

Reviewer comments are in italics. Author responses are in plain text.
As seen in 25.

Reviewer #1 (Comments to the Author)

In this article, Buonaiuto and colleagues proposed and tested a new framework that quantifies the impact of various hypotheses that may explain the variability in flower-leaf sequences (FLS) among species. The subject is timely and of the interest of the readership of New Phytologist by bringing new insights on the phenology and forest trees, and especially on the chronology between flowering and leaf budburst. The proposed framework is novel, and the statistical approach and data used to test such framework is appropriate. However, I do have several concerns that should be considered.

- Say Thanks
- say We are super happy with their contributions and critiques for strengthening the paper

First, I do not think that using the minimum precipitation across the species range is appropriate. The authors mentioned this limit in L311-314, but I think that using other indicators to test the “water limitation” won’t be time-consuming and would provide much more reliable results. For instance, this variable doesn’t differentiate cold-dry and warm-dry climates while we know that high temperatures exacerbate drought through increasing evapotranspiration. There are plenty of drought indices that are available (e.g., Speich 2019), but the authors could simply use P-ETP, that can be calculated annually. Another option would be to use species functional traits that may better reflect their drought tolerance, like the PLC50, the water potential causing 50% of the xylem embolism, or the hydraulic safety margin (Martin St-Paul et al. 2017). Such functional traits are now widely available for a large number of species, e.g., through the TRY database. Also, I would not use the term of “water dynamics” as there is a notion of temporal change in “dynamics”, but rather ‘water limitation’

- We agree
- but Gougherty et al called it “water dynamics”. could add water limitit or waterdynamics with citation and then refer to it as water limitation for the rest of the paper
- For functional traits (species level): Show how many species we lose when we use p50 from try, and Barrett et al
- For functional traits (species level): could do analysis with catagorical drought tolerance (High, Med, Low)..show they correlate better to P50 values from NRCS
- Could do a model with drought indices... to explain interannual variation at HF

Second, one key aspect in the analysis of FLSs relies in the choice of the criteria which determines when the flowering and leaf-out start (as shown in Figure S1 for instance). However, the authors used different criteria for their analyses. (1) With the PEP725 database they used the BBCH60 and BBCH11 for flower and leaf phenology, respectively (L14; methods S1). (2) With the Harvard forest data, they used the time between flowers opening and leaves reaching 75% of their final size. Iâ€™m wondering is the authors could test the sensitivity of their models to the BBCH classes used (with the PEP725 and/or Harvard forest database). Such additional information would strengthen the manuscript without complicating its reading.

- Make a table to apporximate between Harvard forest and bbcch
- Run supplimental analysis and figure for HF using budburst and flower budburst

- add paragraph to supplimnet explain the limitations of the data caused use to choose which bbch stages to use

Third, the authors highlight the interests of adopting an intra-specific approach (L30-31; L210-121 –), a viewpoint that I share. As shown in Figure 3, omitting the inter-annual variation oversimplifies the analysis of FLS, and hinders a deeper analysis of its climatic drivers. I agree that “high levels variation across individuals of the same species could suggest that local adaptation and subtle differences in micro-climate, soil, light radiation, or topography contribute to FLS variation” (L231-233). However, (1) I don’t see how the intra-specific variability is considered in the proposed framework, especially in the models developed. All the explanatory variables are compiled at the species level. (2) There are multiple sources of variation within a species: among populations (e.g., along an elevational gradient; Vitasse et al. 2009), among individuals from a same population (e.g., Denechere et al. 2019), among years, and among individuals and years (individuals can show a different sensitivity to climate). This additional “population” level should be better considered in the framework.

- clarify pooling of individuals in model
- show a flow chart of how quantitative can give allow inquiry at other levels
- add and interannual variation model P-etc to show interannual variation

L39: “their chronology” instead of “the relationships between them”

Fixed in line 36

L46-47: flowers can be white as well. Use “colors” instead of “red and yellow”

Fixed in line 44

L47: how much common is this flowering-first FLS? If this has never been quantified, I would suggest the authors to calculate the percentage of tree species that show hysteresis.

We agree that a quantitative assessment would be helpful. We have added estimates from previous work on FLS. (?) in 46.

L52: –the rate of change differs up to five-fold among species– This is not obvious from the figure 1. I would show these rates in an additional panel (e.g., with boxplots that would also show the with uncertainty in the calculation of this rate)

- Add said box plots

L57-59: –Long-term data also highlight high within-species variability in FLS– and –most research has not addressed this variability–. Looks like a contradiction. Please rephrase

We agree with the reviewer that these statements read as contradictory.. We have adjusted the language for clarity in 56-59.

L85-91: Here the authors can also mention that flowers are more sensitive to drought-induced xylem embolism than leaves: see Zhang Brodrick 2017

- Add a part of a sentence for this to water dynamics

L92-101 and throughout the paper: regarding the “early flowering” hypotheses. Please also consider the recent work of Schermer et al. 2020, which shows that early pollination may be responsible of the fruit masting behavior of some species by “making mast–seeding years rare and unpredictable, which would greatly help in controlling the dynamics of seed consumers”.

- Okay, this can be another citation for early flowering. but actually this could be more directly cited in talking about performance.

L101: remove “the”

Done (101.

L141 “precedes” Done (142.

L324: please give an example of such possible experiment

- come up with an example and add.

L325-326: the difference between performance and fitness here is not clear to me

- Yeah could probably just simplify this whole sentence

L332: “differences in reproductive and vegetative phenological responses to the environment’. Could you give an example with an appropriate reference?

- Cite the peach literature

Figure 1: to support the use of 1980 as breakpoint, cite Kharouba’s paper here, and also Methods S1

- add citation to figure

Figure 4 and elsewhere in the paper: the naming of the variable “Earlier flowering” seems incorrect. It should be “Early flowering” (absolute value).

- Will do this

Figure 4 and Figure S2: the interpretation of such figures is not straightforward for the pollination syndrome. What do positive estimates mean in that case? Wind-pollinated species show more positive FLSs than biotic-pollinated ones?

- this one is a little tricky, but could add text to the figure to indicate wind pollination on the right

Reviewer #2 (Comments to the Author)

In this study, Buanaiuto et al. examined the order of vegetative and reproductive phenology (Flower-leaf sequences, FLS): They reviewed hypotheses put forth to explain FLS and the predictions that can be generated from these hypotheses; They propose a quantitative estimate measured at the level of individualsto examine FLS, and used long-term data from Harvard Forest to test this framework. They found support for multiple hypothesis related to migration and community assembly. They demonstrated that quantitative intra-specific estimates of FLS are useful for robust hypotheses testing.

The manuscript is generally very well written. It addresses a reasonably important and timely topic. It offers a novel view on a phenological metric that has not received much attention. Estimates of FLS in a quantitative manner, at the level of individuals will likely help advance our understanding of the importance of FLS. I am just not entirely convinced that as a trait FLS is important enough to warrant much attention. As highlighted by the authors, this opinion is based on the fact that there is very little empirical evidence that demonstrates the functional significance of variation in FLS.

- Thanks

- We agree, with the critique, yet is it made its way into the phenology literature. Part of the reason we are proposing this framework is to to be give researchers the tools to more appropriately test for such evidence
- As of now, not planning on changing the manuscript to address this in anyway.

While well written, I do think that the manuscript can be made a lot more succinct and shortened considerably. This manuscript seems a lot lengthier than other viewpoints I have read in New Phytologist. We have revised the structure and content of our manuscript , balancing a stronger focus on clarity and brevity while trying to maintain the essential elements of the text. We have shortened this main text to XXX (from YYY in the original submission).

find that there are a few other points that need to be considered to make the hypotheses that have been summarized to explain FLS more complete:

1) The authors need to explicitly consider relationships between flowering and fruiting. (see Primack 1987, Li et al. 2016, Ettinger et al. 2018). For e.g., flowering timing may be driven by selection on fruit dispersal timing and fruit development times. Therefore, large-fruited species that require longer durations for fruit development, will flower early in the spring to have sufficient time for fruit maturation (Mungu.âRosas et al. 2011). In contrast, small fruited species may flower at anytime.

- We think this is a hypothesis for early flowering not FLS.
- We could go back, grab fruit size and look at the correlation. But rereading this makes me feel like the review might be satisfied with verbal contributions to the text.

3) Architectural and Developmental constraints: In terminally flowering species, the timing of flowering is intricately linked to leaf flushing (Borchert 1983, Diggle 1995, Diggle 1999, Rathcke and Lacey 1985). Trees with determinate growth have terminal meristems that differentiate into inflorescence after producing vegetative tissue. In contrast, terminal meristems in plants with indeterminate growth only produce leaves, and inflorescences are produced from lateral meristems, allowing simultaneous flowering and fruiting (Rathcke and Lacey 1985). The same is also true for ramiflorous and cauliflorous flowering species, where timing can be independent of leaf flushing times (Van Schaik et al. 1993).2) Based on a hypothesis of energetic efficiency, one may expect leaf flushing to be related to flowering and fruiting (Van schaik and Pfanne 2005, Primack 1987). Both flowering and fruiting are energetically expensive. Flowering and fruiting at times when leaves are newly matured and at maximum photosynthetic efficiency is energetically optimal as opposed to relying on stored reserves. Flowering before leafing would require the use of stored non-structural carbohydrates (NSC) from the previous growing season.

3) Architectural and Developmental constraints: In terminally flowering species, the timing of flowering is intricately linked to leaf flushing (Borchert 1983, Diggle 1995, Diggle 1999, Rathcke and Lacey 1985). Trees with determinate growth have terminal meristems that differentiate into inflorescence after producing vegetative tissue. In contrast, terminal meristems in plants with indeterminate growth only produce leaves, and inflorescences are produced from lateral meristems, allowing simultaneous flowering and fruiting (Rathcke and Lacey 1985). The same is also true for ramiflorous and cauliflorous flowering species, where timing can be independent of leaf flushing times (Van Schaik et al. 1993).

- For both of these we could add 1 paragrah section or a few sentences to the phylogeny some where that says something like

- *There are obviously energetic constraints and physiological constraints. But Savage found they might not be as strong as theory suggests*
- *These factors may constrain FLS, but are beyond the scope they aren't functional/adaptive hypotheses.*
- *However, our framework would make explorations of these factors better too.*

4) It is unclear why authors did not include all of the five hypotheses outlined in the recent paper by Gougherty (Gougherty 2018)? Specifically, they left out the hypotheses related to cold tolerance, seed mass and xylem anatomy

- *Like 1, the citation in Gougherty explain early flowering, not FLS.*
- *we can do a correlation between cold tolerance and early flowering, and seed mass without grabbing more data. Xylem anatomy is early to get. and make a table for the supplement.*

[It is important for authors to distinguish between temperate and tropical broadleaf deciduous species. The proximate cues and ultimate causes of leaf phenology in these two groups are very different. Currently, it is unclear if the focus of this study is on temperate winter deciduous species or general to both temperate and tropical dry deciduous species. While the title specifies temperate forests, other parts of the manuscript includes general references to deciduous trees (Line 21, 43, figure legends, etc.) that gives the impression that this study pertains to all deciduous woody species. Issues like the problems described with categorizing FLS related to budburst may be relevant for temperate species and not for tropical species which do not have large lags between bud formation, budburst, and open flowers or leafing.

- *add "temperate" before each mention of deciduousness*

Related to the point made above, the hypothesis related to water dynamics seems relevant for seasonally dry tropical forests where leaf flushing and flowering often peak during the dry season when water is limiting. It is unclear how this is relevant for temperate forests. One line of exploration that may be worthwhile to examine relationships between freezing tolerance, wood anatomy, and water dynamics. Water dynamics are related to xylem anatomy, which is related to freezing tolerance (cold tolerance) (Zanne et al. 2014). Relationships with water use may be driven by correlations with xylem anatomy and/or cold tolerance?

- *not really sure what to do here.*
- *could make a hypothesis that this water dynamics hypothesis translates has been repurposed for cold tolerance in the temperate zone and model that? At the very least test these correlations in the above mentioned table.*

Line 23: "FLSs are adaptive" - I am not convinced that there is unequivocal demonstration of this. There is some evidence to suggest that FLS might be adaptive.

As suggested, we tempered our phrasing in .

Line 27: "...concurrent support for multiple hypotheses reflects the complicated history of migration and community assembly". This is speculative, and the inferences of complicated history of migration and community assembly (line 263 onward) are very speculative.

- *True, depending on what other sorts of analyses I have to do, could drop this, or basically just say we find support for multiple hypotheses.*

Line 77: Should be "Fig.2". *Done.*

"Line 161: "...there should be a cost..". *Done in 161.*

Line 189: "Under the current framework, FLS categories are assigned at the species level". What authors are trying to point out is that FLS is determined from species mean leafing and flowering times. Note that one can determine (or quantify) FLS for individuals and still assign a species level estimate.

Paragraph beginning line 189: There are two separate points that are mixed up here. One, that measures of FLS should be done at the level of the individual. Note that one can still average these measures made at the level of individuals to get representative species estimates of FLS. Two, there is a need to understand intra-specific variation in FLS. Note that it is possible to do one and not the other.

It is important to point out that the current estimates/categorization of FLS from species mean leafing and flowering times can give very different results from estimates/categorization done at the level of individuals and averaged to species level values/categorization, especially for species that are asynchronous and have short duration of activity. In contrast species that are synchronous in flowering and leafing, and with relatively long durations of flowering/leafing species and individual level estimates may be similar. Additionally, in considering hypothesis that explain FLS, individual level estimates are more appropriate - selection is going to act at the level of individuals.

- *I'm not sure I following this. But it should like a reason to look below the species level and could include it there.*

Line 214: Authors combine two separate points here - estimates of FLS at multiple scales and quantitative estimation of FLS. I think these are both kept separate. E.g. one can do hierarchical modeling with multi scale categorical estimates of FLS.

- *Yes. This will fall out in the restructuring and clarificaion on methods*

Paragraph beginning line 263: This entire paragraph is very speculative. This can be shortened considerably.

- *Doesn't like the biogeography attempt. Ok. Again this can either be cut or replaces with an alternative analysis.*

*emph*Some of the references in the bibliography are incomplete: Niklas 1985, O'Keefe 2015, Robertson 1985.

- *fix these*

These should get covered in the restructuring

Reviewer #3 (Comments to the Author)

In their Viewpoint paper entitled "Reconciling competing hypotheses regarding flower-leaf sequences in temperate forests for fundamental and global change biology", the authors propose a new quantitative framework to describe flower-leaf sequence (FLS) variation and they illustrate how it could be used to better test competing hypotheses explaining their function. They also advocate for an intra-specific approach that would improve our understanding the fitness consequences of FLS variation. I really enjoyed reading this paper. The ideas introduced are new and well-illustrated. The authors also used a very impressive dataset from 23 tree species in the Harvard Forest to compare their new quantitative FLS framework to the categorical one.

- *Thanks*

I only have one general comment about the new framework for FLSs proposed by the authors. I found that the strongest point presented is the shift from categorical measurements to quantitative one, by recording the dates of phenological events and number of days between phenophases. And, to my opinion, the idea of “individual-level quantification” can be dissociated from this first point. For instance, phenological observations can monitor the “average” BBCH score of several individuals within a plot to get ‘population-level’ quantification, and this description of FLS would still be more informative than a categorical one (for the reasons explained l. 141-151). I agree with the authors that “continuous individual-level quantification” of FLS would be the best data one could analyse, but there is generally a trade-off in the amount of data we can get at the individual versus species levels. I also agree that the intra-specific FLS framework is an exciting new avenue for future FLS research. However, I question its usefulness to clarify the mechanisms underlying inter-specific variation in FLSs (as stated l. 223-233). I think the biological processes driving the intra- and inter-specific variation can highly diverge (it is at least the case for the microevolutionary and macroevolutionary processes). I like the way the ideas are presented in the abstract, and I think the authors should keep this presentation in the main text too (i.e. presenting the “continuous individual-level quantification” as an interesting prospect to address other questions).

- *Agreed. We'll break these ideas up more.*

37-39: the fact that relationship between individual phenological stages affect fitness is the base idea of many plant life-history models, which link individual developmental threshold models across life-stages. In trees, empirical studies have also investigated how leafout and leaf senescence events and their relationship constrain the length of the growing season in trees (Vitasse *et al.* *Func. Ecol.* 2010, Firmat *et al.* *J. Evol. Bio.* 2017). So, this sentence probably needs to be nuanced.

- *Remove the word recently and add citations*
- *add the thing about threshold models across life stages.*

51-52: Fig. 1 shows that this change is only visible (significant) for the “flower-first” species, maybe this should be highlighted here.

- *not sure what to do here. I could mention it..but its only one species so thats not much for precedent.*

emph210: Following my general comment, I wonder if the authors should advocate here for a “continuous individual-level quantification” if such quantitative measures of FLSs are needed across multiple taxonomic scales. I agree that it would be the perfect data to analyse in order to model the different levels of variation and their associated noise. However, how many studies would be able to get this amount of information? From database, we can generally obtain the occurrence date of phenological events at the scale of populations. If the authors want their new framework to be use it would be more careful to present the different degrees of precision toward which FLS measurements can evolve: 1- categorical “continuous”; 2- species-level “population-level”; 3- population-level “individual-level”. The points 1- and 2- may already provide more robust test of the hypotheses for FLS variation.

- *new flo chart addition and talk about the possibilities.*

223-233: I am very confused by this entire paragraph. From my point of view, it would be most interesting to understanding whether the same/different biological processes explaining FLS variation at inter- and intra-specific levels. But I do not think that the comparison of the inter- and intra-specific variation in FLSs could inform about the FLS hypotheses researchers may want to test. Only the intra-specific variation in FLSs can inform about future FLS shifts (+ other parameters like the degree of gene flow among sub-species, hybridisation events etc). Please clarify your ideas about what an intra-specific FLS framework brings.

- *This paragraph will be scrapped. But I like the point that studying FLS at differnt levels can tell you different things about the trait. Should use this.*

Check Fig. S3 caption: “the” repeated several times in a few sentences.

- *fix this*