Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A database with 1 Million individual observations of intra-specific trait variation for 10 European tree species: the OptFORESTS provenance trial database

Jan-Peter George

jan-peter.george@luke.fi

Natural Resources Institute Finland (Luke)

Alin-Madalin Alexandru

National Institute for Research and Development in Forestry “Marin Dracea”, (ICAS)

Samuel Aspalter

Natural Hazards and Landscape (BFW)

Maria Belovarska

Executive Forestry Agency (FSCS)

Benjamin Brachi

National Research Institute for Agriculture, Food and Environment (INRAE)

Regina Chambel

National Institute for Agricultural and Food Research and Technology (INIA)

José Climent

National Institute for Agricultural and Food Research and Technology (INIA)

Arnaud Dowkiw

National Research Institute for Agriculture, Food and Environment (INRAE)

Bruno Fady

National Research Institute for Agriculture, Food and Environment (INRAE)

Irena Fundova

Norwegian Institute of Bioeconomy Research (NIBIO), Norway

Jonas Glatthorn

Swiss Federal Research Institute (WSL), Switzerland

Santiago C. Gonzalez Martinez

National Research Institute for Agriculture, Food and Environment (INRAE)

Carina Heiling

Natural Hazards and Landscape (BFW)

Francisco José Lario

Empresa de Transformación Agraria, S.A. (TRAGSA)

Luka Krajnc
Slovenian Forestry Institute (GIS/SFI)
Jon Kehlet Hansen
University of Copenhagen (UCPH)
Maurizio Marchi
National Research Council (CNR)
Georgeta Mihai
National Institute for Research and Development in Forestry “Marin Dracea”, (ICAS)
Luc Pâques
National Research Institute for Agriculture, Food and Environment (INRAE)
Silvio Schüler
Natural Hazards and Landscape (BFW)
Christian Rellstab
Swiss Federal Research Institute (WSL), Switzerland
Arne Steffenrem
Norwegian Institute of Bioeconomy Research (NIBIO), Norway
Jan Stejskal
Czech University of Life Sciences (CZU), Czech Republic
Emanuel Stoica
National Institute for Research and Development in Forestry “Marin Dracea”, (ICAS)
Eva Vera Jiminez
Empresa de Transformación Agraria, S.A. (TRAGSA)
Egbert Beuker
Natural Resources Institute Finland (Luke)

---

data-descriptor

Keywords:

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9849095/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Intraspecific trait variation is a keystone variable in evolutionary biology as it contains information on past adaptational processes as well as future evolutionary potential of species. Recently, intraspecific trait variation has become an important data source for various research disciplines such as species distribution modelling, eco-evolutionary research, gene-environment associations, tree breeding and many more. Yet, systematic and rangewide data for long-living organisms such as trees are rare. This is partly because only a few old experiments on intraspecific trait variation still exist in Europe. What is more, their data are often stored in non-digital format and are therefore hardly findable, accessible, interoperable and reusable. Here we present a provenance trial dataset aiming to fill this gap by compiling information on intraspecific trait variation for 10 European tree species together with necessary metadata on trial sites and provenance origin. In total, measurements from more than 1,900 provenances growing in 75 trial sites across Europe provide nearly 1 million observations on growth, phenology, survival, competition, fructification, resilience, and other important life history traits. This OptFORESTS provenance trial data base was prepared as part of the European HORIZON OptFORESTS project. It represents one of the largest datasets on phenotypic variation in trees that has been published to date in terms of number of species and provenances tested and is the first initiative aiming at compiling such data across European countries with open access .

Background & Summary

For more than two centuries, foresters, botanists, ecologists, and geneticists have been interested in studying the effects of transplanting seeds and plants from their natural place of origin to novel environments (Langlet, 1971). Such experiments are usually called common garden experiments or, more specifically in forest sciences, being referred to as *provenance trials* (Fady & Rihm 2022). While the predominant scientific purpose of these experiments was and still is to disentangle the relative contributions of environment and genetics to phenotypic variation, these experiments offer now an incredible wealth of further research opportunities in the era of statistical modelling, genomics, and phenomics (Alberto et al. 2013; Leites et al. 2012; Leites & Benito-Garzon 2023; Boshier & Buggs 2014). Possible applications of provenance trial data range from classical genecological studies of fitness traits such as height, survival, and phenology (e.g. Harberd 1957; Lesser & Parker 2004; Clausen et al. 1940) to studying intraspecific differences in climate-growth relationships (e.g. Taeger et al. 2013; Montwe et al. 2015) given that both accessibility, resolution and quality of global climate databases have considerably improved over the last years. Recently, data from provenance trials are also increasingly studied in combination with novel high-resolution molecular markers to assess genotype-environment and genotype-trait associations (Hansen et al. 2020; Mahoney et al. 2019, George et al. 2021). What is more, almost all provenance trials with trees being older than 15–20 years constitute a valuable information source for dendroclimatological studies aiming at detecting intraspecific differences in growth reactions and wood properties to changing climates and extreme climatic events (e.g. Nabais et al. 2018; Sass-Klaassen et al. 2016; Kormann et al. 2024).

In the face of globally changing climatic conditions, for tree populations it may be challenging to avoid local maladaption and fitness decline, because of their slow natural migration capacity and long generation times (Aitken et al. 2008). One possible outcome of this is the human-mediated transfer of seed or pollen to more favourable growing sites in the future so that the genetic legacy in growth and phenology will not lag behind future climates (Williams & Dumrose, 2013). The terms *assisted gene flow* (e.g. Chakraborty et al. 2024) and *assisted migration* (Aitken & Whitlock 2013; Chludill et al. 2025) are used nowadays widely in the scientific community as a keystone concept to describe the movement of forest reproductive material either within or beyond boarders of natural occurrences. While already implemented in some national seed transfer guidelines in North America (e.g. Pedlar et al. 2011), there is not yet a uniform transfer guideline for future seed source deployment in Europe. The OptFORESTS provenance trial database provided here can be of tremendous value for developing such a harmonized seed transfer system for European forestry as it contains information on fitness-related traits and intraspecific genetic variation for 10 tree species (Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) H. Karst), pendunculate oak (*Quercus robur* L.), sessile oak (*Quercus petraea* Liebl.), European larch (*Larix decidua* Mill.), European ash (*Fraxinus excelsior* L.), stone pine (*Pinus pinea* L.), maritime pine (*Pinus pinaster* Ait.), black pine (*Pinus nigra* Arnold), *and* wild cherry (*Prunus avium* L.)). This kind of information is needed to disentangle the relative contributions of environment (i.e trial site), genetics (i.e. provenance), and plasticity (site x provenance) to overall phenotypic diversity. The many site and provenance metadata available in this dataset will allow for detailed and easy downstream analyses of data, for instance geographic locations for climate data retrieval, genetic background of the tested material, and many more.

This paper covers information on the general content of the OptFORESTS provenance trial database, its diversity regarding geographic coverage, measured traits as well as the associated metadata for provenances and trial sites that were tested. It also provides a reproducible workflow example of the dataset, which demonstrates the entire process from downloading the data to a very first genetic model together with customized plotting options and data validation in the R environment. So far, various publicly available databases for provenance trial data exist (e.g. Liepe et al. 2024 ; Risk et al. 2021, Fady et al. 2024), but the number of species and provenances compiled in the OptFORESTS provenance trial database is -to the best of our knowledge- the largest dataset so far. We therefore strongly believe that this dataset will soon gather attention beyond the forestry community and will make a significant contribution to future modelling in vegetation science, genetics, and ecology.

Methods

Study design, selection of species, and availability of trial sites

10 tree species were identified, native to Europe, of which a sufficient number of common garden trials is available. Logically most of these species have considerable economic and ecologial potential for European forestry. Further criteria included the vulnerability of these tree species to climate change and their potential to strengthen resilience in European forests through the exploitation of intraspecific

variation and targeted selection of climatically-adapted seed sources (Chakraborty et al. 2024). Ten different countries provided metadata for a preliminary set of available trial sites, which were potentially suitable for the database. The collected metadata included geographic positions of trial sites and provenances, age of the trial sites, trial design (e.g. block-repetition, single-tree repetition, etc.), number of available trees per provenance and available traits. After review of the provided metadata (Fig. 1, steps i and ii), a final assembly of trial sites was selected from this preliminary set for each species. This final assembly encompassed 75 trial sites with a total of 1,932 provenances. Figure 1 gives an overview of the workflow of data collection, standardization, and processing. Figure 2 illustrates the geographic origins of provenances and locations of trial sites within the biogeographical context of each species.

Standardized collection of data and metadata

Three spreadsheets with instructions were sent out to national data providers to collect measurement data and metadata of trial sites and provenances (**Supplementary Material S1**). This ensured that the data was entered and returned in a standardized and uniform format. Every data sheet was quality checked to ensure that measurement units were consistent within trial sites, that geographic coordinates were in World Geodetic System 1984, that dates were in correct format, and that provenance identifiers in observation files exactly matched those in metadata files. Measurements conducted on 161,681 trees resulted in a total of 966,707 observations (trees x years). Spreadsheets that did not pass the quality check were either returned to data providers or additional information was requested for those cases were only a few inconsistencies remained.

Data processing

Quality-checked data files were transformed into R data files. The RDS file format was chosen to preserve the dataset in form of a data table. Data tables in R are fast and highly optimized alternatives to classical data frames, which can speed up computations significantly for large datasets (Dowle et al. 2019). Four additional columns were then created as unique and standardized identifiers: a unique identifier for each trial site (OptFORESTS trial code), an unique identifier for each provenance in the dataset (OptFORESTS provenance code), and an unique tree identifier consisting of a combined trial site-provenance-block-tree id (OptFORESTS tree id). These unique identifiers permit filtering the dataset fast and efficiently and allow to merge information of the different sub-databases based on common ids (see next section).

Data Records

The OptFORESTS provenance database is organised in two main folders under the open repository Zenodo (https://zenodo.org/records/20433863). The folder ”*Data*” includes the four sub databases referring to trial site metadata, provenance metadata, tree observational data, as well as the survival data (Fig. 3&4). The second folder ”Datasheets & Data dictionary” includes pdf files for each of the four sub-databases with additional information such as an explanation of each column. In addition, a data

dictionary explains the exact meaning, measurement type, and measurement scale of all available traits in the OptFORESTS provenance database.

The core of the OptFORESTS provenance database is formed by the four .rds database files, which can be easily downloaded, processed, and interconnected to each other via linking variables in the R statistical software (R Core Team, 2021) (Fig. 5). The file “OptFORESTS trial site metadata” includes all necessary information about the trial site such as species, country, geographic position, trial site design, establishment year, number of blocks, as well as contact details. The file “OptFORESTS provenance metadata” includes geographic coordinates and altitude information as well as information about the genetic background of the planted material (e.g. open pollinated, half-sib/full-sib, seed orchard material). The “OptFORESTS tree database” and “OptFORESTS survival database” files include the measured trait information and have together more than 1.3 Million entries. However, since not every tree has been measured in every year, the number of numerical trait data is still almost 1 Million. Assessment dates are stored as date variables so that time differences between two or several assessments can be seamless calculated. For each trait and observation, the sub-database includes information tags regarding the mode of measurement (e.g. ”measured”, ”remotely sensed”, ”imputed”) so that potential differences in accuracy and precision can be accounted for in modelling procedures.

The survival data is partly derived from the tree data and transforms single-tree binary survival information into survival rates per block and time period, expressed as annual mortality rate and annual mortality rate corrected for the different lengths in assessment periods (Sheil & May 1996).

Data overview

The dataset presented here includes eco-evolutionary traits of great importance to tree populations. The vast majority of observations represent classic growth parameters such as height, diameter, and crown length. However, unlike many other published datasets, this dataset also includes traits related to reproduction, phenology, survival & management, form and competition, damage, physiology, and others. The user can easily make a thematic preselection by using the OptFORESTS trait theme column to select a theme of interest. Figure 6 provides an overview of the number of observations separately for each thematic category for the available tree species.

Technical Validation

The dataset has undergone two quality checks during the data collection period to verify measurement units, assessment dates, traits, provenance names and geographic positions. Geographic positions of trial sites and provenance origins have been manually plotted and compared with provided location names. Provenances for which locations could not be verified were omitted from the database.

Tree observations were checked for duplicates for each assessment date and for unreasonable growth behaviour in case of all growth traits and survival (i.e. no negative growth rates or survival rates).

However, for some trial sites a very small number of negative growth rates was observed for measurements taken at short succession in relatively young ages, which may be explained by broken terminal shoots or browsing of the trees.

Usage Notes

An important aspect of using this dataset is the fact that data from provenance trials in different countries have been combined. For traits such as height or diameter, data can therefore be easily compared with one another without significant bias. For other traits, however, with a higher degree of measuring complexity (e.g., phenology, reproduction), it will be necessary to thoroughly analyse the metadata of the individual experimental plots as well as the measurement scale of the relevant trait with the help of the data dictionaries before drawing conclusions. We therefore recommend always including relevant metadata as factors in the modelling (e.g., as random effects in generalized mixed linear models). The working example which is provided as **Supplementary file S2** to this paper includes an example on how to account for such variation in a simple provenance transfer model.

Declarations

Data availability

The data is publicly available in the open repository Zenodo under the following URL: https://zenodo.org/records/20433863.

Code Availability

The OptFORESTS database runs under R version 4.3.2 (Eye holes). This requires the R package data.table version 1.15.2. A working example called *demoscript_OptFORESTS_2.0.rmd* is available as **Supplementary Material S2** as a R Markdown file and requires the following packages and versions: ggplot2 (3.5.0), dplyr 1.1.4, tidyverse (2.0.0), lubridate (1.9.3), stringr (1.5.1), stringx (0.2.9), breedR (0.1.2), rnaturalearth (1.0.1), car (3.1-3), and EnvStats (3.1.0).

Acknowledgements, Author Contributions & Competing Interests

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101081774: OptFORESTS - Harnessing forest genetic resources for increasing options in the face of environmental and societal challenges

JPG and EB designed the database and were responsible for collecting, cross-checking, and standardization of data. AA, SA, MB, BB, RC, JC, AD, BF, IF, SGM, CH, FL, LK, JKH, MM, GM, LP, SS, CR, AS, JS, ES & EJ provided data, JPG & EB were responsible for drafting and coordinating the manuscript, all authors commented and revised the manuscript and agreed to the final version to be published.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual review of ecology, evolution, and systematics*, *44*(1), 367-388.
2. Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, *1*(1), 95-111.
3. Alberto, Florian J., et al. "Potential for evolutionary responses to climate change–evidence from tree populations." *Global Change Biology* 19.6 (2013): 1645-1661.
4. Boshier, D., & Buggs, R. J. (2015). The potential for field studies and genomic technologies to enhance resistance and resilience of British tree populations to pests and pathogens. *Forestry: An International Journal of Forest Research*, *88*(1), 27-40.
5. Chakraborty, D., Ciceu, A., Ballian, D., Benito Garzón, M., Bolte, A., Bozic, G., ... & Schueler, S. (2024). Assisted tree migration can preserve the European forest carbon sink under climate change. *Nature Climate Change*, *14*(8), 845-852.
6. Chludil, D., Čepl, J., Steffenrem, A., Stejskal, J., Sagariya, C., Pook, T., ... & Lstibůrek, M. (2025). A Pollen-Based Assisted Migration for Rapid Forest Adaptation. *Global Change Biology*, *31*(1), e70014.
7. Clausen, J., Keck, D. D., & Hiesey, W. M. (1940). Experimental studies on the nature of species. I. Effect of varied environments on western North American plants. *Carnegie Institution of Washington Publication* **520.**
8. Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., ... & Tan, X. (2019). Package 'data. table'. *Extension of 'data. frame*, *596*, 368.
9. Fady, B., Vauthier, D., Rei, F., Scotti-Saintagne, C., Veisse, D., Bernier, F., Severin, L. (2024). Phenotypic trait and dendrometric data (1971-2023) for Pinus nigra Arn., the European black pine, from French common gardens. https://doi.org/10.57745/SKPSYI, Recherche Data Gouv, V2.
10. Fady B., Rihm G., 2022. Arboretums, common gardens and forest tree resilience. New Forests 53, 603–606 (https://doi.org/10.1007/s11056-022-09908-y). https://hal.inrae.fr/hal-03669883v1.
11. George, J. P., Schueler, S., Grabner, M., Karanitsch-Ackerl, S., Mayer, K., Stierschneider, M., ... & van Loo, M. (2021). Looking for the needle in a downsized haystack: Whole-exome sequencing unravels genomic signals of climatic adaptation in Douglas-fir (Pseudotsuga menziesii). *Ecology and Evolution*, *11*(12), 8238-8253.
12. Hansen, O. K., Lamour, K., Whetten, R., Xu, J., Cuenca, A., Islam, M. S., & Nielsen, U. B. (2020). Ad hoc breeding of a genetically depauperate landrace of noble fir (Abies procera Rehder) using SNP genotyping via high-throughput targeted sequencing. Tree Genetics & Genomes, 16(5), 63.
13. Harberd, D. J. (1957). The within population variance in genecological trials. *The New Phytologist*, *56*(3), 269-280.

14. Kormann, J. M., van der Maaten, E., Liesebach, M., Liepe, K. J., & van der Maaten-Theunissen, M. (2024). High risk, high gain? Trade-offs between growth and resistance to extreme events differ in northern red oak (Quercus rubra L.). *Frontiers in Plant Science*, *15*, 1374498.
15. Langlet, O. (1971). Two hundred years genecology. *Taxon*, 653-721.
16. Leites, Laura, and Marta Benito Garzón. "Forest tree species adaptation to climate across biomes: Building on the legacy of ecological genetics to anticipate responses to climate change." *Global Change Biology* 29.17 (2023): 4711-4730.
17. Leites, L. P., Robinson, A. P., Rehfeldt, G. E., Marshall, J. D., & Crookston, N. L. (2012). Height-growth response to climatic changes differs among populations of Douglas-fir: a novel analysis of historic data. *Ecological Applications*, *22*(1), 154-165.
18. Lesser, M. R., & Parker, W. H. (2004). Genetic variation in Picea glauca for growth and phenological traits from provenance tests in Ontario. *Silvae Genetica*, *53*(4), 141-147.
19. Liepe, K. J., Rieckmann, C. A., Mittelberg, H. S., & Liesebach, M. (2024). Phenotypic variation in 1,100 provenances of Picea abies measured over 50 years on 33 German trial sites. *Scientific Data*, *11*(1), 854.
20. Mahony, C. R., MacLachlan, I. R., Lind, B. M., Yoder, J. B., Wang, T., & Aitken, S. N. (2020). Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. *Evolutionary Applications*, *13*(1), 116-131.
21. Montwé, D., Spiecker, H., & Hamann, A. (2015). Five decades of growth in a genetic field trial of Douglas-fir reveal trade-offs between productivity and drought tolerance. *Tree Genetics & Genomes*, *11*(2), 29.
22. Nabais, C., Hansen, J. K., David-Schwartz, R., Klisz, M., Lopez, R., & Rozenberg, P. (2018). The effect of climate on wood density: what provenance trials tell us?. *Forest Ecology and Management*, *408*, 148-156.
23. Pedlar, J., McKenney, D., Beaulieu, J., Colombo, S., McLachlan, J., & O'Neill, G. (2011). The implementation of assisted migrationin Canadian forests. *The Forestry Chronicle*, *87*(6), 766-777.
24. Core Team, R. (2021). R: A language and environment for statistical computing [Software]. *R Foundation for Statistical Computing*.
25. Risk, C., McKenney, D. W., Pedlar, J., & Lu, P. (2021). A compilation of North American tree provenance trials and relevant historical climate data for seven species. *Scientific Data*, *8*(1), 29.
26. Sass-Klaassen, U., Fonti, P., Cherubini, P., Gričar, J., Robert, E. M., Steppe, K., & Bräuning, A. (2016). A tree-centered approach to assess impacts of extreme climatic events on forests. *Frontiers in Plant Science*, *7*, 1069.
27. Sheil, D., & May, R. M. (1996). Mortality and recruitment rate evaluations in heterogeneous tropical forests. *Journal of Ecology*, 91-100.
28. Taeger, S., Zang, C., Liesebach, M., Schneck, V., & Menzel, A. (2013). Impact of climate and drought events on the growth of Scots pine (Pinus sylvestris L.) provenances. *Forest Ecology and Management*, *307*, 30-42.

29. Williams, M. I., & Dumroese, R. K. (2013). Preparing for climate change: forestry and assisted migration. *Journal of Forestry*, *111*(4), 287-297.

Figures

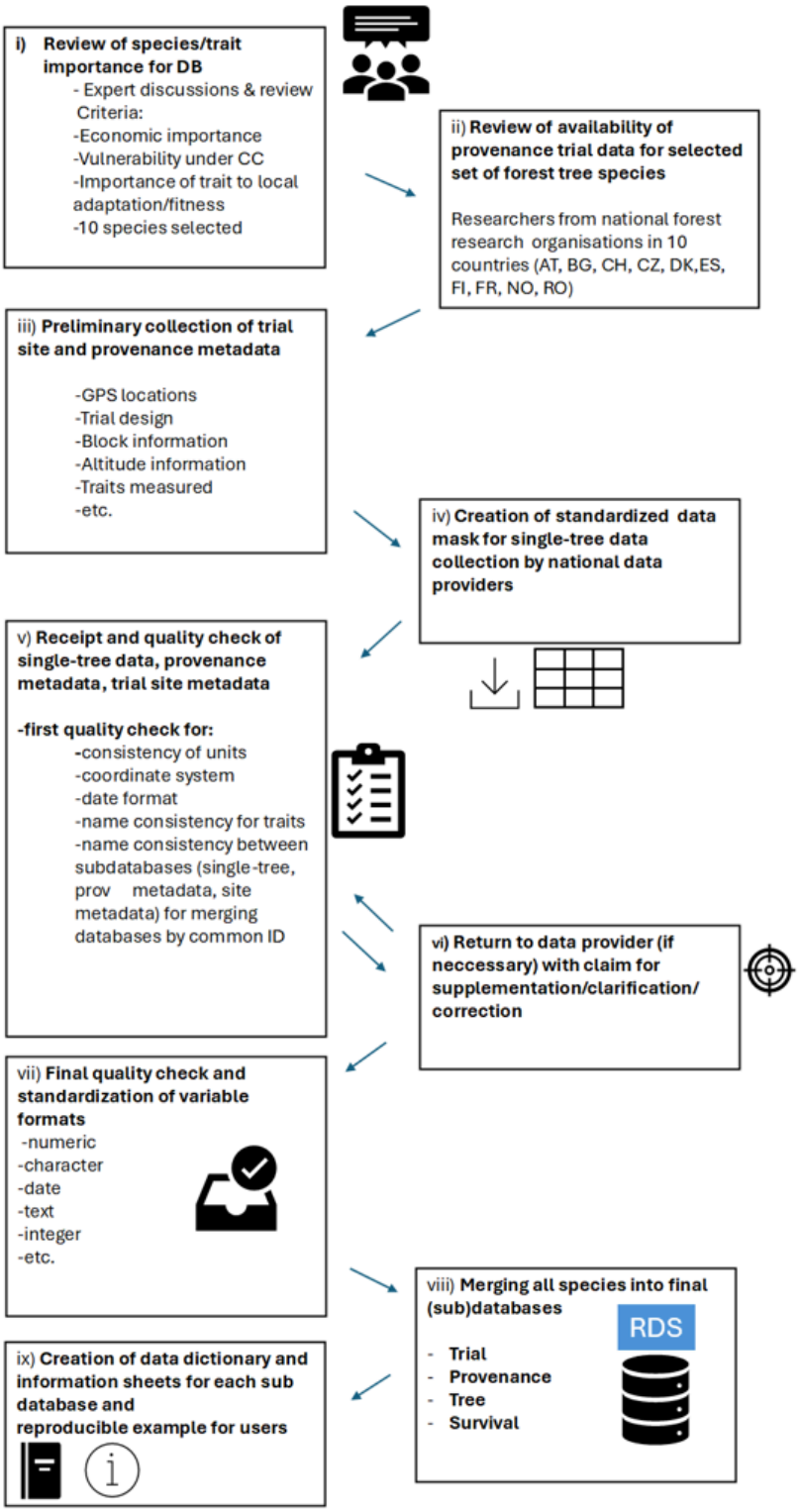


Figure 1

Workflow of data collection, review, and database creation.

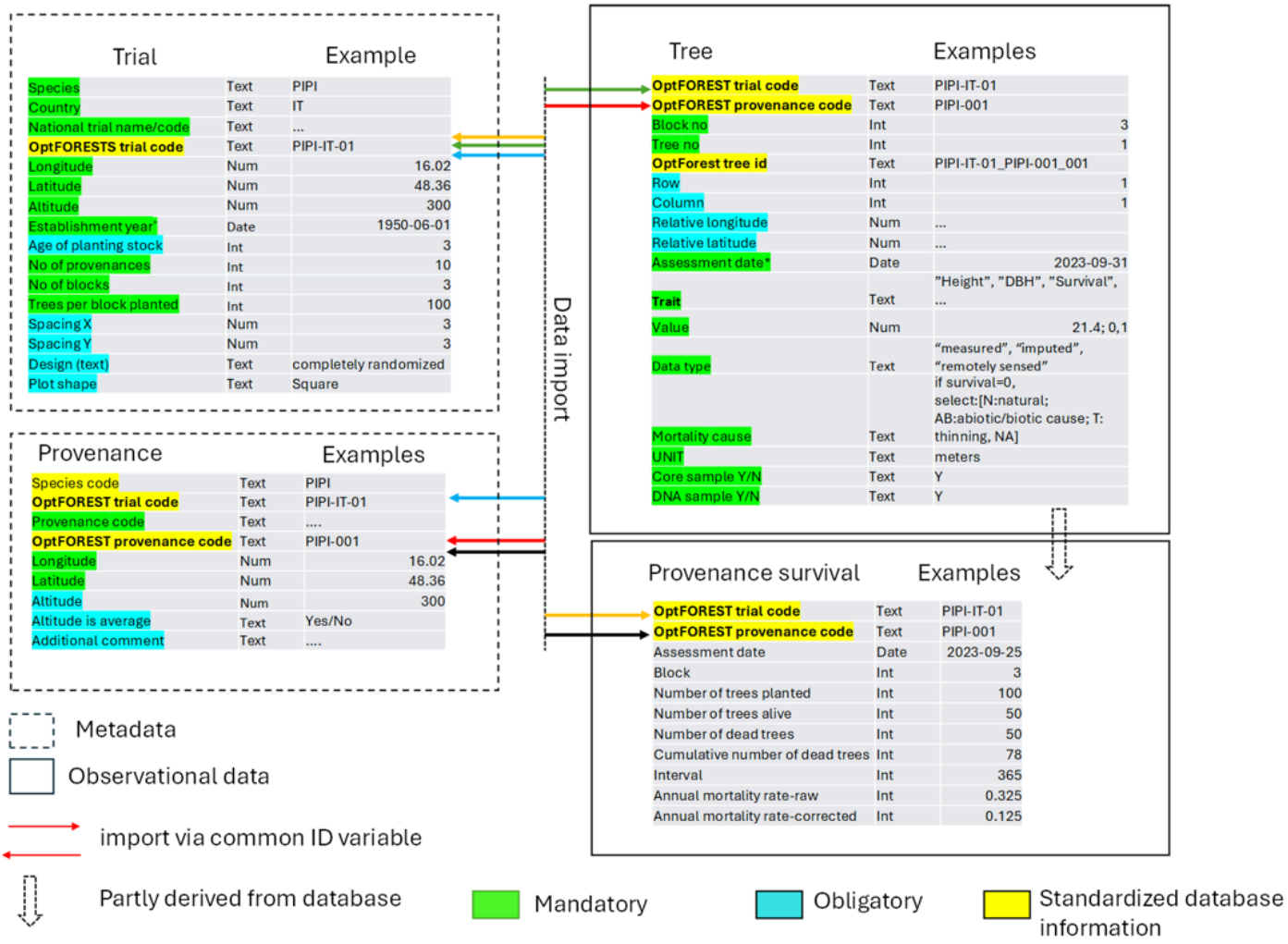


Figure 2

Structure of the database and relationships among sub databases.

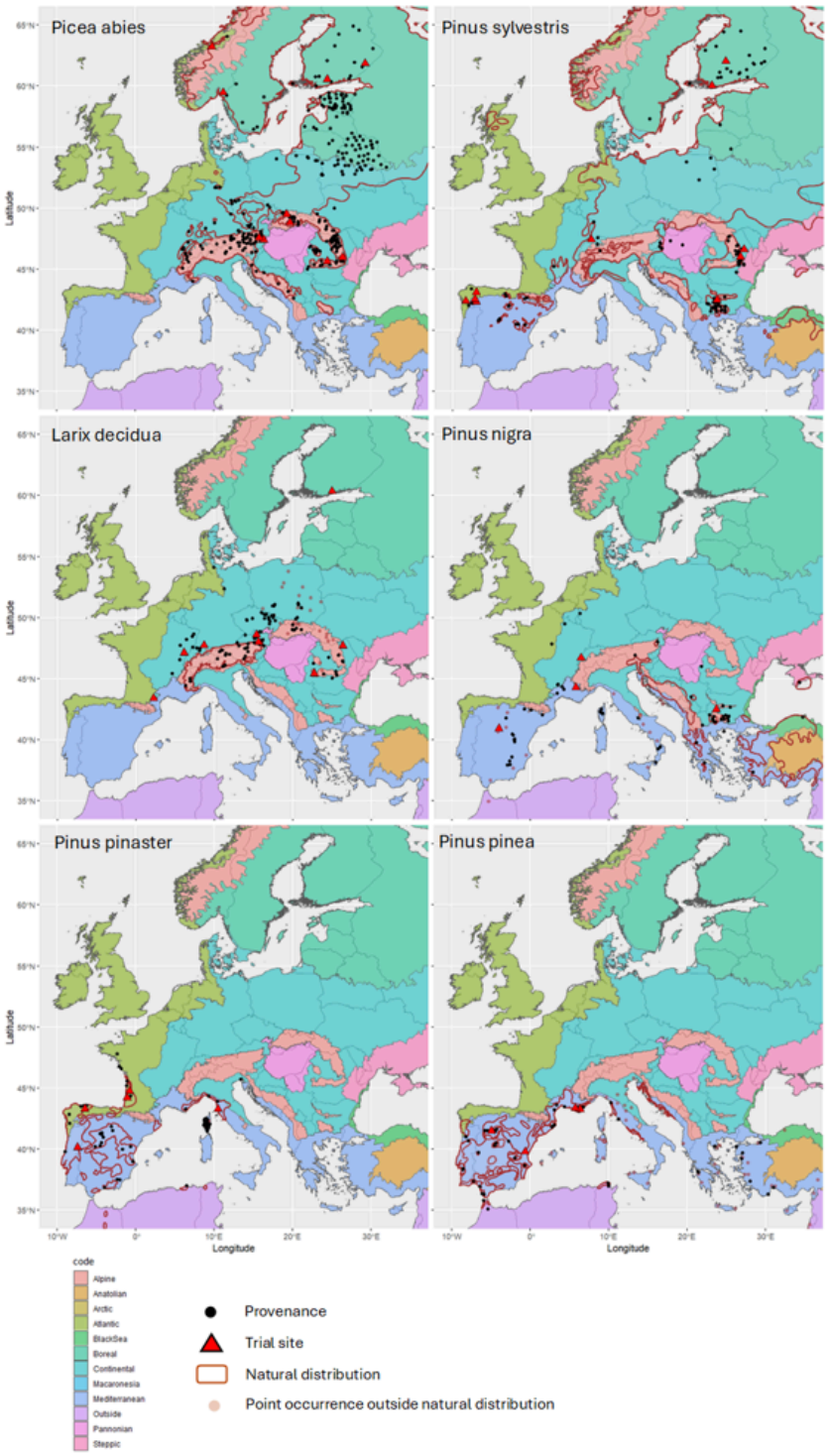


Figure 3

Geographic variation of provenances and trial sites for conifers in the OptFORESTS database

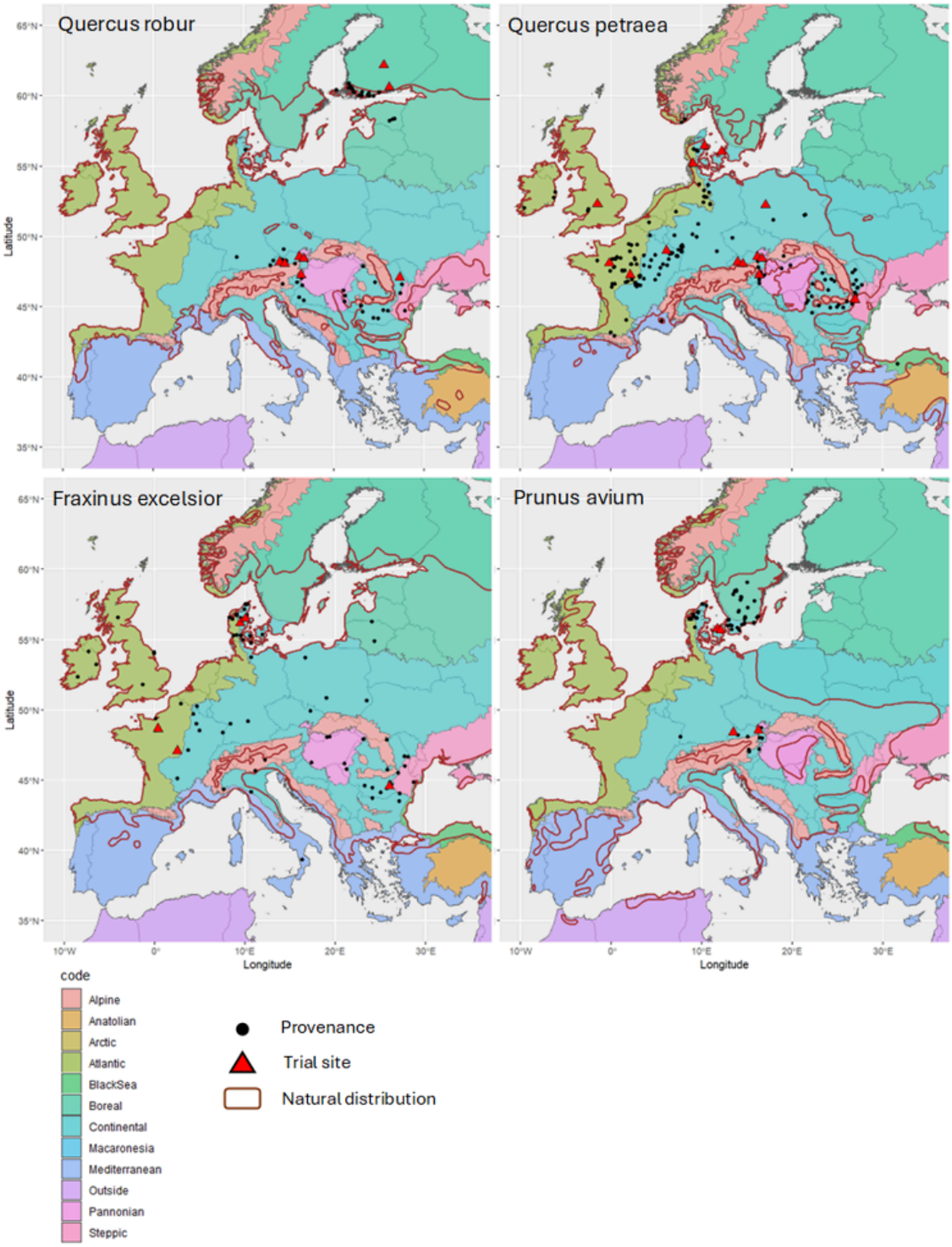


Figure 4

Geographic variation of provenances and trial sites for broadleaves in the OptFORESTS database

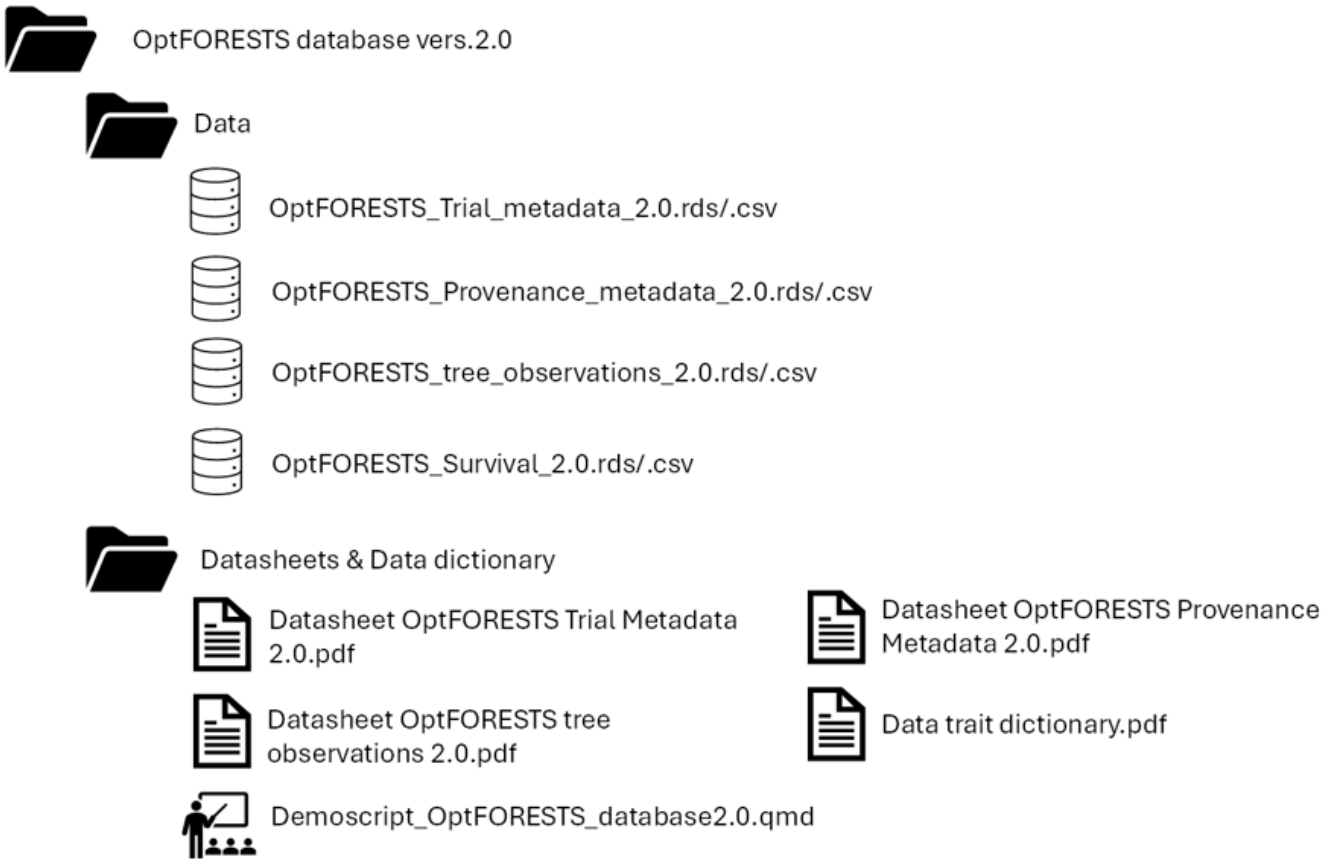


Figure 5

Overview of available folders and files in the OptFORESTS provenance database

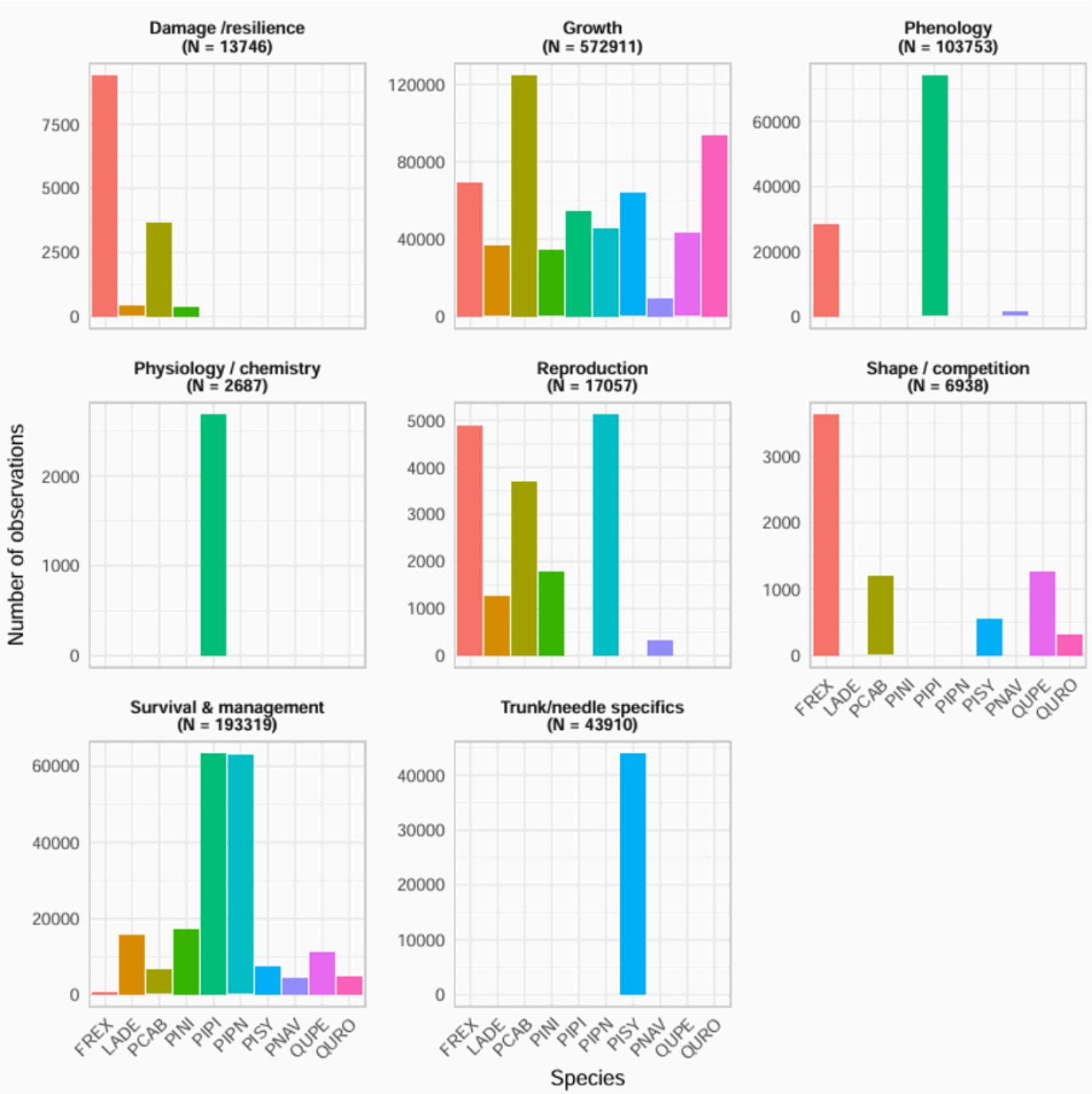


Figure 6

Overview of available trait categories and number of observations for each species in the database.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DemoscriptOptFORESTS2.0.rmd
- SupplementaryMaterialS1.xlsx