Methods

Common Garden

In 2014-2015, we collected seeds from four field sites in northeastern North America spanning approximately a 3.5° latitudinal gradient. The four field sites included Harvard Forest (42.55°N, 72.20°W), the White Mountains (44.11°N, 71.40°W), Second College Grant, (44.79°N, 71.15°W), and St. Hippolyte, QC, CAN (45.98°N, 74.01°W). We transported all seeds back to the Weld Hill Research Building at the Arnold Arboretum in Boston Massachusetts (42.30°N, 71.13°W) where we germinated seeds following standard germination protocols, and grew them to seedling stages in the research greenhouse. In the spring of 2017 we out-planted seedlings to establish the garden.

Plots were regularly weeded and watered throughout the duration of the study and were pruned in the fall of 2020. Survival of *Acer spicatum*, *Acer pensylvanicum*, *Vaccinium myrtilloides*, *Quercus alba*, and *Quercus rubra* was limited and these species are therefore not included in the following analyses. Based on survivorship in the common garden, our subsequent analyses included an average of 14 individuals per species. Our statistical analyses account for the unbalanced design that is frequently occurs in common gardens and other provenance trials.

Phenological monitoring:

For the years of 2018-2019, we made phenological observations of all individuals in the common garden twice per week from February to December. In 2020 due to the COVID 19 pandemic, we monitored once per week from March to November. We describe phenological stages using a modified BBCH scale (1) a common metrics for quantify woody plant phenological progression. We observed all major vegetative stages (budburst BBCH 07, leafout BBCH 15, end of leaf expansion BBCH 19, leaf coloration/drop BBCH 97, reproductive phases flowering BBCH 60-65, fruiting BBCH 72-79 and fruit/cones fully ripe BBCH 89). We added additional phases for budset and labelled full budset as BBCH 102.

Data analysis

To better understand the role that variation in leafout and budset phenology play in determining calendar growing season length among species, populations, and years we fit a Bayesian hierarchical model with a normal (Gaussian) error distribution. We calculated growing season duration by subtracting the day of leafout from the day of budset. We fit an intercept-only model with phenophase timing (leafout, budset) as the response variable with partial pooling across species, populations, and years. We only included observations with both budset and leafout observed on the same plant for this analysis ($n= 595$).

To assess the relationship between variation in leafout timing and calendar and thermal growing seasons we fit two additional regression models with thermal or calendar growing season length as the response variable and day of leafout as the main prediction. To account for species-level differences we included partial pooling on the slope and intercept of species.

We define the thermal growing season as the cumulative growing degree day heat sums between the day of leafout and the day of budset for each species. We calculated daily heat sums using the R package "pollen" (2) using a 10 °C base temperature with minimum and maximum daily temperature data from the Arboretum weather data record: https://arboretum.harvard.edu/research/data-resources/weather/. These weather data only ran through August of 2020, so we combined these data with temperate series from the nearest Network for Environment and Weather Applications (NEWA) weather station: "Boston (Weld Hill)" at https://newa.cornell.edu/all-weather-data-query/.

All models were fit in the R package "brms" (3) on 4 chains with a 4000 iteration warm-up and 1000 sampling iterations on each chain for a total of 4,000 sampling iterations across all chains. We evaluated model fit, with no divergent transitions, rhats less than 1.01 and high effective sample sizes. We performed all analyses in R version 4.1.2 (4).

Acknowledgements:

References

[1] Finn, G., Straszewski, A., and Peterson, V. A general growth stage key for describing trees and woody plants. *Annals of Applied Biology* **151**(1), 127–131, Aug (2007).

[2] Jakub Nowosad. *pollen: Analysis of Aerobiological Data*, (2019).

[3] Bürkner, P.-C. Advanced bayesian multilevel modeling with the r package brms. *R Journal* **10**, 395–411, 07 (2018).

[4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, (2021).

[5] Thornton, M., Shrestha, R., Wei, Y., Thornton, P., and Kao, S.-C. Daymet: Daily surface weather data on a 1-km grid for north america, version 4, (2020).

1 Tables

Species	functional group	n
*Acer pensylvanicum	tree	NA
*Acer spicatum	tree	NA
Alnus incana	shrub	31
Amelanchier canadensis	shrub	6
Aronia melanocarpa	shrub	12
Betula alleghaniensis	tree	24
Betula papyrifera	tree	13
Betula populifolia	tree	24
Diervilla lonicera	shrub	16
Myrica gale	shrub	15
*Quercus alba	tree	NA
*Quercus rubra	tree	NA
Sambucus racemosa	shrub	11
Sorbus americana	shrub	5
Spiraea alba	shrub	19
Spiraea tomentosa	shrub	21
*Vaccinium myrtilloides	shrub	NA
Viburnum cassinoides	shrub	25

Table S1: Common garden focal species. In 2017, plantings were randomized between 16 plot blocks. Individuals that were too small to survive outside were maintained in the growth facilities for an additional year and out-planted in the early spring of 2018. Plots were divided between tree plots which included species *Acer pensylvanicum*, *Amelanchier canadensis*, *Alnus incana*, *Betula papyrifera*, *Betula populifolia*, *Beluta alleghaniensis*, *Quercus alba*, and *Quercus rubra* and shrub plots which included the remaining species and shade cloth. * indicates species that were planted but not included in analyses due to low survivorship.
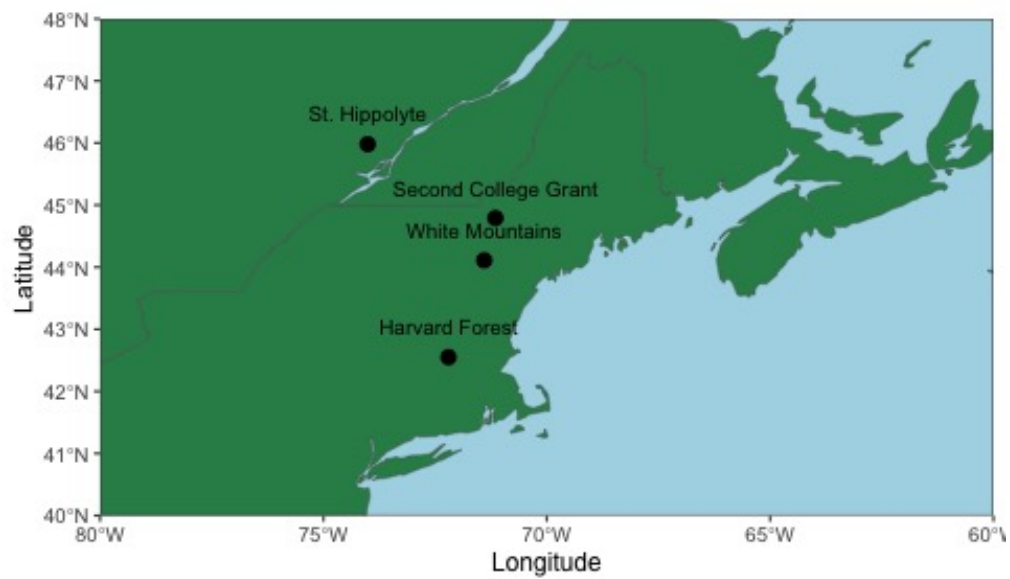
2 Figures

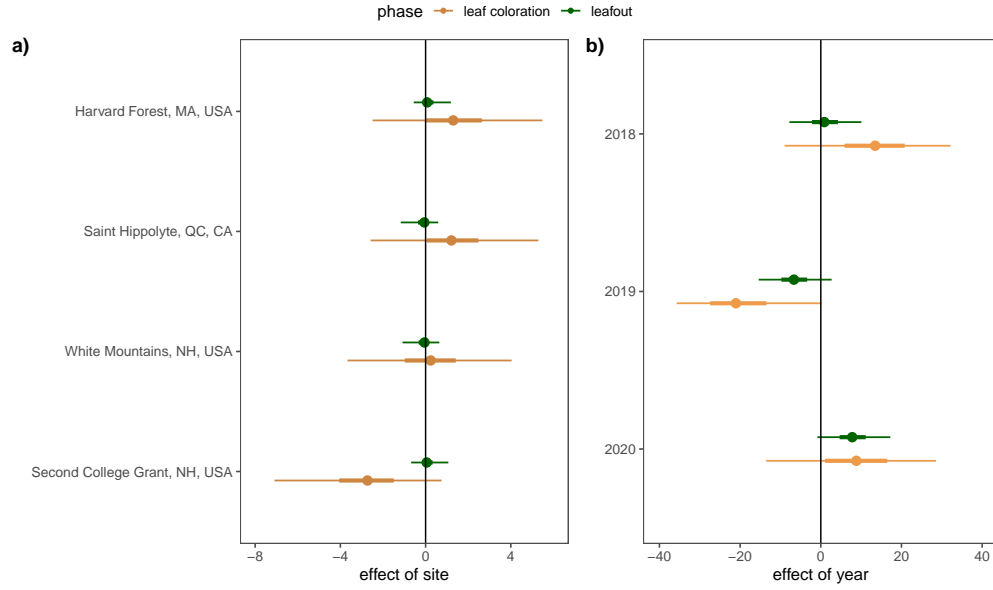Figure S1: A map of the four site sampled in this study

Figure S2: Difference in leafout, leaf coloration and growing season length partitioned between populations (a) and years (b). Points represent the median effect size estimate, and thick and thin bars the 50% and 90% uncertainty intervals.
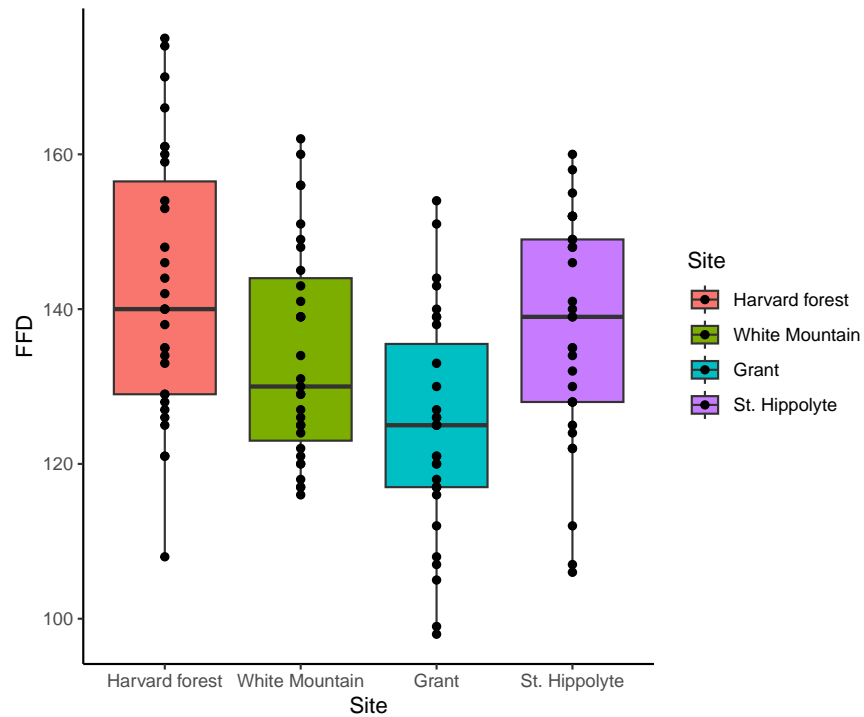

Figure S3: Differences in frost free days (FFD) between sites. Each point represents the number of days between the 90th percentile of last spring frost days and the 90th percentile of first fall frost days for a given year between 1980 and 2010. To calculate FFD, we used the long-term gridded esimates of minimum temperature from Daymet (version 4, 5).
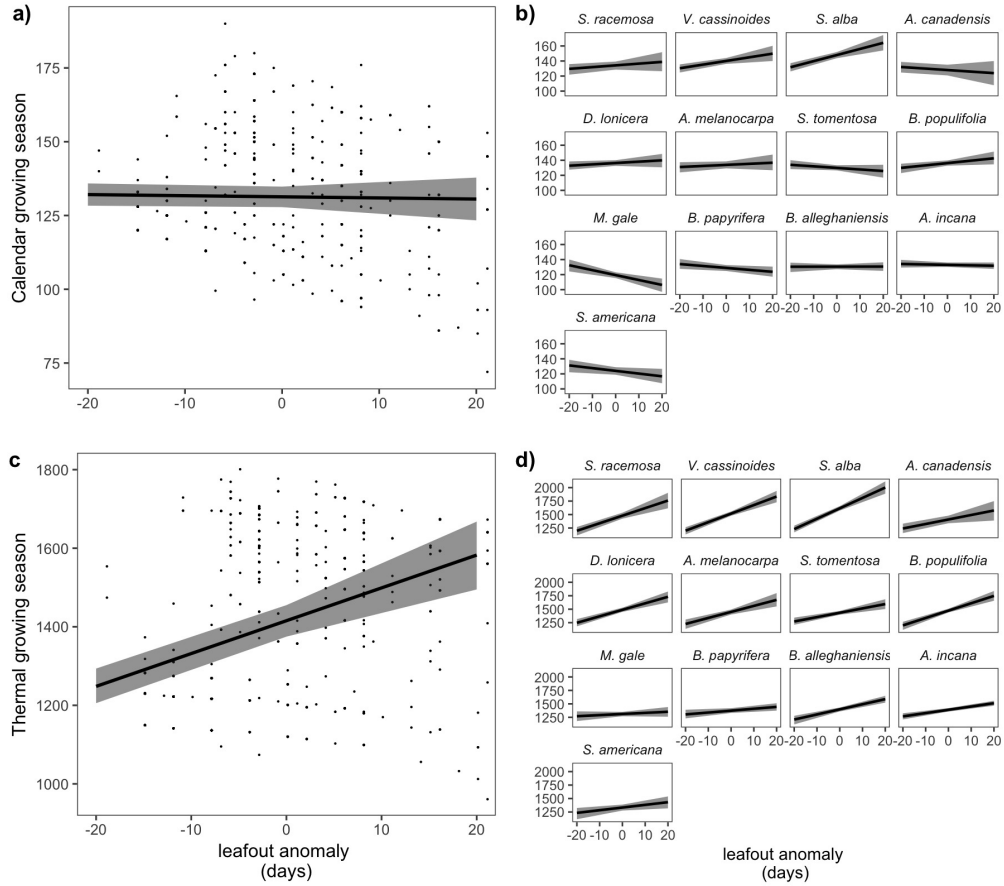
Figure S4: The relationship between the day of leafout and growing season length (defined by leaf coloration) differs between the calendar growing season and the thermal growing season. Later leafout did not affect calendar growing season (a) but this pattern varied across species in our study (b). Increasingly later leafout resulted in a longer thermal growing season (c) though this effect was stronger for species that typically leafout earlier in the season—panels in c) display in the typical order of leafout among species. The trend lines represent the median effect of leafout timing on growing season length and shaded bars the 90% uncertainty interval. Points in a), and c) represent the raw data.