New Phytologist Supporting Information

Article title: Reconciling competing hypotheses regarding flower-leaf sequences in temperate forests for fundamental and global change biology

Authors: D.M. Buonaiuto, I. Morales-Castilla, E.M. Wolkovich

The following Supporting Information is available for this article:
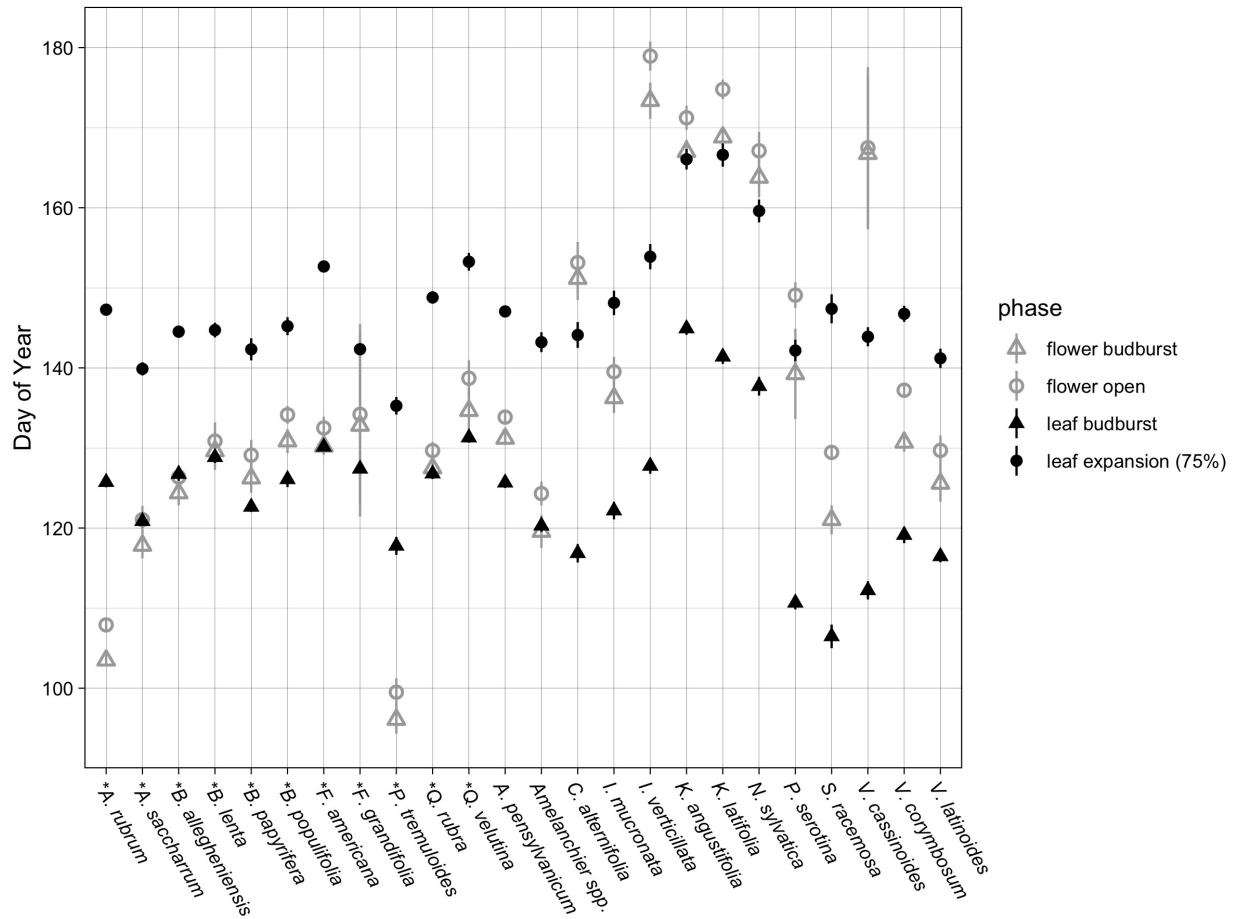
Figures

Figure S1: **Quantitative FLS patterns for woody plants at Harvard Forest in Petersham, MA.** Because phenological sequences consist of several sub-stages it is difficult to unambiguously categorize many species into the current FLS categories. The * accompanying the species name indicates wind-pollinated species.
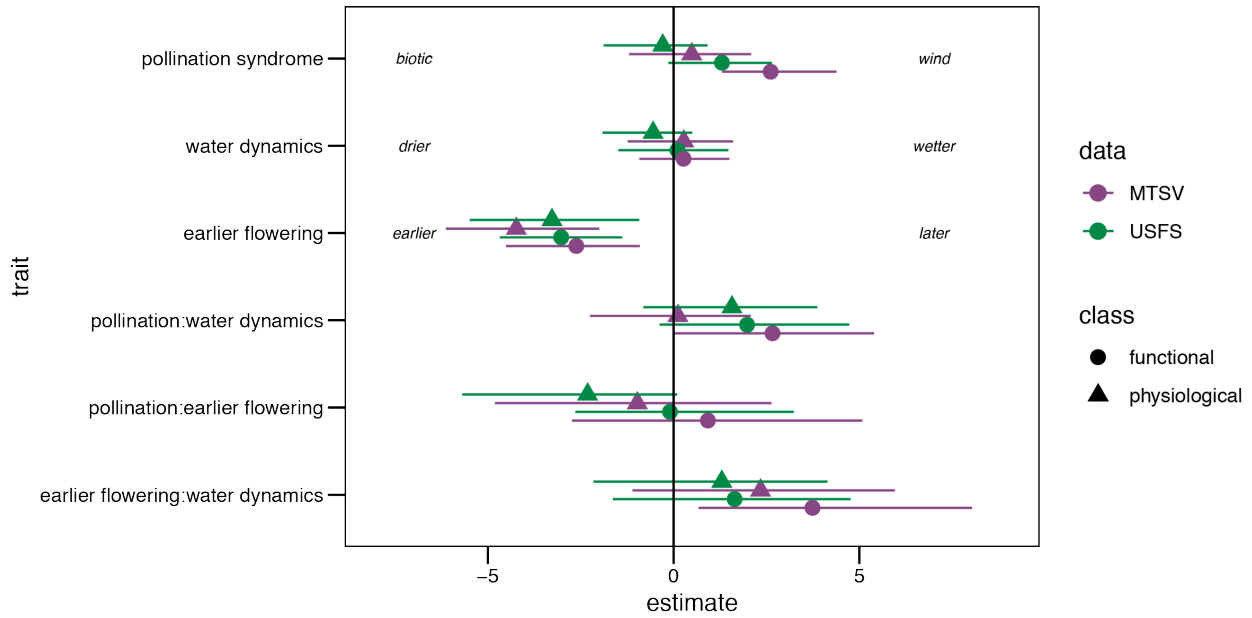
Figure S2: **Mean estimates of the effects of FLS predictors on the likelihood a species is hysteranthous vary across datasets and definitions of FLSs.** We used phylogenetic adjustments and standardized units to make a basic comparison of two datasets (Michigan Trees, Michigan Shrubs and Vines (MTSV) (Barnes *et al.*, 2016; Barnes & Wagner, 1981,2004) and The United States Forest Service's Silvics Manual (USFS) (Burns & Honkala, 1990)) and classes (physiological= no overlap between flowering and leafing, functional= moderate overlap) of FLSs. We chose functional trait predictors related to each of FLS hypothesis. While there is some agreement across models (strong effects of flowering time, no consistent effect interactions between predictors), the effect of other predictors (pollination syndrome, minimum precipitation across species' range) were highly sensitive to how data were defined, potentially biasing any inference from models and compromising the ability to validate the existing FLS hypotheses. Lines represent 95% bootstrap intervals.
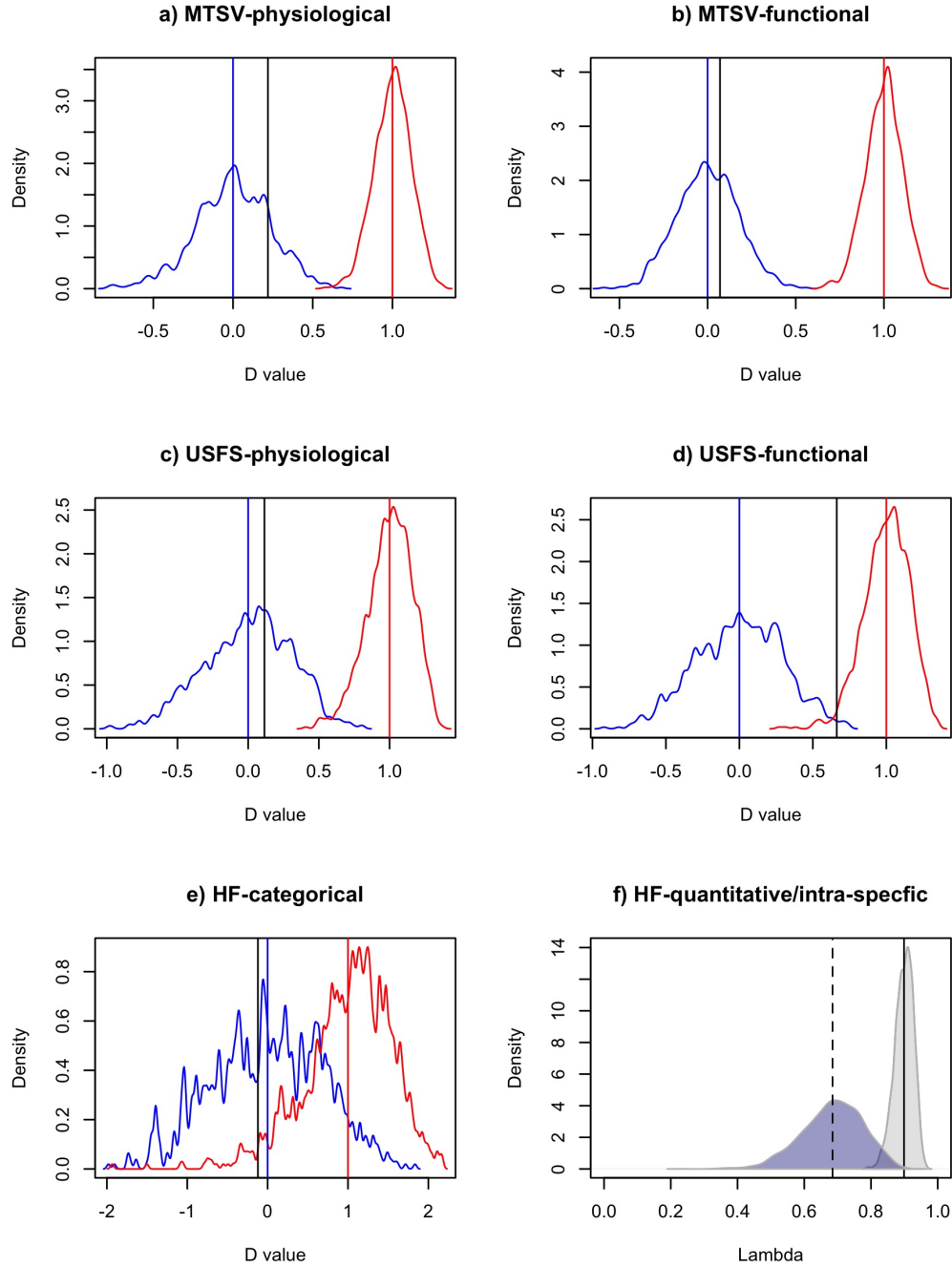
Figure S3: **The phylogenetic signal for FLSs varies between datasets, and is sensitive to how FLS patterns are categorized.** In a)-e),the black vertical line shows the the Fritz's D statistic for binary classifications of FLS estimated from the data, with blue and red lines representing expected D values based on simulations under Brownian threshold model and random model respectively. Panel f) shows the the estimated λ values of FLS from the the continuous modeling framework. The solid line indicates the mean estimate of λ in the intercept only model and the dashed line indicates the mean estimate of λ when all predictors were included in the model. Higher values indicate stronger phylogenetic structure.
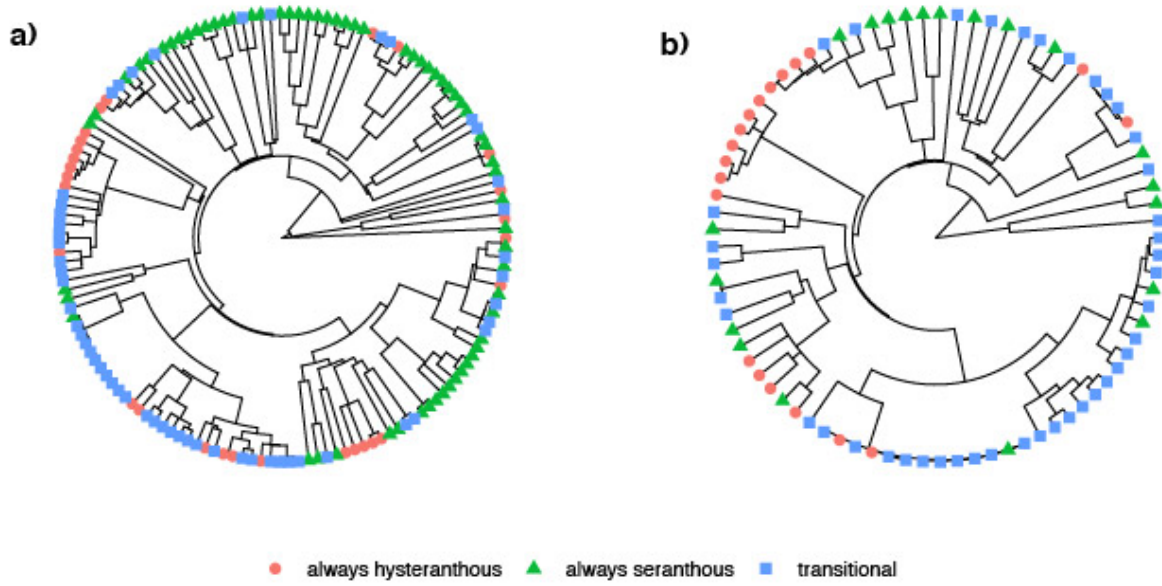
4

Figure S4: **Phylogenetic structure of FLSs in MTSV a) and USFS b) varies significantly depending on how FLSs are defined.** Many species are re-assigned to either hysteranthy or seranthy depending on whether FLSs are defined functionally (partial overlap between flowering and leafing allowed) or physiologically (no overlap between flowering and leafing allowed) (blue squares). This modeling choice dramatically alters FLS patterning across the tree, resulting in an unstable phylogenetic signal for this trait.
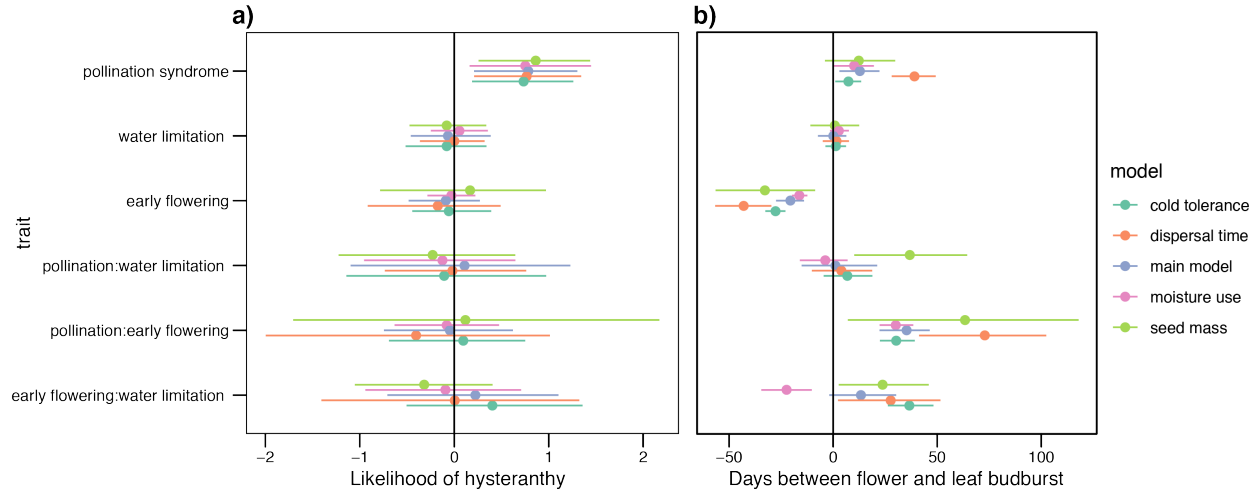
Figure S5: **Model estimates based on categorical (a) and quantitative (b) FLS data of associations between FLSs and functional traits are stable when alternate traits are used to represent the FLS hypotheses.** In addition to our main model which used minimum precipitation across a species' range to represent the water limitation hypothesis and flowering time to represent the early flowering hypothesis, we ran alternative models with cold tolerance and moisture use representing the water limitation hypothesis, and dispersal time and seed mass representing the early flowering hypothesis. Especially for the quantitative model **(b)**, the exact estimates and uncertainty varied depending on which model we used, but the relative strength among predictors representing each hypothesis remained consistent suggesting the drivers of FLS variation are consistent for the suite of traits that may comprise them. Lines represent 95% bootstrap intervals.
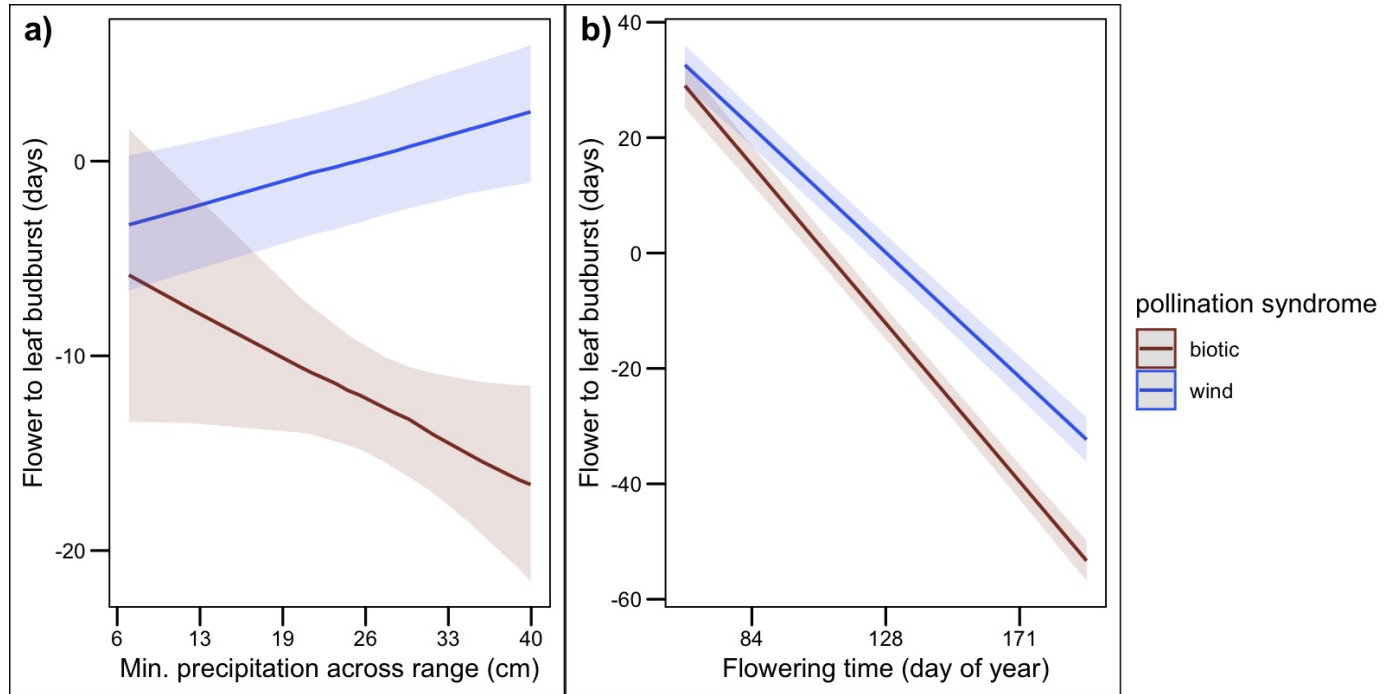
Figure S6: **Results from a quantitative-hierarchical model accounting for individual variation in FLSs suggests that water limitation may be a driver of hysteranthy in biotically-pollinated but not in wind-pollinated taxa (a) and that while wind pollinated species tend to always have a longer FLS interphase, the FLSs of bioticly pollinated species are are more sensitive to absolute flowering time (b).** Here we show model-predicted differences in FLSs as a function the minimum precipitation across a species' range and flowering time for a two generic species with contrasting pollination syndromes. Model projections are conditioned on long term phenological data from Harvard Forest in Petersham, MA (O'Keefe, 2015). The interaction depicted between hydraulic demand and pollination syndrome **(a)** assumes the mean flowering date (May 8) in the community for both functional types and the interaction between flowering time and pollination syndrome depicted in **(b)** is based on mean minimum precipitation values for the species included in the analysis (26 cm). Lines and shaded regions indicate mean estimates and 50% credible intervals respectively. These systematic differences in drivers of FLSs could reflect differences in function of FLSs for wind and biotically-pollinated taxa of temperate forest communities.
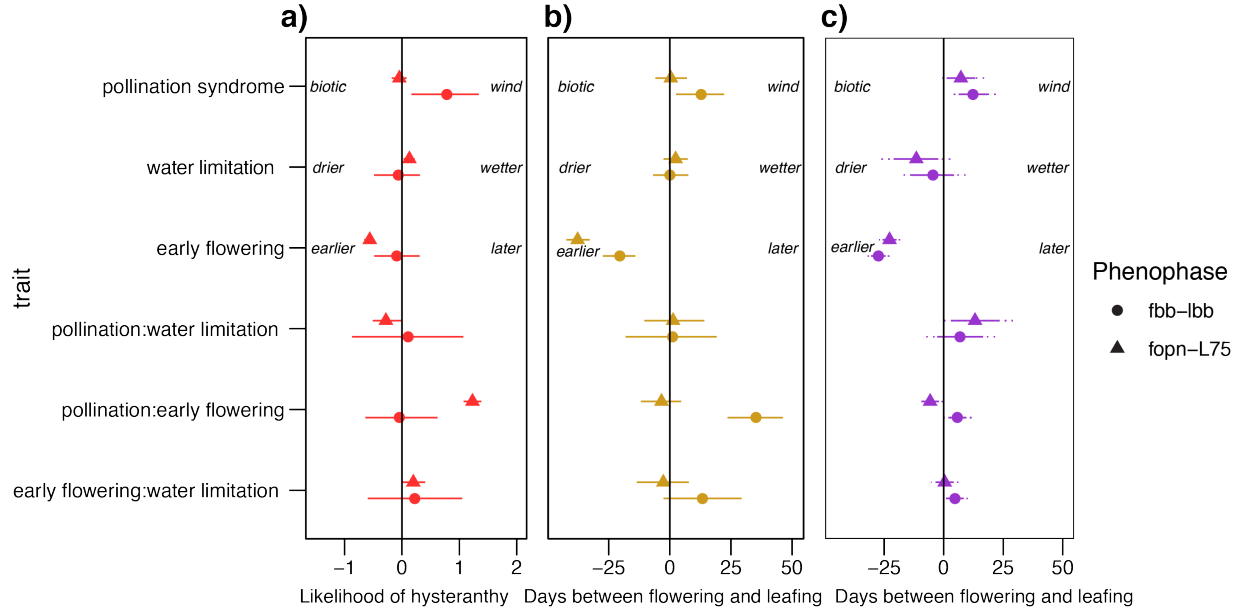
7

Figure S7: **Comparisons of sensitivity of model estimates to the use of different flower and leaf sub-phases for a categorical (a), quantitative (mean trait values), (b) and quantitative hierarchical (c) modeling framework reveal that hierarchical models that explicitly incorporate intra-specific variation may reduce the bias from comparing different sub-phases.** While it is not surprising that trait associations vary in strength when FLSs are defined based on different sub-phases of flowering and leafing, the fact that hierarchical models conditioned on intra-specific variation appear to reduce this bias may allow researchers to compare existing FLS data that are not perfectly standardized.
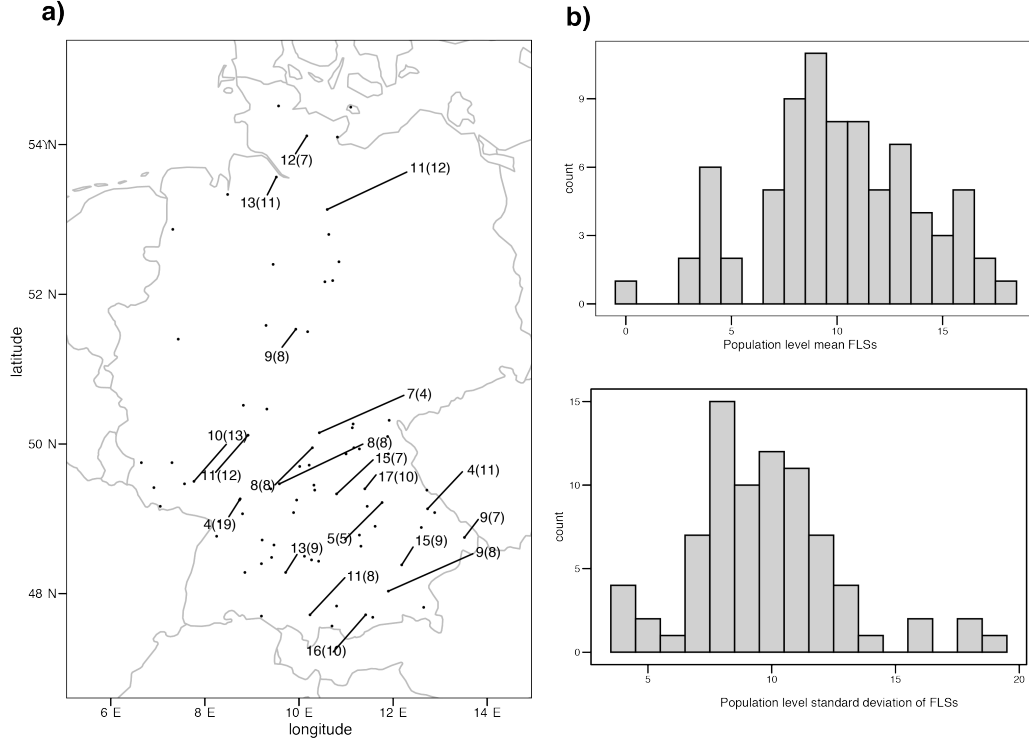
Figure S8: **FLSs vary significantly among populations of of *Fraxinus excelsior* across Germany**. In **(a)** each dot represent a population of *F. excelsior* with more than 50 years of phenological observations. We depict the mean and standard deviation (in parentheses) of the days between flowering and leaf expansion for 20 randomly selection populations. In **(b)**, we show this population level variation for all populations using histograms. Data comes from the PEP725 database (Templ *et al.*, 2018). Observations of FLSs at the population level are critical to better understand how environmental conditions shape FLSs and how FLS variation interacts with larger landscape-scale processes.

Tables

HF phenophase	verbal description	approximated BBCH
fbb	flower buds first broke with petals visible	55
fopn	50% of the flower buds were open	65
bb	50% of the buds were open with visible leaves	11
l75	50% of the leaves were developed to 75% of their final size	15

Table S1: Approximation of phenophases observed in (O'Keefe, 2015) the BBCH scale (Finn *et al.*, 2007)

Methods S1

Climate Change and FLS:

To evaluate how FLS patterns have changed over time in association with climate change we obtained phenological data for four European woody plant species with long term phenology records of both flower (BBCH 60) and leafout phenology (BBCH 11) from the Pan European Phenological Database (Templ *et al.*, 2018). We restricted the data set to include only stations with more than 50 years of data. Following conventions for modeling effects of climate change, we modeled the number of days between flowering and leafing as a function of time for each species, using a hinge model with 1980 as a break point (Kharouba *et al.*, 2018; Stocker *et al.*, 2013). For each species, we display the pre-1980 mean and 95% credible intervals of the time between flowering and leafing and the post-1980 change in mean time between phenophases that can be attributed to climate change and box plots to summarize the uncertainty in the estimates (Fig. 1).

Modeling FLS variation in MTSV and USFS data

For these two categorical, species-level case-studies, we converted verbal descriptions of flower-leaf sequences into a binary response variable. For our more inclusive "functional" definition of hysteranthy, which allows for some overlap between floral and vegetative phenophases, we included species entries with descriptions *"flowers before the leaves"*, *"flowers before or with leaves"* and *"flowers with leaves"* as hysteranthous. Our more restrictive "physiological" hysteranthy definition only included species described as *"flowers before the leaves"* as hysteranthous.

For modeling trait associations with FLSs, we chose three predictors to represent the three adaptive FLS hypotheses; pollination syndrome, average flowering time and minimum precipitation across a species' range. We obtained pollination syndrome and average flowering time information directly from the respective data sources and estimates of minimum precipitation across range from the USDA/NRCS Conservation Plants Characteristics (CPC) database (USDA/NRCS, 2020). We

coded pollination syndrome as biotic- or wind-pollinated, and assigned known ambophilous species in the genus *Salix* as biotic-pollinated. We re-coded flowering time as the average of the range of months of flowering reported in each data source.

For these case studies, we modeled associations between hysteranthy and the trait predictors with all two-way interactions with logistical regressions in phylogenetic generalized linear modeling framework (Ives & Garland, 2010) using the R package "phylolm" (Ho & Ane, 2014). Model results are presented in Fig. S2. Our models incorporated a published angiosperm phylogenetic tree (Zanne *et al.*, 2013) pruned to match the species list for each case study. Species found in the trait data set but not in the original phylogenetic tree were added to the pruned tree at the generic root. In total, 32 species were added to the generic roots for the MTSV data set and eight for the USFS data set. We visualized phylogenetic patterning of FLS across the tree of each case study (Fig. S4). The MTSV analysis was based on trait and FLS data for 147 species and the USFS analysis on 81 species.

We ran the models with 599 bootstrapped re-sampling iterations for each data set (Wilcox, 2010). We standardized all predictors by subtracting the mean and dividing by two standard deviations to allow for a reasonable comparison of effect sizes between the binary and continuous predictors in this model (Gellman & Hill, 2007).

Harvard Forest models

From the publicly available Harvard Forest phenology data (O'Keefe, 2015) we calculated the days between flower budburst and leaf budburst for each individual tree per year in the data. We also calculated FLS timing using alternative phenophases; the days between flowers opening and leaves reaching 75% of their final size. Positive FLS values indicate flowering-first and negative values leafing-first. We transformed these data to compare the inference between categorical, continuous, and intra-specific variation approaches to FLSs. To make the data categorical, we re-coded the continuous FLS measures as binary responses with positive values coded as hysteranthous and

negative values as seranthous. For the continuous data, we calculated the mean FLS timing and day of flowering for each species. For the intra-specific variation model, individual variation in FLSs and flowering time was incorporated into the model as a random effect nested within species. These models used the same predictors as the MTSV and USFS datasets (flowering time, pollination syndrome, minimum precipitation across species' range and all two way interactions between predictors). To test the sensitivity of the models to specific functional traits used to represent the FLS hypotheses, we ran additional models based on alternative functional traits also obtained for the CPC database (USDA/NRCS, 2020), moisture use and and minimum temperature across a species range for the water limitation hypothesis and seed mass (numbers seed/pound) and mean dispersal time (based on estimates from MTSV) for the early flowering hypotheses (see Fig. S5).

The Harvard Forest analysis included 23 species. While taxonomically limited compared to the MTSV and USFS data, this data set included repeated phenology observations per species over time, and within year variation between individuals per species.

For both the categorical and quantitative Harvard Forest models we used a phylogenetic generalized linear modeling framework (Ives & Garland, 2010) using the R package "phylolm" (Ho & Ane, 2014) built on a Bernoulli probability distribution for the categorical model and a Gaussian distribution for the quantitative model. For the intra-specific variation model, we used a Bayesian phylogenetic mixed modeling framework (PMM) (de Villemeruil P. Nakagawa, 2014) using the R package "brms" (Bürkner, 2018). PMM's incorporate the phylogenetic relationship among species as a random effect, utilizing a variance-covariance matrix based on species relationships to account for the non-independence in the model residuals that can be explained by phylogeny. We also included species as an additional random effect to account for non-independence in the residuals than is not due to phylogeny, and included individual as a nested factor within this random intercept to account for the repeat observations over time and differences among individuals. With this model, we ran 4 chains with 4000 iterations and a warm-up of 3000 iterations each, resulting

13

in 4000 total sampling iterations. Models used weakly informative priors on the intercept and error terms. Increasing priors three-fold did not impact the model estimates. As our primary goal was to directly compare the effects each predictor, we standardized these variables to allow for a reasonable comparison between them (Gellman & Hill, 2007). Model results can be found in Fig. 4, Fig. S5. and Fig. S7.

To graphically interpret the interactions in the hierarchical model, we calculated the marginal effects using the R-package "ggeffects" (Daniel Lüdecke, 2018). Fig. S6a shows the water limitation effect of FLS given are based on the mean flowering date (May 8) in the community and Fig. S6b shows the flowering time effect on FLS based on mean minimum precipitation values for the species included in the analysis (26 cm).

Though we make comparisons between the various HF and MTSV/USFS case studies, differences in data structure between the datasets required us to use alternative modeling frameworks. The categorical, mean quantitative HF data as well as the MTSV and USFS data provide one response variable for each species while the intra-specific variation HF data contains several different response values per species, among individuals and years. The current phylogenetic generalized linear model framework can only fit models with one response value per species, while the phylogenetic mixed model in brms may over-fit models with this kind of data structure and performs better on multi-response per species datasets like HF (Bürkner, 2017). At the outset, we ran both model types on each case study and while they do yield different absolute estimates, the patterns we found were consistent across each framework, and we report results from the most appropriate model for each dataset.

Analyses of phylogenetic signals

For all categorical specifications of FLS (MTSV, USFS and HF), we assessed the phylogenetic structure of hysteranthous flowering with Fritz's D-statistic (Fritz & Purvis, 2010) using the "Caper" package (Orme *et al.*, 2013) in R. Fritz's D calculates the sum of changes in estimated node values of a binary trait along edges in a phylogeny and compares this observed value to both a model of phylogenetic randomness and Brownian threshold model. The means of the two data simulations scale values of D to set points of 0 (as phylogenetically conserved) and 1 (random) (Orme *et al.*, 2013). We visualized the distribution of the traits across the tree for the MTSV and USFS datasets using the R package "ggtree" (Yu *et al.*, 2017), (see Fig. S4).

For the intra-specific variation Harvard Forest model, we estimated the phylogenetic signal for FLS (lambda) directly from the PMM model. To estimate lambda, we fit an intercept-only model with the phylogeny covariance matrix as a random effect and obtained the intra-class correlation value which is the phylogenetic signal. We also estimated the phylogenetic signal from the full model which included all predictors, and in both cases the intra-class correlation in the residuals were high. Estimated phylogenetic signals from all case studies are reported in (Fig. S3).

The water limitation hypothesis in wet temperate forests

As mentioned, it is surprising that the water-limitation hypothesis for FLS variation might be applicable in temperate regions where water is rarely limited during the main period of reproductive and vegetative phenological activity (Polgar & Primack, 2011). While recent work found a strong association between drought tolerance traits and FLS variation (Gougherty & Gougherty, 2018), this inconsistency suggests that this hypothesis needs further development to be considered in the temperate zone.

There are a number of modifications to the water limitation hypothesis that may explain its rel-

evance to temperate FLSs. This most simple is that hysteranthy in temperate species reflects a "memory" of an ancestral existence in drier environments and there is no contemporary function relating to plant water status. Hysteranthy may simply be maintained in temperate flora because there is no or weak selection against it. If this is the case, FLS variability should be independent of contemporary water availability. We modified our hierarchical FLS model to test the effect of variation in annual climatic water balance (precipitation - potential evapotranspiration) at Harvard Forest on interannual FLS variation. We calculate water balance from flux tower measurements at Harvard Forest (Hadley & Munger, 2004-), converting latent heat flux to potential evapotranspiration using the R package "bigleaf" (Knauer *et al.*, 2018). We found no significant effect of climatic water balance on FLS variation, further suggesting that FLSs variation seems to be independent of water availability in the temperate zone.

A second possibility is that the drought tolerance conferred by hysteranthy in the dry tropics has been re-purposed to serve a different function in the temperate zone. Cold tolerance, a critical adaptation for life at higher latitudes, may be conferred by many of the same physiological adaptations as for drought tolerance in dry regions (Zanne *et al.*, 2013). For the species in our analyses, we found a strong correlation (0.85) between minimum precipitation and temperature across species' ranges. It is possible that hysteranthy is among the suite of traits that allowed dry-adapted taxa to move into colder temperate regions. Association between hysteranthy and cold adaptation have been found in some families (Gougherty & Gougherty, 2018). Though the mechanism by which hysteranthous flowering may contribute to cold tolerance has not been thoroughly investigated, this correlation remains a promising avenue for applying the water limitation hypothesis in the temperate zone.

References

Barnes BV, Dick CW , Gunn ME. 2016. *Michgan Shrubs & Vines: A guide to species of the Great Lakes Region.* University of Michigan Press.

Barnes BV , Wagner WHJ. 1981,2004. *Michigan Trees: A guide to the Trees of the Great Lakes Region.* University of Michigan Press.

Bürkner PC. 2017. brms-users group. personal communication.

Bürkner PC. 2018. Advanced bayesian multilevel modeling with the r package brms. *R Journal*, **10**: 395–411.

Burns RM , Honkala BH. 1990. Silvics of North America: Volume 2. hardwoods. Tech. rep., United States Department of Agriculture (USDA), Forest Service.

Daniel Lüdecke. 2018. ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, **3**: 772.

Finn GA, Straszewski AE , Peterson V. 2007. A general growth stage key for describing trees and woody plants. *Annals of Applied Biology*, **151**: 127–131.

Fritz SA , Purvis A. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology* **24**: 1042–1051.

Gellman A , Hill J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Gougherty AV , Gougherty SW. 2018. Sequence of flower and leaf emergence in deciduous trees is linked to ecological traits, phylogenetics, and climate. *New Phytologist*, **220**: 121–131.

Hadley J , Munger JW. 2004. Ameriflux us-ha2 harvard forest hemlock site, dataset.

Ho LST , Ane C. 2014. A linear-time algorithm for gaussian and non-gaussian trait evolution models. *Systematic Biology*, **63**: 397–408.

Ives AR , Garland Jr. T. 2010. Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*, **59**: 9–26.

Kharouba HM, Ehrlén J, Gelman A, Bolmgren K, Allen JM, Travers SE , Wolkovich EM. 2018. Global shifts in the phenological synchrony of species interactions over recent decades. *Proceedings of the National Academy of Sciences*, **115**: 5211.

Knauer J, El-Madany TS, Zaehle S, Migliavacca M. 2018. Bigleaf - An R package for the calculation of physical and physiological ecosystem properties from eddy covariance data. *PLoS ONE*, **13**: e0201114.

O'Keefe J. 2015. Phenology of woody species at harvard forest since 1990. *Harvard Forest Data Archive: HF003*.

Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N , Pearse W. 2013. *caper: Comparative Analyses of Phylogenetics and Evolution in R*.

Polgar C , Primack R. 2011. Leaf-out phenology of temperate woody plants: From trees to ecosystems. *New Phytologist*, **191**: 926–41.

Stocker T, Qin D, Plattner GK, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V , Midgley P. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Cambridge, United Kingdom and New York, NY.

Templ B, Koch E, K.Bolmgren, Ungersböck M, Paul A, Scheifinger H , et al.. 2018. Pan European phenological database (pep725): a single point of access for European data. *Int. J. Biometeorology*.

USDA/NRCS. 2020. Conservation plants characteristics.

de Villemeruil P. Nakagawa S. 2014. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*, Springer, New York, chap. General quantitative genetic methods for comparative biology, pp. pp. 287–303.

Wilcox RR. 2010. *Fundamentals of modern statistical methods: Substantially improving power and accuracy.* Springer.

Yu G, Smith D, Zhu H, Guan Y , Lam T. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**: 28–36.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al*. 2013. Three keys to the radiation of angiosperms into freezing environments. *Nature*, **506**: 89–92.