nature climate change

ARTICLES

https://doi.org/10.1038/s41558-020-00917-3

Check for updates

Winter temperatures predominate in spring phenological responses to warming

A. K. Ettinger[1]✉, C. J. Chamberlain[1,2], I. Morales-Castilla[1,2,3], D. M. Buonaiuto[1,2], D. F. B. Flynn[1,2,4], T. Savas[1,2,5], J. A. Samaha[1,2,6] and E. M. Wolkovich[1,2,7]

Research on woody plant species highlights three major cues that shape spring phenological events: chilling, forcing and photoperiod. Increasing research on the phenological impacts of climate change has led to debate over whether chilling and/or photoperiod cues have slowed phenological responses to warming in recent years. Here we use a global meta-analysis of all published experiments to test the relative effects of these cues. Almost all species show strong responses to all three cues, with chilling being the strongest and photoperiod the weakest. Forecasts from our findings for Central Europe suggest that spring phenology will continue to advance, as stalling effects of chilling generally appear above 4 °C warming in this region. Our results unify both sides of the debate over phenological cues: while all species may respond to all cues strongly in experimental conditions, in current environmental conditions the dominant signal of climate change is from increased forcing.

For decades, plant phenology has been one of the most reported and consistent biological imprints of climate change[1]: many temperate plants are leafing and flowering days to weeks earlier with rising temperatures[2,3]. Understanding such shifts is important as phenology shapes community assembly and a suite of ecosystem services, including pollination and carbon sequestration, and scales up to impact projections of climate change itself[4].

As research interest in phenology has progressed, critical discrepancies and uncertainties in our understanding have emerged. Although responses to warming are widespread, showing strong advances on average, there is substantial variation among species and sites[5]. Furthermore, long-term observational studies provide increasing evidence that sensitivities of phenology to temperature are weakening in recent decades[6–8], especially in Europe, where researchers suggest that responses to multiple environmental cues underlie declining temperature sensitivities[9].

Fundamental research in phenology outlines how three major environmental cues, chilling (cool temperatures, generally occurring in the autumn through late winter), forcing (warm temperatures, generally occurring in the late winter through early spring) and photoperiod (daylength), provide multiple routes to budburst each spring, depending on the environment[10]. For example, in some species a cool winter will lower the amount of forcing required to trigger budburst, compared to a warmer winter[11]. Additionally, photoperiod may trigger budburst, given low chilling and/or forcing[12–14]. Research suggests that all three cues may affect spring phenology for many temperate woody species[13,15,16], which could have critical forecasting implications—predicting delays in spring phenology as increased warming reduces chilling in some areas[17] or where earlier budburst shortens the experienced photoperiod. However, there is strong debate, with research suggesting some cues (such as photoperiod) may be effectively absent in some species but dominate in others[12,13,18,19].

Resolving this debate requires overcoming major hurdles to estimate responses to each cue. Studies attempting to estimate cues using long-term observational data (for example, refs. [12,20]) generally fail to overcome the fundamental challenge that cues are strongly correlated in nature. For example, during the seasonal transition from winter to spring at temperate latitudes, forcing and photoperiod usually increase in step for a given location. Average chilling and spring (forcing) temperatures can be positively correlated in space, especially at high latitudes. In contrast to observational studies and field warming studies designed to test higher temperatures in natural conditions[5], experiments using controlled temperature and photoperiod conditions can break down correlations between the cues. These experiments, which generally rely on dormant tree cuttings or dormant plants exposed to temperature and light regimes in growth chambers (Fig. 1), have been shown to replicate whole-plant responses in nature[21]. Such experiments have been conducted for decades (although each experiment generally lasts under a year). They have produced contrasting results, however, potentially due to differences in focal species or study sites[12,14,22–24]. Resolving these discrepancies is critical to accurate predictions of spring phenology, especially as continued climate change will yield warmer temperatures than have been experienced in at least the last 150 years[25–29].

Here, we leverage these short-term controlled environment experiments in a meta-analysis to understand how chilling, forcing and photoperiod determine budburst timing in woody species. We reviewed 201 papers, extracting data from all experiments that reported budburst responses, yielding data from 72 studies and 203 species (Extended Data Fig. 1 and Supplementary Tables 1 and 2). The resulting Observed Spring Phenology Responses in Experimental Environments (OSPREE) database includes studies of dormant plant tissue (grown in greenhouses or taken directly from the field) exposed to experimental temperature and/or photoperiod conditions[30] for which we could identify chilling, forcing and

[1]Arnold Arboretum of Harvard University, Boston, MA, USA. [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [3]GloCEE - Global Change Ecology and Evolution Group, Department of Life Sciences, University of Alcalá, Alcalá de Henares, Spain. [4]US Department of Transport, Volpe National Transportation Systems Center, Cambridge, MA, USA. [5]Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Forest Resources Management, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada. [7]Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada. ✉e-mail: ailene.ettinger@tnc.org

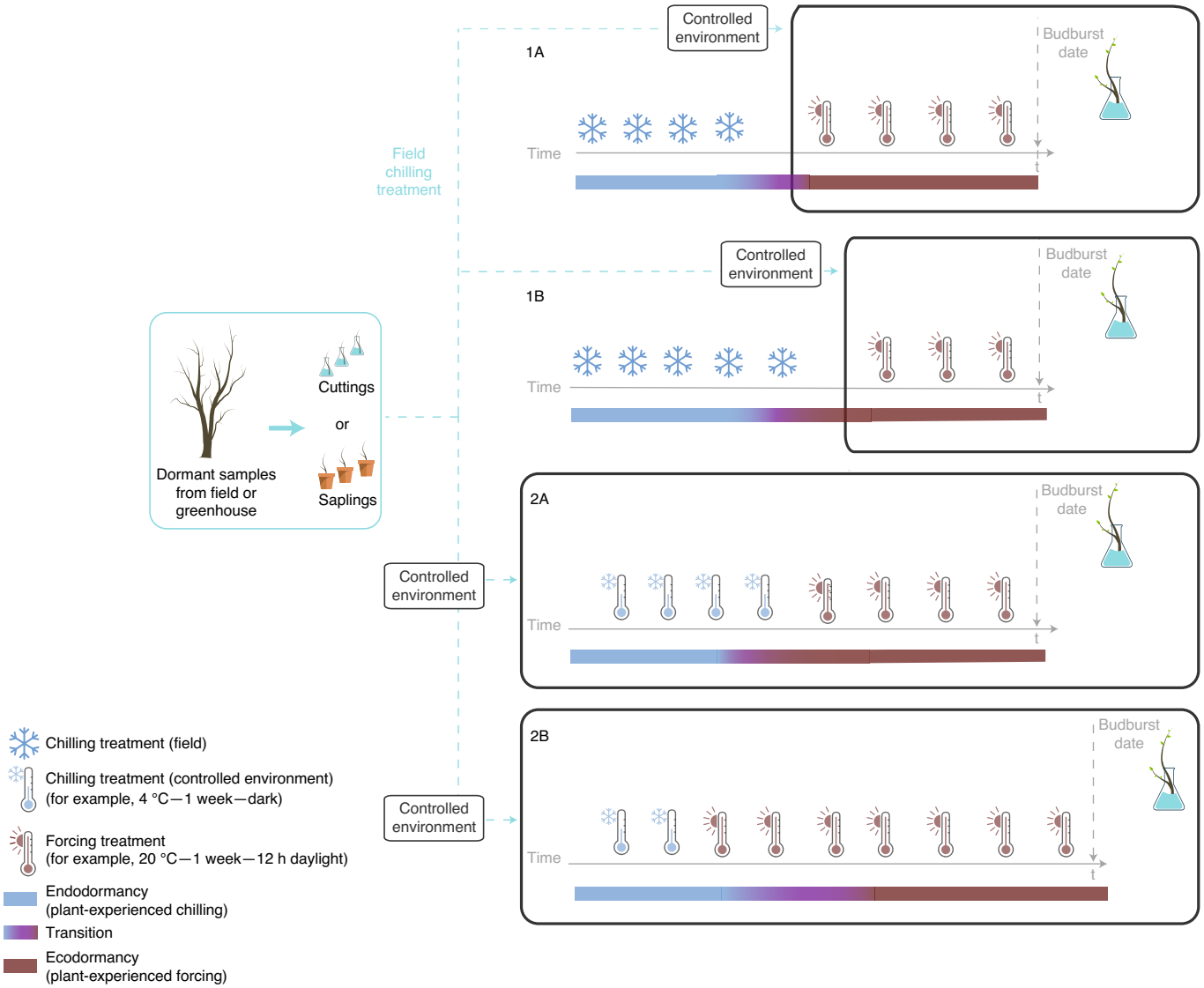


Fig. 1 | Short-term experiments on woody plant phenology manipulate photoperiod and temperature to estimate chilling, forcing and photoperiod cues. Chilling is manipulated by using natural chilling in the field (1A and 1B, in which plant material is collected after different numbers of days in the autumn/winter) and/or experimentally (2A and 2B, in which plant material is placed in controlled environment chambers set to different chilling temperatures and/or durations). Chilling treatments are designed to break plant endodormancy, after which forcing treatments are imposed by moving plant material to warmer temperatures that allow budburst to occur. Ideally, this experimental transition aligns with the physiological shift from endo- to ecodormancy (for example, 1A, although it could also occur with experimentally applied chilling). A challenge with these experiments is that species-specific chilling requirements are rarely known, so experimental treatments may not always align with what the plant experiences (physiological shifts in dormancy). Thus, in some cases, chilling treatments may bridge across what plants experience as both chilling and forcing (1B and 2A, where plants transition into ecodormancy before 'forcing' treatments are applied) or chilling treatments may end before endodormancy is fully broken (2B). In the experiments synthesized here, photoperiod (not shown) is most often manipulated in forcing treatments. Across the 39 studies (found in 28 papers) included in our main model, we found that treatments varied uniquely for each study but some were more common than others, see Extended Data Fig. 3: chilling treatments averaged 71.4 d (range 1–182 d) at an average temperature of 4.4 °C (range 0–16 °C), forcing treatments averaged 15.7 °C (range 5–32 °C).

photoperiod treatments quantitatively (these varied by each study; Extended Data Figs. 2 and 3). Most experiments reported forcing and photoperiod treatments, whereas chilling occurred mainly in the field, although some studies additionally applied chilling before moving plants into forcing conditions (Fig. 1). Because chilling was rarely reported, we calculated an estimate of chilling (both in the field and in experimental conditions), using a common approximation[31], based on a hypothesis of how chilling accumulates[32], with no chilling accumulating below 1.4 °C or above 12.4 °C (throughout the main text we use the term 'chill unit'; Supplementary Information, especially Supplementary Table 4).

We estimated the effects of chilling, forcing and photoperiod using a Bayesian hierarchical model. Our model averages over interactive effects of predictors, including only main effects that we could more robustly estimate given current study designs (Methods). Species are modelled hierarchically, producing estimates of both species-level responses (generally yielding more accurate estimates for well-studied species, such as *Fagus sylvatica* and *Betula pendula*)

and the distribution from which they are drawn, yielding estimates of the overall responses across species (Methods):

$$y_i = \alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α and each of the three β coefficients were modelled at the species level, as follows:

$$\begin{aligned} \alpha_{sp} &\sim N(\mu_\alpha, \sigma_\alpha) \\ \beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing}) \\ \beta_{photoperiod_{sp}} &\sim N(\mu_{photoperiod}, \sigma_{photoperiod}) \\ \beta_{chilling_{sp}} &\sim N(\mu_{chilling}, \sigma_{chilling}) \end{aligned}$$

where i represents each unique observation, sp is the species (or species complex grouping, explained below), α represents the intercept, β terms represent slope estimates and y is the days to budburst since forcing conditions were applied. Some species were represented in only one dataset in the OSPREE database, making it difficult to differentiate statistically between species, study and treatment effects for these taxa. To address this, we focus on estimates (reported as mean with 95% uncertainty intervals, unless otherwise noted) from a model of 65 species, which were included in multiple datasets and treatments (generally this occurred at the species level but in some cases we collapsed species found in only one study into 'complexes' at the level of genera; see OSPREE database). Estimates from this model were generally similar to estimates from a model of all 203 species (Supplementary Tables 5 and 6). To directly compare the effects of chilling, forcing and photoperiod we fit models using standardized predictor variables (following ref. [33] which we refer to as 'standard units') as well as with predictors in their natural units (chill units, °C and hours). We further fit several additional models, including a model testing provenance latitude effects, one testing effects of chilling study design and one testing effects of life stage (see Models for model equations and other details).

Across experiments, all cues—chilling, forcing and photoperiod—advance budburst phenology (Fig. 2 and Supplementary Tables 5 and 6). Chilling was the strongest cue (−8.35 d per standard unit (−11.43 to −5.36) or −2.76 d per chill unit (−3.65 to −1.89)), followed by forcing (−4.35 d per standard unit (−6.65 to −1.92) or −0.8 d per °C of warming (−1.18 to −0.43)) and photoperiod (−2.95 d per standard unit (−5.46 to −0.48) or −0.53 d per hour of daylight (−0.92 to −0.15); Fig. 3, Supplementary Figs. 2 and 4 and Supplementary Tables 5 and 6). While photoperiod had the smallest effect among the three cues, our results contrast with the extensive literature suggesting photoperiod is an unimportant cue for many species[12,34]—instead we found that it was surprisingly large, even when accounting for its interaction with provenance latitude (the latitude of origin for plant material; Extended Data Figs. 5 and 7 and Supplementary Table 10). It was also generally consistent across species (variance = 5.18 d per standard unit), only deviating in *F. sylvatica*, a species well-known for having a large response to photoperiod (which we also found; Fig. 2 and Extended Data Fig. 5). Species responses to chilling were slightly more variable (variance = 7.21 d per standard unit; Fig. 2) than responses to forcing (variance = 5.72 days per standard unit; Fig. 2 and Supplementary Table 5).

As temperature is fundamentally altered by climate change, our finding that different ends of the temperature spectrum—chilling and forcing—have the strongest effects on budburst suggests that understanding these two cues will be critical for forecasting phenology with climate change. Many previous studies attribute advances in budburst to increased forcing[3,11,13,35]. Our results, however, suggest that, across 65 species and 72 experiments, chilling has a greater effect on budburst than forcing (Fig. 2, Extended Data Fig. 5, Supplementary Fig. 3 and Supplementary Tables 5–10). This has not been widely suggested previously, perhaps because little experimental work has directly manipulated chilling and the few studies that have were designed to compare chilling versus photoperiod effects (for example, refs. [12,13,16,22]), not chilling versus forcing effects. Process-based phenological models, however, that explicitly model chilling often find this cue to be most critical (for example, refs. [22,36,37]).

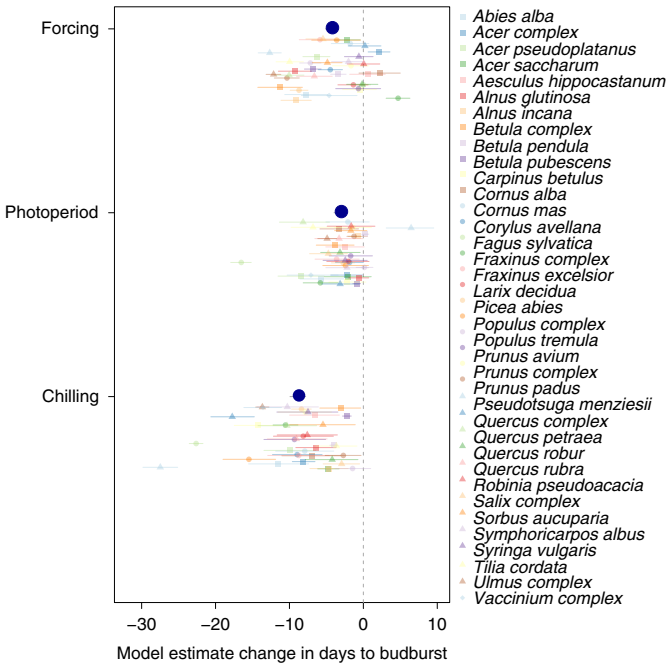


Fig. 2 | Estimated effects of chilling, forcing and photoperiod on budburst timing across 65 species. The species were modelled as 36 separate taxa ('Models' section of Methods) in 39 controlled environment experiments. Using standardized units, which allow comparisons across cues, we show that most species (smaller symbols) are responsive to most cues, with chilling being the strongest cue when considering overall estimates across species (larger, dark blue circles). Overall estimates shown here were generally similar to other model formulations, including using data from 203 species (and 72 studies) and using different methods for calculating chilling (Extended Data Fig. 5, Supplementary Fig. 3 and Supplementary Tables 5–12). Lines represent 50% uncertainty intervals (other intervals provided in Supplementary Tables 5–12).

Despite its apparent importance, chilling and its related physiological stage of endodormancy, are not well understood[10]. Physiologically, plants appear to accumulate forcing only after they have exited endodormancy (and entered ecodormancy; Fig. 1), which is generally thought to occur when chilling requirements have been met[10]. Thus, while researchers generally define 'chilling' and 'forcing' treatments on the basis of temperatures in controlled experiments (including in the studies synthesized here; Fig. 1), fully separating out what plants experience as chilling versus forcing (as well as how this varies across species and sites) will probably require new methods to measure endodormancy and ecodormancy[38].

Until then, researchers must generally rely on modelled estimates of chilling, as we have used here. While we found that applying a different chilling model did not strongly affect our estimates (95% uncertainty intervals of estimates for chilling, photoperiod and forcing effects overlapped using two different chill metrics, Utah and

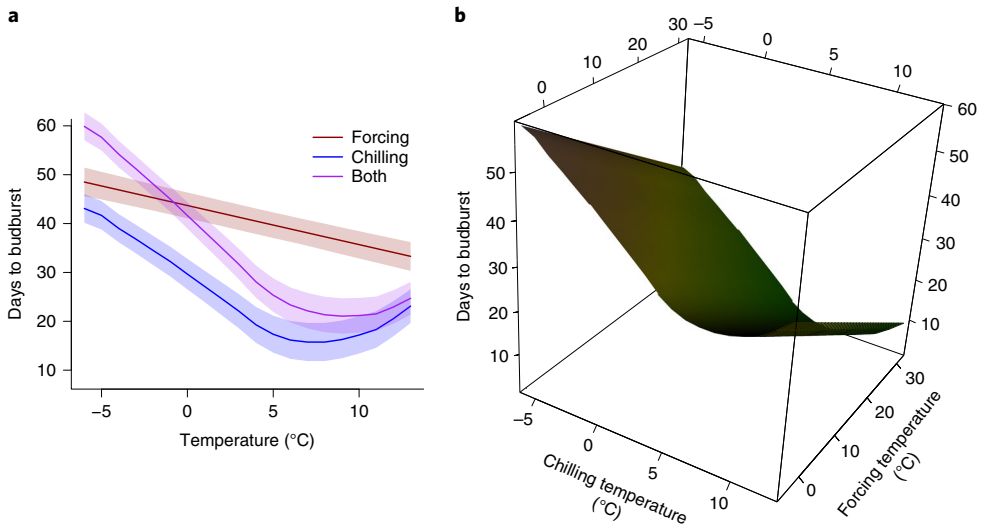


Fig. 3 | Estimates of budburst across a range of forcing temperatures and estimated chilling. Results were converted to a representative mean temperature ('Estimating chilling' in Methods) on the basis of overall estimates of chilling and forcing effects from a meta-analysis of short-term experiments using controlled temperature and/or photoperiod conditions (Fig. 2). Note that days to budburst is relative to experimental methods and thus not comparable to day of year in the field; shading in **a** represents 50% uncertainty intervals. **a**, The effect of chilling temperature on budburst, with forcing kept at the mean level across all experiments (16 °C); the effect of forcing temperature with chilling kept at the mean level across all experiments (1,324 chilling units) and the effect of varying both chilling and forcing temperatures simultaneously. **b**, All possible combinations of chilling and forcing across the experimental conditions. Maximum advances in budburst occur at intermediate chilling temperatures (for example, here at mean winter temperatures of 6–7 °C) and the highest forcing (here at 32 °C). We set photoperiod to 8 h, which is the most common photoperiod treatment in our meta-analysis.

chill portions, and the mean posterior of these estimates varied by about 10% or less between the two metrics; Supplementary Table 5), models of how species accumulate chilling are still poorly developed for forest trees. To date, there have been relatively few tests of the particular temperatures at which species do or do not accumulate chilling. Instead, researchers generally rely on models developed for perennial fruit trees (Utah units; Supplementary Table 4; ref. [31]) and chill portions[39]. These models are themselves hypotheses for how chilling may accumulate and lead to dormancy release but are likely to be inaccurate for many species[32].

Progress on developing chilling models for wild species may be especially slow as only a small portion of studies (13 of the total 72 studies) manipulated chilling directly. Instead many studies estimated chilling effects through sequential removal of tissue from the field followed by exposure to 'forcing' conditions (Fig. 1, 1A and 1B, 25 out of 72 studies; the remaining 34 studies did not appear to manipulate chilling), with the assumption that tissues collected later experience more chilling[40]. This method benefits from more natural chilling conditions but introduces other challenges: first, chilling duration may not always co-vary with the magnitude of total accumulated chilling[32] and, second, photoperiod and other factors also change across the season. Indeed, we found that sequential-removal studies tended to result in later budburst, weaker effects of chilling and stronger effects of forcing compared to estimates from studies that directly manipulated chilling (Supplementary Fig. 3 and Supplementary Table 11; refs. [40,41]). This suggests that a study's design of chilling manipulation impacts both forcing and chilling estimates and further supports that an improved understanding of chilling could in turn alter our understanding of forcing.

Linking such short-term controlled experiments to natural environmental conditions robustly will require more efforts to understand the complex interactions between chilling, forcing and photoperiod that we were not able to quantify in this meta-analysis. Most experimental studies do not test for interactions among all three cues (Supplementary Table 3). Further, many additional factors can affect phenological responses, including ontogeny (Supplementary Table 12; ref. [42]), provenance latitude (Extended Data Fig. 5) and air humidity[43].

Despite these limitations, a simple interpretation of our results does support the widespread hypotheses that chilling and/or photoperiod cues may underlie declining sensitivities to warming in long-term Central European data[6,7,9]. Under these hypotheses, warming increases forcing and thus advances budburst but such advances become muted if warming also causes important declines in chilling and shorter photoperiods experienced near the timing of budburst[36]. This basic agreement between our results—based on short-term experiments with highly controlled conditions—and long-term observational trends integrates across experimental conditions that encompass more extreme scenarios than may be seen in nature (Extended Data Fig. 4 and Supplementary Fig. 2). A more robust comparison requires examining predictions under conditions closer to those found in nature.

Reinterpreting our estimates of effects of chilling, forcing and photoperiod (from experiments) using climate and phenology data that have led to observations of declining temperature sensitivities in Central Europe suggests that chilling and photoperiod are unlikely to cause the observed declines. Our results predict such declines only at extreme warming for most sites (Supplementary Information). In contrast to the common hypothesis that plants experience less chilling with global warming, we found that—for many sites—total estimated chilling increased with warming (Fig. 4a,c), although this varied with local climate before warming (Extended Data Fig. 6 and Supplementary Fig. 5). Portions of Central Europe have experienced more dramatic warming in winter versus summer (refs. [44,45], with variation over time and space). Yet even if warming uniquely occurs in the winter, we found that delays due to decreased chilling only occur at warming above at least 4 °C for most sites, although responses vary by species (Fig. 3

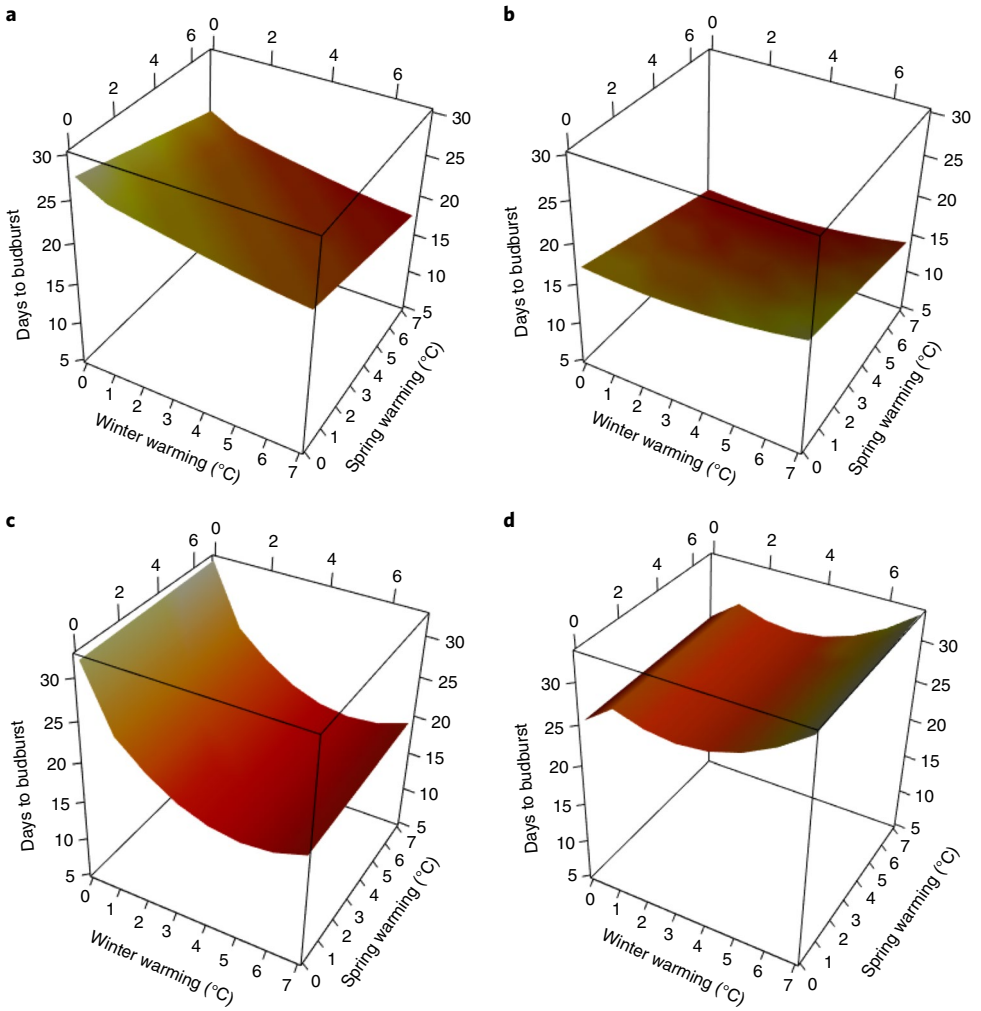


Fig. 4 | Implications of warming on budburst timing varies across species and sites, depending strongly on prewarming climate conditions related to chilling for each site. a–d, Here we show species-level estimates from our model based on a meta-analysis of experiments (Fig. 2) for two common species *B. pendula* (**a**,**b**) and *F. sylvatica* (**c**,**d**), on the basis of climate data from two sites in Central Europe (these two sites were chosen to highlight the diversity of possible budburst responses to warming; see Extended Data Fig. 6 for general trends across many sites in the same region). In some sites, warming increases total chilling estimates (**a**,**c**) leading to greater advances in budburst (compared with forcing alone), whereas warming decreases total chilling estimates in other sites (**b**,**d**), leading to smaller advances and, eventually, delays with substantial warming. See Supplementary Fig. 4 for a simplified two-dimensional version.

and Extended Data Fig. 6). At high warming, predicted declines in sensitivity were due to declines in chilling—photoperiod had comparatively little effect on budburst day of year, even for the photosensitive species *F. sylvatica* (Extended Data Fig. 7).

Our predictions leave open the question of what underlies declining sensitivities across Europe but one possibility is that it may be a statistical artefact of how temperature sensitivities are calculated. Physiologically, budburst is triggered by the accumulation of forcing temperatures during the spring[10,46]. However, researchers today often estimate temperature sensitivities from long-term observational data using a linear regression of annual budburst date versus mean spring temperature or other aggregated temperature metrics (for example, ref. [5]). This approach has the benefit of yielding an easily interpretable metric—days change per °C—but will estimate systematically lower sensitivities given warmer daily temperatures, even with no change in the underlying cues (Extended Data Fig. 8). We found the declining sensitivities observed in European data (Extended Data Fig. 9) are of the same magnitude as those predicted from a statistical artefact (sensitivity declines of 0.8 ± 0.3 d °C^{-1} in European data versus 0.9 ± 0.5 d °C^{-1} in simulations) and the data also show a related decline in leafout variance that would not be immediately predicted from shifting cues (see 'Potential statistical artefacts in declines of temperature sensitivity in observational long-term data' in the Supplementary Information, Extended Data Figs. 9 and 10, and ref. [47] for further details). This statistical artefact is probably not confined to phenological studies; it should apply to any research using a similar days per °C metric to estimate an underlying thermal accumulation model where the thermal sum per day is non-stationary, as is the case with climate change.

Our results unify decades of short-term experiments using controlled temperature and photoperiod conditions, which have shown the importance of chilling, forcing and daylength to determining budburst timing, with long-term observational data, where forcing appears to dominate (for example, ref. [48]). We do not find strong evidence for delaying budburst in the near future. Instead, our predictions suggest budburst will continue to advance in many well-studied European regions in the future. The most dramatic changes in future spring phenology will come from regions where winter warming causes large changes in chilling, with implications for ecosystem services related to phenology. Thus identifying

processes that plants undergo when accumulating chilling versus forcing will be critical for the most accurate forecasts[10,19].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-020-00917-3.

Received: 4 September 2019; Accepted: 12 August 2020;
Published online: 19 October 2020

References

1. IPCC *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (eds Field, C. B. et al.) (Cambridge Univ. Press, 2014).
2. Miller-Rushing, A. J. & Primack, R. B. Global warming and flowering times in Thoreau's Concord: a community perspective. *Ecology* **89**, 332—341 (2008).
3. Menzel, A. et al. European phenological response to climate change matches the warming pattern. *Glob. Change Biol.* **12**, 1969–1976 (2006).
4. Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A. & Schwartz, M. D. Shifting plant phenology in response to global change. *Trends Ecol. Evol.* **22**, 357–365 (2007).
5. Wolkovich, E. M. et al. Warming experiments underpredict plant phenological responses to climate change. *Nature* **485**, 494–497 (2012).
6. Rutishauser, T., Luterbacher, J., Defila, C., Frank, D. & Wanner, H. Swiss spring plant phenology 2007: extremes, a multi-century perspective, and changes in temperature sensitivity. *Geophys. Res. Lett.* **35**, L05703 (2008).
7. Yu, H. Y., Luedeling, E. & Xu, J. C. Winter and spring warming result in delayed spring phenology on the Tibetan Plateau. *Proc. Natl Acad. Sci. USA* **107**, 22151–22156 (2010).
8. Wang, X. et al. No trends in spring and autumn phenology during the global warming hiatus. *Nat. Commun.* **10**, 2389 (2019).
9. Fu, Y. S. H. et al. Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**, 104–107 (2015).
10. Chuine, I. et al. Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Glob. Change Biol.* **22**, 3444–3460 (2016).
11. Harrington, C. A. & Gould, P. J. Tradeoffs between chilling and forcing in satisfying dormancy requirements for Pacific Northwest tree species. *Front. Plant Sci.* **6**, 120 (2015).
12. Zohner, C. M., Benito, B. M., Svenning, J. C. & Renner, S. S. Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nat. Clim. Change* **6**, 1120–1123 (2016).
13. Basler, D. & Körner, C. Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. *Tree Physiol.* **34**, 377–388 (2014).
14. Caffarra, A., Donnelly, A., Chuine, I. & Jones, M. B. Modelling the timing of *Betula pubescens* bud-burst. I. Temperature and photoperiod: a conceptual model. *Clim. Res.* **46**, 147–157 (2011).
15. Flynn, D. F. B. & Wolkovich, E. M. Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytol.* **219**, 1353–1362 (2018).
16. Caffarra, A., Donnelly, A. & Chuine, I. Modelling the timing of *Betula pubescens* budburst. II. Integrating complex effects of photoperiod into process-based models. *Clim. Res.* **46**, 159–170 (2011).
17. Fraga, H., Pinto, J. G. & Santos, J. A. Climate Change projections for chilling and heat forcing conditions in European vineyards and olive orchards: a multi-model assessment. *Climatic Change* **152**, 179–193 (2019).
18. Heide, O. Daylength and thermal time responses of budburst during dormancy release in some northern deciduous trees. *Physiol. Plant.* **88**, 531–540 (1993).
19. Singh, R. K., Svystun, T., AlDahmash, B., Jönsson, A. M. & Bhalerao, R. P. Photoperiod- and temperature-mediated control of phenology in trees—a molecular perspective. *New Phytol.* **213**, 511–524 (2017).
20. Vitasse, Y. & Basler, D. What role for photoperiod in the bud burst phenology of European beech. *Eur. J. For. Res.* **132**, 1–8 (2013).
21. Vitasse, Y. & Basler, D. Is the use of cuttings a good proxy to explore phenological responses of temperate forests in warming and photoperiod experiments? *Tree Physiol.* **34**, 174–183 (2014).
22. Laube, J. et al. Chilling outweighs photoperiod in preventing precocious spring development. *Glob. Change Biol.* **20**, 170–182 (2014).
23. Basler, D. & Körner, C. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agric. For. Meteorol.* **165**, 73–81 (2012).
24. Caffarra, A. & Donnelly, A. The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. *Int. J. Biometeorol.* **55**, 711–721 (2011).
25. Ohlemüller, R., Gritti, E. S., Sykes, M. T. & Thomas, C. D. Towards European climate risk surfaces: the extent and distribution of analogous and non-analogous climates 1931–2100. *Glob. Ecol. Biogeogr.* **15**, 395–405 (2006).
26. Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ.* **5**, 475–482 (2007).
27. Williams, J. W., Jackson, S. T. & Kutzbacht, J. E. Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl Acad. Sci. USA* **104**, 5738–5742 (2007).
28. IPCC *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) (Cambridge Univ. Press, 2013).
29. Xu, Y., Ramanathan, V. & Victor, D. G. Global warming will happen faster than we think. *Nature* **564**, 30–32 (2018).
30. Wolkovich, E. M. et al. *Observed Spring Phenology Responses in Experimental Environments (OSPREE)* (Knowledge Network for Biocomplexity, 2019); https://doi.org/10.5063/F1CZ35KB
31. Richardson, E. A model for estimating the completion of rest for 'Redhaven' and 'Elberta' peach trees. *HortScience* **9**, 331–332 (1974).
32. Dennis, F. Problems in standardizing methods for evaluating the chilling requirements for the breaking of dormancy in buds of woody plants. *HortScience* **38**, 347–350 (2003).
33. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2006).
34. Fu, Y. H. et al. Short photoperiod reduces the temperature sensitivity of leaf-out in saplings of *Fagus sylvatica* but not in horse chestnut. *Glob. Change Biol.* **25**, 1696–1703 (2019).
35. Bradley, N. L., Leopold, A. C., Ross, J. & Huffaker, W. Phenological changes reflect climate change in Wisconsin. *Proc. Natl Acad. Sci. USA* **96**, 9701–9704 (1999).
36. Gauzere, J., Lucas, C., Ronce, O., Davi, H. & Chuine, I. Sensitivity analysis of tree phenology models reveals increasing sensitivity of their predictions to winter chilling temperature and photoperiod with warming climate. *Ecol. Model.* **441**, 108805 (2019).
37. Heide, O. & Prestrud, A. Low temperature, but not photoperiod, controls growth cessation and dormancy induction and release in apple and pear. *Tree Physiol.* **25**, 109–114 (2005).
38. van der Schoot, C., Paul, L. K. & Rinne, P. L. H. The embryonic shoot: a lifeline through winter. *J. Exp. Bot.* **65**, 1699–1712 (2014).
39. Fishman, S., Erez, A. & Couvillon, G. The temperature dependence of dormancy breaking in plants: mathematical analysis of a two-step model involving a cooperative transition. *J. Theor. Biol.* **124**, 473–483 (1987).
40. Weinberger, J. H. et al. Chilling requirements of peach varieties. *Proc. J. Am. Soc. Hort. Sci.* **56**, 122–128 (1950).
41. Polgar, C. A., Primack, R. B., Williams, E. H., Stichter, S. & Hitchcock, C. Climate effects on the flight period of Lycaenid butterflies in Massachusetts. *Biol. Conserv.* **160**, 25–31 (2013).
42. Vitasse, Y. Ontogenic changes rather than difference in temperature cause understory trees to leaf out earlier. *New Phytol.* **198**, 149–155 (2013).
43. Laube, J., Sparks, T. H., Estrella, N. & Menzel, A. Does humidity trigger tree phenology? Proposal for an air humidity based framework for bud development in spring. *New Phytol.* **202**, 350–355 (2014).
44. Li, C., Stevens, B. & Marotzke, J. Eurasian winter cooling in the warming hiatus of 1998–2012. *Geophys. Res. Lett.* **42**, 8131–8139 (2015).
45. Balling, R. C. J., Michaels, P. J. & Knappenberger, P. C. Analysis of winter and summer warming rates in gridded temperature time series. *Clim. Res.* **9**, 175–181 (1998).
46. Hänninen, H. Effects of climatic change on trees from cool and temperate regions: an ecophysiological approach to modelling of bud burst phenology. *Can. J. Bot.* **73**, 183–199 (1995).
47. Güsewell, S., Furrer, R., Gehrig, R. & Pietragalla, B. Changes in temperature sensitivity of spring phenology with recent climate warming in Switzerland are related to shifts of the preseason. *Glob. Change Biol.***23**, 5189–5202 (2017).
48. Roberts, A. M., Tansey, C., Smithers, R. J. & Phillimore, A. B. Predicting a change in the order of spring phenology in temperate forests. *Glob. Change Biol.***21**, 2603–2611 (2015).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

OSPREE database. To conduct this meta-analysis, we followed systematic review methods to facilitate replication and use by other researchers (for example, we include at least 22 of 27 items on the PRISMA checklist, as summarized in Supplementary Appendix 1, ref. [49]). We searched the literature for research papers that experimentally addressed controls of temperature, chilling and/or photoperiod requirements on the spring phenology of woody plant species. To identify phenological experiments that manipulated chilling, forcing and/or photoperiod, we searched both ISI Web of Science and Google Scholar with the following terms:

(1) TOPIC = (budburst OR leafout) AND (photoperiod or daylength) AND temperature*, which yielded 85 publications
(2) TOPIC = (budburst OR leafout) AND dorman*, which yielded 193 publications

The initial searches yielded 201 papers, which we reviewed and assessed for inclusion in the database. To be included, papers needed to focus on woody plants in temperate ecosystems and test for photoperiod and/or temperature effects on budburst, leafout or flowering and we needed to be able to quantify the phenological response to chilling, forcing and/or photoperiod. We used ImageJ to scrape these response data from figures, whenever possible, and added additional relevant information from the tables and text of each manuscript that could not be scraped. Multiple people checked scraping and data-entry, and mis-entered data and other mistakes were cleaned in R.

Our meta-analysis relies on the published literature where positive effects and larger effect sizes may be more likely to be published[50]. Methods such as a funnel plot of effect size versus sample size can help diagnose the potential for such biases but have many drawbacks and complications as well[51–53]. We could not use funnel plot methods here for several reasons: (1) our fundamental study design is based on three factors that can influence plant phenology, thus variation in effect size could be due to other levels of each factor, (2) studies had low sample sizes (75% of data came from studies with treatment sample sizes less than ten) and most often sample size was not reported (that is, in 25 out of 39 studies in the model with well-represented species) and finally (3) our modelling approach (see below) is designed to standardize and regularize data, thus it will pool some extreme effects that may arise from publication bias. Further, we note that these environmental cues have a firm physiological basis—thus, multiple lines of evidence (outside of publication bias) support that most studies should find an effect of (at least) chilling and forcing.

Some species are only represented in one dataset in the OSPREE database. In these instances, it is not possible to statistically differentiate between species, study and treatment effects. To address this, we combined species found in only one study into 'complexes' at the level of genera—such that each taxonomic unit we use in our model occurs across multiple studies (and treatments). Thus our taxonomic units of analysis are 'species complexes', which are either species represented in >1 dataset or complexes combining multiple species within a genus that are each singly represented in the dataset. Species represented in only one dataset with no congenerics in other datasets were excluded from most of our analyses, except when analysing 'all species'.

Although all studies measured days to budburst, many communicated results differently (for example, days to budburst, degree-days to budburst, percentage budburst and number of leaves). We standardized papers to common units whenever possible (details below) and further restricted studies to those for which forcing, chilling and photoperiod treatments could be quantitatively identified. For this paper, we focus on studies measuring days to budburst. This subset of OSPREE includes data across 72 experiments (in 49 papers; Supplementary Table 1), 39 years and 203 species (Supplementary Table 2 and Extended Data Fig. 1). These experiments span a wide range of chilling, forcing and photoperiod treatment levels (Extended Data Fig. 2) and many test for interactions between two of these cues (Supplementary Table 3). This subset of OSPREE is freely available on the Knowledge Network for Biocomplexity[30] and we hope other researchers will find it useful.

Defining budburst. Most studies defined budburst as initial 'green tips' (33 out of 49 papers). Select studies defined budburst as a specific increment of growth (for example, '0.5 cm of new growth') or as bud swell, leaf emergence, leaf unfolded, open bud scales or petiole emerged. The remaining papers (4 of 49) did not include a definition of budburst. Most papers using the above definitions (34 of 49) required only one bud to have met the defined criteria of budburst; however, the remaining studies implemented specific thresholds to be met (10–100% of all buds on an individual needed to have bursted bud). For studies that quantified multiple measurements of percentage budburst over time (days), we extracted one value of 'days to budburst' of these multiple measurements to make them comparable to other studies. To extract this summary value, we selected the days to budburst when percentage budburst was closest to 90%, including estimates as low as 49.5% budburst.

Estimating chilling. Chilling was reported far less in the OSPREE database than forcing and photoperiod. Although not all studies applied multiple treatments of forcing or photoperiod they generally all maintained and explicitly defined the forcing temperatures and daylengths applied in their treatments. In contrast, we found that most studies did not experimentally apply chilling by manipulating duration or temperature of chilling in controlled environments, nor did most quantify the total chilling imposed in their experiment. We therefore calculated the total chilling imposed by all studies; it would otherwise have been impossible to provide estimates with only experimental chilling given the rarity of such study designs (Extended Data Fig. 3).

To estimate total chilling we combined chilling from the field (chilling before plant material was brought into controlled environment conditions) and experimental chilling (chilling that plant material experienced in controlled environment conditions) into two widely used metrics of chilling: Utah units (Supplementary Table 4) and dynamic chill portions[32,54]. We used the `chillR` package v.0.70.17 in R v.3.6.0 (refs. [55,56]) to calculate both Utah units and dynamic chill portions from time series of hourly temperature data. To estimate field chilling, we generated hourly time series from a European-wide gridded dataset[57] and a North American gridded climate dataset[58], from which we extracted daily minimum and maximum temperature from the grid cells and dates during which experiments were conducted. For experimental chilling, we used reported chilling treatments to generate time series of hourly temperature data.

In the formulation we used, Utah chilling units accumulated the most at temperatures between 2.4 and 9.1 °C but slightly less at temperatures between 1.4 and 2.4 °C and from 9.1 to 12.4 °C. Utah units were reduced when temperatures fell below or exceeded this range (Supplementary Table 4). Chill portions accumulated when temperatures were between 0 and 7.2 °C. We note that these models for chilling (both of which were originally developed for peach species) are hypotheses for how chilling may accumulate to affect the process of endodormancy release but are likely to be inaccurate for many species. These models are, however, some of our current best approximations and versions of them are routinely applied to forest trees (for example, ref. [59]). We found the effects of chilling and other cues remained qualitatively consistent across the two methods of estimating total chilling (95% uncertainty intervals of estimates for all cues in the standardized models overlapped; Supplementary Table 5).

We wished to explore model predictions across a wide range of experimental temperature conditions (chilling and forcing temperatures) applied by studies included in the OSPREE database (Fig. 3). To do this, we needed to convert chilling temperature to total chilling units, which could be input into our model. There is no straightforward conversion between chilling temperature and total chilling, since the duration a temperature is applied affects chilling (Extended Data Fig. 4). We therefore made these conversions using two alternative approaches and we present both. For one approach, we generated daily time series of a range of experimental chilling temperatures for a range of durations spanning those in the OSPREE database (from −10 to 16 °C for 7–240 d). We averaged across the range of durations for each temperature to get one chilling estimate per chilling temperature (Extended Data Fig. 4 and Supplementary Fig. 2). For the alternative approach, we used historical climate data from a gridded climate dataset (E-OBS, ref. [57]) to estimate chilling and used these historical relationships between mean winter temperature and total chilling to convert chilling temperature to a representative amount of total chilling (Fig. 3). We present this alternative approach in the main text as it is more closely related to field chilling conditions, which was by far the most common type of chilling across experiments.

Estimating forcing and photoperiod. Our studies included a diversity of designs for applying forcing and/or photoperiod experimentally, including studies that imposed constant forcing temperatures and forcing temperatures that varied between day and night. Additionally several studies applied forcing or photoperiod using a 'ramped' design, such that treatments increased or decreased gradually over time throughout the duration of the application. For all studies we used the daylength of light as our photoperiod estimate (for example, a study with 8 h of light and 16 h of dark was recorded as '8'). For forcing, we used the temperature applied when forcing temperatures were constant (the same temperature was applied 24 h per day); if forcing varied with photoperiod, we estimated the mean daily temperature weighted by the hours that temperature was applied. Similarly, for studies that ramped forcing, we calculated a weighted average of forcing temperature over the period from when forcing treatments were applied until budburst day. For studies that ramped photoperiod, we used the photoperiod conditions that individuals initially experienced (for example, studies with photoperiod lengthening from 6 h until budburst, we recorded as '6'). When forcing and photoperiod treatments were reported as ambient, we used the E-OBS dataset to estimate mean forcing temperature and the R package `geosphere` to estimate daylength associated with each date and latitude[57].

Models. We fit four overall models: the main budburst model fit to species in OSPREE that measured days to budburst, the latitude model, which included only studies that had provenance latitude information, a model to examine how the design of chilling treatments affects estimated effects and a model to test for life-stage differences in budburst responses. Given the complexity of our meta-analytic data, we fit each model separately and present the main model in the main text as it was designed to best estimate chilling, forcing and photoperiod cues (our primary goal here). The other models represent subsets of the data in the main model that allow more direct tests of relevant, related questions.

As our primary goal was to directly compare the effects of chilling, forcing and photoperiod we standardized these predictor variables[33]. This was necessary because the range and scale of each predictor varied widely (total chilling across the 28 papers included in our main model ranged from −244 to 4,724 Utah units, forcing ranged from 5 to 32 °C and photoperiod ranged from 6 to 24 h). We followed well-established methods of subtracting the mean and dividing by the standard deviation[33] to yield z-score values for all predictor variables (total chilling units, forcing temperatures and photoperiods in the experiments). In addition to these models with standardized predictors (Supplementary Table 5), we also fit models in which predictors were not standardized (Supplementary Table 6) so that estimates could be more easily interpreted on their natural scales. For all figures in which predictors are shown on their natural scales, we use estimates from models in which predictors were not standardized.

All models were fit using the programming language Stan (www.mc-stan.org) (ref. [60]), accessed via the rstan package v.2.18.0 in R v.3.6.0 (refs. [55,61]). Stan provides efficient MCMC sampling via a No-U-Turn Hamiltonian Monte Carlo approach (more details can be found in refs. [62,60]). We validated our models using test data, then fit the models described below. In all models, i represents each unique observation, sp is the species or species complex grouping, α terms represent intercepts, β terms represent slope estimates and y is the days to budburst since forcing conditions were applied.

Main budburst model.

$$y_i = \alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α and each of the three β coefficients were modelled at the species level, as follows:

$$\begin{aligned}
\alpha_{sp} &\sim N(\mu_\alpha, \sigma_\alpha)\\
\beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing})\\
\beta_{photoperiod_{sp}} &\sim N(\mu_{photoperiod}, \sigma_{photoperiod})\\
\beta_{chilling_{sp}} &\sim N(\mu_{chilling}, \sigma_{chilling})
\end{aligned}$$

We applied this model to both a dataset with 203 species ('all species'), as well as with 65 species grouped into 36 taxa or 'species complexes' (Supplementary Tables 5 and 6) and a model excluding a single study[12] because this study contains 112 species (Supplementary Table 9). We present estimates from the model fit to the reduced dataset in the main text (including for Figs. 2 and 3) as these estimates summarize across species that were more well-represented in multiple papers and study designs, and thus are likely to be more accurate estimates (more details in OSPREE database). On the basis of our modelling approach, species from fewer studies will be pooled towards the overall mean. The reduced dataset model also excluded studies which reported only 'ambient' forcing and photoperiod; these studies were included in the dataset with 203 species ('all species' model).

Latitude model. Given continuing debate over the role of photoperiod on budburst timing across a species' latitudinal range (for example, refs. [12,63]), we examined the effect of including provenance latitude in a model similar to our main one but designed to estimate effects of provenance latitude. This model estimated the effects of each phenological cue (chilling, forcing and photoperiod) on days to budburst (as in the main model), in addition to the effect of provenance latitude (the latitude of origin of plant material used in the experiment) and the interaction of photoperiod and provenance latitude. We include this interaction because photoperiod effects are expected to vary by latitude and this interaction may have important implications under climate change[63–65].We followed the methods above for including species or species complex (see OSPREE database), including only species and species complexes that had multiple provenance locations across different latitudes. This yielded the following model:

$$\begin{aligned}
y_i = \; & \alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \beta_{latitude_{sp[i]}}\\
& + \beta_{photoperiod:latitude_{sp[i]}} + \epsilon_i
\end{aligned}$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α and each of the five β coefficients were modelled at the species level, as follows:

$$\begin{aligned}
\alpha_{sp} &\sim N(\mu_\alpha, \sigma_\alpha)\\
\beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing})\\
\beta_{photoperiod_{sp}} &\sim N(\mu_{photoperiod}, \sigma_{photoperiod})\\
\beta_{chilling_{sp}} &\sim N(\mu_{chilling}, \sigma_{chilling})\\
\beta_{latitude_{sp}} &\sim N(\mu_{latitude}, \sigma_{latitude})\\
\beta_{photoperiod:latitude_{sp}} &\sim N(\mu_{photoperiod:latitude}, \sigma_{photoperiod:latitude})
\end{aligned}$$

The latitude model is summarized in Supplementary Table 10 and Extended Data Fig. 5.

Chilling study-design model. As we found chilling to be the strongest cue and given how few studies directly manipulate it (Extended Data Fig. 3), we also used a subset of our data to estimate how a study's experimental design for chilling impacts model estimates. For this, we included only species or species complexes used in both experiments that used the Weinberger method (in this method plant tissue is sequentially removed from the field and then exposed to 'forcing' conditions, with the assumption that tissues collected later experience more field chilling[40]) and those that experimentally manipulated chilling (by varying chilling temperatures and/or the duration of chilling conditions). We defined Weinberger studies as those with two or more field sample dates, each 2 or more weeks apart, that did not otherwise manipulate chilling. The chilling study-design model was thus:

$$\begin{aligned}
y_i = \; & \alpha_{sp[i]} + \beta_{forcing} + \beta_{photoperiod} + \beta_{chilling} + \beta_{chillmethod} +\\
& \beta_{forcing:chillmethod} + \beta_{photoperiod:chillmethod} + \beta_{chilling:chillmethod} + \epsilon_i
\end{aligned}$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α coefficients were modelled at the species level, as follows:

$$\alpha_{sp} \sim N(\mu_\alpha, \sigma_\alpha)$$

The chilling design model is summarized in Supplementary Table 11 and Supplementary Fig. 3.

Life stage model. Previous research has found differences in spring phenology across life stages (for example, juvenile versus adult trees[42]). We tested for differences in days to budburst across life stages. We followed the guidelines above for including species or species complex (see OSPREE database), including only the species and species complexes used in experiments involving plant material from adult trees as well as juvenile life stages (seedlings or saplings). The life-stage model was thus:

$$y_i = \alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \beta_{stage_{sp[i]}} + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α and each of the four β coefficients were modelled at the species level, as follows:

$$\begin{aligned}
\alpha_{sp} &\sim N(\mu_\alpha, \sigma_\alpha)\\
\beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing})\\
\beta_{photoperiod_{sp}} &\sim N(\mu_{photoperiod}, \sigma_{photoperiod})\\
\beta_{chilling_{sp}} &\sim N(\mu_{chilling}, \sigma_{chilling})\\
\beta_{stage_{sp}} &\sim N(\mu_{stage}, \sigma_{stage})
\end{aligned}$$

The life-stage model is summarized in Supplementary Table 12.

For all models, we chose weakly informative priors; increasing the priors threefold did not change the model results. We ran four chains simultaneously, each with 2,500 sampling iterations (1,500 of which were used for warm-up), yielding 4,000 posterior samples for each parameter. We assessed model performance through $\hat{R}$ close to 1 and high n_{eff} (4,000 for most parameters but as low as 713 for a few parameters in the latitude model), as well as visual consideration of chain convergence and posteriors[62].

In our figures we show means ± 50% uncertainty intervals from our models (Figs. 2 and 3, Extended Data Fig. 5 and 7, Supplementary Figs. 3 and 4) because our focus here is on the most likely value for each parameter (for example, estimated response to forcing) and because they are computationally stable[60,62]. See Supplementary Tables 5–12 for 95% uncertainty intervals.

Modelling limitations based on experimental designs. An ideal model to predict budburst would potentially include (but is not limited to): interactions between cues, sigmoidal or other nonlinearities to assess potential threshold effects, provenance location, methodological details (for example, if plant material was whole plant versus twigs or whether temperatures were constant or varied each day and so on) and measurement error. As with all models, we were limited in how many parameters we could estimate given available data. Thus we focused on species differences and used additional models to assess some of the potentially largest other effects (latitude, methods of estimating chilling and life stage). We were unable to estimate interactions between cues in our meta-analysis because very few studies design experiments to test for interactions between chilling, forcing and photoperiod (Supplementary Table 3). Most experiments in our dataset, however, did include interactions between at least two cues (Supplementary Table 3); we fit our main budburst model to this subset of experiments (Supplementary Table 8), which resulted in qualitatively

similar estimates to those of the model fit to the full set of experiments (Supplementary Table 5).

As our focus is on experiments, which—by design—often impose high variation in phenological cues, we expected a linear model for chilling, forcing and photoperiod would be most appropriate. Nonlinear models, however, are often appropriate for phenological cues, especially in nature, where chilling may typically be very high or extremely short photoperiods are rare. Thus we tested a nonlinear (sigmoidal) model on the OSPREE data[66]. As chilling was the least experimentally manipulated in our database, we examined whether a sigmoidal curve for chilling would be more appropriate but found that it was a poorer fit than a comparable all-linear model ($R^2 = 0.53$ versus 0.57), did not qualitatively alter estimates of forcing (−0.83 versus −0.79) or photoperiod (−0.25 versus −0.54) and led to non-biologically relevant estimates of chilling. Fitting meaningful nonlinear models to experimental data may require more data and/or data at very high and low chilling, forcing and photoperiod values than are currently available.

The few studies that did incorporate interactions generally used the Weinberger method, which is not designed to robustly tease apart the effects of multiple cues (Supplementary Table 11 and Supplementary Fig. 3). Similarly we found variation in thermoperiodicity and variation in study material were too infrequent to test for effects with current data; although our life-stage model found no large differences in days to budburst between material from adults ($\mu_\alpha = 25.29$) versus juveniles ($\beta_{stage} = 24.2$; 50% uncertainty intervals overlap), Supplementary Table 12. Our estimated effects therefore average over interactions[33] but identifying them in future research will be critical to understanding and predicting budburst. This will be particularly challenging for forcing and chilling, as a lack of information on endodormancy requirements makes disentangling forcing from chilling conditions impossible with current data[10].

Our model does not include measurement error because these data were not possible to include for 25 out of 39 experiments included in the main OSPREE model with well-represented species. For those studies that did report measurement error, the error was generally small relative to the magnitude of the responses (for example, standard deviation was on average 12.07% of the response variable for studies for which standard deviation was extracted). Thus, it is unlikely that adding measurement error to our analyses would have a large effect on our estimates[67].

Data availability

The OSPREE budburst database used in this manuscript is publicly archived in the Knowledge Network for Biocomplexity[30].

Code availability

The code for models used in this manuscript is publicly archived in the Knowledge Network for Biocomplexity[30].

References

49. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* **151**, 264–269 (2009).
50. Kicinski, M. Publication bias in recent meta-analyses. *PLoS ONE* **8**, e81823 (2013).
51. Gurevitch, J., Morrow, L. L., Wallace, A. & Walsh, J. S. A meta-analysis of competition in field experiments. *Am. Nat.* **140**, 539–572 (1992).
52. Gurevitch, J. & Hedges, L. V. Statistical issues in ecological meta-analyses. *Ecology* **80**, 1142–1149 (1999).
53. Lin, L. F. & Chu, H. T. Quantifying publication bias in meta-analysis. *Biometrics* **74**, 785–794 (2018).
54. Luedeling, E. & Brown, P. H. A global analysis of the comparability of winter chill models for fruit and nut trees. *Int. J. Biometeorol.* **55**, 411–421 (2011).
55. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2017).
56. Luedeling, E. chillR: statistical methods for phenology analysis in temperate fruit trees. R package version 0.70.17 (2019).
57. Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. & Jones, P. D. An ensemble version of the E-OBS temperature and precipitation data sets. *J. Geophys. Res. Atmos.* **123**, 9391–9409 (2018).
58. Livneh, B.et al. A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013. *Sci. Data* **2**, 150042 (2015).
59. Harrington, C. A., Gould, P. J. & St Clair, J. B. Modeling the effects of winter environment on dormancy release of Douglas-fir. *For. Ecol. Manag.* **259**, 798–808 (2010).
60. Carpenter, B. et al. Stan: a probabilistic programming language. *J. Stat. Softw.* https://doi.org/10.18637/jss.v076.i01(2017).
61. Stan Development Team. RStan: the R interface to Stan. R package version 2.17.3 (2018).
62. Gelman, A. et al. *Bayesian Data Analysis* (CRC Press, 2014).
63. Gauzere, J. et al. Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: *Fagus sylvatica* and *Quercus petraea*. *Agric. For. Meteorol.* **244**, 9–20 (2017).
64. Saikkonen, K. et al. Climate change-driven species' range shifts filtered by photoperiodism. *Nat. Clim. Change* **2**, 239 (2012).
65. Way, D. A. & Montgomery, R. A. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant Cell Environ.* **38**, 1725–1736 (2015).
66. Chuine, I., Garcia de Cortazar Atauri, I., Hanninen, H. & Kramer, K. in *Phenology: An Integrative Environmental Science* (ed. Schwartz M.) 275–293 (Springer, 2013).
67. Stan Development Team *Stan User's Guide* v.2.19 (Stan, 2019).

Acknowledgements

We thank the many researchers who conducted the experiments synthesized in this manuscript. We thank B. Cook for help with climate data, E. Forrestel for assisting with data scraping and C. Zohner for sharing tables. We also thank J. Davies, S. Elmendorf and J. HilleRisLambers for helpful comments that improved the manuscript. The National Science Foundation (grant no. DBI 14-01854 to A.K.E.), NSERC Discovery Award (grant no. RGPIN-05038 to E.M.W.), Canada Research Chair in Temporal Ecology (E.M.W.) and Spanish Ministry for Science and Innovation (grant no. CGL2017-86926-P and PID2019/109711RJ-I00 to I.M.-C.) provided funding. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author contributions

D.F.B.F., T.S. and E.M.W. conceived of the OSPREE database, which gave rise to this manuscript. All authors worked tirelessly to build the database and all contributed data analysis and/or code. C.J.C., D.M.B., E.M.W., I.M.-C. and A.K.E. created the figures. A.K.E. and E.M.W. wrote most of the manuscript, with substantial contributions from C.J.C., D.M.B. and I.M.-C. All authors reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

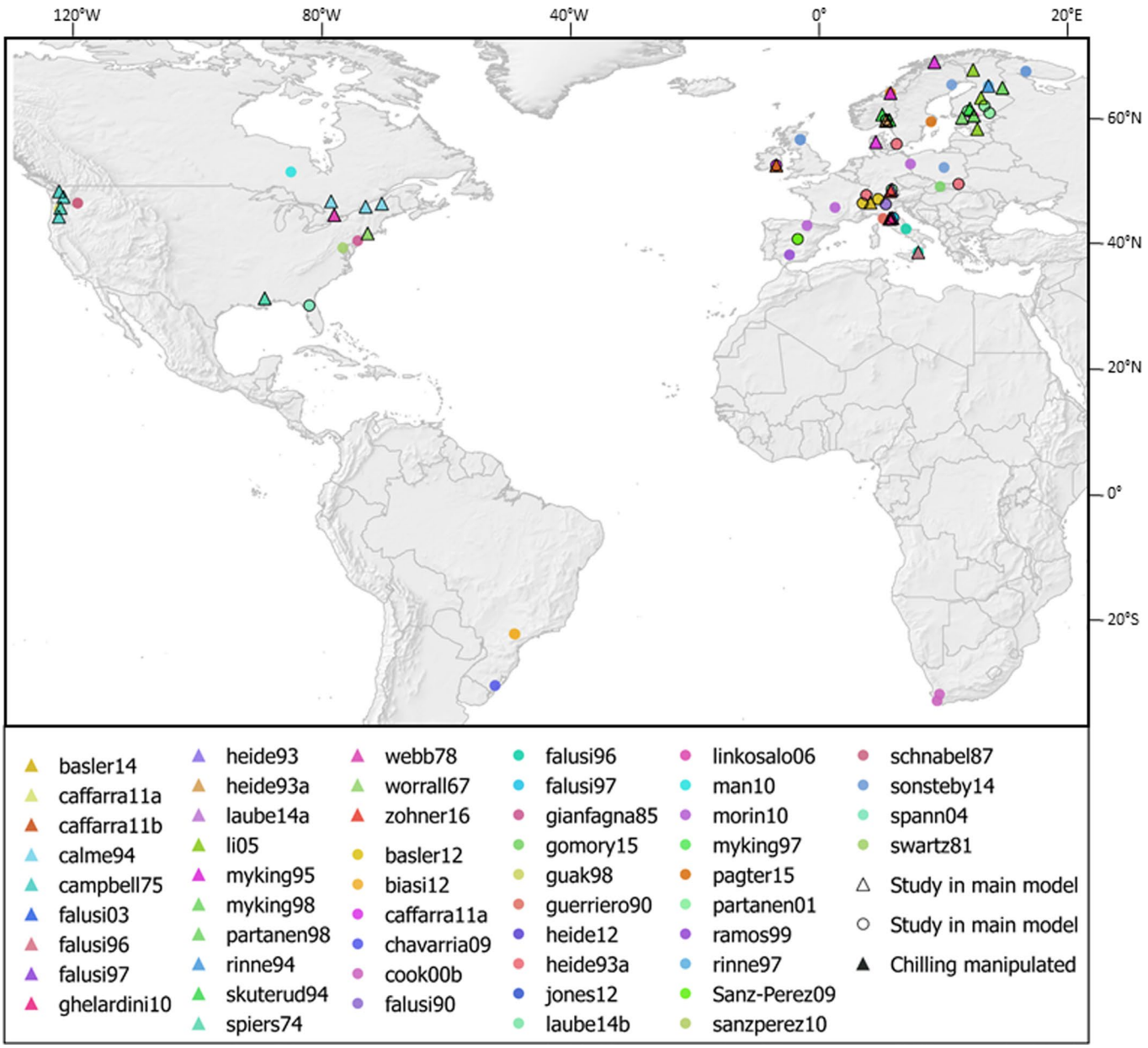
Extended data is available for this paper at https://doi.org/10.1038/s41558-020-00917-3.

Supplementary information is available for this paper at https://doi.org/10.1038/s41558-020-00917-3.

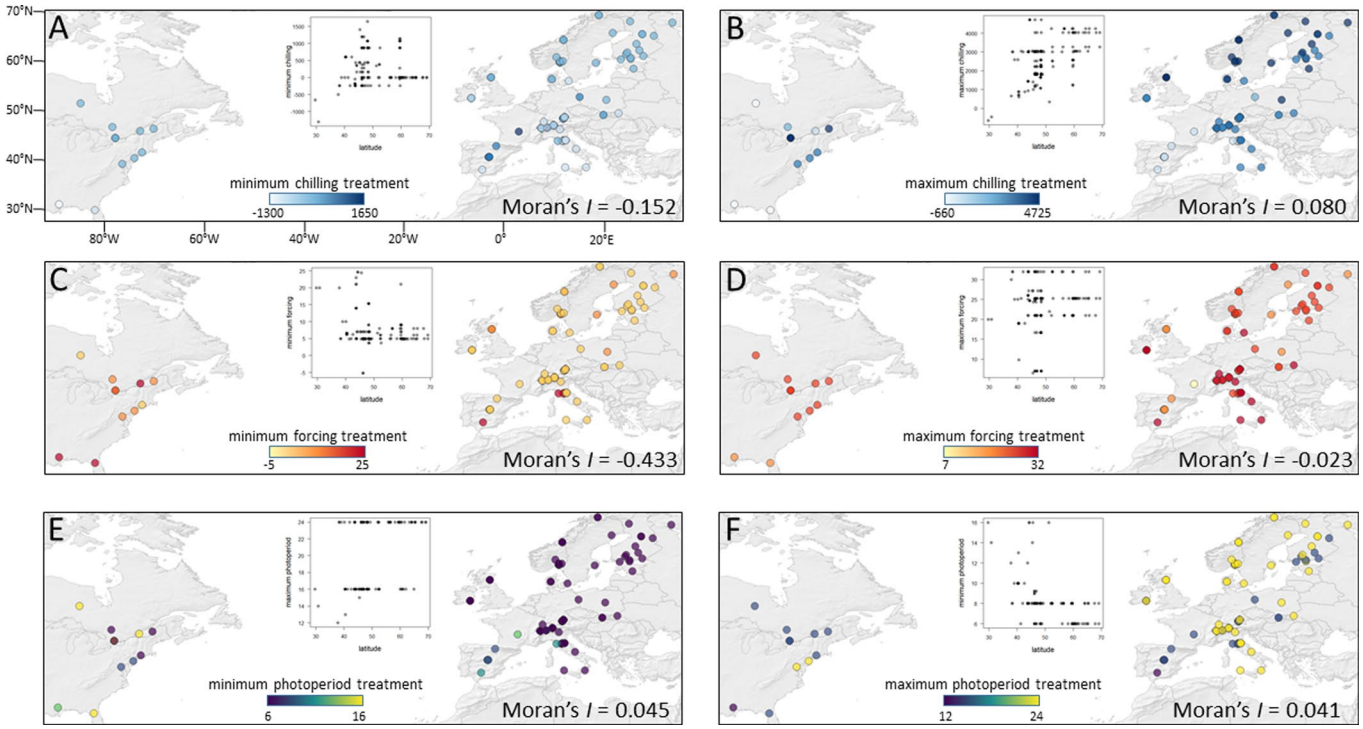
Correspondence and requests for materials should be addressed to A.K.E.

Peer review information *Nature Climate Change* thanks Albert Phillimore, Constantin Zohner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

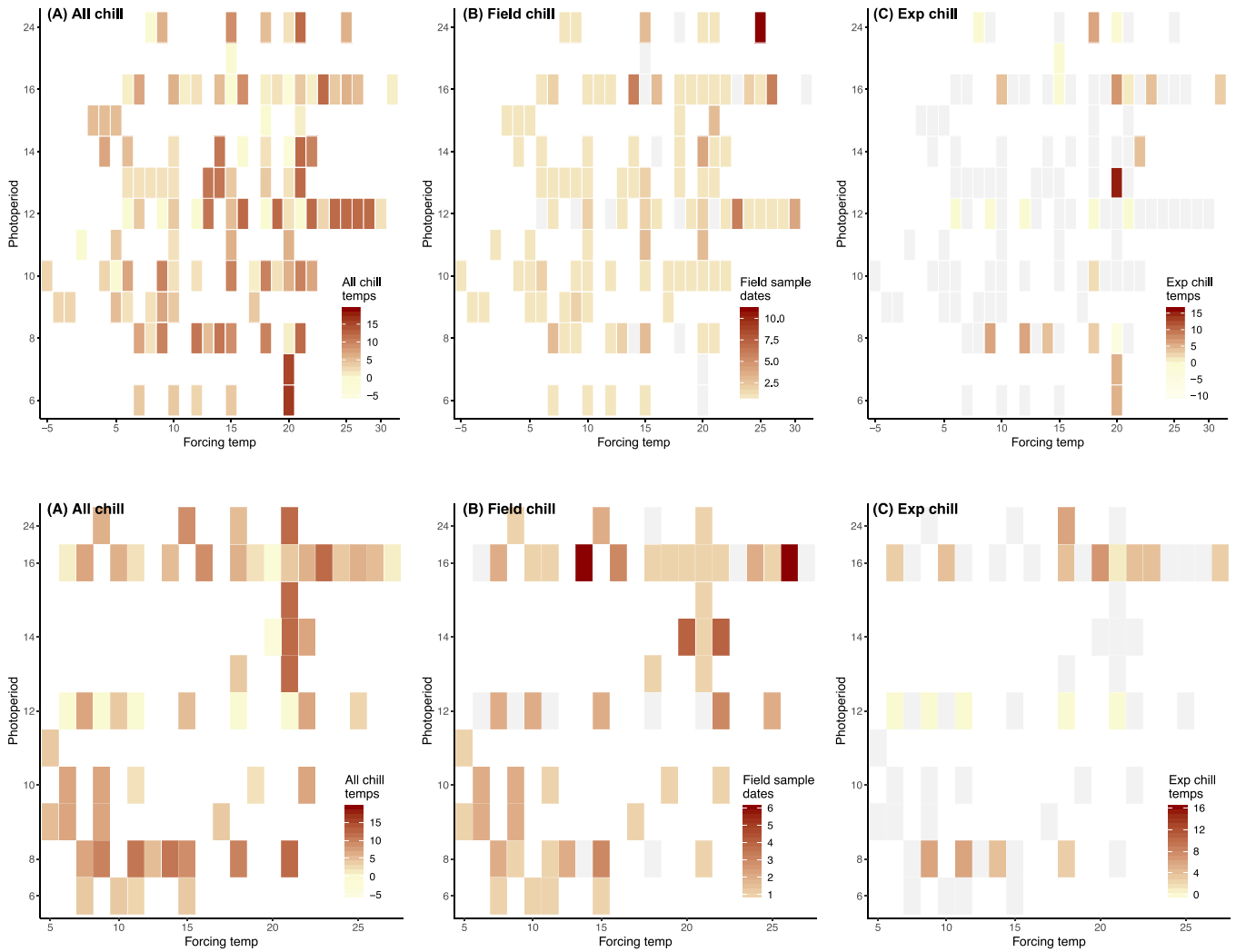
Reprints and permissions information is available at www.nature.com/reprints.



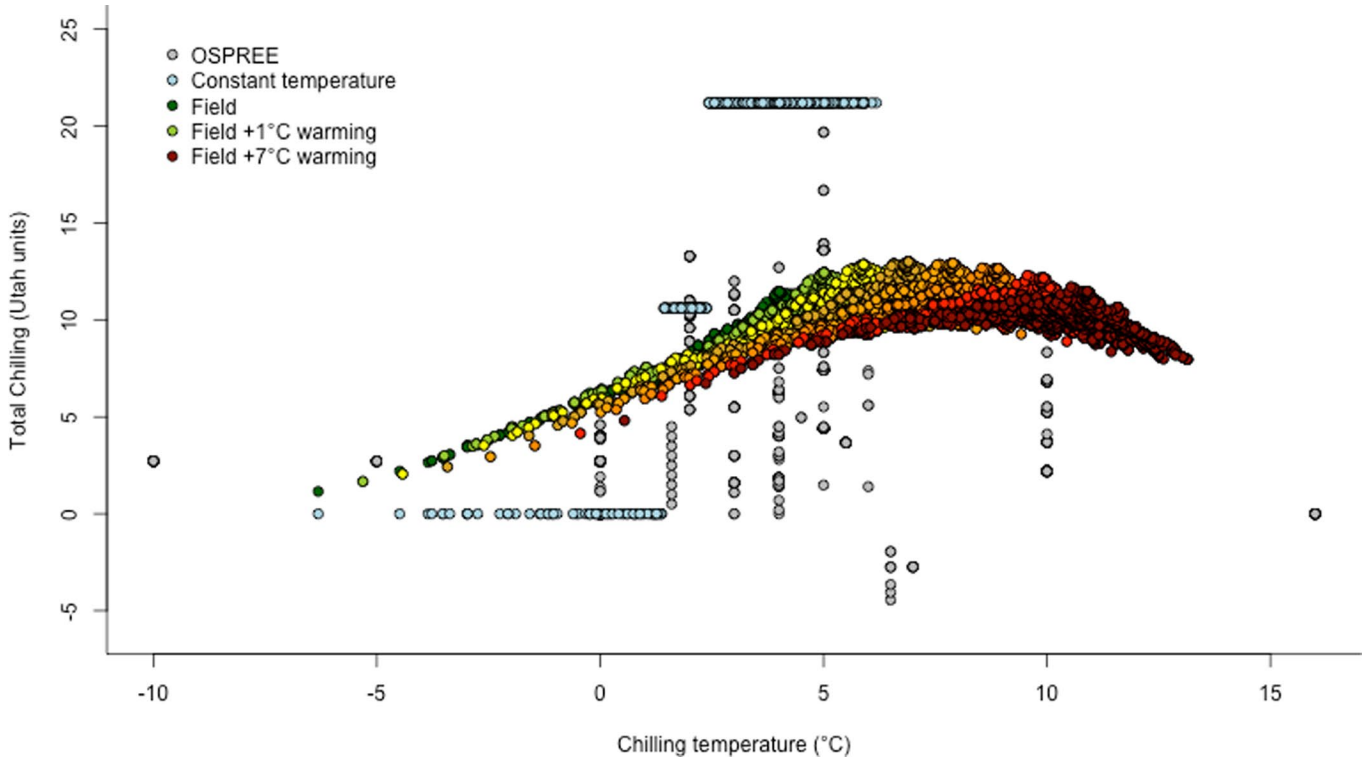
Extended Data Fig. 1 | Map of days to budburst experiments in the OSPREE database. Legend shows each dataset included in the main OSPREE model with all species and treatments (Supplementary Tables 5, 6); symbols outlined in black represent datasets included in the main budburst model; triangles represent studies in which chilling was manipulated experimentally or through multiple field sample dates; circles represent studies that did not manipulate chilling. See Supplementary Table 1 for the reference associated with each dataset.



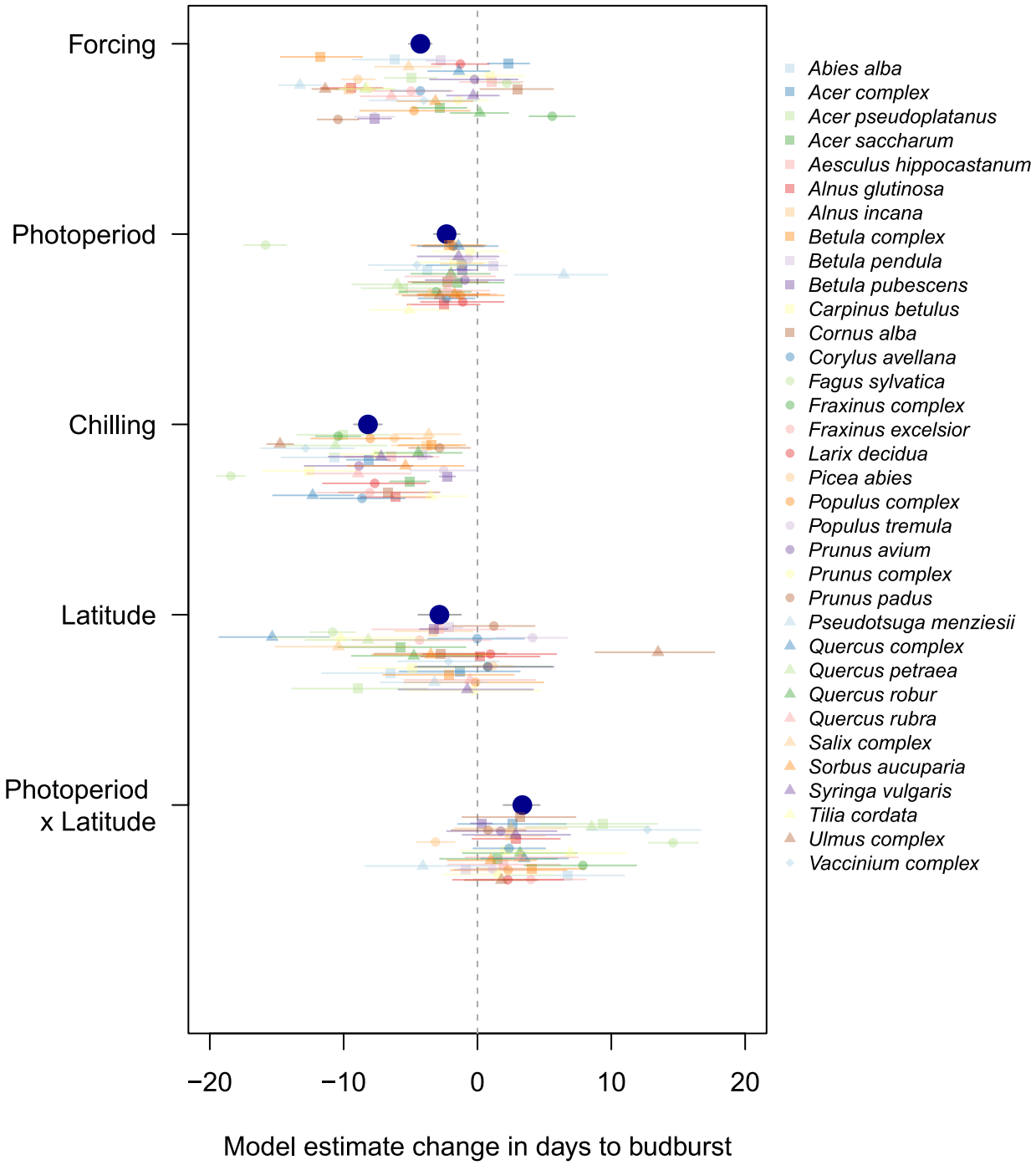
Extended Data Fig. 2 | Map of maximum and minimum chilling, forcing, and photoperiod treatments, across all data included in our main model, and the locations where each experiment was conducted. We did not find strong positive spatial autocorrelation—that is, indicating higher similarity in the treatments applied to experiments in nearby locations- in either minimum (A,C,E) or maximum (B,D,F) treatments, as measured by Moran's I (shown here for European sites). Insets show relationship of each cue's treatment level with the latitude of the experiment.



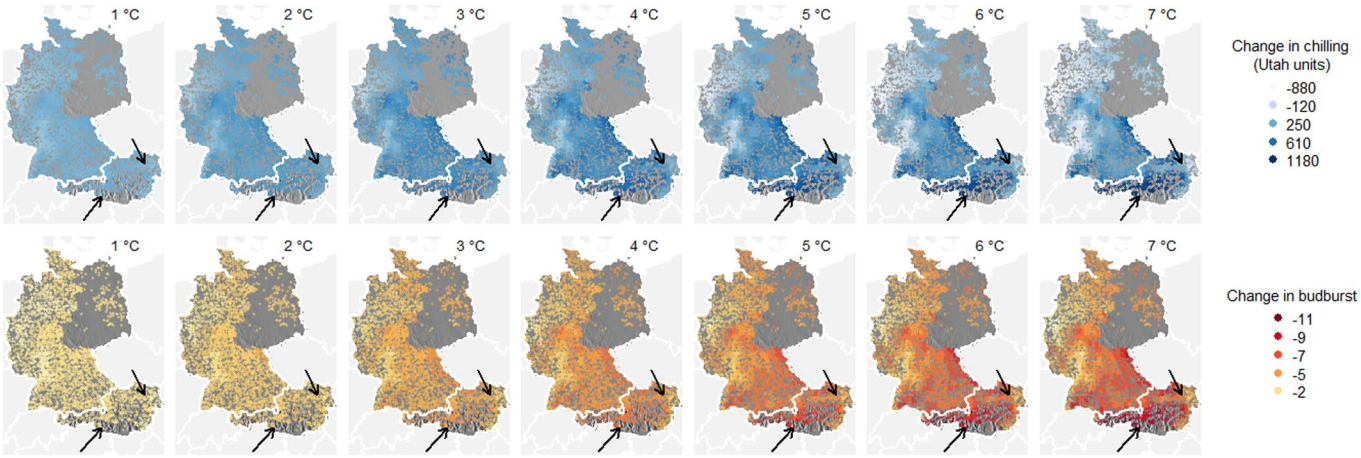
Extended Data Fig. 3 | The diversity of study designs used in analyses. Heatmaps show the range and commonness of different forcing (x-axis in all panels) by photoperiod (y-axis in all panels) combinations and with which chilling they were combined. In (A, top and bottom) we show our estimated chill units, which integrate across field (B, top and bottom) and experimental chilling (C, top and bottom). The top row shows all data included in the full model with 203 species, while the bottom row shows the data included in the main model with a subset of species well-represented across treatments and studies. Grey squares indicate a treatment was not applied (that is, the prevalence of grey squares in (C) highlights how few studies include experimental chilling). Field sample dates are counted as any reported sampling dates more than 14 days apart.



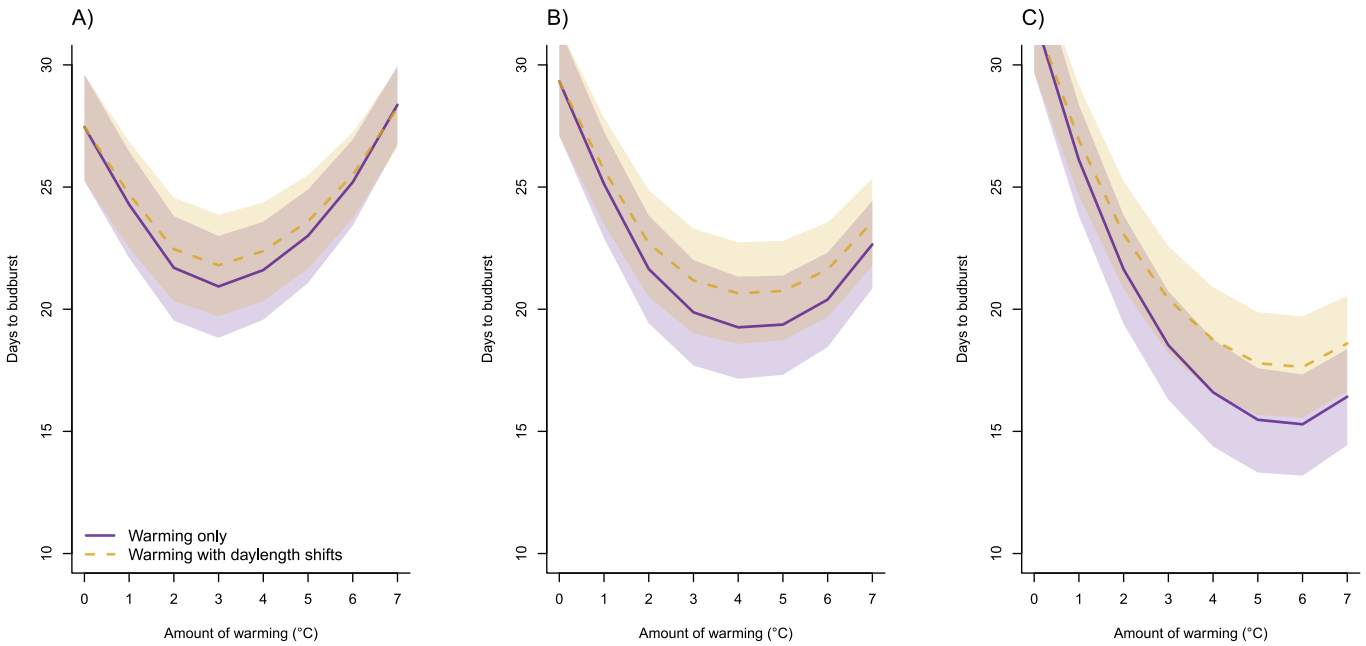
Extended Data Fig. 4 | Chilling accumulates differently in experiments with constant temperatures versus natural systems, in which temperature is more strongly correlated with chilling. See 'Estimating chilling' in Methods for a detailed description of 'Field' climate data, for which we use historical climate data from Europe. Fig. 3 uses 'Field' relationships (that is, climate data and relationships from field chilling conditions to convert chill temperature to total chilling), whereas Supplementary Fig. 2 uses 'Constant temperature' conditions (analogous to most experimental conditions) to estimate total chilling.



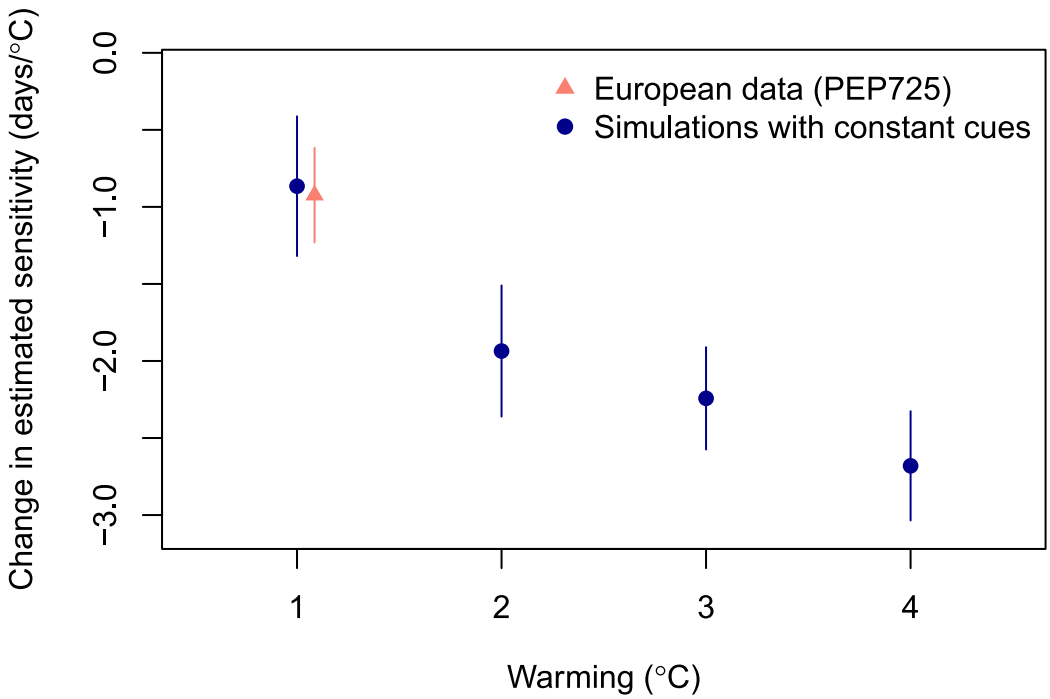
Extended Data Fig. 5 | Estimates for effects of chilling exceeded estimates for forcing, photoperiod, provenance latitude, and the interaction between latitude and photoperiod, for most species, in the latitude budburst model. Using Utah units for chilling (Supplementary Table 10) and standardized predictor variables, which allow comparisons across cues, we show that, as with the main budburst model (Fig. 2), most species (smaller symbols) are responsive to most cues. Chilling is the strongest cue when considering overall estimates across species (larger, dark blue circles).



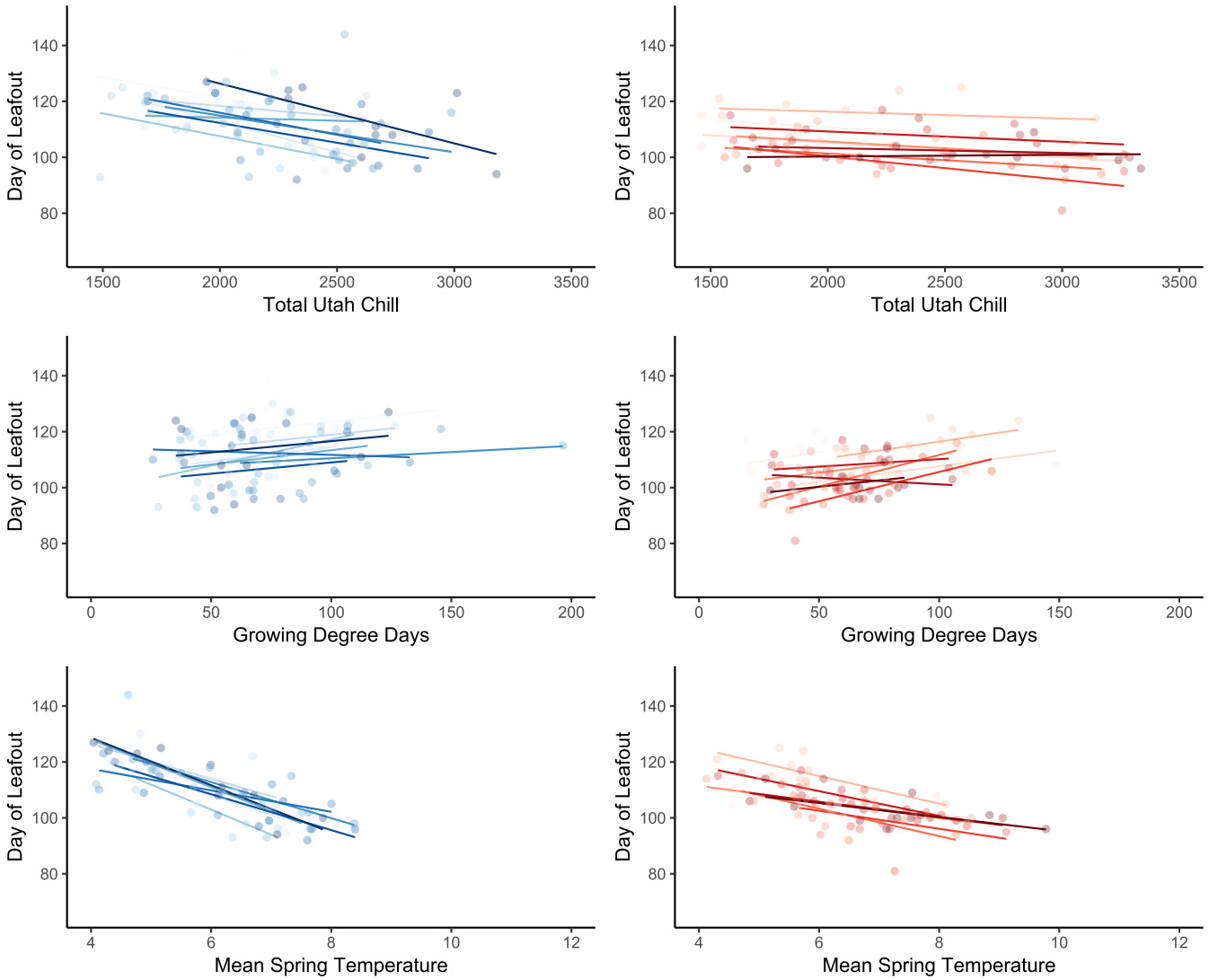
Extended Data Fig. 6 | Forecasted changes in chilling and spring phenology vary with amount of warming across European locations included in the PEP725 database. Changes in chilling (top panel) and budburst for *Betula pendula* (bottom panel) are calculated relative to the mean chilling and leafout dates during a prewarming time period (1951-1960) for each location. Arrows indicate sites shown in Fig. 4a and Supplementary Fig. 4A (latitude = 46.8°N, longitude = 12.8°E, 659 m above sea level) and Fig. 4b and Supplementary Fig. 4B (latitude = 48.3°N, longitude = 15.8°E, 210 m above sea level).



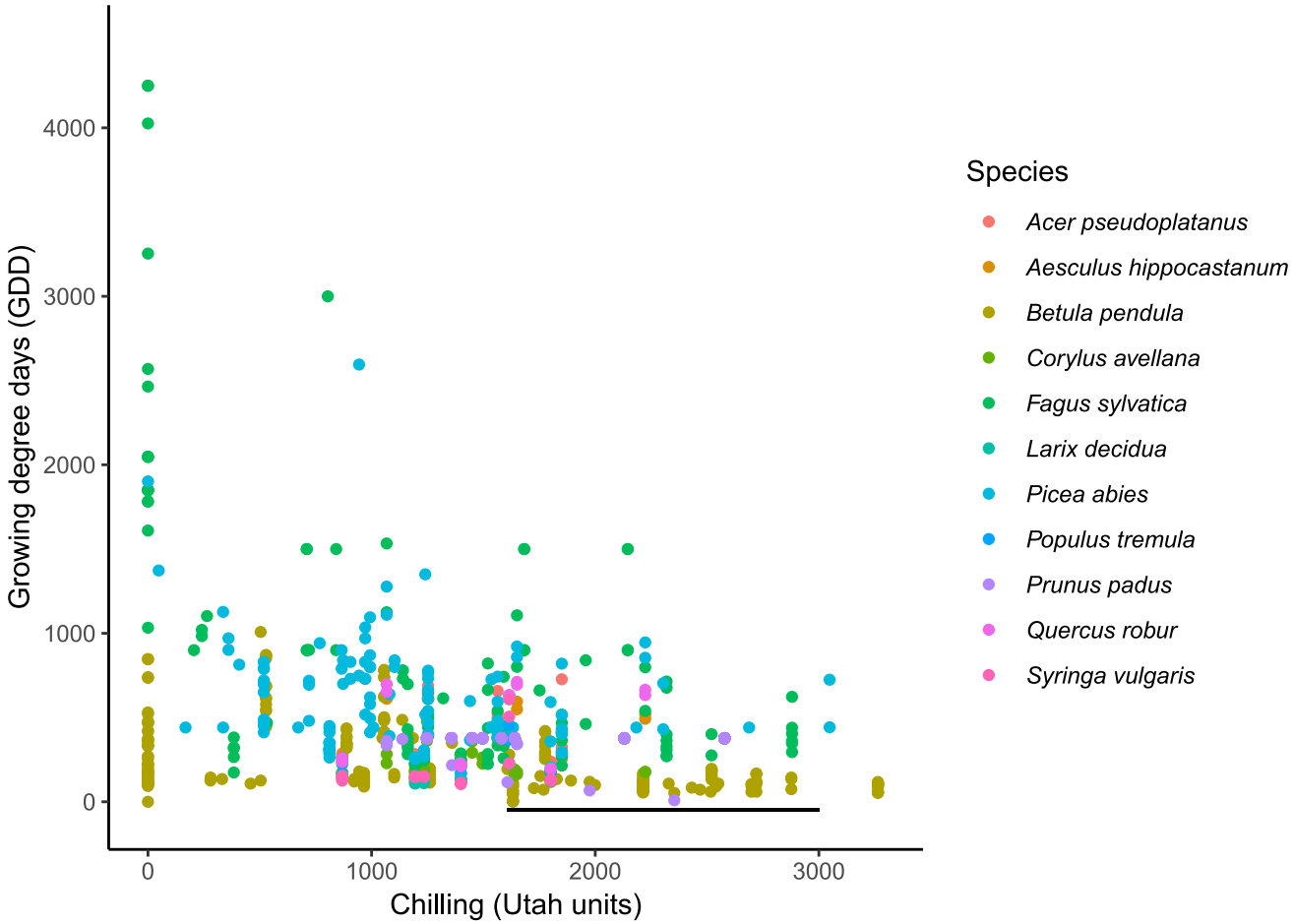
Extended Data Fig. 7 | Budburst is affected by climate change induced shifts in photoperiod, especially at high latitudes, although effects vary by site and are minor compared to effects of warming. We show forecasted effects of varying levels of warming on *Fagus sylvatica*, the most photoperiod-sensitive species in our database, across three latitudes within its range, as predicted by the latitude model. The low latitude site (A) is located at 46.8°N, 15.7°E; the mid-latitude site (B) is located at 47.7°N, 16.3°E; and the high-latitude (C) site is located at 48.8°N, 15.4°E.



Extended Data Fig. 8 | Declining sensitivities observed in long-term European data for a suite of common trees may be explained by a statistical artefact. We compared the sensitivity estimated from linear regressions of day of leafout versus mean spring temperature (estimated thus as days/°C) from PEP725 data for *Betula pendula* from 45 sites ('European data') with estimated declines using simulations where the cues were held constant but spring temperatures warmed by 1-4°C ('Simulations') and found the estimated temperature sensitivity measured as days/°C declined even though the underlying cues had not changed. See *Potential statistical artefacts in declines of temperature sensitivity in observational long-term data* in the Supplementary Information for further details.



Extended Data Fig. 9 | Day of leafout varies with chilling, growing degree-days, and mean spring temperature. These relationships are shown prewarming (left panels, 1951-1960) and post-warming (right panels, 2000-2010) for PEP725 sites in Germany where *Betula pendula* phenology has been monitored for decades.



Extended Data Fig. 10 | Growing degree days (GDD) versus chill units at the time of budburst from the OSPREE database for common species in the PEP725 long-term phenological database. The black line shows the range of chilling (10-90% quantiles) accumulated from 1 September to 1 March for 45 sites for *Betula pendula* (see also *Potential statistical artefacts in declines of temperature sensitivity in observational long-term data* in the Supplementary Information). We calculated GDD here as the average daily forcing temperature multiplied by days to budburst.