https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

LegacyPollen 1.0: A taxonomically harmonized global Late Quaternary pollen dataset of 2831 records with standardized chronologies

Ulrike Herzschuh[1,2,3], Chenzhi Li[1,2], Thomas Böhmer[1], Alexander K. Postl[1], Birgit Heim[1], Andrei A. Andreev[1], Xianyong Cao[1,4], Mareike Wieczorek[1], Jian Ni[5,6]

[1] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Polar Terrestrial Environmental Systems, Telegrafenberg A45, 14473 Potsdam, Germany

[2] Institute of Environmental Science and Geography, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

[3] Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

[4] Alpine Paleoecology and Human Adaptation Group (ALPHA), State Key Laboratory of Tibetan Plateau Earth System, and Resources and Environment (TPESRE), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, 100101 Beijing, China

[5] College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua, Zhejiang 321004, China

[6] Jinhua Mountain Observation and Research Station for Subtropical Forest Ecosystems, Jinhua, Zhejiang 321004, China

Correspondence: Ulrike Herzschuh (Ulrike.Herzschuh@awi.de)

Abstract. Here we describe the LegacyPollen 1.0, a dataset of 2831 fossil pollen records with metadata, harmonized taxonomy, and standardized chronologies. A total of 1032 records originate from North America, 1075 from Europe, 488 from Asia, 150 from Latin America, 54 from Africa, and 32 from the Indo-Pacific. The pollen data cover the Late Quaternary (mostly the Holocene). The original 10,110 pollen taxa names (including variations in the notations) were harmonized to 1002 taxa, with woody taxa and major herbaceous taxa to genus level and other herbaceous taxa to family level. The dataset

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Earth System Science Data
Open Access Discussions

is valuable for synthesis studies such as taxa areal changes, vegetation dynamics, human impact (e.g., deforestation), and climate change at global or continental scales. The harmonized pollen and metadata as well as the harmonization table are available from PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.929773; Herzschuh et al., 2021a). R code for the harmonization is provided at Zenodo (https://doi.org/10.5281/zenodo.5910972; Herzschuh et al., 2022) so that datasets at a customized harmonization level can be easily established.

1 Introduction

Global and regional palaeo-proxy databases and repositories are fundamental to palaeoclimatological and palaeoenvironmental synthesis studies and Earth system model validation (Gaillard et al., 2010; Trondman et al., 2015). Several continental fossil pollen databases have been successfully established (Gajewski, 2008), for example, the European Pollen Database (EPD), the North American Pollen Database (NAPD, http://www.ncdc.noaa.gov/paleo/napd.html) or the Latin American Pollen Database (LAPD, http://www.latinamericapollendb.com). In recent years, efforts have been made to integrate such databases into the Neotoma Paleoecology Database (https://www.neotomadb.org/; Williams et al., 2018), which provides a global collection of pollen data among other palaeoenvironmental proxy data. Furthermore, fossil pollen datasets for China and Mongolia (Cao et al., 2013; Herzschuh et al., 2019) and Siberia (Cao et al., 2020) have been compiled.

The numerous pollen records available in open databases, however, are not yet consistent concerning data type (e.g., pollen counts or percentages), pollen taxonomy, and nomenclature (Fyfe et al., 2009; Cao et al., 2013) and neither were their metadata approved and harmonized. For example, palynologists identify pollen taxa to different taxonomic levels ranging from (sub)species to order, depending on the purpose of their study and the differentiability and preservation of the pollen grains. Accordingly, some efforts have been made to harmonize taxonomies of pollen taxa in the databases (Fyfe et al., 2009; Giesecke et al., 2019; Githumbi et al., 2021; Mottl et al., 2021). However, a general framework is needed that can be applied to existing and newly published records.

Here we present LegacyPollen 1.0, a global taxonomically harmonized pollen dataset along with standardized metadata from 2831 sites for which recent chronologies have also been established (Li et al., 2022). This dataset is based on a general framework and implemented in R, which allows customized datasets to be built as well as the inclusion of new pollen records. The LegacyPollen 1.0 dataset is available at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.929773; Herzschuh et al., 2021a) and provides both count and percentage pollen data. We also provide the R code and the taxa harmonization table at Zenodo (https://doi.org/10.5281/zenodo.5910972; Herzschuh et al., 2022).

2. Methods

2.1 Data sources

We initially downloaded 3147 late Quaternary fossil pollen records (including dating) from the Neotoma Paleoecology Database (Neotoma hereafter; last access: April 2021) using the *Neotoma* package in R (Goring et al., 2019; R Core Team, 2020). As the spatial coverage of Neotoma records in certain regions is poor, for example, in China and Siberia, these records were supplemented by 324 records compiled by Herzschuh et al. (2019) and Cao et al. (2013, 2020) and a few new records (AWI). Out of this pool, we selected 2831 records, including both raw (94.2%) and digitized (5.8%) data, for which standardized chronologies could be established (Li et al., 2022).

2.2 Metadata processing

After checking the metadata of all records from the Neotoma and Asian datasets, we implemented the following modifications: 1) we evaluated the units of the provided depth information (meter to centimeter) of all records and contacted Neotoma to correct the depth information of one record (Dataset-ID 27027); 2) we checked each record's archive type (e.g., peat, lake) based on its site description from Neotoma or original publication; and 3) we integrated two records (Dataset-ID 835, 3127) into a combined record (Dataset-ID 70001).

We collected the sample ages from the chronologies provided by Li et al. (2022), which were newly established for all 2831 records using a standardized approach. Li et al. (2022) present estimated ages for each centimeter. For those records with sample depth at a sub-centimeter scale, we applied a linear interpolation to assign ages for each sample, performed in R (R Core Team, 2020).

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

2.3 Pollen data processing

2.3.1 Pollen taxa harmonization

Only terrestrial pollen taxa (including *Cyperaceae*) were taken into account whilst excluding aquatic pollen taxa as well as spores from mosses, ferns, fungi, and algae. First, we standardized the taxon nomenclature. We set up a master table containing all pollen taxa names from the 2831 records and made names consistent (e.g., '*betula*' to '*Betula*'), italics for all taxa under family level (e.g., '*Artemisia*' to '*Artemisia*'), abbreviation (e.g., '*P. pumila*' to '*Pinus pumila*'), synonym (e.g., 'Gramineae' to 'Poaceae'), wrong spelling (e.g., '*Aluns*' to '*Alnus*'). This master table is published in a machine-readable data format on PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.929773; in the "*Further details*" section; Herzschuh et al., 2021a). Second, we harmonized the pollen taxa according to the classification of the *Angiosperm Phylogeny Group IV* system, a modern molecular-based flowering plant taxonomy system (The Angiosperm Phylogeny Group et al., 2016). Woody taxa were harmonized to genus level as well as some very common herbaceous taxa such as *Artemisia*, *Thalictrum,* and *Rumex*. All other herbaceous taxa were harmonized to the family level. The various pollen taxa of heather plants were summarized at the order level as Ericales.

2.3.2 Pollen data type standardization

Although most pollen records contain the count data (in the following named the 'raw' data), the 'pollen counts' for those without raw pollen counts were back-calculated using the pollen percentages and assuming a terrestrial pollen sum of 300 pollen grains, as most of the publications do not provide a pollen sum. Alternatively, the back-calculation of the pollen sum could be based on more elaborated methods (e.g., the *countSum* R-function (https://github.com/richardjtelford/countSum). We replaced the original taxon name with its harmonized name and summed up all counts of the harmonized taxa for each sample. As we only consider terrestrial plant taxa, some samples in records may contain no pollen counts, and those samples were excluded from the harmonized dataset. We then recalculated the terrestrial pollen percentages for each sample based on their total sum.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

3. Structure of the LegacyPollen 1.0 dataset

3.1. Structure of metadata

The LegacyPollen 1.0 metadata of 2831 records are provided for each pollen sample. These include the dataset identifier (ID) (LegacyPollen 1.0), event name (mostly equivalent to the Neotoma or sample name codes), if available, site ID (in the source datasets), data source, site name, geographical coordinates, site description (from original publication/Neotoma), archive type (e.g., peat, lake sediment core), source of data, and pollen data type (raw counts/percentages). All site-specific metadata are available at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.929773; Herzschuh et al., 2021a) in the "*Further details*" section ("*Description of sampling sites*"). Sample-specific metadata including depth, sample age (according to Li et al., 2022; minimum age, maximum age, mean age, median age) are provided in the pollen data files at PANGAEA.

3.2 Structure of pollen data

The LegacyPollen 1.0 dataset contains one pollen sample in a row and 1002 harmonized taxon names in columns. To ease data handling, data files were separated for pollen count data and pollen percentages and files for each region (Western North America, Eastern North America, Europe, Asia, Latin America, Africa, and Indo-Pacific) are provided separately in both .CSV and TXT format. In total, 28 pollen data files are published at PANGAEA (https://doi.pangaea.de/10.1594/PANGAEA.929773; Herzschuh et al., 2021a) and can be joined by the dataset ID with other data products. Furthermore, we also provide the taxa harmonization table at PANGAEA.

4. Dataset assessment

4.1 Spatial and temporal coverage of the dataset

Of the 2831 records included in LegacyPollen 1.0, 670 records originate from Eastern North America (<105°W; Williams et al., 2000), 362 from Western North America, 1075 from Europe, 488 from Asia, 150 from Latin America, 54 from Africa, and 32 from the Indo-Pacific (Fig. 1). Most records (2659

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

records, 93.9%) are in the Northern Hemisphere, where the main vegetation and climate zones are covered.

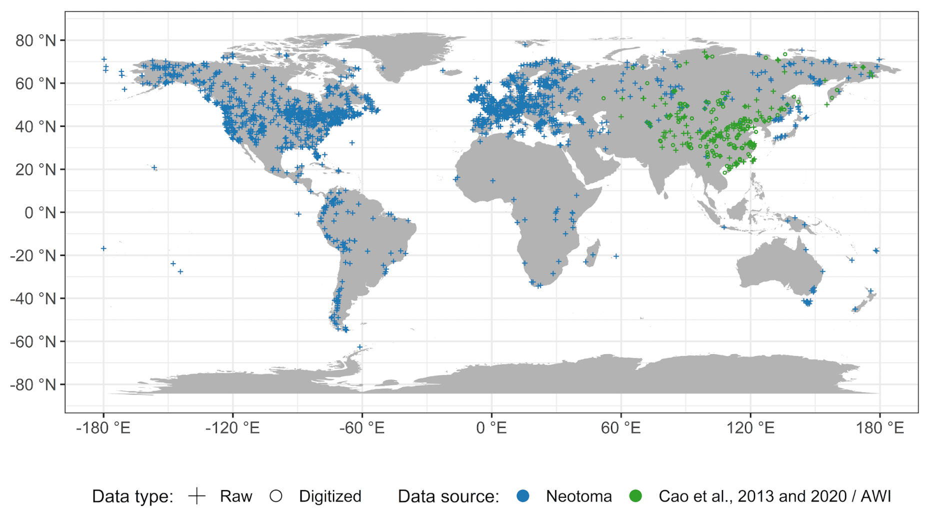


Figure 1. Map of the 2831 records for which standardized chronologies were established by source and data type.

As shown in Fig. 2, only 5.8% of the records are available from periods before the Last Glacial Maximum (>26.5 cal ka BP), 10.2% cover part of the Last Glacial Maximum (26.5-19.0 cal ka BP; Clark et al., 2009), and 45.7% cover part of the Last Deglaciation (ca. 19.0-11.7 cal. ka BP; Clark et al., 2012). Almost all records (97.8%) cover part of the Holocene among them 65.2, 79.5, 89.5% cover the early Holocene (11.7-8.2 cal. ka BP), middle Holocene (8.2-4.2 cal. ka BP), and late Holocene (4.2-0 cal. ka BP), respectively.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

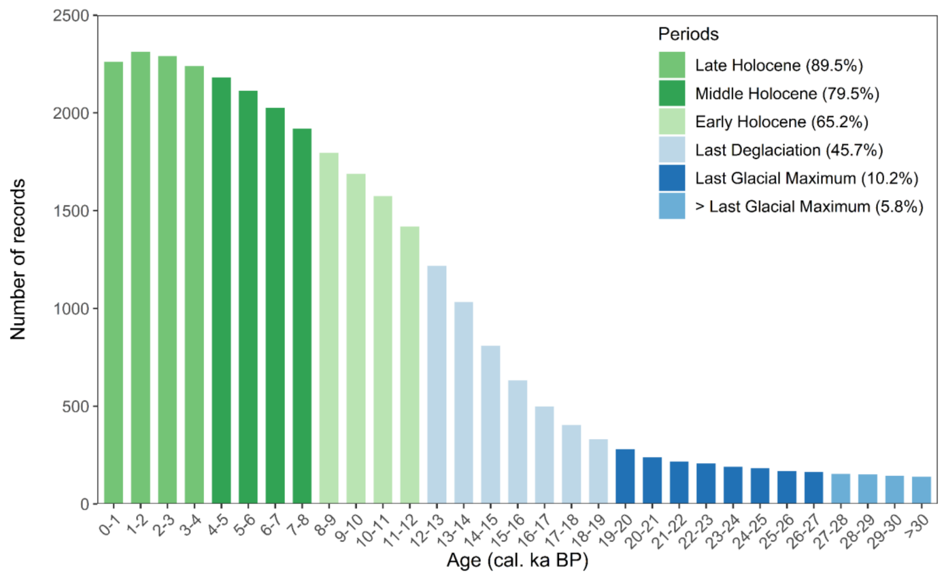


Figure 2. Histogram showing the number of available records in distinct time slices.

4.2 Harmonized taxonomy

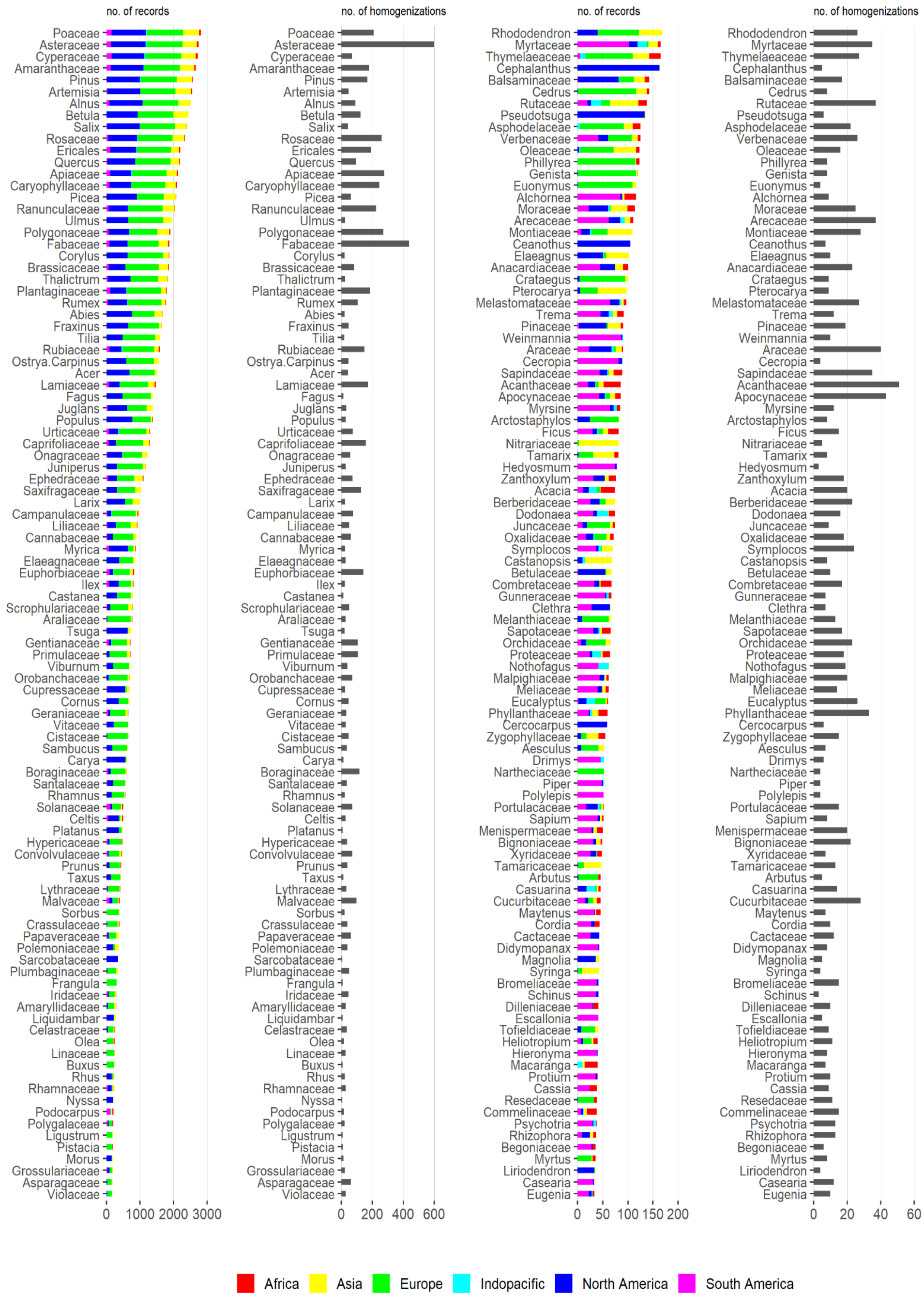
A total of 10,110 terrestrial pollen taxa or taxa notations were obtained from the 2831 records, which we condensed to 1002 families or genera through taxonomic harmonization (Fig. 3; Appendix Fig. 1). On average, 10.8 original taxa or taxa notations are covered by one harmonized pollen taxon, ranging from 1 to 599 (median: 2). Overall, Asteraceae (599), Fabaceae (437), and Apiaceae (276) are the pollen taxa with most variants.

The biggest difference in taxa names and notations before and after harmonization can be found in Europe with a mean of 42 variants per harmonized taxon and in Eastern and Western North America (average of 22) with both continents also exhibiting the highest record density (Fig. 4). A high amount of tropical and subtropical tree and shrub taxa can be found in the Southern Hemisphere, which are harmonized to genus level and therefore subsume to fewer harmonized taxa, and overall have a higher taxa diversity than the Northern Hemisphere continents. In the Southern Hemisphere, the most taxa and variants are harmonized for Fabaceae as this is the most common family found in tropical rainforests and dry forests of Latin America and Africa.

Europe has the most harmonizations of herbaceous taxa from open landscapes: e.g., Asteraceae, Apiaceae, or Caryophyllaceae. In North America and Asia, several species or species groups of major

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

woody taxa are harmonized to their respective genus level, e.g., *Alnus* and *Acer* in North America, or *Betula* and *Quercus* in Asia. The *Pinus Haploxylon* and *Diploxylon* subgenera are subsumed into the genus level *Pinus*, as the differentiation to subgenera level is not provided consistently.



https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Figure 3. Number of records with taxa occurrences (per continent) and number of subsumed variants per harmonized taxon. The figure shows the top 200 taxa with the highest number of records in the dataset. A full overview of all taxa is given in Appendix Fig. 1.

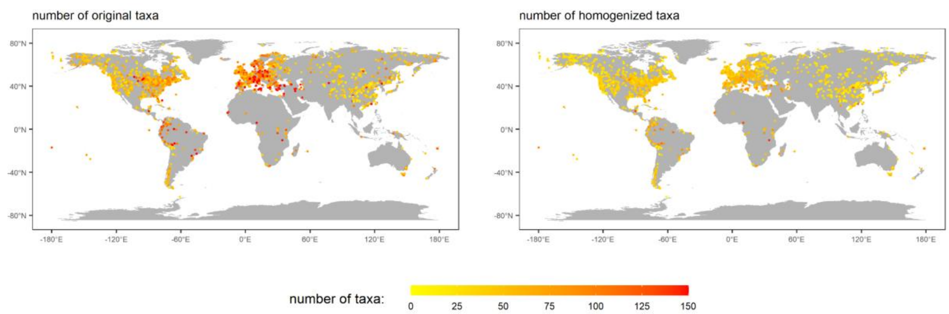


Figure 4. Number of taxa before and after harmonization (number of taxa > 150 were all grouped into the class of 150).

5. Discussion

5.1 Quality of the LegacyPollen 1.0 dataset

To our knowledge, LegacyPollen 1.0 is the largest harmonized fossil pollen dataset including more than twice the number of records included in previously published datasets (e.g., Fyfe et al. (2009): 1032 records; Trondman et al. (2015): 636 records; Marsicek et al. (2018): 642 records; Giesecke et al. (2019): 749 records; Mottl et al. (2021): 1181 records; Githumbi et al. (2021): 1128 records). Several regions have poor pollen-record coverage either because no records are available due to the scarcity of suitable archives (e.g., continental interiors) or because available records were not compiled and integrated into Neotoma. Ongoing initiatives on compilation of pollen data from Africa and Latin America will allow a straightforward extension of the LegacyPollen 1.0 dataset using the provided framework.

Representing a further advantage, the LegacyPollen 1.0 dataset is accompanied by consistent metadata allowing for subsetting of the dataset. Aside from information about the location and archive type, the metadata also includes sample ages that were inferred from recently revised chronologies (Li et al., 2022) along with their age uncertainties (i.e., output from BACON; Blaauw and Christen, 2011) and the framework and R code also allows a customized reestablishment of the age-depth models.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

Generally, temporal coverage is good since about 14 cal. ka BP. Rather few records cover the glacial period, which is mainly due to an absence of archives as many lakes and peatlands were dry, or covered by ice-sheets. Many Asian records cover the Marine Isotope Stage 3 compared with Europe and North America.

Taxonomic harmonization is required for multi-site synthesis studies (Fyfe et al., 2009; Trondman et al., 2015; Marsicek et al., 2018; Herzschuh et al., 2019; Routson et al., 2019; Githumbi et al., 2021; Mottl et al., 2021; Zheng et al., 2021). This is particularly true when numerical approaches are applied that measure compositional dissimilarity between pollen spectra, for example, between fossil and modern sites for climate reconstructions using the Modern Analogue Technique or regression methods, or among fossil records for beta-diversity studies (Birks et al., 2012). If taxa are not harmonized, an inferred high dissimilarity between two spectra may originate just from differences in taxa nomenclature. On the other hand, if all taxa are harmonized to a too high taxonomic level, the ecological signal might be lost (Giesecke et al., 2019). We applied an intermediate level of harmonization taking growth-form (i.e., woody vs. non-woody) as additional guidance. We assume that our approach best reflects the typical presentation of pollen data which is mainly limited by the pollen morphological features visible at 400x magnification using light microscopy and the typical precision in taxa identification of most pollen analysts.

5.2 Potential uses of LegacyPollen 1.0

LegacyPollen 1.0 can be used for a variety of palaeoenvironmental synthesis studies including reconstructions of taxa distributions, climate, and biome change, which can be used for palaeo-model validation (Gaillard et al., 2010; Cao et al., 2013; Trondman et al., 2015; Cao et al., 2020; Mottl et al., 2021).

Plant taxa distribution changes based on mapping of pollen taxa can yield information about glacial refugia and past migration patterns, as, for example, previously implemented for *Quercus* (Brewer et al., 2002), *Picea* (van der Knaap et al., 2005; Zhou and Li, 2012), *Larix* (Cao et al., 2020), east Asian tree taxa (Cao et al., 2015), and European broad-leaf forest (Woodbridge et al., 2014; Fyfe et al., 2015). With the establishment of LegacyPollen 1.0, a Northern Hemisphere-wide analysis of past changes in distributional ranges is now possible, as would help, for example, to better understand the different post-glacial colonization patterns of *Larix* in Europe, North America, and Siberia (Herzschuh, 2020).

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Such understanding of past range changes can underpin conservation management via the use of species distribution modeling at a broad scale enhanced by the higher spatial resolution and larger extent of LegacyPollen 1.0.

Studies aiming at broad-scale pollen-based vegetation reconstructions can benefit from the harmonized LegacyPollen 1.0 dataset including via biomization approaches (Prentice et al., 1996), multi-site ordination or classification approaches (e.g., two-way indicator species analysis; Hill, 1996; Fletcher and Thomas, 2007; Connor and Kvavadze, 2009), or approaches relating modern to fossil datasets (e.g., Modern Analogue Technique; Overpeck et al., 1985). Furthermore, quantitative vegetation reconstructions (e.g., Regional Estimates of Vegetation Abundance from Large Sites (REVEALS) model; Sugita, 2007) can be easily implemented, as a synthesis of relative pollen productivity estimates is already available for the Northern Hemisphere (Wieczorek and Herzschuh, 2020). Such quantitative information about taxa covers changes that can be directly compared to vegetation model outputs (Dallmeyer et al., 2021) at regional to continental scales, which is a potentially more accurate approach than translating pollen and model outputs first biomes (Cao et al., 2019).

Pollen-based climate reconstructions are the backbone of palaeoclimate synthesis studies for the continents (Marcott et al., 2013; Marsicek et al., 2018; Routson et al., 2019; Kaufman et al., 2020a, b). The reconstruction of mean annual temperature (T_{ann}), mean annual precipitation (P_{ann}), and mean temperature of July (T_{July}) using LegacyPollen 1.0 as input is an ongoing project (Herzschuh et al., 2021b). This will substantially increase the number of records and close data gaps in the global temperature datasets and thus enable the evaluation of climate simulations at a hemispheric scale (Wu et al., 2013; Hao et al., 2019). It will contribute to the “Holocene conundrum” debate (Liu et al., 2014) and to the discussion of the relationship between temperature and precipitation change (Trenberth, 2011; Routson et al., 2019).

Human activities are an important driver of vegetation change in addition to climate and other natural forces (Ellis and Ramankutty, 2008; Mottl et al., 2021; Pavlik et al., 2021). At the Holocene time-scale, deforestation is of particular relevance which, with the help of the LegacyPollen 1.0 dataset, can now be investigated at the hemispheric scale. The harmonized chronologies of the LegacyPollen 1.0 dataset allow for the analysis of similarities and dissimilarities in the temporal pattern of deforestation between continents.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

6 Data and code availability

The data are published in the PANGAEA repository under https://doi.pangaea.de/10.1594/PANGAEA.929773 in the "*Other version*" section (Herzschuh et al., 2021a) in both comma-separated values (.CSV) and tab-delimited text (.TXT) formats for LegacyPollen 1.0 dataset of counts per continent and LegacyPollen 1.0 dataset of percentages per continent. Site and pollen metadata, as well as a taxa harmonization master table, are provided in the "*Further details*" section.

The R code for taxa harmonization is stored on Zenodo (https://doi.org/10.5281/zenodo.5910972; Herzschuh et al., 2022) along with an example dataset. Downloading pollen data from the Neotoma Paleoecology Database, harmonizing the pollen taxa, and assigning ages to sample depth data to create customized datasets can thus be easily done.

Author contributions: UH had the idea, set up the implementation plan, led the study and wrote a first version of the manuscript together with CL and TB. CL, TB, AP implemented the harmonization supervised by UH and AA. BH and MW supervised the setup of the dataset and its upload to the repository and documentation. All authors contributed to the final version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to express our gratitude to all the palynologists and geologists who, either directly or indirectly by providing their work to the Neotoma Paleoecology Database, contributed pollen data and chronologies to the dataset. The work of data contributors, data stewards, and the Neotoma community is gratefully acknowledged. We also thank Cathy Jenks for language editing on a previous version of the paper.

Financial support. This research has been supported by the PalMod Initiative (01LP1510C to UH) and the European Research Council (ERC Glacial Legacy 772852 to UH). TB is supported by the German Federal Ministry of Education and Research (BMBF) as a Research for Sustainability initiative (FONA; https://www.fona.de/en) through the PalMod Phase II project (grant no. FKZ: 01LP1926D). CL holds a scholarship from the Chinese Scholarship Council (grant no. 201908130165).

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

References

Birks, H. J. B., Lotter, A. F., Juggins, S., and Smol, J. P. (eds): Tracking Environmental Change Using Lake Sediments (Vol. 5): Data Handling and Numerical Techniques, Springer Science & Business Media, 751 pp., 2012.

Blaauw, M. and Christen, J. A.: Flexible paleoclimate age-depth models using an autoregressive gamma process, Bayesian Anal., 6, 457–474, https://doi.org/10.1214/11-BA618, 2011.

Brewer, S., Cheddadi, R., de Beaulieu, J. L., and Reille, M.: The spread of deciduous *Quercus* throughout Europe since the last glacial period, For. Ecol. Manag., 156, 27–48, https://doi.org/10.1016/S0378-1127(01)00646-6, 2002.

Cao, X., Herzschuh, U., Ni, J., Zhao, Y., and Böhmer, T.: Spatial and temporal distributions of major tree taxa in eastern continental Asia during the last 22,000 years, Holocene, 25, 79–91, https://doi.org/10.1177/0959683614556385, 2015.

Cao, X., Ni, J., Herzschuh, U., Wang, Y., and Zhao, Y.: A late Quaternary pollen dataset from eastern continental Asia for vegetation and climate reconstructions: Set up and evaluation, Rev. Palaeobot. Palynol., 194, 21–37, https://doi.org/10.1016/j.revpalbo.2013.02.003, 2013.

Cao, X., Tian, F., Andreev, A., Anderson, P. M., Lozhkin, A. V., Bezrukova, E., Ni, J., Rudaya, N., Stobbe, A., Wieczorek, M., and Herzschuh, U.: A taxonomically harmonized and temporally standardized fossil pollen dataset from Siberia covering the last 40 kyr, Earth Syst. Sci. Data, 12, 119–135, https://doi.org/10.5194/essd-12-119-2020, 2020.

Cao, X., Tian, F., Dallmeyer, A., and Herzschuh, U.: Northern Hemisphere biome changes (>30°N) since 40 cal ka BP and their driving factors inferred from model-data comparisons, Quat. Sci. Rev., 220, 291–309, https://doi.org/10.1016/j.quascirev.2019.07.034, 2019.

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., and McCabe, A. M.: The Last Glacial Maximum, Science, https://doi.org/10.1126/science.1172873, 2009.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Clark, P. U., Shakun, J. D., Baker, P. A., Bartlein, P. J., Brewer, S., Brook, E., Carlson, A. E., Cheng, H., Kaufman, D. S., Liu, Z., Marchitto, T. M., Mix, A. C., Morrill, C., Otto-Bliesner, B. L., Pahnke, K., Russell, J. M., Whitlock, C., Adkins, J. F., Blois, J. L., Clark, J., Colman, S. M., Curry, W. B., Flower, B. P., He, F., Johnson, T. C., Lynch-Stieglitz, J., Markgraf, V., McManus, J., Mitrovica, J. X., Moreno, P. I., and Williams, J. W.: Global climate evolution during the last deglaciation, PNAS, 109, E1134–E1142, https://doi.org/10.1073/pnas.1116619109, 2012.

Connor, S. E. and Kvavadze, E. V.: Modelling late Quaternary changes in plant distribution, vegetation and climate using pollen data from Georgia, Caucasus, J. Biogeogr., 36, 529–545, https://doi.org/10.1111/j.1365-2699.2008.02019.x, 2009.

Dallmeyer, A., Claussen, M., Lorenz, S. J., Sigl, M., Toohey, M., and Herzschuh, U.: Holocene vegetation transitions and their climatic drivers in MPI-ESM1.2, Clim. Past, 17, 2481–2513, https://doi.org/10.5194/cp-17-2481-2021, 2021.

Ellis, E. C. and Ramankutty, N.: Putting people in the map: anthropogenic biomes of the world, Front. Ecol. Environ., 6, 439–447, https://doi.org/10.1890/070062, 2008.

Fletcher, M.-S. and Thomas, I.: Holocene vegetation and climate change from near Lake Pedder, south-west Tasmania, Australia, J. Biogeogr., 34, 665–677, https://doi.org/10.1111/j.1365-2699.2006.01659.x, 2007.

Fyfe, R. M., de Beaulieu, J.-L., Binney, H., Bradshaw, R. H. W., Brewer, S., Le Flao, A., Finsinger, W., Gaillard, M.-J., Giesecke, T., Gil-Romera, G., Grimm, E. C., Huntley, B., Kunes, P., Kühl, N., Leydet, M., Lotter, A. F., Tarasov, P. E., and Tonkov, S.: The European Pollen Database: past efforts and current activities, Veg. Hist. Archaeobot., 18, 417–424, https://doi.org/10.1007/s00334-009-0215-9, 2009.

Fyfe, R. M., Woodbridge, J., and Roberts, N.: From forest to farmland: pollen-inferred land cover change across Europe using the pseudobiomization approach, Glob. Chang. Biol., 21, 1197–1212, https://doi.org/10.1111/gcb.12776, 2015.

Gaillard, M.-J., Sugita, S., Mazier, F., Trondman, A.-K., Broström, A., Hickler, T., Kaplan, J. O., Kjellström, E., Kokfelt, U., Kuneš, P., Lemmen, C., Miller, P., Olofsson, J., Poska, A., Rundgren, M., Smith, B., Strandberg, G., Fyfe, R., Nielsen, A. B., Alenius, T., Balakauskas, L., Barnekow, L., Birks, H. J. B., Bjune, A., Björkman, L., Giesecke, T., Hjelle, K., Kalnina, L., Kangur, M., van der Knaap, W. O., Koff, T., Lagerås, P., Latałowa, M., Leydet, M., Lechterbeck, J., Lindbladh, M., Odgaard, B., Peglar, S., Segerström, U., von Stedingk, H., and Seppä, H.: Holocene land-cover reconstructions for studies on land cover-climate feedbacks, Clim. Past, 6, 483–499, https://doi.org/10.5194/cp-6-483-2010, 2010.

Gajewski, K.: The Global Pollen Database in biogeographical and palaeoclimatic studies, Prog. Phys. Geogr., 32, 379–402, https://doi.org/10.1177/0309133308096029, 2008.

Giesecke, T., Wolters, S., van Leeuwen, J. F. N., van der Knaap, P. W. O., Leydet, M., and Brewer, S.: Postglacial change of the floristic diversity gradient in Europe, Nat. Commun., 10, 5422, https://doi.org/10.1038/s41467-019-13233-y, 2019.

Githumbi, E., Fyfe, R., Gaillard, M.-J., Trondman, A.-K., Mazier, F., Nielsen, A.-B., Poska, A., Sugita, S., Theuerkauf, M., Woodbridge, J., Azuara, J., Feurdean, A., Grindean, R., Lebreton, V., Marquer, L., Nebout-Combourieu, N., Stancikaite, M., Tanţău, I., Tonkov, S., Shumilovskikh, L., and the LandClimII Data Contributors: European pollen-based REVEALS land-cover reconstructions for the Holocene: methodology, mapping and potentials, Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2021-269, in review, 2021.

Goring, S. J., Simpson, G. L., Marsicek, J. P., Ram, K., and Sosalla, K.: Neotoma: access to the Neotoma Paleoecological Database through R, R package version 1.7.4, https://CRAN.R-project.org/package=neotoma, 2019.

Hao, Z., Phillips, T. J., Hao, F., and Wu, X.: Changes in the dependence between global precipitation and temperature from observations and model simulations, Int. J. Climatol., 39, 4895–4906, https://doi.org/10.1002/joc.6111, 2019.

Herzschuh, U., Böhmer, T., Li, C., Cao, X., Heim, B., and Wieczorek, M.: Global taxonomically harmonized pollen data collection with revised chronologies (LegacyPollen 1.0), PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.929773, 2021a.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Herzschuh, U., Böhmer, T., Li, C., and Cao, X.: Northern Hemisphere temperature and precipitation reconstruction from taxonomically harmonized pollen data set with revised chronologies using WA-PLS and MAT (LegacyClimate 1.0), PANGAEA, https://doi.pangaea.de/10.1594/PANGAEA.930512, 2021b.

Herzschuh, U., Cao, X., Laepple, T., Dallmeyer, A., Telford, R. J., Ni, J., Chen, F., Kong, Z., Liu, G., Liu, K.-B., Liu, X., Stebich, M., Tang, L., Tian, F., Wang, Y., Wischnewski, J., Xu, Q., Yan, S., Yang, Z., Yu, G., Zhang, Y., Zhao, Y., and Zheng, Z.: Position and orientation of the westerly jet determined Holocene rainfall patterns in China, Nat. Commun., 10, 2376, https://doi.org/10.1038/s41467-019-09866-8, 2019.

Herzschuh, U.: Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests, Glob. Ecol. Biogeogr., 29, 198–206, https://doi.org/10.1111/geb.13018, 2020.

Herzschuh, U., Li, C., Böhmer, T., Postl, A. K., Heim, B., Andreev, A. A., Cao, X., and Wieczorek, M.: LegacyPollen 1.0: A taxonomically harmonized global Late Quaternary pollen dataset of 2831 records with standardized chronologies, Zenodo, https://doi.org/10.5281/zenodo.5910972, 2022.

Hill, T. R.: Description, classification and ordination of the dominant vegetation communities, Cathedral Peak, KwaZulu-Natal Drakensberg, S. Afr. J. Bot., 62, 263–269, https://doi.org/10.1016/S0254-6299(15)30655-4, 1996.

Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., Axford, Y., Brussel, T., Cartapanis, O., Chase, B., Dawson, A., de Vernal, A., Engels, S., Jonkers, L., Marsicek, J., Moffa-Sánchez, P., Morrill, C., Orsi, A., Rehfeld, K., Saunders, K., Sommer, P. S., Thomas, E., Tonello, M., Tóth, M., Vachula, R., Andreev, A., Bertrand, S., Biskaborn, B., Bringué, M., Brooks, S., Caniupán, M., Chevalier, M., Cwynar, L., Emile-Geay, J., Fegyveresi, J., Feurdean, A., Finsinger, W., Fortin, M.-C., Foster, L., Fox, M., Gajewski, K., Grosjean, M., Hausmann, S., Heinrichs, M., Holmes, N., Ilyashuk, B., Ilyashuk, E., Juggins, S., Khider, D., Koinig, K., Langdon, P., Larocque-Tobler, I., Li, J., Lotter, A., Luoto, T., Mackay, A., Magyari, E., Malevich, S., Mark, B., Massaferro, J., Montade, V., Nazarova, L., Novenko, E., Pařil, P., Pearson, E., Peros, M., Pienitz, R., Płóciennik, M., Porinchu, D., Potito, A., Rees, A., Reinemann, S., Roberts, S., Rolland, N., Salonen, S., Self, A., Seppä, H., Shala, S., St-Jacques, J.-M., Stenni, B., Syrykh, L., Tarrats, P., Taylor, K., van den Bos,

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

V., Velle, G., Wahl, E., Walker, I., Wilmshurst, J., Zhang, E., and Zhilich, S.: A global database of Holocene paleotemperature records, Sci. Data, 7, 115, https://doi.org/10.1038/s41597-020-0445-3, 2020a.

Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P. S., Heiri, O., and Davis, B.: Holocene global mean surface temperature, a multi-method reconstruction approach, Sci. Data, 7, 201, https://doi.org/10.1038/s41597-020-0445-3, 2020b.

Li, C., Postl, A. K., Böhmer, T., Cao, X., Dolman, A. M., and Herzschuh, U.: Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0), Earth Syst. Sci. Data Discuss. [preprint], https://doi.org/10.5194/essd-2021-212, in review, 2022.

Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann, G., Zheng, W., and Timm, O. E.: The Holocene temperature conundrum, PNAS, 111, E3501–E3505, https://doi.org/10.1073/pnas.1407229111, 2014.

Marcott, S. A., Shakun, J. D., Clark, P. U., and Mix, A. C.: A reconstruction of regional and global temperature for the past 11,300 years, Science, https://doi.org/10.1126/science.1228026, 2013.

Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., and Brewer, S.: Reconciling divergent trends and millennial variations in Holocene temperatures, Nature, 554, 92–96, https://doi.org/10.1038/nature25464, 2018.

Mottl, O., Flantua, S. G. A., Bhatta, K. P., Felde, V. A., Giesecke, T., Goring, S., Grimm, E. C., Haberle, S., Hooghiemstra, H., Ivory, S., Kuneš, P., Wolters, S., Seddon, A. W. R., and Williams, J. W.: Global acceleration in rates of vegetation change over the past 18,000 years, Science, 372, 860–864, https://doi.org/10.1126/science.abg1685, 2021.

Overpeck, J. T., Webb, T., and Prentice, I. C.: Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs, Quat. Res., 23, 87–108, https://doi.org/10.1016/0033-5894(85)90074-2, 1985.

Pavlik, B. M., Louderback, L. A., Vernon, K. B., Yaworsky, P. M., Wilson, C., Clifford, A., and Codding, B. F.: Plant species richness at archaeological sites suggests ecological legacy of Indigenous

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

subsistence on the Colorado Plateau, PNAS, 118, e2025047118, https://doi.org/10.1073/pnas.2025047118, 2021.

Prentice, I. C., Guiot, J., Huntley, B., Jolly, D., and Cheddadi, R.: Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka, Clim. Dyn., 12, 185–194, https://doi.org/10.1007/BF00211617, 1996.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available online at: https://www.R-project.org/, 2020.

Routson, C. C., McKay, N. P., Kaufman, D. S., Erb, M. P., Goosse, H., Shuman, B. N., Rodysill, J. R., and Ault, T.: Mid-latitude net precipitation decreased with Arctic warming during the Holocene, Nature, 568, 83–87, https://doi.org/10.1038/s41586-019-1060-3, 2019.

Sugita, S.: Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition, Holocene, 17, 229–241, https://doi.org/10.1177/0959683607075837, 2007.

The Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., and Stevens, P. F.: An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV, Bot. J. Linn. Soc., 181, 1–20, https://doi.org/10.1111/boj.12385, 2016.

Trenberth, K. E.: Changes in precipitation with climate change, Clim. Res., 47, 123–138, https://doi.org/10.3354/cr00953, 2011.

Trondman, A.-K., Gaillard, M.-J., Mazier, F., Sugita, S., Fyfe, R., Nielsen, A. B., Twiddle, C., Barratt, P., Birks, H. J. B., Bjune, A. E., Björkman, L., Broström, A., Caseldine, C., David, R., Dodson, J., Dörfler, W., Fischer, E., van Geel, B., Giesecke, T., Hultberg, T., Kalnina, L., Kangur, M., van der Knaap, P., Koff, T., Kuneš, P., Lagerås, P., Latałowa, M., Lechterbeck, J., Leroyer, C., Leydet, M., Lindbladh, M., Marquer, L., Mitchell, F. J. G., Odgaard, B. V., Peglar, S. M., Persson, T., Poska, A., Rösch, M., Seppä, H., Veski, S., and Wick, L.: Pollen-based quantitative reconstructions of Holocene regional vegetation cover (plant-functional types and land-cover types) in Europe suitable for climate modelling, Glob. Chang. Biol., 21, 676–697, https://doi.org/10.1111/gcb.12737, 2015.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

van der Knaap, W. O., van Leeuwen, J. F. N., Finsinger, W., Gobet, E., Pini, R., Schweizer, A., Valsecchi, V., and Ammann, B.: Migration and population expansion of *Abies*, *Fagus*, *Picea*, and *Quercus* since 15000 years in and across the Alps, based on pollen-percentage threshold values, Quat. Sci. Rev., 24, 645–680, https://doi.org/10.1016/j.quascirev.2004.06.013, 2005.

Wieczorek, M. and Herzschuh, U.: Compilation of relative pollen productivity (RPP) estimates and taxonomically harmonized RPP datasets for single continents and Northern Hemisphere extratropics, Earth Syst. Sci. Data, 12, 3515–3528, https://doi.org/10.5194/essd-12-3515-2020, 2020.

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth, R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., Nichols, J., Purdum, T., Roth, R. E., Stryker, M., and Takahara, H.: The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource, Quat. Res., 89, 156–177, https://doi.org/10.1017/qua.2017.105, 2018.

Williams, J. W., Webb III, T., Richard, P. H., and Newby, P.: Late Quaternary biomes of Canada and the eastern United States, J. Biogeogr., 27, 585–607, https://doi.org/10.1046/j.1365-2699.2000.00428.x, 2000.

Woodbridge, J., Fyfe, R. M., and Roberts, N.: A comparison of remotely sensed and pollen-based approaches to mapping Europe's land cover, J. Biogeogr., 41, 2080–2092, https://doi.org/10.1111/jbi.12353, 2014.

Wu, R., Chen, J., and Wen, Z.: Precipitation-surface temperature relationship in the IPCC CMIP5 models, Adv. Atmos. Sci., 30, 766–778, https://doi.org/10.1007/s00376-012-2130-8, 2013.

Zheng, Z., Ma, T., Roberts, P., Li, Z., Yue, Y., Peng, H., Huang, K., Han, Z., Wan, Q., Zhang, Y., Zhang, X., Zheng, Y., and Saito, Y.: Anthropogenic impacts on Late Holocene land-cover change and floristic biodiversity loss in tropical southeastern Asia, PNAS, 118, e2022210118, https://doi.org/10.1073/pnas.2022210118, 2021.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

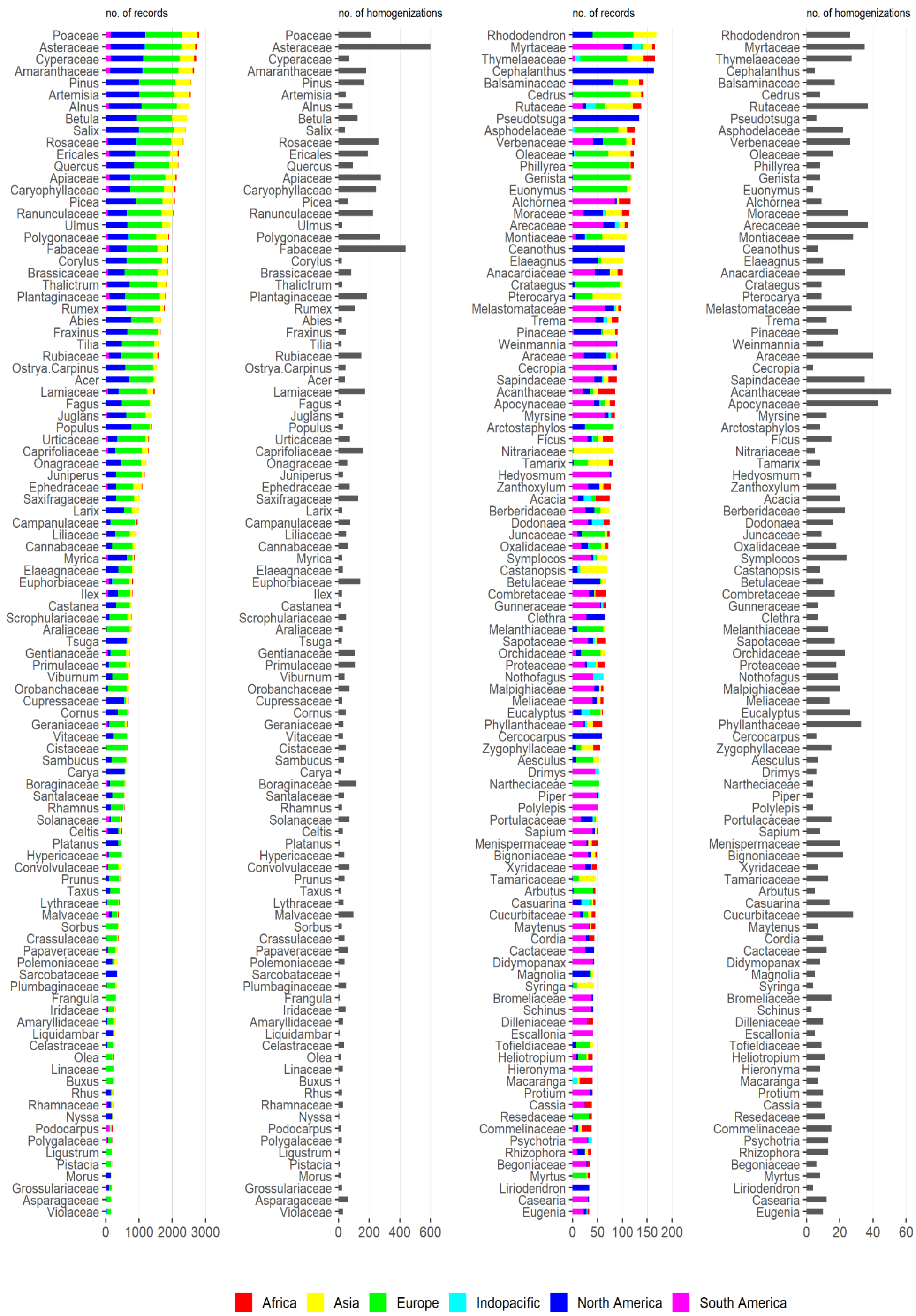
Open Access Earth System Science Data Discussions

455 Zhou, X. and Li, X.: Variations in spruce (*Picea* sp.) distribution in the Chinese Loess Plateau and surrounding areas during the Holocene, Holocene, 22, 687–696, https://doi.org/10.1177/0959683611400195, 2012.

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

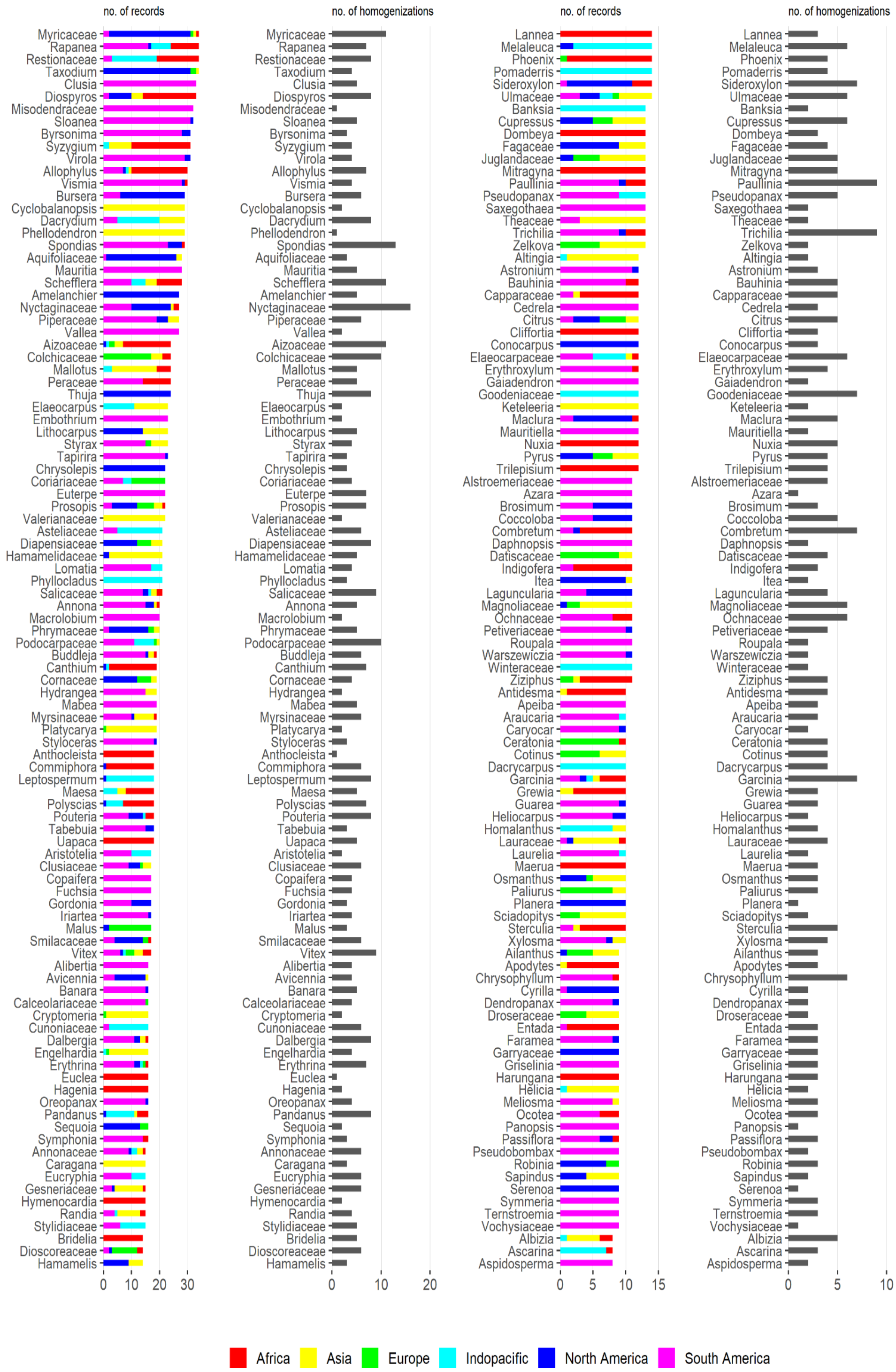
Open Access Earth System Science Data Discussions

Appendix Figures



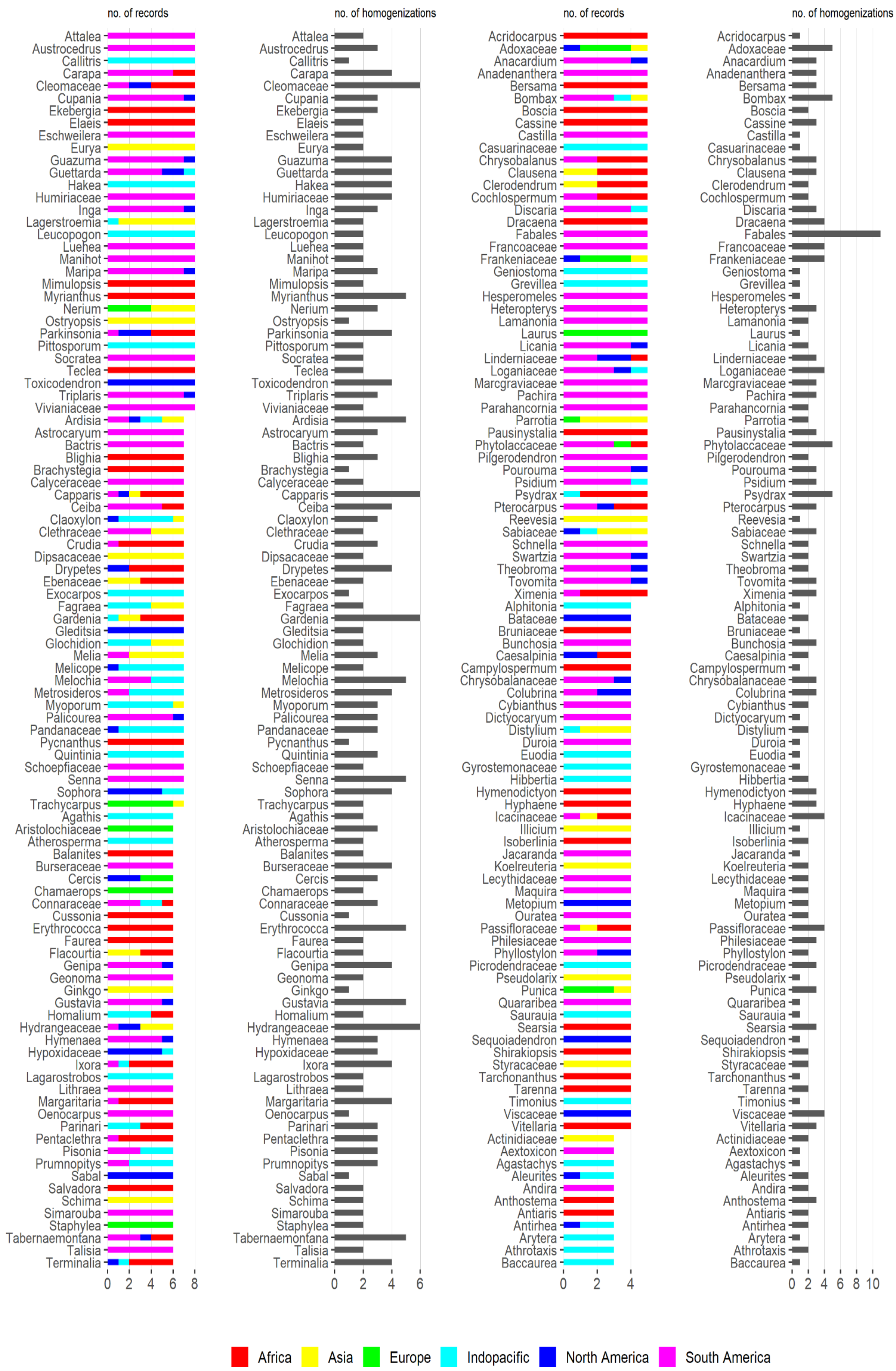
https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions



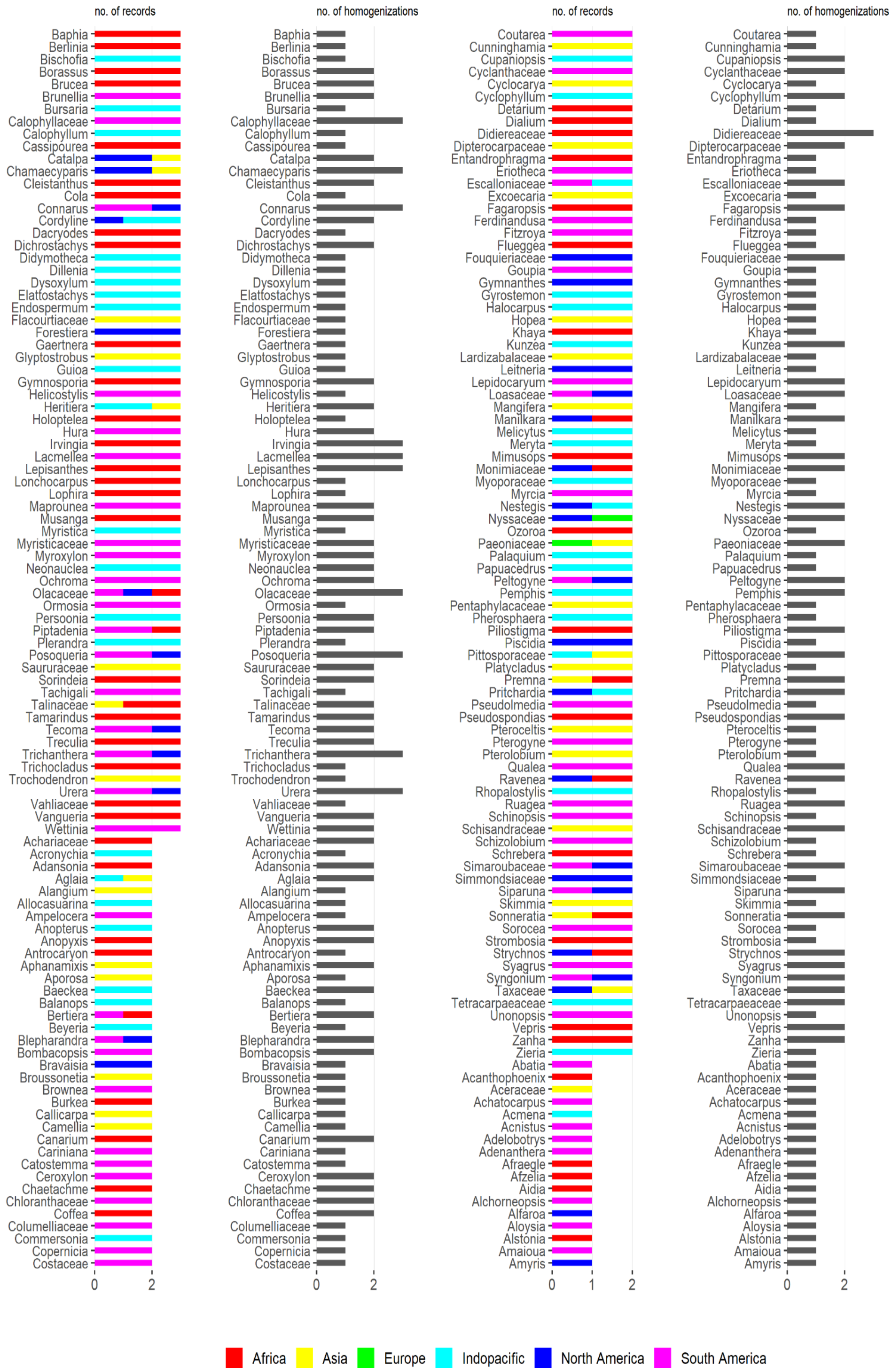
https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions



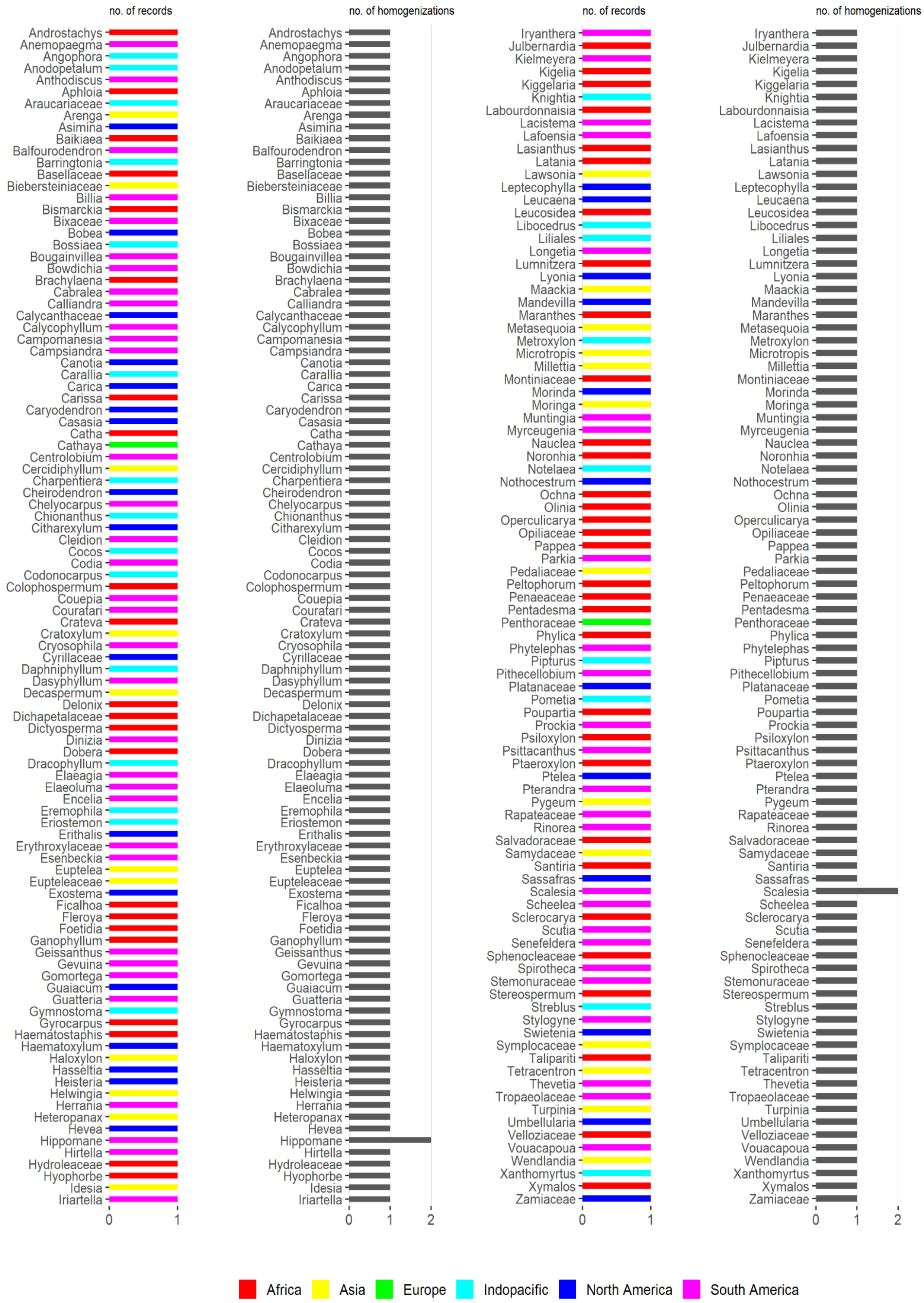
461

https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022



https://doi.org/10.5194/essd-2022-37
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions



Appendix Figure 1 (complete Figure 3). Number of records with taxa occurrences (per continent) and number of harmonizations per taxon (full taxon list).