

*Comments to the Author*

*This article examines various theories of hysteresis (HA) in species of plums using herbarium specimens from North America. The main results are species exhibiting HA tend to be found in more arid environments and have smaller flowers. The authors suggest further lines of evidence for investigating HA. This is an original paper using novel methods to investigate HA, especially the methods for scoring specimens. I looked forward to reading this paper as the approach had the potential to provide important new insights.*

We thank the reviewer for their thoughtful comments, and are glad they find our paper original and methods novel. We feel that their feedback has helped improved this submission.

*The most important goal of the paper was to investigate the relationship between aridity and HA. However, these species flower in spring in Eastern North America where there is abundant rainfall in the spring and the ground is typically saturated with water. So it was not apparent why aridity was being investigated.*

We agree that it is unexpected that aridity is related to hysteresis in wet temperate region we are studying, and state this explicitly in our introduction. Despite this, the water limitation hypothesis is one of the most developed functional hypotheses of hysteresis and two recent analyses, published in this journal have found support for this hypothesis in these regions (). This unexpected result in previous studies is one of the major reasons we developed this high-resolution study to better understand the dynamics of this relationship.

*Also, it is well known that species in this region flower in spring in response to temperature, so why use an aridity index that combines temperature and precipitation? It would seem far better to use temperature and precipitation separately, and such data is readily available; my prediction is that the authors would get the same results or even better results by just using temperature; that is, species in warmer places would show greater HA.*

As the reviewer suggest we have added two new analyses which substitute 100+ year temperature and precipitation averages for our aridity index. The times scales on which these data were collected make direct comparison complicated, but we feel these new analyses were extremely helpful for addressing nuances between the environmental conditions that shaped the biogeographic/evolutionary patterns we are interested in and the contemporary environmental conditions that influence plasticity in flower-leaf sequences, which has resulted in several new figures (Fig. 3, 5, plus two supplement ones I need to do), and a new focal area in our Discussion (lines 320-355). **could just end here or go on.**

Because several climatology studies are clear that precipitation and temperature separately are less reliable for capturing drought conditions (**need help finding these citations**) than aridity indices, we feel that our model with the pdsi index remains the better approach for testing the water limitation hypothesis, while the temperature and precipitation models are most useful for addressing the points about intra-specific variation suggested by the reviewer below.

*And for the PDSI index, the authors need to clearly state what the units mean, and what the difference is between negative and positive values.*

We appreciate this point and now elaborate on the pdsi index in lines (to do) .

*The focus of the results is on differences among species in HA. It was puzzling to me why the authors did not look at variation in HA within species. Within a particular species, did individuals in more arid (or warmer) places and warmer years have a greater tendency to show HA? This is something that the authors already have the data to do. So why not do it?*

We think this is a really interesting question, and have now added three new analyses, resulting in a new figure (Fig. 5) and a more nuanced discussion about the way climate has influenced flower-leaf sequences at different temporal and taxonomic scales.

*The authors have developed a model that is overly complicated for looking at the results so that it is not possible to determine what the results mean. In the results, the authors say: "parameter estimates from*

*this model were  $\beta PDSI : 0.47, UI89[0.96, 0.01]$ ,  $\beta petallength : 0.14, UI89[0.54, 0.24]$   $\beta PDSI \times petallength : 0.14, UI89[0.91, 0.65]$ ” ‘ What these results mean is hard to say, even for scientists working in this field. From the results, one gets the impression that HA is strongly influenced by aridity and somewhat by petal size, but then in the Discussion, the authors say:*

*“The comparatively small window of leafless reproductive development in our temperate clade may, in part, explain why the association we observed between hysternanthy and aridity was relatively weak and variable.” So, HA is not particularly explained by aridity and probably not explained at all by petal size, contrary to what is implied in the results.*

We appreciate the reviewers point here, and feel the confusion arises from the line in our discussion where in retrospect, we did a poor job trying to contextualize our statistically meaningful model results with their biological significance. We have removed this language and move this subject material to an earlier part of the Discussion where we discuss the implications of our sensitivity analyses (lines

We appreciate that our modeling approach is complex, which stems both from our desire to model these data most appropriately (an ordinal model that accounts for phylogenetic structure) and comments from previous reviewers that asked for more complex models to for climate change shifts in our species phenology model, and interactions between trait predictors in our models testing the hysternanthy hypotheses. We hope that our the changes that we have made to they way we interpret our Results clarifies their meaning.

*Also, Figure S3 shows what appear to be all the data in the study as two data clouds with minimal or no relationship between HA and aridity, and no apparent relationship between HA and petal size. This is contrary to what is stated in the results.*

*What is needed is a simple figure showing the PDSI on the x-axis and the HA index on the y axis for the mean of each of the plum species, perhaps with a regression line going through the points. This would show how strong the relationship is and make it clear to the readers. This should be substituted in for the current Figure 3. Also a similar figure for HA and petal length. Perhaps such figures could be added for all of the Prunus species.*

We appreciate the reviewers clear suggestions to make our figures more interpretable and have added points representing the mean trait values of plums species to our marginal effect plots in figure 3 and S3 (need to do this for the supplement). We also report the parameter estimates and uncertainty intervals from all model coefficients (in text lines X and Figures XYZ) so that readers have concrete information about the strength of the relationships presented in our figures.

*Perhaps a simpler explanation for the results is that flower buds are more sensitive to temperature than leaf buds, and HA is more noticeable in warm locations and warm years than in colder locations and colder years. And this could be true both among species and within species.*

We appreciate this point, and at the reviewers suggestion, we’ve conducted new analyses investigating the relation of long-term temperature averages and interspecific patterns of flower-leaf sequences as well as the relation between temperature variation and flower-leaf sequence variation within species. These analyses are presented in figure 3 and 5 and discussed in lines X-Y. Interspecific differences in hysternanthy were not strongly associated with temperature differences, but variation in hysternanthy was related to temperature, which we now discuss in lines (X)

*A major focus of the paper is looking at the relationship of petal size and inflorescence size in relation to HA. The author argue that these flower characters provide an indication of pollinator attractiveness. However, when these Prunus species are in flower, the plants are covered with masses of flowers in clusters; it is not the individual flower which is the unit of attraction in these species, but clusters of flowers, clusters of inflorescences and even whole plants. Perhaps this is why flower size is not significant in the results. As a consequence, the title is misleading. This paper is about flower characteristics, and not about pollination success as implied in the title.*

We strongly agree with point the reviewer raises that the unit of “attractiveness” for these species more than single flowers or even inflorescences, and greatly, appreciate their point that this might explain the weak relationship between petal size and hysteranthly, which we have added to our discussion (lines 359-364). We agree that flower size may not be the best proxy for attractiveness, but the herbarium specimens we used in this study don’t provide any better way to quantify flowering intensity at the whole plant level. This limitation was one of the reasons we also included analyses of with the larger genus, using inflorescence size as a predictor, which we feel is a better proxy for attractiveness, and also yeilded a strong relationship with flower-leaf sequence variation. **Decide of we want to just change the title or add a sentence to defend not doing so.**

*The authors might also consider other explanations for HA, such that is it driven by the time needed for fruit maturation and characteristics of the leaves. Do species that exhibit HA have larger fruits and more drought-resistant leaves than non-HA species?*

We thank the reviewer for this point. We originally included an analysis of fruit size in this study—a previous reviewer raised a major concern that this explanation for hysteranthly was not a example of correlated selection between flower-leaf sequences and traits, but a case of selection on flowering time under which hysteranthly would be a incidental by-product that couldn’t be robustly tested with our approach. We were convinced by this reviewer that these explanations should be provided as an example in our introduction of the null hypotheses (line 80). We think that this explanation for hysteranthly is a plausible one, and hope to explore it further in future work with other methodological approaches.