

Phylogenetic estimates of species-level phenology improve ecological forecasting

March 29, 2024

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 S. Sodhi³ & 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

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

*Corresponding author: ignacio.moralesc@uah.es

1 Abstract

2 Our ability to adapt to climate change requires accurate ecological forecasting of shifts in key
3 ecosystem services, such as carbon storage and biodiversity maintenance. Current forecasts,
4 however, have failed to capture important variability in biological responses, especially across
5 species. Here, we present a novel method using Bayesian hierarchical phylogenetic models to
6 overcome this challenge and estimate species-level responses. We apply our method to pheno-
7 logical experiments manipulating temperature and daylength cues of woody plant species. We
8 show species differences dwarfed the average differences between cues, suggesting the current fo-
9 cus on comparing across cues may be misdirected. Further, our model provides insights on how
10 evolutionary history has shaped responses to cues, indicating stronger evolutionary dynamics
11 in responses to temperature—especially cool winter temperatures—than daylength. Our new
12 approach provides a major advance in ecological forecasting, with implications for predicting
13 the impacts of climate change and other anthropogenic forces on ecosystems.

¹⁴ **Introduction**

¹⁵ The biological impacts of climate change will have major implications for ecosystem functioning
¹⁶ and stability. With rising global temperatures many species have shifted their geographic distri-
¹⁷ butions northward in space and recurring life-history events—their phenology—earlier in time
¹⁸ (IPCC, 2014; Parmesan and Yohe, 2003), against a background of high variability. These shifts
¹⁹ have cascading consequences on many ecosystem services including carbon storage, making both
²⁰ mitigation and human adaptation to future warming dependent on accurate ecological forecasts
²¹ (Richardson et al., 2013).

²² While ecological forecasting has improved over recent years (Dietze, 2017; Lewis et al., 2022),
²³ it remains a challenge to reproduce the high variability observed in biological responses such
²⁴ as phenology, physiology or demography to environmental cues (IPCC, 2014). Some of this
²⁵ variability results from the complexity of climate change itself, including regional and seasonal
²⁶ variation in warming that underlies average trends alongside shifts in other climate axes (e.g.,
²⁷ precipitation). Much of it, however, could be driven by species-specific variation, reflecting
²⁸ evolved differences in species sensitivities to underlying environmental cues and their interac-
²⁹ tions. Unfortunately, we can only estimate the sensitivities to cues for a few well-studied species
³⁰ (Chuine and Regniere, 2017; Ettinger et al., 2020). In the absence of detailed data on individual
³¹ species, species groupings (e.g., functional groups) have improved ecosystem models (Moorcroft
³² et al., 2001; Griffith et al., 2020), but still capture only a fraction of the important variability
³³ (Fuccillo Battle et al., 2022).

³⁴ Recent efforts that have attempted to model species-specific responses to the environment (Diez
³⁵ et al., 2012) are often restricted by data availability—especially the common problem that data
³⁶ are often prevalent for some species and sparse across others. The rise of Bayesian hierarchical
³⁷ models can allow inference across species in such cases. However, underlying most hierarchical
³⁸ models is an implicit assumption that species are exchangeable (all species represent samples
³⁹ drawn from the same underlying distribution, Gelman and Hill, 2006), and they thus partially
⁴⁰ pool ('shrink') towards estimates for species with the most data and least variable responses,
⁴¹ making inference at the species-level unreliable (Ettinger et al., 2020). More reliable estimates
⁴² of species-level responses would allow us to better incorporate species differences into models of
⁴³ ecosystem change.

⁴⁴ Including the evolutionary history of species relationships in models of species responses could
⁴⁵ provide more robust species-level estimates than current approaches and a better understanding
⁴⁶ of the evolutionary constraints that might limit adaptation to change. For example, strong
⁴⁷ phylogenetic niche conservatism (Wiens et al., 2010) could potentially inhibit adaptive responses
⁴⁸ by drawing species back to an evolutionary conserved optimum, which is sub-optimal under new
⁴⁹ conditions. While incorporating such evolutionary history is traditionally seen as necessary,
⁵⁰ either as a statistical correction or to better understand species evolutionary history, the use of
⁵¹ such phylogenetic information should also improve model fitting and forecasts (Freckleton et al.,

52 2002).

53 Research using long-term observational data has highlighted the role that evolutionary history
54 may play in structuring plant phenological responses—which are critical to accurate forecasts
55 of carbon storage. Phylogenetic signal in plant phenology, including dates of budburst, leafout
56 and first flowering (Kochmer and Handel, 1986; Willis et al., 2008; Davies et al., 2013), suggests
57 that more closely related species share more similar phenologies, likely reflecting evolutionary
58 conservatism in responses to common cues. There are two broad explanations for why we
59 might expect phylogenetic conservatism in phenological traits. First, close relatives will tend to
60 share similar ecologies and physiologies, and thus be sensitive to similar environmental pressures.
61 Second, close relatives derive from common geographic centers of origin, and thus their ancestors
62 will have been exposed to—and have adapted to—similar environmental cues (Davies et al.,
63 2013). However, approaches using traditional phylogenetic comparative methods, have produced
64 conflicting results, with some studies reporting evidence of phylogenetic structure in phenology-
65 linked species declines (e.g., Willis et al., 2008) and in some phenophases, but not others (e.g.,
66 CaraDonna and Inouye, 2014), and in responses to some cues, but not others (e.g., Yang et al.,
67 2021). In addition, evidence for phylogenetic conservatism of phenological responses appears to
68 depend on method and species, even varying between sites with overlapping species sets (e.g.,
69 Rafferty and Nabity, 2017), which violates the fundamental idea of shared evolutionary history
70 (the common ancestor of two sets of species cannot possess two separate evolutionary histories
71 for the same trait). Thus, a first challenge is how to better integrate evolutionary history into
72 multi-species models of plant phenological responses.

73 Generating robust ecological forecasts requires addressing a second major hurdle—underlying
74 environmental cues that are complex and interacting. Decades of research have informed our
75 understanding of how species use environmental cues to time their phenotypic responses with
76 the temporal distribution of key resources while avoiding periods of high stress (Larcher, 1980;
77 Bonamour et al., 2019). Commonly, however, responses to environmental cues, and their evolu-
78 tion, are studied individually, linking a given phenotypic response to a single cue, for example,
79 time of leafout responding to summed heat during early spring (Davies et al., 2013). These
80 efforts fail to capture the more likely scenario for most phenotypic traits in which multiple cues
81 interacting along evolutionary history have shaped species responses (Ackerly, 2009). For many
82 plant species, phenological events are determined by a combination of temperature and light
83 (Chuine and Regnieri, 2017), with additional factors (e.g., other cues—like humidity, or species
84 physiology—vasculature or leaf structure) likely further mediating species responses. Although
85 these mediating factors are not well understood (Chuine and Regnieri, 2017), they can be ac-
86 counted for in models either as latent processes or by allowing non-stationarity in responses
87 across species (Davies et al., 2019).

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

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

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

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

Results & Discussion

Most species respond to all three primary cues—forcing, chilling, and photoperiod (Fig. 1, Supporting Table ??)—with responses to chilling approximately five-fold greater than to photoperiod (phenological advances of 6.9 days per standardized unit vs 1.2 days, for chilling and photoperiod, respectively; see Table ??). We estimated lower average responses to temperature compared to a model without phylogeny (model slopes for forcing and chilling decreased by 18% and 22%, respectively); responses to chilling and forcing were also more similar when including phylogeny (though chilling was still greater: 6.9 vs. 6.1 per standard unit), which contrasts with

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

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

139 *Differences across clades & cues*

140 The large differences across species produced striking differences between clades. For example,
141 several groups—oaks and beeches (Fagaceae), elms (Ulmaceae) and buckthorns (Rhamnaceae)—
142 are highly sensitive to chilling while others—rhododendrons (Ericaceae), butterfly bushes (Scro-
143 phulariaceae) and spindles (Celastraceae)—show little to no response to chilling (Fig. 1a).
144 Similar clade-level variation was observed for forcing, where some of these clades—e.g., Eri-
145 caceae, Rhamnaceae, Ulmaceae, or Fagaceae—were particularly sensitive and others, such as
146 the Sapindaceae, Cornaceae or Juglandaceae, show little response (Fig. 1b).

147 Some species responded strongly to both temperature cues, which could suggest the existence of
148 syndromes where the genetic basis for responses to one cue—e.g., forcing—has been selected for
149 alongside responses to another cue—e.g., chilling. This could occur if selection operates jointly
150 on responses to both cues; for example, if sensitivity to multiple cues provides greater insurance
151 against leafing out before the last frost (Bonamour et al., 2019; Wolkovich and Donahue, 2021).
152 Additionally, linkage or pleiotropism among loci associated with different cues (Nakagawa et al.,
153 2005) could induce across-cue correlations. However, the correlation in species responses across
154 cues was generally weak ($r = 0.31$; between forcing and chilling) and some genera, such as *Tilia*
155 and *Rhododendron* (Ericaceae), displayed strong responses to forcing but weak responses to
156 chilling, while others, such as *Acer* (Sapindaceae), show moderately strong responses to chilling
157 but weak responses to forcing (Fig. 1). Species sensitivity to one cue, thus, does not constrain
158 sensitivity to another cue, and it seems selection can operate independently on responses to
159 different cues (Bonamour et al., 2019).

160 In contrast to temperature cues (chilling and forcing), species-level responses to photoperiod were
161 almost uniform across species. This consistency provides novel insight on a large debate over
162 the prevalence of photoperiod cues in temperate trees, where previous experiments (Basler and
163 Körner, 2012; Zohner et al., 2016) and models (e.g., Hunter and Lechowicz, 1992; Schaber and
164 Badeck, 2003) suggested important variability across species that may constrain the responses
165 of certain species to warming (Way and Montgomery, 2015). Our results indicate variability is

166 limited to a handful of species in Fagaceae, which have been particularly well studied, especially
167 *Fagus sylvatica* (e.g., Basler and Körner, 2012; Zohner et al., 2016; Kramer et al., 2017). As
168 *Fagus sylvatica* is nearly five times more sensitive to photoperiod than most other measured tree
169 species, our results caution against using it to draw inferences of photoperiod responses more
170 widely. These same few species are also where most evidence of local adaptation in photoperiod
171 cues for spring phenology comes from (e.g., Kramer et al., 2017), in contrast with common
172 garden studies of other species, which find little evidence of local adaptation in spring (but not
173 fall) phenology (Aitken and Bemmels, 2016). The uniformity of response to photoperiod in our
174 results supports this latter view of generally low local adaptation in photoperiod cues for spring
175 phenology (i.e., if local adaptation were high in photoperiod cues, we would have expected more
176 variability).

177 *Phylogenetic structure of phenological cues*

178 Variation—or lack thereof—in cues across species and clades provides possible insights into
179 the evolution of cues across the phylogeny. While responses to each cue were phylogenetically
180 structured, with closely related species exhibiting more similar sensitivities than distantly related
181 species, the strength of phylogenetic conservatism in response differed between cues (Fig. 2).
182 Responses to temperature (forcing and chilling) were moderately structured ($\lambda = 0.65$ and
183 $\lambda = 0.54$, for forcing and chilling, respectively). Phylogenetic structure in species responses to
184 photoperiod was comparatively weak ($\lambda = 0.4$; see Fig. 2, Table ??).

185 Differences among species in their temperature responses represent shifts in the slope of the
186 relationship between the observed phenology and the cue. The observed phylogenetic structure
187 in temperature responses (forcing and chilling) would be consistent with an interaction with a
188 latent trait that moderates responses, and which also covaries with phylogeny (Davies et al.,
189 2019). This fits fundamentally with the idea that early-season phenology plays a critical role in
190 shaping species temporal niches (Gotelli and Graves, 1996) and thus should covary with a suite
191 of life-history traits, including whether species are early-active with rapid return on investment
192 traits, or start later in the season and have traits associated with higher competitive abilities
193 (e.g., Grime, 1977; Wolkovich and Donahue, 2021).

194 Weak phylogenetic signal in photoperiod sensitivity (Fig. 2) might seem at odds with the unifor-
195 mity of species response—i.e., there is very little variation in the responses to photoperiod across
196 species. However, somewhat counterintuitively, both uniform and random responses can man-
197 ifest as low phylogenetic signal when indexed by Brownian motion expectations (Wiens et al.,
198 2010). While rapid local adaptation within species might erase the phylogenetic structure in
199 photoperiod responses, it does not agree with the uniformity we find in species responses. How-
200 ever, if responses to photoperiod evolved early in plants, as seems likely (Serrano-Bueno et al.,
201 2017), and subsequent selection on photoperiod sensitivity was constrained by stabilizing selec-
202 tion operating on other life-history attributes sensitive to photoperiod (e.g., Rinne et al., 1994;
203 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 σ for photoperiod responses (Fig. 2). Here, as in more traditional phylogenetic comparative methods, σ represents the rate of evolution, and thus our results suggest photoperiod responses are also evolving slower than temperature responses (see Fig. ??).

Phylogenetic conservatism (high λ) and slow evolutionary rates (low σ) 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 (λ) in forcing is higher compared to other cues and its evolutionary rate (σ) is lower than that estimated for chilling. This is misleading, however, as estimates of λ are independent from the rate of evolution, and macroevolutionary rates are estimated on phylogenetic trees that integrate across millions of years of evolutionary history, and thus do not necessarily inform us of maximum possible rates of evolution over much shorter timescales. Our estimates are thus more useful in providing unique insights into the evolutionary history of phenological cues, and emphasize the critical importance of incorporating species-level differences in ecological forecasts.

220 *Forecasting species-level responses*

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

The consequences of including shared evolutionary history in forecasting are most apparent for poorly sampled species nested within more well-sampled clades. For example, forecasts for *Acer campestre*, which has only 6 observations, shift by up to 35% in the number of days until budburst after forcing starts, when comparing our phylogenetically informed model to one without phylogeny (see Fig. 4 for details). 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. ??). Thus,

traditional (non-phylogenetically informed) approaches that partially pool across species (most hierarchical models in ecology, e.g., Flynn and Wolkovich, 2018; 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. ??). Another advantage of our Bayesian approach is that we are also better able to accommodate imprecision in the data that informs our model, which might arise from multiple sources, including measurement or experimental error, and the general stochasticity associated with limited sample sizes and unbalanced species representation. Critically, by partially pooling across species and weighting by phylogeny, we gain strength from species estimates that are informed by more data, such as within *Betula* and *Fagaceae*, but avoid skewing estimates for phylogenetically distant clade that may have been exposed to different selective regimes. We found species estimates were robust through cross-validation—the phylogenetic model better predicted observed values for held-out data, and yielded more stable species coefficients compared to a hierarchical model (Figs. ??-??; see “Leave-One-Clade-Out model cross validation”).

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, 2023).

While we focused on spring phenology here, our new approach suggests a path forward for more general forecasting of species-level climate change responses. Our results show how including the phylogenetic relationship of species in a mechanistic model of underlying cues can overcome major limitations of most current hierarchical models—correcting biased model estimates, 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.

274 **Methods & Materials**

275 **Phenological and Phylogenetic Data**

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

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

285 We scraped data from all papers of woody species that tested for photoperiod and/or tempera-
286 ture effects on budburst, leafout, or flowering, resulting in data from 155 experiments across 97
287 papers in the updated database. Ettinger et al. (2020), which used a portion (72 experiments
288 across 49 papers) of the earlier OSPREE database, provides extensive methods on database
289 creation and cleaning.

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

300 Across experiments chilling treatments were often fully or partially applied in the field, thus we
301 estimated chilling—both in the field and applied in controlled environments—using Utah units
302 with the **chillR** package. We estimated field chilling from 1 September to the date given for when
303 samples were taken from the field using daily temperature data (converted to hourly) from both
304 European (E-OBS, version 16, calculating the average of minimum and maximum daily temps,
305 Cornes et al., 2018) and North American (v3, Sheffield et al., 2006) gridded climate datasets.
306 We also converted experimental chilling into Utah chill units, based on reported treatments
307 (for studies with a mix of field and experimental treatments, we added field and experimentally
308 applied Utah units). To avoid numerical instability in our models (from having predictor values

309 on very different scales), we divided Utah units by 240 (roughly equivalent to 10 days of average
310 chilling).

311 We report the Utah model because a small number of studies reported chilling only in Utah
312 units, thus using this common metric allowed us to include the most data. The Utah model
313 relies on the assumption that temperatures between 1.4 and 15.9C affect endodormancy release
314 differently, though recent findings show possibly similar effects for a wide range of temperatures
315 (-2 to 10C, see Baumgarten et al., 2021). Because chilling is a latent process an accurate model
316 of it, especially for the 191 species in our dataset, is not currently possible (Ettinger et al.,
317 2020). We found consistent results, however, using another common model of chilling—chill
318 portions, suggesting our results are robust to the exact chilling metric used (see Tables ?? and
319 ?? in Supporting Information).

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

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

344

345 **Bayesian hierarchical phylogenetic model**

346 Commonly used phylogenetic regression methods today (e.g., phylogenetic generalized least
347 squares models, PGLS, Freckleton et al. (2002); phylogenetic mixed models, PMM, Hous-
348 worth et al. (2004)) were originally conceived as statistical corrections for phylogenetic non-
349 independence across observations—generally species—thus allowing multi-species studies to meet
350 the assumptions of linear regression (Freckleton et al., 2002). These corrections incorporated
351 phylogenetic structure by estimating the magnitude of a transformation of a variance-covariance
352 (VCV) matrix whose elements were derived from the amount of evolutionary history (branch
353 lengths) shared between species on a phylogeny. The most commonly used transformation was
354 Pagel’s λ —a multiplier of the off-diagonal elements—where estimates of $\lambda = 1$ essentially left the
355 VCV untransformed and suggested that the residuals of the regression had phylogenetic signal
356 consistent with Brownian motion; estimates of $\lambda = 0$ suggested no phylogenetic signal. Be-
357 cause the original aim of these methods was to correct for statistical bias introduced by shared
358 evolutionary history among species, the underlying assumption of phylogenetic regressions is
359 that phylogenetic relatedness would only affect either model residuals (in PGLS approaches,
360 Freckleton et al., 2002) or model intercepts (e.g., in many PMM approaches, Housworth et al.,
361 2004).

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

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

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

where

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

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

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

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

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

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

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

where the means of the multivariate normal distributions are root trait values (i.e., values of cue responses prior to evolving across a phylogenetic tree) and $\boldsymbol{\Sigma}_i$ are $n \times n$ phylogenetic variance-covariance matrices of the form:

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

where σ_i^2 is the rate of evolution across a tree for a given trait or predictor (here assumed to be constant along all branches), λ_i scales branch lengths and therefore is a measure of the phylogenetic signal or extent of phylogenetic relatedness on each model parameter (i.e., α_j , $\beta_{force,j}$, $\beta_{photo,j}$), 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 ??-??).

To assess if the phylogenetic mixed model (PMM) presents any advantages with respect commonly used hierarchical mixed models (HMM; see e.g., Ettinger et al., 2020), beyond fitting evolutionary parameters to model predictors, we compare results of PMM and HMM. HMM is a simplified version of PMM where off-diagonal elements of the variance-covariance phylogenetic matrices are multiplied by zero ($\lambda = 0$). Both models account for differences in sample sizes and variances for each species, by partially pooling across all data while at the same time providing species-level estimates; however the PMM will pool more strongly to closely-related species

394 when λ is high. Additional grouping factors beyond species could be added to these models. For
395 example, similar approaches can be used to estimate study or location effects.

396 **References**

- 397 Ackerly, D. 2009. Conservatism and diversification of plant functional traits: Evolutionary rates
398 versus phylogenetic signal. *Proceedings of the National Academy of Sciences of the United
399 States of America* 106:19699–19706.
- 400 Aitken, S. N., and J. B. Bemmels. 2016. Time to get moving: assisted gene flow of forest trees.
401 *Evol Appl* 9:271–90.
- 402 Azeez, A., and A. P. Sane. 2015. Photoperiodic growth control in perennial trees. *Plant signaling
403 & behavior* 10:e1087631.
- 404 Basler, D., and C. Körner. 2012. Photoperiod sensitivity of bud burst in 14 temperate forest
405 tree species. *Agricultural and Forest Meteorology* 165:73–81.
- 406 Baumgarten, F., C. M. Zohner, A. Gessler, and Y. Vitasse. 2021. Chilled to be forced: the best
407 dose to wake up buds from winter dormancy. *New Phytologist* 230:1366–1377.
- 408 Bennett, J. M., J. Sunday, P. Calosi, F. Villalobos, B. Martínez, R. Molina-Venegas, M. B.
409 Araújo, A. C. Algar, S. Clusella-Trullas, B. A. Hawkins, et al. 2021. The evolution of critical
410 thermal limits of life on earth. *Nature communications* 12:1–9.
- 411 Bonamour, S., L. M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in
412 response to climate change: the importance of cue variation. *Philosophical Transactions of
413 the Royal Society B-Biological Sciences* 374.
- 414 Buonaiuto, D. M., M. Donahue, and E. M. Wolkovich. 2023. Experimental designs for testing the
415 interactive effects of temperature and light in ecology: the problem of periodicity. *Functional
416 Ecology* .
- 417 CaraDonna, P. J., and D. W. Inouye. 2014. Phenological responses to climate change do not
418 exhibit phylogenetic signal in a subalpine plant community. *Ecology* 96:355–361.
- 419 Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals.
420 *Annual Review of Ecology, Evolution, and Systematics* 48:159–182.
- 421 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble
422 version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical
423 Research: Atmospheres* 123:9391–9409.
- 424 Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism
425 in plant phenology. *Journal of Ecology* 101:1520–1530.

- 426 Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted
427 regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology*
428 and biogeography 28:275–285.
- 429 Dietze, M. 2017. Ecological forecasting. *In Ecological Forecasting*. Princeton University Press.
- 430 Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins,
431 C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to
432 community patterns. *Ecology letters* 15:545–553.
- 433 Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha,
434 and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses
435 to warming. *Nature Climate Change* pages 1–6.
- 436 Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich.
437 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist*
438 230:462–474.
- 439 Flynn, D. F. B., and E. M. Wolkovich. 2018. Temperature and photoperiod drive spring phe-
440 nology across all species in a temperate forest community. *New Phytologist* 219:1353–1362.
- 441 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative
442 data: a test and review of evidence. *The American Naturalist* 160:712–726.
- 443 Fuccillo Battle, K., A. Duhon, C. R. Vispo, T. M. Crimmins, T. N. Rosenstiel, L. L. Armstrong-
444 Davies, and C. E. de Rivera. 2022. Citizen science across two centuries reveals phenological
445 change among plant species and functional groups in the northeastern us. *Journal of Ecology*
446 110:1757–1774.
- 447 Gelman, A., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models.
448 Cambridge University Press.
- 449 Gotelli, N. J., and G. R. Graves. 1996. The temporal niche. Pages 95–111 *in* N. J. Gotelli and
450 G. R. Graves, eds. *Null Models in Ecology*. Smithsonian Institution, Washington, DC.
- 451 Griffith, D. M., C. P. Osborne, E. J. Edwards, S. Bachle, D. J. Beerling, W. J. Bond, T. J. Gal-
452 laher, B. R. Helliker, C. E. R. Lehmann, L. Leatherman, J. B. Nippert, S. Pau, F. Qiu, W. J.
453 Riley, M. D. Smith, C. A. E. Stromberg, L. Taylor, M. Ungerer, and C. J. Still. 2020. Lineage-
454 based functional types: characterising functional diversity to enhance the representation of
455 ecological behaviour in land surface models. *New Phytologist* 228:15–23.
- 456 Grime, J. P. 1977. Evidence for existence of 3 primary strategies in plants and its relevance to
457 ecological and evolutionary theory. *American Naturalist* 111:1169–1194.
- 458 Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The*
459 *American Naturalist* 163:84–96.

- 460 Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of budburst in temperate trees.
461 Journal of Applied Ecology 29:597–604.
- 462 IPCC. 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge Uni-
463 versity Press, Cambridge, United Kingdom and New York, NY, USA.
- 464 Kochmer, J. P., and S. N. Handel. 1986. Constraints and competition in the evolution of flowering
465 phenology. Ecological monographs 56:303–325.
- 466 Kramer, K., A. Ducousoo, D. Gomory, J. K. Hansen, L. Ionita, M. Liesebach, A. Lorent,
467 S. Schueler, M. Sulkowska, S. de Vries, and G. von Wuehlisch. 2017. Chilling and forc-
468 ing requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between
469 provenances and are phenotypically plastic. Agricultural and Forest Meteorology 234:172–181.
- 470 Larcher, W. 1980. Plant Physiological Ecology. Springer-Verlag.
- 471 Laube, J., T. H. Sparks, N. Estrella, J. Höfler, D. P. Ankerst, and A. Menzel. 2014. Chilling
472 outweighs photoperiod in preventing precocious spring development. Global Change Biology
473 20:170–182.
- 474 Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins,
475 M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, et al. 2022. The power of forecasts to
476 advance ecological theory. Methods in Ecology and Evolution .
- 477 Molina-Venegas, R., I. Morales-Castilla, and M. Á. Rodríguez. 2023. Unreliable prediction of
478 b-vitamin source species. Nature Plants 9:31–33.
- 479 Molina-Venegas, R., J. C. Moreno-Saiz, I. Castro Parga, T. J. Davies, P. R. Peres-Neto, and
480 M. Á. Rodríguez. 2018. Assessing among-lineage variability in phylogenetic imputation of
481 functional trait datasets. Ecography 41:1740–1749.
- 482 Moorcroft, P., G. Hurtt, and S. Pacala. 2001. A method for scaling vegetation dynamics: The
483 ecosystem demography model (ed). Ecological Monographs 71:557–585.
- 484 Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005.
485 Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological
486 model. Theoretical and Applied Genetics 110:778–786.
- 487 Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts
488 across natural systems. Nature 421:37.
- 489 Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker,
490 and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. Bioinformatics
491 31:2888–2890.

- 492 Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering
493 time under climate change. *Journal of Ecology* 105:627–633.
- 494 Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey. 2013.
495 Climate change, phenology, and phenological control of vegetation feedbacks to the climate
496 system. *Agricultural and Forest Meteorology* 169:156–173.
- 497 Rinne, P., A. Saarelainen, and O. Junntila. 1994. Growth cessation and bud dormancy in
498 relation to ABA level in seedlings and coppice shoots of *Betula pubescens* as affected by a
499 short photoperiod, water stress and chilling. *Physiologia Plantarum* 90:451–458.
- 500 Schaber, J., and F. Badeck. 2003. Physiology-based phenology models for forest tree species in
501 germany. *International Journal of Biometeorology* 47:193–201. International Conference on
502 the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- 503 Serrano-Bueno, G., F. J. Romero-Campero, E. Lucas-Reina, J. M. Romero, and F. Valverde.
504 2017. Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology*
505 37:10–17.
- 506 Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global
507 dataset of meteorological forcings for land surface modeling. *Journal of Climate* 19:3088–3111.
- 508 Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny.
509 *American journal of botany* 105:302–314.
- 510 Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of
511 energetic scaling across the vertebrate tree of life. *The American Naturalist* 190:185–199.
- 512 Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, perfor-
513 mance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- 514 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.
515 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism
516 as an emerging principle in ecology and conservation biology. *Ecology letters* 13:1310–1324.
- 517 Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to
518 warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*
519 111:7906–7913.
- 520 Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phyloge-
521 netic patterns of species loss in thoreau’s woods are driven by climate change. *Proceedings of*
522 *the National Academy of Sciences* 105:17029–17033.
- 523 Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and
524 communities in non-stationary environments. *Biological Reviews* .

- 525 Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and
526 J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-
527 PREE). doi:10.5063/F1CZ35KB.
- 528 Yang, Z., Y. Du, M. Shen, N. Jiang, E. Liang, W. Zhu, Y. Wang, and W. Zhao. 2021. Phy-
529 logenetic conservatism in heat requirement of leaf-out phenology, rather than temperature
530 sensitivity, in tibetan plateau. Agricultural and Forest Meteorology 304.
- 531 Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to
532 constrain climate-driven shifts in leaf-out times of northern woody plants. Nature Climate
533 Change 6:1120–1123.

534 **Data and code availability statement**

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

540 **Tables and Figures**

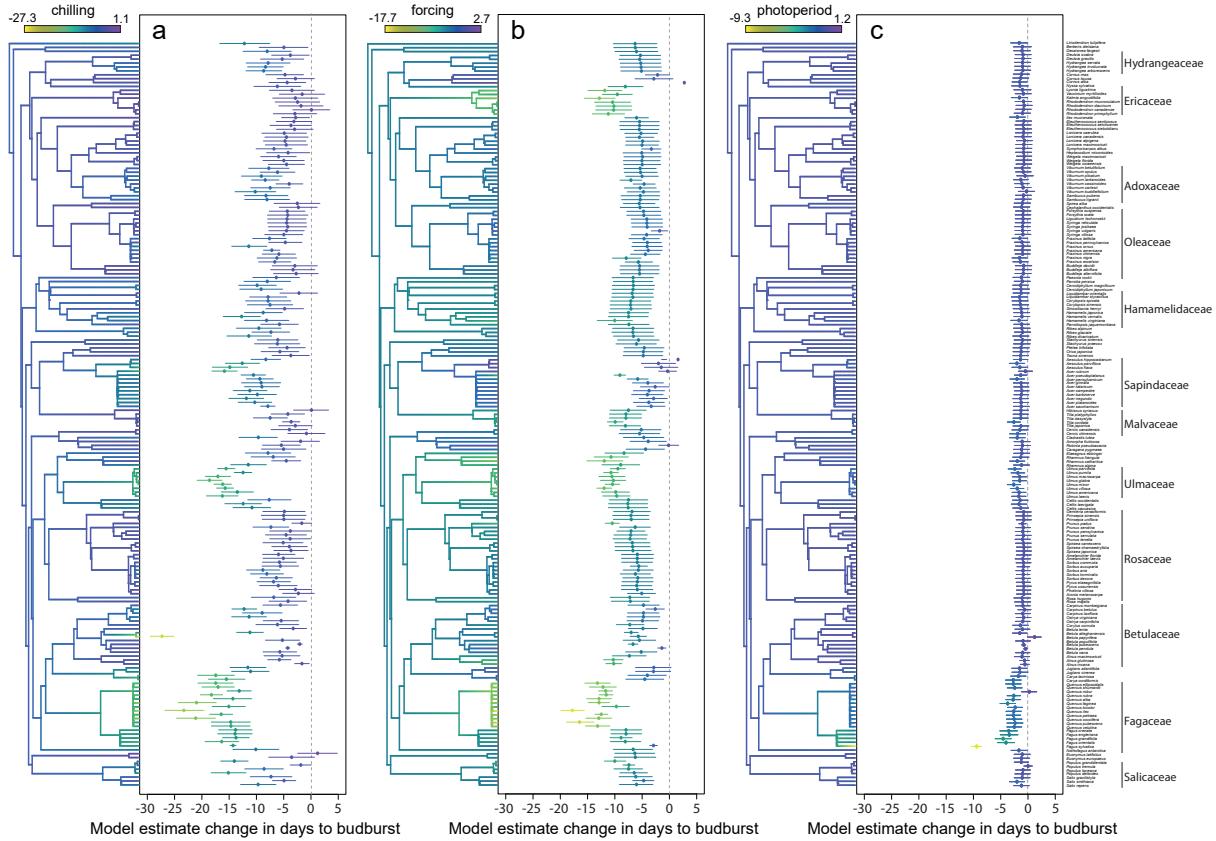


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit (z -transformation) of the cues across 191 woody species. The same phylogenetic tree is shown in each panel, colored according to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

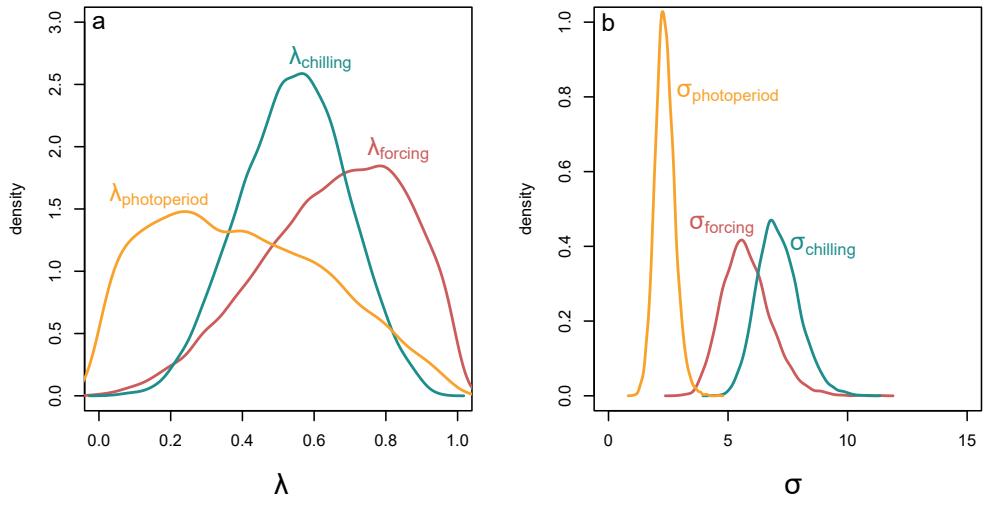


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters λ and σ estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to λ (a) and σ (b) from the phylogenetic model.

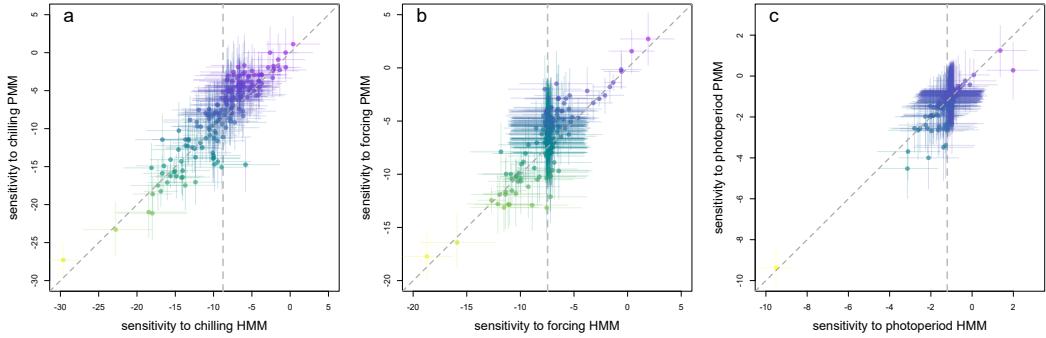


Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue (phylogenetic mixed model, PMM; y -axis), and the more commonly used hierarchical mixed model (HMM) where species are exchangeable (where λ is constrained to be equal to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

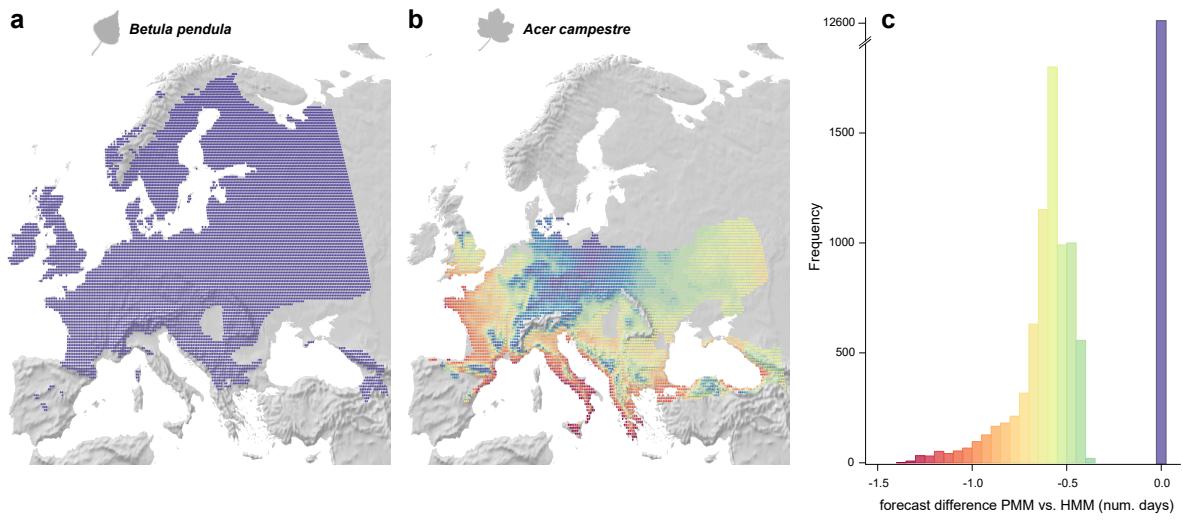


Figure 4: Comparison of forecasts of phenological shifts (i.e., computed as the difference between predictions under current climate vs. a 2°C warmer climate) resulting from a phylogenetic (PMM) and a non-phylogenetic but hierarchical (HMM) approach. Differences in forecasted shifts are negligible for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