

# Ref. JVS-R-05435 - Response to reviewers

Eduardo Fernández Pascual

12/8/2020

## Editor's comments

Thanks for your contribution. Two reviewers think your work is valuable but it is necessary you better define the biome (and geographic) scope for inclusion in the database, and also revisit the available data that contributes to your database. You'll see that one reviewer thinks you've missed lots of relevant information. Please address these concerns. If you can do this, the quality of the database will be greatly improved.

Please, find below the response to the comments of the reviewers.

## Referee 1's comments

This manuscript suggests publication and description of an important data-base on germination features for temperate woodland plants, one of the largest and best documented to date. The data-base has been well prepared, described in detail and two figures summarize geographical extent, variation and mean values. The amount of data is enormous, and will be an important tool to analyse features of regeneration by seed for forest ecosystems in the world. The authors restrict their data-base to "Northern Hemisphere" but include on the map biomes and records from Chile, Australia and New Zealand, this should be either corrected in text or in the figure to make it coherent.

We restricted the collection of data to frequent species occurring in vegetation plots of the northern hemisphere. However, the web search for germination data found records of experiments performed with material of those species that had been collected in the southern hemisphere. In other words, the database is for the nemoral flora of the northern hemisphere, but the records come from both that area and other areas where the species can be found. We have added two sentences to the Summary of contents section of the results to clarify this point: *Although the frequent species list was created using only relevés from the northern hemisphere, the web search of germination articles found studies of those species that had been performed in the southern hemisphere. These records are included in the database, and their geographical origin is reported.*

As to the extent of the biomes it should be stated somewhere that all of these forests undergo frosts (except most of Australia). The ecological description of the biome used might be further be extended with abiotic factors that separate this biome from adjacent ones (frost, drought, etc).

We have included this descriptive sentence in the introduction: *Generally, these forests are characterized by some frost events during the winter while summers are moist and do not suffer drought for any relevant length of time, conditions that favour the dominance of broadleaf deciduous trees.*

Probably the WWF forest biome source used is not the most relevant here, a map based on the Köppen climate classification or similar would simplify the understanding of the map and also probably better match records to the biome map.

We agree that the Köppen climate classification would probably be more recognisable for many readers. However, we used the WWF classification for a practical reason: it includes a biogeographical regionalization

(the ecoregions) which we used to stratify the random sampling of the vegetation plots from which we constructed the frequent species list (please see the methods section for a longer description of the procedure). We did this to get a homogeneous sampling across the biogeographical regions of the biome. Since this is the regionalization that we used to create the species list, we believe that we should keep the WWF map to delimit the study area. However, to facilitate understanding with the Köppen system, we have included this sentence in the introduction: *Deciduous forests are dominant in temperate latitudes with mild to cold winters and relatively warm and wet summers, corresponding with the Cf climate in Köppen's classification (mild temperate, fully humid).*

L57: may be You can use “nemoral biome” explaining it as “Temperate Broadleaf and Mixed Forests biome of the northern hemisphere” and avoiding a long and difficult to follow abbreviation? (TBMFNNH) or “temperate woodlands” ? This might help to read the text

We have adopted the *nemoral biome* abbreviation in the text, as suggested.

L193 : may be replace Sown\_by\_container by Number\_per\_container

Done.

L195 : you want to say the “last day” when seeds were counted?

Yes, corrected.

L195: do you have data on the scarification (=permeabilizing seed coats) method? (physical abrasion, acid treatment, heat shock)?

Sorry, but we did not record this information.

L201: do you have data on the duration of stratification?

Yes, when this information was reported in the original paper. It is recorded in the column *Stratification\_days*.

L211: why “germinable”? Some of them might have been dead at the beginning? May be “Number\_seed” used?

Changed to *Number\_seeds*.

L173: do you maintain a version and backup system for this database in the case updates may come?

The database is maintained in a GitHub repository, which provides a version control system. The repository will be made open when the manuscript is accepted for publication. A website will also be provided, where it will be possible (i) to access the last version of the database; (ii) to be informed about updates; and (iii) to get in contact, to join and contribute data to the database. Please see the last sentences that have been added at the end of the discussion.

L219: do You think this app can be extended to add data to the data-base?

The app functionality is limited to visualization of the dataset and, as far as we know at this moment, cannot be expanded into a GUI to submit and incorporate data to the database. However, the addition of new data is possible and easy to do using the R code that is behind the app. The website where the app and the database are hosted will provide information on how to contribute data to the database.

## Referee 2's comments

The idea for a deciduous forest database for seed germination is excellent. Some things need to be made more clear than they are now. The list of included species is relatively short, when one considers the size of the flora of the deciduous forest biome and the amount of work that has been done on seed germination of members of this biome. Further, species from subalpine coniferous species have been included, and they need to be separated from

the deciduous forest species. Lines 57. Whose vegetation classification system are you using? What is the source for the name Temperate Broadleaf and Mixed Forests?

This is the official name of the WWF biome, which we used to sample the vegetation plots that provided the species list (please see the methods, and the response to referee 1). For clarity, we have also added the WWF reference at the first mention of the name in the introduction.

Line 64. “Baskin, 1995” is not a good reference to use here. Why not the Baskin and Baskin 2014 book on seeds? This reference is not correct on line 64 or in the Literature Cited. It should be Baskin et al., 1995. The authors of this paper are Baskin, C. C., Meyer, S.E. and Baskin, J.M. Also, this paper is not about species that grow in the deciduous forests per se.

We have changed the reference to the Baskin book, as suggested.

Lines 86-88. You need to explain to the reader why you have included subalpine coniferous forests of the Cascade Mountains in the same category as deciduous (nemoral) forests of USA, China and Europe. What vegetation classification systems puts these very different kinds of forests together?

The referee is right that the Cascadia ecoregions do not belong in the TBMF biome. The truth is that these regions were included in the search for practical reasons related to the way the project’s workflow was organised, and are unrelated to the deciduous forest biome. Once we had germination records for species from these regions, we had to make the difficult decision of either keeping them (as they are indeed valuable data which might be of interest to the users of the database) or removing them (to keep the homogeneity of the deciduous forest concept).

We have tried to explain this better in the methods section. We have also included a column with the biome variable in the dataset, so the records belonging to species from Cascadia can be identified and separated at any time. Alternatively, if the editor considers it more reasonable, those records could be removed from the dataset and the manuscript.

Pages 21-500 What is the meaning of all these pages? These must be references you consulted, but there is no explanation.

These are the supporting materials, please see the Supporting Information section (which is after the bibliography) for a description of the files. We tried to upload these files in specific formats (a csv data file and BibTeX bibliographies) but the online submission system forced us to transform them to pdf and appended them at the end of the manuscript. We are sorry by this inconvenience.

I was shocked when I saw your list of species. It is not very complete for the deciduous forests, even for eastern North America. There must be over 100 papers from the Baskin’s work that you have not included. You should take a look at the Baskin and Baskin (2014) book on seeds. Look at the list of papers of their own work that they cite. I think the idea behind this project is noble and the product will be of great use to people. However, at this point I think the project is far from being complete. Also, I think mixing species from the deciduous forests with those from subalpine coniferous forests will create confusion for people in the future who try to use this data base.

We agree that the database is not complete, but we argue that a project of this nature can never be truly complete. As we discuss with the other referee (please also see the new sentences at the end of the discussion) this is a database that intends to be alive, being updated and expanded with future projects, and welcoming collaborators who want to contribute new datasets.

We followed a systematic approach to search for data, with the purpose of prioritising work and maximising its utility for vegetation science. In short, this approach consisted in (i) obtaining a representative number of vegetation plots for the biome from sPlot (<https://www.idiv.de/en/splot.html>); and (ii) using the plots to produce a list of frequent species (i.e. species that occur in at least 5 % of the plots of a region). Please see the methods for more details.

This systematic approach may leave out relevant information on species of the biome (as the case mentioned by the referee) for three reasons: (1) the species is not recorded in the vegetation databases; (2) the species

has been recorded, but it is infrequent; or (3) the species is recorded and frequent and goes into our list, but the relevant germination papers are not found by our web search because the species names is not mentioned in the title or the abstract. While points 1 and 2 are intentionally excluded by our methodology, point 3 would be a failure of the methodology (specifically, a failure of the search string to find a relevant paper).

For this reason, we had applied a protocol to detect the importance of point 3. Incidentally, this protocol used the Baskin book. The protocol consisted in the following: After completing the database, we took the species in our frequent species list for which we had not found relevant papers and searched them in the taxonomical index of the Baskin book. There were 273 species with missing data and mentioned by the Baskins in their index. We took a random sample of 10% of these species and searched the original references in the Baskin book, and found that all of them were either inaccessible (national journals or old conference proceedings not available online) or had no relevant data (i.e. provided no primary germination data that we could retrieve). From this trial, we concluded that our search string had found a reasonable amount of the available information, by comparing it to the best reference possible, the review of the Baskins.

Therefore, the articles mentioned by the referee mostly pertain to species that we intentionally left out of the systematic approach, because of their low representation in vegetation databases. Although of course the approach suggested by the referee is valid, and the information on those works valuable, it is also opportunistic (i.e. it would include references that we are personally aware of, instead of the systematic set of references found by the search based on occurrence in vegetation plots). It is difficult to see where this opportunistic approach would justifiably stop looking for more papers, and consider the database complete.

We believe that the amount of data we provide (primary data extracted from more than 600 references) justifies its utility and the publication of this report as a first building block of a database that will be expanded and maintained. The utility of the dataset is specially justified because, by focusing on frequent species from vegetation databases, it maximises its value for vegetation scientists. As it stands now, it is a good sample of frequent species of the deciduous forest of the northern hemisphere, with a good representation of, for example, many of the dominant tree species.