

Phylogenetic estimates of species-level phenology improve ecological forecasting

June 1, 2023

Authors:

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

Author affiliations:

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

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

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

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

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

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

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

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

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

¹¹ 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.

¹ Introduction

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

⁹ While ecological forecasting has improved over recent years (Dietze, 2017; Lewis et al., 2022), it
¹⁰ remains a challenge to reproduce the high variability observed in responses (IPCC, 2014). Some
¹¹ of this variability results from the complexity of climate change itself, including regional and
¹² seasonal variation in warming that underlies average trends alongside shifts in other climate axes
¹³ (e.g. precipitation). Much of it, however, is driven by species-specific variation, reflecting evolved
¹⁴ differences in species' sensitivities to underlying environmental cues and their interactions, which
¹⁵ we know for only a few well-studied species (Chuine and Regniere, 2017; Ettinger et al., 2020).
¹⁶ In the absence of detailed data on individual species, species groupings (e.g., functional groups)
¹⁷ have improved ecosystem models (Moorcroft et al., 2001; Griffith et al., 2020), but still capture
¹⁸ only a fraction of the important variability (?). Improving forecasts, will require models that
¹⁹ accurately predict species-level differences in responses to complex environmental change.

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

²⁹ Including the evolutionary history of species relationships in models of species responses could
³⁰ at once provide more robust species-level estimates than current approaches and a better un-
³¹ derstanding of the evolutionary constraints that might limit future adaptation to change. For
³² example, strong phylogenetic niche conservatism (Wiens et al., 2010) could potentially inhibit
³³ adaptive responses by drawing species back to an evolutionary conserved optimum, which is
³⁴ sub-optimal under new conditions.

³⁵ Research using long-term observational data has especially highlighted the role that evolution-
³⁶ ary history may play in structuring plant phenological responses—which are critical to accurate
³⁷ forecasts of carbon storage. Phylogenetic signal in plant phenology (Kochmer and Handel, 1986;

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

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

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

74 Here we present a novel Bayesian framework that extends upon phylogenetic mixed models
75 (Housworth et al., 2004) to examine how chilling, forcing (both metrics of temperature) and
76 photoperiod together determine plant phenology. By allowing non-stationarity in species re-

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

⁸⁵ We illustrate our method with an unprecedented dataset on phenological responses to environ-
⁸⁶ mental cues (chilling, forcing and photoperiod) determined experimentally for 191 deciduous
⁸⁷ woody species (by far the most studied group of species in phenology experiments, see Ettinger
⁸⁸ et al., 2020), in an updated version of the Observed Spring Phenology Responses in Experi-
⁸⁹ mental Environments (OSPREE) database (Wolkovich et al., 2019). These data combined with the
⁹⁰ Smith and Brown (2018) megatree adjusted to our species and our modeling approach allows
⁹¹ us to address the common question of which cue has the largest effect on budburst and at the
⁹² same time provide robust estimates of how cues vary across species. Using spring phenology, we
⁹³ identify historical regime shifts (Uyeda et al., 2017) in phenological responses, and highlight how
⁹⁴ our approach could advance forecasting of other critical responses to ongoing global change.

⁹⁵ Results & Discussion

⁹⁶ Most species respond to all three primary cues—forcing, chilling, and photoperiod (Fig. 1,
⁹⁷ Supporting Table S1)—with responses to chilling approximately five-fold greater than to pho-
⁹⁸ toperiod (phenological advances of 6.9 days per standardized unit vs 1.2 days, for chilling and
⁹⁹ photoperiod, respectively; see Table S1). We estimated lower average responses to temperature
¹⁰⁰ compared to a model without phylogeny (model slopes for forcing and chilling decreased by 19%
¹⁰¹ and 21%, respectively); responses to chilling and forcing were also more similar when including
¹⁰² phylogeny (though chilling was still greater: 6.9 vs. 6 per standard unit), which contrasts with
¹⁰³ previous results suggesting chilling responses are much greater than forcing (Laube et al., 2014;
¹⁰⁴ Ettinger et al., 2020).

¹⁰⁵ These average estimates, however, fail to capture the large differences in species' responses to
¹⁰⁶ both chilling and forcing (Fig. 1, Supporting Table S3). By allowing species responses to
¹⁰⁷ vary, based on a model including their shared evolutionary history, we found species differences
¹⁰⁸ dwarfed the mean differences between cues, especially temperature cues (Fig. 1). The largest cue
¹⁰⁹ in magnitude—chilling—varied 25-fold between species, while variation to forcing varied 6-fold.
¹¹⁰ This variation indicates large differences between chilling and forcing occur at the species-level
¹¹¹ and not across species (as previously suggested Laube et al., 2014; Ettinger et al., 2020). These
¹¹² results highlight why robust phenological forecasts must account for both the complexity of
¹¹³ multiple cues and species-level variation in responses to them.

114 *Differences across clades & cues*

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

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

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

153 *Phylogenetic structure of phenological cues*

154 Variation—or lack thereof—in cues across species and clades provides possible insights into
155 the evolution of cues across the phylogeny. While responses to each cue were phylogenetically
156 structured, with closely related species exhibiting more similar sensitivities than distantly related
157 species, the strength of phylogenetic conservatism in response differed between cues (Fig. 2).
158 Responses to temperature (forcing and chilling) were moderately structured ($\lambda = 0.65$ and
159 $\lambda = 0.54$, for forcing and chilling, respectively). Phylogenetic structure in species responses to
160 photoperiod was comparatively weak ($\lambda = 0.4$) (see Fig. 2, Table S1).

161 Differences among species in their temperature responses represent shifts in the slope of the
162 relationship between the observed phenology and the cue. The observed phylogenetic structure
163 in temperature responses (forcing and chilling) would be consistent with an interaction with a
164 latent trait that moderates responses, and which also covaries with phylogeny (Davies et al.,
165 2019). This fits fundamentally with the idea that early-season phenology plays a critical role in
166 shaping species temporal niches (Gotelli and Graves, 1996) and thus should covary with a suite
167 of life-history traits, including whether species are early-active with rapid return on investment
168 traits, or start later in the season and have traits associated with higher competitive abilities
169 (e.g., Grime, 1977; Wolkovich and Donahue, 2021).

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

184 Phylogenetic conservatism (high λ) and slow evolutionary rates (low σ) in traits has sometimes
185 been interpreted as indicative of evolutionary constraints to adaptive change (Wiens et al., 2010;
186 Bennett et al., 2021). If this were the case, we might then conclude that species where responses
187 are dominated by forcing cues might be more vulnerable to future warming because phylogenetic
188 conservatism (λ) in forcing is higher compared to other cues and its evolutionary rate (σ) is lower
189 than that estimated for chilling. This is misleading, however, as estimates of λ are independent
190 from the rate of evolution, and macroevolutionary rates are estimated on phylogenetic trees that

integrate across millions of years of evolutionary history, and thus do not necessarily inform us of maximum possible rates of evolution over much shorter timescales. Indeed, there is accumulating evidence for rapid evolution to shifting climates (Bradshaw and Holzapfel, 2006; Franks et al., 2014). Our estimates are thus more useful in providing unique insights into the evolutionary history of phenological cues, and emphasize the critical importance of incorporating species-level differences in ecological forecasts

197 *Forecasting species-level responses*

Our results highlight that species-level variability can be extremely high—when properly estimated. Our approach, which partially pooled species responses based on their shared evolutionary history, estimated substantially higher variation across species compared with more widely used hierarchical models. This was especially noticeable in temperature responses (for chilling variance across species means, $var(\beta_{chill,j})$ from eqn. 2), was estimated as 23.86 in the phylogenetic model, versus 17.41 in the non-phylogenetic model; variance across means, $var(\beta_{force,j})$: 8.88 compared to 5.05) while photoperiod, which had low phylogenetic structure, was more similar across approaches (variance across means, $var(\beta_{photo,j})$: 0.82 compared to 0.62).

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

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

The contrasts between temperature and photoperiod responses—in both their variability across species and phylogenetic structure—have important implications for generating multi-species forecasts. Notably, responses to photoperiod appear weaker, more uniform across species, and less phylogenetically constrained compared to temperature. For temperature responses, the large variability among species makes predicting species-level responses challenging, but the phylogenetic structure in responses lets us borrow information from close relatives to improve our predictions. However, given that Brownian motion (our assumed model of evolution) is an

extremely noisy process, we recommend imputation only for missing taxa that are closely related to other well sampled species or clades (Molina-Venegas et al., 2018; ?).

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

240 **Methods & Materials**

241 **Phenological and Phylogenetic Data**

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

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

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

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

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

273

274 **Bayesian hierarchical phylogenetic model**

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

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

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

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

where

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

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

296 Predictors X_{chill} , X_{force} , X_{photo} are standardized chilling, forcing, and photoperiod, and their
 297 effects on the phenology of species j are determined by parameters $\beta_{chill,j}$, $\beta_{force,j}$, $\beta_{photo,j}$,
 298 representing species' responses (or sensitivities) to each of the cues. These responses, including
 299 the species-specific intercept α_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 Σ_i are $n \times n$ phylogenetic variance-covariance matrices of the form:

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

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

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

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

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

- 323 Bennett, J. M., J. Sunday, P. Calosi, F. Villalobos, B. Martínez, R. Molina-Venegas, M. B.
324 Araújo, A. C. Algar, S. Clusella-Trullas, B. A. Hawkins, et al. 2021. The evolution of critical
325 thermal limits of life on earth. *Nature communications* 12:1–9.
- 326 Bonamour, S., L. M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in
327 response to climate change: the importance of cue variation. *Philosophical Transactions of
328 the Royal Society B-Biological Sciences* 374.
- 329 Bradshaw, W. E., and C. M. Holzapfel. 2006. Evolutionary response to rapid climate change.
330 *Science* 312:1477–1478.
- 331 CaraDonna, P. J., and D. W. Inouye. 2014. Phenological responses to climate change do not
332 exhibit phylogenetic signal in a subalpine plant community. *Ecology* 96:355–361.
- 333 Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals.
334 *Annual Review of Ecology, Evolution, and Systematics* 48:159–182.
- 335 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble
336 version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical
337 Research: Atmospheres* 123:9391–9409.
- 338 Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism
339 in plant phenology. *Journal of Ecology* 101:1520–1530.
- 340 Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted
341 regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology
342 and biogeography* 28:275–285.
- 343 Dietze, M. 2017. Ecological forecasting. *In* *Ecological Forecasting*. Princeton University Press.
- 344 Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins,
345 C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to
346 community patterns. *Ecology letters* 15:545–553.
- 347 Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha,
348 and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses
349 to warming. *Nature Climate Change* pages 1–6.
- 350 Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich.
351 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist*
352 230:462–474.
- 353 Franks, S. J., J. J. Weber, and S. N. Aitken. 2014. Evolutionary and plastic responses to climate
354 change in terrestrial plant populations. *Evolutionary Applications* 7:123–139.

- 355 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative
356 data: a test and review of evidence. *The American Naturalist* 160:712–726.
- 357 Gelman, A., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models.
358 Cambridge University Press.
- 359 Gotelli, N. J., and G. R. Graves. 1996. The temporal niche. Pages 95–111 in N. J. Gotelli and
360 G. R. Graves, eds. Null Models in Ecology. Smithsonian Institution, Washington, DC.
- 361 Griffith, D. M., C. P. Osborne, E. J. Edwards, S. Bachle, D. J. Beerling, W. J. Bond, T. J. Gal-
362 laher, B. R. Helliker, C. E. R. Lehmann, L. Leatherman, J. B. Nippert, S. Pau, F. Qiu, W. J.
363 Riley, M. D. Smith, C. A. E. Stromberg, L. Taylor, M. Ungerer, and C. J. Still. 2020. Lineage-
364 based functional types: characterising functional diversity to enhance the representation of
365 ecological behaviour in land surface models. *New Phytologist* 228:15–23.
- 366 Grime, J. P. 1977. Evidence for existence of 3 primary strategies in plants and its relevance to
367 ecological and evolutionary theory. *American Naturalist* 111:1169–1194.
- 368 Hadfield, J. D. 2010. Mcmc methods for multi-response generalized linear mixed models: the
369 mcmcglmm r package. *Journal of statistical software* 33:1–22.
- 370 Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The*
371 *American Naturalist* 163:84–96.
- 372 Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of budburst in temperate trees.
373 *Journal of Applied Ecology* 29:597–604.
- 374 IPCC. 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge Uni-
375 versity Press, Cambridge, United Kingdom and New York, NY, USA.
- 376 Ives, A. R., and M. R. Helmus. 2011. Generalized linear mixed models for phylogenetic analyses
377 of community structure. *Ecological Monographs* 81:511–525.
- 378 Kochmer, J. P., and S. N. Handel. 1986. Constraints and competition in the evolution of flowering
379 phenology. *Ecological monographs* 56:303–325.
- 380 Kramer, K., A. Ducousoo, D. Gomory, J. K. Hansen, L. Ionita, M. Liesebach, A. Lorent,
381 S. Schueler, M. Sulkowska, S. de Vries, and G. von Wuehlisch. 2017. Chilling and forc-
382 ing requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between
383 provenances and are phenotypically plastic. *Agricultural and Forest Meteorology* 234:172–181.
- 384 Larcher, W. 1980. Plant Physiological Ecology. Springer-Verlag.
- 385 Laube, J., T. H. Sparks, N. Estrella, J. Höfler, D. P. Ankerst, and A. Menzel. 2014. Chilling
386 outweighs photoperiod in preventing precocious spring development. *Global Change Biology*
387 20:170–182.

- 388 Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins,
389 M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, et al. 2022. The power of forecasts to
390 advance ecological theory. *Methods in Ecology and Evolution* .
- 391 Molina-Venegas, R., J. C. Moreno-Saiz, I. Castro Parga, T. J. Davies, P. R. Peres-Neto, and
392 M. Á. Rodríguez. 2018. Assessing among-lineage variability in phylogenetic imputation of
393 functional trait datasets. *Ecography* 41:1740–1749.
- 394 Moorcroft, P., G. Hurtt, and S. Pacala. 2001. A method for scaling vegetation dynamics: The
395 ecosystem demography model (ed). *Ecological Monographs* 71:557–585.
- 396 Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005.
397 Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological
398 model. *Theoretical and Applied Genetics* 110:778–786.
- 399 Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts
400 across natural systems. *Nature* 421:37.
- 401 Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker,
402 and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. *Bioinformatics*
403 31:2888–2890.
- 404 Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering
405 time under climate change. *Journal of Ecology* 105:627–633.
- 406 Rinne, P., A. Saarelainen, and O. Juntila. 1994. Growth cessation and bud dormancy in
407 relation to ABA level in seedlings and coppice shoots of *Betula pubescens* as affected by a
408 short photoperiod, water stress and chilling. *Physiologia Plantarum* 90:451–458.
- 409 Schaber, J., and F. Badeck. 2003. Physiology-based phenology models for forest tree species in
410 germany. *International Journal of Biometeorology* 47:193–201. International Conference on
411 the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- 412 Serrano-Bueno, G., F. J. Romero-Campero, E. Lucas-Reina, J. M. Romero, and F. Valverde.
413 2017. Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology*
414 37:10–17.
- 415 Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global
416 dataset of meteorological forcings for land surface modeling. *Journal of Climate* 19:3088–3111.
- 417 Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny.
418 *American journal of botany* 105:302–314.
- 419 Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of
420 energetic scaling across the vertebrate tree of life. *The American Naturalist* 190:185–199.

- 421 Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, perfor-
422 mance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- 423 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.
424 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism
425 as an emerging principle in ecology and conservation biology. *Ecology letters* 13:1310–1324.
- 426 Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to
427 warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*
428 111:7906–7913.
- 429 Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phyloge-
430 netic patterns of species loss in thoreau’s woods are driven by climate change. *Proceedings of*
431 *the National Academy of Sciences* 105:17029–17033.
- 432 Wolkovich, E. M., and E. E. Cleland. 2014. Phenological niches and the future of invaded
433 ecosystems with climate change. *AoB Plants* 6:plu013.
- 434 Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and
435 communities in non-stationary environments. *Biological Reviews* .
- 436 Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and
437 J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-
438 PREE). doi:10.5063/F1CZ35KB.
- 439 Yang, Z., Y. Du, M. Shen, N. Jiang, E. Liang, W. Zhu, Y. Wang, and W. Zhao. 2021. Phy-
440 logenetic conservatism in heat requirement of leaf-out phenology, rather than temperature
441 sensitivity, in tibetan plateau. *Agricultural and Forest Meteorology* 304.
- 442 Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to
443 constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate
444 Change* 6:1120–1123.

⁴⁴⁵ **Tables and Figures**

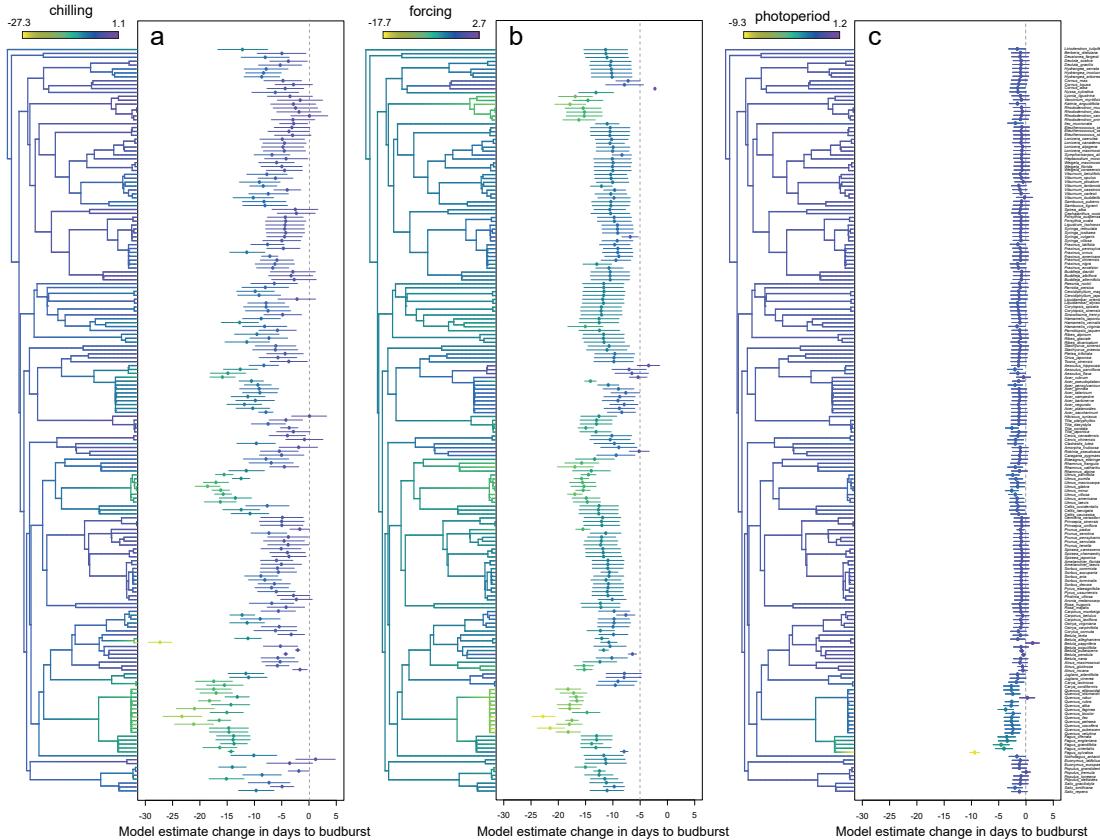


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit (z-transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored according to an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

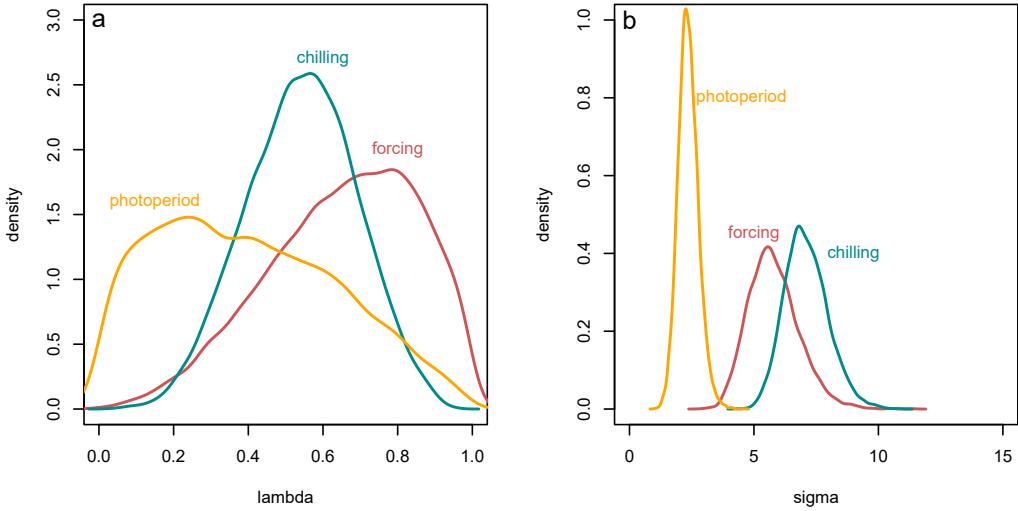


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters λ and σ estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to λ (a) and σ (b) from the phylogenetic model.

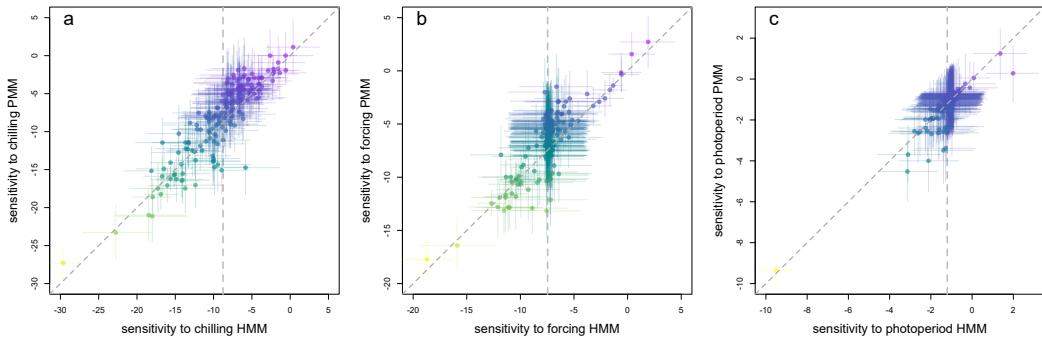


Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue (y -axis), and the more commonly used hierarchical model where species are exchangeable (where λ is constrained to be equal to zero, x -axis). While species with large amounts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

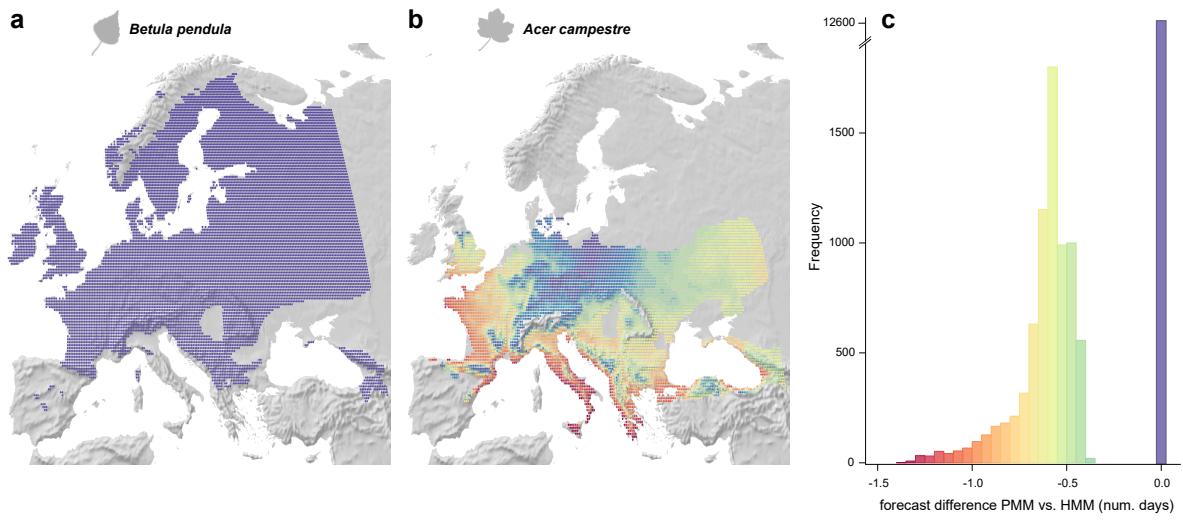


Figure 4: Comparison of forecasts of phenological shifts (i.e., computed as the difference between predictions under current climate vs. a 2°C warmer climate) resulting from a phylogenetic (PMM) and a non-phylogenetic (HMM) approach. Differences in forecasted shifts are negligible for well sampled species (*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.