation (grant no. PID2019/109711RJ-I00 to I.M.-C.) well as from Comunidad de Madrid and University of Alcalá (grant no. CM/BG/2021-003 to I.M.-C.). E.M.W. acknowledges funding by The Canada Research Chair in Temporal Ecology (E.M.W.). Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Author Contributions Statement

I.M.-C., D.M.B., C.J.C., A.K.E. and E.M.W. conceived the manuscript. All authors worked to clean the database and conducted literature review. I.M.-C., T.J.D., E.M.W., G.L., D.L. and W.D.P. contributed data analysis and/or code. I.M.-C., D.M.B. and C.J.C. created the figures. I.M.-C., T.J.D. and E.M.W. wrote an initial draft of the manuscript. All authors reviewed and revised the manuscript.

Competing Interests Statement

The authors declare no competing interests.

Figure Legends/Captions

Figure 1: **Phenological sensitivity to three environmental cues across 191 woody species estimated by a Phylogenetic Mixed Model.** The environmental cues are chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit (z -transformation) of the cues. The database used to fit the Phylogenetic Mixed Model comprised 44 studies, 191 species and 2940 observations. 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 as mean values \pm 50% uncertainty intervals. Note that the color scale varies in each panel. Total tree depth is 81 My.

Figure 2: **Posterior distributions of phylogenetic parameters.** The lines show density plots comparing the lambda (λ) and sigma (σ) parameters 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. The λ parameter measures phylogenetic signal in each estimate of cue sensitivity (or how close species are similarly sensitive to a given cue, in a scale ranging from 0 to 1). The σ parameter indicates evolutionary rate or accumulation of phenotypic variation with evolutionary time.

Figure 3: **Correlations between parameters estimated by the phylogentic and non-phylogenetic model.** The phylogenetic model (phylogenetic mixed model, PMM; y -axis) accounts for phylogenetic structure on each phenological cue, and the more commonly used hierarchical mixed model (HMM; x -axis) where species are exchangeable (where λ is constrained to be equal to 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 grey vertical lines). The strength and prevalence of pulling across species is particularly obvious for forcing (b). Panels correspond to sensitivity to chilling (a), forcing (b), and photoperiod (c). Dashed grey 1:1 lines also shown. Species estimates are shown as mean values \pm 50% uncertainty intervals with estimate colors are in the same scale as in 2. Note that uncertainties of each species are higher in HMM than in PMM (Extended Data Fig. Extended Data 2). The database used to fit both models comprised 44 studies, 191 species and 2940 observations.

Figure 4: **Comparison of forecasts of phenological shifts resulting from a phylogenetic (PMM) and a non-phylogenetic but hierarchical (HMM) approach.** Phenological shifts were computed as the difference between predictions under current climate vs. a 2°C warmer climate). Differences in forecasted shifts are negligible for well sampled species (*Betula pendula*, $n = 311$, a), but can be substantially different for poorly sampled species in well-sampled clades (*Acer campestre*, $n = 6$, b). The maps show the difference in number of days between the shifts predicted by PMM and HMM, with values colored according to histograms in panel c (days here are relative to start of forcing conditions, not calendar days). See Supplementary Information for details on forecast calculation.

References

References

- [1] IPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014).
- [2] Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37 (2003).
- [3] Richardson, A. D. *et al.* Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology* **169**, 156–173 (2013).
- [4] Dietze, M. Ecological forecasting. In *Ecological Forecasting* (Princeton University Press, 2017).
- [5] Lewis, A. S. *et al.* The power of forecasts to advance ecological theory. *Methods in Ecology and Evolution* (2022).
- [6] Chuine, I. & Regniere, J. Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics* **48**, 159–182 (2017).
- [7] Ettinger, A. *et al.* Winter temperatures predominate in spring phenological responses to warming. *Nature Climate Change* 1–6 (2020).
- [8] Moorcroft, P., Hurtt, G. & Pacala, S. A method for scaling vegetation dynamics: The ecosystem demography model (ed). *Ecological Monographs* **71**, 557–585 (2001).
- [9] Griffith, D. M. *et al.* Lineage-based functional types: characterising functional diversity to enhance the representation of ecological behaviour in land surface models. *New Phytologist* **228**, 15–23 (2020).
- [10] Fuccillo Battle, K. *et al.* Citizen science across two centuries reveals phenological change among plant species and functional groups in the northeastern us. *Journal of Ecology* **110**, 1757–1774 (2022). <https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2745.13926>.
- [11] Diez, J. M. *et al.* Forecasting phenology: from species variability to community patterns. *Ecology letters* **15**, 545–553 (2012).
- [12] Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models* (Cambridge University Press, 2006).
- [13] Wiens, J. J. *et al.* Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology letters* **13**, 1310–1324 (2010).
- [14] Freckleton, R. P., Harvey, P. H. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* **160**, 712–726 (2002).
- [15] Kochmer, J. P. & Handel, S. N. Constraints and competition in the evolution of flowering phenology. *Ecological monographs* **56**, 303–325 (1986).
- [16] Willis, C. G., Ruhfel, B., Primack, R. B., Miller-Rushing, A. J. & Davis, C. C. Phylogenetic patterns of species loss in thoreau’s woods are driven by climate change. *Proceedings of the National Academy of Sciences* **105**, 17029–17033 (2008).

- [17] Davies, T., Wolkovich, E., Kraft, N., Salamin, N. & Travers, S. E. Phylogenetic conservatism in plant phenology. *Journal of Ecology* **101**, 1520–1530 (2013).
- [18] CaraDonna, P. J. & Inouye, D. W. Phenological responses to climate change do not exhibit phylogenetic signal in a subalpine plant community. *Ecology* **96**, 355–361 (2014).
- [19] Yang, Z. *et al.* Phylogenetic conservatism in heat requirement of leaf-out phenology, rather than temperature sensitivity, in tibetan plateau. *Agricultural and Forest Meteorology* **304** (2021).
- [20] Rafferty, N. E. & Nabity, P. D. A global test for phylogenetic signal in shifts in flowering time under climate change. *Journal of Ecology* **105**, 627–633 (2017).
- [21] Larcher, W. *Plant Physiological Ecology* (Springer-Verlag, 1980).
- [22] Bonamour, S., Chevin, L. M., Charmantier, A. & Teplitsky, C. Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B-Biological Sciences* **374** (2019).
- [23] Ackerly, D. 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 (2009).
- [24] Davies, T. J., Regetz, J., Wolkovich, E. M. & McGill, B. J. Phylogenetically weighted regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology and biogeography* **28**, 275–285 (2019).
- [25] Ettinger, A. K., Buonaiuto, D. M., Chamberlain, C. J., Morales-Castilla, I. & Wolkovich, E. M. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist* **230**, 462–474 (2021).
- [26] Housworth, E. A., Martins, E. P. & Lynch, M. The phylogenetic mixed model. *The American Naturalist* **163**, 84–96 (2004).
- [27] Uyeda, J. C., Pennell, M. W., Miller, E. T., Maia, R. & McClain, C. R. The evolution of energetic scaling across the vertebrate tree of life. *The American Naturalist* **190**, 185–199 (2017).
- [28] Wolkovich, E. M. *et al.* Observed Spring Phenology Responses in Experimental Environments (OS-PREE). doi:10.5063/F1CZ35KB (2019).
- [29] Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *American journal of botany* **105**, 302–314 (2018).
- [30] Laube, J. *et al.* Chilling outweighs photoperiod in preventing precocious spring development. *Global Change Biology* **20**, 170–182 (2014).
- [31] Wolkovich, E. M. & Donahue, M. J. How phenological tracking shapes species and communities in non-stationary environments. *Biological Reviews* (2021).
- [32] Nakagawa, H. *et al.* Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological model. *Theoretical and Applied Genetics* **110**, 778–786 (2005).
- [33] Basler, D. & Körner, C. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology* **165**, 73–81 (2012).
- [34] Zohner, C. M., Benito, B. M., Svenning, J. C. & Renner, S. S. Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate Change* **6**, 1120–1123 (2016).

- [35] Hunter, A. F. & Lechowicz, M. J. Predicting the timing of budburst in temperate trees. *Journal of Applied Ecology* **29**, 597–604 (1992).
- [36] Schaber, J. & Badeck, F. Physiology-based phenology models for forest tree species in germany. *International Journal of Biometeorology* **47**, 193–201 (2003). International Conference on the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- [37] Way, D. A. & Montgomery, R. A. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment* **38**, 1725–1736 (2015).
- [38] Kramer, K. *et al.* 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 (2017).
- [39] Aitken, S. N. & Bemmels, J. B. Time to get moving: assisted gene flow of forest trees. *Evol Appl* **9**, 271–90 (2016).
- [40] Gotelli, N. J. & Graves, G. R. The temporal niche. In Gotelli, N. J. & Graves, G. R. (eds.) *Null Models in Ecology*, 95–111 (Smithsonian Institution, Washington, DC, 1996).
- [41] Grime, J. P. Evidence for existence of 3 primary strategies in plants and its relevance to ecological and evolutionary theory. *American Naturalist* **111**, 1169–1194 (1977).
- [42] Serrano-Bueno, G., Romero-Campero, F. J., Lucas-Reina, E., Romero, J. M. & Valverde, F. Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology* **37**, 10–17 (2017).
- [43] Rinne, P., Saarelainen, A. & Junttila, O. 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 (1994).
- [44] Wilczek, A. M., Cooper, M. D., Korves, T. M. & Schmitt, J. Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **111**, 7906–7913 (2014). <https://www.pnas.org/content/111/22/7906.full.pdf>.
- [45] Azeez, A. & Sane, A. P. Photoperiodic growth control in perennial trees. *Plant signaling & behavior* **10**, e1087631 (2015).
- [46] Bennett, J. M. *et al.* The evolution of critical thermal limits of life on earth. *Nature communications* **12**, 1–9 (2021).
- [47] Flynn, D. F. B. & Wolkovich, E. M. Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytologist* **219**, 1353–1362 (2018).
- [48] Molina-Venegas, R. *et al.* Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. *Ecography* **41**, 1740–1749 (2018).
- [49] Molina-Venegas, R., Morales-Castilla, I. & Rodríguez, M. Á. Unreliable prediction of b-vitamin source species. *Nature Plants* **9**, 31–33 (2023).
- [50] Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. & Jones, P. D. An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres* **123**, 9391–9409 (2018).

- 360 [51] Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of
361 meteorological forcings for land surface modeling. *Journal of Climate* **19**, 3088–3111 (2006).
- 362 [52] Baumgarten, F., Zohner, C. M., Gessler, A. & Vitasse, Y. Chilled to be forced: the best dose to wake
363 up buds from winter dormancy. *New Phytologist* **230**, 1366–1377 (2021).
- 364 [53] Buonaiuto, D. M., Donahue, M. & Wolkovich, E. M. Experimental designs for testing the interactive
365 effects of temperature and light in ecology: the problem of periodicity. *Functional Ecology* (2023).
- 366 [54] Pearse, W. D. *et al.* *Pez*: Phylogenetics for the environmental sciences. *Bioinformatics* **31**, 2888–2890
367 (2015).

Methods

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 (28) 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 data from 155 experiments across 97 papers in the updated database. (author?) (7), which used a portion (72 experiments across 49 papers) of the earlier OSPREE database, provides extensive methods on database creation and cleaning.

We focused on angiosperms (as gymnosperms are very poorly represented in spring phenology experiments), and included all budburst experiments where we could quantify chilling, forcing and photoperiod levels, resulting in 44 studies from 33 papers, and 2940 data points. In our dataset most studies come from Europe ($n=37$) and a few from North America ($n=7$). The same bias towards Europe is found across the full OSPREE dataset with less North American ($n=19$) than European ($n=60$) studies and only 3 studies located in the Southern Hemisphere. Given our need of daily gridded data for estimating chilling we only include studies from Europe and North America (see Fig. S3). Our final dataset is both geographically and taxonomically limited, mirroring the existing literature on phenological experiments and highlighting a critical need to expand this literature.

Across experiments chilling treatments were often fully or partially applied in the field, thus we estimated chilling—both in the field and applied in controlled environments—using Utah units with the `chillR` package. We estimated field chilling from 1 September to the date given for when samples were taken from the field using daily temperature data (converted to hourly) from both European (E-OBS, version 16, calculating the average of minimum and maximum daily temps, 50) and North American (v3, 51) gridded climate datasets. We also converted experimental chilling into Utah chill units, based on reported treatments (for studies with a mix of field and experimental treatments, we added field and experimentally applied Utah units). To avoid numerical instability in our models (from having predictor values on very different scales), we divided Utah units by 240 (roughly equivalent to 10 days of average chilling).

We report the Utah model because a small number of studies reported chilling only in Utah units, thus using this common metric allowed us to include the most data. The Utah model relies on the assumption that temperatures between 1.4 and 15.9°C affect endodormancy release differently, though recent findings show possibly similar effects for a wide range of temperatures (-2 to 10°C, see 52). Because chilling is a latent process an accurate model of it, especially for the 191 species in our dataset, is not currently possible (7). We found consistent results, however, using another common model of chilling—chill portions, suggesting our results are robust to the exact chilling metric used (see Tables S4 and S5 in Supplementary Information).

Forcing and photoperiod treatments occurred after chilling treatments; we report photoperiod as the length

of light and weighted these treatments by the reported photo- and thermo-periodicity (53). Most studies reported two temperatures per day across the whole experiment, one for day and night, but some had ramped temperatures and/or photoperiods (or other complexities). In these cases we built an hourly model of the full treatment period until budburst and took the mean value. For a phylogenetic tree, we pruned the megatree for seed plants (29) to extract the sub-tree containing only the species present in the OSPREE dataset, species not included in the megatree were added to the congeneric basal node age (using the function ‘congeneric.merge’ in 54), and assigned branch lengths to maintain tree ultrametricity. This addition of species to the tree can introduce polytomies (multifurcations) when many species are added to the same ancestral node, as was the case for *Acer*, for which several species were included in the OSPREE dataset but the megatree lacked species-level resolution within the genus. In total, our pruned tree had 8 polytomies affecting 46 out of 191 species. Our Bayesian hierarchical model, described below, is informed by the phylogenetic structure describing species evolutionary relationships. Errors in phylogenetic topology and branching times could thus impact model estimates, although if errors were large the contribution of phylogeny would simply be scaled to zero. To assess whether the inclusion of polytomies in our data biased model estimates, we ran sensitivity analyses excluding these species from models (see Table S8 in Supplementary Information). Our approach assumes a tree where branch lengths represent time, but it could be possible to change this assumption. For example, if the genes underlying plant responses to particular cues were known, branch lengths would directly represent mutational changes along gene sequences. In the absence of such detailed gene specific data, evolutionary time provides a useful proxy for species differences.

Bayesian hierarchical phylogenetic model

Commonly used phylogenetic regression methods today (e.g., phylogenetic generalized least squares models, PGLS, (14); phylogenetic mixed models, PMM, (26)) 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 (14). These corrections incorporated phylogenetic structure by estimating the magnitude of a transformation of a variance-covariance (VCV) matrix whose elements were derived from the amount of evolutionary history (branch lengths) shared between species on a phylogeny. The most commonly used transformation was Pagel’s λ —a multiplier of the off-diagonal elements—where estimates of $\lambda = 1$ essentially left the VCV untransformed and suggested that the residuals of the regression had phylogenetic signal consistent with Brownian motion; estimates of $\lambda = 0$ suggested no phylogenetic signal. Because the original aim of these methods was to correct for statistical bias introduced by shared evolutionary history among species, the underlying assumption of phylogenetic regressions is that phylogenetic relatedness would only affect either model residuals (in PGLS approaches, 14) or model intercepts (e.g., in many PMM approaches, 26).

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:

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

where

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

and 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:

$$\begin{aligned} \boldsymbol{\alpha} &= [\alpha_1, \dots, \alpha_n]^T \text{ such that } \boldsymbol{\alpha} \sim \mathcal{N}(\mu_{\alpha}, \boldsymbol{\Sigma}_{\alpha}) \\ \boldsymbol{\beta}_{chill} &= [\beta_{chill,1}, \dots, \beta_{chill,n}]^T \text{ such that } \boldsymbol{\beta}_{chill} \sim \mathcal{N}(\mu_{\beta_{chill}}, \boldsymbol{\Sigma}_{\beta_{chill}}) \\ \boldsymbol{\beta}_{force} &= [\beta_{force,1}, \dots, \beta_{force,n}]^T \text{ such that } \boldsymbol{\beta}_{force} \sim \mathcal{N}(\mu_{\beta_{force}}, \boldsymbol{\Sigma}_{\beta_{force}}) \\ \boldsymbol{\beta}_{photo} &= [\beta_{photo,1}, \dots, \beta_{photo,n}]^T \text{ such that } \boldsymbol{\beta}_{photo} \sim \mathcal{N}(\mu_{\beta_{photo}}, \boldsymbol{\Sigma}_{\beta_{photo}}) \end{aligned} \quad (3)$$

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:

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

where σ_i^2 is the rate of evolution across a tree for a given trait or predictor (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:

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

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. We fit all models to our data using **RStan** using 4 chains of 4,000 iterations with a warmup of 2,000 each (resulting 8,000 posterior samples), and assessed fit via \hat{R} near 1 and adequate effective sample sizes (see Tables S2-S3).

To assess if the phylogenetic mixed model (PMM) presents any advantages with respect commonly used hierarchical mixed models (HMM; see e.g., 7), beyond fitting evolutionary parameters to model predictors,

we compare results of PMM and HMM. HMM is a simplified version of PMM where off-diagonal elements of the variance-covariance phylogenetic matrices are multiplied by zero ($\lambda = 0$). Both models account for differences in sample sizes and variances for each species, by partially pooling across all data while at the same time providing species-level estimates; however the PMM will pool more strongly to closely-related species when λ is high. Additional grouping factors beyond species could be added to these models. For example, similar approaches can be used to estimate study or location effects.

Data availability

Our data comprises phenological responses of 191 species to experimental treatments of temperature and daylight. Data were subset from a larger existing dataset published by the authors (OSPREE, doi:10.5063/F1CZ35KB), which assembles the largest available dataset on experimentally determined woody plant phenology. The data included all budburst experiments where we could quantify chilling, forcing and photoperiod levels for target species, resulting in 44 studies from 33 papers. All data supporting the findings of this study are publicly available in GitHub <https://zenodo.org/records/10902899>.

Code availability

Data analyses were performed in R and Bayesian Hierarchical and Phylogenetic Mixed Models (HMM and PMM, respectively) were performed within the Stan environment. All Stan and R custom code to run models and visualize results is publicly available in GitHub <https://zenodo.org/records/10902899>.

483 **Methods-only references**