Winter temperatures predominate in spring phenological responses to warming

A. K. Ettinger[1,a], 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]

[1]Arnold Arboretum of Harvard University, Boston, Massachusetts 02131, USA

[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

[3]Global Change Ecology and Evolution Group, Department of Life Sciences, University of Alcalà, Alcalà de Henares 28805, Spain

[4]U.S. DOT Volpe National Transportation Systems Center, Cambridge, Massachusetts, USA

[5]Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[6]Forest Resources Management, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada

[7]Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada

[a]Corresponding author; current email: ailene.ettinger@tnc.org; phone: 781-296-4821; mailing address: 74 Wall Street, Seattle, Washington USA

August 3, 2020

Abstract

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 slow 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.

Main text

For decades, plant phenology has been one of the most reported and consistent biological imprints of climate change *(IPCC:2014sm-1)*: many temperate plants are leafing and flowering days to weeks earlier with rising temperatures *(millerrushing2008-1menzel2006; -1)*. 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 *(Cleland:2007or-1)*.

As research interest in phenology has progressed, critical discrepancies and uncertainties in our understanding have emerged. Though responses to warming are widespread, showing strong advances on average, there is substantial variation among species and sites *(Wolkovich:2012n-1)*. Furthermore, long-term observational studies provide increasing evidence that sensitivities of phenology to temperature are weakening in recent decades *(Rutishauser:2008-1yu2010; -1wang2019; -1)*, especially in Europe, where researchers suggest that responses to multiple environmental cues underlie declining temperature sensitivities *(fu2015-1)*.

Fundamental research in phenology outlines how three major environmental cues, chilling (cool temperatures, generally occurring in the fall 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 *(chuine2016-1)*. For example, in some species a cool winter will lower the amount of forcing required to trigger budburst, compared to a warmer winter *(harrington2015-1)*. Additionally, photoperiod may trigger budburst, given low chilling and/or forcing

2

₃₀ *(zohner2016-1Basler:2014aa; -1Caffarra:2011b; -1)*. Research suggests that all three cues may affect spring

₃₁ phenology for many temperate woody species *(flynn2018-1Basler:2014aa; -1Caffarra:2011qf; -1)*, which could

₃₂ have critical forecasting implications—predicting delays in spring phenology as increased warming reduces

₃₃ chilling in some areas *(fraga2019-1)* or where earlier budburst shortens the experienced photoperiod. How-

₃₄ ever, there is strong debate, with research suggesting some cues—such as photoperiod—may be effectively

₃₅ absent in some species, but dominate in others *(zohner2016-1Heide:1993; -1Basler:2014aa; -1Singh:2017; -1)*.

₃₆ Resolving this debate requires overcoming major hurdles to estimate responses to each cue. Studies at-

₃₇ tempting to estimate cues using long-term observational data *(e.g., zohner2016-1vitasse2013; -1)* generally

₃₈ fail to overcome the fundamental challenge that cues are strongly correlated in nature (*e.g.*, during the sea-

₃₉ sonal transition from winter to spring at temperate latitudes, forcing and photoperiod usually increase in

₄₀ step for a given location; average chilling and spring (forcing) temperatures can be positively correlated in

₄₁ space, especially at high latitudes, see Fig. ED6). In contrast to observational studies and field warming

₄₂ studies designed to test higher temperatures in natural conditions *(Wolkovich:2012n-1)*, experiments using

₄₃ controlled temperature and photoperiod conditions can break down correlations between the cues. These

₄₄ experiments, which generally rely on dormant tree cuttings or dormant plants exposed to temperature and

₄₅ light regimes in growth chambers (Fig. 1), have been shown to replicate whole-plant responses in nature

₄₆ *(vitasse2014-1)*. Such experiments have been conducted for decades (though each experiment generally

₄₇ lasts under a year). They have produced contrasting results, however, potentially due to differences in fo-

₄₈ cal species or study sites *(zohner2016-1Caffarra:2011b; -1Laube:2014a; -1Basler:2012; -1Caffarra:2011a; -1)*.

₄₉ Resolving these discrepancies is critical to accurate predictions of spring phenology, especially as continued

₅₀ climate change will yield warmer temperatures than have been experienced in at least the last 150 years

₅₁ *(ohlemuller2006-1williams2007; -1williams2007b; -1ipcc2013; -1xu2018; -1)*.

₅₂ Here, we leverage these short-term controlled environment experiments in a meta-analysis to understand how

₅₃ chilling, forcing, and photoperiod determine budburst timing in woody species. We reviewed 201 papers,

₅₄ extracting data from all experiments that reported budburst responses, yielding data from 72 studies and

₅₅ 203 species (Fig. ED1, Tables S1, S2). The resulting Observed Spring Phenology Responses in Experimental

₅₆ Environments (OSPREE) database includes studies of dormant plant tissue (grown in greenhouses or taken di-

₅₇ rectly from the field) exposed to experimental temperature and/or photoperiod conditions *(wolkovich2019-1)*

₅₈ for which we could identify chilling, forcing, and photoperiod treatments quantitatively (these varied by each

₅₉ study, see Fig. ED3). Most experiments reported forcing and photoperiod treatments, whereas chilling oc-

60 curred mainly in the field, though some studies additionally applied chilling before moving plants into forcing

61 conditions (Fig. 1). Because chilling was rarely reported, we calculated an estimate of chilling (both in

62 the field and in experimental conditions), using a common approximation *(richardson1974-1)*, based on a

63 hypothesis of how chilling accumulates *(dennis2003-1)*, with no chilling accumulating below 1.4°C or above

64 12.4°C (throughout the main text we use the term 'chill unit,' see Supplemental Materials, especially Table

65 S4, for details).

66 We estimated the effects of chilling, forcing, and photoperiod using a Bayesian hierarchical model. Our

67 model averages over interactive effects of predictors, including only main effects that we could more robustly

68 estimate given current study designs (see *Methods*). Species are modeled hierarchically, producing estimates

69 of both species-level responses (generally yielding more accurate estimates for well-studied species, such as

70 *Fagus sylvatica* and *Betula pendula*), and the distribution from which they are drawn, yielding estimates of

71 the overall responses across species (see *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 modeled at the species level, as follows:

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

$$\beta_{forcing_{sp}} \sim N(\mu_{forcing}, \sigma_{forcing})$$

$$\beta_{photoperiod_{sp}} \sim N(\mu_{photoperiod}, \sigma_{photoperiod})$$

$$\beta_{chilling_{sp}} \sim N(\mu_{chilling}, \sigma_{chilling})$$

72 where i represents each unique observation, sp is the species (or species complex grouping, explained below),

73 α represents the intercept, β terms represent slope estimates, and y is the days to budburst since forcing

4

conditions were applied. Some species were represented in only one dataset in the OSPREE database, making it difficult to statistically differentiate 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 taxa, 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 *The Observed Spring Phenology Responses in Experimental Environments (OSPREE) database* in Methods for details). Estimates from this model were generally similar to estimates from a model of all 203 species (Tables S5, S6). To directly compare the effects of chilling, forcing and photoperiod we fit models using standardized predictor variables *(*following gelman2006-1, which we refer to as "standard units") and predictors in their natural units (chill units, °C, 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* section of *Methods* for model equations and other details).

Across experiments, all cues—chilling, forcing, and photoperiod—advance budburst phenology (Fig. 2, Tables S5, S6). Chilling was the strongest cue (-8.35 days/standard unit [-11.43 to -5.36] or -2.76 days per chill unit [-3.65 to -1.89]), followed by forcing (-4.35 days/standard unit [-6.65 to -1.92] or -0.8 days per °C of warming, [-1.18 to -0.43]), and photoperiod (-2.95 days/standard unit [-5.46 to -0.48] or -0.53 days per hour of daylight [-0.92 to -0.15]; Fig.3, S2, S4; Tables S5, S6; see Supplemental Materials for more details). 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 *(zohner2016-1fu2019; -1)*—instead we found it was surprisingly large, even when accounting for its interaction with provenance latitude (*i.e.*, the latitude of origin for plant material; see Supplemental Materials, Fig. ED5, ED7 & Table S10 for details). It was also generally consistent across species (variance = 5.18 days/standard unit), only deviating in *Fagus sylvatica*, a species well-known for having a large response to photoperiod (which we also found, Fig. 2, ED5). Species responses to chilling were slightly more variable (variance = 7.21 days/standard unit, Fig. 2) than responses to forcing (variance = 5.72 days/standard unit Fig. 2, Table S5).

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 *(menzel2006-1harrington2015; -1Basler:2014aa; -1bradley1999; -1)*. Our results, however, suggest that, across 65 species and 72 experiments, chilling has a greater effect on budburst

than forcing (Fig. 2, ED5, S3; Tables S5-S10). 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 *(e.g., zohner2016-1Basler:2014aa; -1Caffarra:2011qf; -1Laube:2014a; -1)*, not chilling versus forcing effects. Process-based phenological models, however, that explicitly model chilling often find this cue to be most critical *(e.g., gauzere2019-1Laube:2014a; -1Heide:2005aa; -1)*.

Despite its apparent importance, chilling and its related physiological stage of endodormancy, are not well understood *(chuine2016-1)*. 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 *(chuine2016-1)*. Thus, while researchers generally define "chilling" and "forcing" treatments based on temperatures in controlled experiments (including in the studies synthesized here, see Fig. 1), fully separating out what plants experience as chilling versus forcing (as well as how this varies across species and sites) will likely require new methods to measure endo- and ecodormancy *(vanderschoot2014-1)*.

Until then, researchers must generally rely on modeled estimates of chilling, as we have used here. While we found that applying a different chilling model did not strongly affect our estimates (*i.e.*, 95% uncertainty intervals of estimates for chilling, photoperiod, and forcing effects overlapped using two different chill metrics, Utah and chill portions, and the mean posterior of these estimates varied by about 10% or less between the two metrics, see Table S5), 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 (*i.e.,* Utah, Table S4, richardson1974-1) and chill portions *(fishman1987-1)*. These models are themselves hypotheses for how chilling may accumulate and lead to dormancy release, but are likely to be inaccurate for many species *(dennis2003-1)*.

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. 1A,B, 25 out of 72; the remaining 34 studies did not appear to manipulate chilling), with the assumption that tissues collected later experience more chilling *(weinberger1950-1)*. 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 *(dennis2003-1)*, 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 (Fig. S3, Table S11 *(weinberger1950-1polgar2013; -1)*. 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 between all three cues (Table S3). Further, many additional factors can affect phenological responses, including ontogeny (Table S12) *(vitasse2013ont-1)*, provenance latitude (Fig. ED5), and air humidity *(Laube:2014b-1)*.

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 *(Rutishauser:2008-1yu2010; -1fu2015; -1)*. 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 *(gauzere2019-1)*. 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 (Fig. ED4, S2). 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 (see Supplemental Materials). 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), though this varied with local climate prior to warming (Fig. ED6, ED6, S5). Portions of Central Europe have experienced more dramatic warming in winter versus summer *(li2015-1balling1998; -1,* 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, though responses vary by species (Fig. 3, ED6). 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* (Fig. ED7).

Our predictions leave open the question of what underlies declining sensitivities across Europe, but one possibility is that it may be a statistical artifact of how temperature sensitivities are calculated. Physiologically, budburst is triggered by the accumulation of forcing temperatures during the spring *(chuine2016-1hanninen1995; -1)*. 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 *(e.g., Wolkovich:2012n-1)*. 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 (Fig. ED8). We found the declining sensitivities observed in European data are of the same magnitude as those predicted from a statistical artifact (sensitivity declines of 0.8±0.3 days/°C in European data versus 0.9±0.5 days/°C 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 artifacts in declines of temperature sensitivity in observational long-term data* in the Supplemental Materials and gusewell2017-1, for further details). This statistical artifact is likely not confined to phenological studies; it should apply to any research using a similar days/°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 *(e.g., roberts2015-1)*. 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 *(chuine2016-1Singh:2017; -1)*.

Correspondence Statement

Please direct correspondence and requests related to this article to the corresponding author, A. K. Ettinger, ailene.ettinger@tnc.org

Statement of Author Contributions

DF, TS, and EW 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. CC, DB, EW, IM, and AE created the figures. AE and EW wrote the majority of the manuscript, with substantial contributions from CC, DB, and IM; all authors reviewed and revised the manuscript.

Data & Code Availability

The OSPREE budburst database and code for models fit in this manuscript are publicly archived at KNB, doi:10.5063/F1QV3JQR (wolkovich2019-1).

Acknowledgements

References

IPCC:2014smIPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014). millerrushing2008Miller-Rushing, A. J. & Primack, R. B. Global warming and flowering times in Thoreau's Concord: A community perspective. *Ecology* **89**, 332—341 (2008). menzel2006Menzel, A. *et al.* European phenological response to climate change matches the warming pattern. *Global Change Biology* **12**, 1969–1976 (2006). Cleland:2007orCleland, E. E., Chuine, I., Menzel, A., Mooney, H. A. & Schwartz, M. D. Shifting plant phenology in response to global change. *Trends in Ecology & Evolution* **22**, 357–365 (2007). Wolkovich:2012nWolkovich, E. M. *et al.* Warming experiments underpredict plant phenological responses to climate change. *Nature* **485**, 494–497 (2012). Rutishauser:2008Rutishauser, T., Luterbacher, J., Defila, C., Frank, D. & Wanner, H. Swiss spring plant phenology 2007: Extremes, a multi-century perspective, and changes in temperature sensitivity. *Geophysical Research Letters* **35**, L05703 (2008). yu2010Yu, H. Y., Luedeling, E. & Xu, J. C. Winter and spring warming result in delayed spring phenology on the Tibetan Plateau. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22151–22156 (2010). wang2019Wang, X. *et al.* No trends in spring and autumn phenology during the global warming hiatus. *Nature communications* **10**, 2389 (2019). fu2015Fu, Y. S. H. *et al.* Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**, 104–107 (2015). chuine2016Chuine, I. *et al.* Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Global Change Biology* **22**, 3444–3460 (2016). harrington2015Harrington, C. A. & Gould, P. J. Tradeoffs between chilling and forcing in satisfying dormancy requirements for pacific northwest tree species. *Frontiers in Plant Science* **6**, 120 (2015). zohner2016Zohner, 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. *Nature Climate Change* **6**, 1120–1123 (2016). Basler:2014aaBasler, D. & Körner, C. Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. *Tree Physiology* **34**, 377–388 (2014). Caffarra:2011bCaffarra, A., Donnelly, A., Chuine, I. & Jones, M. B. Modelling the timing of *Betula pubescens* bud-burst. I. Temperature and photoperiod: A conceptual model. *Climate Research* **46**, 147 (2011). flynn2018Flynn, D. F. B. & Wolkovich, E. M. Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytologist* **219**, 1353–1362 (2018). Caffarra:2011qfCaffarra, A., Donnelly, A. & Chuine, I. Modelling the timing of *Betula pubescens* bud-burst. II. Integrating complex effects of photoperiod into process-based models. *Climate Research* **46**,

159–170 (2011). fraga2019Fraga, 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* 1–15 (2019). Heide:1993Heide, O. Daylength and thermal time responses of budburst during dormancy release in some northern deciduous trees. *Physiologia Plantarum* **88**, 531–540 (1993). Singh:2017Singh, 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 Phytologist* **213**, 511–524 (2017). vitasse2013Vitasse, Y. & Basler, D. What role for photoperiod in the bud burst phenology of European beech. *European Journal of Forest Research* **132**, 1–8 (2013). vitasse2014Vitasse, Y. & Basler, D. Is the use of cuttings a good proxy to explore phenological responses of temperate forests in warming and photoperiod experiments? *Tree physiology* **34**, 174–183 (2014). Laube:2014aLaube, J. *et al.* Chilling outweighs photoperiod in preventing precocious spring development. *Global Change Biology* **20**, 170–182 (2014). Basler:2012Basler, D. & Körner, C. Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology* **165**, 73–81 (2012). Caffarra:2011aCaffarra, A. & Donnelly, A. The ecological significance of phenology in four different tree species: effects of light and temperature on bud burst. *International Journal of Biometeorology* **55**, 711–721 (2011). ohlemuller2006Ohlemüller, 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. *Global Ecology and Biogeography* **15**, 395–405 (2006). williams2007Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment* **5**, 475–482 (2007). williams2007bWilliams, J. W., Jackson, S. T. & Kutzbacht, J. E. Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5738–5742 (2007). ipcc2013Stocker, T., Qin, D. & Platner, G. Climate change 2013: The physical science basis. *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013)* (2013). xu2018Xu, Y., Ramanathan, V. & Victor, D. G. Global warming will happen faster than we think (2018). wolkovich2019Wolkovich, E. M. *et al.* Observed Spring Phenology Responses in Experimental Environments (OSPREE). doi:10.5063/F1QV3JQR (2019). richardson1974Richardson, E. A model for estimating the completion of rest for 'Redhaven' and 'Elberta' peach trees. *HortScience* **9**, 331–332 (1974). dennis2003Dennis, 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). gelman2006Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models*

11

²⁷² (Cambridge University Press, 2006). fu2019Fu, Y. H. *et al.* Short photoperiod reduces the temperature

²⁷³ sensitivity of leaf-out in saplings of *Fagus sylvatica* but not in horse chestnut. *Global Change Biology*

²⁷⁴ **25**, 1696–1703 (2019). bradley1999Bradley, N. L., Leopold, A. C., Ross, J. & Huffaker, W. Phenological

²⁷⁵ changes reflect climate change in Wisconsin. *Proceedings of the National Academy of Sciences* **96**, 9701–

²⁷⁶ 9704 (1999). gauzere2019Gauzere, J., Lucas, C., Ronce, O., Davi, H. & Chuine, I. Sensitivity analysis of

²⁷⁷ tree phenology models reveals increasing sensitivity of their predictions to winter chilling temperature

²⁷⁸ and photoperiod with warming climate. *Ecological Modelling* **441**, 108805 (2019). Heide:2005aaHeide, O.

²⁷⁹ & Prestrud, A. Low temperature, but not photoperiod, controls growth cessation and dormancy induc-

²⁸⁰ tion and release in apple and pear. *Tree Physiology* **25**, 109–114 (2005). vanderschoot2014van der Schoot,

²⁸¹ C., Paul, L. K. & Rinne, P. L. H. The embryonic shoot: a lifeline through winter. *Journal of Experimen-*

²⁸² *tal Botany* **65**, 1699–1712 (2014). fishman1987Fishman, S., Erez, A. & Couvillon, G. The temperature

²⁸³ dependence of dormancy breaking in plants: mathematical analysis of a two-step model involving a

²⁸⁴ cooperative transition. *Journal of Theoretical Biology* **124**, 473–483 (1987). weinberger1950Weinberger,

²⁸⁵ J. H. *et al.* Chilling requirements of peach varieties. In *Proceedings. American Society for Horticultural*

²⁸⁶ *Science*, vol. 56, 122–28 (1950). polgar2013Polgar, C. A., Primack, R. B., Williams, E. H., Stichter, S.

²⁸⁷ & Hitchcock, C. Climate effects on the flight period of Lycaenid butterflies in Massachusetts. *Biolog-*

²⁸⁸ *ical Conservation* **160**, 25–31 (2013). vitasse2013ontVitasse, Y. Ontogenic changes rather than differ-

²⁸⁹ ence in temperature cause understory trees to leaf out earlier. *New Phytologist* **198**, 149–155 (2013).

²⁹⁰ Laube:2014bLaube, J., Sparks, T. H., Estrella, N. & Menzel, A. Does humidity trigger tree phenology?

²⁹¹ Proposal for an air humidity based framework for bud development in spring. *New Phytologist* **202**,

²⁹² 350–355 (2014). li2015Li, C., Stevens, B. & Marotzke, J. Eurasian winter cooling in the warming hiatus

²⁹³ of 1998–2012. *Geophysical Research Letters* **42**, 8131–8139 (2015). balling1998Balling, R. C. J., Michaels,

²⁹⁴ P. J. & Knappenberger, P. C. Analysis of winter and summer warming rates in gridded temperature

²⁹⁵ time series. *Climate Research* **9**, 175–181 (1998). hanninen1995Hänninen, H. Effects of climatic change

²⁹⁶ on trees from cool and temperate regions: an ecophysiological approach to modelling of bud burst phe-

²⁹⁷ nology. *Canadian Journal of Botany* **73**, 183–199 (1995). gusewell2017Güsewell, S., Furrer, R., Gehrig,

²⁹⁸ R. & Pietragalla, B. Changes in temperature sensitivity of spring phenology with recent climate warm-

²⁹⁹ ing in Switzerland are related to shifts of the preseason. *Global Change Biology* **23**, 5189–5202 (2017).

³⁰⁰ roberts2015Roberts, A. M., Tansey, C., Smithers, R. J. & Phillimore, A. B. Predicting a change in the

³⁰¹ order of spring phenology in temperate forests. *Global Change Biology* **21**, 2603–2611 (2015).
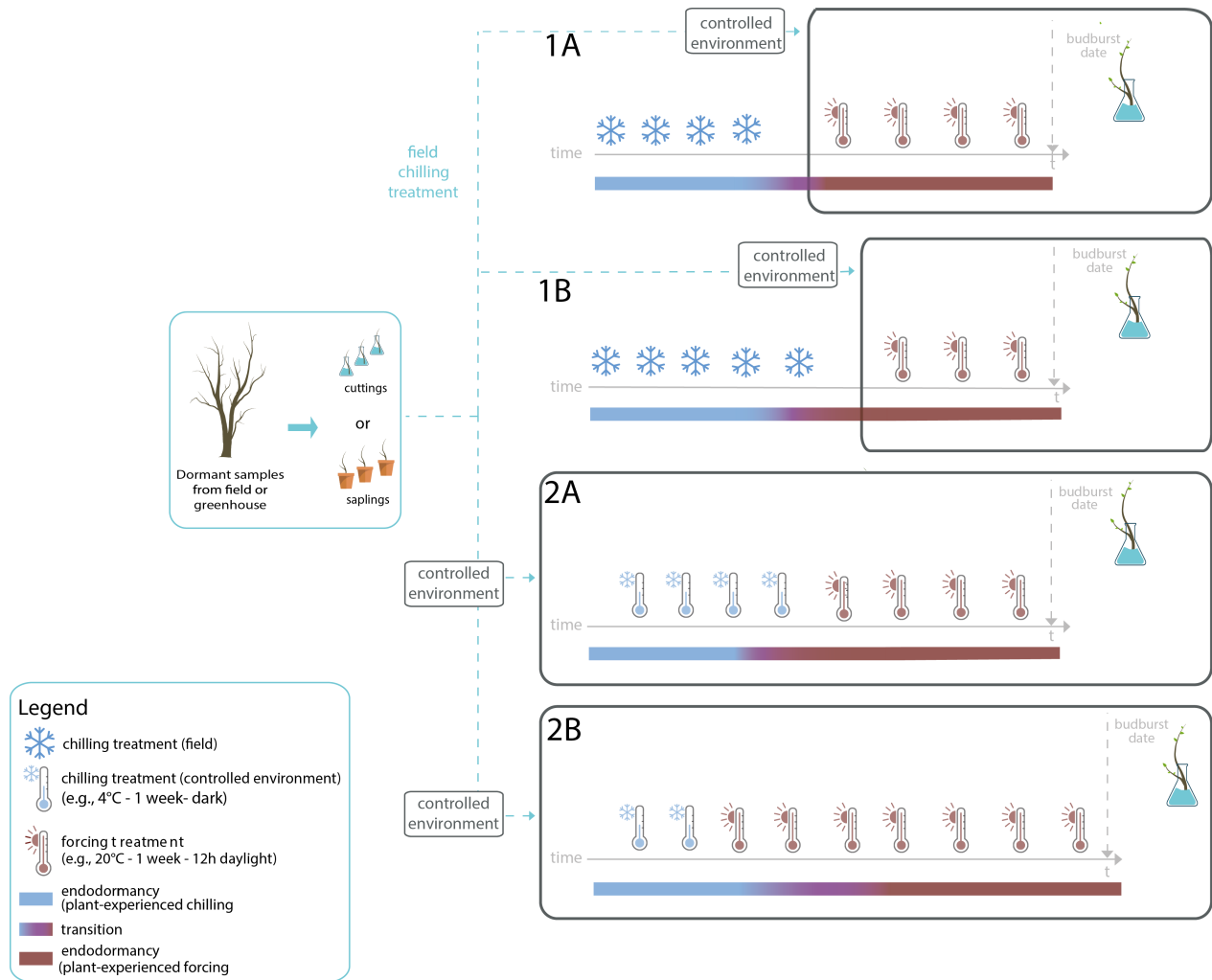
Figures

Figure 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-B, in which plant material is collected after different numbers of days in the fall/winter) and/or experimentally (2A-B, 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 (e.g., 1A, though 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 (*i.e.*, 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 treatments varied uniquely for each study, but some were more common than others, see Fig. ED3: chilling treatments across the averaged 71.4 days (range: 1-182 days) at an average temperature of 4.4 °C (range: 0-16 °C), forcing treatments averaged 15.7 °C (range: 5 to 32 °C).
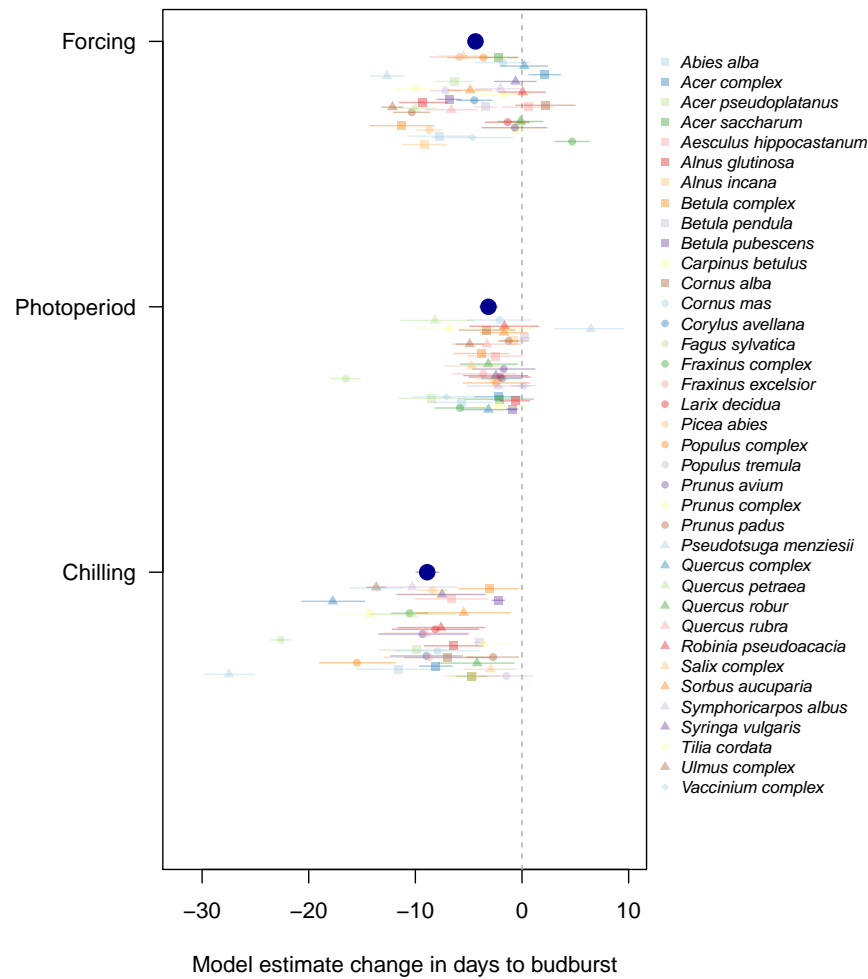
Figure 2: **Estimated effects of chilling, forcing, and photoperiod on budburst timing across 65 species (modeled as 36 separate taxa, see the _Models_ section of _Methods_) in 39 controlled environment experiments.** Using standardized units, which allow comparisons across cues, we show that most species (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 (Fig. ED5, S3; Tables S5-S12). Lines represent 50% uncertainty intervals (other intervals provided in Tables S5-S12)

.

Methods: Winter temperatures predominate in spring phenological responses to warming

The Observed Spring Phenology Responses in Experimental Environments (OS-PREE) database

To conduct this meta-analysis, we followed systematic review methods to facilitate replication and use by other researchers *(e.g.,* we include at least 22/27 items on the PRISMA checklist, as summarized in Appendix
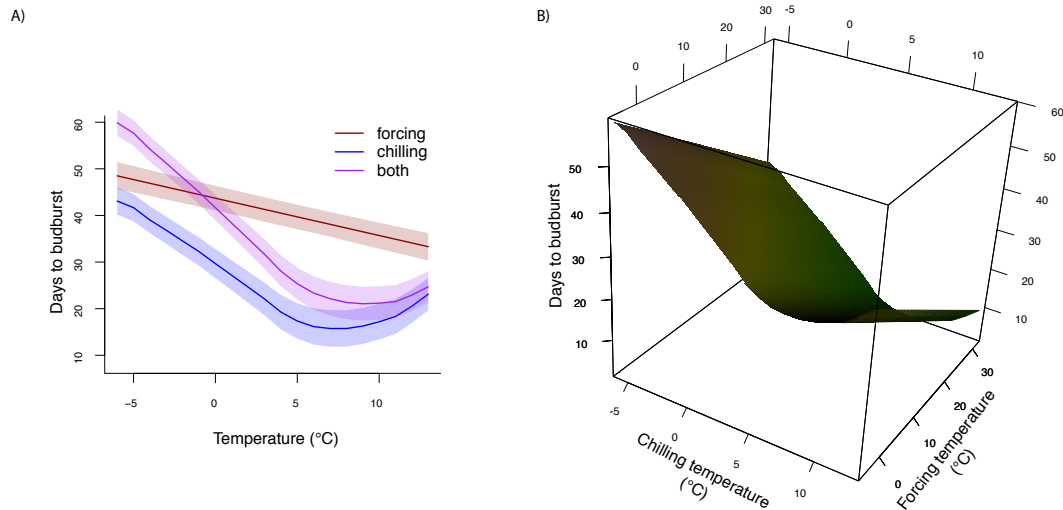
Figure 3: **Estimates of budburst across a range of forcing temperatures and estimated chilling** (converted to a representative mean temperature, see *Estimating chilling* in Methods) based on overall estimates of chilling and forcing effects from a meta-analysis of short-term experiments using controlled 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. Panel A shows 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 (1324 chilling units), and the effect of varying both chilling and forcing temperatures simultaneously. Panel B shows all possible combinations of chilling and forcing across the experimental conditions. Maximum advances in budburst occur at intermediate chilling temperatures (*e.g.*, here at mean winter temperatures of 6-7°C) and the highest forcing (here at 32°C). We set photoperiod to eight hours, which is the most common photoperiod treatment in our meta-analysis.

309   1, *moher2009-1).* We searched the literature for research papers that experimentally addressed controls of

310   temperature, chilling, and/or photoperiod requirements on the spring phenology of woody plant species. To

311   identify phenological experiments that manipulated chilling, forcing, and/or photoperiod, we searched both

312   ISI Web of Science and Google Scholar with the following terms:

313   **[40]** TOPIC = (budburst OR leaf-out) AND (photoperiod or daylength) AND temperature*, which yielded

314   85 publications

315   2. TOPIC = (budburst OR leaf-out) AND dorman*, which yielded 193 publications

316   The initial searches yielded 201 papers, which we reviewed and assessed for inclusion in the database. To be

317   included, papers needed to focus on woody plants in temperate ecosystems and test for photoperiod and/or
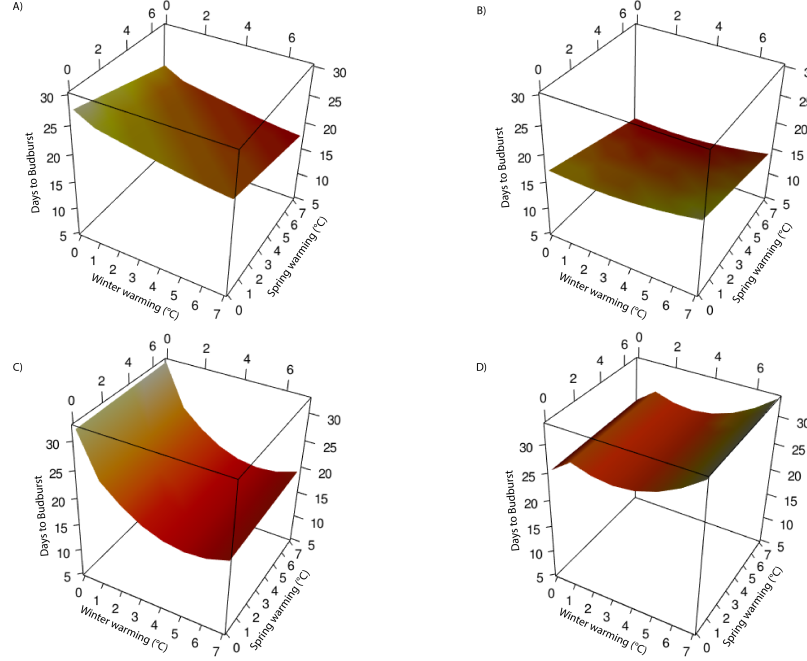
Figure 4: **Implications of warming on budburst timing varies across species and sites**, depending strongly on pre-warming climate conditions related to chilling for each site. Here we show species-level estimates from our model based on a meta-analysis of experiments (Fig. 2) for two common species *Betula pendula* (A, B) and *Fagus sylvatica* (C, D), based on climate data from two sites in Central Europe (these two sites chosen to highlight the diversity of possible budburst responses to warming, see Fig. ED6 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 to forcing alone), whereas warming decreases total chilling estimates in other sites (B, D), leading to smaller advances and, eventually, delays with substantial warming. See Fig. S4 in the Supplemental Materials for a simplified two-dimensional version.

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 be published *(Kicinski2013-1)*. 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 *(Gurevitch1992-1Gurevitch1999; -1Lin2018; -1)*. 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 10) and most often sample size was not reported (*i.e.*, in 25 out of 39 studies in the model with well-represented species), and finally (3) our modeling 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 analyzing "all species."

Although all studies measured days to budburst, many communicated results differently (*e.g.*, days to budburst, degree-days to budburst, percent budburst, 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, Table S1), 39 years, and 203 species (Table S2, Fig. ED1). These experiments span a wide range of chilling, forcing, and photoperiod treatment levels (Fig. ED2), and many test for interactions between two of these cues (Table S3). This subset of OSRPREE is freely available on KNB *(wolkovich2019-1)* and we hope other researchers will find it useful.

Defining budburst

Most studies defined budburst as initial "green tips" (33/49 papers). Select studies defined budburst as a specific increment of growth (*e.g.*, "0.5 cm of new growth") or as bud swell, leaf emergence, leaf unfolded, open bud scales, or petiole emerged. The remaining papers (4/49) did not include a definition of budburst. The majority of papers using the above definitions (34/49) required only one bud to have met the defined criteria of budburst, however, the remaining studies implemented specific thresholds to be met (*i.e.*, 10-100% of all buds on an individual needed to have bursted bud). For studies that quantified multiple measurements of percent

18

359 budburst over time (days), we extracted one value of "days to budburst" of these multiple measurements to

360 make them comparable to other studies. To extract this summary value, we selected the days to budburst

361 when percent budburst was closest to 90%, including estimates as low as 49.5% budburst.

Estimating chilling

363 Chilling was reported far less in the OSPREE database than forcing and photoperiod. Although not all

364 studies applied multiple treatments of forcing or photoperiod they generally all maintained and explicitly

365 defined the forcing temperatures and daylengths applied in their treatments. In contrast, we found that most

366 studies did not experimentally apply chilling by manipulating duration or temperature of chilling in controlled

367 environments, nor did most quantify the total chilling imposed in their experiment. We therefore calculated

368 the total chilling imposed by all studies; it would otherwise have been impossible to provide estimates with

369 only experimental chilling given the rarity of such study designs (Fig. ED3).

370 To estimate total chilling we combined chilling from the field (*i.e.*, chilling before plant material was brought

371 into controlled environment conditions) and experimental chilling (*i.e.*, chilling that plant material expe-

372 rienced in controlled environment conditions) into two widely used metrics of chilling: Utah units (Table

373 S4) and dynamic chill portions *(dennis2003-1Luedeling:2011qe; -1)*. We used the `chillR` package (version

374 0.70.17) in R *(Rcore:2017-1chillR2019; -1)*, version 3.6.0, to calculate both Utah units and dynamic chill

375 portions from time-series of hourly temperature data. To estimate field chilling, we generated hourly time

376 series from a European-wide gridded dataset *(cornes2018-1)* and a North American gridded climate dataset

377 *(livneh2015-1)*, from which we extracted daily minimum and maximum temperature from the grid cells

378 and dates during which experiments were conducted. For experimental chilling, we used reported chilling

379 treatments to generate time-series of hourly temperature data.

380 In the formulation we used, Utah chilling units accumulated the most at temperatures between 2.4-9.1°C

381 but slightly less at temperatures between 1.4-2.4°C and from 9.1-12.4°C. Utah units were reduced when

382 temperatures fell below or exceeded this range (Table S4). Chill portions accumulated when temperatures

383 were between 0 and 7.2°C. We note that these models for chilling (both of which were originally developed

384 for peach species) are *hypotheses* for how chilling may accumulate to affect the process of endodormancy

385 release, but are likely to be inaccurate for many species. These models are, however, some of our current

386 best approximations, and versions of them are routinely applied to forest trees *(e.g., Harrington:2010-1)*.

387 We found the effects of chilling and other cues remained qualitatively consistent across the two methods of

388 estimating total chilling (*i.e.*, 95% uncertainty intervals of estimates for all cues in the standardized models

19

overlapped, Table S5).

We wished to explore model predictions across a wide range of experimental temperature conditions (*i.e.*, 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 (Fig. ED4). 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 to 240 days). We averaged across the range of durations for each temperature to get one chilling estimate per chilling temperature (Fig. ED4, S2). For the alternative approach, we used historical climate data from a gridded climate dataset *(*E-OBS*, cornes2018-1)* 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 & 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 (*e.g.*, a study with 8 hours of light and 16 hours of dark was recorded as "8"). For forcing, we used the temperature applied when forcing temperatures were constant (*i.e.*, the same temperature was applied 24 hours 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 (*e.g.*, studies with photoperiod lengthening from 6 hours 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 *(cornes2018-1)*.

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 *(gelman2006-1)*. 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 4724 Utah units; forcing ranged from 5 to 32°C, photoperiod ranged from 6 to 24 hours). We followed well-established methods of subtracting the mean and dividing by the standard deviation *(gelman2006-1)* 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 (Table S5), we also fit models in which predictors were not standardized (Table S6) 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` *(Carpenter:2016aa-1)* (`www.mc-stan.org`), accessed via the `rstan` package (version 2.18.0) in R *(Rcore:2017-1rstan2018; -1)*, version 3.6.0. `Stan` provides efficient MCMC sampling via a No-U-Turn Hamiltonian Monte Carlo approach (more details can be found in *(BDA-1)* and in *(Carpenter:2016aa-1)*). 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.

1. Main budburst model:

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

,

21

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

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

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

$$\beta_{forcing_{sp}} \sim N(\mu_{forcing}, \sigma_{forcing})$$

$$\beta_{photoperiod_{sp}} \sim N(\mu_{photoperiod}, \sigma_{photoperiod})$$

$$\beta_{chillingsp} \sim N(\mu_{chilling}, \sigma_{chilling})$$

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" (Tables S5, S6) and a model excluding a single study (zohner2016-1) because this study contains 112 species (Table S9). We present estimates from the model fit to the reduced dataset in the main text (including for Figs. 2-3) 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 above in section describing the OSPREE database). Based on our modeling 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).

2. <u>Latitude model</u>:

Given continuing debate over the role of photoperiod on budburst timing across a species' latitudinal range *(e.g., zohner2016-1gauzere2017; -1)*, 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, photoperiod) on days to budburst (as in the main model), in addition to the effect of provenance latitude (i..e., 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 *(gauzere2017-1saikkonen2012; -1way2015; -1)*.

We followed the methods above for including species or species complex (see *Observed Spring Phenology Responses in Experimental Environments (OSPREE) database* section above), including only species

22

and species complexes that had multiple provenance locations across different latitudes. This yielded the following model:

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

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

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

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

$$\beta_{forcing_{sp}} \sim N(\mu_{forcing}, \sigma_{forcing})$$

$$\beta_{photoperiod_{sp}} \sim N(\mu_{photoperiod}, \sigma_{photoperiod})$$

$$\beta_{chilling_{sp}} \sim N(\mu_{chilling}, \sigma_{chilling})$$

$$\beta_{latitude_{sp}} \sim N(\mu_{latitude}, \sigma_{latitude})$$

$$\beta_{photoperiod:latitude_{sp}} \sim N(\mu_{photoperiod:latitude}, \sigma_{photoperiod:latitude})$$

The latitude model is summarized in Table S10 and Fig. ED5.

3. Chilling study design model: As we found chilling to be the strongest cue, and given how few studies directly manipulate it (Fig. ED3), 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 employed 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 *(weinberger1950-1)* and those that experimentally manipulated chilling (*i.e.*, 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 two or more weeks

23

apart, that did not otherwise manipulate chilling. The chilling study-design model was thus:

$$y_i = \alpha_{sp[i]} + \beta_{forcing} + \beta_{photoperiod} + \beta_{chilling} + \beta_{chillmethod} +$$

$$\beta_{forcing:chillmethod} + \beta_{photoperiod:chillmethod} + \beta_{chilling:chillmethod} + \epsilon_i,$$

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

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

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

466   The chilling design model is summarized in Table S11 and and Fig. S3.

467 4. <u>Life stage model</u>: Previous research has found differences in spring phenology across life stages *(e.g.,*
468   *juvenile versus adult trees vitasse2013ont-1)*. We tested for differences in days to budburst across life
469   stages.

We followed the guidelines above for including species or species complex (see *Observed Spring Phenology Responses in Experimental Environments (OSPREE) database* section above), 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 modeled at the species level, as follows:

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

$$\beta_{forcing_{sp}} \sim N(\mu_{forcing}, \sigma_{forcing})$$

$$\beta_{photoperiod_{sp}} \sim N(\mu_{photoperiod}, \sigma_{photoperiod})$$

$$\beta_{chilling_{sp}} \sim N(\mu_{chilling}, \sigma_{chilling})$$

$$\beta_{stage_{sp}} \sim N(\mu_{stage}, \sigma_{stage})$$

470   The life-stage model is summarized in Table S12.

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

476 In our figures we show means $\pm$ 50% uncertainty intervals from our models (Figs. 2, 3, ED5, S3, S4, ED7),
477 because our focus here is on the most likely value for each parameter (*e.g.*, estimated response to forcing)
478 and because they are computationally stable *(Carpenter:2016aa-1BDA; -1)*. See Tables S5- S12 for 95%
479 uncertainty intervals.

480 *Modeling limitations based on experimental designs*

481 An ideal model to predict budburst would potentially include (but is not limited to): interactions between
482 cues, sigmoidal or other non-linearities to assess potential threshold effects, provenance location, method-
483 ological details (*e.g.*, if plant material was whole plant versus twigs, or whether temperatures were constant
484 or varied each day, etc.), and measurement error. As with all models, though, we were limited in how many
485 parameters we could estimate given available data. Thus we focused on species differences and used addi-
486 tional models to assess some of the potentially largest other effects (latitude, methods of estimating chilling,
487 life stage). We were unable to estimate interactions between cues in our meta-analysis because very few
488 studies design experiments to test for interactions between chilling, forcing, and photoperiod (Table S3).
489 Most experiments in our dataset, however, did include interactions between at least two cues (Table S3); we

fit our main budburst model to this subset of experiments (Table S8), which resulted in qualitatively similar estimates to those of the model fit to the full set of experiments (Table S5).

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. Non-linear 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 non-linear (sigmoidal) model on the OSPREE data *(pmp-1)*. 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-squared = 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 non-linear 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 (Table S11, Fig. S3). Similarly we found variation in thermoperiodicity and variation in study material were too infrequent to test for effects with current data (though our life-stage model found no large differences in days to budburst between material from adults [μ_α = 25.29] versus juveniles [β_{stage} = 24.2; 50% uncertainty intervals overlap], Table S12). Our estimated effects therefore average over interactions *(gelman2006-1)*, 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 *(chuine2016-1)*.

Our model does not include measurement error because these data were not possible to include for 25 out of 39 experiments included in the 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 (*e.g.*, 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 *(rstan2019-1)*.

References

moher2009Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine* **151**, 264–269 (2009). Kicinski2013Kicinski, M. Publication bias in recent meta-analyses. *Plos One* **8** (2013). Gurevitch1992Gurevitch, J., Morrow, L. L., Wallace, A. & Walsh, J. S. A meta-analysis of competition in field experiments. *American Naturalist* **140**, 539–572 (1992). Gurevitch1999Gurevitch, J. & Hedges, L. V. Statistical issues in ecological meta-analyses. *Ecology* **80**, 1142–1149 (1999). Lin2018Lin, L. F. & Chu, H. T. Quantifying publication bias in meta-analysis. *Biometrics* **74**, 785–794 (2018). wolkovich2019Wolkovich, E. M. *et al.* Observed Spring Phenology Responses in Experimental Environments (OSPREE). doi:10.5063/F1QV3JQR (2019). dennis2003Dennis, 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). Luedeling:2011qeLuedeling, E. & Brown, P. H. A global analysis of the comparability of winter chill models for fruit and nut trees. *International Journal of Biometeorology* **55**, 411–421 (2011). Rcore:2017R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2017). chillR2019Luedeling, E. *chillR: Statistical Methods for Phenology Analysis in Temperate Fruit Trees* (2019). R package version 0.70.17. cornes2018Cornes, 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. *Journal of Geophysical Research: Atmospheres* **123**, 9391–9409 (2018). livneh2015Livneh, B. *et al.* A spatially comprehensive, hydrometeorological data set for mexico, the us, and southern canada 1950–2013. *Scientific data* **2**, 1–12 (2015). Harrington:2010Harrington, C. A., Gould, P. J. & St Clair, J. B. Modeling the effects of winter environment on dormancy release of Douglas-fir. *Forest Ecology and Management* **259**, 798–808 (2010). gelman2006Gelman, A. & Hill, J. *Data analysis using regression and multilevel/hierarchical models* (Cambridge University Press, 2006). Carpenter:2016aaCarpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **76**, 10.18637/jss.v076.i01 (2017). rstan2018Stan Development Team. RStan: the R interface to Stan (2018). R package version 2.17.3. BDAGelman, A. *et al. Bayesian Data Analysis* (CRC Press, New York, 2014), 3rd edn. zohner2016Zohner, 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. *Nature Climate Change* **6**, 1120–1123 (2016). gauzere2017Gauzere, 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. Agricultural and*

Forest Meteorology **244**, 9–20 (2017). saikkonen2012Saikkonen, K. *et al.* Climate change-driven species' range shifts filtered by photoperiodism. *Nature Climate Change* **2**, 239 (2012). way2015Way, D. A. & Montgomery, R. A. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment* **38**, 1725–1736 (2015). weinberger1950Weinberger, J. H. *et al.* Chilling requirements of peach varieties. In *Proceedings. American Society for Horticultural Science*, vol. 56, 122–28 (1950). vitasse2013ontVitasse, Y. Ontogenic changes rather than difference in temperature cause understory trees to leaf out earlier. *New Phytologist* **198**, 149–155 (2013). pmpChuine, I., Garcia de Cortazar Atauri, I., Hanninen, H. & Kramer, K. *Plant development models*, 275–293. (Kluwer, Dordrecht, the Netherlands, 2013). chuine2016Chuine, I. *et al.* Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Global Change Biology* **22**, 3444–3460 (2016). rstan2019Stan Development Team. Stan user's guide, version 2.19 (2019).