

Unravelling the phenology-phylogeny tangle.

November 18, 2022

Authors:

The Wolkovich Lab in 2019 & collaborators ^{1,2,3,4}

Author affiliations:

¹Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4;

²Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA;

³Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA;

⁴Edificio Ciencias, Campus Universitario 28805 Alcalá de Henares, Madrid, Spain

*Corresponding author: ignacio.moralesc@uah.es

Abstract

Plants have evolved responses to environmental cues able to inform them about the temporal distribution of key resources—i.e. energy and light. The responses to individual cues such as forcing (or spring warming) have shown to be subjected to some degree of evolutionary conservatism. Yet, plants do not respond to isolated cues but to a combination of interacting cues, which difficults accurate predictions of phenology in the face of environmental change. Whether and how evolution has constrained phenological responses to combinations of interacting cues is not yet understood even when this knowledge could enhance model predictions and inform how different plant lineages have adapted to environmental change along their evolutionary histories. Here we use Bayesian hierarchical models and the most complete dataset on tree species phenological responses measured in experimental conditions to: (a) test if phenological responses to three major interacting cues are conserved phylogenetically when modelled jointly, (b) compare the phylogenetic signal in the responses to different cues and, (c) test whether coefficient estimates differ between models assuming phylogenetic independence among species and models that explicitly incorporate phylogeny. Results show non-random phylogenetic structuring of phenological responses, highly variable across species and cues. More interestingly, regression coefficients shift when models control for phylogenetic effects, particularly so for forcing, which emerges as the cue more strongly associated to phenology. Taken together, our results suggest that phylogeny should be incorporated into studies modelling multi-species phenological responses, as such responses have been jointly constrained through evolution and thus are not independent.

Introduction

1. Predicting species responses to recent anthropogenic climate change presents a major challenge to ecological forecasting
 - (a) On average global temperatures are warming, but there is regional and seasonal variation behind this average and shifts along other climate axes such as precipitation are more complex than simple increases
 - (b) Over the past few decades empirical studies have suggested that plants are shifting in their geographical distributions, moving to more northern latitudes and elevations, and shifting the timing of their life cycles
 - (c) However, responses are highly variable across species.
 - (d) Understanding how different species lineages have evolved their phenotypic responses to the combined effects of environmental change would greatly aid prediction

2. Environmental cues matter as they inform organisms about the temporal distribution of key resources.
 - (a) Responses (and their evolution) to cues are usually studied individually assuming that a given phenotypic response (e.g. time of leafout) is linked to a single cue, when likely multiple ones operate interactively (and have done so across evolutionary history) to shape that response.
 - (b) For example, light, nutrients and water often determine growth rates and size
 - (c) Or how the timing of many recurring life cycle events (phenology) is determined by a combination of temperature and light.
 - (d) modelling complex responses to multiple interactive cues is challenging, and a place where this becomes obvious is in predicting phenology.
3. Phenology makes for an ideal study case of species' responses to interacting environmental cues.
 - (a) Phenology is a critical trait to studying biological responses to climate change.
 - (b) In temperate species its cue system is generally known: forcing, chilling, photoperiod
 - (c) It is amongst the few phenotypic characters (if any other exists) for which there are multi-species experimental data on its responses to the three major environmental cues.
 - (d) Out of these three major cues that affect plants, few multi-species analyses have considered all three simultaneously, with repeated consensus that chilling and forcing would prevail, but would this pattern hold if evolution/phylogeny was accounted for?
4. Phenology x phylogeny quick review ... (not sure we need this anymore?)
 - (a) It is evolutionary conserved (to some extent, review antecedents).
 - (b) Research in this area has focused on the phenotype (phenological event or its shifts) rather than on the cues—i.e. how shifts in the environment trigger species responses. For example, first flowering is significantly conserved (Davies et al., 2013).
 - (c) Whether shifts in first flowering over time are also conserved is less clear and all inference to date is based on observational data, where geographical signals in phylogeny may drive trends attributed to phenology (e.g., Rafferty and Nabity, 2017), as discussed in (Davies et al., 2013).
 - (d) Further, additional questions remain open: in a multi-species context, have specific lineages adapted more strongly to some of the cues? or to any combination of cues? Is there any cue that is particularly labile?
 - (e) Answering these questions may: (i) inform about the need to account for phylogeny in phenological models and predictions, and (ii) expand our knowledge on how phenological responses have been constrained so far, which would be relevant in a context where species' sensitivities to warming temperatures seem to decline.

5. Method's review paragraph: Current methods advanced our understanding of how specific lineages have adapted phenotypic responses to the environment, but don't capture the complexity and non-stationarity of responses that depend on interacting environmental cues.

- (a) Traditional phylogenetic comparative methods (such as PGLS) allow us to fit models of species phenotypic traits while correcting for the non-independence among species.
- (b) But they do not allow the response to environmental cues to evolve over time while operating in concert with other cues.
- (c) However, exploratory methods show that responses can shift dramatically across a phylogenetic tree in a non-stationary fashion (Davies et al., 2019)

6. here's one possible solution

- (a) we model phenological responses to its known environmental drivers using a formulation that accounts explicitly for phylogenetic structuring in how these responses associate with each of the cues (forcing, chilling, photoperiod).
- (b) we model the driver of the variation, not only are interested in measuring that variation (as other methods do).
- (c) as a bonus, our modelling scheme provides cue-specific parameters (λ and σ) that inform us about how responses to each cue have evolved.
- (d) Common phylogenetic regression accounts for phylogenetic relationships as a grouping factor either explicitly (PMM) or implicitly (PGLS). Here we present one possible approach that accounts for more complex interactions going on among predictors, which would be reflected in the species-level slopes being allowed to vary as a function of the phylogeny, rather than keeping slopes constant and only allowing the intercepts (or residuals) to vary.
- (e) We ignore whether this is important and maybe current models are fine.
- (f) In a first attempt at establishing whether or not it is important, we compare results from a common hierarchical model with partial pooling on the slopes that does not allow for phylogenetic constraints to affect slope estimates against results from a phylogenetic hierarchical model allowing phylogeny to constrain partially pooled slopes.
- (g) We do so for an unprecedented dataset on phenological responses to environmental cues determined experimentally.

7. Questions rather than specific hypotheses

- (a) Based on previous research on phylogenetic signal of phenological responses, we expect non-random phylogenetic structuring of the responses to environmental cues (Davies et al., 2013; Rafferty and Nabity, 2017; Joly et al., 2019) and expect that temperature-related cues display higher phylogenetic signal than photoperiod because the latter has remained more constant through evolutionary time. Yet, rather than specific hypotheses for different lineage-level responses, our work aims at exploring and discussing the following questions:
- i. Do we need to account for phylogeny in multi-species, multi-cue modelling of the magnitude (strength) and variation of phenological responses to cues? This is, we worry about what are the biggest cues, and we think we may know which are those but if we have the wrong model, we may make the wrong inference or get estimates wrong.
 - ii. If so, can accounting for phylogeny shed light on the ongoing debate on declining sensitivities? For example, if particular lineages have very different evolutionary constraints on their responses to the cues, they may also display very different declines in their sensitivities to the cues.
 - iii. How can we interpret lambdas and sigmas for each cue?
 - iv. What are the implications for phenological predictions and forecasts?
 - v. Is this approach transferable to different taxa or biological responses?

Methods

Phenological and Phylogenetic Data

1. Description of the OSPREE database (where it comes from, number of species, studies, etc.).
2. We analyze 5 different subsets of species in the OSPREE database to explore differences across taxa (effect of gymnosperms?) and to test to what extent data resolution affects the results:
 - (a) Species grouped in generic complexes, to ensure enough cross-treatment data, as in Ettinger et al. (under review) (including 52 complexes)[flags.for.mainmodel=T]
 - (b) All species in the main model (including 117 species resulting from) [flags.for.mainmodel=T]
 - (c) All angiosperm species in the main model (including 110 species)[flags.for.mainmodel=T]
 - (d) All species in the latest version of OSPREE (including 231 species resulting from) [flags.for.allspmodel=T]
 - (e) All angiosperm species in the latest version of OSPREE (including 215 species)[flags.for.allspmodel=T]
3. Two phylogenetic hypotheses have been considered to build a tree containing the species in OSPREE. First the vascular plant megatree by Zanne et al. (2014);Nature and, second the megatree by Smith & Brown (2019);AJB.
4. The backbone phylogenies were pruned to contain only the studied species in each subset.
5. Species not in the backbone phylogeny were added as polytomies at the generic level (using the function *congeneric.merge*; (Pearse et al., 2015)).
6. To build a phylogeny for species complexes, the terminal branches of species belonging to the same complexes were collapsed.

Lizzie will
write this!

The Bayesian hierarchical phylogenetic model

For each of n species, we assumed that data were generated from the following sampling distribution:

$$y_j \sim \mathcal{N}(\mu_j, \sigma_e^2) \quad (1)$$

where

$$\mu_j = \alpha_j + \beta_{1,j}X_1 + \beta_{2,j}X_2 + \beta_{3,j}X_3 \quad (2)$$

Predictors X_1 , X_2 , X_3 are standardized forcing, chilling, and photoperiod, and their effects on the phenology of species j are determined by parameters $\beta_{1,j}$, $\beta_{2,j}$, $\beta_{3,j}$ representing traits. These traits, including the species-specific intercept α_j , are elements of the following normal random vectors:

$$\begin{aligned}\boldsymbol{\alpha} &= \{\alpha_1, \dots, \alpha_n\}^T \text{ such that } \boldsymbol{\alpha} \sim \mathcal{N}(\mu_{\alpha}, \boldsymbol{\Sigma}_{\alpha}) \\ \boldsymbol{\beta}_1 &= \{\beta_{1,1}, \dots, \beta_{1,n}\}^T \text{ such that } \boldsymbol{\beta}_1 \sim \mathcal{N}(\mu_{\beta_1}, \boldsymbol{\Sigma}_{\beta_1}) \\ \boldsymbol{\beta}_2 &= \{\beta_{2,1}, \dots, \beta_{2,n}\}^T \text{ such that } \boldsymbol{\beta}_2 \sim \mathcal{N}(\mu_{\beta_2}, \boldsymbol{\Sigma}_{\beta_2}) \\ \boldsymbol{\beta}_3 &= \{\beta_{3,1}, \dots, \beta_{3,n}\}^T \text{ such that } \boldsymbol{\beta}_3 \sim \mathcal{N}(\mu_{\beta_3}, \boldsymbol{\Sigma}_{\beta_3})\end{aligned}\tag{3}$$

where the means of the multivariate normal distributions are root trait values (i.e., trait values prior to evolving across a phylogenetic tree) and $\boldsymbol{\Sigma}_i$ are $n \times n$ phylogenetic variance-covariance matrices of the form:

$$\begin{bmatrix} \sigma_i^2 & \lambda_i \times \sigma_i \times \rho_{12} & \dots & \lambda_i \times \sigma_i \times \rho_{1n} \\ \lambda_i \times \sigma_i \times \rho_{21} & \sigma_i^2 & \dots & \lambda_i \times \sigma_i \times \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_i \times \sigma_i \times \rho_{n1} & \lambda_i \times \sigma_i \times \rho_{n2} & \dots & \sigma_i^2 \end{bmatrix}\tag{4}$$

where σ_i^2 is the rate of evolution across a tree for trait i (here assumed to be constant along all branches), λ_i scales branch lengths and therefore is a measure of the “phylogenetic signal” within a species trait, and ρ_{xy} is the phylogenetic correlation between species x and y , or the fraction of the tree shared by the two species.

The above specification is exactly equivalent to writing equation 2 in terms of root trait values and residuals, such that:

$$\mu_j = \mu_{\alpha} + \mu_{\beta_1} X_1 + \mu_{\beta_2} X_2 + \mu_{\beta_3} X_3 + e_{\alpha_j} + e_{\beta_{1,j}} + e_{\beta_{2,j}} + e_{\beta_{3,j}}\tag{5}$$

where the residual error terms (e.g., e_{α_j}) are elements of normal random vectors from multivariate normal distributions centered on 0 with the same phylogenetic variance-covariance matrices as in equation 4.

Interpretation of λ_i

In contrast to classic approaches to controlling for phylogenetic non-independence of analysis units (i.e. species), see (Freckleton et al., 2002), where Pagel’s Pagel (1999) λ is assumed constant across multiple predictors if those enter a PGLS model, our approach retrieves

The λ in our models is analogous to, but not fully equivalent to Pagel's Pagel (1999) λ parameter (Housworth et al., 2004), constrained to range from 0 to 1, with values of 0 indicating absence of phylogenetic relatedness, and values of 1 indicating *Brownian Motion* evolution (BM). This is because in our approach, a λ is estimated for each predictor in the model whilst in PGLS and similar approaches, λ is computed simultaneously across the predictor matrix.

Phylogeny in forecasts of phenology

This sections needs to be fleshed out, but first we need to think and decide how to proceed (or if we want to proceed at all):

1. How we define the two scenarios (regular scenario; climate change scenario; see below)?
2. For which subset of species do we test it?
3. Are we predicting with and without phylogeny? I'm still not sure about how to do this.

Results

take-home from the model

1. most species are sensitive to all cues
2. report sensitivities
3. something else?

this is why we think our model outperform others

1. how our model compares with lambda0 model (no phylo) in terms of variances, accuracy, etc
2. how our results compare with previous results in the literature (Ailene, PWR paper, Lizzie's and Simon's)...

phylogenetic results paragraph 1 - signals

1. how our model compares with lambda0 model (no phylo) in terms of variances, accuracy, etc
2. how our results compare with previous results in the literature (Ailene, PWR paper, Lizzie's and Simon's)...

Phylogenetic signal in phenological responses

1. Phenological responses to the three studied cues are overall phylogenetically conserved but estimates of phylogenetic signal differ strongly across species subsets (angio vs. gymno).
2. When angiosperm species (from main model) are considered, responses to forcing are more conserved ($\lambda = 0.64$) than responses to chilling ($\lambda = 0.66$) or to photoperiod ($\lambda = 0.35$) (see Figure 4).
3. When gymnosperm species are considered, all responses to cues are similarly low (yet different from zero): forcing ($\lambda = 0.36$), chilling ($\lambda = 0.32$) and photoperiod ($\lambda = 0.37$) and show almost overlapping posterior distributions, which may be driven by a low number of species (19) ??).

Budburst models, phylogenetic vs. non-phylogenetic

1. Here goes text comparing results with lambda $\bar{0}$ against results with estimated lambda.

Discussion

To be fleshed out.

1. Random discussion points with no home, yet ...
 - (a) This is a case where phylogeny makes a big difference! Changes overall forcing cues?
 - (b) Reduced uncertainty in species estimates (I think?) with including phylogeny (goes with above point perhaps also)
 - (c) Even with phylogeny added FagSyl is still freakish for photoperiod cue ... suggesting we've been studying an extreme species as one of our focal species (maybe?)

References

- Davies, T., E. Wolkovich, N. Kraft, N. Salamin, and S. E. Travers. 2013. Phylogenetic conservatism in plant phenology. *Journal of Ecology* 101:1520–1530.
- Davies, T. J., J. Regetz, E. M. Wolkovich, and B. J. McGill. 2019. Phylogenetically weighted regression: A method for modelling non-stationarity on evolutionary trees. *Global ecology and biogeography* 28:275–285.

- Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist* 160:712–726.
- Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *The American Naturalist* 163:84–96.
- Joly, S., D. F. Flynn, and E. M. Wolkovich. 2019. On the importance of accounting for intraspecific genomic relatedness in multi-species studies. *Methods in Ecology and Evolution* .
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Pearse, W. D., M. W. Cadotte, J. Cavender-Bares, A. R. Ives, C. M. Tucker, S. C. Walker, and M. R. Helmus. 2015. Pez: Phylogenetics for the environmental sciences. *Bioinformatics* 31:2888–2890.
- Rafferty, N. E., and P. D. Nabity. 2017. A global test for phylogenetic signal in shifts in flowering time under climate change. *Journal of Ecology* 105:627–633.

Tables and Figures

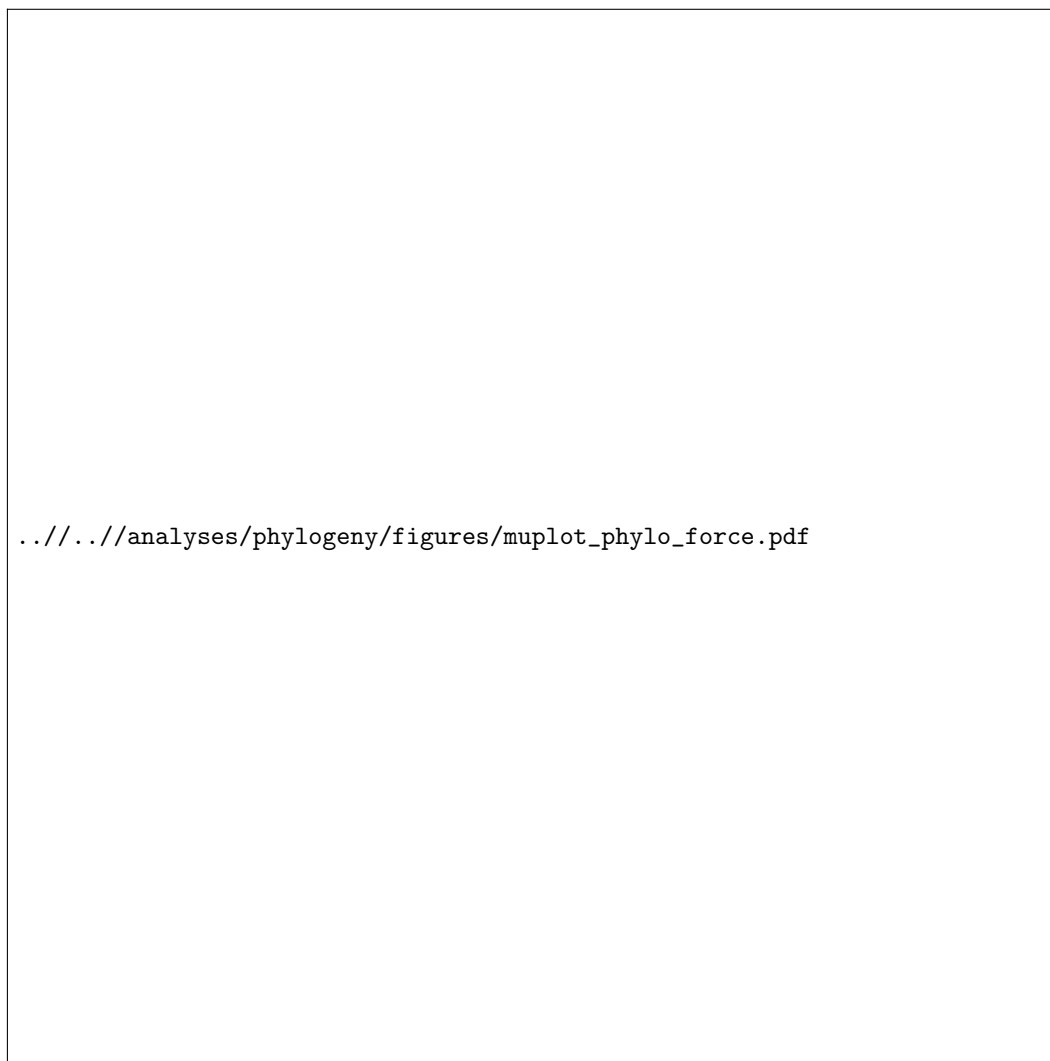


Figure 1: Cue sensitivity estimation by hierarchical phylogenetic model showing slopes for forcing, for 194 angiosperm species.

../../../../analyses/phylogeny/figures/muplot_phylo_chill.pdf

Figure 2: Cue sensitivity estimation by hierarchical phylogenetic model showing slopes for chilling, for 194 angiosperm species.

../../../../analyses/phylogeny/figures/muplot_phylo_photo.pdf

Figure 3: Cue sensitivity estimation by hierarchical phylogenetic model showing slopes for photoperiod, for 194 angiosperm species.

../../../../analyses/phylogeny/figures/lambda_sigmals_density_noout.pdf

Figure 4: Posterior distribution of phylogenetic signal measured by lambda for each cue included as a predictor in the model for angiosperms: forcing (red), chilling (blue), photoperiod (orange) and for the model intercept (grey).

../../../../analyses/phylogeny/figures/sigmas_density_lambda0.pdf

Figure 5: Posterior distribution of sigma for each cue included as a predictor in the model for angiosperms: forcing (red), chilling (blue), photoperiod (orange) and for the model intercept (grey).

../../../../analyses/phylogeny/figures/muplot_phylo_force_lambda0.pdf

Figure 6: Cue sensitivity estimation by hierarchical phylogenetic model showing slopes for forcing making lambda $\bar{0}$, for 194 angiosperm species.

../../../../analyses/phylogeny/figures/muplot_phylo_chilll_lambda0.pdf

Figure 7: Cue sensitivity estimation by hierarchical phylogenetic model showing slopes for chilling making lambda $\bar{0}$, for 194 angiosperm species.

../../../../analyses/phylogeny/figures/muplot_phylo_photo_lambda0.pdf

Figure 8: Cue sensitivity estimation by hierarchical phylogenetic model showing slopes for photoperiod making λ_0 , for 194 angiosperm species.