

# Phylogenetic estimates of species-level phenology improve ecological forecasting

June 2, 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> Imperial College, UK

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

## Abstract

Adaptation to ongoing climate change hinges on accurate ecological forecasting to predict shifts in key ecosystem services, such as carbon storage and biodiversity maintenance. 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 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 responses (IPCC, 2014). Some  
<sup>11</sup> of this variability results from the complexity of climate change itself, including regional and  
<sup>12</sup> seasonal variation in warming that underlies average trends alongside shifts in other climate axes  
<sup>13</sup> (e.g. precipitation). Much of it, however, is driven by species-specific variation, reflecting evolved  
<sup>14</sup> differences in species' sensitivities to underlying environmental cues and their interactions, which  
<sup>15</sup> we know for only a few well-studied species (Chuine and Regniere, 2017; Ettinger et al., 2020).  
<sup>16</sup> In the absence of detailed data on individual species, species groupings (e.g., functional groups)  
<sup>17</sup> have improved ecosystem models (Moorcroft et al., 2001; Griffith et al., 2020), but still capture  
<sup>18</sup> only a fraction of the important variability (Fuccillo Battle et al., 2022). Improving forecasts,  
<sup>19</sup> will require models that accurately predict species-level differences in responses to complex  
<sup>20</sup> environmental change.

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

<sup>30</sup> Including the evolutionary history of species relationships in models of species responses could  
<sup>31</sup> at once provide more robust species-level estimates than current approaches and a better un-  
<sup>32</sup> derstanding of the evolutionary constraints that might limit future adaptation to change. For  
<sup>33</sup> example, strong phylogenetic niche conservatism (Wiens et al., 2010) could potentially inhibit  
<sup>34</sup> adaptive responses by drawing species back to an evolutionary conserved optimum, which is  
<sup>35</sup> sub-optimal under new conditions.

<sup>36</sup> Research using long-term observational data has especially highlighted the role that evolution-  
<sup>37</sup> ary history may play in structuring plant phenological responses—which are critical to accurate

38 forecasts of carbon storage. Phylogenetic signal in plant phenology (Kochmer and Handel, 1986;  
39 Willis et al., 2008; Davies et al., 2013) suggests species responses to cues have diverged over  
40 macro-evolutionary timescales, helping explain species present day differences. However, ap-  
41 proaches using traditional phylogenetic comparative methods, have produced conflicting results  
42 (e.g., Willis et al., 2008; CaraDonna and Inouye, 2014; Yang et al., 2021). In addition, evi-  
43 dence for phylogenetic conservatism of phenological responses appears to depend on method  
44 and species, even varying between sites with overlapping species sets (e.g., Rafferty and Nabity,  
45 2017), which violates the fundamental idea of shared evolutionary history (the common ancestor  
46 of two sets of species cannot possess two separate evolutionary histories for the same trait). A  
47 first challenge is thus how to better integrate evolutionary history into multi-species models of  
48 plant phenological responses.

49 Generating robust ecological forecasts requires addressing a second major hurdle—underlying  
50 environmental cues that are complex and interacting. Decades of research have informed our  
51 understanding of how species use environmental cues to time their phenotypic responses with  
52 the temporal distribution of key resources and to avoid periods of high abiotic or biotic stress  
53 (Larcher, 1980; Bonamour et al., 2019). Commonly, however, responses to environmental cues,  
54 and their evolution, are studied individually, linking a given phenotypic response to a single cue,  
55 for example, time of leafout responding to summed heat during early spring (Davies et al., 2013).  
56 These efforts fail to capture the more likely scenario for most phenotypic traits in which multiple  
57 cues interacting along evolutionary history have shaped species' responses (Ackerly, 2009). For  
58 many plant species, phenological events are determined by a combination of temperature and  
59 light (Chuine and Regniere, 2017), with additional factors (e.g., other cues—like humidity, or  
60 species physiology—vasculature or SLA) likely further mediating species responses. Although  
61 these mediating factors are often less well understood (Chuine and Regniere, 2017), they can be  
62 accounted for in models either as latent processes or by allowing non-stationarity in responses  
63 across species (Davies et al., 2019).

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

75 Here we present a novel Bayesian framework that extends upon phylogenetic mixed models  
76 (Housworth et al., 2004) to examine how chilling, forcing (both metrics of temperature) and

77 photoperiod together determine plant phenology. By allowing non-stationarity in species re-  
78 sponses across the phylogeny (Davies et al., 2019), our model departs from most previous work  
79 and assumptions of traditional phylogenetic comparative methods (e.g. Freckleton et al., 2002;  
80 Ives and Helmus, 2011; Hadfield, 2010), and moves towards integrating evolutionary history in  
81 models of phenological responses to environmental change. To understand how evolution has  
82 shaped the cues underlying shifting phenology with climate change (Uyeda et al., 2017), we ex-  
83 plicitly incorporate phylogenetic structure across model intercepts and slopes (that is, allowing  
84 a separate model of evolutionary history for chilling, forcing and photoperiod, see full Methods  
85 & Materials for a complete description).

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

## 96 Results & Discussion

97 Most species respond to all three primary cues—forcing, chilling, and photoperiod (Fig. 1,  
98 Supporting Table S1)—with responses to chilling approximately five-fold greater than to pho-  
99 toperiod (phenological advances of 6.9 days per standardized unit vs 1.2 days, for chilling and  
100 photoperiod, respectively; see Table S1). We estimated lower average responses to temperature  
101 compared to a model without phylogeny (model slopes for forcing and chilling decreased by 18%  
102 and 22%, respectively); responses to chilling and forcing were also more similar when including  
103 phylogeny (though chilling was still greater: 6.9 vs. 6.1 per standard unit), which contrasts with  
104 previous results suggesting chilling responses are much greater than forcing (Laube et al., 2014;  
105 Ettinger et al., 2020).

106 These average estimates, however, fail to capture the large differences in species' responses to  
107 both chilling and forcing (Fig. 1, Supporting Table S3). By allowing species responses to  
108 vary, based on a model including their shared evolutionary history, we found species differences  
109 dwarfed the mean differences between cues, especially temperature cues (Fig. 1). The largest cue  
110 in magnitude—chilling—varied 24-fold between species, while variation to forcing varied 7-fold.  
111 This variation indicates large differences between chilling and forcing occur at the species-level  
112 and not across species (as previously suggested Laube et al., 2014; Ettinger et al., 2020). These  
113 results highlight why robust phenological forecasts must account for both the complexity of

<sup>114</sup> multiple cues and species-level variation in responses to them.

<sup>115</sup> *Differences across clades & cues*

<sup>116</sup> The large differences across species produced striking differences between clades. For example,  
<sup>117</sup> several groups—oaks and beeches (Fagaceae), elms (Ulmaceae) and buckthorns (Rhamnaceae)—  
<sup>118</sup> are highly sensitive to chilling while others—rhododendrons (Ericaceae), butterfly bushes (Scro-  
<sup>119</sup> phulariaceae) and spindles (Celastraceae)—show little to no response to chilling (Fig. 1a).  
<sup>120</sup> Similar clade-level variation was observed for forcing, where some of these clades—e.g., Eri-  
<sup>121</sup> caceae, Rhamnaceae, Ulmaceae, or Fagaceae—were particularly sensitive (advancing their bud-  
<sup>122</sup> burst more than 10 days per standardized unit of forcing) and others, such as the Sapindaceae,  
<sup>123</sup> Cornaceae or Juglandaceae, show little response (Fig. 1b).

<sup>124</sup> Some species responded strongly to both temperature cues, which could suggest the existence  
<sup>125</sup> of syndromes where the genetic basis for responses to one cue—e.g. forcing—has been selected  
<sup>126</sup> for alongside responses to another cue—e.g. chilling. This could occur if selection  
<sup>127</sup> operates jointly on responses to both cues; for example, if sensitivity to multiple cues provides  
<sup>128</sup> greater insurance against leafing out before the last frost (Bonamour et al., 2019; Wolkovich  
<sup>129</sup> and Donahue, 2021). Additionally, linkage or pleiotropism among loci associated with different  
<sup>130</sup> cues (Nakagawa et al., 2005) could induce these correlations. However, the correlation in species  
<sup>131</sup> responses across cues was generally weak ( $r = 0.31$ ; between forcing and chilling) and some  
<sup>132</sup> genera, such as *Tilia* and *Rhododendron* (Ericaceae), displayed strong responses to forcing but  
<sup>133</sup> weak responses to chilling, while others, such as *Acer* (Sapindaceae), show moderately strong  
<sup>134</sup> responses to chilling but weak responses to forcing (Fig. 1). Species sensitivity to one cue, thus,  
<sup>135</sup> does not constrain sensitivity to another cue, and it seems selection can operate independently  
<sup>136</sup> on responses to different cues (Bonamour et al., 2019).

<sup>137</sup> In contrast to temperature cues (chilling and forcing), species-level responses to photoperiod were  
<sup>138</sup> almost uniform across species. This consistency provides novel insight on a large debate over  
<sup>139</sup> the prevalence of photoperiod cues in temperate trees, where previous experiments (Basler and  
<sup>140</sup> Körner, 2012; Zohner et al., 2016) and models (e.g., Hunter and Lechowicz, 1992; Schaber and  
<sup>141</sup> Badeck, 2003) suggested important variability across species that may constrain the responses  
<sup>142</sup> of certain species to warming (Way and Montgomery, 2015). Our results indicate variability is  
<sup>143</sup> limited to a handful of species in Fagaceae, which have been particularly well studied, especially  
<sup>144</sup> *Fagus sylvatica* (e.g., Basler and Körner, 2012; Zohner et al., 2016; Kramer et al., 2017). As  
<sup>145</sup> *Fagus sylvatica* is nearly five times more sensitive to photoperiod than most other measured tree  
<sup>146</sup> species, our results caution against using it to draw inferences of photoperiod responses more  
<sup>147</sup> widely. These same few species are also where most evidence of local adaptation in photoperiod  
<sup>148</sup> cues for spring phenology comes from (e.g., Kramer et al., 2017), in contrast with common  
<sup>149</sup> garden studies of other species, which found little evidence of local adaptation in spring (but not  
<sup>150</sup> fall) phenology (Aitken and Bemmels, 2016). The uniformity of response to photoperiod in our  
<sup>151</sup> 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 S1).

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. However, somewhat counterintuitively, both uniform and random responses can manifest as low phylogenetic signal when indexed by Brownian motion expectations (Wiens et al., 2010). While rapid local adaptation within species might erase the phylogenetic structure in photoperiod responses, it does not agree with the uniformity we find in species' responses. However, if responses to photoperiod evolved early in plants, as seems likely (Serrano-Bueno et al., 2017), and subsequent selection on photoperiod sensitivity was constrained by stabilizing selection operating on other life-history attributes sensitive to photoperiod (e.g., Rinne et al., 1994; Wilczek et al., 2014; Azeez and Sane, 2015), we would predict both low interspecific variation and weak phylogenetic signal in responses, matching observations. This latter interpretation is also consistent with our estimates of lower  $\sigma$  for photoperiod responses (Fig. 2). Here, as in more traditional phylogenetic comparative methods,  $\sigma$  represents the rate of evolution, and thus our results suggest photoperiod responses are also evolving slower than temperature responses (see Fig. S8).

Phylogenetic conservatism (high  $\lambda$ ) and slow evolutionary rates (low  $\sigma$ ) in traits has sometimes been interpreted as indicative of evolutionary constraints to adaptive change (Wiens et al., 2010; Bennett et al., 2021). If this were the case, we might then conclude that species where responses are dominated by forcing cues might be more vulnerable to future warming because phylogenetic conservatism ( $\lambda$ ) in forcing is higher compared to other cues and its evolutionary rate ( $\sigma$ ) is lower

190 than that estimated for chilling. This is misleading, however, as estimates of  $\lambda$  are independent  
191 from the rate of evolution, and macroevolutionary rates are estimated on phylogenetic trees that  
192 integrate across millions of years of evolutionary history, and thus do not necessarily inform us of  
193 maximum possible rates of evolution over much shorter timescales. Indeed, there is accumulating  
194 evidence for rapid evolution to shifting climates (Bradshaw and Holzapfel, 2006; Franks et al.,  
195 2014). Our estimates are thus more useful in providing unique insights into the evolutionary  
196 history of phenological cues, and emphasize the critical importance of incorporating species-level  
197 differences in ecological forecasts

198 *Forecasting species-level responses*

199 Our results highlight that species-level variability can be extremely high—when properly esti-  
200 mated. Our approach, which partially pooled species responses based on their shared evolution-  
201 ary history, estimated substantially higher variation across species compared with more widely  
202 used hierarchical models. This was especially noticeable in temperature responses (for chilling  
203 variance across species means,  $var(\beta_{chill,j}$  from eqn. 2), was estimated as 23.55 in the phyloge-  
204 netic model, versus 17.47 in the non-phylogenetic model; variance across means,  $var(\beta_{force,j})$ :  
205 8.75 compared to 5.01) while photoperiod, which had low phylogenetic structure, was more  
206 similar across approaches (variance across means,  $var(\beta_{photo,j})$ : 0.83 compared to 0.64).

207 The consequences of including shared evolutionary history in forecasting are most apparent for  
208 poorly sampled species nested within more well-sampled clades. For example, forecasts for *Acer*  
209 *campestre*, which has only 6 observations, shift by up to 35% in our phylogenetically informed  
210 model compared to one without phylogeny (Fig. 4). In contrast, forecasts for *Betula pendula*,  
211 which is one of the most sampled species, are nearly identical across models (Fig. 4). This occurs  
212 because cue estimates for *Acer campestre* in the phylogenetically informed model are strongly  
213 influenced by other *Acer* species, which diverge from other clades. In the non-phylogenetically  
214 informed model all species are equally exchangeable and thus *Acer campestre* is pulled strongly  
215 towards well-sampled species, such as *Betula pendula* ( $n = 311$ ), leading to forecasted shifts that  
216 are more similar across all species (Fig. 4).

217 The increase in variability across species in our model with phylogenetic structure also decreased  
218 the uncertainty in estimates for each individual species temperature responses (Fig. S3). Thus,  
219 traditional (non-phylogenetically informed) approaches that partially pool across species (most  
220 hierarchical models in ecology, e.g., Flynn and Wolkovich, 2018; Ettinger et al., 2020) may also  
221 lead to less precise predictions and forecasts of phenology for individual species, although overall  
222 model accuracy might still appear reasonable (see Fig. S4).

223 The contrasts between temperature and photoperiod responses—in both their variability across  
224 species and phylogenetic structure—have important implications for generating multi-species  
225 forecasts. Notably, responses to photoperiod appear weaker, more uniform across species, and  
226 less phylogenetically constrained compared to temperature. For temperature responses, the  
227 large variability among species makes predicting species-level responses challenging, but the

228 phylogenetic structure in responses lets us borrow information from close relatives to improve  
229 our predictions. However, given that Brownian motion (our assumed model of evolution) is an  
230 extremely noisy process, we recommend imputation only for missing taxa that are closely related  
231 to other well sampled species or clades (Molina-Venegas et al., 2018, 2023).

232 While we focused on spring phenology here, our approach suggests a path forward for more  
233 general forecasting of species-level climate change responses. Our results show how including  
234 the phylogenetic relationship of species in a mechanistic model of underlying cues can overcome  
235 major limitations of most current hierarchical models—correcting biased model estimates, prop-  
236 erly estimating the full variability across species and reducing uncertainty around individual  
237 species estimates—while at once providing insight into the evolutionary history of biological  
238 responses. Using this approach improved forecasts of phenological responses to climate change  
239 and could help anticipate impacts on critical ecosystem services from species-level shifts and  
240 thus aid mitigation and human adaption to warming.

241 **Methods & Materials**

242 **Phenological and Phylogenetic Data**

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

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

252 We scraped data from all papers of woody species that tested for photoperiod and/or tempera-  
253 ture effects on budburst, leafout, or flowering, resulting in data from 155 experiments across 97  
254 papers in the updated database. Ettinger et al. (2020) used a portion (72 experiments across  
255 49 papers) of the earlier OSPREE database and provides extensive methods on the database  
256 creation and cleaning. For our analysis here, we focused on angiosperms (as gymnosperms are  
257 very poorly represented in spring phenology experiments), and included all budburst experi-  
258 ments where we could quantify chilling, forcing and photoperiod levels, resulting in 44 studies  
259 from 33 papers.

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

267 For a phylogenetic tree, we pruned the phylogenetic megatree for seed plants (Smith and Brown,  
268 2018) to extract a subset of the phylogenetic tree containing only the angiosperm species in the  
269 OSPREE dataset, then added species that were not present in the megatree as polytomies at the  
270 generic level (using the function ‘congeneric.merge’ in Pearse et al., 2015)), with a branch length  
271 of zero. Polytomies represent 26.8% of the fulldataset. To test for the ability of polytomies to  
272 bias our results we run sensitivity analyses excluding these species from models (which lead to  
273 191 species; see Supporting Information).

274

275    **Bayesian hierarchical phylogenetic model**

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

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

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

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

where

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

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

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

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

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

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

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

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  $\Sigma_i$  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  $i$  (here assumed to be constant along all branches),  $\lambda_i$  scales branch lengths and therefore is a measure of the “phylogenetic signal” or extent of phylogenetic relatedness on each model parameter (i.e.,  $\alpha_j$ ,  $\beta_{force,j}$ ,  $\beta_{photo,j}$ ), 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 S1-S2).

## 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.

- 324 Bennett, J. M., J. Sunday, P. Calosi, F. Villalobos, B. Martínez, R. Molina-Venegas, M. B.  
325 Araújo, A. C. Algar, S. Clusella-Trullas, B. A. Hawkins, et al. 2021. The evolution of critical  
326 thermal limits of life on earth. *Nature communications* 12:1–9.
- 327 Bonamour, S., L. M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in  
328 response to climate change: the importance of cue variation. *Philosophical Transactions of  
329 the Royal Society B-Biological Sciences* 374.
- 330 Bradshaw, W. E., and C. M. Holzapfel. 2006. Evolutionary response to rapid climate change.  
331 *Science* 312:1477–1478.
- 332 CaraDonna, P. J., and D. W. Inouye. 2014. Phenological responses to climate change do not  
333 exhibit phylogenetic signal in a subalpine plant community. *Ecology* 96:355–361.
- 334 Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals.  
335 *Annual Review of Ecology, Evolution, and Systematics* 48:159–182.
- 336 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble  
337 version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical  
338 Research: Atmospheres* 123:9391–9409.
- 339 Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism  
340 in plant phenology. *Journal of Ecology* 101:1520–1530.
- 341 Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted  
342 regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology  
343 and biogeography* 28:275–285.
- 344 Dietze, M. 2017. Ecological forecasting. *In* *Ecological Forecasting*. Princeton University Press.
- 345 Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins,  
346 C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to  
347 community patterns. *Ecology letters* 15:545–553.
- 348 Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha,  
349 and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses  
350 to warming. *Nature Climate Change* pages 1–6.
- 351 Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich.  
352 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist*  
353 230:462–474.
- 354 Flynn, D. F. B., and E. M. Wolkovich. 2018. Temperature and photoperiod drive spring phe-  
355 nology across all species in a temperate forest community. *New Phytologist* 219:1353–1362.

- 356 Franks, S. J., J. J. Weber, and S. N. Aitken. 2014. Evolutionary and plastic responses to climate  
357 change in terrestrial plant populations. *Evolutionary Applications* 7:123–139.
- 358 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative  
359 data: a test and review of evidence. *The American Naturalist* 160:712–726.
- 360 Fuccillo Battle, K., A. Duhon, C. R. Vispo, T. M. Crimmins, T. N. Rosenstiel, L. L. Armstrong-  
361 Davies, and C. E. de Rivera. 2022. Citizen science across two centuries reveals phenological  
362 change among plant species and functional groups in the northeastern us. *Journal of Ecology*  
363 110:1757–1774.
- 364 Gelman, A., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models.  
365 Cambridge University Press.
- 366 Gotelli, N. J., and G. R. Graves. 1996. The temporal niche. Pages 95–111 in N. J. Gotelli and  
367 G. R. Graves, eds. *Null Models in Ecology*. Smithsonian Institution, Washington, DC.
- 368 Griffith, D. M., C. P. Osborne, E. J. Edwards, S. Bachle, D. J. Beerling, W. J. Bond, T. J. Gal-  
369 laher, B. R. Helliker, C. E. R. Lehmann, L. Leatherman, J. B. Nippert, S. Pau, F. Qiu, W. J.  
370 Riley, M. D. Smith, C. A. E. Stromberg, L. Taylor, M. Ungerer, and C. J. Still. 2020. Lineage-  
371 based functional types: characterising functional diversity to enhance the representation of  
372 ecological behaviour in land surface models. *New Phytologist* 228:15–23.
- 373 Grime, J. P. 1977. Evidence for existence of 3 primary strategies in plants and its relevance to  
374 ecological and evolutionary theory. *American Naturalist* 111:1169–1194.
- 375 Hadfield, J. D. 2010. Mcmc methods for multi-response generalized linear mixed models: the  
376 mcmcglmm r package. *Journal of statistical software* 33:1–22.
- 377 Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The*  
378 *American Naturalist* 163:84–96.
- 379 Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of budburst in temperate trees.  
380 *Journal of Applied Ecology* 29:597–604.
- 381 IPCC. 2014. *Climate Change 2014: Impacts, Adaptation, and Vulnerability*. Cambridge Uni-  
382 versity Press, Cambridge, United Kingdom and New York, NY, USA.
- 383 Ives, A. R., and M. R. Helmus. 2011. Generalized linear mixed models for phylogenetic analyses  
384 of community structure. *Ecological Monographs* 81:511–525.
- 385 Kochmer, J. P., and S. N. Handel. 1986. Constraints and competition in the evolution of flowering  
386 phenology. *Ecological monographs* 56:303–325.

- 387 Kramer, K., A. Ducousoo, D. Gomory, J. K. Hansen, L. Ionita, M. Liesebach, A. Lorent,  
388 S. Schueler, M. Sulkowska, S. de Vries, and G. von Wuehlisch. 2017. Chilling and forcing  
389 requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between  
390 provenances and are phenotypically plastic. Agricultural and Forest Meteorology 234:172–181.
- 391 Larcher, W. 1980. Plant Physiological Ecology. Springer-Verlag.
- 392 Laube, J., T. H. Sparks, N. Estrella, J. Höfler, D. P. Ankerst, and A. Menzel. 2014. Chilling  
393 outweighs photoperiod in preventing precocious spring development. Global Change Biology  
394 20:170–182.
- 395 Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins,  
396 M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, et al. 2022. The power of forecasts to  
397 advance ecological theory. Methods in Ecology and Evolution .
- 398 Molina-Venegas, R., I. Morales-Castilla, and M. Á. Rodríguez. 2023. Unreliable prediction of  
399 b-vitamin source species. Nature Plants 9:31–33.
- 400 Molina-Venegas, R., J. C. Moreno-Saiz, I. Castro Parga, T. J. Davies, P. R. Peres-Neto, and  
401 M. Á. Rodríguez. 2018. Assessing among-lineage variability in phylogenetic imputation of  
402 functional trait datasets. Ecography 41:1740–1749.
- 403 Moorcroft, P., G. Hurtt, and S. Pacala. 2001. A method for scaling vegetation dynamics: The  
404 ecosystem demography model (ed). Ecological Monographs 71:557–585.
- 405 Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005.  
406 Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological  
407 model. Theoretical and Applied Genetics 110:778–786.
- 408 Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts  
409 across natural systems. Nature 421:37.
- 410 Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker,  
411 and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. Bioinformatics  
412 31:2888–2890.
- 413 Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering  
414 time under climate change. Journal of Ecology 105:627–633.
- 415 Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey. 2013.  
416 Climate change, phenology, and phenological control of vegetation feedbacks to the climate  
417 system. Agricultural and Forest Meteorology 169:156–173.
- 418 Rinne, P., A. Saarelainen, and O. Juntila. 1994. Growth cessation and bud dormancy in  
419 relation to ABA level in seedlings and coppice shoots of *Betula pubescens* as affected by a  
420 short photoperiod, water stress and chilling. Physiologia Plantarum 90:451–458.

- 421 Schaber, J., and F. Badeck. 2003. Physiology-based phenology models for forest tree species in  
422 germany. International Journal of Biometeorology 47:193–201. International Conference on  
423 the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- 424 Serrano-Bueno, G., F. J. Romero-Campero, E. Lucas-Reina, J. M. Romero, and F. Valverde.  
425 2017. Evolution of photoperiod sensing in plants and algae. Current Opinion in Plant Biology  
426 37:10–17.
- 427 Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global  
428 dataset of meteorological forcings for land surface modeling. Journal of Climate 19:3088–3111.
- 429 Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny.  
430 American journal of botany 105:302–314.
- 431 Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of  
432 energetic scaling across the vertebrate tree of life. The American Naturalist 190:185–199.
- 433 Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, perfor-  
434 mance and migration in a warming world. Plant, Cell & Environment 38:1725–1736.
- 435 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.  
436 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism  
437 as an emerging principle in ecology and conservation biology. Ecology letters 13:1310–1324.
- 438 Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to  
439 warming climate in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences  
440 111:7906–7913.
- 441 Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phyloge-  
442 netic patterns of species loss in thoreau’s woods are driven by climate change. Proceedings of  
443 the National Academy of Sciences 105:17029–17033.
- 444 Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and  
445 communities in non-stationary environments. Biological Reviews .
- 446 Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and  
447 J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-  
448 PREE). doi:10.5063/F1CZ35KB.
- 449 Yang, Z., Y. Du, M. Shen, N. Jiang, E. Liang, W. Zhu, Y. Wang, and W. Zhao. 2021. Phy-  
450 logenetic conservatism in heat requirement of leaf-out phenology, rather than temperature  
451 sensitivity, in tibetan plateau. Agricultural and Forest Meteorology 304.
- 452 Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to  
453 constrain climate-driven shifts in leaf-out times of northern woody plants. Nature Climate  
454 Change 6:1120–1123.

<sup>455</sup> **Tables and Figures**

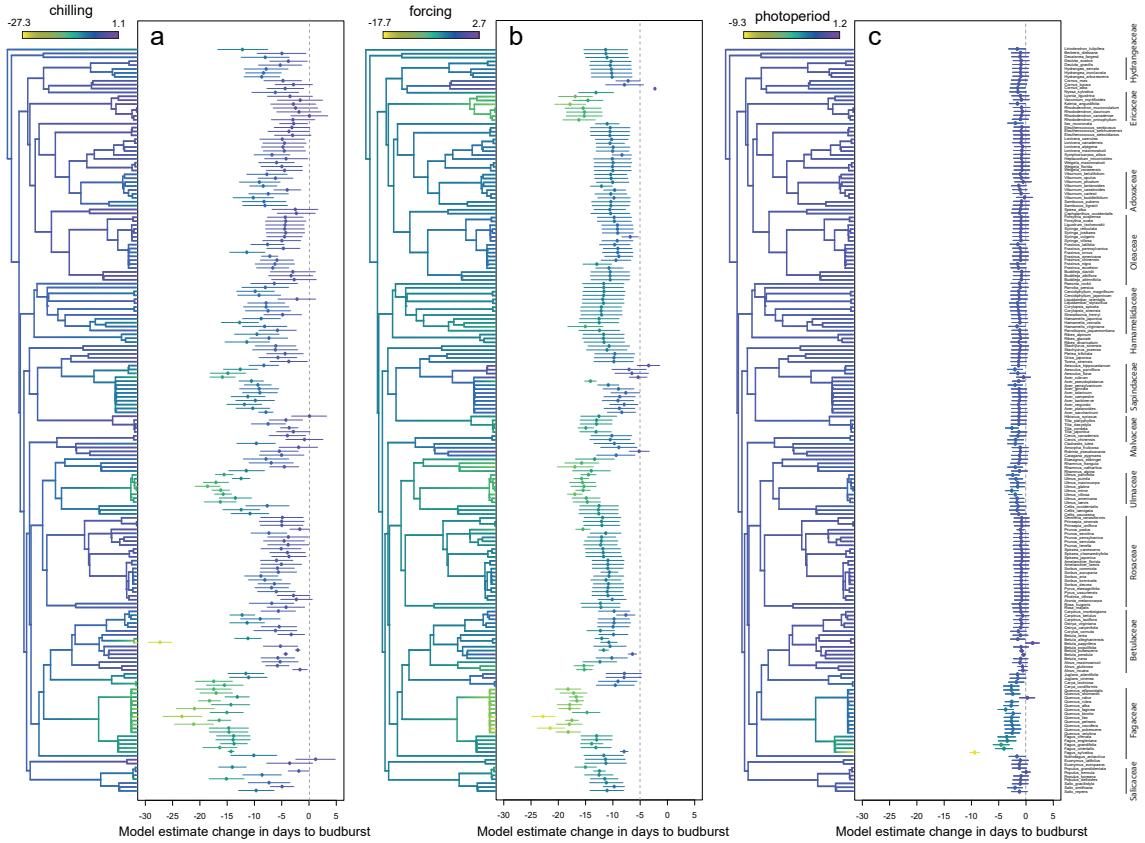


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit (z-transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored according to an 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.

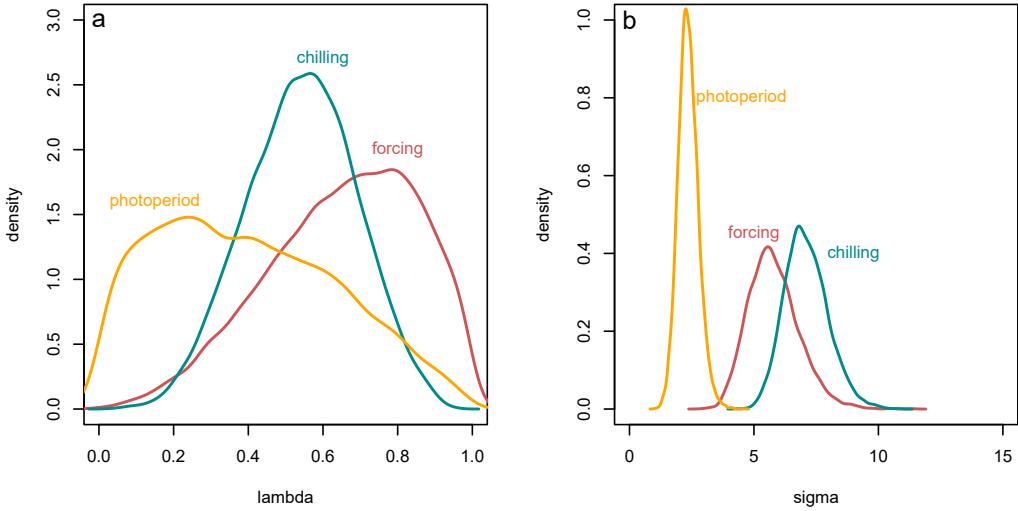


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters  $\lambda$  and  $\sigma$  estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to  $\lambda$  (a) and  $\sigma$  (b) from the phylogenetic model.

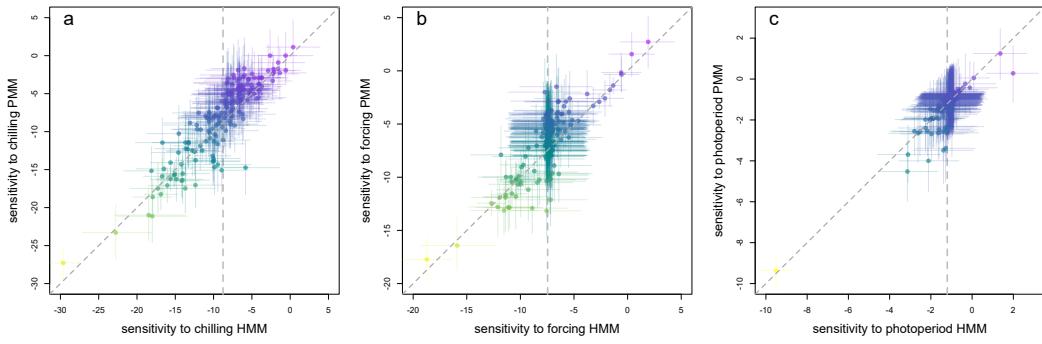


Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue ( $y$ -axis), and the more commonly used hierarchical model where species are exchangeable (where  $\lambda$  is constrained to be equal to zero,  $x$ -axis). While species 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.

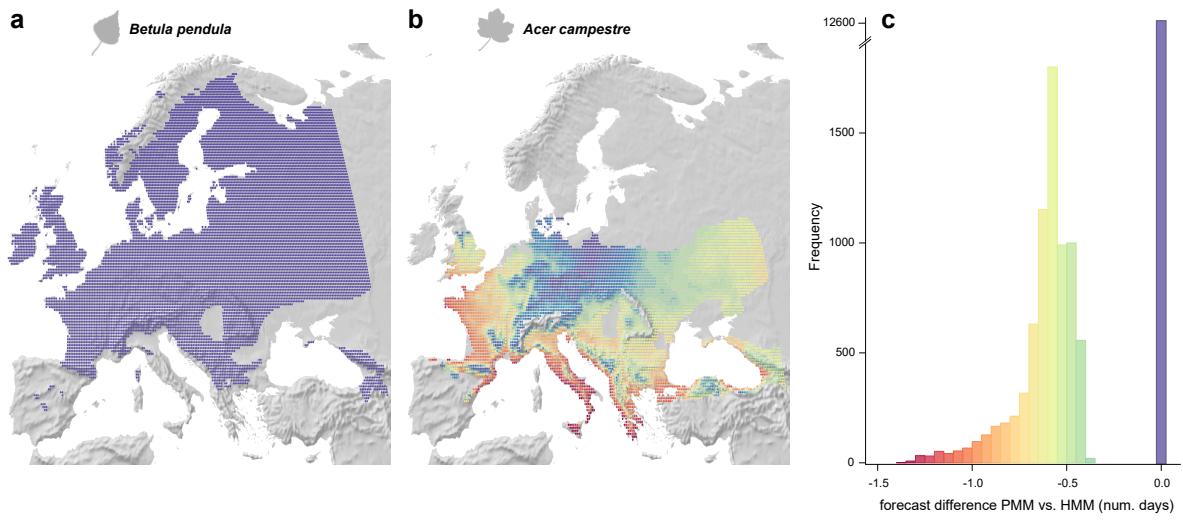


Figure 4: Comparison of forecasts of phenological shifts (i.e., computed as the difference between predictions under current climate vs. a  $2^{\circ}\text{C}$  warmer climate) resulting from a phylogenetic (PMM) and a non-phylogenetic (HMM) approach. Differences in forecasted shifts are negligible for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