

Phylogenetic estimates of species-level phenology improve ecological forecasting

October 12, 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 Regniere, 2017;
¹⁷ Ettinger et al., 2020). In the absence of detailed data on individual species, species groupings
¹⁸ (e.g., functional groups) have improved ecosystem models (Moorcroft et al., 2001; Griffith et al.,
¹⁹ 2020), but still capture only a fraction of the important variability (Fuccillo Battle et al., 2022).

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

²⁹ Including the evolutionary history of species relationships in models of species responses could
³⁰ provide more robust species-level estimates than current approaches and a better understanding
³¹ of the evolutionary constraints that might limit adaptation to change. For example, strong
³² phylogenetic niche conservatism (Wiens et al., 2010) could potentially inhibit adaptive responses
³³ by drawing species back to an evolutionary conserved optimum, which is sub-optimal under new
³⁴ conditions. While incorporating such evolutionary history is traditionally seen as necessary
³⁵ either as a statistical correction or to better understand species' evolutionary history, the use of
³⁶ such phylogenetic information should also improve model fitting and forecasts (Freckleton et al.,
³⁷ 2002).

38 Research using long-term observational data has highlighted the role that evolutionary history
39 may play in structuring plant phenological responses—which are critical to accurate forecasts
40 of carbon storage. Phylogenetic signal in plant phenology including dates of budburst, leafout
41 and first flowering (Kochmer and Handel, 1986; Willis et al., 2008; Davies et al., 2013), suggests
42 species responses to cues have diverged over macro-evolutionary timescales, helping explain
43 species present day differences. However, approaches using traditional phylogenetic compara-
44 tive methods, have produced conflicting results, with some studies reporting evidence of phy-
45 logenetic structure in phenology-linked species declines ((?, e.g.,) illis2008phylogenetic, and in
46 some phenophases, such as first flowering, but not others, such as peak or last flowering (?,
47 e.g.,) araDonna2015, and for some cues but not others (? , e.g.,) ang2021afm. In addition, ev-
48 idence for phylogenetic conservatism of phenological responses appears to depend on method
49 and species, even varying between sites with overlapping species sets (e.g., Rafferty and Nabity,
50 2017), which violates the fundamental idea of shared evolutionary history (the common ancestor
51 of two sets of species cannot possess two separate evolutionary histories for the same trait). A
52 first challenge is thus how to better integrate evolutionary history into multi-species models of
53 plant phenological responses.

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

69 Spring plant phenology may represent our best opportunity to improve forecasts of species'
70 responses to interacting environmental cues. Beyond being the most studied biological impact
71 of climate change, the primary cue system is well established (Chuine and Regniere, 2017),
72 especially for temperate woody species where phenology is generally thought to be determined
73 by two components of temperature—chilling (cool temperatures during dormancy period over
74 winter) and forcing (warm temperatures, generally in the spring)—and photoperiod (Ettinger
75 et al., 2021). Plant phenology is also one of few phenotypic traits with extensive experimental
76 data on responses to multiple environmental cues across species. Recent multi-species analyses

⁷⁷ considering forcing, chilling and photoperiod have shown that chilling and forcing together often
⁷⁸ determine complex non-linear responses to warming, but cannot forecast beyond several well-
⁷⁹ studied species (Ettinger et al., 2020).

⁸⁰ Here we present a novel Bayesian framework that extends upon phylogenetic mixed models
⁸¹ (Housworth et al., 2004) to examine how chilling, forcing (both metrics of temperature) and
⁸² photoperiod together determine spring plant phenology. By allowing non-stationarity in species
⁸³ responses across the phylogeny (Davies et al., 2019), our model departs from previous work
⁸⁴ and assumptions of traditional phylogenetic comparative methods concerned with phylogenetic
⁸⁵ correction (e.g. Freckleton et al., 2002), and moves towards integrating evolutionary history
⁸⁶ in models of phenological responses to environmental change. To understand how evolution
⁸⁷ has shaped the cues underlying shifting phenology with climate change (Uyeda et al., 2017), we
⁸⁸ explicitly incorporate phylogenetic structure across model intercepts and slopes (that is, allowing
⁸⁹ a separate model of evolutionary history for chilling, forcing and photoperiod, see Methods &
⁹⁰ Materials for a complete description).

⁹¹ We illustrate our method with an unprecedented dataset on phenological responses to environ-
⁹² mental cues (chilling, forcing and photoperiod) determined experimentally for 191 deciduous
⁹³ woody species (by far the most studied group of species in phenology experiments, see Ettinger
⁹⁴ et al., 2020), in an updated version of the Observed Spring Phenology Responses in Experi-
⁹⁵ mental Environments (OSPREE) database (Wolkovich et al., 2019). These data combined with
⁹⁶ the Smith and Brown (2018) megatree adjusted to our species and modeling approach allows
⁹⁷ us to address the common question of which cue has the largest effect on budburst and, at the
⁹⁸ same time, provide robust estimates of how cues vary across species. Using spring phenology,
⁹⁹ we identify historical regime shifts (Uyeda et al., 2017) in phenological responses, and highlight
¹⁰⁰ how our approach could advance forecasting of other critical responses to ongoing global change.

¹⁰¹ Results & Discussion

¹⁰² Most species respond to all three primary cues—forcing, chilling, and photoperiod (Fig. 1,
¹⁰³ Supporting Table ??)—with responses to chilling approximately five-fold greater than to pho-
¹⁰⁴ toperiod (phenological advances of 6.9 days per standardized unit vs 1.2 days, for chilling and
¹⁰⁵ photoperiod, respectively; see Table ??). We estimated lower average responses to temperature
¹⁰⁶ compared to a model without phylogeny (model slopes for forcing and chilling decreased by 18%
¹⁰⁷ and 22%, respectively); responses to chilling and forcing were also more similar when including
¹⁰⁸ phylogeny (though chilling was still greater: 6.9 vs. 6.1 per standard unit), which contrasts with
¹⁰⁹ previous results suggesting chilling responses are much greater than forcing (Laube et al., 2014;
¹¹⁰ Ettinger et al., 2020).

¹¹¹ These average estimates, however, fail to capture the large differences in species' responses
¹¹² to both chilling and forcing (Fig. 1, Supporting Table ??). By allowing species responses to

113 vary, based on a model including their shared evolutionary history, we found species differences
114 dwarfed the mean differences between cues, especially temperature cues (Fig. 1). The largest cue
115 in magnitude—chilling—varied 24-fold between species, while variation to forcing varied 7-fold.
116 This variation indicates large differences between chilling and forcing occur at the species-level
117 and not across species (as previously suggested Laube et al., 2014; Ettinger et al., 2020). These
118 results highlight why robust phenological forecasts must account for both the complexity of
119 multiple cues and species-level variation in responses to them.

120 *Differences across clades & cues*

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

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

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

species, our results caution against using it to draw inferences of photoperiod responses more widely. These same few species are also where most evidence of local adaptation in photoperiod cues for spring phenology comes from (e.g., Kramer et al., 2017), in contrast with common garden studies of other species, which find little evidence of local adaptation in spring (but not fall) phenology (Aitken and Bemmels, 2016). The uniformity of response to photoperiod in our results supports this latter view of generally low local adaptation in photoperiod cues for spring phenology (i.e., if local adaptation were high in photoperiod cues we would have expected more variability across species).

159 *Phylogenetic structure of phenological cues*

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

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

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

189 our results suggest photoperiod responses are also evolving slower than temperature responses
190 (see Fig. ??).

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

204 *Forecasting species-level responses*

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

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

224 The increase in variability across species in our model with phylogenetic structure also decreased
225 the uncertainty in estimates for each individual species temperature responses (Fig. ??). Thus,
226 traditional (non-phylogenetically informed) approaches that partially pool across species (most

227 hierarchical models in ecology, e.g., Flynn and Wolkovich, 2018; Ettinger et al., 2020) may also
228 lead to less precise predictions and forecasts of phenology for individual species, although overall
229 model accuracy might still appear reasonable (see Fig. ??).

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

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

248 **Methods & Materials**

249 **Phenological and Phylogenetic Data**

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

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

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

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

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

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

286 For a phylogenetic tree, we pruned the phylogenetic megatree for seed plants (Smith and Brown,
287 2018) to extract a subset of the phylogenetic tree containing only the angiosperm species in the
288 OSPREE dataset, then added species that were not present in the megatree as polytomies (i.e.,
289 comb-like phylogenetic relationships assuming that species within the polytomy originated simul-
290 taneously from their common ancestor) at the generic level (using the function ‘congeneric.merge’
291 in Pearse et al., 2015), with a branch length of the congeneric basal node age. Our pruned tree
292 had 8 Polytomies affecting 46 out of 191 species. To test for the ability of polytomies to bias our
293 results we ran sensitivity analyses excluding these species from models (see Supporting Infor-
294 mation). Our approach uses a tree where time (along phylogenetic branch lengths) is assumed
295 as a valid proxy for species differences, but it could be possible to change this assumption. For
296 example, if the genes underlying plant responses to particular cues were known, branch lengths
297 would directly represent mutational changes along gene sequences. In the absence of such de-
298 tailed gene specific data, evolutionary time provides a useful proxy for species differences.

299

300 Bayesian hierarchical phylogenetic model

301 Commonly used phylogenetic regression methods today (e.g., phylogenetic generalized least
302 squares models, PGLS, Freckleton et al. (2002); phylogenetic mixed models, PMM, Hous-
303 worth et al. (2004)) were originally conceived as statistical corrections for phylogenetic non-
304 independence across observations—generally species—thus allowing multi-species studies to meet
305 the assumptions of linear regression (Freckleton et al., 2002). These corrections incorporated
306 phylogenetic structure by estimating the magnitude of a transformation of a variance-covariance
307 (VCV) matrix whose elements were derived from the amount of evolutionary history (branch
308 lengths) shared between species on a phylogeny. The most commonly used transformation was
309 Pagel’s λ —a multiplier of the off-diagonal elements—where estimates of $\lambda = 1$ essentially left the
310 VCV untransformed and suggested that the residuals of the regression had phylogenetic signal
311 consistent with Brownian motion; estimates of $\lambda = 0$ suggested no phylogenetic signal. Be-
312 cause the original aim of these methods was to correct for statistical bias introduced by shared
313 evolutionary history among species, the underlying assumption of phylogenetic regressions is
314 that phylogenetic relatedness would only affect either model residuals (in PGLS approaches,
315 Freckleton et al., 2002) or model intercepts (e.g., in many PMM approaches, Housworth et al.,
316 2004).

317 Because our aim is to understand how evolution may have imprinted biological responses to
318 multiple interactive cues, our approach expands the above methods by explicitly incorporating
319 phylogenetic structure across model intercepts and slopes. Doing so allows explicitly estimating

320 the amount of phylogenetic relatedness in species' sensitivities to each cue, when these sensitiv-
321 ities are modelled in a multi-predictor regression setting.

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

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

where

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

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

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

$$\begin{aligned} \boldsymbol{\alpha} &= [\alpha_1, \dots, \alpha_n]^T \text{ such that } \boldsymbol{\alpha} \sim \mathcal{N}(\mu_\alpha, \Sigma_{\boldsymbol{\alpha}}) \\ \boldsymbol{\beta}_{chill} &= [\beta_{chill,1}, \dots, \beta_{chill,n}]^T \text{ such that } \boldsymbol{\beta}_{chill} \sim \mathcal{N}(\mu_{\beta_{chill}}, \Sigma_{\boldsymbol{\beta}_{chill}}) \\ \boldsymbol{\beta}_{force} &= [\beta_{force,1}, \dots, \beta_{force,n}]^T \text{ such that } \boldsymbol{\beta}_{force} \sim \mathcal{N}(\mu_{\beta_{force}}, \Sigma_{\boldsymbol{\beta}_{force}}) \\ \boldsymbol{\beta}_{photo} &= [\beta_{photo,1}, \dots, \beta_{photo,n}]^T \text{ such that } \boldsymbol{\beta}_{photo} \sim \mathcal{N}(\mu_{\beta_{photo}}, \Sigma_{\boldsymbol{\beta}_{photo}}) \end{aligned} \quad (3)$$

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

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

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

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

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

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

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

349 References

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

- 367 Bradshaw, W. E., and C. M. Holzapfel. 2006. Evolutionary response to rapid climate change.
368 Science 312:1477–1478.
- 369 Chuine, I., and J. Regniere. 2017. Process-based models of phenology for plants and animals.
370 Annual Review of Ecology, Evolution, and Systematics 48:159–182.
- 371 Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble
372 version of the E-OBS temperature and precipitation data sets. Journal of Geophysical
373 Research: Atmospheres 123:9391–9409.
- 374 Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism
375 in plant phenology. Journal of Ecology 101:1520–1530.
- 376 Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted
377 regression: A method for modelling non-stationarity on evolutionary trees. Global ecology
378 and biogeography 28:275–285.
- 379 Dietze, M. 2017. Ecological forecasting. In Ecological Forecasting. Princeton University Press.
- 380 Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins,
381 C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: from species variability to
382 community patterns. Ecology letters 15:545–553.
- 383 Ettinger, A., C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. Flynn, T. Savas, J. Samaha,
384 and E. Wolkovich. 2020. Winter temperatures predominate in spring phenological responses
385 to warming. Nature Climate Change pages 1–6.
- 386 Ettinger, A. K., D. M. Buonaiuto, C. J. Chamberlain, I. Morales-Castilla, and E. M. Wolkovich.
387 2021. Spatial and temporal shifts in photoperiod with climate change. New Phytologist
388 230:462–474.
- 389 Flynn, D. F. B., and E. M. Wolkovich. 2018. Temperature and photoperiod drive spring phe-
390 nology across all species in a temperate forest community. New Phytologist 219:1353–1362.
- 391 Franks, S. J., J. J. Weber, and S. N. Aitken. 2014. Evolutionary and plastic responses to climate
392 change in terrestrial plant populations. Evolutionary Applications 7:123–139.
- 393 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative
394 data: a test and review of evidence. The American Naturalist 160:712–726.
- 395 Fuccillo Battle, K., A. Duhon, C. R. Vispo, T. M. Crimmins, T. N. Rosenstiel, L. L. Armstrong-
396 Davies, and C. E. de Rivera. 2022. Citizen science across two centuries reveals phenological
397 change among plant species and functional groups in the northeastern us. Journal of Ecology
398 110:1757–1774.

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

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

- 464 Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, perfor-
465 mance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- 466 Wiens, J. J., D. D. Ackerly, A. P. Allen, B. L. Anacker, L. B. Buckley, H. V. Cornell, E. I.
467 Damschen, T. Jonathan Davies, J.-A. Grytnes, S. P. Harrison, et al. 2010. Niche conservatism
468 as an emerging principle in ecology and conservation biology. *Ecology letters* 13:1310–1324.
- 469 Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to
470 warming climate in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*
471 111:7906–7913.
- 472 Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phyloge-
473 netic patterns of species loss in thoreau’s woods are driven by climate change. *Proceedings of*
474 *the National Academy of Sciences* 105:17029–17033.
- 475 Wolkovich, E. M., and M. J. Donahue. 2021. How phenological tracking shapes species and
476 communities in non-stationary environments. *Biological Reviews* .
- 477 Wolkovich, E. M., A. K. Ettinger, D. Flynn, T. Savas, C. Chamberlain, D. Buonaiuto, and
478 J. Samaha. 2019. Observed Spring Phenology Responses in Experimental Environments (OS-
479 PREE). doi:10.5063/F1CZ35KB.
- 480 Zohner, C. M., B. M. Benito, J. C. Svensson, 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**

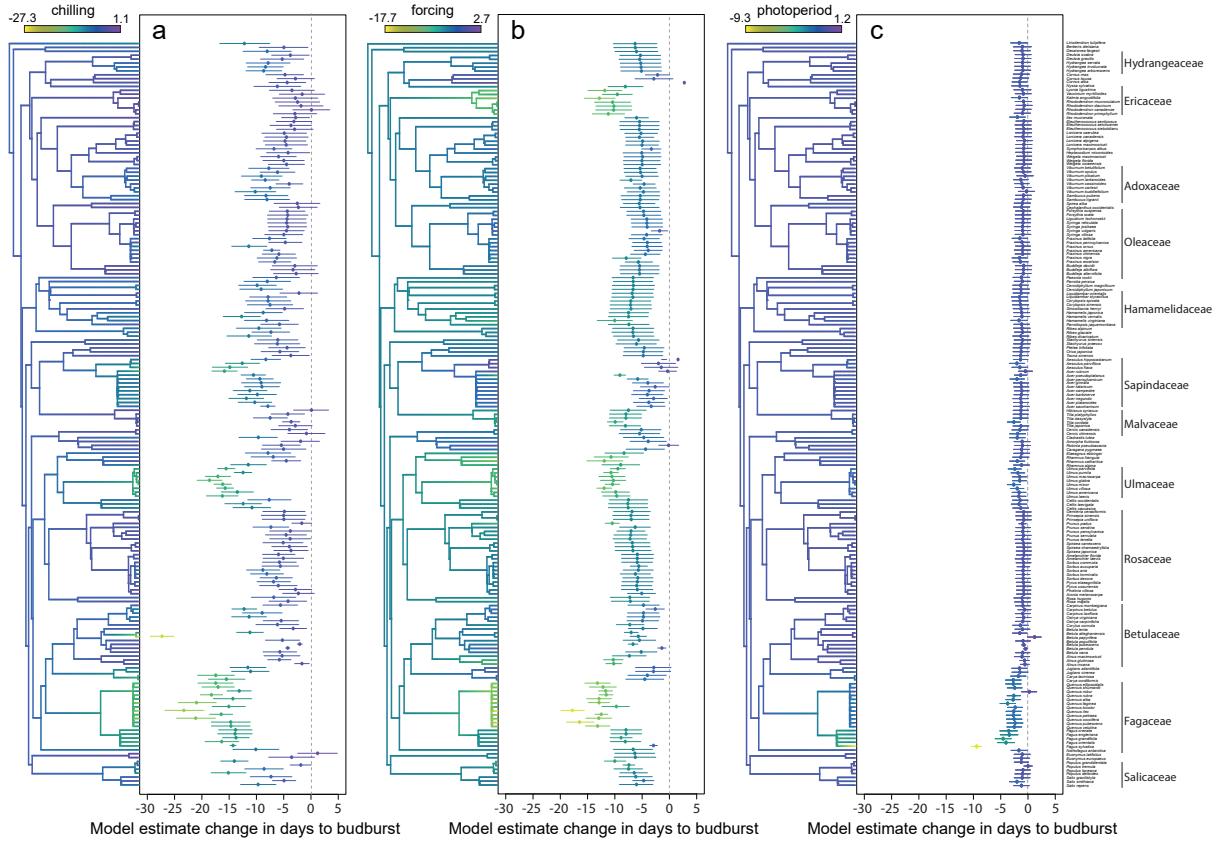


Figure 1: Phenological sensitivity to three environmental cues, chilling (a), forcing (b) and photoperiod (c) measured as change in days to budburst per standardized unit (z -transformation) of the cues across 191 tree species. The same phylogenetic tree is shown in each panel, colored according to an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

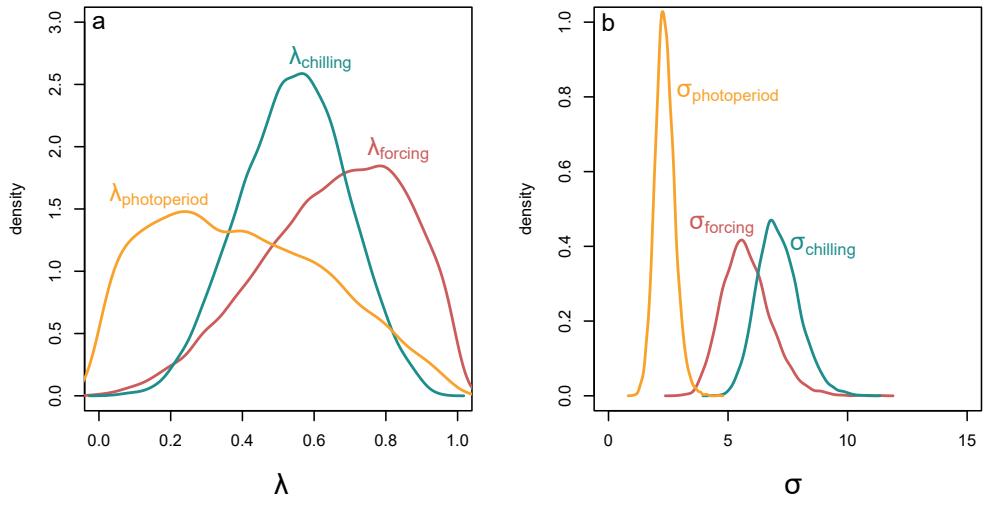


Figure 2: Density plots comparing the posterior distributions of phylogenetic parameters λ and σ estimated for each cue in the model: chilling (blue), forcing (red), and photoperiod (orange). Panels correspond to λ (a) and σ (b) from the phylogenetic model.

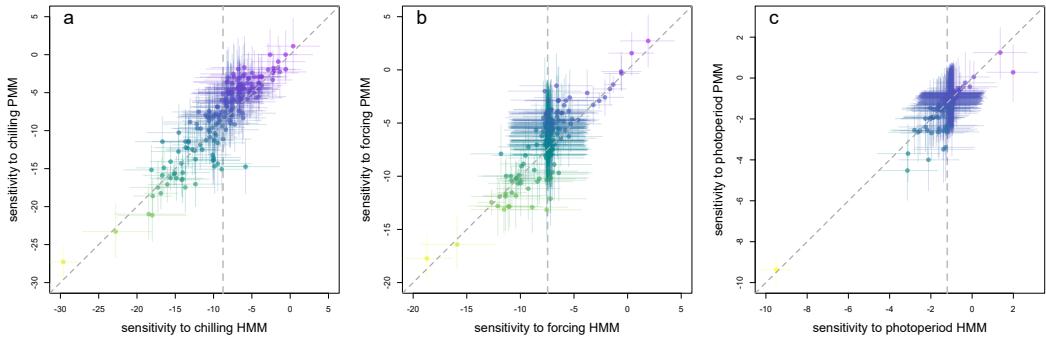


Figure 3: Correlations between model parameters as estimated by the model including phylogenetic structure on each phenological cue (phylogenetic mixed model, PMM; y -axis), and the more commonly used hierarchical mixed model (HMM) where species are exchangeable (where λ is constrained to be equal to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

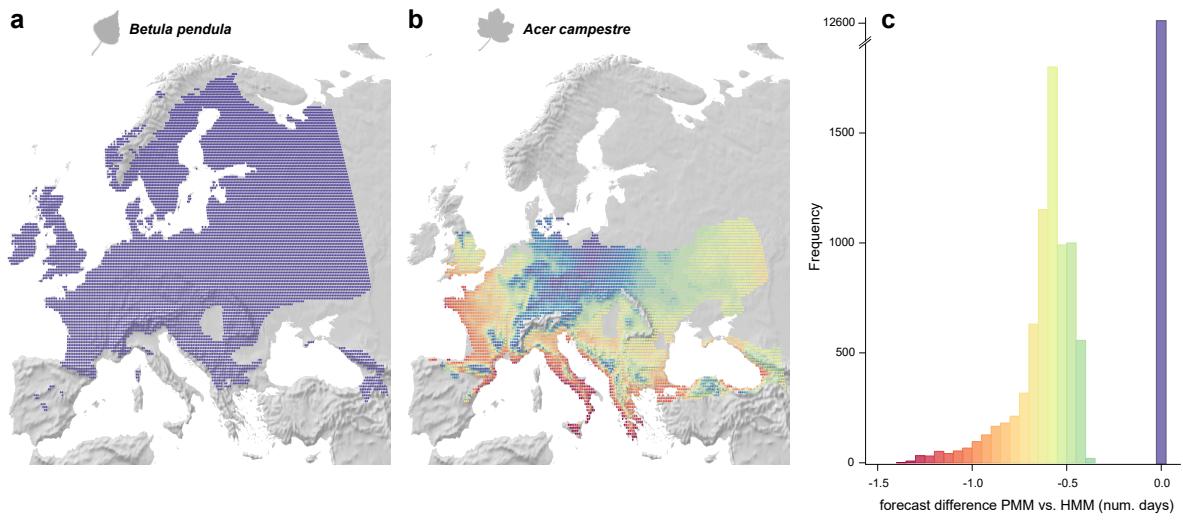


Figure 4: Comparison of forecasts of phenological shifts (i.e., computed as the difference between predictions under current climate vs. a 2°C warmer climate) resulting from a phylogenetic (PMM) and a non-phylogenetic (HMM) approach. Differences in forecasted shifts are negligible for well sampled species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