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 hysteranthy 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 pleased 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.

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 hysteranthy, 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 a precise characterization of the occurrence of hysteranthy. These lines now read:

> This flowering-first phenological sequence, known as hysteranthy, proteranthy or precocious flowering, is apparent in temperate deciduous forests around the globe.

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 span many decades of collection (1844-2020), which is necessary to generate a large enough sample size to give our analyses statistical power. We have added this information at line 106.

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. We detail these methods in lines 126-130. In this case, year did not have a significant effect on patterns of hysteranthy.

We have also added a paragraph to our Discussion regarding this point in lines 325-331.

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 that 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 (lines 110-115), and now include a supplemental table with the sample sizes for each species (Tab. S1Sample sizes of each for each species used in this studytable.1).

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 at line 109 .

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 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 hysteranthy 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 think the these changes suggested by the reviewer have made our methods much easier to follower, even without an additional visual aid, Howver if the reviwer still feels that a conceptual figure would enhance the clarity of the study, we would be have to discuss this further.

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.

In the revised version of the changed our verbal descriptions of the vegetative BBCH stages to more closely align with how they are described in Finn *et al.* (2007). These changes can be found at lines 141-151.

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 (lines 160-163).

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.

We now include reference to the original tree toplogy in the figure.

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

We have added the sample size to our Methods section (line 190).

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

We have added this information to the figure captions.

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 hysteranthy and flowering duration and time of flowering. We now mention this more explicit in our paper at line 123. We think this is a really interesting question to persue further, to do so, we think you would really need to understand both spatial bias in the dataset and species-level differences in range size, which we think is beyond the capacity of our models to account for in this context.

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. **TO DO**

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 thank the Reviewer 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 be sound. We appreciate the Reviewer's concerns, and have done our best to address them 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 hysteranthy 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 hysteranthy 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, and agree that the fruit maturation hypothesis fits more broadly within the category of "independent but differential selection".

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 hypotheses 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 (line 76), 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 hysteranthy, 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 also agree 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 hysteranthy and floral display size in our revised manuscript (lines 67-73) and **include additional citations to better anchor them in the literature. If we need to**

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 hysteranthy on drought and flower size respectively. However, the logic way to construct models seem to be to examine effects of drought on hysteranthy rather than effects of hysteranthy 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 we did not present clear justification for some of the modeling choices we made in our original submission. In the new manuscript, we now include several new analyses more similar to the ones suggested by the reviewer and add text to clarify our modeling choices. We address detail thse changes below.

Moreover, using a model with hysteranthy as the dependent variable, rather than the current models, would also allow assessing the simultaneous effects of drought and flower size on hysteranthy, something that is not done currently. This is essential as there, as pointed out by the authors, are very good reason that drought and hysteranthy 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 hysteranthy 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 wau to do this as our hysteranthy indices, PDSI and petal size variables were measured on different individual and have different sample sizes (FLS:1000, PDSI:2305 Petal:2757).

We now explicitly make this point out in lines 167-lines 169 and have executed new analyses to address this. First, to capture the additive and interactive effects of our predictors on hysteranthy we now utilize a model where we use species-level means of PDSI and petal size and test their relationship to our hysteranthy index using a beta regression framework. This is the analyses we now present in our main text and Fig. 3Relationships between hysteranthy index scores and environmental and biological traits based on Bayesian Beta regression for the 13 species of the American Plums. Panel a) shows the estimated effect size estimates of each predictor and their interaction on the mean shape parameter of the beta distribution. Points indicate the mean effects and the thick and thin bars represent the 50 and 89% uncertainty intervals. b) and c) depict the conditional effects of each predictor and hysteranthy index values. Blue lines indicate the mean estimate

and grey fill the 89% uncertainty intervalsfigure.caption.22.

This approach captures the important interaction between the predictors, but the tradeoff is that it losses the within-species variation in the environmental/morphological traits and the phylogenetic structure. For this reason, we also include versons of our original analyses in which we model the relationship between hysteranthy and each trait seperately, which allows us to interchange the placement of the dependent and independent variable (**citation for this?**) and used phylogenetic mixed models. We now include these complementary analyses in Fig. S3Relationships between the hysteranthy index scores and environmental and biological traits based on Bayesian phylogenetic mixed models where the relationship between hysteranthy score and PDSI a) and mean petal length b) were each modeled seperately. Solid lines indicate the mean posterior estimate and shaded areas 4000 draws from the posterior distrubtion as a display of uncertainty. The points are jittered along the x-axis only for visibility. This plots complements the model presented in Fig. 3Relationships between hysteranthy index scores and environmental and biological traits based on Bayesian Beta regression for the 13 species of the American Plums. Panel a) shows the estimated effect size estimates of each predictor and their interaction on the mean shape parameter of the beta distribution. Points indicate the mean effects and the thick and thin bars represent the 50 and 89% uncertainty intervals. b) and c) depict the conditional effects of each predictor and hysteranthy index values. Blue lines indicate the mean estimate and grey fill the 89% uncertainty intervalsfigure.caption.22figure.3, and as Extended Methods in our Supporting Information.

We should note that we found that the coefficient interaction between mean aridity and mean petal size was not a significant driver of hysteranthy, making the qualitative inferences from these two approaches similar, with strong associations between hysteranthy and aridity and both approach and weaker associations between hysteranthy 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, hysteranthy co-varies with flowering time. In fact, selection for hysteranthy is likely to largely occur through selection for earlier flowering. Adjusting for day of observation will thus statistically remove some on the variation in hysteranthy 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 co-variate (see lines 154-157). While these differences do change the hysteranthy likelihood estimates for some species, the relative patterns among species do not change greatly, nor do the relationships between hysteranthy and the trait predictors. We have included the results from the model without day of flowering as a co-variate, in the Supporting Information for comparison (Tab. Tablestable.1,Tab. Tablestable.2,Fig. S2Predicted likelihood that a species would be in flower during each vegetative BBCH phase for each species in the American plums. Points are the mean likelihood while bars represent 89% uncertainty intervals. Hysteranthy index values in each box were derived from the summed likelihood species would be found at BBCH stages 0-11 (purple fill)figure.2).

We have chosen to keep the model with day of flowering as a predictor as our main analyses because we feel it a biological relevant explanatory variable that also helps control from temporal observer bias in herbaria records. Unlike in well-designed observational studies or experiments , hebaria record observations are not consistent across the season, and, therefor, trying to estimate flower-leaf sequences with accounting fore the unevenness of observations within season could misrepresent the estimates. We have added a sentence explaining this to our Methods section (line 124) and include histograms of observations across the season for each species in the Supporting Information, Fig. Figuressection*.2.

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 hysteranthy. 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 141-151), 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 on 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 issues was also raised by Reviewer 1. We have added information about collection dates to our Methods (line 106.), 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 co-variate in out main model following convention from phenological change studies using 1980 as a hinge point (detailed in lines 126-130 of the revised manuscript). Sample year did not have a significant effect on patterns of hysteranthy (Fig. 2Predicted likelihood that a species would be in flower during vegetative BBCH phases for 13 species of the American plums. Panel a) depicts the influence of among season (year) and within season (day of year) temporal predictors on the likelihood would be be at earlier or later vegetative BBCH phases during flowering. Points are the mean effect size estimates while think and thin bars represent the 50 and 89% uncertainty intervals respectively. Panel b) depicts the predicted likelihood that each species would be hysteranthous across their flowering season. The colored shapes represents how the likelihood changes over each species flowering season and the boxed numerical values represent a hysteranthy index, or the total average likelihood a species would express hysteranthy throughout their flowering period. We defined hysteranthy as having open flowers at BBCH 0-BBCH 11 (leaf buds closed-start of leaf unfolding). See Tab. **??** for comparisons between the hysteranthous index scores reported here and alternative modeling approaches. figure.caption.21a), 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 (Guo *et al.*, 2023).

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 hysteranthy.

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 our Introduction and Discussion seconds. We now only discuss climate change in the context of our model covariate (see above) in lines 325-331, and include the new citation from Guo *et al.* (2023) suggesting the stability of flower-leaf sequences we observed in our data is consistent with wider observations.

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.

We have added this information the the Abstract.

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 in line uniq1. It now reads:

Woody perennials are among a subset of plant types with the unique ability to seasonally begin reproduction prior to vegetative growth.

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

We have replace this statement with a more precise one: "can confer performance advantages...".

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 possibly 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 approached, 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-time 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 iy.

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 at lines 267-273.

Lines 178-185: Why did you sue a different number of c 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. We have added a sentence to explicitly make this point at line 205.

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.

With the incorporation of our new continuous flower-leaf sequence index, we have re-written the Results sections. We hope that the reviewer finds this presentation more clear.

Line 217: Why However?

This section is no longer included in our revised manuscript.

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

We have removed this sentence.

Line 222: .. through the predictions ??

As the reviewer suggested about we have worked to tighten the language of this summary paragraph of our Discussion, and agree the meaning of this statement was not clear in our original submission. We hope the stylistic changes we've made here (lines 243-251) read more clearly in this version.

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

We thank the reviewer for this point and have removed this subheading.

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

We have elaborated on this in lines 246.

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

We agree with the Reviewer that this would be a very interesting contribution to our understanding of these phenological sequences. However, because our study was based on herbaria records that were collected systematically across space and time without any repeat sample, we feel that we cannot robustly partition this variation with the present data. We have added a paragraph clarifying this limitation and suggesting it as an area for further study at lines 314-318.