

Phylogenetic estimates of species-level phenology improve ecological forecasting

Received: 23 June 2023

Accepted: 22 July 2024

Published online: 15 August 2024



Ignacio Morales-Castilla¹✉, T. J. Davies^{2,3}, Geoffrey Legault³, D. M. Buonaiuto^{4,5,6}, Catherine J. Chamberlain^{4,5,7}, Ailene K. Ettinger^{5,8}, Mira Garner³, Faith A. M. Jones^{3,9}, Deirdre Loughnan³, William D. Pearse¹⁰, Darwin S. Sodhi³ & E. M. Wolkovich^{3,4,5}

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 new method using Bayesian hierarchical phylogenetic models and show that species-level differences are larger than the average differences between cues. Applying our method to phenological experiments manipulating temperature and day length we show 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. We thus illustrate how a focus on certain clades can bias prediction, but that predictions may be improved by integrating information on phylogeny to better estimate species-level responses. Our approach provides an advance in ecological forecasting, with implications for predicting the impacts of climate change and other anthropogenic forces on ecosystems.

The biological impacts of climate change will have major implications for ecosystem functioning and stability. With rising global temperatures, many species have shifted their geographical 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 forecasts³.

Although 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 cues¹. 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 (for example, 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 (for example, functional groups) have improved ecosystem models^{8,9} but still capture only a fraction of the important variability¹⁰.

Recent efforts that have attempted to model species-specific responses to the environment¹¹ 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

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, Vancouver, British Columbia, Canada. ³Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada. ⁴Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. ⁵Arnold Arboretum of Harvard University, Boston, MA, USA. ⁶Department of Environmental Conservation, University of Massachusetts—Amherst, Amherst, MA, USA. ⁷The Nature Conservancy, Durham, NC, USA. ⁸The Nature Conservancy of Washington, Seattle, WA, USA. ⁹Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, Umeå, Sweden. ¹⁰Department of Life Sciences, Imperial College London, Berkshire, UK. ✉e-mail: ignacio.moralesc@uah.es

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 distribution¹²). As such, they partially pool (shrink) towards estimates for species with the most data and least variable responses, making inference at the species level unreliable⁷. 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 conservatism¹³ could inhibit adaptive responses by drawing species back to an evolutionarily conserved optimum, which is suboptimal under new conditions. Although 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 forecasts¹⁴.

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, leaf-out and first flowering^{15–17}, suggests that more closely related species share more similar phenologies, probably 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 geographical centres of origin and thus their ancestors will have been exposed to—and have adapted to—similar environmental cues¹⁷.

However, approaches using traditional phylogenetic comparative methods have produced conflicting results, with some studies reporting evidence of phylogenetic structure in phenology-linked species declines (for example, ref. 16) and in some phenophases but not others (for example, ref. 18) and in responses to some cues but not others (for example, ref. 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 (for example, 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 multispecies models of plant phenological responses.

Generating robust ecological forecasts requires addressing a second major hurdle: the 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 leaf-out responding to summed heat during early spring¹⁷. These efforts fail to capture the more likely scenario for most phenotypic traits in which multiple cues interacting along evolutionary history have shaped species responses²³. For many plant species, phenological events are determined by a combination of temperature and light⁶, with additional factors (for example, other cues such as humidity or species physiology, vasculature or leaf structure) probably further mediating species responses. Although these mediating factors are not well understood⁶, they can be accounted for in models either as latent processes or by allowing non-stationarity in responses across species²⁴.

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 established⁶, especially for temperate woody species for which 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 photoperiod²⁵. Plant phenology is also one of few phenotypic traits with extensive experimental data on responses to multiple environmental cues across species. Recent multispecies analyses considering forcing, chilling and photoperiod have shown that chilling and forcing together often determine complex nonlinear responses to warming but cannot forecast beyond several well-studied species⁷.

Here we present a Bayesian framework that extends on phylogenetic mixed models (PMMs)²⁶ to examine how chilling and forcing (both metrics of temperature) and photoperiod together determine spring plant phenology. By allowing non-stationarity in species responses across the phylogeny²⁴, our model departs from previous work and assumptions of traditional phylogenetic comparative methods concerned with phylogenetic correction (for example, ref. 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 change²⁷, 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 experiments⁷) in an updated version of the Observed Spring Phenology Responses in Experimental Environments (OSPREE) database²⁸. These data combined with a published plant megatree²⁹ (that is, hypothesized phylogenetic relationships for a large number of species and built from multiple smaller clade-level trees) adjusted to our species and modelling 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 shifts²⁷ in phenological responses and highlight how our approach could advance forecasting of other critical responses to ongoing global change.

Responses to cues across species

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 and Supplementary Table 2), with responses to chilling approximately fivefold greater than that to photoperiod (phenological advances of 6.9 days per standardized unit versus 1.2 days for chilling and photoperiod, respectively; Supplementary Table 2). We estimated lower average responses to temperature compared with a model without phylogeny (model slopes for forcing and chilling decreased by 18% and 22%, respectively; Extended Data Fig. 1); responses to chilling and forcing were also more similar when including phylogeny (although chilling was still greater: 6.9 versus 6.1 per standard unit), which contrasts with previous results suggesting chilling responses are much greater than forcing^{7,30}.

These average estimates, however, fail to capture the large differences in species responses to both chilling and forcing (Fig. 1 and Supplementary Table 6). 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, whereas variation to forcing varied 7-fold. This variation indicates large differences between chilling and forcing

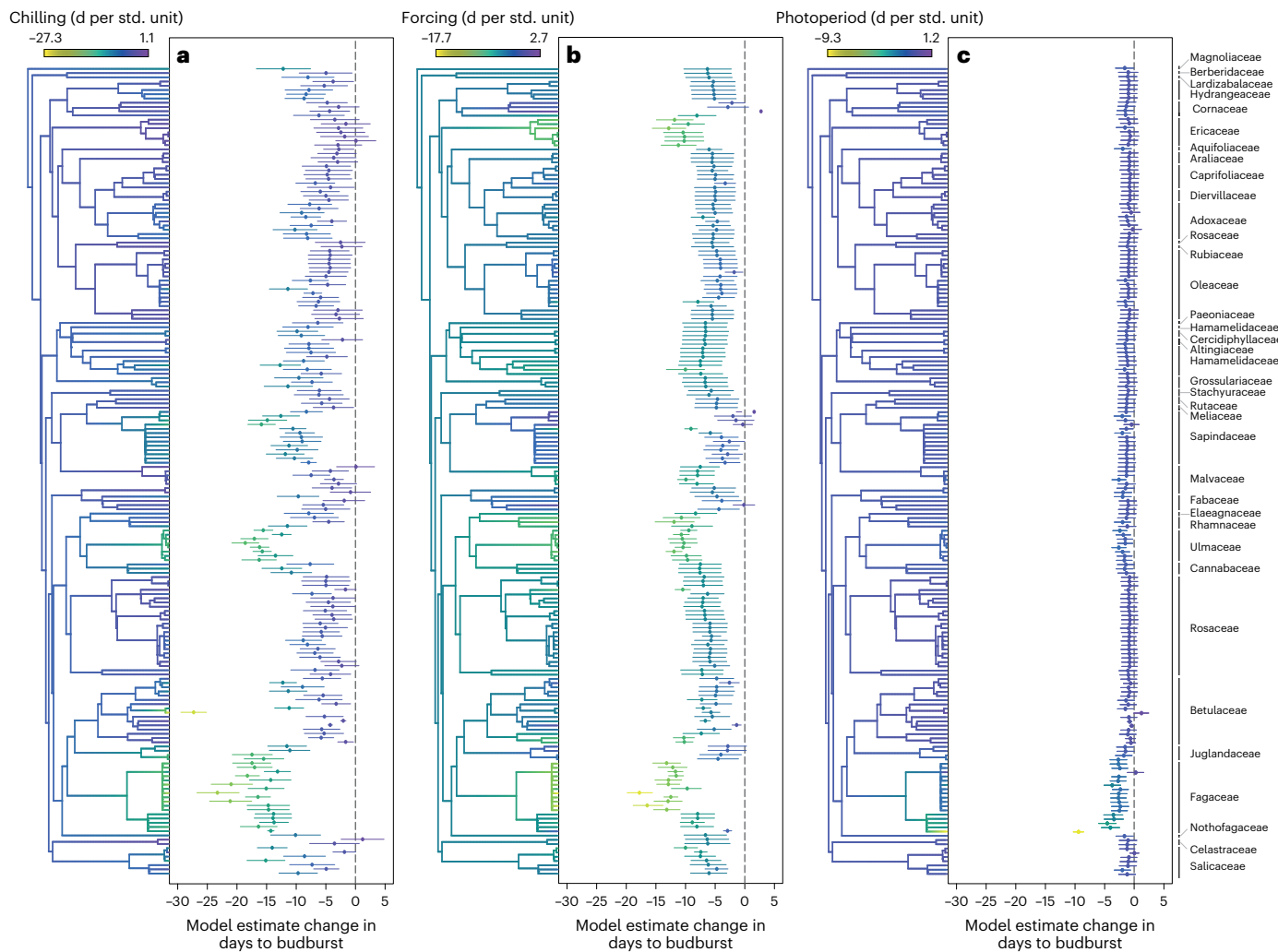


Fig. 1 | Phenological sensitivity to 3 environmental cues across 191 woody species estimated by a PMM. a–c, The environmental cues are chilling (a), forcing (b) and photoperiod (c), measured as change in days to budburst per standardized unit (d per std. unit, z-transformation) of the cues. The database used to fit the PMM comprised 44 studies, 191 species and 2,940 observations. The same phylogenetic tree is shown in each panel, coloured according to

an estimation of ancestral character states, where the states at the tips represent 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 colour scale varies in each panel. Total tree depth is 81 Myr.

occur due to differences found at the species level rather than due to differences across species (for example, the average effect across species as previously suggested^{7,30}). 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 and 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, whereas 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: some of these clades (for example, 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 in which the genetic basis for responses to one cue (for example, forcing) has been selected for alongside responses to another cue (for example, 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}. In addition, linkage or pleiotropism among loci associated with different cues³² 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. 4) and some genera, such as *Tilia* and *Rhododendron* (Ericaceae), showed strong responses to forcing but weak responses to chilling, whereas others, such as *Acer* (Sapindaceae), show moderately strong responses to chilling but weak responses to forcing (Fig. 1). Thus, species sensitivity to one cue does not constrain sensitivity to another cue and it seems selection can operate independently on responses to different cues²².

In contrast to temperature cues (chilling and forcing), species-level responses to photoperiod were almost uniform across species. This consistency provides insight on a large debate over the prevalence of photoperiod cues in temperate trees in which previous experiments^{33,34} and models (for example, refs. 35,36) suggested important variability across species that may constrain the responses of certain species to warming³⁷. Our results indicate variability is limited to a handful of species in Fagaceae, which have been particularly well studied, especially

Fagus sylvatica (for example, refs. 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 the source of most evidence of local adaptation in photoperiod cues for spring phenology (for example, ref. 38), in contrast with common garden studies of other species, which find little evidence of local adaptation in spring (but not autumn) phenology³⁹. The uniformity of response to photoperiod in our results supports this latter view of generally low local adaptation in photoperiod cues for spring phenology (that is, if local adaptation was 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. Although responses to each cue were phylogenetically structured, with closely related species showing 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 (as measured by Pagel's λ ; $\lambda = 0.65$ and $\lambda = 0.54$ for forcing and chilling, respectively). Phylogenetic structure in species responses to photoperiod was comparatively weak ($\lambda = 0.4$; Fig. 2 and Supplementary Table 2).

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 that also covaries with phylogeny²⁴. This fits fundamentally with the idea that early-season phenology plays a critical role in shaping species temporal niches⁴⁰ 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 (for example, refs. 31,41).

Weak phylogenetic signal in photoperiod sensitivity (Fig. 2) might seem at odds with the uniformity of species response—that is, 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 expectations¹³. Although 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 probable⁴², and subsequent selection on photoperiod sensitivity was constrained by stabilizing selection operating on other life-history attributes sensitive to photoperiod (for example, refs. 43–45), we would predict both low interspecific variation and weak phylogenetic signal in responses, which would match observations. This latter interpretation is also consistent with our estimates of lower σ^2 for photoperiod responses (Fig. 2). Here, as in more traditional phylogenetic comparative methods, σ^2 represents the rate of evolution and thus our results suggest that photoperiod responses are also evolving slower than temperature responses (Supplementary Fig. 6).

Phylogenetic conservatism (high λ) and slow evolutionary rates (low σ^2) 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 in which responses are dominated by forcing cues might be more vulnerable to future warming because phylogenetic conservatism (λ) in forcing is higher compared with other cues and its evolutionary rate (σ^2) is lower than that estimated for chilling. This is misleading, however, as estimates of λ are independent of 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

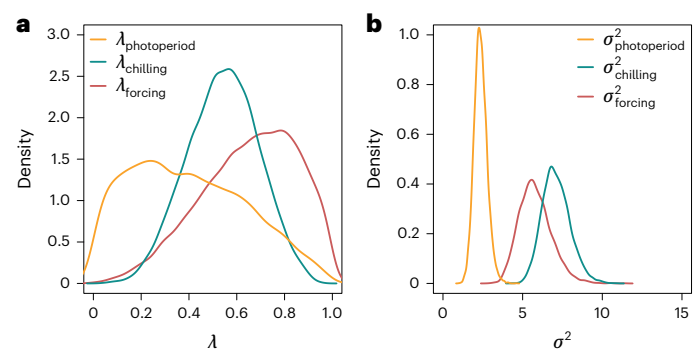


Fig. 2 | Posterior distributions of phylogenetic parameters. a, b. The lines show density plots comparing the λ (a) and σ^2 (b) parameters estimated for each cue in the phylogenetic model: chilling (blue), forcing (red) and photoperiod (orange). The λ parameter measures phylogenetic signal in each estimate of cue sensitivity (or how close species are similarly sensitive to a given cue, in a scale ranging from zero to one). The σ^2 parameter indicates evolutionary rate or accumulation of phenotypic variation with evolutionary time.

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 that estimated from more widely used hierarchical models. This was especially noticeable in temperature responses (chilling variance across species means ($\text{var}(\beta_{\text{chill},j})$) (from equation (2)) was estimated as 23.55 in the phylogenetic model versus 17.47 in the non-phylogenetic model; forcing variance across means ($\text{var}(\beta_{\text{force},j})$) was 8.75 compared with 5.01), whereas photoperiod, which had low phylogenetic structure, was more similar across approaches (photoperiod variance across means ($\text{var}(\beta_{\text{photo},j})$) was 0.83 compared with 0.64).

The consequences of including shared evolutionary history in forecasting are most apparent for poorly sampled species nested within more well-sampled clades (Figs. 3 and 4 and Extended Data Fig. 2). 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 (Fig. 4 and Extended Data Fig. 2). In contrast, forecasts for *Betula pendula*, which is one of the most sampled species, are nearly identical across models (Fig. 4 and Extended Data Fig. 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 (Figs. 3 and 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 Fig. 3). Thus, traditional (non-phylogenetically informed) approaches that partially pool across species (most hierarchical models in ecology; for example, refs. 7,47) may also lead to less precise predictions and forecasts of phenology for individual species, although overall model accuracy might still appear reasonable (Extended Data Fig. 4). Another advantage of our Bayesian approach is that we are also better able to accommodate imprecision in the data that inform our model, which might arise from

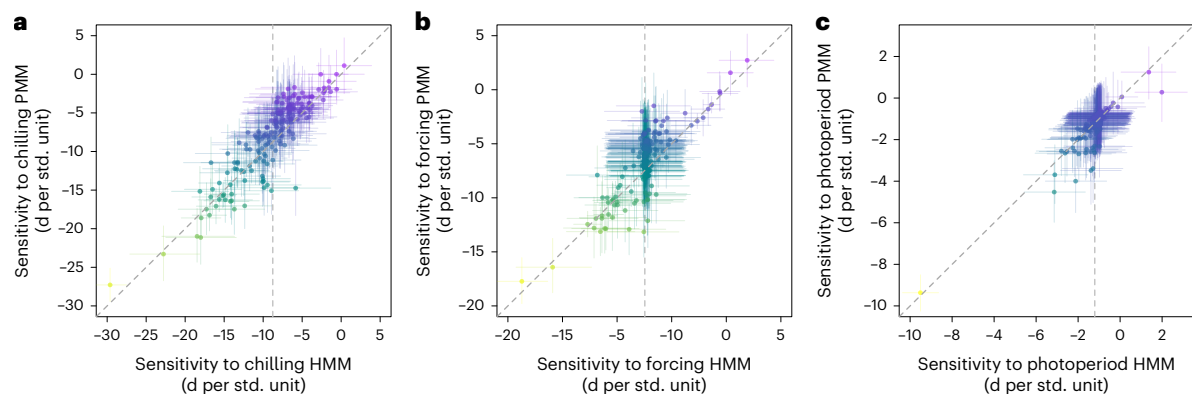


Fig. 3 | Correlations between parameters estimated by the phylogenetic and non-phylogenetic model. The PMM (y axis) accounts for phylogenetic structure on each phenological cue and the more commonly used HMM (x axis) assumes species are exchangeable (where λ is constrained to be equal to zero, x axis). **a–c**, The correlations for sensitivity to chilling (**a**), forcing (**b**) and photoperiod (**c**) are shown. Although 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**). Dashed grey 1:1 lines also shown. Species estimates are shown as mean values \pm 50% uncertainty intervals with estimate colours in the same scale as in Fig. 1. Note that uncertainties of each species are higher in the HMM than in the PMM (Extended Data Fig. 2). The database used to fit both models comprised 44 studies, 191 species and 2,940 observations.

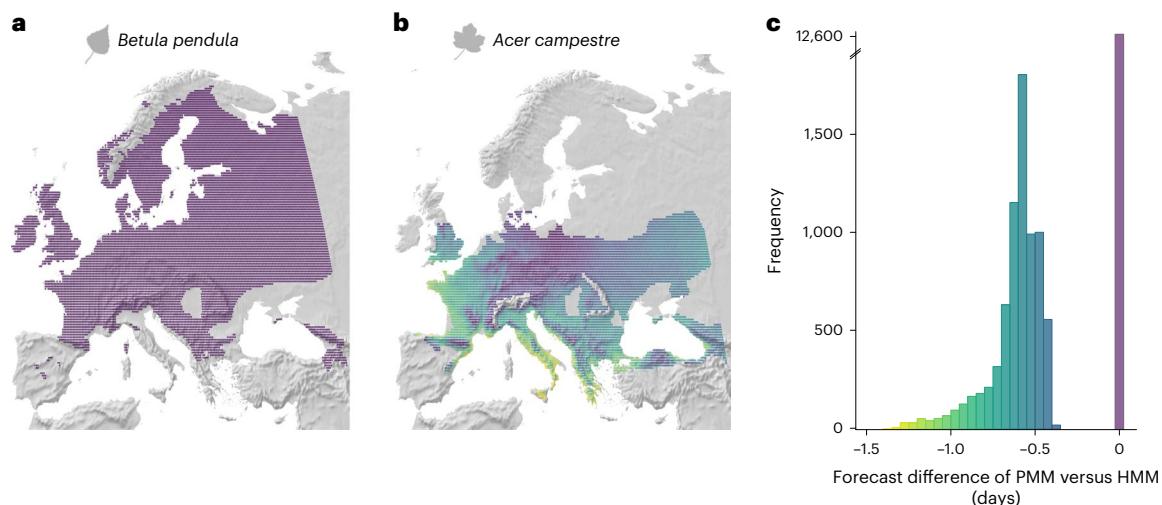


Fig. 4 | Comparison of forecasts of phenological shifts resulting from a phylogenetic (PMM) and a non-phylogenetic hierarchical (HMM) approach. Phenological shifts were computed as the difference between predictions under the current climate versus a 2 °C warmer climate. **a,b**, Differences in forecasted 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 the number of days between the shifts predicted by PMM and HMM, with values coloured according to the histograms in **c** (days here are relative to start of forcing conditions, not calendar days). Histograms compare the distributions of forecasted values for each species depicted in the maps (**c**). See Supplementary Information for details on forecast calculation. Basemaps in **a** and **b** from Natural Earth.

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 clades 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 with a hierarchical model (Extended Data Figs. 5 and 6; see ‘Leave-one-clade-out model cross-validation’ in Supplementary Information).

The contrasts between temperature and photoperiod responses—in both their variability across species and phylogenetic structure—have important implications for generating multispecies forecasts. Notably, responses to photoperiod appear weaker, more uniform

across species and less phylogenetically constrained compared with responses 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}.

Although 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. 7), estimating the full variability across species and reducing uncertainty around individual species estimates—while also 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 adaptation to warming.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41558-024-02102-2>.

References

- IPCC *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (eds Field, C. B. et al) (Cambridge Univ. Press, 2014).
- Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
- Richardson, A. D. et al. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric. For. Meteorol.* **169**, 156–173 (2013).
- Dietze, M. *Ecological Forecasting* (Princeton Univ. Press, 2017).
- Lewis, A. S. et al. The power of forecasts to advance ecological theory. *Methods Ecol. Evol.* **14**, 746–756 (2023).
- Chuine, I. & Regniere, J. Process-based models of phenology for plants and animals. *Annu. Rev. Ecol. Evol. Syst.* **48**, 159–182 (2017).
- Ettinger, A. et al. Winter temperatures predominate in spring phenological responses to warming. *Nat. Clim. Change* **10**, 1137–1142 (2020).
- Moorcroft, P., Hurtt, G. & Pacala, S. A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecol. Monogr.* **71**, 557–585 (2001).
- Griffith, D. M. et al. Lineage-based functional types: characterising functional diversity to enhance the representation of ecological behaviour in land surface models. *New Phytol.* **228**, 15–23 (2020).
- Fuccillo Battle, K. et al. Citizen science across two centuries reveals phenological change among plant species and functional groups in the northeastern US. *J. Ecol.* **110**, 1757–1774 (2022).
- Diez, J. M. et al. Forecasting phenology: from species variability to community patterns. *Ecol. Lett.* **15**, 545–553 (2012).
- Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2006).
- Wiens, J. J. et al. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* **13**, 1310–1324 (2010).
- Freckleton, R. P., Harvey, P. H. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
- Kochmer, J. P. & Handel, S. N. Constraints and competition in the evolution of flowering phenology. *Ecol. Monogr.* **56**, 303–325 (1986).
- 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. *Proc. Natl Acad. Sci. USA* **105**, 17029–17033 (2008).
- Davies, T., Wolkovich, E., Kraft, N., Salamin, N. & Travers, S. E. Phylogenetic conservatism in plant phenology. *J. Ecol.* **101**, 1520–1530 (2013).
- 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).
- Yang, Z. et al. Phylogenetic conservatism in heat requirement of leaf-out phenology, rather than temperature sensitivity, in Tibetan plateau. *Agric. For. Meteorol.* **304**, 108413 (2021).
- Rafferty, N. E. & Nability, P. D. A global test for phylogenetic signal in shifts in flowering time under climate change. *J. Ecol.* **105**, 627–633 (2017).
- Larcher, W. *Plant Physiological Ecology* (Springer, 1980).
- Bonamour, S., Chevin, L. M., Charmanier, A. & Teplitsky, C. Phenotypic plasticity in response to climate change: the importance of cue variation. *Philos. Trans. R. Soc. B* **374**, 20180178 (2019).
- Ackerly, D. Conservatism and diversification of plant functional traits: evolutionary rates versus phylogenetic signal. *Proc. Natl Acad. Sci. USA* **106**, 19699–19706 (2009).
- Davies, T. J., Regetz, J., Wolkovich, E. M. & McGill, B. J. Phylogenetically weighted regression: a method for modelling non-stationarity on evolutionary trees. *Glob. Ecol. Biogeogr.* **28**, 275–285 (2019).
- 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 Phytol.* **230**, 462–474 (2021).
- Housworth, E. A., Martins, E. P. & Lynch, M. The phylogenetic mixed model. *Am. Nat.* **163**, 84–96 (2004).
- 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. *Am. Nat.* **190**, 185–199 (2017).
- Wolkovich, E. M. et al. Observed Spring Phenology Responses in Experimental Environments (OSPREE). *Knowledge Network for Biocomplexity* <https://doi.org/10.5063/FICZ35KB> (2019).
- Smith, S. A. & Brown, J. W. Constructing a broadly inclusive seed plant phylogeny. *Am. J. Bot.* **105**, 302–314 (2018).
- Laube, J. et al. Chilling outweighs photoperiod in preventing precocious spring development. *Glob. Change Biol.* **20**, 170–182 (2014).
- Wolkovich, E. M. & Donahue, M. J. How phenological tracking shapes species and communities in non-stationary environments. *Biol. Rev.* **96**, 2810–2827 (2021).
- Nakagawa, H. et al. Flowering response of rice to photoperiod and temperature: a QTL analysis using a phenological model. *Theor. Appl. Genet.* **110**, 778–786 (2005).
- Basler, D. & Körner, C. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agric. For. Meteorol.* **165**, 73–81 (2012).
- 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. *Nat. Clim. Change* **6**, 1120–1123 (2016).
- Hunter, A. F. & Lechowicz, M. J. Predicting the timing of budburst in temperate trees. *J. Appl. Ecol.* **29**, 597–604 (1992).
- Schaber, J. & Badeck, F. Physiology-based phenology models for forest tree species in Germany. *Int. J. Biometeorol.* **47**, 193–201 (2003).
- Way, D. A. & Montgomery, R. A. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant Cell Environ.* **38**, 1725–1736 (2015).
- 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. *Agric. For. Meteorol.* **234**, 172–181 (2017).
- Aitken, S. N. & Bemmels, J. B. Time to get moving: assisted gene flow of forest trees. *Evol. Appl.* **9**, 271–290 (2016).
- Gotelli, N. J. & Graves, G. R. In *Null Models in Ecology* (eds Gotelli, N. J. & Graves, G. R.) 95–111 (Smithsonian Institution, 1996).
- Grime, J. P. Evidence for existence of 3 primary strategies in plants and its relevance to ecological and evolutionary theory. *Am. Nat.* **111**, 1169–1194 (1977).
- Serrano-Bueno, G., Romero-Campero, F. J., Lucas-Reina, E., Romero, J. M. & Valverde, F. Evolution of photoperiod sensing in plants and algae. *Curr. Opin. Plant Biol.* **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. *Physiol. Plant.* **90**, 451–458 (1994).
 44. Wilczek, A. M., Cooper, M. D., Korves, T. M. & Schmitt, J. Lagging adaptation to warming climate in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **111**, 7906–7913 (2014).
 45. Azeez, A. & Sane, A. P. Photoperiodic growth control in perennial trees. *Plant Signal. Behav.* **10**, e1087631 (2015).
 46. Bennett, J. M. et al. The evolution of critical thermal limits of life on earth. *Nat. Commun.* **12**, 1198 (2021).
 47. Flynn, D. F. B. & Wolkovich, E. M. Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytol.* **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. *Nat. Plants* **9**, 31–33 (2023).
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.
- © The Author(s), under exclusive licence to Springer Nature Limited 2024

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 OSPREE database. In July 2019 we updated a previous version of this database²⁸ by reviewing all articles found through searching ISI Web of Science and Google Scholar with the following terms:

- (1) TOPIC = (budburst OR leaf-out) AND (photoperiod OR day-length) AND temperature*, which yielded 623 publications
- (2) TOPIC = (budburst OR leaf-out) AND dorman*, which yielded 270 publications

We scraped data from all articles of woody species that tested for photoperiod and/or temperature effects on budburst, leaf-out or flowering, resulting in data from 155 experiments across 97 articles in the updated database. Ref. 7, which used a portion (72 experiments across 49 articles) of the previous OSPREE database, provides extensive methods on database creation and cleaning.

We focused on angiosperms (because gymnosperms are very poorly represented in spring phenology experiments) and included all budburst experiments from which we could quantify chilling, forcing and photoperiod levels, resulting in 44 studies from 33 articles and 2,940 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 (Supplementary Fig. 3). 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 v.16; calculating the average of minimum and maximum daily temperatures⁵⁰) and North American (v.3)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 only reported chilling 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 °C and 15.9 °C affect endodormancy release differently, although recent findings show possibly similar effects for a wide range of temperatures (−2 °C to 10 °C; ref. 52). As chilling is a latent process, an accurate model of it, especially for the 191 species in our dataset, is not currently possible⁷. We found consistent results, however, using another common model of chilling—chill portions, suggesting our results are robust to the exact chilling metric used (Supplementary Tables 4 and 5).

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-periodicity⁵³. Most studies reported two temperatures per day across the whole experiment, one for day and one for 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 plants²⁹ to extract the subtree 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 ref. 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 (Supplementary Table 8). 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 (for example, phylogenetic generalized least squares models¹⁴ and PMMs²⁶) were originally conceived as statistical corrections for phylogenetic non-independence across observations—generally species, thus allowing multispecies studies to meet the assumptions of linear regression¹⁴. 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 matrix untransformed and suggested that the residuals of the regression had phylogenetic signal consistent with Brownian motion; estimates of $\lambda = 0$ suggested no phylogenetic signal. As 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 phylogenetic generalized least square approaches¹⁴) or model intercepts (for example, in many PMM approaches²⁶).

As 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. This allows us to explicitly estimate the amount of phylogenetic relatedness in species sensitivities to each cue when these sensitivities are modelled in a multipredictor regression setting.

For each observation i of species j , we assumed that the timing of phenological events was generated from the following sampling distribution:

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

where

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

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

Predictors X_{chill} , X_{force} and X_{photo} are standardized chilling, forcing and photoperiod, and their effects on the phenology of species j are determined by parameters $\beta_{\text{chill},j}$, $\beta_{\text{force},j}$ and $\beta_{\text{photo},j}$, 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:

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

where the means of the multivariate normal distributions are root trait values (that is, values of cue responses before evolving across a phylogenetic tree) and $\boldsymbol{\Sigma}_i$ are $n \times n$ phylogenetic VCV 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 (that is, α_j , $\beta_{\text{force},j}$, $\beta_{\text{photo},j}$ and $\beta_{\text{chill},j}$), 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:

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

where the residual phylogenetic error terms (for example, e_{α_j}) are elements of normal random vectors from multivariate normal distributions centred on 0 with the same phylogenetic VCV matrices as in equation (4). Model code, including priors used here, are given in the Supplementary Information. We fit all models to our data using RStan using 4 chains of 4,000 iterations with a warm-up of 2,000 each (resulting 8,000 posterior samples) and assessed fit via \hat{R} near 1 and adequate effective sample sizes (Supplementary Tables 2 and 3).

To assess if the PMM presents any advantages with respect to commonly used hierarchical mixed models (HMMs; for example, ref. 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 VCV 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.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Our data comprise the phenological responses of 191 species to experimental treatments of temperature and daylight. Data were subset from

a larger existing dataset published by the authors (OSPREE²⁸), which assembles the largest available dataset on experimentally determined woody plant phenology. The data included all budburst experiments from which we could quantify chilling, forcing and photoperiod levels for target species, resulting in 44 studies from 33 articles. All data supporting the findings of this study are publicly available via Zenodo at <https://zenodo.org/records/10902899> (ref. 55). Source data are provided with this paper.

Code availability

Data analyses were performed in R. HMM and PMM were performed within the Stan environment. All Stan and R custom code to run models and visualize results is publicly available via Zenodo at <https://zenodo.org/records/10902899> (ref. 55).

References

- 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. *J. Geophys. Res.* **123**, 9391–9409 (2018).
- Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).
- Baumgarten, F., Zohner, C. M., Gessler, A. & Vitas, Y. Chilled to be forced: the best dose to wake up buds from winter dormancy. *New Phytol.* **230**, 1366–1377 (2021).
- Buonaiuto, D. M., Donahue, M. & Wolkovich, E. M. Experimental designs for testing the interactive effects of temperature and light in ecology: the problem of periodicity. *Funct. Ecol.* **37**, 1747–1756 (2023).
- Pearse, W. D. et al. Pez: phylogenetics for the environmental sciences. *Bioinformatics* **31**, 2888–2890 (2015).
- Morales-Castilla, I. MoralesCastilla/PhenoPhyloMM: initial release. Zenodo <https://doi.org/10.5281/zenodo.10902899> (2024).

Acknowledgements

We thank the many researchers who conducted the experiments and built the phylogeny used in this Article. I.M.-C. acknowledges funding from the Spanish Ministry for Science and Innovation (grant number PID2019/109711RJ-I00 to I.M.-C.) and from the Comunidad de Madrid and University of Alcalá (grant number CM/BG/2021-003 to I.M.-C.). E.M.W. acknowledges funding from 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

I.M.-C., D.M.B., C.J.C., A.K.E. and E.M.W. conceived the study. 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 paper. I.M.-C., T.J.D., G.L., D.M.B., C.J.C., A.K.E., M.G., F.A.M.J., D.L., W.D.P., D.S.S. and E.M.W. reviewed and revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

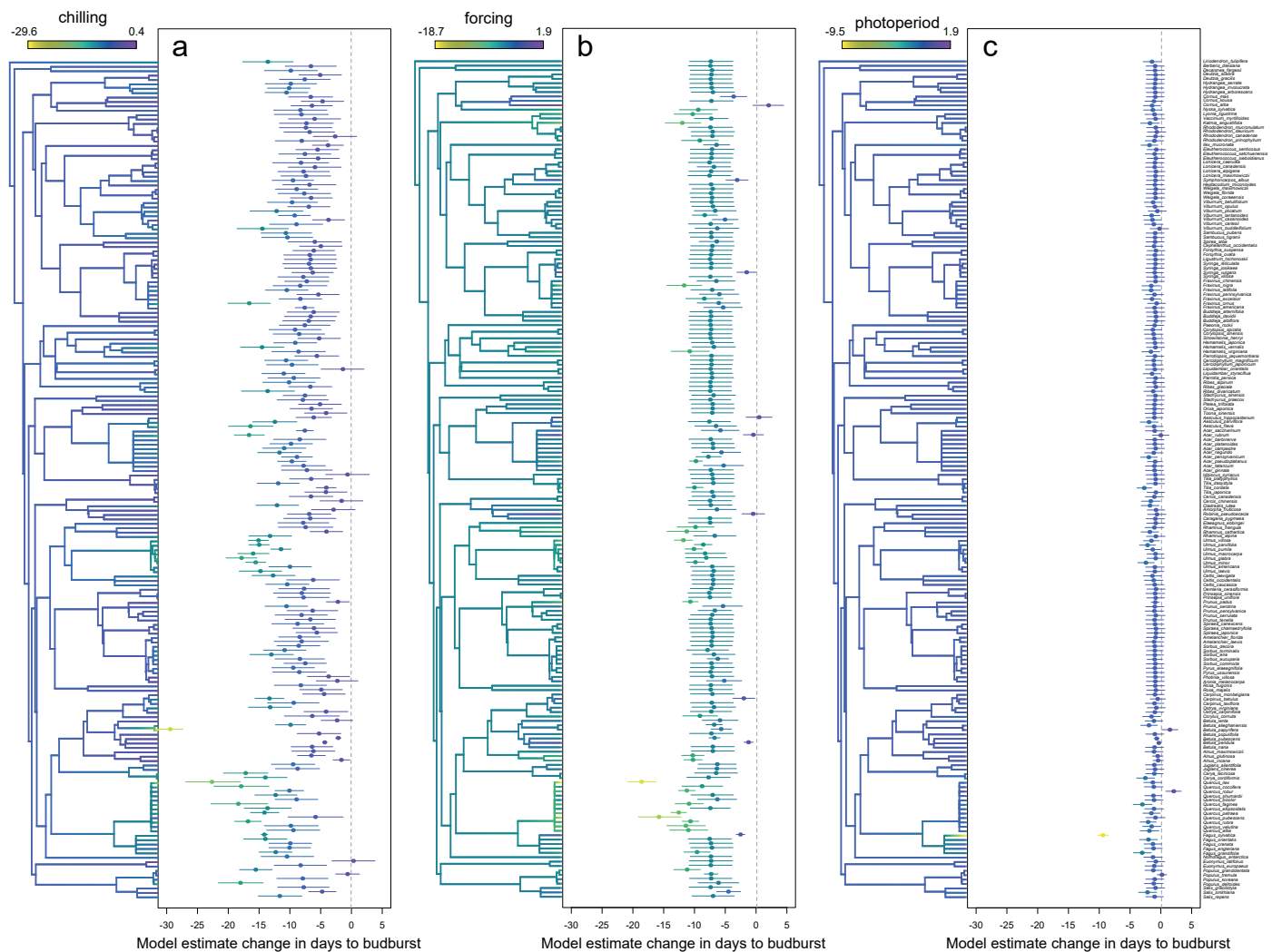
Extended data is available for this paper at <https://doi.org/10.1038/s41558-024-02102-2>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41558-024-02102-2>.

Correspondence and requests for materials should be addressed to Ignacio Morales-Castilla.

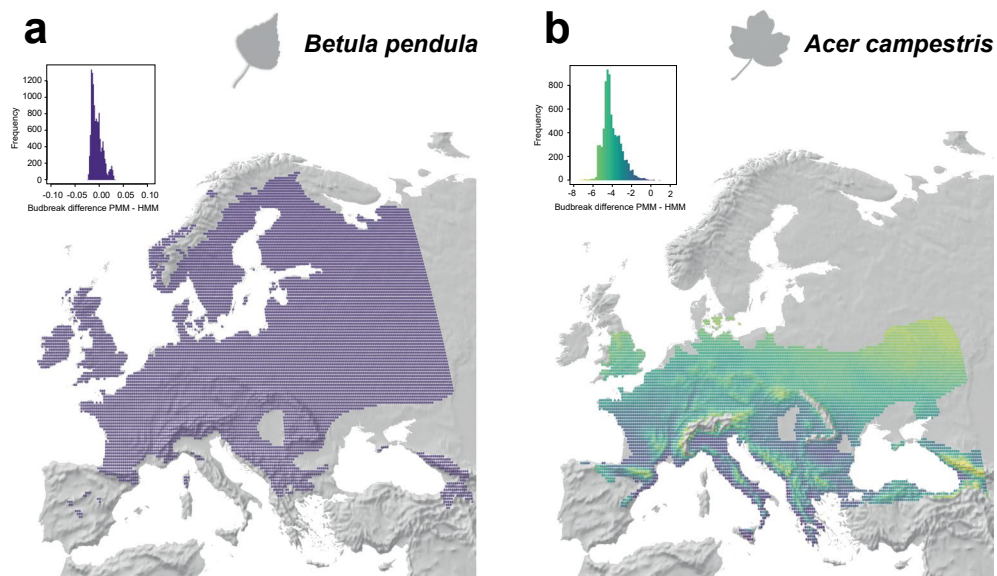
Peer review information *Nature Climate Change* thanks Junhu Dai, Susan Everingham and Daijiang Li for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



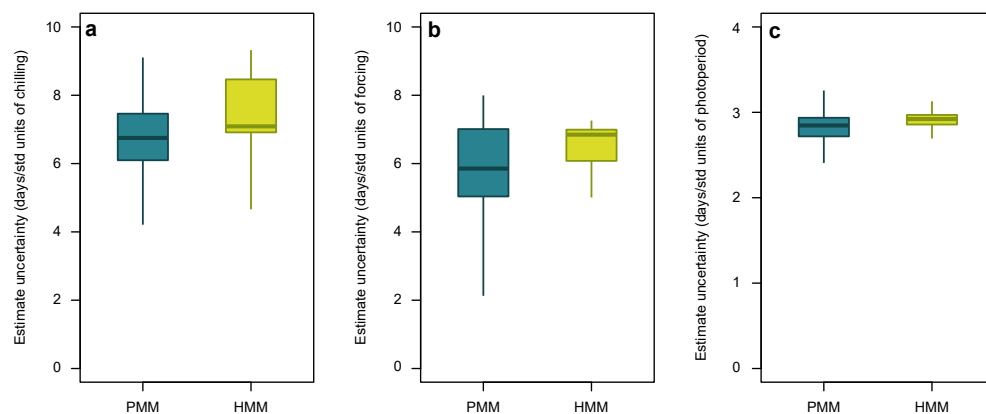
Extended Data Fig. 1 | Phenological sensitivity to three environmental cues across 191 woody species estimated by a non-phylogenetic Hierarchical Mixed Model. Non-phylogenetic phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured in change in days to budburst per standardized unit (z-transformation) of the cues across 191 tree species. Sensitivity estimates are computed by commonly used hierarchical model where phylogenetic distances are not accounted for ($\lambda = 0$). The database

used to fit the 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. Species sensitivities are shown as mean values \pm 50% uncertainty intervals in the diagrams. Note that the color scale varies in each panel. Total tree depth is 81 My.



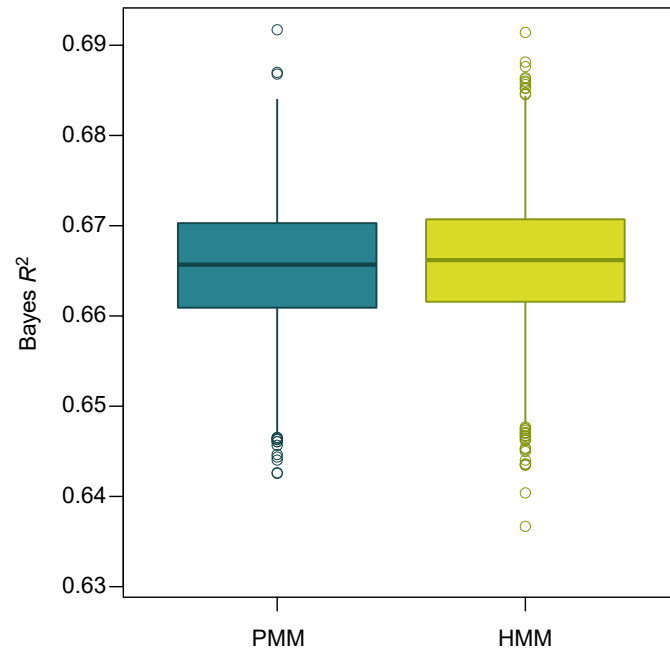
Extended Data Fig. 2 | Comparison of phenological forecasts by phylogenetic and non-phylogenetic models. Maps comparing projections of phylogenetic (PMM) against non-phylogenetic (HMM) models into the European distributions of two overlapping species, one well represented in the dataset *Betula pendula*

(a) and one underrepresented *Acer campestre* (b). The color scale shown in maps and histograms reflects budbreak differences between models where days are relative to start of forcing conditions, not calendar days.



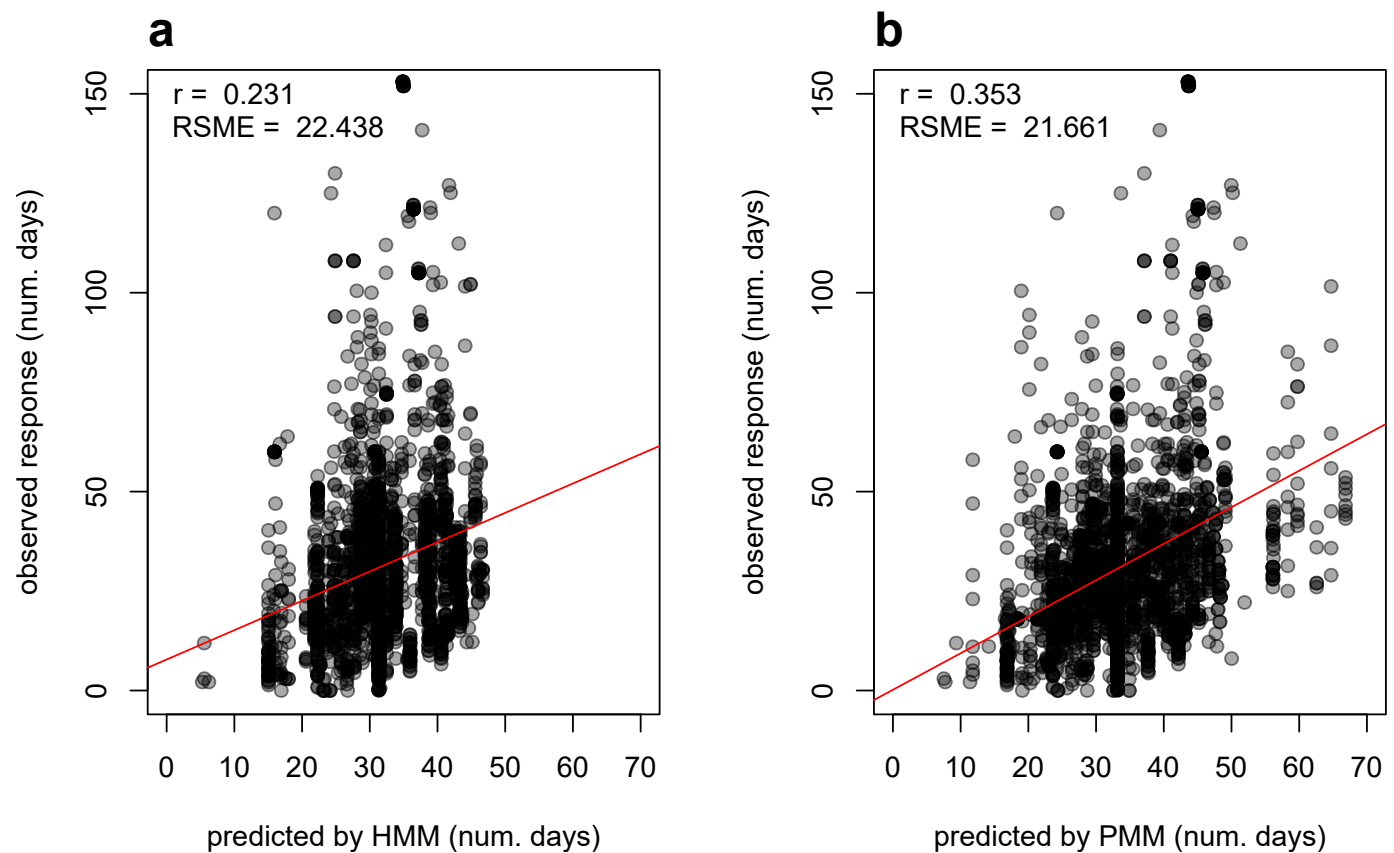
Extended Data Fig. 3 | Comparison of uncertainty around cue sensitivities estimated by the phylogenetic and non-phylogenetic models. Species uncertainties in model estimation of sensitivity to chilling (a), forcing (b) and photoperiod (c) of individual species between the phylogenetic model with estimated λ (PMM), and the non-phylogenetic model with $\lambda = 0$ (HMM).

The non-phylogenetic model increases uncertainty. Box plots are calculated for ($n = 191$) uncertainty measurements corresponding to each species in our dataset. Box upper and lower limits represent the first and third quartiles (25th and 75th percentiles, respectively), the median is represented as the horizontal line internal to the box and whiskers reach 1.5 times the interquartile range.



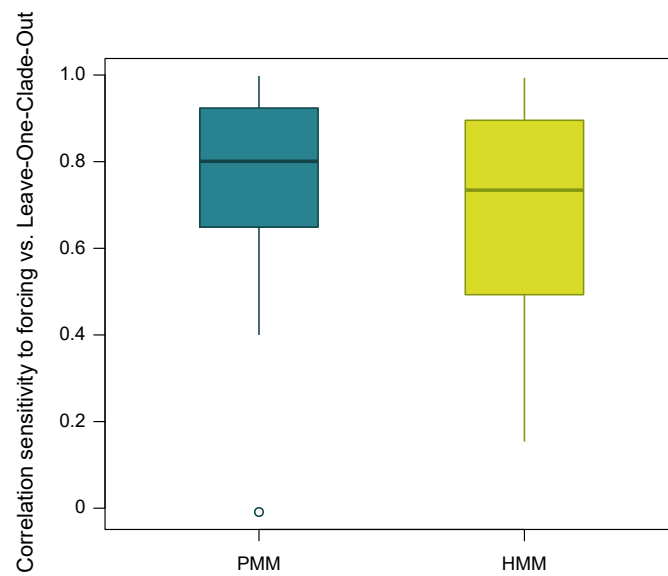
Extended Data Fig. 4 | Comparison of overall model accuracy between the phylogenetic model and the non-phylogenetic models. Accuracy is measured by Bayes R^2 . There are no differences in accuracy even if individual species estimates markedly differ between models. Bayes R^2 was computed across the posterior distributions of predicted values for each model (effective sample size

is, $n = 636$ for PMM, and $n = 649$ for HMM). Box upper and lower limits represent the first and third quartiles (25th and 75th percentiles, respectively), the median is represented as the horizontal line internal to the box and whiskers reach 1.5 times the interquartile range.



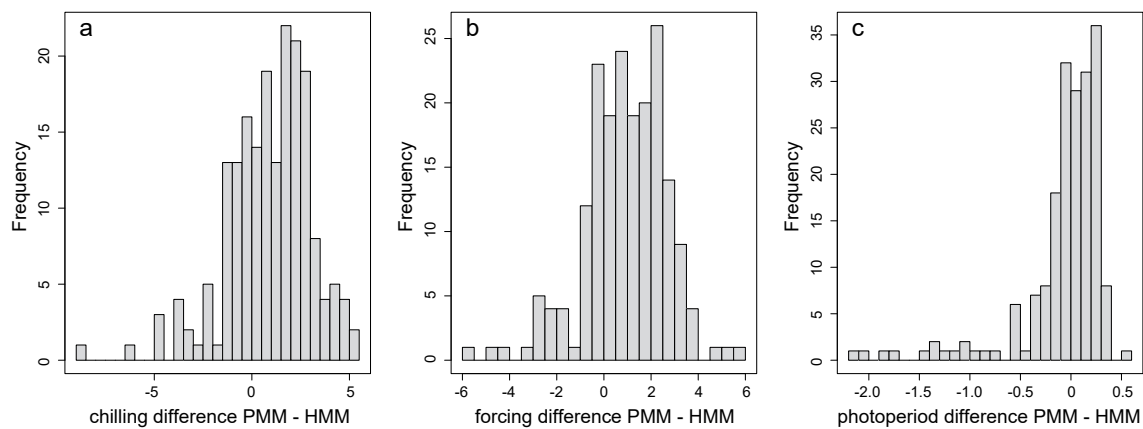
Extended Data Fig. 5 | Comparison of observed vs. predicted values of our response variable (number of days to budburst) for held out species in cross-validation analyses. Predicted values for days to budburst are modelled by the more traditional hierarchical mixed model (HMM; panel **a**) and the phylogenetic

mixed model we present (PMM; panel **b**). Observed and predicted values according to the leave-one-clade-out scheme (see Extended Methods section in Supplementary Information) were for the 2435 observations belonging to the 112 species in the 25 genera with more records.



Extended Data Fig. 6 | Comparison of correlations between cue sensitivities based on a leave-one-clade-out cross validation. Correlations are computed between estimated slopes for species in subset models after leaving out a genus, and the full model including forcing as the predictor, for both the phylogenetic mixed model (PMM) and the hierarchical mixed model (HMM). Box plots are

computed across ($n = 25$) iterations where one genus with at least 2 species was left out. Box upper and lower limits represent the first and third quartiles (25th and 75th percentiles, respectively), the median is represented as the horizontal line internal to the box and whiskers reach 1.5 times the interquartile range.



Extended Data Fig. 7 | Bias in estimation of sensitivity to environmental cues by non-phylogenetic vs. phylogenetic models. Histograms show the difference between the phylogenetic model (PMM) with estimated λ against the non-phylogenetic (HMM) model with $\lambda = 0$, in their estimations of species sensitivities

to three environmental cues: chilling (a), forcing (b) and photoperiod (c). Positive values indicate that sensitivities estimated by the non-phylogenetic model are smaller than those estimated by the phylogenetic model.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a

Confirmed
- ☐

☒

The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒

☐

A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐

☒

The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☐

☒

A description of all covariates tested
- ☐

☒

A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐

☒

A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒

☐

For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
Give P values as exact values whenever suitable.
- ☐

☒

For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐

☒

For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐

☒

Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- Data collection

All data collection and data cleaning tasks were performed in R 4.0.4
- Data analysis

Data analyses were performed in R 4.0.4 and Bayesian Hierarchical and Phylogenetic Mixed Models (HMM and PMM, respectively) were performed within the Stan environment. All custom code will be made available through GitHub.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

- All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
 - A description of any restrictions on data availability
 - For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data supporting the findings of this study are publicly available (see Supporting Information for details and online links).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences

☐ Behavioural & social sciences

☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

This study analyzes phenological responses of 191 woody plant species to known environmental drivers (i.e., temperature, daylight) while accounting for both species variability and phylogenetic relationships. Specifically, it models time to budbreak as a function of forcing, chilling and photoperiod in a hierarchical setting, where phylogenetic structure is allowed to inform species level sensitivities to each cue.

Research sample

The research sample comprises phenological responses of 191 species to experimental treatments of temperature and daylight. Data were extracted from an existing dataset published by the authors (OSPREE, doi:10.5063/F1CZ35KB), which assembles the largest available dataset on experimentally determined woody plant phenology. The analysis focuses 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.

Sampling strategy

All data available at the intersection between species in the OSPREE dataset and in the Smith and Brown (2018) phylogenetic megatree, were utilized to parameterize phenological models and thus, no extra subsampling procedures were applied.

Data collection

n/a

Timing and spatial scale

The database includes experimental observations for specimens sampled within the 1974-2019 period, for temperate species distributed across North America and Europe.

Data exclusions

No data were excluded from analyses.

Reproducibility

The findings of our modelling approach are reproducible given the code and data that accompanies our work (see data and code availability statement). Reproducibility of results from Bayesian models have been verified by three co-authors who run the models in their computers and checked that results were compatible.

Randomization

n/a

Blinding

Blinding was not relevant to our study's approach.

Did the study involve field work?

☐ Yes

☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging