

Scheiter et al. delve into the utilization of crowd-sourced trait data for defining global biomes, investigating the suitability of trait data for biome classification and identifying the most relevant traits for this purpose. While the study is intriguing and significant, I have some comments to offer.

Thank you for the helpful and constructive comments. As suggested by the reviewer, we will add additional analyses to show the relation between traits and biomes and emphasize the ecological interpretation. We will further move analyses on different PFTs and trait subsets to the supplementary materials such that the main text is more focused. We will also include a schematic figure to illustrate the steps of our analysis. All details are provided in our responses to the comments.

Major comments:

- *The findings presented in this study offer valuable insights into the fundamental ecological mechanisms driving plant biogeography and global distribution strategies. However, while the results are intriguing and significant, the interpretation from an ecological perspective could be further emphasized. Additionally, it would be beneficial for the authors to underscore the importance of their study and the implications of their findings, particularly in informing our understanding of global biome classification and its ecological implications. Strengthening the ecological context and emphasizing the practical implications of the findings would significantly enhance the impact and relevance of the study.*

We agree that the ecological interpretation of the results can be strengthened. Currently, the manuscript focuses on the methodological aspects and how the method employed here may be used to generate new ecological findings. We will revise the manuscript with a stronger ecological perspective. Specifically, we will analyze co-variation of traits across biomes in more detail, for example by including plots showing the position of different biomes in the trait space. Therefore, we will conduct a principle component analysis and, based on the loadings of traits and their ranking, illustrate the locations of biomes in the 2-dimensional trait space of selected traits. Such plots will allow us to identify which trait combinations are characteristic for different biomes. We will also revise the discussion to account for these analyses and explain patterns of trait covariation (for example, co-occurrence of tall trees with deep roots and low SLA in tropical forests, indicating intense light competition and the potential for rooting niche separation). We will further underscore the relevance of our results for biome classification by stating more explicitly which traits should be used for classification.

- *The findings regarding the differences in trait distribution and occurrence between biomes are indeed intriguing, but they have been explored somewhat superficially in the manuscript. While I understand that this topic is complex, questions 2 and 4 require a more in-depth exploration throughout the manuscript. As a reader, I was expecting these interesting questions to be thoroughly addressed. For instance, the observed tendency of tropical forests to have plants with greater height could be explored in more detail in the discussion section. One possible explanation for this phenomenon could be light competition.*

Following the suggestion of the reviewer, we will revise the discussion to refer more explicitly to the questions in the introduction. Questions 2 and 4 ask about which and how many traits are required for classification and about characteristic trait values per biome. To deal with question 2, we will strengthen the discussion on traits that turned out to be important in our analysis. These traits agree with previous results on the leaf/plant economic spectrum as well as with previous modeling result that link traits and PFTs (Verheijen et al. 2016). To deal with question 4, we will include additional analysis and figures showing the relation between biomes and traits in the trait space. We will discuss the patterns and explain why different trait combinations co-occur (see also previous point).

- *It would be valuable to delve deeper into why the identified distinguishing traits are significant and what they represent in terms of plant strategies and ecological functions. Hence, you can offer insights into their ecological relevance and their role in shaping biome characteristics. Specifically, you could discuss how these traits contribute to plant adaptation to specific environmental conditions, resource acquisition strategies, and ecosystem functioning. While you have begun to address this topic, emphasizing the ecological significance of the identified traits and highlighting avenues for future research would strengthen the overall argument and underscore the importance of your findings.*

As suggested, we will analyze ecological strategies and co-occurrence of traits in different biomes in more detail in the revised version. Therefore, we will plot the location of biomes in trait space obtained by a principal component analysis. This analysis will illustrate which traits are most relevant to explain the occurrence of different biomes. We will interpret these additional results in the discussion and explain mechanisms causing such patterns (e.g., light competition in dense forests, or water competition, or tradeoffs between acquisitive vs conservative strategies).

- *It's essential to clarify the methodology, especially in section 2.4, to improve understanding for readers.*

We agree that the reasons for conducting this analysis need to be described better. We think that this analysis adds an important aspect to our study, and we will therefore keep it in the revised version of the ms. We will (1) better motivate the analysis, (2) use trait subsets with higher performance ($\kappa > 0.6$ in the sensitivity analysis in 2.3, following a comment by reviewer 1), and (3) move Fig. 4 to the appendix and replace it by a novel analysis of the relation between traits and biomes. We will further use this trait set with higher kappa for analyses in 2.5 and 2.6.

- *The conclusion and a significant portion of the discussion focusing on models may seem disconnected from the main purpose of the paper as outlined in the abstract and manuscript title. As a reader, I felt somewhat puzzled. Notably, none of the four questions posed at the end of the introduction mention the theme of modeling. To mitigate this discrepancy, you could consider clarifying from the outset that modeling will be a significant aspect of the study. Alternatively, you may choose to shift the focus of the discussion to emphasize other aspects that were highlighted earlier in the manuscript, aligning more closely with the stated objectives and questions.*

Given that the cluster analyses applied in the study are models, we assume that the comment regarding “none of the four questions ... mention ... modeling” refers to DGVMs and previous results with aDGVM2 mentioned in the discussion. We agree that this is not the main topic of the manuscript and not reflected in one of the questions. Nonetheless, we found it relevant to compare our results with previous results, both from the analysis of observational data and from modeling.

We will shorten the comparison with aDGVM2 results and make sure that these aspects are not disconnected from the analyses of our study. We will further discuss our results in the context of Verheijen et al. (2016) who used kernel estimation to link traits and PFT and found that similar traits are important. We think that the trait data derived by combining TRY and GBIF are valuable for improving DGVMs, but this is beyond the scope of the presented study and should be addressed in a follow-up study.

Minor comments:

In the abstract the use of the “31 different biome maps” is not clear, is it used to perform the supervised cluster analyses or it is used to evaluate the results from this analyses?

We will reword the abstract and make clear that the 31 observation-based biome maps are used to inform the supervised cluster analysis of traits into biomes and for the evaluation of the biome maps derived by the cluster analysis. When a single data set is used for model training and testing, it is common to split the data set into a subset for training and a subset for testing. However, as we do not use our models for extrapolations or for predictions with novel data, we argue that our approach is valid.

Enhancing the fluidity of the introduction by providing more ecological context and emphasizing the significance of generating this type of map would improve the transition to the research questions.

Many different biome maps are available and were developed using different approaches and data sources. We argue that traits are an additional data source that inform on biophysiological differences among plant communities that can be used generate novel biome maps. This functional trait perspective might provide a more plant-oriented and meaningful understanding and characterization of biomes than, for instance, a biome perspective that is solely based on climate or remotely sensed reflectance. To create such biome maps, we need to conduct two steps, (1) use traits for biome classification and (2) extrapolate from the spatially heterogeneous trait data to a biome map covering the entire land surface by making use of the biomes obtained from clustering, bioclimatic variables and species distribution models. We will better explain this context in the revision.

158 and 159 are more appropriate for Materials and Methods section

We think that the extrapolation to global maps should be kept in the introduction, and we will better motivate this step (see also previous comment). The method section describes how we did this.

It's not clear in the methods why you "created three different maps for each trait" and why you filtered the observations according to PFTs before spatially aggregating the trait values.

In line 78, the differentiation of the trait space per PFT might provide added value for biome classification, instead of having only information across all PFT. This step was done before aggregating trait values to be able to conduct the filtering based on the trait data from TRY and keeping as much of the information from TRY as possible.

In section 2.2 I didn't understand the difference between 3) and 4).

In case 3, we did not differentiate between woody and non-woody plants to obtain a community trait value per grid cell. In cases 1 and 2, trait values for a grid cell were calculated separately for non-woody and woody plants, resulting in two trait maps. In case 4, we combined the woody and non-woody trait maps in one biome prediction model. Hence, we include more detailed information on PFT-specific trait values.

We will better explain difference between 3 and 4 in the revision.

In section 2.3, specifying the total number of traits considered would enhance clarity. If I don't miss anything, it is only specified in the abstract.

By merging TRY and GBIF, data for 33 traits are available and included in the analysis. We will add this information in 2.3.

It is unclear why the analysis described in section 2.4 was performed if the most important traits were identified in section 2.3.

The analyses in 2.4 are conducted to make clear that the selection of traits for classification matters and that the selection of the biome map used to inform the cluster analysis matters. This analysis was based on the important traits identified by the sensitivity analysis in 2.3. We will clarify this in the revision and move the results into the supplementary materials.

Fig. 3: Parallel coordinates plots are already complex visualizations to comprehend, and as it stands, I don't think the figure adds much value, unless it is thoroughly explored and contextualized in the text.

This figure illustrates that the same set of traits is important (i.e., has a high rank) for all numbers of traits included in the cluster analysis. Similarly, the same set of traits shows low rank. We therefore think that this figure is important as it allows the selection of traits for subsequent analyses (2.4, 2.5, 2.6) using only a subset of traits. We will better explain the figure in the revision and provide context in the text.

Fig. 5: The background color (gray) on the map is too similar to the light blue, making it difficult to interpret the figure.

We will change the colors for clarity.

Why is there such an extensive comparison with the results from aDGVM2? This was not anticipated based on the main questions posed in the Introduction section.

We will shorten or entirely remove the section on aDGVM2 and therefore focus more on the ecological interpretation of the main results.

In Section 4.2, you briefly touch on the importance of examining trait covariation, highlighting its significance compared to focusing solely on individual traits. However, given that some of your results seem to directly address this aspect, it would be beneficial to delve deeper into this topic and explicitly connect it with the obtained results.

We agree that the aspect of trait co-variation was not discussed in depth, and we will strengthen this aspect in the revised version. Our results show co-variation of different traits across biomes in the trait space and a principal component analysis. For example, tropical forests are characterized by the highest vegetation, low SLA, high leaf N and deep roots. These traits represent dominance of tall trees with evergreen phenology (typically associated with low SLA). Tall vegetation also indicated intense light competition. Deep roots indicate that substantial amounts of water percolate into deeper soil layers and may indicate rooting niche separation between tall trees with deep roots and understory vegetation with shallow roots. Deep roots and height can be correlated because tall plants need deep roots for mechanic stability. Shallow roots in boreal forests indicate that those ecosystems are less limited by water or nutrients and more by energy.

While I acknowledge that analyzing the occurrence and abundance of traits in each biome may not be the primary focus of your manuscript, it presents an opportunity for a valuable complementary analysis. By examining the diversity of traits within each biome, specifically by assessing the variance in their distribution, you can provide additional insights into the functional composition and ecological characteristics of these biomes. This analysis could help elucidate patterns of trait diversity across different environmental gradients and provide a deeper understanding of the ecological processes driving biome differentiation.

We agree that such an analysis would be very interesting. However, the trait data we used are means for entire 0.5° grid cells such that diversity measures of biomes would represent variability in those mean values and not the full trait space per biome. Therefore, the original TRY data, the merged TRY+GBIF data (i.e., without averaging per grid cell) would be more suitable. In addition to the trait means used in our study, ranges and standard deviations of traits are also available for all traits and could be analyzed to study diversity. Yet, we think that such an analysis would go beyond the scope of the current manuscript, and it should therefore be analyzed in a separate study. We will mention this point in the discussion.

Writing errors:

I believe a parenthesis is missing for the reference in the following sentences:

- 1. "Biomes are commonly used to represent major vegetation formations and to map their biogeographic distributions. Multiple biome maps were developed based on a variety of different data sources Beierkuhnlein and Fischer (2021)."*
- 2. "Despite the increasing availability of trait data in databases such TRY (Kattge et al., 2020) and extrapolated global biome maps Wolf et al. (2022); Boonman et al. (2020), a systematic assessment of the performance of traits for biome classification and an identification of the most appropriate traits remain elusive."*

Thanks for spotting those errors, we will correct the brackets.

References

- Boonman CCF, Huijbregts MAJ, Benitez-Lopez A, Schipper AM, Thuiller W, Santini L (2022) Trait-based projections of climate change effects on global biome distributions. *Diversity and Distributions*, 28, 25–37.
- Bruehlheide H, Dengler J, Purschke O, et al. (2018) Global trait–environment relationships of plant communities. *Nature Ecology & Evolution*, 2, 1906–1917.
- Diaz S, Kattge J, Cornelissen JHC, et al. (2016) The global spectrum of plant form and function. *Nature*, 529, 167–171.
- Verheijen LM, Aerts R, Bönisch G, Kattge J, Van Bodegom PM (2016) Variation in trait trade-offs allows differentiation among predefined plant functional types: implications for predictive ecology. *New Phytologist*, 209, 563-575.
- Wright IJ, Reich PB, Westoby M, et al. (2004) The worldwide leaf economics spectrum. *Nature*, 428, 821–827.