Herbarium specimens reveal substantial and unexpected variation in phenological sensitivity across the eastern United States

Daniel S. Park1,†, Ian Breckheimer1,†, Alex C. Williams2, Edith Law2, Aaron M. Ellison3 and Charles C. Davis1

1Department of Organismic and Evolutionary Biology and Harvard University Herbaria, Harvard University, Cambridge, MA 02138, USA
2David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
3Harvard Forest, Harvard University, Petersham, MA 01366, USA

Phenology is a key biological trait that can determine an organism’s survival and provides one of the clearest indicators of the effects of recent climatic change. Long time-series observations of plant phenology collected at continental scales could clarify latitudinal and regional patterns of plant responses and illuminate drivers of that variation, but few such datasets exist. Here, we use the web tool CrowdCurio to crowdsource phenological data from over 7000 herbarium specimens representing 30 diverse flowering plant species distributed across the eastern United States. Our results, spanning 120 years and generated from over 2000 crowdsourcers, illustrate numerous aspects of continental-scale plant reproductive phenology. First, they support prior studies that found plant reproductive phenology significantly advances in response to warming, especially for early-flowering species. Second, they reveal that fruiting in populations from warmer, lower latitudes is significantly more phenologically sensitive to temperature than that for populations from colder, higher-latitude regions. Last, we found that variation in phenological sensitivities to climate within species between regions was of similar magnitude to variation between species. Overall, our results suggest that phenological responses to anthropogenic climate change will be heterogeneous within communities and across regions, with large amounts of regional variability driven by local adaptation, phenotypic plasticity and differences in species assemblages. As millions of imaged herbarium specimens become available online, they will play an increasingly critical role in revealing large-scale patterns within assemblages and across continents that ultimately can improve forecasts of the impacts of climatic change on the structure and function of ecosystems.

This article is part of the theme issue ‘Biological collections for understanding biodiversity in the Anthropocene’.

1. Introduction

Ecosystems on every continent have been affected by local, regional and global changes in climate, especially increases in temperature [1]. Changes in phenology—the timing of life-history events—are among the most conspicuous and well-documented species responses to climatic change, especially for plants [2–7]. Phenological disruption has already impacted species’ local persistence and community diversity [8–10], which may have widespread consequences for critical ecosystem processes, including carbon sequestration [11–13], ecosystem–atmosphere interactions [14] and trophic interactions [15–28].

Despite these trends, our knowledge of plant phenological responses to climatic change remains inadequate. In particular, although phenological responses may differ among species with different functional or life-history traits and biogeographical origins [29–32], long-term observational datasets...
to assess such trends are limited in geographical, temporal and taxonomic scope [33]. Many of these data track woody plant species of the Northern Hemisphere (most commonly, abundant tree species), and only for the last approximately 40 years (but see [34,35]). These biases limit our understanding of variation in phenological responses across species and biomes. Furthermore, although population-level variation in phenology has been demonstrated for a few species [36–38], there are very few studies that quantify both inter- and intraspecific variation in phenological response [32,33].

Variability in species’ phenology is particularly relevant because climatic change is not geographically uniform. For example, high-latitude regions are warming faster than subtropical and tropical ones [1,39]. Short growing seasons also may cause high-latitude ecosystems to be especially sensitive to temperature, leading to stronger selective pressures for populations to initiate growth as soon as conditions become favourable in early spring. Additionally, plants adapted to highly variable climates may exhibit higher phenological thresholds to temperature, as it provides a less reliable signal [40]. Thus, the effects of climatic change on species’ phenology may differ across their ranges depending on variation in phenological sensitivity. Variability in phenological responses to climatic change within species may also alter patterns of gene flow, which could either promote or counteract adaptive evolution via the sharing of locally (mal)adapted alleles [41–44]. Recent studies have shown that plant phenology may be more responsive in more northern-ranging populations where there are more variable and extreme climates, especially during the early part of the growing season [44,45]. Although these studies are suggestive, they are restricted in spatial scale and taxonomic scope, and broad regional patterns of phenological response to climate may differ from patterns at smaller, local scales.

Herbarium specimens represent snapshots of phenology (i.e. flowering and fruiting) at a specific place and time, and have shown tremendous promise to increase the spatial, temporal and taxonomic resolution of phenological data [46–48]. They provide rich historical depth, wide geographical scope and taxonomic diversity, all of which allow researchers to track long-term changes of vast numbers of species and communities through space and time [4,46–49]. Despite their representation of phenological responses [48,50], herbarium specimens have been used less frequently than other data sources, such as field observations, to address phenological change, in part because they have been inaccessible to many researchers [46,51]. However, the widespread digitization of herbarium collections [52] combined with new approaches to collecting [46,53] and analysing [54,55] phenological data derived from herbarium specimens has the power to transform our understanding of plant responses to global climatic change.

Here, we applied a newly developed web-enabled crowdsourcing platform, CrowdCurio: Thoreau’s Field Notes (https://www.crowdcurio.com/) [46], to examine more than 7000 specimens of 30 phylogenetically diverse flowering plant species. The Thoreau’s Field Notes module facilitates the rapid quantification of phenological traits via image annotation and has been demonstrated to yield reliable data regardless of the level of expertise among crowdsourcers (i.e. expert versus non-expert scoring) [46]. We used these crowdsourced data to infer the magnitude, direction, and variability in reproductive phenological responses to spring temperature across 23° of latitude in the eastern United States. We examined both native and introduced plant species from northern coniferous forests, eastern deciduous forests, subtropical evergreen forests, grasslands, wetlands, alpine meadows and aquatic plant communities. Environmental conditions in this region vary considerably across species’ ranges, and populations may exhibit substantial variation in phenological response across this latitudinal gradient. Our overall goals with this study were: (i) to demonstrate the power of characterizing phenology from herbarium data using an efficient and rapid workflow that leverages a nearly fully mobilized online flora of the eastern United States [56,57]; (ii) to greatly increase the taxonomic and ecological diversity of species sampled for this purpose (from woody perennials to herbaceous annuals across a range of biomes); and (iii) to sample species with broad latitudinal ranges to assess regional and inter- and intra-species variation in phenological responses.

2. Methods

(a) Specimen data collection

We examined phenological responses of species using digitized specimens from two of the most comprehensive digitized regional floras in the world, the Consortium of Northeastern Herbaria (CNH; http://portal.neherbaria.org/portal/) [56] and Southeast Regional Network of Expertise and Collections (SERNEC; http://serneckportal.org/portal/index.php) [57]. These two online portals include more than six million digitized herbarium records, including specimen images. Our criteria for selecting angiosperm species for analysis were that specimens: (i) included at least county-level location data; (ii) included at least 50 unique collections across space and time; (iii) were of species with relatively easily identifiable and quantifiable reproductive structures; and (iv) were from species with broad latitudinal ranges sufficient to enable quantification of population-level variation.

Applying these criteria yielded 30 species with varying life-history traits, growth forms, native status and general reproductive seasonality (e.g. early- versus late-spring flowering). We downloaded over 10 000 digital herbarium specimen images of these species from CNH and SERNEC, removed duplicate, misidentified or sterile specimens, and those with notable insect damage on reproductive structures, extensive physical damage or poor preservation. We also removed all 30 specimens from Florida, which were geographical and climatic outliers. Our final dataset comprised 7722 specimens and spanned 120 years across 512 United States counties (table 1). Species’ life-history (annual versus perennial), growth form (woody versus herbaceous) and native status (native versus introduced) were inferred from the United States Department of Agriculture PLANTS Database (https://plants.usda.gov/). For individual specimen metadata, see electronic supplementary material, table S1.

(b) Crowdsourcing phenological data collection

The phenological state (phenophase) of a plant can be inferred from the presence and quantity of relevant structures, such as leaves, flowers or fruits [46,48]. Past researchers generally have focused on the presence or the absence of a single structure or trait (e.g. [58]) or applied majority estimates for scoring a single phenophase (e.g. [49]). Here, we quantified data for two reproductive phenophases, flowering and fruiting. Specimens were scored as flowering if open flowers represented greater than or equal to 50% of the total reproductive structure count.
and scored as fruiting if they had less than 50% flowers and buds and at least one fruit present. We used the Thoreau’s Field Notes instance of CrowdCurio to crowdsource phenological data from digitized herbarium specimens [46]. Citizen scientists hired through Amazon’s Mechanical Turk service (MTurk; https://www.mturk.com/) counted the number of buds, flowers and fruits observed for a set of 10 specimen images. Participants first watched a short (1 min) instructional video on how to score phenological traits using CrowdCurio and then were provided with three tutorial images of each reproductive structure for every species. The 2364 anonymous participants were compensated at a rate of $0.12 per image.

To provide an estimate of measurement error, each 10-image set was scored by a single crowdsourcer included nine unique images and a single duplicate image randomly selected from the other nine [55]. We estimated the reliability score for each participant based on the data for each 10-image set by dividing the absolute difference in organ counts for each phenophase by the total count of that specimen across the two duplicate specimens and subtracting this value from 1 (1 – (count1 – count2)/(count1 + count2)) [55]. Reliability scores range from zero (unreliable/inconsistent) to one (reliable/consistent). Participants who reported no organs on one sheet and a non-zero number of the same organ on the duplicate sheet were assigned a reliability score of zero for that organ (i.e. the lowest reliability score). We conservatively selected the lowest reliability score among the three calculated for each organ per participant and assigned it to each participant as their final score. That is, if a participant got a reliability score equal to zero on one organ, they would be assigned a reliability score of zero for all organs. Such data were also spot checked for suspicious outliers manually and removed such data.

Table 1. The number of specimens and different categorical traits of examined species. FL and FR refer to the number of specimens classified as flowering and fruiting, respectively.

<table>
<thead>
<tr>
<th>family</th>
<th>species</th>
<th>time span (years)</th>
<th>FL</th>
<th>FR</th>
<th>lifespan</th>
<th>growth form</th>
<th>status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranunculaceae</td>
<td>Anemone canadensis L.</td>
<td>116</td>
<td>41</td>
<td>59</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Ranunculaceae</td>
<td>Anemone hepatica L.</td>
<td>95</td>
<td>40</td>
<td>60</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Ranunculaceae</td>
<td>Aquilegia canadensis L.</td>
<td>119</td>
<td>251</td>
<td>238</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>Bidens vulgata Greene</td>
<td>119</td>
<td>83</td>
<td>136</td>
<td>annual</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Celastraceae</td>
<td>Celastrus orbiculatus Thumb.</td>
<td>103</td>
<td>151</td>
<td>220</td>
<td>perennial</td>
<td>herbaceous</td>
<td>introduced</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>Centaurea stoebe Tausch</td>
<td>111</td>
<td>93</td>
<td>161</td>
<td>perennial</td>
<td>herbaceous</td>
<td>introduced</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>Cirsium arvense (L.) Scop.</td>
<td>118</td>
<td>186</td>
<td>171</td>
<td>perennial</td>
<td>herbaceous</td>
<td>introduced</td>
</tr>
<tr>
<td>Asteraceae</td>
<td>Cirsium discolor (Muhl. ex Willd.)</td>
<td>117</td>
<td>46</td>
<td>93</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Geraniaceae</td>
<td>Geranium maculatum L.</td>
<td>119</td>
<td>489</td>
<td>513</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Geraniaceae</td>
<td>Geranium robertianum L.</td>
<td>119</td>
<td>48</td>
<td>307</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Xanthorrhoeaceae</td>
<td>Hemerocallis fulva (L.) L.</td>
<td>115</td>
<td>144</td>
<td>45</td>
<td>perennial</td>
<td>herbaceous</td>
<td>introduced</td>
</tr>
<tr>
<td>Malvaceae</td>
<td>Hibiscus moscheutos L.</td>
<td>119</td>
<td>105</td>
<td>100</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Balsaminaceae</td>
<td>Impatiens capensis Meerb.</td>
<td>120</td>
<td>153</td>
<td>501</td>
<td>annual</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Iridaceae</td>
<td>Iris pseudacorus L.</td>
<td>117</td>
<td>90</td>
<td>66</td>
<td>perennial</td>
<td>herbaceous</td>
<td>introduced</td>
</tr>
<tr>
<td>Iridaceae</td>
<td>Iris versicolor L.</td>
<td>119</td>
<td>344</td>
<td>185</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Liliaceae</td>
<td>Lilium canadense L.</td>
<td>117</td>
<td>139</td>
<td>27</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Caprifoliaceae</td>
<td>Lonicera × bella Zab.</td>
<td>107</td>
<td>37</td>
<td>62</td>
<td>perennial</td>
<td>woody</td>
<td>introduced</td>
</tr>
<tr>
<td>Caprifoliaceae</td>
<td>Lonicera canadensis Bartram ex Marshall</td>
<td>120</td>
<td>194</td>
<td>201</td>
<td>perennial</td>
<td>woody</td>
<td>native</td>
</tr>
<tr>
<td>Caprifoliaceae</td>
<td>Lonicera japonica Thunb.</td>
<td>116</td>
<td>329</td>
<td>115</td>
<td>perennial</td>
<td>woody</td>
<td>introduced</td>
</tr>
<tr>
<td>Rosaceae</td>
<td>Malus pumila Mill.</td>
<td>118</td>
<td>74</td>
<td>40</td>
<td>perennial</td>
<td>woody</td>
<td>introduced</td>
</tr>
<tr>
<td>Malvaceae</td>
<td>Malva neglecta Wallr.</td>
<td>116</td>
<td>25</td>
<td>140</td>
<td>perennial</td>
<td>herbaceous</td>
<td>introduced</td>
</tr>
<tr>
<td>Onagraceae</td>
<td>Oenothera perennis L.</td>
<td>120</td>
<td>194</td>
<td>214</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Orobanchaceae</td>
<td>Orobanche uniflora L.</td>
<td>118</td>
<td>213</td>
<td>105</td>
<td>annual</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Rosaceae</td>
<td>Rosa gallica L.</td>
<td>108</td>
<td>45</td>
<td>17</td>
<td>perennial</td>
<td>woody</td>
<td>introduced</td>
</tr>
<tr>
<td>Rosaceae</td>
<td>Rubus odoratus L.</td>
<td>120</td>
<td>176</td>
<td>318</td>
<td>perennial</td>
<td>woody</td>
<td>native</td>
</tr>
<tr>
<td>Sarraceniaceae</td>
<td>Sarracenia purpurea L.</td>
<td>119</td>
<td>234</td>
<td>75</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Iridaceae</td>
<td>Sigynchinium mucronatum Michx.</td>
<td>117</td>
<td>86</td>
<td>157</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Solanaceae</td>
<td>Solanum rostratum Dunal</td>
<td>115</td>
<td>21</td>
<td>85</td>
<td>annual</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Melanthiaceae</td>
<td>Trillium grandiflorum (Michx.) Salisb.</td>
<td>119</td>
<td>129</td>
<td>40</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
<tr>
<td>Melanthiaceae</td>
<td>Trillium undulatum Willd.</td>
<td>120</td>
<td>402</td>
<td>156</td>
<td>perennial</td>
<td>herbaceous</td>
<td>native</td>
</tr>
</tbody>
</table>
(c) Historical climate data

We used estimates of historic (1895–2016) average monthly temperature and precipitation data at 4 km resolution from PRISM (product AN81 m; http://prism.oregonstate.edu/), which provide high-resolution time-series estimates of climatic elements for the contiguous United States. As accurate locality data are not available for the majority of historic specimen records [59], we used county as our geographical unit of analysis. The vast majority (79%) of specimens used in this study were collected before 1980, and while 72% of the specimens used in this study had associated coordinate data, at least 91% of those coordinates had been georeferenced post hoc (e.g. assigned county or township centroid coordinates), and thus may not represent precise sampling locales. For each county and year, we estimated the mean monthly temperature, precipitation and elevation, and assigned these values to each specimen [59]. Although counties can vary in size and climate, counties in states along the Atlantic coast of the United States are generally small in size and geographically homogeneous. We estimated within-county climatic heterogeneity as the standard deviation of estimated monthly climatic values across each county and year and included it in our initial analyses, but coefficients had Bayesian credibility intervals that were not credibly different from zero, so we dropped these terms from our final models.

(d) Statistical analyses

Phenological sensitivity to spring temperature—defined as the mean of March, April and May temperatures [46]—was defined as the slope of the linear relationship between the day of year (DOY) of a phenophase and the spring temperature of the corresponding location and year (shifts in days per degree Celsius change: days/°C) [44,46]. These months have been used to define spring across the east coast of the United States [46,60]. To calculate phenological sensitivity, we binned our specimen data into both broad climatic zones [61,62] and finer-scale National Ecological Observatory Network (NEON) domains. Our data comprised two climatic zones (cold/very cold; mixed-humid/hot-humid—hereafter referred to as cold and mixed-warm) and five NEON ecoclimatic domains (NE, northeast; MA, mid-Atlantic; AP, Appalachians & Cumberland Plateau; OZ, Ozarks Complex; SE, southeast; electronic supplementary material, table S2). We also estimated phenological sensitivity to elevation as the slope of the linear relationship between the DOY of a phenophase and metres above sea level (m a.s.l.).

We estimated the mean timing of flowering and fruiting phenophases, and environmental influences on them, using Bayesian hierarchical linear regression models [63]. In our models, species, region, county and observer were considered random effects, while spring temperature and county elevation were covariates. The hierarchical nature of the model, in which the phenological responses of individual species were assumed to be drawn from statistical distributions instead of fixed estimates [64], allowed us to better estimate their climatic sensitivities. These models also more accurately quantified uncertainty in our estimates and partitioned the variance in phenological timing and phenological sensitivity within and between species and regions.

We fitted two models for each phenological phase. ‘Model 1’ estimated species-specific phenological sensitivities and partitioned their variances. ‘Model 2’ provided a more powerful comparison between phenological sensitivities found in the warmer, lower latitudes of our study area and those in the cooler, higher-latitude regions (figure 1).

In Model 1, the dependent variable was the DOY for which a given phenological phase (flowering or fruiting) was recorded for the ith specimen. $\text{DOY}_{ij}$ was assumed to be normally distributed, with mean $\mu_{ij}$ and species-specific variance $\sigma_{ij}$:

$$\text{DOY}_{ij} \sim N(\mu_{ij}, \sigma_{ij}).$$  \hspace{1cm} (2.1)

The linear predictor $\mu_{ij}$ was estimated as a function of covariates, including mean spring (March–May) air temperatures ($\text{Spring}_T$) and the average elevation of the county in which the specimen was recorded ($\text{Elev}_c$). Additional intercept terms ($\alpha_1$–$\alpha_9$) were added for each species ($s$), region ($r$), species combination ($sr$), county ($c$) and observer ($o$). The full expression for estimating $\mu_{ij}$ was

$$\mu_{ij} = \alpha_1 + \alpha_2 s + \alpha_3 s + \alpha_4 s + \alpha_5 s + \beta_1 r + \beta_2 r + \beta_3 r \cdot \text{Spring}_T + \beta_4 r \cdot \text{Spring}_T + \beta_5 r \cdot \text{Elev}_c.$$

(2.2)

Species-specific slope and intercept terms ($\alpha_{1r}$, $\beta_1$, $\beta_2$, $\beta_3$) were drawn from normal distributions, with species assemblage means $\mu_{s1}$, $\mu_{s2}$ and $\mu_{s3}$ and hypervariances $\sigma_{a1}$, $\sigma_{a2}$ and $\sigma_{a3}$.

$$\alpha_{1r} \sim N(\mu_{s1}, \sigma_{a1}),$$

$$\beta_1 \sim N(\mu_{s2}, \sigma_{a2}),$$

$$\beta_2 \sim N(\mu_{s3}, \sigma_{a3}).$$

(2.3)

and

$$\beta_3 \sim N(\mu_{s4}, \sigma_{a4}).$$

(2.4)

Region and species × region slopes ($\beta_{2r}$, $\beta_{3r}$) and region, species × region, county and observer intercepts ($\alpha_{2r}$, $\alpha_{3r}$, $\alpha_{4r}$, $\alpha_{5r}$) in equation (2.2) were drawn from zero-centred normal distributions with hypervariances $\sigma_{o1}$, $\sigma_{o2}$, $\sigma_{o3}$, $\sigma_{o4}$ and $\sigma_{o5}$, respectively. The species-specific sampling variation ($\sigma_{o}$) terms in equation (2.1) were estimated independently to account for differences in the duration of flowering and fruiting phases between species.

The three different groups of slopes estimated for spring temperature decomposed variation in phenological sensitivity into components representing between-species variability ($\beta_{1s}$), between-region variability ($\beta_{2r}$) and within-species variability across regions ($\beta_{3sr}$). The accompanying hypervariances ($\sigma_{o1}$, $\sigma_{o2}$, $\sigma_{o3}$, $\sigma_{o4}$) directly represented these different sources of variability; comparing their relative magnitudes quantified the contributions of each source of variation to overall variation in phenological sensitivity. This model structure also provided estimates of the contributions of species turnover to differences in phenological sensitivity between regions. We estimated these contributions by analysing the output of Model 1, computing phenological sensitivities for each observation for each iteration of our model. We then used the mean and standard deviation of these estimates for each region to create region-specific estimates of mean phenological sensitivities and their variability. We assessed the contribution of community turnover by comparing estimates that included all three climate sensitivity terms ($\beta_{1s} + \beta_{2r} + \beta_{3sr}$) with estimates that included only the terms that represent species-level variability in climate sensitivity ($\beta_{1s}$). This strategy allowed us to infer what the mean phenological sensitivities would be across regions in the hypothetical case that they differed only in species composition, and species responded identically to climate across their ranges.

Model 2 differed from Model 1 in treating the region term ($\beta_{2r}$) as a two-level fixed effect representing the climatic region from which the specimen was drawn.

$$\mu_{ij} = \alpha_2 s + \alpha_3 s + \alpha_4 s + \alpha_5 s + \beta_1 r + \beta_2 r \cdot \text{Spring}_T + \beta_3 r \cdot \text{Spring}_T + \beta_4 r \cdot \text{Elev}_c.$$

(2.6)

Model 2 maximized statistical power to compare overall phenological sensitivities of species between warmer, more southerly parts of our study area and cooler, more northerly areas. Instead of treating region-specific slopes and intercepts as normally distributed random grouping factors, we represented species × region, county and observer terms as zero-centred and normally distributed. This structure allowed more direct inference about overall differences in phenological sensitivity between cool and
mixed-warm climate regions (differences between estimates of $\beta_{2r}$ for the two regions). Unlike Model 1, however, Model 2 lacks species-specific parameters ($\alpha_{1s}$ and $\beta_{1s}$) and did not provide species-specific estimates of phenological sensitivity or permit comparison of variability in phenological sensitivity within and between species.

We estimated all parameters of the two models using Hamiltonian Monte Carlo (HMC) [65] implemented in Stan (v. 2.17.3) [66] called from the rstan interface [67] in R [68]. HMC is a form of Markov chain Monte Carlo (MCMC) that efficiently estimates hierarchical Bayesian regression models for larger datasets like ours [69]. We used relatively uninformative prior distributions: zero-centred normal priors for slopes and intercepts, and truncated normal distributions for variances and hypervariances. To account for sampling behaviour and simplify prior choices, we scaled and centred the response variable DOY$_i$ and all continuous predictors by subtracting the mean and dividing by the standard deviation of each variable. For each model run for each phenophase, we estimated parameters using four MCMC chains of 4000 iterations each and discarded the first 2000 iterations of each chain (as burn-in). We assessed convergence of each model both visually and with the Gelman–Rubin statistic ($\hat{r} < 1.1$ for all parameters). We also assessed good model fit using visual posterior predictive checks implemented in the bayesplot R package [70]. All parameter estimates were based on at least 1000 effective posterior samples. Estimates reported in the results were back-transformed to the original data scale to facilitate illustration and interpretation.

Code and data for reproducing these analyses are archived by Harvard Forest [71].

### 3. Results

Our focal species spanned wide geographical and climatic space (figure 1). They demonstrated diverse patterns of phenology and significant variation in responses to climate across species and geographies. Using Model 1, estimated mean (non-leap-year) flower timing at 7.4°C and 216 m a.s.l. (mean collection conditions for the specimens) varied from 10 May (Day 130, Anemone hepatica) to 10 September (Day 253, Bidens vulgata) for flowering and 22 May (Day 142, A. hepatica) to 14 September (Day 257, B. vulgata) for fruiting (figure 2). The average lag time between flowering and fruiting across all species was approximately 20 days. Most species flowered and fruited earlier with warmer spring temperatures (assemblage mean $\beta_{2s}$ = 2.56 days/8°C, 95% CI $\beta_{2s}$ = 2.36 to 2.148, figure 2), and these responses were credibly different from zero (posterior probability > 0.95) for 21 out of 30 species for flowering and 15 out of 30 species for fruiting (electronic supplementary material, tables S3 and S4).

---

**Figure 1.** Distribution of herbarium specimens across geographical and environmental space, colour-coded by NEON domain (a,c) or climate region (b,d). Panels (a) and (c) show specimens in flower, and panels (b) and (d) show specimens in fruit. Specimens are referenced to county centroids, and jitter has been added to coordinates to reduce over-plotting. Average spring (March–May) air temperatures are strongly negatively correlated with latitude (Pearson correlation $r = -0.92$, panel c), but elevation and latitude are largely independent (panel d).
For both flowering and fruiting, species with earlier reproductive phenologies were substantially more sensitive to spring temperature than species that flowered and fruited later in the season. This sensitivity manifested in a strong positive correlation between mean flowering and fruiting date and spring temperature sensitivity, with a slope of 0.018 days/°C per day for flowering and 0.023 days/°C per day for fruiting (figure 2a,b). These slopes were different from zero with greater than 99% posterior probability. We also found that, all other conditions being equal, flowering and fruiting came earlier at higher elevations (community means −1.71 to −0.11 days earlier per 100 m greater elevation for flowering, and −2.07 to 0.14 days earlier for fruiting, figure 2c,d). These effects were credibly different from zero for 7 of 30 species for flowering and 8 of 30 species for fruiting, but elevation influenced early-flowering and late-flowering species approximately equally.

Species in the warm and mixed-temperate climatic regions showed greater mean sensitivities to spring temperature and also greater variability (standard deviation) in climate sensitivity between species than in the cool-temperate region (figure 3b). All differences between cool and mixed-warm climatic regions had a posterior probability greater than 0.95 except for mean differences in flowering, where differences had a posterior probability of 0.87. These qualitative patterns were robust to an alternative spatial binning strategy that used only latitude, and not climate, to differentiate more northerly and southerly regions (electronic supplementary material, figure S3).

Overall differences between cool, northern and warm, southerly parts of the study area were accompanied by large amounts of regional variation not explained by climate or latitude (figure 4). For example, using Model 1 we estimated that plants in the Ozark Complex and mid-Atlantic NEON domains were substantially more phenologically sensitive than those in the northeast and Appalachians (figure 4a), while the northeast and Ozark Complex had plant assemblages with greater variability in phenological sensitivity (figure 4c). Differences in phenological sensitivity between regions were not adequately explained by differences in species composition, as per-sample weighted means computed using only species effects ($\beta_{1s}$ in equation (2.2)), did not show strong regional differences (figure 4b,d).

Our hierarchical approach allowed us to compare within-species, between-species and between-region sources of variability for both mean flowering time and sensitivity to spring temperature (figure 5). This analysis shows that between-species variation dominated variability in mean flowering time (figure 5a), but there was a similar amount of variation in phenological sensitivity within species between regions to that seen between species for both flowering and fruiting (figure 5b).
with probabilities greater than or equal to 0.99 are depicted with '**'. Comparisons with a posterior probability greater than or equal to 0.8 and less than 0.95 are depicted with a '⋆', and comparisons with probabilities greater than or equal to 0.95 and less than 0.99 are depicted with '*' and within species across their latitudinal ranges and (iii) that phenological sensitivity to temperature tends to be higher in the warmer, more stable climates of lower-latitude regions.

4. Discussion

Our analyses revealed that (i) plant species from the eastern United States exhibit advanced timing of flowering and fruiting in response to warmer spring temperatures, (ii) the magnitude of these responses varies significantly between and within species across their latitudinal ranges and (iii) that phenological sensitivity to temperature tends to be higher in the warmer, more stable climates of lower-latitude regions.

(a) Differential responses to spring warming across species

Consistent with previous field observations of community phenology, we found that reproductive phenology of flowering plants accelerated with warming spring temperatures (e.g. [46,72,73]; but see [63]). The average number of days of phenological advancement per degree increase in temperature (−2.56 days/°C) that we observed also fell within previous estimates [46,74,75]. All else being equal, flowering and fruiting tended to occur earlier at higher elevations. Higher elevations tend to be relatively colder and have shorter growing seasons, which exert pressure for species to initiate growth as soon as conditions become favourable [44,76–78].

Despite these general trends, we observed significant variation among species in their responses to warming. In general, early-flowering and early-fruiting species were more sensitive to spring temperatures than late-flowering/fruiting species (figure 2), a pattern also observed at smaller scales [75,79]. Warming-induced leaf budburst advancement has been suggested to be less prominent in late-flushing species compared with early-flushing ones owing to their greater chilling requirements [80]. Similar mechanisms may affect flowering and fruiting, where advances in the flowering date of late-flowering species caused by spring warming would be smaller than those of early-flowering species, which would manifest as weaker phenological responses to temperature in late-flowering species. Further, as flowering and fruiting events later in the year are more separated from spring climatic conditions, there is an extended window of time in which other factors could affect or modify reproductive timing. For instance, late-flowering species may be more sensitive to photoperiod or precipitation.

A large amount of variability in phenological sensitivity across species suggests that phenological responses to climatic change will be heterogeneous within communities. This could cause temporal reorganization of the structure and composition of plant communities, potentially altering direct and indirect interactions among plant species and between plant and animal species, and ecosystem services [24,34,81–83].

(b) Phenological sensitivity to spring temperature tends to decrease with latitude

The consequences of phenological shifts can be further complicated by intraspecific variation in phenological sensitivity to environmental cues [33,38]. For instance, using 20 years of observational data, Prevèy et al. [44] found that the phenological sensitivity to temperature of tundra plants at colder, higher latitudes was greater than at warmer, lower latitudes. However, contrary to such studies, we found that plants in warmer, lower-latitude regions tended to be more phenologically sensitive to temperature, especially for fruiting (figure 3). We hypothesize that this is due to the lower and less predictable winter and spring climates of the northeastern United States. In such environments, dynamic phenological tracking of spring temperatures (i.e. high phenological sensitivity to temperature) presents high risks to reproductive success, because warm periods may often be followed by periods of dramatic chilling [40]. At lower latitudes, the advent and progression of spring is less variable and average temperatures are higher; thus phenological tracking of temperature is less risky (electronic supplementary material, figures S1 and S2). Indeed, species exhibited a larger amount of variability in their responses to temperature in the warmer, lower-latitude parts of their ranges.

Climate and phenology might play different roles in filtering species assemblages in regions with longer growing seasons than in regions where the growing season is short and reproductive phenologies are strongly constrained by shorter freeze-free periods [35]. Indeed, studies synthesizing plot-level observational data have suggested phenological sensitivity of plant communities to warming may be positively correlated with mean annual temperature, but
negatively correlated with seasonal temperature range (i.e. variability) in Europe [84] and China [85]. Alternatively, it is possible that plants in warmer climates exist closer to their response thresholds in terms of phenology, and thus react more dramatically to small changes in temperature. However, Körner & Basler [86] noted that cherry cultivars from regions with less variable spring temperatures flowered earlier in common gardens, suggesting phenological sensitivity does vary with climate. Plants in regions with high spring temperature variability also tend to be less phenologically sensitive in terms of leaf out and bud break to temperature than those in less variable climates [40]. Our results demonstrating that phenological sensitivity to temperature is higher in areas with low standard deviation of intra-annual temperature and inter-annual variation in spring temperature and high mean annual temperature support these findings.

(c) Consequences of variation in phenological responses across species ranges

Our results imply that with equal warming, individuals in lower-latitude populations will advance their reproductive phenology more dramatically than those at higher latitudes. This observed variation in phenological response may reflect adaptation to local climatic conditions, especially in annual species. We found a large amount of regional variation in phenological sensitivity that was not clearly linked to climate or latitude. These regional differences were not explained by species turnover, but rather suggest the presence of inter-population variation driven by local adaptation or phenological plasticity (figure 4). Further, within-species variation in phenological sensitivity between regions was of similar magnitude to differences in sensitivity between species (figure 5). Other studies examining leaf out and senescence in trees also have shown that individuals from geographically and climatically separated populations differ in their phenology even when grown in common gardens [40,87,88]. Because the eastern United States is experiencing geographically variable climatic change, the heterogeneity in phenological responses to warming that we observed within and among species may have important consequences for plant communities in the near future. Colder, climatically variable high-latitude regions are experiencing disproportionate warming and climatic homogenization (i.e. reduced standard variation of intra-annual temperature), while warmer, climatically less variable more southerly regions are experiencing increases in intra-annual temperature variability (electronic supplementary material, figure S1). These climatic changes could alter patterns of overlap in reproductive timing among species in a community and across their individual ranges. Changes in phenological overlap across ranges could have direct consequences for adaptive evolution and species resilience to current and impending climatic changes, as

Figure 4. Differences in sensitivity to spring temperatures between NEON domains are driven by regional variation due to local adaptation or plasticity, not community turnover between regions. This applies to both differences in species assemblage mean phenological sensitivities (a,b) and assemblage standard deviations (c,d). Panels (a) and (b) represent best estimates of regional variations in phenological sensitivity incorporating species identity, NEON Domain and NEON Domain × species identity as random effects ($\beta_{100}, \beta_{200}, \beta_{300}$) and panels (c) and (d) represent estimates incorporating only species-level effects ($\beta_{100}^s$). This means that in panels (c) and (d), phenological sensitivity is assumed to be constant within a species across domains. Estimates for flowering are represented in black and estimates for fruiting are represented in grey. Thick and thin bars represent 50 and 80% posterior credible intervals, respectively, from a hierarchical Bayesian model (Model 1).
gene flow may increase among previously (temporally) iso-
lated populations of some species and decrease among
others [41–43,89–91]. Decreased phenological overlap could
generically isolate fringe populations, potentially leading to
local extirpation and species range contractions. Moreover,
phenologically sensitive plants at lower latitudes may
especially be at risk due to increasingly variable temperatures
and increased probabilities of phenological mismatch with
mutualists [92]. However, we cannot fully rule out the
possibility that individuals of long-lived perennial species
may also be able to acclimate their phenological sensitivity
to changing climatic conditions over longer periods of time.

Some of the variation we observe in phenological sensitiv-
ty to temperature across and within species may be due
to differences in microclimate. However, the lack of accurate
location data for most historic specimens limited our ability
to infer fine-scale climate, necessitating coarser, county-level
analyses. Also, we cannot ignore the possibility that the pheno-
lological trends we observed are unique to the species that
we studied and/or reflect biases in herbarium collections
[93]. To minimize effects of spatial bias and uncertainty, we
studied common, well-collected species and accounted for
climatic heterogeneity present in each sampling locale in
our models. However, these taxa are not necessarily represen-
tative of species assemblages across regions, and our analyses
do not account explicitly for spatial and temporal sampling
biases; some regional differences could be due to differing
patterns of collection across space and time. Including
county-level random effects as we have done here minimizes
the impacts of these biases but does not eliminate them
altogether. It is possible that different patterns of pheno-
logical sensitivity may be observed across species ranges
depending on how climatic and/or geographical regions
are delimited, though testing an alternative threshold yielded
similar results, suggesting that the patterns we observe are
robust to spatial binning choices. Additionally, we addressed
crowdsourcing bias by including crowdsourcer random
effects and removing observations for crowdsourcers with
low reliability scores even though phenological data
collected by citizen scientists do not differ significantly in
quality from those collected by experts [46,55]. Lastly,
although spring temperature is a critical driver of flowering
phenology in temperate climates, we cannot fully exclude
the possibility that other variables correlated with latitude
or mean spring temperature may determine observed
variance in phenological sensitivity [46,73,75,94–97]. For
example, spring temperature tends to be highly correlated
with mean annual and mean monthly temperatures in east-
ern North America (electronic supplementary material,
figure S4). Photoperiod or snow melt may further influence
and alter species phenological responses [98–103]. Future
research into how these environmental cues interact to trigger
phenological events is necessary and will greatly improve our
understanding of plant phenology.

5. Conclusion

Building on previous phenological research by scoring mul-
tiple phenological traits across over 8000 herbarium
specimens spanning 120 years, we have demonstrated that
phenological sensitivity can vary greatly across species’
ranges. This variance may be attributed to adaptation or
acclimation to local climates. The large amount of within-
species variation in phenological sensitivity that we observed
underlines the complex and contingent nature of phenological
sensitivities. Phenological responses of individual species to
climate are not stable phenotypic traits, but instead emerge
from a multitude of potentially reciprocal interactions between
organisms and their environment. Populations in different
regions could have differences in frequencies of genes that
control how climate affects the timing of development or
alterations in microhabitat distributions between regions that
alter how populations experience local climate. The regions
themselves may have differences in unmeasured environ-
mental factors that interact with responses to temperature or
differences in species interactions that may alter phenological
signalling. The circumstances and extent to which these or
other factors explain regional variation in species responses
to climate is currently unknown. To untangle the roles of
ecological and evolutionary processes governing the hetero-
geneous phenological responses of plant species to warming,
researchers will have to take advantage of new techniques
and datasets. In addition to continued field observations and
laboratory analysis of mechanisms responsible for flowering
and fruiting, herbarium specimens can provide a comprehen-
sive, nuanced picture of phenological responses to ongoing
climatic change across many species. Our study further
demonstrates that we can now harness the treasure trove of

Figure 5. Variance components from Model 1, allowing comparisons of vari-
ability between NEON domains, between species and within species. Variance in
the intercept (a) represents variability in flowering (black) or fruiting times (grey)
at 7.4°C spring temperature and 216 m elevation above sea level, median con-
ditions for the specimens. Variance in phenological sensitivity to spring air
temperature (days°C) represents variability in the slope of the linear relationship
between spring temperature and flowering or fruiting times.
information in herbaria across the world to examine hundreds, if not thousands of species across myriad plant lineages, habitats and regions. Such efforts will be critical to enhance our ability to forecast future changes in plant assemblages across space and time in an era of accelerating climate change [104,105].

Ethics. The use and collection of data by citizen scientists were approved by an ethics review committee at the University of Waterloo (ORE no. 21647).

Data accessibility. Data and R code are available from the Harvard Forest Data Archive (http://harvardforest.fas.harvard.edu/dataarchive, dataset HF309).

Authors’ contributions. C.C.D. conceived the study; D.S.P., C.C.D., A.C.W. and E.L. designed the study; D.S.P. and A.C.W. collected data; D.S.P., A.M.E. and I.B. analysed the data; D.S.P. drafted the first version of the manuscript; C.C.D. and D.S.P. made first substantial revisions to this draft, and all authors contributed to subsequent revisions.

Competing interests. We declare we have no competing interests.

Funding. This study was funded as part of the New England Vascular Plant Project to C.C.D. (NSF-DBI: EF0128835), NSF-DEB 1754584 to C.C.D. and A.M.E., and an NSERC Discovery Grant to E.L. (RGPIN-2015-04545). A.M.E.’s participation in this project was supported by Harvard Forest. D.S.P.’s contribution was supported by the Harvard University Herbaria and NSF-DEB 1754584. I.B.’s contribution was supported by the National Science Foundation’s Postdoctoral Research Fellowship in Biology (NSF-DBI-1711936).

Acknowledgements. The authors thank X. Feng, S. Worthington, and members of the Davis lab for their invaluable insights and comments on the project and manuscript.

References


52. Zhang Y, Bienloy L, Georgopoulous PG. 2014 Climate change effect on Betula (birch) and Quercus (oak) pollen seasons in the United States. *Int. J. Biometeorol.* 58, 909 – 919. (doi:10.1007/s00484-017-1674-7)


71. Park DS, Breckheimer I, Williams AC, Law E, Ellison AM, Davis CC. 2018 Assessing plant phenological patterns in the eastern United States over the last 120 years. Harvard Forest Data Archive, HF309. (doi:10.6073/pasta/f6aaf72b5e56f7948586d5e2c3f559)


