

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 hysteroanthony 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.

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.

l.15 Many trees in temperate forests produce flowers before their leaves emerge. ll.36-37 This flowering-first phenological sequence, known as hysteroanthony, proteranthony 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.

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 shifts influenced by global warming.

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.

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.

ll.134-144 I believe this is an important passage that explains the criteria for determining whether each species is hysteroanthous, 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?

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?

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.

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.

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?

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

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).

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.

The two basic ways in which selection might favor hysteranthly 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 hysteranthly 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.

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.

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.

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.

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, hysteranthly co-varies with flowering time. In fact, selection for hysteranthly is likely to largely occur through selection for earlier flowering. Adjusting for day of observation will thus statistically remove some on the variation in hysteranthly 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.

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 hysteranthly. The methods used seem to imply an

unnecessary loss of information through categorizing a continuous variable, without stating any reason for doing this.

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.

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.

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.

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

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

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?

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?

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

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.