

# Phylogenetic estimates of species-level phenology improve ecological forecasting

October 14, 2023

## Authors:

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

## *Author affiliations:*

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

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

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

<sup>4</sup>Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA

<sup>5</sup>Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA

<sup>6</sup>Department of Environmental Conservation, University of Massachusetts-Amherst, 160 Holdsworth Way, Amherst, MA, USA

<sup>7</sup>The Nature Conservancy, 334 Blackwell St Ste 300, Durham, NC, USA

<sup>8</sup>The Nature Conservancy of Washington, 74 Wall Street, Seattle, WA USA

<sup>10</sup>Department of Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden

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

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

## Abstract

Our ability to adapt to climate change requires accurate ecological forecasting to predict shifts in key ecosystem services, such as carbon storage and biodiversity maintenance. Current forecasts, however, have generally failed to capture the important variability in biological responses, especially observed across species. Here, using novel Bayesian hierarchical phylogenetic models, we overcome this challenge to estimate species-level responses across 191 temperate woody plant species for which we had experimental data on how their phenology respond to two major environmental cues of spring phenology: temperature and daylength. We find that variation is greater across species than across cues, suggesting the current focus on identifying which cue is the strongest may be less important than predicting how each species responds to a combination of cues. Further, our model provides insights on how evolutionary history has shaped responses to cues, suggesting stronger evolutionary dynamics in responses to temperature—especially cool winter temperatures—than daylength. Our approach provides a major advance in ecological forecasting, with implications for predicting the impacts of climate change and other anthropogenic forces on species interactions and ecosystem functioning.

## <sup>1</sup> Introduction

<sup>2</sup> The biological impacts of climate change will have major implications for ecosystem functioning  
<sup>3</sup> and stability. With rising global temperatures many species have shifted their geographic distri-  
<sup>4</sup> butions northward in space and recurring life-history events—their phenology—earlier in time  
<sup>5</sup> (IPCC, 2014; Parmesan and Yohe, 2003), against a background of high variability. These shifts  
<sup>6</sup> have cascading consequences on many ecosystem services including carbon storage, making both  
<sup>7</sup> mitigation and human adaptation to future warming dependent on accurate ecological forecasts  
<sup>8</sup> (Richardson et al., 2013).

<sup>9</sup> While ecological forecasting has improved over recent years (Dietze, 2017; Lewis et al., 2022), it  
<sup>10</sup> remains a challenge to reproduce the high variability observed in biological responses such as phe-  
<sup>11</sup> nology, physiology or demography to environmental cues (IPCC, 2014). Some of this variability  
<sup>12</sup> results from the complexity of climate change itself, including regional and seasonal variation in  
<sup>13</sup> warming that underlies average trends alongside shifts in other climate axes (e.g. precipitation).  
<sup>14</sup> Much of it, could be driven by species-specific variation, reflecting evolved differences in species'  
<sup>15</sup> sensitivities to underlying environmental cues and their interactions. Unfortunately, we can  
<sup>16</sup> only estimate the sensitivities to cues for a few well-studied species (Chuine and Regniere, 2017;  
<sup>17</sup> Ettinger et al., 2020). In the absence of detailed data on individual species, species groupings  
<sup>18</sup> (e.g., functional groups) have improved ecosystem models (Moorcroft et al., 2001; Griffith et al.,  
<sup>19</sup> 2020), but still capture only a fraction of the important variability (Fuccillo Battle et al., 2022).

<sup>20</sup> Recent efforts that have attempted to model species-specific responses to the environment (Diez  
<sup>21</sup> et al., 2012) are often restricted by data availability—especially the common problem that data  
<sup>22</sup> are often prevalent for some species and sparse across others. The rise of Bayesian hierarchical  
<sup>23</sup> models can allow inference across species in such cases. However, underlying most hierarchical  
<sup>24</sup> models is an implicit assumption that species are exchangeable (all species represent samples  
<sup>25</sup> drawn form the same underlying distribution, Gelman and Hill, 2006), and they thus partially  
<sup>26</sup> pool ('shrink') towards estimates for species with the most data and least variable responses,  
<sup>27</sup> making inference at the species-level unreliable (Ettinger et al., 2020). More reliable estimates  
<sup>28</sup> of species-level responses would allow us to better incorporate species differences into models of  
<sup>29</sup> ecosystem change.

<sup>30</sup> Including the evolutionary history of species relationships in models of species responses could  
<sup>31</sup> provide more robust species-level estimates than current approaches and a better understanding  
<sup>32</sup> of the evolutionary constraints that might limit adaptation to change. For example, strong  
<sup>33</sup> phylogenetic niche conservatism (Wiens et al., 2010) could potentially inhibit adaptive responses  
<sup>34</sup> by drawing species back to an evolutionary conserved optimum, which is sub-optimal under new  
<sup>35</sup> conditions. While incorporating such evolutionary history is traditionally seen as necessary  
<sup>36</sup> either as a statistical correction or to better understand species' evolutionary history, the use of  
<sup>37</sup> such phylogenetic information should also improve model fitting and forecasts (Freckleton et al.,  
<sup>38</sup> 2002).

39 Research using long-term observational data has highlighted the role that evolutionary history  
40 may play in structuring plant phenological responses—which are critical to accurate forecasts  
41 of carbon storage. Phylogenetic signal in plant phenology, including dates of budburst, leafout  
42 and first flowering (Kochmer and Handel, 1986; Willis et al., 2008; Davies et al., 2013), suggests  
43 that more closely related species share more similar phenologies, likely reflecting evolutionary  
44 conservatism in responses to common cues. There are two broad explanations for why we might  
45 expect phylogenetic conservatism in phenological traits. First, close relatives will tend to share  
46 similar ecologies and physiologies, and thus be sensitive to similar environmental pressures.  
47 Second, close relatives derive from common geographic centers of origin, and thus their an-  
48 cestors will have been exposed to—and have adapted to—similar environmental cues (Davies  
49 et al., 2013). However, approaches using traditional phylogenetic comparative methods, have  
50 produced conflicting results, with some studies reporting evidence of phylogenetic structure in  
51 phenology-linked species declines (e.g., Willis et al., 2008) and in some phenophases, such as first  
52 flowering, but not others, such as peak or last flowering (e.g., CaraDonna and Inouye, 2014),  
53 and in responses to some cues but not others (e.g., Yang et al., 2021). In addition, evidence for  
54 phylogenetic conservatism of phenological responses appears to depend on method and species,  
55 even varying between sites with overlapping species sets (e.g., Rafferty and Nabity, 2017), which  
56 violates the fundamental idea of shared evolutionary history (the common ancestor of two sets of  
57 species cannot possess two separate evolutionary histories for the same trait). A first challenge is  
58 thus how to better integrate evolutionary history into multi-species models of plant phenological  
59 responses.

60 Generating robust ecological forecasts requires addressing a second major hurdle—underlying  
61 environmental cues that are complex and interacting. Decades of research have informed our  
62 understanding of how species use environmental cues to time their phenotypic responses with  
63 the temporal distribution of key resources and to avoid periods of high abiotic or biotic stress  
64 (Larcher, 1980; Bonamour et al., 2019). Commonly, however, responses to environmental cues,  
65 and their evolution, are studied individually, linking a given phenotypic response to a single cue,  
66 for example, time of leafout responding to summed heat during early spring (Davies et al., 2013).  
67 These efforts fail to capture the more likely scenario for most phenotypic traits in which multiple  
68 cues interacting along evolutionary history have shaped species' responses (Ackerly, 2009). For  
69 many plant species, phenological events are determined by a combination of temperature and  
70 light (Chuine and Regnieri, 2017), with additional factors (e.g., other cues—like humidity, or  
71 species physiology—vasculature or leaf structure) likely further mediating species responses.  
72 Although these mediating factors are not well understood (Chuine and Regnieri, 2017), they  
73 can be accounted for in models either as latent processes or by allowing non-stationarity in  
74 responses across species (Davies et al., 2019).

75 Spring plant phenology may represent our best opportunity to improve forecasts of species'  
76 responses to interacting environmental cues. Beyond being the most studied biological impact  
77 of climate change, the primary cue system is well established (Chuine and Regnieri, 2017),

especially for temperate woody species where phenology is generally thought to be determined by two components of temperature—chilling (cool temperatures during dormancy period over winter) and forcing (warm temperatures, generally in the spring)—and photoperiod (Ettinger et al., 2021). Plant phenology is also one of few phenotypic traits with extensive experimental data on responses to multiple environmental cues across species. Recent multi-species analyses considering forcing, chilling and photoperiod have shown that chilling and forcing together often determine complex non-linear responses to warming, but cannot forecast beyond several well-studied species (Ettinger et al., 2020).

Here we present a novel Bayesian framework that extends upon phylogenetic mixed models (Housworth et al., 2004) to examine how chilling, forcing (both metrics of temperature) and photoperiod together determine spring plant phenology. By allowing non-stationarity in species responses across the phylogeny (Davies et al., 2019), our model departs from previous work and assumptions of traditional phylogenetic comparative methods concerned with phylogenetic correction (e.g. Freckleton et al., 2002), and moves towards integrating evolutionary history in models of phenological responses to environmental change. To understand how evolution has shaped the cues underlying shifting phenology with climate change (Uyeda et al., 2017), we explicitly incorporate phylogenetic structure across model intercepts and slopes (that is, allowing a separate model of evolutionary history for chilling, forcing and photoperiod, see Methods & Materials for a complete description).

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

## Results & Discussion

Most species respond to all three primary cues—forcing, chilling, and photoperiod (Fig. 1, Supporting 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 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); 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

115 previous results suggesting chilling responses are much greater than forcing (Laube et al., 2014;  
116 Ettinger et al., 2020).

117 These average estimates, however, fail to capture the large differences in species' responses to  
118 both chilling and forcing (Fig. 1, Supporting Table S6). By allowing species responses to  
119 vary, based on a model including their shared evolutionary history, we found species differences  
120 dwarfed the mean differences between cues, especially temperature cues (Fig. 1). The largest cue  
121 in magnitude—chilling—varied 24-fold between species, while variation to forcing varied 7-fold.  
122 This variation indicates large differences between chilling and forcing occur at the species-level  
123 and not across species (as previously suggested Laube et al., 2014; Ettinger et al., 2020). These  
124 results highlight why robust phenological forecasts must account for both the complexity of  
125 multiple cues and species-level variation in responses to them.

126 *Differences across clades & cues*

127 The large differences across species produced striking differences between clades. For example,  
128 several groups—oaks and beeches (Fagaceae), elms (Ulmaceae) and buckthorns (Rhamnaceae)—  
129 are highly sensitive to chilling while others—rhododendrons (Ericaceae), butterfly bushes (Scro-  
130 phulariaceae) and spindles (Celastraceae)—show little to no response to chilling (Fig. 1a).  
131 Similar clade-level variation was observed for forcing, where some of these clades—e.g., Eri-  
132 caceae, Rhamnaceae, Ulmaceae, or Fagaceae—were particularly sensitive (advancing their bud-  
133 burst more than 10 days per standardized unit of forcing) and others, such as the Sapindaceae,  
134 Cornaceae or Juglandaceae, show little response (Fig. 1b).

135 Some species responded strongly to both temperature cues, which could suggest the existence  
136 of syndromes where the genetic basis for responses to one cue—e.g. forcing—has been selected  
137 for alongside responses to another cue—e.g. chilling. This could occur if selection  
138 operates jointly on responses to both cues; for example, if sensitivity to multiple cues provides  
139 greater insurance against leafing out before the last frost (Bonamour et al., 2019; Wolkovich and  
140 Donahue, 2021). Additionally, linkage or pleiotropism among loci associated with different cues  
141 (Nakagawa et al., 2005) could induce across-cue correlations. However, the correlation in species  
142 responses across cues was generally weak ( $r = 0.31$ ; between forcing and chilling) and some  
143 genera, such as *Tilia* and *Rhododendron* (Ericaceae), displayed strong responses to forcing but  
144 weak responses to chilling, while others, such as *Acer* (Sapindaceae), show moderately strong  
145 responses to chilling but weak responses to forcing (Fig. 1). Species sensitivity to one cue, thus,  
146 does not constrain sensitivity to another cue, and it seems selection can operate independently  
147 on responses to different cues (Bonamour et al., 2019).

148 In contrast to temperature cues (chilling and forcing), species-level responses to photoperiod were  
149 almost uniform across species. This consistency provides novel insight on a large debate over  
150 the prevalence of photoperiod cues in temperate trees, where previous experiments (Basler and  
151 Körner, 2012; Zohner et al., 2016) and models (e.g., Hunter and Lechowicz, 1992; Schaber and  
152 Badeck, 2003) suggested important variability across species that may constrain the responses

of certain species to warming (Way and Montgomery, 2015). Our results indicate variability is limited to a handful of species in Fagaceae, which have been particularly well studied, especially *Fagus sylvatica* (e.g., Basler and Körner, 2012; Zohner et al., 2016; Kramer et al., 2017). As *Fagus sylvatica* is nearly five times more sensitive to photoperiod than most other measured tree species, our results caution against using it to draw inferences of photoperiod responses more widely. These same few species are also where most evidence of local adaptation in photoperiod cues for spring phenology comes from (e.g., Kramer et al., 2017), in contrast with common garden studies of other species, which find little evidence of local adaptation in spring (but not fall) phenology (Aitken and Bemmels, 2016). 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 across species).

#### *Phylogenetic structure of phenological cues*

Variation—or lack thereof—in cues across species and clades provides possible insights into the evolution of cues across the phylogeny. While responses to each cue were phylogenetically structured, with closely related species exhibiting more similar sensitivities than distantly related species, the strength of phylogenetic conservatism in response 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, 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 (Davies et al., 2019). This fits fundamentally with the idea that early-season phenology plays a critical role in shaping species temporal niches (Gotelli and Graves, 1996) and thus should covary with a suite of life-history traits, including whether species are early-active with rapid return on investment traits, or start later in the season and have traits associated with higher competitive abilities (e.g., Grime, 1977; Wolkovich and Donahue, 2021).

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 (Wiens et al., 2010). While rapid local adaptation within species might erase the phylogenetic structure in photoperiod responses, it does not agree with the uniformity we find in species' responses. However, if responses to photoperiod evolved early in plants, as seems likely (Serrano-Bueno et al., 2017), and subsequent selection on photoperiod sensitivity was constrained by stabilizing selection operating on other life-history attributes sensitive to photoperiod (e.g., Rinne et al., 1994;

191 Wilczek et al., 2014; Azeez and Sane, 2015), we would predict both low interspecific variation  
192 and weak phylogenetic signal in responses, matching observations. This latter interpretation is  
193 also consistent with our estimates of lower  $\sigma$  for photoperiod responses (Fig. 2). Here, as in  
194 more traditional phylogenetic comparative methods,  $\sigma$  represents the rate of evolution, and thus  
195 our results suggest photoperiod responses are also evolving slower than temperature responses  
196 (see Fig. S12).

197 Phylogenetic conservatism (high  $\lambda$ ) and slow evolutionary rates (low  $\sigma$ ) in traits has sometimes  
198 been interpreted as indicative of evolutionary constraints to adaptive change (Wiens et al., 2010;  
199 Bennett et al., 2021). If this were the case, we might then conclude that species where responses  
200 are dominated by forcing cues might be more vulnerable to future warming because phylogenetic  
201 conservatism ( $\lambda$ ) in forcing is higher compared to other cues and its evolutionary rate ( $\sigma$ ) is lower  
202 than that estimated for chilling. This is misleading, however, as estimates of  $\lambda$  are independent  
203 from the rate of evolution, and macroevolutionary rates are estimated on phylogenetic trees that  
204 integrate across millions of years of evolutionary history, and thus do not necessarily inform us of  
205 maximum possible rates of evolution over much shorter timescales. Indeed, there is accumulating  
206 evidence for rapid evolution to shifting climates (Bradshaw and Holzapfel, 2006; Franks et al.,  
207 2014). Our estimates are thus more useful in providing unique insights into the evolutionary  
208 history of phenological cues, and emphasize the critical importance of incorporating species-level  
209 differences in ecological forecasts.

210 *Forecasting species-level responses*

211 Our results highlight that species-level variability can be extremely high—when properly esti-  
212 mated. Our approach, which partially pooled species responses based on their shared evolution-  
213 ary history, estimated substantially higher variation across species compared with more widely  
214 used hierarchical models. This was especially noticeable in temperature responses (for chilling  
215 variance across species means,  $var(\beta_{chill,j}$  from eqn. 2), was estimated as 23.55 in the phyloge-  
216 netic model, versus 17.47 in the non-phylogenetic model; variance across means,  $var(\beta_{force,j})$ :  
217 8.75 compared to 5.01) while photoperiod, which had low phylogenetic structure, was more  
218 similar across approaches (variance across means,  $var(\beta_{photo,j})$ : 0.83 compared to 0.64).

219 The consequences of including shared evolutionary history in forecasting are most apparent  
220 for poorly sampled species nested within more well-sampled clades. For example, forecasts  
221 for *Acer campestre*, which has only 6 observations, shift by up to 35% in the number of days  
222 until budburst after forcing starts, when comparing our phylogenetically informed model to one  
223 without phylogeny (see Fig. 4 for details). In contrast, forecasts for *Betula pendula*, which is one  
224 of the most sampled species, are nearly identical across models (Fig. 4). This occurs because  
225 cue estimates for *Acer campestre* in the phylogenetically informed model are strongly influenced  
226 by other *Acer* species, which diverge from other clades. In the non-phylogenetically informed  
227 model all species are equally exchangeable and thus *Acer campestre* is pulled strongly towards  
228 well-sampled species, such as *Betula pendula* ( $n = 311$ ), leading to forecasted shifts that are

229 more similar across all species (Fig. 4).

230 The increase in variability across species in our model with phylogenetic structure also decreased  
231 the uncertainty in estimates for each individual species temperature responses (Fig. S7). Thus,  
232 traditional (non-phylogenetically informed) approaches that partially pool across species (most  
233 hierarchical models in ecology, e.g., Flynn and Wolkovich, 2018; Ettinger et al., 2020) may also  
234 lead to less precise predictions and forecasts of phenology for individual species, although overall  
235 model accuracy might still appear reasonable (see Fig. S8). Another advantage of our Bayesian  
236 approach is that we are also better able to accommodate imprecision in the data that informs  
237 our model, which might arise from multiple sources, including measurement or experimental  
238 error, and the general stochasticity associated with limited sample sizes and unbalanced species  
239 representation. Critically, by partially pooling across species and weighting by phylogeny, we  
240 gain strength from species estimates that are informed by more data, such as within *Betula*  
241 and *Fagaceae*, but avoid skewing estimates for phylogenetically distant clade that may have  
242 been exposed to different selective regimes. Robustness of species stimates was confirmed by  
243 cross-validation analyses showing how PMM coefficients were more stable to clade removal than  
244 HMM coefficients (Fig. S4) and how PMM predictions were more strongly related to observed  
245 values for left out clades (Fig. S3; see “Leave-One-Clade-Out model cross validation” section in  
246 Supporting Information).

247 The contrasts between temperature and photoperiod responses—in both their variability across  
248 species and phylogenetic structure—have important implications for generating multi-species  
249 forecasts. Notably, responses to photoperiod appear weaker, more uniform across species, and  
250 less phylogenetically constrained compared to temperature. For temperature responses, the  
251 large variability among species makes predicting species-level responses challenging, but the  
252 phylogenetic structure in responses lets us borrow information from close relatives to improve  
253 our predictions. However, given that Brownian motion (our assumed model of evolution) is an  
254 extremely noisy process, we recommend imputation only for missing taxa that are closely related  
255 to other well sampled species or clades (Molina-Venegas et al., 2018, 2023).

256 While we focused on spring phenology here, our approach suggests a path forward for more  
257 general forecasting of species-level climate change responses. Our results show how including  
258 the phylogenetic relationship of species in a mechanistic model of underlying cues can overcome  
259 major limitations of most current hierarchical models—correcting biased model estimates, es-  
260 timating the full variability across species and reducing uncertainty around individual species  
261 estimates—while at once providing insight into the evolutionary history of biological responses.  
262 Using this approach improved forecasts of phenological responses to climate change and could  
263 help anticipate impacts on critical ecosystem services from species-level shifts and thus aid mit-  
264 igation and human adaption to warming.

265 **Methods & Materials**

266 **Phenological and Phylogenetic Data**

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

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

275 2. TOPIC = (budburst OR leaf-out) AND dorman\*, which yielded 270 publications

276 We scraped data from all papers of woody species that tested for photoperiod and/or tempera-  
277 ture effects on budburst, leafout, or flowering, resulting in data from 155 experiments across 97  
278 papers in the updated database. Ettinger et al. (2020) used a portion (72 experiments across  
279 49 papers) of the earlier OSPREE database and provides extensive methods on the database  
280 creation and cleaning. For our analysis here, we focused on angiosperms (as gymnosperms are  
281 very poorly represented in spring phenology experiments), and included all budburst experi-  
282 ments where we could quantify chilling, forcing and photoperiod levels, resulting in 44 studies  
283 from 33 papers, resulting in 2940 data points. In our dataset most studies come from Europe  
284 ( $n=37$ ) and a few from North America ( $n=7$ ). The same bias towards Europe is found across  
285 the full OSPREE dataset with less North American ( $n=19$ ) than European ( $n=60$ ) studies and  
286 only 3 studies located in the Southern Hemisphere. Given our need of daily gridded data for  
287 chilling we only include studies from Europe and North America, with most of these sites in  
288 temperate areas and a few in European Mediterranean areas (see Map in Supporting Fig. S1).  
289 Our final dataset is both geographically and taxonomically limited, mirroring existing literature  
290 on phenological experiments. Enlarging this body of literature will be critical to address new  
291 exciting research questions such as whether variability in cue sensitivity across species is larger  
292 in temperate than in tropical latitudes.

293 Across experiments chilling treatments were often fully or partially applied in the field, thus  
294 we estimated chilling—both in the field and applied in controlled environments—ourselves, es-  
295 timating Utah units (unless given) with the `chillR` package. For field chilling we used daily  
296 temperature data (converted to hourly) from both European (E-OBS, version 16, calculating the  
297 average of minimum and maximum daily temps, Cornes et al., 2018) and North American (v3,  
298 Sheffield et al., 2006) gridded climate datasets and the date given for when samples were taken  
299 from the field. We also converted experimental chilling into Utah chill units, based on reported  
300 treatments (for studies with a mix of field and experimental treatment, we added field and ex-

301 perimentally applied Utah units). To avoid numerical instability in our models (from having  
302 predictor values on very different scales), we divided Utah units by 240 (roughly equivalent to  
303 10 days of average chilling).

304 We report the Utah model because a small number of studies reported chilling only in Utah  
305 units, thus using this common metric allowed us to include the most data. The Utah model  
306 relies on the assumption that temperatures between 1.4 and 15.9 affect endodormancy release  
307 differently, though recent findings show possibly similar effects for a wide range of temperatures  
308 (-2 to 10; Baumgarten et al. (2021)). Because chilling is a latent process an accurate model of  
309 it (especially for the 191 species in our dataset) is not currently possible (Ettinger et al., 2020).  
310 However, we found consistent results using another common model of chilling—chill portions,  
311 suggesting our results are robust to the exact chilling metric used (see Tables S4 and S5 in  
312 Supporting Information).

313 Forcing and photoperiod treatments occurred after chilling treatments; we report photoperiod as  
314 the length of light and weighted these treatments by the reported photo- and thermo-periodicity  
315 (Buonaiuto et al., 2023). Most studies reported two temperatures per day across the whole  
316 experiment, one for day and night, but some had ramped temperatures and/or photoperiods (or  
317 other complexities). In these cases we built an hourly model of the full treatment period until  
318 budburst and took the mean value.

319 For a phylogenetic tree, we pruned the megatree for seed plants (Smith and Brown, 2018)  
320 to extract the sub-tree containing only the species present in the OSPREE dataset, species  
321 not included in the megatree were added to the congeneric basal node age (using the function  
322 ‘congeneric.merge’ in Pearse et al., 2015), and assigned branch lengths to maintain tree ultra-  
323 metricity. This addition of species to the tree can introduce polytomies (multifurcations) when  
324 many species are added to the same ancestral node, as was the case for *Acer*, for which several  
325 species where included in the OSPREE dataset but the megatree lacked species-level resolution  
326 within the genus. In total, our pruned tree had 8 polytomies affecting 46 out of 191 species.  
327 Our Bayesian hierarchical model, described below, is informed by the phylogenetic structure de-  
328 scribing species evolutionary relationships. Errors in phylogenetic topology and branching times  
329 could thus impact model estimates, although if errors were large the contribution of phylogeny  
330 would simply be scaled to zero. To assess whether the inclusion of polytomies in our data biased  
331 model estimates, we ran sensitivity analyses excluding these species from models (see Table S8  
332 in Supporting Information). Our approach assumes a tree where branch lengths represent time,  
333 but it could be possible to change this assumption. For example, if the genes underlying plant  
334 responses to particular cues were known, branch lengths would directly represent mutational  
335 changes along gene sequences. In the absence of such detailed gene specific data, evolutionary  
336 time provides a useful proxy for species differences.

337

338 **Bayesian hierarchical phylogenetic model**

339 Commonly used phylogenetic regression methods today (e.g., phylogenetic generalized least  
340 squares models, PGLS, Freckleton et al. (2002); phylogenetic mixed models, PMM, Hous-  
341 worth et al. (2004)) were originally conceived as statistical corrections for phylogenetic non-  
342 independence across observations—generally species—thus allowing multi-species studies to meet  
343 the assumptions of linear regression (Freckleton et al., 2002). These corrections incorporated  
344 phylogenetic structure by estimating the magnitude of a transformation of a variance-covariance  
345 (VCV) matrix whose elements were derived from the amount of evolutionary history (branch  
346 lengths) shared between species on a phylogeny. The most commonly used transformation was  
347 Pagel’s  $\lambda$ —a multiplier of the off-diagonal elements—where estimates of  $\lambda = 1$  essentially left the  
348 VCV untransformed and suggested that the residuals of the regression had phylogenetic signal  
349 consistent with Brownian motion; estimates of  $\lambda = 0$  suggested no phylogenetic signal. Be-  
350 cause the original aim of these methods was to correct for statistical bias introduced by shared  
351 evolutionary history among species, the underlying assumption of phylogenetic regressions is  
352 that phylogenetic relatedness would only affect either model residuals (in PGLS approaches,  
353 Freckleton et al., 2002) or model intercepts (e.g., in many PMM approaches, Housworth et al.,  
354 2004).

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

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

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

where

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

362 and  $\sigma_e^2$  represents random error unrelated to the phylogeny.

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

$$\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_n]^T \text{ such that } \boldsymbol{\alpha} \sim \mathcal{N}(\mu_{\alpha}, \boldsymbol{\Sigma}_{\alpha}) \quad (3)$$

$$\boldsymbol{\beta}_{chill} = [\beta_{chill,1}, \dots, \beta_{chill,n}]^T \text{ such that } \boldsymbol{\beta}_{chill} \sim \mathcal{N}(\mu_{\beta_{chill}}, \boldsymbol{\Sigma}_{\beta_{chill}})$$

$$\boldsymbol{\beta}_{force} = [\beta_{force,1}, \dots, \beta_{force,n}]^T \text{ such that } \boldsymbol{\beta}_{force} \sim \mathcal{N}(\mu_{\beta_{force}}, \boldsymbol{\Sigma}_{\beta_{force}})$$

$$\boldsymbol{\beta}_{photo} = [\beta_{photo,1}, \dots, \beta_{photo,n}]^T \text{ such that } \boldsymbol{\beta}_{photo} \sim \mathcal{N}(\mu_{\beta_{photo}}, \boldsymbol{\Sigma}_{\beta_{photo}})$$

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

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

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

where the residual phylogenetic error terms (e.g.,  $e_{\alpha_j}$ ) are elements of normal random vectors from multivariate normal distributions centered on 0 with the same phylogenetic variance-covariance matrices as in equation 4. 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., Ettinger et al. (2020)), 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  $\lambda$  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; however, many species occurred in only one study and location in our dataset, making it difficult to separate these effects, thus we do not include study as a separate parameter here (and thus average over it in our model estimates).

## References

- Ackerly, D. 2009. Conservatism and diversification of plant functional traits: Evolutionary rates versus phylogenetic signal. *Proceedings of the National Academy of Sciences of the United States of America* 106:19699–19706.
- Aitken, S. N., and J. B. Bemmels. 2016. Time to get moving: assisted gene flow of forest trees. *Evol Appl* 9:271–90.
- Azeez, A., and A. P. Sane. 2015. Photoperiodic growth control in perennial trees. *Plant signaling & behavior* 10:e1087631.
- Basler, D., and C. Körner. 2012. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology* 165:73–81.
- Baumgarten, F., C. M. Zohner, A. Gessler, and Y. Vitasse. 2021. Chilled to be forced: the best dose to wake up buds from winter dormancy. *New Phytologist* 230:1366–1377.
- Bennett, J. M., J. Sunday, P. Calosi, F. Villalobos, B. Martínez, R. Molina-Venegas, M. B. Araújo, A. C. Algar, S. Clusella-Trullas, B. A. Hawkins, et al. 2021. The evolution of critical thermal limits of life on earth. *Nature communications* 12:1–9.
- Bonamour, S., L. M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in response to climate change: the importance of cue variation. *Philosophical Transactions of the Royal Society B-Biological Sciences* 374.
- Bradshaw, W. E., and C. M. Holzapfel. 2006. Evolutionary response to rapid climate change. *Science* 312:1477–1478.
- Buonaiuto, D. M., M. Donahue, and E. M. Wolkovich. 2023. Experimental designs for testing the interactive effects of temperature and light in ecology: the problem of periodicity. *Functional Ecology* .
- CaraDonna, P. J., and D. W. Inouye. 2014. Phenological responses to climate change do not exhibit phylogenetic signal in a subalpine plant community. *Ecology* 96:355–361.
- Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics* 48:159–182.

- 419 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble  
420 version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical*  
421 *Research: Atmospheres* 123:9391–9409.
- 422 Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism  
423 in plant phenology. *Journal of Ecology* 101:1520–1530.
- 424 Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted  
425 regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology*  
426 and biogeography 28:275–285.
- 427 Dietze, M. 2017. Ecological forecasting. In *Ecological Forecasting*. Princeton University Press.
- 428 Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins,  
429 C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to  
430 community patterns. *Ecology letters* 15:545–553.
- 431 Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha,  
432 and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses  
433 to warming. *Nature Climate Change* pages 1–6.
- 434 Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich.  
435 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist*  
436 230:462–474.
- 437 Flynn, D. F. B., and E. M. Wolkovich. 2018. Temperature and photoperiod drive spring phe-  
438 nology across all species in a temperate forest community. *New Phytologist* 219:1353–1362.
- 439 Franks, S. J., J. J. Weber, and S. N. Aitken. 2014. Evolutionary and plastic responses to climate  
440 change in terrestrial plant populations. *Evolutionary Applications* 7:123–139.
- 441 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative  
442 data: a test and review of evidence. *The American Naturalist* 160:712–726.
- 443 Fuccillo Battle, K., A. Duhon, C. R. Vispo, T. M. Crimmins, T. N. Rosenstiel, L. L. Armstrong-  
444 Davies, and C. E. de Rivera. 2022. Citizen science across two centuries reveals phenological  
445 change among plant species and functional groups in the northeastern us. *Journal of Ecology*  
446 110:1757–1774.
- 447 Gelman, A., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models.  
448 Cambridge University Press.
- 449 Gotelli, N. J., and G. R. Graves. 1996. The temporal niche. Pages 95–111 in N. J. Gotelli and  
450 G. R. Graves, eds. *Null Models in Ecology*. Smithsonian Institution, Washington, DC.

- 451 Griffith, D. M., C. P. Osborne, E. J. Edwards, S. Bachle, D. J. Beerling, W. J. Bond, T. J. Gal-  
452 laher, B. R. Helliker, C. E. R. Lehmann, L. Leatherman, J. B. Nippert, S. Pau, F. Qiu, W. J.  
453 Riley, M. D. Smith, C. A. E. Stromberg, L. Taylor, M. Ungerer, and C. J. Still. 2020. Lineage-  
454 based functional types: characterising functional diversity to enhance the representation of  
455 ecological behaviour in land surface models. *New Phytologist* 228:15–23.
- 456 Grime, J. P. 1977. Evidence for existence of 3 primary strategies in plants and its relevance to  
457 ecological and evolutionary theory. *American Naturalist* 111:1169–1194.
- 458 Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The  
459 American Naturalist* 163:84–96.
- 460 Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of budburst in temperate trees.  
461 *Journal of Applied Ecology* 29:597–604.
- 462 IPCC. 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge Uni-  
463 versity Press, Cambridge, United Kingdom and New York, NY, USA.
- 464 Kochmer, J. P., and S. N. Handel. 1986. Constraints and competition in the evolution of flowering  
465 phenology. *Ecological monographs* 56:303–325.
- 466 Kramer, K., A. Ducousoo, D. Gomory, J. K. Hansen, L. Ionita, M. Liesebach, A. Lorent,  
467 S. Schueler, M. Sulkowska, S. de Vries, and G. von Wuehlisch. 2017. Chilling and forc-  
468 ing requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between  
469 provenances and are phenotypically plastic. *Agricultural and Forest Meteorology* 234:172–181.
- 470 Larcher, W. 1980. Plant Physiological Ecology. Springer-Verlag.
- 471 Laube, J., T. H. Sparks, N. Estrella, J. Höfler, D. P. Ankerst, and A. Menzel. 2014. Chilling  
472 outweighs photoperiod in preventing precocious spring development. *Global Change Biology*  
473 20:170–182.
- 474 Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins,  
475 M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, et al. 2022. The power of forecasts to  
476 advance ecological theory. *Methods in Ecology and Evolution* .
- 477 Molina-Venegas, R., I. Morales-Castilla, and M. Á. Rodríguez. 2023. Unreliable prediction of  
478 b-vitamin source species. *Nature Plants* 9:31–33.
- 479 Molina-Venegas, R., J. C. Moreno-Saiz, I. Castro Parga, T. J. Davies, P. R. Peres-Neto, and  
480 M. Á. Rodríguez. 2018. Assessing among-lineage variability in phylogenetic imputation of  
481 functional trait datasets. *Ecography* 41:1740–1749.
- 482 Moorcroft, P., G. Hurtt, and S. Pacala. 2001. A method for scaling vegetation dynamics: The  
483 ecosystem demography model (ed). *Ecological Monographs* 71:557–585.

- 484 Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005.  
485 Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological  
486 model. *Theoretical and Applied Genetics* 110:778–786.
- 487 Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts  
488 across natural systems. *Nature* 421:37.
- 489 Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker,  
490 and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. *Bioinformatics*  
491 31:2888–2890.
- 492 Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering  
493 time under climate change. *Journal of Ecology* 105:627–633.
- 494 Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey. 2013.  
495 Climate change, phenology, and phenological control of vegetation feedbacks to the climate  
496 system. *Agricultural and Forest Meteorology* 169:156–173.
- 497 Rinne, P., A. Saarelainen, and O. Junntila. 1994. Growth cessation and bud dormancy in  
498 relation to ABA level in seedlings and coppice shoots of *Betula pubescens* as affected by a  
499 short photoperiod, water stress and chilling. *Physiologia Plantarum* 90:451–458.
- 500 Schaber, J., and F. Badeck. 2003. Physiology-based phenology models for forest tree species in  
501 germany. *International Journal of Biometeorology* 47:193–201. International Conference on  
502 the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- 503 Serrano-Bueno, G., F. J. Romero-Campero, E. Lucas-Reina, J. M. Romero, and F. Valverde.  
504 2017. Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology*  
505 37:10–17.
- 506 Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global  
507 dataset of meteorological forcings for land surface modeling. *Journal of Climate* 19:3088–3111.
- 508 Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny.  
509 *American journal of botany* 105:302–314.
- 510 Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of  
511 energetic scaling across the vertebrate tree of life. *The American Naturalist* 190:185–199.
- 512 Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, perfor-  
513 mance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- 514 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.  
515 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism  
516 as an emerging principle in ecology and conservation biology. *Ecology letters* 13:1310–1324.

- 517 Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to  
518 warming climate in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences  
519 111:7906–7913.
- 520 Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phyloge-  
521 netic patterns of species loss in thoreau’s woods are driven by climate change. Proceedings of  
522 the National Academy of Sciences 105:17029–17033.
- 523 Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and  
524 communities in non-stationary environments. Biological Reviews .
- 525 Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and  
526 J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-  
527 PREE). doi:10.5063/F1CZ35KB.
- 528 Yang, Z., Y. Du, M. Shen, N. Jiang, E. Liang, W. Zhu, Y. Wang, and W. Zhao. 2021. Phy-  
529 logenetic conservatism in heat requirement of leaf-out phenology, rather than temperature  
530 sensitivity, in tibetan plateau. Agricultural and Forest Meteorology 304.
- 531 Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to  
532 constrain climate-driven shifts in leaf-out times of northern woody plants. Nature Climate  
533 Change 6:1120–1123.

534 **Data and code availability statement**

535 All data supporting the findings of this study are publicly available. Data analyses were per-  
536 formed in R and Bayesian Hierarchical and Phylogenetic Mixed Models (HMM and PMM,  
537 respectively) were performed within the Stan environment. All custom code will be made avail-  
538 able through a GitHub repository for this project (see Supporting Information for details and  
539 online links).

540 **Tables and Figures**

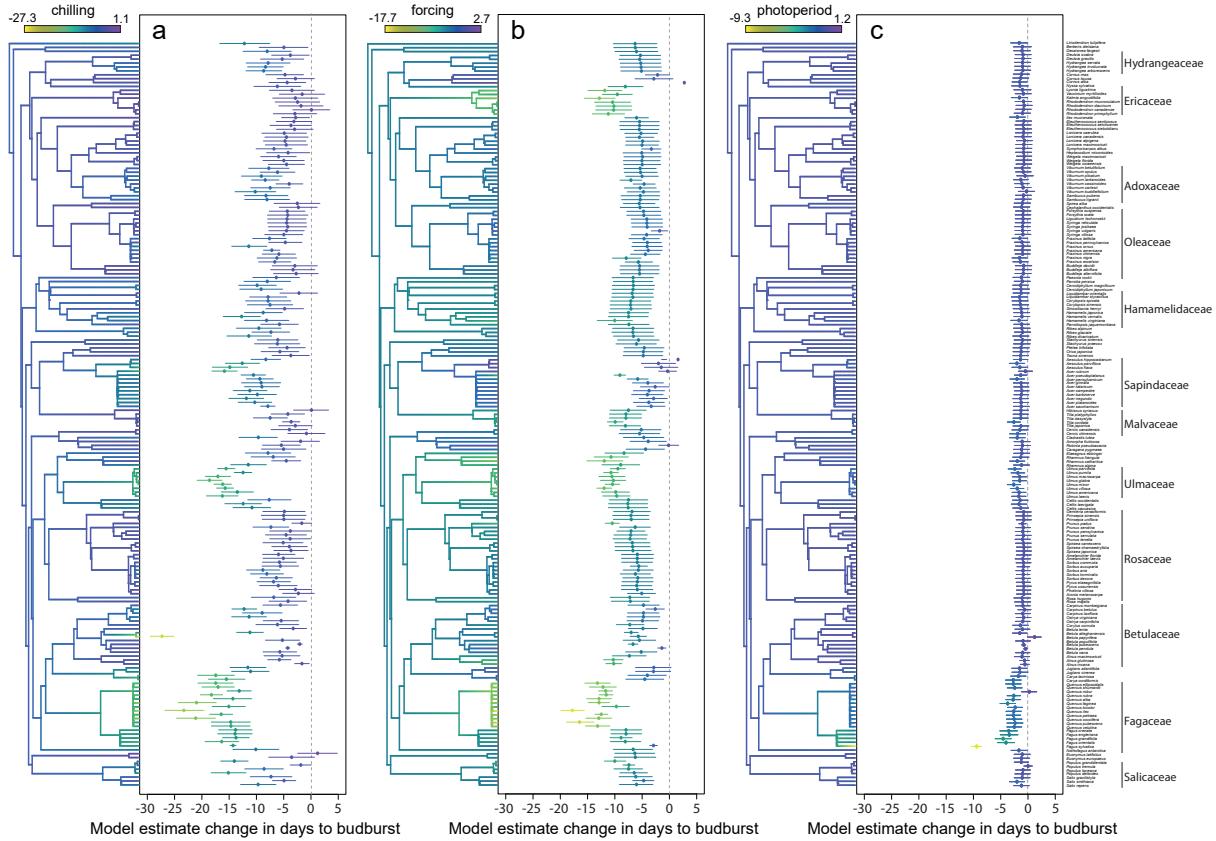


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit ( $z$ -transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored according to an estimation of ancestral character states, being the states at the tips the species' sensitivities to a cue, as estimated by our hierarchical phylogenetic model. Species sensitivities are shown along with 50% uncertainty intervals in the diagrams. Note that the color scale varies in each panel. Total tree depth is 81 My.

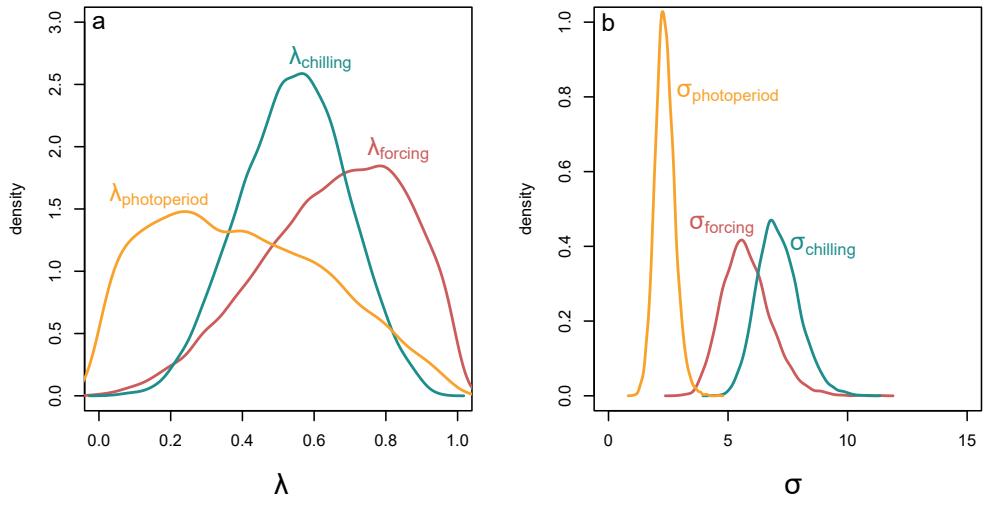


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters  $\lambda$  and  $\sigma$  estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to  $\lambda$  (a) and  $\sigma$  (b) from the phylogenetic model.

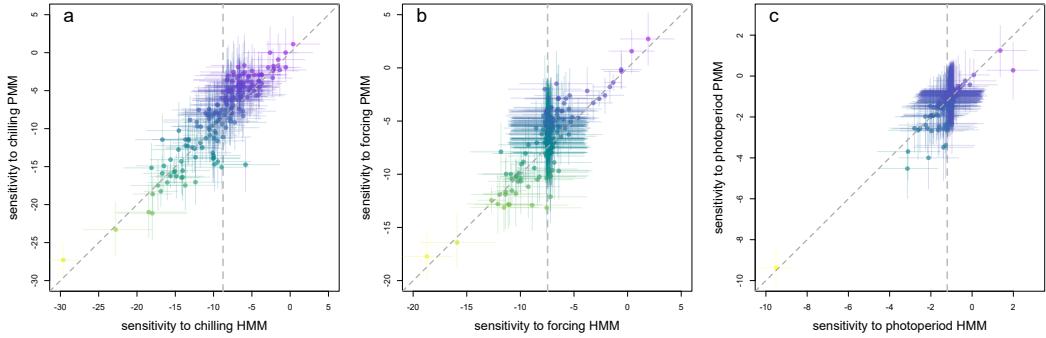


Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue (phylogenetic mixed model, PMM;  $y$ -axis), and the more commonly used hierarchical mixed model (HMM) where species are exchangeable (where  $\lambda$  is constrained to 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. Estimate colors are in the same scale as in 2, and error lines correspond to 50% uncertainty intervals.

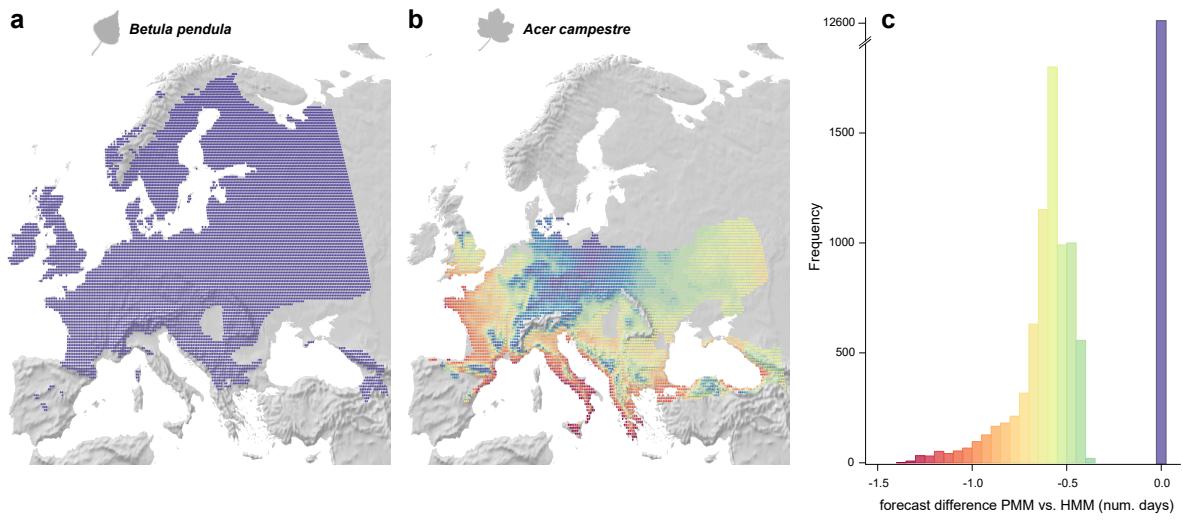


Figure 4: Comparison of forecasts of phenological shifts (i.e., computed as the difference between predictions under current climate vs. a  $2^{\circ}\text{C}$  warmer climate) resulting from a phylogenetic (PMM) and a non-phylogenetic (HMM) approach. Differences in forecasted shifts are negligible for 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 Material for details on forecast calculation.