Reviewer comments are italicized and our detailed explanation of changes are in regular text.

Handling Editor's comments:

Comments for Authors: I appreciate the authors' efforts to address the reviewer comments by undertaking substantial revisions. This version is improved in many ways. However, the review process has identified some lingering issues. Most of these issues relate to a lack of clarity in the methods and confusion around how the results have been interpreted, concerns that are similar to those expressed in the first round of review. As a side note, I agree with Reviewer 2 that the frequent use of em dashes in the text is distracting and reduces clarity. Please consider splitting the content into multiple sentences, offsetting phrases with commas, or otherwise editing to reduce the reliance on em dashes.

We are glad that our previous edits addressed many of the reviewers comments and improved the manuscript. To address the lingering issues, we have made additional revisions throughout the text, including extensive revisions to the methods and the results. We have also revised our sentence structure to remove em dashes throughout the text to improve clarity.

Reviewer 1 – comments:

Thank you very much for the extensive revision of the paper.

We thank the reviewer for acknowledging the extent of the revisions and are happy to see they have addressed most of the reviewer comments.

Minor Comments:

L241: I would recommend to write "Of the studied leaft traits" instead of "our leaf traits".

Done (line 240).

Figure 1: Why are the first bullet points in each text box written in italic?

Our aim was to highlight the difference between our predictions pertaining to cues from our predictions about triats. But we see how this could cause confusion and have made all predictions plain text.

Reviewer 2 – comments

I like the idea and the approach of the study and further think it is an interesting topic with current relevance. Reading the manuscript, I had some trouble understanding the main results, primarily due to a lacking definition of what a large/small response to a cue means and contradicting sentences that confused me. The Discussion would benefit from an additional subsection or a restructuring to facilitate the reading flow and logica of the text. Please find more information in the attached document where major and minor issues are addressed in detail.

We are happy to hear the reviewer believes the topic is interesting and current and that they like approach used. We agree with the reviewer that would could be more explicit in how we define larege/small responses to cues. In revising the text in the results (line **??** and line **??**), we believe we have provided additional information that improves the clarity of our results

and addresses any lingering confusion. We have also made extensive changes to the discussion to improve the flow and logic (line **??**).

General:

The authors compiled a large dataset containing phenological and other trait data for trees from three different databases to link responses in budburst with whole plant, leaf and seed functional traits. The overall aim is to potentially predict phenological change in budburst based on other more easily measurable traits (SLA, plant height, seed mass, leaf N content), by placing the responses of plants to environmental cues such as forcing, chilling and photoperiod in the context of a functional trait framework. The authors found that the budburst response was linked to traits being related to nutrient acquisition strategies, with early budbursting species being linked to an acquisitive strategy and late budbursting species to a rather conservative strategy. I like the idea of the study and think the dataset is a suitable approach that advances the current understanding of budburst response to climate change. Therefore, I also believe that the study meets the aims and scope of the Journal of Ecology and may be suitable for publication after addressing certain parts and improving unclarities in the manuscript.
We thank the reviewer for this positive feedback and are happy to hear our manuscript would be suitable for publication in the Journal of Ecology.

Major comments:

It seems necessary to me to state clearly that the study focusses on trees only. Regarding references and contextualisation, it needs to be clearly differentiated between herb and tree traits (incl. phenology) because they behave differently (see e.g., supplementary of Díaz et al. 2016 where it is directly compared). Sometimes, these two very different growth forms, and the growth strategies that accompany them, seem to be seen as equal here.

We agree with the reviewer that we could be clearer in differentiating work on herbaceous plants from trees and now do so on line 72, line **??**, and line **??**.
The study needs more clarification of the Methods and Results for readers that are not familiar with the statistics used here, mainly regarding the modelling part, and a more thorough explanation of the variables used to infer phenology-trait relationships. Until the end I found it hard to understand what a large response to a cue actually means here. Traditionally, I understand a large response as high sensitivity to a cue but certain sentences throughout the manuscript imply the opposite, which is confusing and makes an understanding of the main findings hard. For example in L299 "Species with small cue responses, an indication of earlier budburst, ..." or in L332 "with smaller responses to all cues, especially chilling and photoperiod, [species] would tend to advance more with warming (Guy, 2014). Our results suggest that these same species are likely to have acquisitive traits". A definition needs to be stated clear and simple in the Methods or Results to be able to follow later on. E.g., "a strong response /large coefficient to a cue indicates that budburst happens later/earlier. Consequently, a weak response (towards zero) indicates...".

We agree the results of this study are complicated and have revised the text to help with their clarity.
The difference between response and sensitivity?
Minor comments:

Summary

This may be preference, but I would suggest replacing the dashes by commas so that misunderstandings are avoided (e.g. in L7: "..., such as temperature-changes..." vs. "..., such as temperature, changes...").

We have revised the text to remove em dashes throughout, including on line **??**, line **??**, and line **??**.

L22: I suggest to replace "higher nitrogen leaves" by "higher leaf nitrogen content"

Done (R2$_{l}nc$)

Introduction

L40: remove opening bracket before e.g. and shift to before the reference

Done (line 40)

L44: unclear what is meant with architectures. Tree architecture, morphology, branching? The cited reference (Flynn & Wolkovich, 2018) compares trees and shrubs, i.e., growth form, but does not mention architecture explicitly, so please be more explicit here.

We agree that using the same language as Flynn & Wolkovich (2018) would improve clarity and have revised the text to now refer to "growth form" on line 44.

L59: Sporbert et al. (2022) refers to herbaceous species which should be added because they have different phenology and other trait strategies, compared to trees

We have revised the text one line 72 to specify which references refer to woody versus herbaceous species.
L71: remove comma before opening bracket

Done (line 71)
L72: Rauschkolb et al. (2024); Sporbert et al. (2022) use data from Botanical gardens but focus on herbaceous species. They behave different to tree species, so it needs to be added somewhere that their studies focus on herbs.

R2$_{t}reeHerb$1
L73: whether the problem of proximate drivers causing phenological variation can be more easily ignored depends on not only on a limited number of sites but rather on the environmental conditions these sites reflect. Consider a semi-dry grassland, a mesophilic grassland, and an urban park, all sharing some species and all being spatially close to each other but phenology and other traits, as well as their relationship may differ strongly. In other words, intraspecific variation is strongly habitat-specific and can even exceed interspecific variation.

L92: What exactly is meant by "strong gradients in frost risk or nutrient availability in spring"? Do you mean compared to the whole year or within spring or between sites in spring? Fig. 1 refers to a gradient in growth strategies based on traits, so please explain what you mean by gradient in the text. Also are there any references backing up Fig. 1?

L96: Bucher & Rosbakh, 2021 also focused on herbaceous plants only which, again, behave

different to trees. This has to be mentioned here.

L105: LNC not spelled out in the text before. Would be possible to do in L99.
Consider switching paragraph 4 and 6 to improve reading flow and logical structure as they seem to be topic wise related.

Methods

L124-136 seem irrelevant for the present study. Already described in Ettinger et al., 2020. Rather briefly describe the data OSPREE contains and link to the original publication (Wolkovich et al., 2019), as is done with TRY and BIEN later.

L144: Which data did you use? If the updated one then just mention this date maybe.

L150: Why do you think the DBH standard height is a good proxy for being an adult individual? In forest understories, many individuals reach this height without being adult, really, so I think this needs some more justification. Otherwise you could run into the issue of comparing rejuvenation and adult trees, likely affecting your results.

L165: Fig S2 mentioned after Fig S3 (L157).

L170: Subsampling of height to reduce the influence of most frequently measured species on the model. What about the other traits? Was no subsampling needed here because species were measured more or less equally?

L190: How do you obtain measurement error variance?

L191: What is N?

L193: What is meant by the T in $..^{T}$? Total? Does the (4) refer to one of the formulas?

L194: How do you obtain alpha(grand trait) as being independent of species and study-level offsets from that trait value?

L216: s. L190

Results

Fig. 2: What does a large and a small response to a cue, respectively, mean? This is needed to explain thoroughly as it is the basis for further discussion. Does a large response mean, that budburst happens way earlier or does it mean that a low cue intensity is needed to trigger budburst? Please add a simple explanatory sentence to get this right as reader.

L244-246: indication of SLA and relationship to photoperiod rather fits to Discussion

L250-252: move to Discussion.

L256: add space before "For height"

L254-267: variation =/= variance. What you show is variance as derived from your model. Variance is just one measure of variation. Please stay consistent here.

L264: use directly $\sigma 2$study or delete bracket as it was mentionend in the beginning of the paragraph

L266: s. L264

L265: move to Discussion.

Discussion

L273: "the trait effects of height and LNC were associated with earlier or later phenology". Which one is related to earlier, which one to later phenology? This is formulated a bit confusing here.

L275: I miss a discrimination between early and late budbursting species here or in the Results. From the text only, I understand that the ones responding strongly to the cues are the late budbursting species and oppositely for the early ones. I am not sure, though, if this is a general assumption based on literature or if this is derived from the results. If it is part of the results, what is the threshold to classify early vs late species? An overview of which species is considered early and late and an average of their relationships with traits associated to the different growth strategies would improve the understanding of this part. Also I think that some references to back up the these relationships of budbursting timing with cue intensity are useful.

L285: The study is definitely more global than local scale studies but since basically South America, Africa, and Australia is missing, I would rather call it intercontinental or large scale.

L295: Have you checked whether this isn't done already? Usually study site is included as random effect in modelling, so this should account for study-site related variation, right? Also, what does this imply for interspecific variation? That interspecific differences are stronger than trait differences between sites/forest stands? Could this also be a result of different species occurring at different study sites or did you consider overlapping species only?

L300: shorter heights not for response to forcing. Any idea why the relationship is different for this cue?

L314: Any examples for the multiple aspects of species growth and adaptations from the literature? What about intraspecific variation?

L315: the response to forcing may be large but according to Table S2 it is a weak relationship, which you also state in your Results. I don't quite understand why this point is being picked up here as it was a "significant" relationship.

L315: If a large response means high sensitivity to cues, then a large response for short trees, i.e., fast budburst when spring temperature rises only a bit, would rather increase the change of frost damage under a late spring frost event, right? If this is correct, then the large response could hint to a phenological adaptation of understory trees (since you included trees $>$ 1.38 m height) to light variability throughout the year.

L316: SLA is an adaptation to competition for light as leaf area increases with high SLA to

capture more photons per area leaf. That way, it is not suprising that you find the strongest relationship with photoperiod only. However, I would have expected as well that photoperiod and temperature are quite closely related, so you could discuss towards this direction.

L327: I think you should mention that Macgregor et al, 2019 focus on Lepidoptera (butterflies s.l.), and not on plants which can be misleading. I suggest you stick to relationships of phenological change and performance in woody plants here.

L331: Forest species or forest understory species would be more accurate here.

L332: Temperature and frost may change under future climate but what about photoperiod? Isn't this a rather restricting cue in a sense that it does not change substantially in future but when species are sensitive to forcing and to photoperiod at the same time, then it can become a phenological issue when spring temperature advances but photoperiod stays the same. This could be discussed as well here in terms of growth strategies. The referenced Guy (2014) refers to some climax species being phenologically inflexible which could result in a disadvantage under future climate conditions.

L341: I agree. Can you give some specific examples how exactly this could help management decisions?

L358: If we know which species are most vulnerable because they fail to adapt phenologically, how would you prevent their loss in communities by being outcompeted by better adapted species? I think "Develop more effective management practices" needs some further explanation here.

General comments to discussion:

I feel that the last three paragraphs could be summarised in a "conclusion and outlook part" or in an "application part" as they aim to improve management decisions and practices.

Do you have any explanation why SLA and LNC behave so differently in predicting response to photoperiod, even though they tend to be highly correlated traits? The estimated model fits and raw data do not work as well for tree height as for the other traits. Why? What are the consequences for inference? The data used here stem from forests, so I miss some habitat specific discussion. It is discussed rather generally but especially the cue photoperiod is particularly important in deciduous forest understories and successions and may be different in coniferous forests, shrub- and grasslands, or other habitats.

Tables

Table S2-5: From the caption, I take that the 50% and 90% UI are provided but what is shown are values for 5% to 95% of the posterior distribution. This is confusing and would benefit from some more explanation for readers that are not familiar with Bayesian approaches, i.e., add some explanation that 5-95% equals the 90% UI and 25-75% the 50% UI.

Figures

Fig. 2 shows the estimated species-level response with the 50% UI but it is stated in the Methods in L226 that you show the 90% UI, and the 50% UI in the supplementary. This is confusing. Furthermore, maybe add log10 to the x axis of Fig. 2 for seed mass as at the first

glimpse a negative seed mass is confusing.

Also, if the relationship is weak or crosses zero, then I strongly encourage to also show this somehow in the figure by e.g., dashed lines or removal of the response line. Otherwise, it is just confusing and the tables in the supplementary need to be studied carefully and compared to Fig. 2 for each trait.

To read this figure properly, I would appreciate some annotation in the plot stating what a small or large response mean, e.g., earlier budburst, later budburst.

In Fig. S3, the number of unique traits is different (8 and 11, instead of 10 and 13 as stated in the text). Please correct where needed.

Fig. S5: I find it difficult to understand this figure. How do I assess the effect size of a trait on budburst? When I understand correctly, then it related to the 50% UI of the full model vs. trait effect = 0. What does trait effect = 0 mean, though? I think, some clarification in the caption or in the Methods is beneficial to correctly read the figure.