University of Massachusetts Amherst

Dear Dr. Rausher,

Please consider this revised manuscript re-submission, now titled: "Ecological drivers of flower-leaf sequences: aridity and proxies for pollinator attraction select for flowering-first in the American Plums" as a "Full paper" in *New Phytologist*.

Our paper tests two hypotheses that seek to explain why some woody plants species have evolved to flower before they leafout each season. Our findings—which support both the **Water limitation hypothesis** (Gougherty & Gougherty, 2018; Buonaiuto *et al.*, 2021) and the **Insect visibility hypothesis** (Janzen, 1967)—are timely and relevant for understanding phenological adaptations that may enhance species' fitness in a changing environment.

Our submission has now been reviewed by three anonymous reviewers that have helped to shape this manuscript into an impactful study that we think would be of great interest to readers of *New Phytologist*. At our first round feedback, Reviewers requested several additions to our modeling approach and sensitivity analyses. This effort required the re-analysis of all our data from which we generated 4 new main text figures and several additional figures and tables in the Supporting Information that showed our finding were robust to alternate analytical approaches.

In the current round of feedback, Reviewer 2 reported that our paper was original, and used novel methods with potential to provide important new insights about phenological sequences of woody plants. Based on their feedback, we have made improvements to several of our results figures (Fig. 3 and S3), added critical text in our Discussion to delve into the nuances of our findings across taxonomic space and over the climatic history of the past two millennia, and updated our manuscript title to more concretely represent our findings. We feels these changes that we made based on the reviewers' comments helped further refine our manuscript.

Review 2 also suggested that using simple climate metrics like annual temperature and precipitation could provide equal or better inference on the hypotheses we tested. We explored the possibility of using many climate metrics to test the predictions of our hypotheses, and were convinced by the substantial evidence in the literature that precipitation and temperature separately are poor indicators of plant available water/aridity conditions (Moles *et al.*, 2014; Piedallu *et al.*, 2013; Hickler *et al.*, 2009), which is the critical feature of the water-limitation hypothesis. Drought indices such as the one we use in our study are considerably more reliable in this regard (Moles *et al.*, 2014; Dai *et al.*, 2004; Mika *et al.*, 2005), so we feel that adding alternate models would not benefit our analyses. Because the reviewer pointed out that a strong temperature association could also be a signal of our null hypothesis, we did perform a preliminary analyses substituting temperature for PDSI in our models. We found this version of the model explained considerably less variation than our original model and was highly sensitive to outliers, so we decided not to pursue this analysis further.

It was clear that Reviewer 2 is very knowledgeable about flowering phenology and interested in many related questions to those we addressed in our study including patterns of intra-specific variation, physiological responses to climate variation, and alternative hypotheses to the ones we tested. While we share the reviewer's

interest in these questions, they are not the ones we set out to address, and our approach to data collection was not designed for investigating these topics. We feel that to address these questions well, we would need to acquire several new data sets and perform numerous new analyses that would constitute a whole new study, and do not feel that we should pursue them as additions to our existing study which already includes several analyses across multiple taxonomic scales.

We'd like to highlight that Reviewer 2 did not identify any clear flaws in our choice of predictors, analyses or results, and their comments that we have not been able to incorporate at this stage are primarily ideas about substantial extensions of our topic. We feel that it would be more appropriate for these inquiries to receive due attention as their own research questions in future works, and we hope that you agree with this assessment.

The main text of this manuscript is 4,830 words in length and it contains four figures. It is co-authored by T.J. Davies, S. Collins and E.M. Wolkovich, and is not under consideration elsewhere. We hope that you will now find it suitable for publication in *New Phytologist*, and look forward to hearing from you.

Daniel Buonaiuto

References

Buonaiuto DM, Morales-Castilla I , Wolkovich EM. 2021. Reconciling competing hypotheses regarding flower–leaf sequences in temperate forests for fundamental and global change biology. *New Phytologist*, **229**: 1206–1214.

Dai A, Trenberth KE , Qian T. 2004. A global dataset of palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology*, **5**: 1117–1130.

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.

Hickler T, Fronzek S, Araújo MB, Schweiger O, Thuiller W , Sykes MT. 2009. An ecosystem model-based estimate of changes in water availability differs from water proxies that are commonly used in species distribution models. *Global Ecology and Biogeography*, **18**: 304–313.

Janzen DH. 1967. Synchronization of sexual reproduction of trees within the dry season in central america. *Evolution*, **21**: 620–637.

Mika J, Horváth S, Makra L , Dunkel Z. 2005. The palmer drought severity index (pdsi) as an indicator of soil moisture. *Physics and Chemistry of the Earth, Parts A/B/C*, **30**: 223–230. Agrometeorology 2003.

Moles AT, Perkins SE, Laffan SW, Flores-Moreno H, Awasthy M, Tindall ML, Sack L, Pitman A, Kattge J, Aarssen LW et al. 2014. Which is a better predictor of plant traits: temperature or precipitation? *Journal of Vegetation Science*, **25**: 1167–1180.

Piedallu C, Gégout JC, Perez V , Lebourgeois F. 2013. Soil water balance performs better than climatic water variables in tree species distribution modelling. *Global Ecology and Biogeography*, **22**: 470–482.

Reviewer comments are in italics; author responses are in plain text.

Comments to the Author

This article examines various theories of hysteranthy (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 topic original and methods novel. We feel that their feedback has helped improve 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 not immediately intuitive that aridity may be related to hysteranthy in the wet temperate region we are studying, and discuss this explicitly in our introduction (line 64). However, the water limitation hypothesis is one of the most developed functional hypotheses of hysteranthy and two recent analyses, published in this journal, have found support for this hypothesis in these regions (Gougherty & Gougherty, 2018; Buonaiuto *et al.*, 2021). We now also address this point in our Discussion (lines 313-318), where we offer a possible explanation for this un-intuitive pattern with a new figure in our Supporting Information (Fig. S4).

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.

We appreciate this point, and during the development of this project, we spent many hours searching for the appropriate climate variable to test the prediction of water limitation hypothesis that hysteranthous species should be found in more arid areas that non-hysteranthous ones. Numerous sources consistently report that temperature and precipitation separately are not reliable metrics for capturing plant available water, and that drought indices that combine precipitation and temperature perform better (Moles *et al.*, 2014; Piedallu *et al.*, 2013; Hickler *et al.*, 2009). We ultimately chose PDSI as our metric because 1) it is a widely-used drought index that is relevant to plant water status (Dai *et al.*, 2004; Mika *et al.*, 2005), and 2) it can be reconstructed through dendrochronology (Cook *et al.*, 2010), providing climate information that is on a more relevant timescale to the process of range filling that is related to the predictions of our hypotheses. Using PDSI, we were able to obtain a 2000+ year record of aridity measurements, but the longest times series of temperature and precipitation data we could find with the spatial coverage necessary for our analyses was 115 years. For these reason we favor using a long-term aridity index like PDSI to test the water limitation hypothesis than contemporary precipitation and temperature.

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.

Thanks for this point. We now elaborate on the PDSI index in lines 175-177.

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 hope to address it in future work. From the outset, we

were interested in understanding species-level differences in hysteranthy, and designed our sampling approach accordingly. We do not feel that we have large enough sample sizes within each species to answer detailed questions about intra-specific variation adequately, especially given the challenges that arise from working with patchy spatio-temporal data without repeat sampling. We acknowledge this limitation in lines 348-352, and are excited about the possibilities for addressing this question using a different approach to sampling herbarium data in the future.

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 $\beta PDSI : 0.47, UI89[0.96, 0.01]$, $\beta petallength : 0.14, UI89[0.54, 0.24]$ $\beta PDSI x petallength : 0.14, UI89[0.91, 0.65])$" ' 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 hysteranthy 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 reviewer's point here, and feel the confusion arises from the line quoted in our Discussion, where—in retrospect—we did a poor job contextualizing our statistically meaningful model results with their biological significance. We have rephrased this section of our Discussion, which now relies more heavily on the materials presented in lines 319-331 to contextualize the biological implications of our analyses.

We also appreciate that our modeling approach may appear complex, which stems both from our desire to model these data most appropriately (with an ordinal model that accounts for phylogenetic structure and latent temporal variation) and comments from previous reviewers that asked for more complex models to account for climate change shifts in our species phenology model, and interactions between trait predictors in our models testing the hysteranthy hypotheses. We hope that the changes we made to the Discussion better reinforce the reviewer's correct interpretation of our findings from the Results section.

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 plum species to our marginal effect plots in Figures 3 and S3. We also report the parameter estimates and uncertainty intervals for all model coefficients (in text lines 252, 260, and Tab. S3) 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 while we do not feel we can reliably assess drivers of intra-specific variation with our data, we agree that differences in temperature responses between flower and leaf buds relate to the null hypothesis we presented in our manuscript.

Following this suggestion, we performed some preliminary analyses to check if a model with average temper-

ature explained variation in hysteranthy better than petal size and aridity. When we re-fit our main model with average annual temperature instead of PDSI, it explained considerably less variation than our original model (Bayesian R-squared value of the temperature model was 30% lower than R-squared value of the PDSI model). This temperature model was also highly sensitive to outliers. Given the weaker explanatory power and the fact that the temperature records we were able to obtain had the same issues of data truncation we discussed above, we decided against proceeding with this analysis.

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 have changed our title to reflect the reviewers feedback here, as we agree with point the reviewer raises that the unit of "attractiveness" for these species is more than single flowers or even inflorescences, and we are only measuring some proxies for pollinator attraction. We appreciate their point that this might explain the weak relationship between petal size and hysteranthy, which we have added to our discussion (lines 284-288). 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 the larger genus using inflorescence size as a predictor as we felt this was a better proxy for attractiveness, and also yielded a strong relationship with flower-leaf sequence variation.

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. While our estimates for the influence of PDSI and flower display size were robust with the inclusion of this additional predictor—a previous reviewer raised a major concern that this explanation for hysteranthy was not a example of correlated selection between flower-leaf sequences and traits, but a case of selection on flowering time under which hysteranthy would be a incidental by-product that couldn't be robustly tested with our approach. We were convinced by this reviewer that this potential explanation for hysteranthy should, instead, be given as an example of a null hypothesis (line 80).

References

Buonaiuto DM, Morales-Castilla I , Wolkovich EM. 2021. Reconciling competing hypotheses regarding flower–leaf sequences in temperate forests for fundamental and global change biology. *New Phytologist*, **229**: 1206–1214.

Cook ER, Seager R, Heim Jr RR, Vose RS, Herweijer C , Woodhouse C. 2010. Megadroughts in north america: placing ipcc projections of hydroclimatic change in a long-term palaeoclimate context. *Journal of Quaternary Science*, **25**: 48–61.

Dai A, Trenberth KE , Qian T. 2004. A global dataset of palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology*, **5**: 1117–1130.

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.

Hickler T, Fronzek S, Araújo MB, Schweiger O, Thuiller W , Sykes MT. 2009. An ecosystem model-based estimate of changes in water availability differs from water proxies that are commonly used in species distribution models. *Global Ecology and Biogeography*, **18**: 304–313.

Mika J, Horváth S, Makra L , Dunkel Z. 2005. The palmer drought severity index (pdsi) as an indicator of soil moisture. *Physics and Chemistry of the Earth, Parts A/B/C*, **30**: 223–230. Agrometeorology 2003.

Moles AT, Perkins SE, Laffan SW, Flores-Moreno H, Awasthy M, Tindall ML, Sack L, Pitman A, Kattge J, Aarssen LW *et al*. 2014. Which is a better predictor of plant traits: temperature or precipitation? *Journal of Vegetation Science*, **25**: 1167–1180.

Piedallu C, Gégout JC, Perez V , Lebourgeois F. 2013. Soil water balance performs better than climatic water variables in tree species distribution modelling. *Global Ecology and Biogeography*, **22**: 470–482.