

Referee: 1

Comments to the Author

Summary

*This manuscript quantified flower-leaf sequence variation in the American plums, a clade of insect-pollinated species, using herbaria specimens and Bayesian hierarchical modeling and revealed hysternanthy was associated with aridity and smaller floral displays.*

General Comments

*Overall, I believe this paper addresses a significant and timely topic, and the analyses conducted were intriguing. However, it is crucial to provide additional information as the current explanation is inadequate and challenging to comprehend.*

We thank the reviewer for their efforts reviewing our manuscript and are please that they found our topic interesting and relevant. We also appreciate the reviewer's perspective on the aspects of our study that required more development and clarification. We have done our best to address the issues highlighted by the reviewer and feel that their comments have helped us generate an improved manuscript that expressed our analyses and their implications more clearly. We detail the changes we've made below.

Concerns

*I feel that the title, insect-pollinated temperate trees, is not appropriate for this study as only Prunus species was included in this study.*

This is a good point. We have modified the title to better reflect our study system. The title of the manuscript is now:

Ecological drivers of flower-leaf sequences: aridity and pollination success select for flowering-first in the American Plums

*l.15 Many trees in temperate forests produce flowers before their leaves emerge. ll.36-37 This flowering-first phenological sequence, known as hysternanthy, proteranthy or precocious flowering, is particularly common in temperate deciduous forests around the globe (Rathcke Lacey, 1985). Those sentences are misleading as flowerleaf sequence is a typical pattern for temperate trees. According to Buonaiuto et al. (2021), leaf out before flowering is a typical model of plant life history.*

Thanks for highlighting this. We have adjusted these statements to be more precise characterization of the occurrence of hysternanthy. These lines now read:

**Paste new text here**

*ll.110-114 Please clarify the collected years of sampled specimens. Are those specimens collected in a year or over the years? It may be difficult to pool the data if some specimens collected more than 50 years ago due to phenological sifts influenced by global warming.*

We appreciate this point. Our samples to span many decades of collection (X-Y), which is necessary to generate a large enough sample size to give our analyses statistical power. **We have added this information.**

Both reviewers raised the potentially issue of climate change affecting the patterns of flower-leaf sequences we sought to quantify, and for this reason, we have now included year of observation as a co-variate in our main model. Following convention from phenological change studies we, used 1980 as a hinge. With this approach year did not have a significant effect on patterns of hysternanthy, which is consistent with recent large-scale observations that despite advances in both flowering and leafout, the interval between these phases has remained relatively stable ().

*ll.114-115 In total, we evaluated the phenology of 2521 specimens, but only specimens with visible flowers were included in this analysis (n=1009). Doing some simple math, 16 species x 200 specimens/species is 3200 specimens in total. Why are there less than 3200 specimens evaluated? Also, please provide the numbers of the analysis-included specimens for each species.*

We thank the reviewer for catching this. The total sample size is less than 3200, because for several species, the database contained significantly fewer than 200 collected specimens. We have added this point to the presentation of our Methods, and now include a supplemental table with the sample sizes for each species.

We also would like to highlight that one of the reasons we chose models that partially-pool coefficient estimates on species is to account for this lack of balance among species in our dataset.

*l.113 Authors are encouraged to add a little more explanations related to BBCH scale so that readers can understand it without referring to the original references.*

We agree this would make our text easier to follow and have added a more concrete explanation of this scale to our Methods.

*ll.134-144 I believe this is an important passage that explains the criteria for determining whether each species is hysteranthous, but it took me some time to understand it. The mixing of similar numerical values such as the percentage of flowering seasonal quantile, the percentage of probability distribution, the BBCH scale, and the aggregated index made it challenging to comprehend. Could you create a diagram using several species as examples to facilitate a more intuitive understanding?*

We can see why the development of our initial index was difficult to follow and, in retrospect, we agree that building our index of seasonal quantiles wasn't the most straightforward approach. In the new version of the paper, we instead present a continuous index based on the likelihood a species' flowers appear before leaf development through across their whole flowering season. We believe this approach is equally robust and substantially more clear and useful.

We want to highlight that this index is based on the same underlying data and statistical model as our previous one, and that it yields similar results. The major difference is that we predict hysteranthly likelihood across the whole flowering season for each species, rather than just at each major quantile, which we feel is yields more precise and interpretable index than our previous one.

We also appreciate the reviewer's suggestion to provide a more conceptual figure that depicts how we classified species as hysteranthous based on a few representative examples for our species. We have developed this as Figure X, and hope the reviewer finds it useful.

*Also, it is hard to understand the followings: -Why is it needed the 0%, 25% 50% and 75% quantiles of their flowering period? -Does probability distribution mean flowering-probability distribution?*

With our new continuous index described above, we no longer use these quantiles. We also feel that this improved index circumvents the confusion between the statistical distributions and quantification of flowering probability that the reviewer highlights here.

*According to Finn et al. (2007), BBCH 00 is Dormancy: buds closed and covered by scales and 09 is Buds show green tips. I think 01 is suitable for bud development and 07 is suitable for bud break.*

I am not sure why this is in reference to.

*l. 147 Authors are encouraged to add a little more explanations related to PDSI so that readers can understand it without referring to the original references.*

We agree, and have added this an explanation of PDSI to the Methods.

*ll.197-206 Please mention the topology of Prunus species on phylogenetic tree in Fig.1b because it is also one of the results of this study.*

ll.167-173 Please indicate the number of species included to the analysis.

Fig.2-4 Please indicate sample sizes of each analysis.

Fig.2 Fig. S1 I think the flowering season and length are different among those species. Is there any trend such as hysteranthous species bloom in early spring but serathous species bloom in late spring or summer?

We agree there is a potentially interesting relationship between hysteranthous and flowering duration and time of flowering. It is true that the three most hysteranthous species in our analysis flower earlier and longer than other in our dataset. However, these species also tend to have more southern distributions, where flowering typically occurs earlier, and growing seasons are longer. We think this is an important nuance and have added a paragraph to our Discussion detailing this relationship.

Fig.4b Readers with red-green color blindness may have difficulty distinguishing between four colors.

We appreciate you identifying this and have changed our color scheme in all colored figures.

Citation: please check citation style. -l.120 (de Villemeruil P. Nakagawa, 2014)

-and and are mixed.

-Order lists of references in date order (oldest first).

Thanks. We have made these formatting changes.

Referee: 2

Comments to the Author This paper examines the adaptive value of flowering before leaf-out in trees, using data on variation across species from herbarium specimens. and analyses accounting for phylogenetic relationships.

The question addressed is important and interesting, and the approach taken appears basically sound. I do, however, have several concerns with the manuscript.

We thankful to the Reviwer for their time spent with our manuscript, and for providing important feedback. We are please that they found the topic important and interesting, and our general approach to the topic sound. We appreciata the Reviwer's concerns and have done our best to address this in the revised version of the manuscript. We feel their comments have improved the quality of our study, and we detail the changes we've made below.

The two basic ways in which selection might favor hysteranthous are correlational selection, and independent but differential selection on timing of flowering and timing of leaf-out. In the introduction, the authors touch upon this difference, but the reasoning in the manuscript appears a bit confused as Fruit maturation hypothesis, similar to the first two hypotheses, is treated as an example of correlational selection. However, the mechanisms supposed to explain hysteranthous through this mechanism does not involve correlational selection, and thus seems to fall within the category null explanations discussed in the following paragraph.

Overall, the role of fruit size as a third hypothesis to be tested is a bit unclear. While it is brought up as a separate hypothesis in the introduction, very little is said about it in subsequent sections.

We appreciate the reviewer's clear summary of the different ways selection and operate on flower leaf sequence. We found this framing very helpful and have added it to our introduction of the hypotheses.

We included the fruit maturation hypothesis in our original submission because it has been explicitly stated in the literature, but we agree with the reviewers point that this hypothesis better aligns with out null explanations than the two hpyotheses that we treat in greater detail. Because of this point, and the fact that we had low confidence in the inference from our fruit size analyses due to sample size limitations, in our revised manuscript we now mention this hypothesis as an example of our null hypotheses (**lines X**), and have removed our tests of it from our analyses.

The insect visibility hypothesis, as presented, appears a bit simplistic as the argument basically assumes that

*pollen limitation is equal among species and environments. However, if some species occur in environments that are associated with more severe pollen limitation, then you might expect that these species experience strong selection for increased visibility both through selection for an increased degree of hysteranthly, and through selection for larger floral displays. Responses to this selection would then result in a positive rather than a negative correlation.*

We agree that the insect visibility hypothesis feels overly simplistic. We feel this is in part because it has not been well developed in the literature. We with the reviewer's alternative presentation of how selection could operate to increase the size of hysteranthous flower and we had in fact presented the hypothesis this was in previous drafts of our submission. We now present the possibilities for both the positive and negative associations between hysteranthly and floral display size in our revised manuscript and include additional citations to better anchor them in the literature.

*One major problem I have with the analyses used to test the hypotheses, is that the statistical models used do not seem to correspond to the predicted causal relationships. Currently the models examine effects hysteranthly on drought and flower size respectively. However, the logic way to construct models seem to be to examine effects of drought on hysteranthly rather than effects of hysteranthly on drought. It is thus difficult to understand why the authors use the models they do, and no motivation for the chosen model structure is provided.*

We thank the reviewer for this point— we agree that some of the modeling choices we made in our original were unconventional, and we did not present clear justification for these decision. In the new manuscript, we have included additional analyses more similar to the ones suggested by the reviewer and added text to clarify our modeling choices. We address each of the the reviews concerns in more detail in the sections below.

We originally chose to to specify hysteranthly as the independent variable and traits (petal size or drought), as the dependent variable because this allowed us to account for phylogenetic autocorrelation in the petal size and PDSI in our analyses. This decision hinges on the fact that for regression models with only one predictor, we are essentially estimating the correlation coefficient between them, so *statistically*, the assignment of dependent and independent variable will not effect this relationship (). However we agree with the Reviewer's point that modeling each predictor separately was not ideal, and the that this was of assigning independent and dependent variables not as biologically intuitive, so we have move these analyses to the Supporting Information, to complement the modeling approach the Reviewer suggested below.

*Moreover, using a model with hysteranthly as the dependent variable, rather than the current models, would also allow assessing the simultaneous effects of drought and flower size on hysteranthly, something that is not done currently. This is essential as there, as pointed out by the authors, are very good reason that drought and hysteranthly are correlated. Models examining the effects of both traits are therefore essential to disentangle the effects of each trait. In fact, you might expect drought to affect hysteranthly both directly and indirectly through effects on flower size. Anyway, I think that statistical models that better reflect causal relationships, and that examine the effects of drought and flower size simultaneously are necessary.*

We agree with the Reviewer that our ideal model would include both traits as a predictors and their interactions to better understand their additive and interactive relationship to flower-leaf sequences. However, our data does not proved a straight-forward was to do this as FLS, PDSI and petal size were measured on different individual and have different sample sizes (FLS:1009, PDSI: Petal).

A way around this to to first estimate the mean PDSI and petal size for each species and uses these values in a model. We now have added a new analysis with this approach, reflected in Figure X.

We emphasize a tradeoff of this approach and out original one is that by using trait means as predictors, this model is based on only 13 rows of data (n species), cannot account for intra-specific variability in the predictors or phylogenetic correlations among them. Because of this, we have decided to keep our original analyses in the Supporting Information to complement this new analysis, and have also added more text to our methods explaining the pros and cons of these complementary approaches.

We also emphasize that the coefficient interaction between mean aridity and mean petal size was not a significant driver of hysteresis, making the qualitative inferences from these two approaches are the same, with strong associations between hysteresis and aridity and both approach and weaker associations between hysteresis and petal size.

*On the other hand, I find it highly questionable to include day of observation as a covariate in the models. This is because, as stated by the authors, hysteresis co-varies with flowering time. In fact, selection for hysteresis is likely to largely occur through selection for earlier flowering. Adjusting for day of observation will thus statistically remove some of the variation in hysteresis and bias the results. At the least, I would like to see analyses both with and without day of flowering as a covariate in order to be able to judge the effects of including it.*

For the updated version of this manuscript, we have followed the Reviewer's suggestion and produced analyses both with and without day of flowering as a covariate. While these differences do change the hysteresis likelihood estimates for some species, the relative patterns among species do not change greatly, nor do the relationships between hysteresis and the trait predictors. We have included the results from the model without day of flowering as a covariate, in the Supporting Information for comparison.

We have chosen to keep the model with day of flowering as a predictor as our main analyses because we feel it a biologically relevant explanatory variable that also helps control for temporal observer bias in herbaria records. Unlike in well-designed observational studies or experiments, herbaria record observations are not consistent across the season, and, therefore, trying to estimate flower-leaf sequences with accounting for the unevenness of observations within season could misrepresent the estimates. We have added a sentence explaining this to our Methods section and include histograms of observations across the season for each species in the Supporting Information.

*I also do not understand why the authors used the approach they did to generate a five-level index rather than using continuous functions and continuous values of hysteresis. The methods used seem to imply an unnecessary loss of information through categorizing a continuous variable, without stating any reason for doing this.*

This concern was shared by Reviewer 1, and we agree that developing this five level index masked important variation that could be captured in a continuous index. In this version, we now provide a continuous index (detailed in lines X-X), and redone all of our analyses with this metric.

*The authors in several places stress that the flower-sequence is likely to change as a result of climate change. However, climate and phenology has already changed considerably, meaning that herbarium specimens for the same species collected during different periods are likely to differ both with regards to absolute and relative phenology. Thus, if collection dates for herbarium specimens are unevenly distributed over the study period, or differ among species, then this might constitute a significant problem for the analyses. It is therefore a major short-coming that the manuscript does not contain any information about how collection dates were distributed, or even during what time period the used specimens were collected. I think that this key information must be provided, and that analyses need to account for differences in collection date.*

This issue was also raised by Reviewer 1. We have added a histogram of collection dates to our supporting information, and take steps to account for collection dates in our model. As we detailed above in our response to Reviewer 1, we have included year of sample as a covariate in our main model following convention from phenological change studies using 1980 as a hinge point. Sample year did not have a significant effect on patterns of hysteresis, which is consistent with recent findings that the interval between these phases has remained relatively stable for most species in the face of recent climate change ().

*In several places, the authors refer to the importance for the current study to understand the effects of ongoing global change. For example, on lines 56-57 they state that the study .. offers insights into how shifting flower-leaf sequences may impact species demography and species distributions as climate continues to change. Although, I agree that almost all basic knowledge about the biology of species will be helpful in this*

*respect, I do not see that this study is particularly important in this respect. I thus suggest to down-tune this type of arguments and rely on the importance of this study for our basic understanding of hysteresis.*

We appreciate this point, and agree that this study may have more relevance to understanding the basic ecology and evolution of phenological sequences than the implications for climate change. We have followed the Reviewers recommendation and shifted the context of our study away from climate change in several places in Both our Introduction and Discussion sections.

*Other comments, in order of appearance in the manuscript:*

*Abstract: It would be useful to indicate what hypotheses that were tested, as well as the general methods used to test them. Will do.*

*Lines 35-36: This is not correct. Flowering before leaf-out occurs also in non-woody species.*

This is an important point. We have amended this statement.

*Lines 37-41: Explain what kinds of functional significance you refer to here.*

TBD.

*Lines 146-150: I assume that drought indices have changed over this considerable time period. What was the reason for using this time period, and for using the same time period for all records?*

We agree that our decision to truncate the drought index at for this period was arbitrary. In the new version, we use the full 2000 year record provided by our data source.

While we also agree it is certainly possible that aridity levels at a given location have changed over this time period, we see no clear way coherently assign different evaluation periods for different records, or truncate specific records based on change at a location. While this could be possible with extending spatio-temporal modeling approaches, we do not have any baseline information about how long a population has been present at a given locality, which we feel would make it difficult to root any attempt to fine-tune the time period we evaluate in ecological sound. If the reviewer has any suggestions for a straightforward way to do this we would certainly be open to trying it.

*Lines 151-154: The unit relevant for attracting pollinators is probably not only the individual flower, but entire inflorescences, or even trees. Would it thus be useful to examine effects of an attraction parameter that includes also, for example, the number of flowers per inflorescence?*

We agree with this point, and in our analyses of the larger *Prunus* genus we use do indeed use number of flowers/inflorescence as a predictor. By contrast, the American plums all have solitary flowers, which is why we used petal length as our predictor. We have added text to discuss this nuance in our Discussion.

*Lines 178-185: Why did you use a different number of categories for this analysis?*

This choice was based on data availability. As we mention in our methods, our flower-leaf sequence data for this group come from qualitative descriptions of which only four levels are available.

*Lines 209-210: I do not really understand the meaning of this. You also seem to use different wording to explain the same effect in the three paragraphs of the results section.*

*Line 217: Why However?*

*Lines 220-221: I suggest to skip sentences like these and let the results talk for themselves.*

*Line 222: .. through the predictions ??*

*Line 225: Strange subheading given that this is the topic of the entire paper?*

*Line 228: Unclear exactly what trade-offs you refer to here.*

*Lines 262-291: Here, I think that it would have been interesting to provide quantitative information about how much of the total variation in hysteresis that was within vs. among species.*

We can't. THIS IS NOT AN OBSERVATIONAL STUDY.