

Phylogenetic estimates of species-level phenology improve ecological forecasting

October 11, 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 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

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 across 191 temperate woody plant species for which we had experimental data on how their phenology respond to two major environmental cues of spring phenology: temperature and daylength. We find that variation is greater across species than across cues, suggesting the current focus on identifying which cue is the strongest may be less important than predicting how each species responds to a combination of cues. Further, our model provides insights on how evolutionary history has shaped responses to cues, suggesting stronger evolutionary dynamics in responses to temperature—especially cool winter temperatures—than daylength. Our approach provides a major advance in ecological forecasting, with implications for predicting the impacts of climate change and other anthropogenic forces on species interactions and ecosystem functioning.

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 (Richardson et al., 2013).

⁹ While ecological forecasting has improved over recent years (Dietze, 2017; Lewis et al., 2022),
¹⁰ it remains a challenge to reproduce the high variability observed in biological responses such as
¹¹ phenology, physiology or demography to environmental cues (IPCC, 2014). Some of this variabil-
¹² ity results from the complexity of climate change itself, including regional and seasonal variation
¹³ in warming that underlies average trends alongside shifts in other climate axes (e.g. precipi-
¹⁴ tation). Much of it, could be driven by species-specific variation, reflecting evolved differences
¹⁵ in species' sensitivities to underlying environmental cues and their interactions. Unfortunately,
¹⁶ we only know the sensitivities to cues of a few well-studied species (Chuine and Regnieri, 2017;
¹⁷ Ettinger et al., 2020). In the absence of detailed data on individual species, species group-
¹⁸ ings (e.g., functional groups) have improved ecosystem models (Moorecroft et al., 2001; Griffith
¹⁹ et al., 2020), but still capture only a fraction of the important variability (Fuccillo Battle et al.,
²⁰ 2022). Improving forecasts, will require models that accurately predict species-level differences
²¹ in responses to complex environmental change.

²² Recent efforts that have attempted to model species-specific responses to the environment (Diez
²³ et al., 2012) are often restricted by data availability—especially the common problem that data
²⁴ are often prevalent for some species and sparse across others. The rise of Bayesian hierarchical
²⁵ models can allow inference across species in such cases. However, underlying most hierarchical
²⁶ models is an implicit assumption that species are exchangeable (Gelman and Hill, 2006), and they
²⁷ thus partially pool ('shrink') towards estimates for species with the most data and least variable
²⁸ responses, making inference at the species-level unreliable (Ettinger et al., 2020). More reliable
²⁹ estimates of species-level responses would allow us to better incorporate species differences into
³⁰ models of ecosystem change.

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

39 2002).

40 Research using long-term observational data has highlighted the role that evolutionary history
41 may play in structuring plant phenological responses—which are critical to accurate forecasts
42 of carbon storage. Phylogenetic signal in plant phenology including dates of budburst, leafout
43 and first flowering (Kochmer and Handel, 1986; Willis et al., 2008; Davies et al., 2013), suggests
44 species responses to cues have diverged over macro-evolutionary timescales, helping explain
45 species present day differences. However, approaches using traditional phylogenetic comparative
46 methods, have produced conflicting results (e.g., Willis et al., 2008; CaraDonna and Inouye, 2014;
47 Yang et al., 2021). In addition, evidence for phylogenetic conservatism of phenological responses
48 appears to depend on method and species, even varying between sites with overlapping species
49 sets (e.g., Rafferty and Nabity, 2017), which violates the fundamental idea of shared evolutionary
50 history (the common ancestor of two sets of species cannot possess two separate evolutionary
51 histories for the same trait). A first challenge is thus how to better integrate evolutionary history
52 into multi-species models of plant phenological responses.

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

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

⁷⁸ studied species (Ettinger et al., 2020).

⁷⁹ Here we present a novel Bayesian framework that extends upon phylogenetic mixed models
⁸⁰ (Housworth et al., 2004) to examine how chilling, forcing (both metrics of temperature) and
⁸¹ photoperiod together determine spring plant phenology. By allowing non-stationarity in species
⁸² responses across the phylogeny (Davies et al., 2019), our model departs from previous work
⁸³ and assumptions of traditional phylogenetic comparative methods concerned with phylogenetic
⁸⁴ correction (e.g. Freckleton et al., 2002), and moves towards integrating evolutionary history
⁸⁵ in models of phenological responses to environmental change. To understand how evolution
⁸⁶ has shaped the cues underlying shifting phenology with climate change (Uyeda et al., 2017), we
⁸⁷ explicitly incorporate phylogenetic structure across model intercepts and slopes (that is, allowing
⁸⁸ a separate model of evolutionary history for chilling, forcing and photoperiod, see Methods &
⁸⁹ Materials for a complete description).

⁹⁰ We illustrate our method with an unprecedented dataset on phenological responses to environ-
⁹¹ mental cues (chilling, forcing and photoperiod) determined experimentally for 191 deciduous
⁹² woody species (by far the most studied group of species in phenology experiments, see Ettinger
⁹³ et al., 2020), in an updated version of the Observed Spring Phenology Responses in Experi-
⁹⁴ mental Environments (OSPREE) database (Wolkovich et al., 2019). These data combined with
⁹⁵ the Smith and Brown (2018) megatree adjusted to our species and modeling approach allows
⁹⁶ us to address the common question of which cue has the largest effect on budburst and, at the
⁹⁷ same time, provide robust estimates of how cues vary across species. Using spring phenology,
⁹⁸ we identify historical regime shifts (Uyeda et al., 2017) in phenological responses, and highlight
⁹⁹ how our approach could advance forecasting of other critical responses to ongoing global change.

¹⁰⁰ Results & Discussion

¹⁰¹ Most species respond to all three primary cues—forcing, chilling, and photoperiod (Fig. 1,
¹⁰² Supporting Table S1)—with responses to chilling approximately five-fold greater than to pho-
¹⁰³ toperiod (phenological advances of 6.9 days per standardized unit vs 1.2 days, for chilling and
¹⁰⁴ photoperiod, respectively; see Table S1). We estimated lower average responses to temperature
¹⁰⁵ compared to a model without phylogeny (model slopes for forcing and chilling decreased by 18%
¹⁰⁶ and 22%, respectively); responses to chilling and forcing were also more similar when including
¹⁰⁷ phylogeny (though chilling was still greater: 6.9 vs. 6.1 per standard unit), which contrasts with
¹⁰⁸ previous results suggesting chilling responses are much greater than forcing (Laube et al., 2014;
¹⁰⁹ Ettinger et al., 2020).

¹¹⁰ These average estimates, however, fail to capture the large differences in species' responses to
¹¹¹ both chilling and forcing (Fig. 1, Supporting Table S5). By allowing species responses to
¹¹² vary, based on a model including their shared evolutionary history, we found species differences
¹¹³ dwarfed the mean differences between cues, especially temperature cues (Fig. 1). The largest cue

114 in magnitude—chilling—varied 24-fold between species, while variation to forcing varied 7-fold.
115 This variation indicates large differences between chilling and forcing occur at the species-level
116 and not across species (as previously suggested Laube et al., 2014; Ettinger et al., 2020). These
117 results highlight why robust phenological forecasts must account for both the complexity of
118 multiple cues and species-level variation in responses to them.

119 *Differences across clades & cues*

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

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

141 In contrast to temperature cues (chilling and forcing), species-level responses to photoperiod were
142 almost uniform across species. This consistency provides novel insight on a large debate over
143 the prevalence of photoperiod cues in temperate trees, where previous experiments (Basler and
144 Körner, 2012; Zohner et al., 2016) and models (e.g., Hunter and Lechowicz, 1992; Schaber and
145 Badeck, 2003) suggested important variability across species that may constrain the responses
146 of certain species to warming (Way and Montgomery, 2015). Our results indicate variability is
147 limited to a handful of species in Fagaceae, which have been particularly well studied, especially
148 *Fagus sylvatica* (e.g., Basler and Körner, 2012; Zohner et al., 2016; Kramer et al., 2017). As
149 *Fagus sylvatica* is nearly five times more sensitive to photoperiod than most other measured tree
150 species, our results caution against using it to draw inferences of photoperiod responses more
151 widely. These same few species are also where most evidence of local adaptation in photoperiod

152 cues for spring phenology comes from (e.g., Kramer et al., 2017), in contrast with common
153 garden studies of other species, which find little evidence of local adaptation in spring (but not
154 fall) phenology (Aitken and Bemmels, 2016). The uniformity of response to photoperiod in our
155 results supports this latter view of generally low local adaptation in photoperiod cues for spring
156 phenology (i.e., if local adaptation were high in photoperiod cues we would have expected more
157 variability across species).

158 *Phylogenetic structure of phenological cues*

159 Variation—or lack thereof—in cues across species and clades provides possible insights into
160 the evolution of cues across the phylogeny. While responses to each cue were phylogenetically
161 structured, with closely related species exhibiting more similar sensitivities than distantly related
162 species, the strength of phylogenetic conservatism in response differed between cues (Fig. 2).
163 Responses to temperature (forcing and chilling) were moderately structured ($\lambda = 0.65$ and
164 $\lambda = 0.54$, for forcing and chilling, respectively). Phylogenetic structure in species responses to
165 photoperiod was comparatively weak ($\lambda = 0.4$) (see Fig. 2, Table S1).

166 Differences among species in their temperature responses represent shifts in the slope of the
167 relationship between the observed phenology and the cue. The observed phylogenetic structure
168 in temperature responses (forcing and chilling) would be consistent with an interaction with a
169 latent trait that moderates responses, and which also covaries with phylogeny (Davies et al.,
170 2019). This fits fundamentally with the idea that early-season phenology plays a critical role in
171 shaping species temporal niches (Gotelli and Graves, 1996) and thus should covary with a suite
172 of life-history traits, including whether species are early-active with rapid return on investment
173 traits, or start later in the season and have traits associated with higher competitive abilities
174 (e.g., Grime, 1977; Wolkovich and Donahue, 2021).

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

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

203 *Forecasting species-level responses*

204 Our results highlight that species-level variability can be extremely high—when properly esti-
205 mated. Our approach, which partially pooled species responses based on their shared evolution-
206 ary history, estimated substantially higher variation across species compared with more widely
207 used hierarchical models. This was especially noticeable in temperature responses (for chilling
208 variance across species means, $var(\beta_{chill,j}$ from eqn. 2), was estimated as 23.55 in the phyloge-
209 netic model, versus 17.47 in the non-phylogenetic model; variance across means, $var(\beta_{force,j})$:
210 8.75 compared to 5.01) while photoperiod, which had low phylogenetic structure, was more
211 similar across approaches (variance across means, $var(\beta_{photo,j})$: 0.83 compared to 0.64).

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

223 The increase in variability across species in our model with phylogenetic structure also decreased
224 the uncertainty in estimates for each individual species temperature responses (Fig. S3). Thus,
225 traditional (non-phylogenetically informed) approaches that partially pool across species (most
226 hierarchical models in ecology, e.g., Flynn and Wolkovich, 2018; Ettinger et al., 2020) may also
227 lead to less precise predictions and forecasts of phenology for individual species, although overall

228 model accuracy might still appear reasonable (see Fig. S4).

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

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

247 **Methods & Materials**

248 **Phenological and Phylogenetic Data**

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

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

258 We scraped data from all papers of woody species that tested for photoperiod and/or tempera-
259 ture effects on budburst, leafout, or flowering, resulting in data from 155 experiments across 97
260 papers in the updated database. Ettinger et al. (2020) used a portion (72 experiments across
261 49 papers) of the earlier OSPREE database and provides extensive methods on the database
262 creation and cleaning. For our analysis here, we focused on angiosperms (as gymnosperms are
263 very poorly represented in spring phenology experiments), and included all budburst experi-
264 ments where we could quantify chilling, forcing and photoperiod levels, resulting in 44 studies
265 from 33 papers, resulting in 2940 data points.

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

273 The Utah model, as other approaches to estimate chilling, has important shortcomings as it
274 relies on the assumption that temperatures between 1.4 and 15.9 C affect dormancy release
275 differently, while recent findings show similar effects on dormancy releases for a wide range of
276 temperatures (-2 to 10; Baumgarten et al. (2021)). Unfortunately, we lack a fully versatile model
277 that accurately estimates dormancy release for each 191 species in our dataset, likely because
278 we still need deeper understanding of the processes underlying chilling (Ettinger et al., 2020).
279 Nonetheless, to test the sensitivity of our results to our choice of modelling chilling, we ran
280 models using chill portions instead of Utah units, and found consistent results (see Tables SXX
281 and SXI in Supporting Information).

282 To address different day/night or ramped temperatures we take a weighted hourly average over
283 the full period of forcing. We also adjusted for differences caused by co-varying thermo- and
284 photo-periodicity for both the forcing temperatures and photoperiod values.

285 For a phylogenetic tree, we pruned the phylogenetic megatree for seed plants (Smith and Brown,
286 2018) to extract a subset of the phylogenetic tree containing only the angiosperm species in the
287 OSPREE dataset, then added species that were not present in the megatree as polytomies (i.e.,
288 comb-like phylogenetic relationships assuming that species within the polytomy originated simul-
289 taneously from their common ancestor) at the generic level (using the function ‘congeneric.merge’
290 in Pearse et al., 2015), with a branch length of the congeneric basal node age. Our pruned tree
291 had 8 Polytomies affecting 46 out of 191 species. To test for the ability of polytomies to bias
292 our results we ran sensitivity analyses excluding these species from models (see Supporting In-
293 formation).

294

295 Bayesian hierarchical phylogenetic model

296 Commonly used phylogenetic regression methods today (e.g., phylogenetic generalized least
297 squares models, PGLS, Freckleton et al. (2002); phylogenetic mixed models, PMM, Hous-
298 worth et al. (2004)) were originally conceived as statistical corrections for phylogenetic non-
299 independence across observations—generally species—thus allowing multi-species studies to meet
300 the assumptions of linear regression (Freckleton et al., 2002). These corrections incorporated
301 phylogenetic structure by estimating the magnitude of a transformation of a variance-covariance
302 (VCV) matrix whose elements were derived from the amount of evolutionary history (branch
303 lengths) shared between species on a phylogeny. The most commonly used transformation was
304 Pagel’s λ —a multiplier of the off-diagonal elements—where estimates of $\lambda = 1$ essentially left the
305 VCV untransformed and suggested that the residuals of the regression had phylogenetic signal
306 consistent with Brownian motion; estimates of $\lambda = 0$ suggested no phylogenetic signal. Be-
307 cause the original aim of these methods was to correct for statistical bias introduced by shared
308 evolutionary history among species, the underlying assumption of phylogenetic regressions is
309 that phylogenetic relatedness would only affect either model residuals (in PGLS approaches,
310 Freckleton et al., 2002) or model intercepts (e.g., in many PMM approaches, Housworth et al.,
311 2004).

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

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

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

where

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

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

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

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

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

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

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

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 trait k (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 , $\beta_{force,j}$, $\beta_{force,j}$, $\beta_{photo,j}$), 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:

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

331 where the residual phylogenetic error terms (e.g., e_{α_j}) are elements of normal random vec-
332 tors from multivariate normal distributions centered on 0 with the same phylogenetic variance-
333 covariance matrices as in equation 4. Model code, including priors used here, are given in the
334 supplement. We fit all models to our data using RStan using 4 chains of 4,000 iterations with
335 a warmup of 2,000 each (resulting 8,000 posterior samples), and assessed fit via \hat{R} near 1 and
336 adequate effective sample sizes (see Tables S1-S2).

337 To assess if the phylogenetic mixed model (PMM) presents any advantages with respect com-
338 monly used hierarchical mixed models (HMM; see e.g., Ettinger et al. (2020)), beyond fitting
339 evolutionary parameters to model predictors, we compare results of PMM and HMM. HMM is
340 a simplified version of PMM where off-diagonal elements of the variance-covariance phylogenetic
341 matrices are multiplied by zero ($\lambda = 0$). Since many species occurred in only one study, making
342 it difficult to separate the effects of study and species, thus we do not include study as a separate
343 parameter here and average over it in our model estimates.

344 References

- 345 Ackerly, D. 2009. Conservatism and diversification of plant functional traits: Evolutionary rates
346 versus phylogenetic signal. Proceedings of the National Academy of Sciences of the United
347 States of America 106:19699–19706.
- 348 Aitken, S. N., and J. B. Bemmels. 2016. Time to get moving: assisted gene flow of forest trees.
349 Evol Appl 9:271–90.
- 350 Azeez, A., and A. P. Sane. 2015. Photoperiodic growth control in perennial trees. Plant signaling
351 & behavior 10:e1087631.
- 352 Basler, D., and C. Körner. 2012. Photoperiod sensitivity of bud burst in 14 temperate forest
353 tree species. Agricultural and Forest Meteorology 165:73–81.
- 354 Baumgarten, F., C. M. Zohner, A. Gessler, and Y. Vitasse. 2021. Chilled to be forced: the best
355 dose to wake up buds from winter dormancy. New Phytologist 230:1366–1377.
- 356 Bennett, J. M., J. Sunday, P. Calosi, F. Villalobos, B. Martínez, R. Molina-Venegas, M. B.
357 Araújo, A. C. Algar, S. Clusella-Trullas, B. A. Hawkins, et al. 2021. The evolution of critical
358 thermal limits of life on earth. Nature communications 12:1–9.
- 359 Bonamour, S., L. M. Chevin, A. Charmantier, and C. Teplitsky. 2019. Phenotypic plasticity in
360 response to climate change: the importance of cue variation. Philosophical Transactions of
361 the Royal Society B-Biological Sciences 374.
- 362 Bradshaw, W. E., and C. M. Holzapfel. 2006. Evolutionary response to rapid climate change.
363 Science 312:1477–1478.

- 364 CaraDonna, P. J., and D. W. Inouye. 2014. Phenological responses to climate change do not
365 exhibit phylogenetic signal in a subalpine plant community. *Ecology* 96:355–361.
- 366 Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals.
367 *Annual Review of Ecology, Evolution, and Systematics* 48:159–182.
- 368 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble
369 version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical
370 Research: Atmospheres* 123:9391–9409.
- 371 Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism
372 in plant phenology. *Journal of Ecology* 101:1520–1530.
- 373 Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted
374 regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology
375 and biogeography* 28:275–285.
- 376 Dietze, M. 2017. Ecological forecasting. *In* *Ecological Forecasting*. Princeton University Press.
- 377 Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins,
378 C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to
379 community patterns. *Ecology letters* 15:545–553.
- 380 Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha,
381 and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses
382 to warming. *Nature Climate Change* pages 1–6.
- 383 Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich.
384 2021. Spatial and temporal shifts in photoperiod with climate change. *New Phytologist*
385 230:462–474.
- 386 Flynn, D. F. B., and E. M. Wolkovich. 2018. Temperature and photoperiod drive spring phe-
387 nology across all species in a temperate forest community. *New Phytologist* 219:1353–1362.
- 388 Franks, S. J., J. J. Weber, and S. N. Aitken. 2014. Evolutionary and plastic responses to climate
389 change in terrestrial plant populations. *Evolutionary Applications* 7:123–139.
- 390 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative
391 data: a test and review of evidence. *The American Naturalist* 160:712–726.
- 392 Fuccillo Battle, K., A. Duhon, C. R. Vispo, T. M. Crimmins, T. N. Rosenstiel, L. L. Armstrong-
393 Davies, and C. E. de Rivera. 2022. Citizen science across two centuries reveals phenological
394 change among plant species and functional groups in the northeastern us. *Journal of Ecology*
395 110:1757–1774.

- 396 Gelman, A., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models.
397 Cambridge University Press.
- 398 Gotelli, N. J., and G. R. Graves. 1996. The temporal niche. Pages 95–111 in N. J. Gotelli and
399 G. R. Graves, eds. Null Models in Ecology. Smithsonian Institution, Washington, DC.
- 400 Griffith, D. M., C. P. Osborne, E. J. Edwards, S. Bachle, D. J. Beerling, W. J. Bond, T. J. Gal-
401 laher, B. R. Helliker, C. E. R. Lehmann, L. Leatherman, J. B. Nippert, S. Pau, F. Qiu, W. J.
402 Riley, M. D. Smith, C. A. E. Stromberg, L. Taylor, M. Ungerer, and C. J. Still. 2020. Lineage-
403 based functional types: characterising functional diversity to enhance the representation of
404 ecological behaviour in land surface models. *New Phytologist* 228:15–23.
- 405 Grime, J. P. 1977. Evidence for existence of 3 primary strategies in plants and its relevance to
406 ecological and evolutionary theory. *American Naturalist* 111:1169–1194.
- 407 Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The
408 American Naturalist* 163:84–96.
- 409 Hunter, A. F., and M. J. Lechowicz. 1992. Predicting the timing of budburst in temperate trees.
410 *Journal of Applied Ecology* 29:597–604.
- 411 IPCC. 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge Uni-
412 versity Press, Cambridge, United Kingdom and New York, NY, USA.
- 413 Kochmer, J. P., and S. N. Handel. 1986. Constraints and competition in the evolution of flowering
414 phenology. *Ecological monographs* 56:303–325.
- 415 Kramer, K., A. Ducousoo, D. Gomory, J. K. Hansen, L. Ionita, M. Liesebach, A. Lorent,
416 S. Schueler, M. Sulkowska, S. de Vries, and G. von Wuehlisch. 2017. Chilling and forc-
417 ing requirements for foliage bud burst of European beech (*Fagus sylvatica* L.) differ between
418 provenances and are phenotypically plastic. *Agricultural and Forest Meteorology* 234:172–181.
- 419 Larcher, W. 1980. Plant Physiological Ecology. Springer-Verlag.
- 420 Laube, J., T. H. Sparks, N. Estrella, J. Höfler, D. P. Ankerst, and A. Menzel. 2014. Chilling
421 outweighs photoperiod in preventing precocious spring development. *Global Change Biology*
422 20:170–182.
- 423 Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins,
424 M. C. Dietze, A. S. Gallinat, N. Juvigny-Khenafou, et al. 2022. The power of forecasts to
425 advance ecological theory. *Methods in Ecology and Evolution* .
- 426 Molina-Venegas, R., I. Morales-Castilla, and M. Á. Rodríguez. 2023. Unreliable prediction of
427 b-vitamin source species. *Nature Plants* 9:31–33.

- 428 Molina-Venegas, R., J. C. Moreno-Saiz, I. Castro Parga, T. J. Davies, P. R. Peres-Neto, and
429 M. Á. Rodríguez. 2018. Assessing among-lineage variability in phylogenetic imputation of
430 functional trait datasets. *Ecography* 41:1740–1749.
- 431 Moorcroft, P., G. Hurtt, and S. Pacala. 2001. A method for scaling vegetation dynamics: The
432 ecosystem demography model (ed). *Ecological Monographs* 71:557–585.
- 433 Nakagawa, H., J. Yamagishi, N. Miyamoto, M. Motoyama, M. Yano, and K. Nemoto. 2005.
434 Flowering response of rice to photoperiod and temperature: a qtl analysis using a phenological
435 model. *Theoretical and Applied Genetics* 110:778–786.
- 436 Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts
437 across natural systems. *Nature* 421:37.
- 438 Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker,
439 and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. *Bioinformatics*
440 31:2888–2890.
- 441 Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering
442 time under climate change. *Journal of Ecology* 105:627–633.
- 443 Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey. 2013.
444 Climate change, phenology, and phenological control of vegetation feedbacks to the climate
445 system. *Agricultural and Forest Meteorology* 169:156–173.
- 446 Rinne, P., A. Saarelainen, and O. Juntila. 1994. Growth cessation and bud dormancy in
447 relation to ABA level in seedlings and coppice shoots of *Betula pubescens* as affected by a
448 short photoperiod, water stress and chilling. *Physiologia Plantarum* 90:451–458.
- 449 Schaber, J., and F. Badeck. 2003. Physiology-based phenology models for forest tree species in
450 germany. *International Journal of Biometeorology* 47:193–201. International Conference on
451 the Times they are a Changin, WAGENINGEN, NETHERLANDS, DEC 05-07, 2001.
- 452 Serrano-Bueno, G., F. J. Romero-Campero, E. Lucas-Reina, J. M. Romero, and F. Valverde.
453 2017. Evolution of photoperiod sensing in plants and algae. *Current Opinion in Plant Biology*
454 37:10–17.
- 455 Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global
456 dataset of meteorological forcings for land surface modeling. *Journal of Climate* 19:3088–3111.
- 457 Smith, S. A., and J. W. Brown. 2018. Constructing a broadly inclusive seed plant phylogeny.
458 *American journal of botany* 105:302–314.
- 459 Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of
460 energetic scaling across the vertebrate tree of life. *The American Naturalist* 190:185–199.

- 461 Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, perfor-
462 mance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- 463 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.
464 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism
465 as an emerging principle in ecology and conservation biology. *Ecology letters* 13:1310–1324.
- 466 Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to
467 warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*
468 111:7906–7913.
- 469 Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phyloge-
470 netic patterns of species loss in thoreau’s woods are driven by climate change. *Proceedings of*
471 *the National Academy of Sciences* 105:17029–17033.
- 472 Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and
473 communities in non-stationary environments. *Biological Reviews* .
- 474 Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and
475 J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-
476 PREE). doi:10.5063/F1CZ35KB.
- 477 Yang, Z., Y. Du, M. Shen, N. Jiang, E. Liang, W. Zhu, Y. Wang, and W. Zhao. 2021. Phy-
478 logenetic conservatism in heat requirement of leaf-out phenology, rather than temperature
479 sensitivity, in tibetan plateau. *Agricultural and Forest Meteorology* 304.
- 480 Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to
481 constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate*
482 *Change* 6:1120–1123.

⁴⁸³ **Tables and Figures**

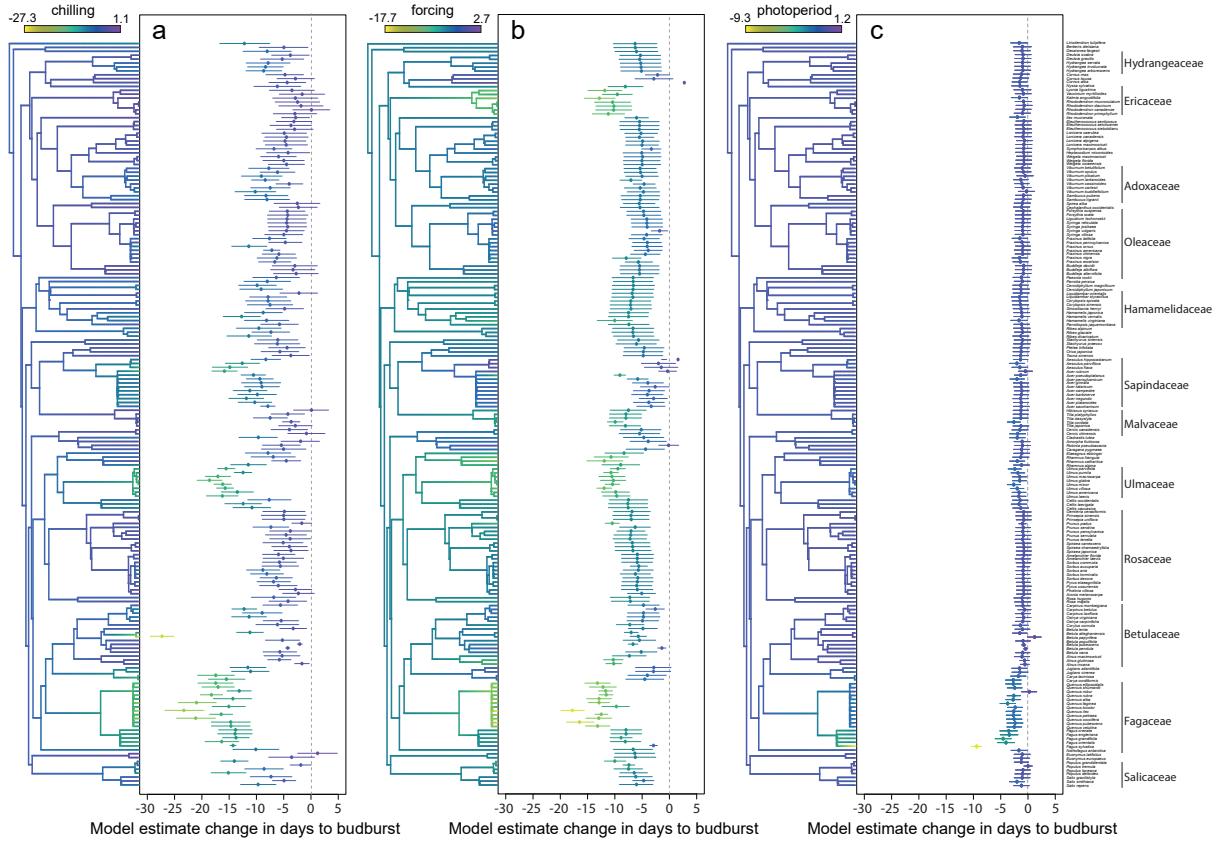


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit (z -transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored according to an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

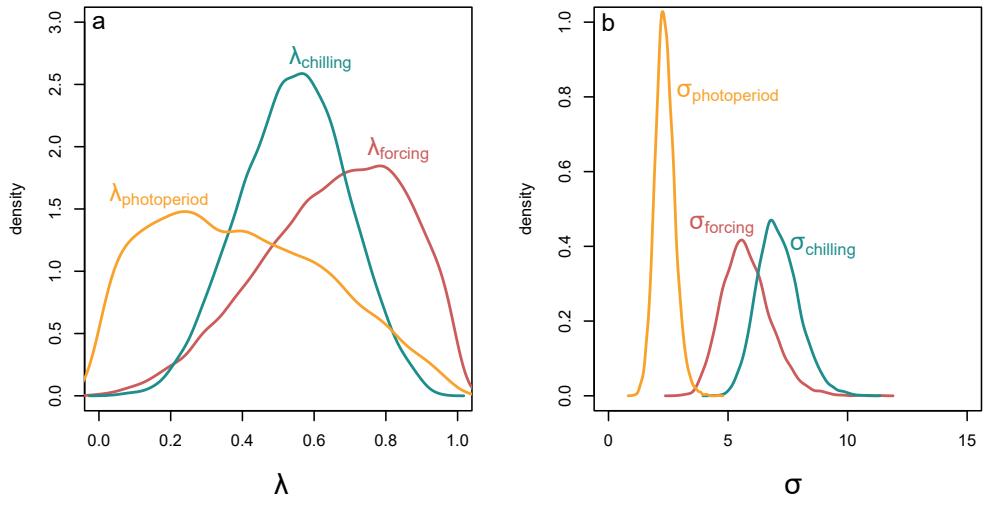


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters λ and σ estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to λ (a) and σ (b) from the phylogenetic model.

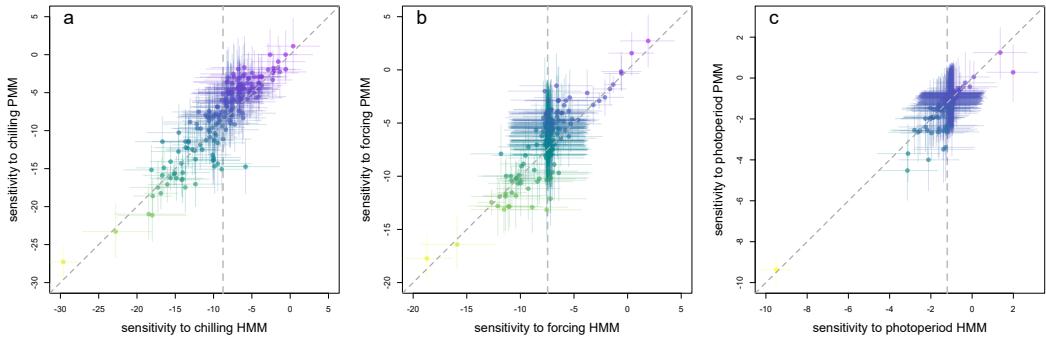


Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue (phylogenetic mixed model, PMM; y -axis), and the more commonly used hierarchical mixed model (HMM) where species are exchangeable (where λ is constrained to be equal to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

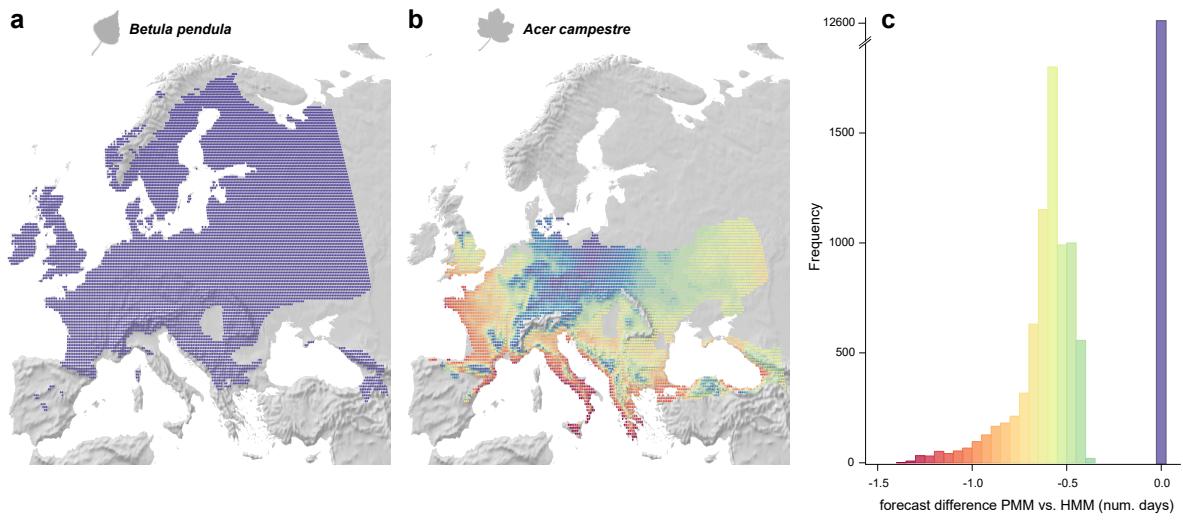


Figure 4: Comparison of forecasts of phenological shifts (i.e., computed as the difference between predictions under current climate vs. a 2°C warmer climate) resulting from a phylogenetic (PMM) and a non-phylogenetic (HMM) approach. Differences in forecasted shifts are negligible for well sampled species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