

Reviewer comments are in italics. Author responses are in plain text.

Comments from the Editor

I write you in regard to your submitted manuscript NPH-V-2020-32938 “Reconciling competing hypotheses regarding flower-leaf sequences in temperate forests for fundamental and global change biology”, which has now been reviewed. We found three reviewers and they all saw merit in your manuscript. While all reviewers were quite enthusiastic about the topic and your impressive dataset, they point towards several aspects of the methodology that could use more information justifying the choices you have made and the potential effects these choices might have on your results. Finally, each reviewer also has some further, more specific concerns to convey (see below), including a suggestion for shortening the manuscript. Although resolving these issues will require some careful reworking and potentially some reanalysis, I think you should be able to handle these suggestions through revision and, ultimately, this manuscript will make a valuable contribution to the literature.

We thank the editor for the work and attention they have already given to our manuscript and for providing us the opportunity to improve the manuscript and resubmit it to *New Phytologist*. We are pleased that all the reviewers saw the value of our work, and have worked hard to address their concerns and improve our manuscript based on their feedback. To do this, we have restructured the manuscript to clarify the methodology behind and utility of our framework. We have also included several additional analyses in both the main text and Supporting Information to evaluate the sensitivity of our framework to the various modeling choices we made and provided several alternatives for comparison (our findings are robust to these alternative specifications). We have done our best to resolve all the issues raised by the reviewers which we present in detail below.

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.

We thank the reviewer for their interest in the topic and find their concerns and suggestions helpful towards improving our manuscript. To address the reviewer’s concerns we have re-worked the main text and added new analyses to the main manuscript and Supporting Information. We provide more details regarding the changes below.

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 thank the reviewer for this important point, and we agree that minimum precipitation across a range is an imperfect functional trait to represent the water limitation hypothesis for the reasons the reviewer discussed. For this analysis, we tried to balance using functional traits that covered the broadest taxonomic range while choosing traits that best encompassed the hypotheses. We explored several alternative measures and performed four additional analyses to address the reviewer concerns.

First, we explored the possibility of using other functional traits, many of the them suggested by the reviewer including ψ_{50} , and ψ_{50} safety margins, from the TRY database and two large datasets from Bartlett *et al.* (2012) and Choat *et al.* (2012). Even with these large databases, the coverage for our species was very low, and would dramatically reduce the species n for our models, as can be seen in the table below (which we are also happy to include in the supplement if requested).

Overlapping species entries	MTSV	USFS	HF
Original	147	81	23
Choate 2012	29	22	7
Bartlett 2012	7	7	5
TRY	34	26	7

Despite limited available data, we agree with reviewer’s comments and have included results from an alternative model with two different drought-related functional traits, species’ moisture use and because of the high correlation between drought and cold tolerance (Zanne *et al.*, 2013), minimum temperature across a species’ range. It should also be noted that moisture use and minimum precipitation were the two functional traits used from the NCS database to represent the “water dynamics” hypothesis in Gougherty & Gougherty (2018), that were found to be significant drivers of FLS variation. We have plotted these results alongside the main text models in the Supporting Information (Fig. S5). As can be seen the alternative predictor did not change the results of our analyses substantially. Since “moisture use” is a categorical trait we have decided, in keeping with our focus on quantitative measurements, to use minimum precipitation as the predictor we use in the main text models, though we mention the alternative predictors and their results in the main text (with extended results in the Supporting Information).

We also appreciated the reviewer’s suggestion to explore common-use drought indices in our analyses. Since these indices are usually determined at the landscape scale, given our available data, their use may be most appropriate for exploring their role in inter-annual variation in FLSs. We performed an additional analysis investigating the relationship between annual P-ETP and FLS variation calculated from flux tower data. While we found no significance to this relationship, even this result is a useful metric for evaluating the relevance of water limitation hypothesis in the temperate zone and we have added this analysis to an additional section “The water limitation hypothesis in wet temperate forests” in the Supporting Information.

Finally, we agree that there are more precise ways to describe the “water dynamics hypothesis”. We used this term because it has been used in previous publications (Gougherty & Gougherty, 2018). As per the reviewer’s suggestion we changed our description of the hypothesis to the “water limitation hypothesis” throughout the manuscript.

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. I'm wondering if 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.

The reviewer makes an excellent point here. For our analysis using PEP725 data, we chose the flowering and leafing sub-phases that had the most available data and, as the reviewer noted, the Harvard Forest dataset uses alternative phenophases descriptions that do not perfectly correspond to established BBCH stages. To clarify how our Harvard Forest analyses should be considered in relation to the BBCH scale, we have provided a table with our best approximations for the relationship between Harvard Forest phenophases and the BBCH scale based on the publicly available meta-data for this dataset (see Tab. S1).

Additionally, for our main analysis of the Harvard Forest data we now present models with the more recognizable phenophases “leaf budburst” and “flower budburst” in the main text (see Fig. 4), and provided a sensitivity analysis comparing these results to model with FLSs calculated from the time of “flowers open” to “leaves 75% final size” (see Fig. S7). Interestingly, we found that estimates from our intra-specific hierarchical model were the most robust to these changes and have added a discussion of this topic explicitly in the main text of the manuscript in line 282-288. We thank the reviewer for this suggestion, which we believe provides greater insight into the benefit of this new framework.

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.

We agree with the reviewer that “intra-specific analyses” is an ambiguous descriptor and that, in the main text, it wasn't clear enough how we incorporated these levels of organization into our modeling framework. For our analyses of the Harvard Forest data, we used hierarchical models to account for variation among individuals, and within individuals over time. We have now improved the Methods section in our Supporting Information to clarify these methods.

To further address the role of individual variation in our analyses, we now also now provide an additional analysis in which we compare a quantitative model with the mean trait values for all species to hierarchical models accounting for inter-individual and inter-annual variation in FLSs and predictors for which data was available at these levels (see Fig. 4).

We also appreciate the reviewer's observation that the multiple sources of intra-specific variation within a species could be more clearly addressed in our manuscript. We have changed our main text to more explicitly discuss the implication of variation at the individual and population level in line 306-310. We also agree that population level variation in particular is an interesting application that could be well tested with this framework. Our main source of data came from a single population, so a detailed analysis of population level

variation is difficult in this paper. We have, however, made sure to include a more explicit treatment of this level of variation in the aforementioned lines, and created a new figure depicting differences in the mean and variation of FLS's from one species in the PEP725 database (see Supporting Information Fig. S8) to provide an example of the kind of population level data that may be leveraged in future analyses.

L39: "their chronology" instead of "the relationships between them"

Fixed in line 33.

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

Fixed in line 42.

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 hysteranthly.

We agree that a quantitative assessment would be helpful. We have added estimates from previous work on FLS (Gougherty & Gougherty, 2018) in line 43.

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)

We appreciate this suggestion and have added boxplots to the figure.

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

We have adjusted the language for clarity in line 55-58.

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

We appreciate the reviewer pointing us to this reference and have added it to line 81.

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".

We thank the reviewer for pointing us to this reference. We have added it to our discussion of the role of dispersal events in the early flowering hypothesis in line 92 and our discussion of the link between inter-annual variability and performance in line 237.

L101: remove “the”

We have made this change. (Line 97)

L141 “precedes”

We have made this change. (Line 151)

L324: please give an example of such possible experiment

We now provide a conceptual example of an experiment in line 318.

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

We have removed this line of discussion from the current version of the manuscript.

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

We have simplified our discussion of environment, physiology and FLS in line 310, stating simply that: “Variation among and within individuals provides insights regarding micro-climate effects, heritability, selection and plasticity for FLSs”.

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

We have added this citation to the figure.

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

We agree with the reviewer and have made these changes throughout the manuscript and figures.

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?

We thank the reviewer for pointing this out—their interpretation is correct. For clarity, we have annotated these figures (Figures 4, S2 and S7) to indicate which factor of a given predictor is associated with increasing time between flowering and leafing.

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 individuals to 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.

We appreciate the reviewer's assessment of the implications of our manuscript and share their concern that the functional significance FLSs may not compare to that of other phenological stages. Yet, this idea that they are important pervades the literature (e.g. Rathcke & Lacey, 1985; Gougherty & Gougherty, 2018) though, as the reviewer points out, there is little evidence for this.

At its core, the purpose of this manuscript and our proposed framework is to give researchers the tools to robustly evaluate the functional significance of FLS variation. We feel there is enough theory suggesting FLS variation may be important in at least some species that its significance cannot be dismissed before the adequate evidence is gathered and we have added to the manuscript a statement of this uncertainty in line 239.

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 thank the reviewer for this feedback. While our original submission was well below the 6,500 word maximum for Viewpoint articles, we have made efforts to condense and streamline our revised manuscript further while still incorporating the important additional analyses and points requested by several reviewers. We have reduced the figures in the main text to four and shortened the main text from approximately 3,500 words in the original submission to approximately 3,300. We believe the current submission balances length with clarity of message, though we can work to condense further if requested.

I 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:

We have attempted to incorporate this reviewer's points into both our discussion of the FLS hypotheses and several supplemental analyses that can be found in the Supporting Information. Our adjustments to the manuscript based on each specific point can be found below.

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 (Munigua-Rosas et al. 2011). In contrast, small fruited species may flower at anytime.

We agree with the reviewer and the citations provided that fruiting characteristics are a well established

proximate cause for selection on early flowering. We did not originally include these factors in our discussion as we considered them to be drivers of early flowering but not mechanistically related to FLS variation. However, we now agree with the reviewer that a more detailed treatment of these proximate drivers of early flowering would enrich our discussion of this hypothesis.

We have thus added explicit references to these drivers in lines 91-94, as well as incorporated them into alternative models using dispersal time or seed mass as a proxy for the early flowering hypothesis and included them in the Supporting Information (see Fig. S5). We found that using these predictors did not strongly affect our model inference but increased the uncertainty around the model estimates.

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.

This is an important point that we did not state explicitly enough in our original submission though we believe it highlights a fundamental FLS tradeoff, and explains why not all species would evolve a flowering-first FLS. We have added text to address this in line 45.

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).

We agree with the reviewer that physical constraints are an important part of any discussion about the evolution of any trait including FLSs, and that adaptive hypotheses are only part of the picture. Because of our interest in FLSs, performance and climate change, we focused on these adaptive hypotheses, but physical constraints, like phylogeny, are no doubt important drivers of the diversity of FLS patterns in the temperate zone. We have expanded our discussion of phylogenetic constraints to review constraints more broadly—including physical constraints—based on the references suggested by the reviewer, in line 100-113.

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 fruit characteristics, these three hypotheses tend to appear in the literature as proximate drivers of early flowering rather than FLS variation and therefore should be related closely to flowering time. As seen in the table below, correlations between these predictors were variable and generally low; we have included additional analyses to model the effects of seed characteristics (fruit development and seed mass) and cold tolerance in the Supporting Information. Gougherty & Gougherty (2018) found little support for the xylem anatomy hypothesis, but we felt that this hypothesis should be considered a “physical constraint” and have added it to that section of the main text, see line 114.

	data	fruit.development	seed.mass	xylem_anatomy	cold.tol
1	MTSV	0.303	-0.008	-0.026	0.180
2	HF	0.295	0.067	-0.154	0.622

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.

This is a good point. As our paper is focused on temperate deciduous species, we have clarified this throughout the manuscript by adding the modifier “temperate” to any references to deciduous woody plants.

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?

We fully agree with the reviewer that water limitation is an unlikely driver of FLS variation in the temperate zone. However, this hypothesis found support in Gougherty & Gougherty (2018), which is one of the most detailed analyses of FLSs to date, suggesting that it merits being grappled with. We have added text to emphasize that this hypothesis may need further development to be applied to temperate species in line 86 and a more detailed treatment of this development in the Supporting Information in a new section called “The water limitation hypothesis in wet temperate forests”.

We think the suggestion that there may be a relationship between FLSs and water limitation in the dry-tropics and FLSs and cold tolerance in the temperate zone is a good one. We found there to be a high correlation between our main metric for water limitation (minimum precipitation across range) and metrics of cold tolerance (minimum temperature across range). We have included an additional model substituting cold tolerance for water limitation in the Supporting Information (Fig. S5), and discussed this connection explicitly in our new added section “The water limitation hypothesis in wet temperate forests”.

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.

We have adjusted this language accordingly.

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.

We have adjusted these lines to limit our speculation into what is no doubt a much more complicated bio-geographic pattern that we originally portrayed. This line (line 278) now reads simply, “Together, these systematic differences suggest that flowering-first FLSs in these functional groups may have evolved under radically different environments and converged in temperate forests.”

Line 77: Should be "Fig.2".

We have fixed this typo.

Line 161: "...there should be a cost..".

We have fixed this typo in line 170.

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.

As the reviewer points out, we agree it was a vague description of how FLS patterns are generally described. To their point, we have provided an example of this method (to “determine (or quantify) FLS for individuals and still assign a species level estimate”) that can be found in Fig. 4b).

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.

This is a very important point that was highlighted by all three reviewers. As such, we have changed the structure of our manuscript to treat these two aspects of our framework separately. Each component of the framework now has its own subsection in the main text.

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.

We thank the reviewer for stressing this point and agree with their assessment. To address this we have added analyses to our main text comparing a quantitative FLS model based on species means to a hierarchical model accounting for individual variation in FLS in Fig. 4 b) and c).

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.

We agree. We now draw stronger distinctions between the utility of quantitative measures of FLS and multi-scale FLS observations throughout the manuscript.

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

We appreciate the reviewer's suggestion to refrain from straying toward speculation. As mentioned above we have limited our speculation on this topic in line 278.

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

We appreciate the reviewer calling our attention to these references and have filled them in accordingly.

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.

We thank the reviewer for their positive comments about our manuscript and are glad they found it enjoyable to read.

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).

We appreciate this feedback from the reviewer, and as we mentioned above, we agree that continuous measures of FLSs, and measures below the species level are two separate components of our framework that provide different benefits on their own.

To address the reviewer's general suggestion for a closer adherence to the structure of the abstract in the main text we now break our presentation of our framework into two subsections; first discussing the benefits of quantitative measures and second how observations below the species level allows for novel research avenues regarding FLSs. As the reviewer also pointed out, we agree that inter- and intra- specific may offer unique insights towards a more complete understanding of FLS variation, but may not relate to the evolutionary mechanisms driving FLS variation at different scale. In our new version of the manuscript, we no longer suggest that intra-specific observations can serve to clarify inter-specific mechanisms (formerly line 223-233), and rather discuss these scales separately throughout the manuscript.

l. 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.

We appreciate the reviewer highlighting that the importance of phenological sequences has been acknowledged for a longer time than our language suggested. We have removed the temporal qualifier from this statement in line 33 and added the citations suggested by the reviewer. It now reads: "It is not only individual phenological stages that affect these processes, but also their chronology."

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

We agree with the reviewer that this is an interesting point and is worth highlighting. In our revised manuscript we call attention to this pattern in line 51.

l. 210: 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 to continuous; 2- species-level to population-level; 3- population-level to individual-level. The points 1- and 2- may already provide more robust test of the hypotheses for FLS variation.

As mentioned before we have taken the reviewer's suggestion to separate our arguments about the merits of quantitative measures of FLSs and the values of intra-specific data and re-written the manuscript around this structure.

l. 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.

The reviewer's point is well taken here. We have attempted to highlight the value of intra-specific observations more clearly by devoting a subsection of the paper to this topic beginning in line 301.

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

We have fixed these typos and checked for repeated words throughout the manuscript.

References

- Bartlett MK, Scoffoni C , Sack L. 2012.** The determinants of leaf turgor loss point and prediction of drought tolerance of species and biomes: a global meta-analysis. *Ecology Letters*, **15**: 393–405.
- Choat B, Jansen S, Brodribb TJ, Cochard H, Delzon S, Bhaskar R, Bucci SJ, Feild TS, Gleason SM, Hacke UG *et al.* 2012.** Global convergence in the vulnerability of forests to drought. *Nature*, **491**: 752–755.
- 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.
- Rathcke B , Lacey EP. 1985.** Phenological patterns of terrestrial plants. *Annual Review of Ecology and Systematics*, **16**: 179–214.
- Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB *et al.* 2013.** Three keys to the radiation of angiosperms into freezing environments. *Nature*, **506**: 89–92.