Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography — A Journal of Macroecology

DATA ARTICLE OPEN ACCESS

A Comprehensive Database of Expert-Curated Occurrences for the Genus *Carex* L. (Cyperaceae)

María Sanz-Arnal[1,2] | Pablo García-Moro[1] | Carmen Benítez-Benítez[3] | Marina Coca-de-la-Iglesia[4] | Angélica Gallego-Narbón[4,5] | Consolación Barciela[1] | Fabrizio Bartolucci[6] | Prabin Bhandari[7,8,9] | Matthew Bradley[10] | Asunción Cano[11] | Antoine Derouaux[12] | Sabina Donadío[13] | Marcial Escudero[3] | Mariela Fabbroni[14] | Kerry A. Ford[15] | Gabriele Galasso[16] | Sebastian Gebauer[17,18,19] | M. Socorro González-Elizondo[20] | David Hamon[21] | Matthias H. Hoffmann[22] | Xiao-Feng Jin[23] | Jacob Koopman[24] | Bangze Li[7,8,9] | Raúl Lois[1,25] | Yi-Fei Lu[21,26] | Modesto Luceño[1] | José Ignacio Márquez-Corro[1] | Santiago Martín-Bravo[1] | Attila Mesterházy[27] | Mónica Míguez[1] | Ana Morales-Alonso[1,2] | A. Muthama Muasya[28] | Paulo Muñoz-Schüler[1,29] | Robert F. C. Naczi[30] | Nora Oleas[31] | Luciana Pereira-Silva[32] | Radomír Řepka[33] | Anton A. Reznicek[34] | Katie K. Sanbonmatsu[7] | Enmily Sánchez[31] | Daniel Spalink[7] | Arne Strid[35] | Pieter Vanormelingen[36] | Filip Verloove[37] | Karen L. Wilson[38] | Okihito Yano[39] | Shuren Zhang[8,9] | Pedro Jiménez-Mejías[1]

Correspondence: María Sanz-Arnal (maria.sanza@urjc.es)

Received: 11 March 2025 | **Revised:** 25 July 2025 | **Accepted:** 4 September 2025

Handling Editor: Anne Bjorkman

Funding: This work was supported by the project PID2020-113897GB-I00 (DANZ, to P.J.-M. and S.M.-B.) and a Ramón y Cajal postdoctoral contract (RYC2021-031238-I) for P.J.-M, both from the Spanish Ministry of Science and Innovation. Also, we have been funded by the Regional Government of Madrid, Spain (Macondo, SI1/PIJ/2019-00333), the Smithsonian Postdoctoral Fellowship program, and a Youth Guarantee contract towards P.G.-M. (PEJ-2020-AI/AMB-18719, Regional Government of Madrid). C.B.-B., M.S.-A., and P. G.-M. were supported by a Predoctoral Fellowship grant from the Ministry of Science, Innovation and Universities (FPU16/01257, PREP2023-C1FPI2024-02) and from the Ministry of Science and Innovation (PRE2021-100655), respectively. A. M.-A. was financed by the Bentham-Moxon Trust grant 2021 (Kew Royal Botanic Gardens, UK), and M.E. was sponsored by the "Beca Iberoamérica. Jóvenes Profesores e Investigadores del Programa Santander Universidades".

Keywords: biodiversity databases | *Carex* | data quality control | expert-curation | GBIF | geographic occurrences | sedges

ABSTRACT

Motivation: Geographic occurrences are essential for biodiversity studies, but publicly available repositories like GBIF often contain errors and biases, especially for taxonomically complex groups like *Carex* L. (Cyperaceae). This work provides an expert-curated global dataset of occurrences compiled from different sources to enhance data accuracy and usability. The final dataset includes 384,067 occurrences of 1790 *Carex* species.

Main Types of Variables Contained: The dataset includes species occurrence records with geographic coordinates, taxonomic identifications, and curation flags (e.g., introduced, erroneous records).

Spatial Location and Grain: The dataset covers a global scale, using the WGS84 projection. Spatial resolution is standardised to a minimum of three decimal degrees (~1 km, if possible).

Time Period and Grain: Online records span from 1950 to 2020, but some manually georeferenced records are earlier (1850). There is also fieldwork data after 2020, specifically up to 2023.

Major Taxa and Level of Measurement: Cyperaceae: *Carex*. Most records have species-level identification, and some of them are identified at subspecies or variety levels.

Software Format: Data are supplied as comma-separated values files with UTF-8 encoding.

For affiliations refer to page 8.

1 | Introduction

Geographic occurrence records constitute an increasingly scientific resource widely used in conservation-related topics, including species management (Palacio et al. 2021; Moreira et al. 2022), biological invasions (Eckert et al. 2020; Santamarina et al. 2023), analyses of diversity (Naczi et al. 2020; Sanbonmatsu and Spalink 2022), and range shifts under climate change (Fourcade 2016; Spalink et al. 2018; Benítez-Benítez et al. 2022). They are also essential for analysing species' bioclimatic niches and their evolutionary dynamics (Benítez-Benítez, Martín-Bravo, et al. 2021; Benítez-Benítez, Otero, et al. 2021; Coca-de-la-Iglesia et al. 2022; Mejía et al. 2022). The primary publicly available resource of occurrence data is the Global Biodiversity Information Facility (GBIF 2023), which compiles information from museum collections, publications, and from projects or institutions. Additionally, it hosts quality observations recovered by community science platforms, such as iNaturalist (https://www.inaturalist.org/) and Observation.org (https://observation.org/). However, GBIF data are often taxonomically and spatially biased and contain errors related to species identifications and locations. Thus, further curation is often required to maximise its utility for scientific studies (Hortal et al. 2007).

The genus *Carex* L. (Cyperaceae) is a challenging group in terms of its systematic structure and species identification. *Carex* records on GBIF and other public repositories frequently contain errors due to several factors. First, *Carex* is remarkably diverse with more than 2000 species, positioning it among the top three largest angiosperm plant genera. This excludes agamosperm genera (those that form seed without prior fertilisation and are therefore conformed by microspecies, Global Carex Group et al. 2021; Govaerts et al. 2022). Second, its intricate taxonomy poses a well-recognised challenge for both professional and amateur botanists alike, and it hampers the reliable identification of specimens by non-experts. This is due to the extremely reduced morphology of key taxonomic characters, and it is aggravated by the recurrent homoplasy across the genus (Global Carex Group et al. 2016): species distantly related lineages may exhibit similar characteristics, making them difficult to distinguish and leading to mislabeling under the same species name, which becomes a 'collective' taxonomic hotchpotch. Third, *Carex* is under constant revision as systematic knowledge advances. The embedding of smaller sedge genera within *Carex* has been known since early molecular studies (Starr et al. 1999; Roalson et al. 2001; Hendrichs, Michalski, et al. 2004; Hendrichs, Oberwinkler, et al. 2004; Global Carex Group et al. 2016). However, it was not until 2015 that *Carex* became monophyletic (Global Carex Group 2015) after formally absorbing the closely related genera (i.e., satellite genera *Cymophyllus* Mack., *Kobresia* Willd., *Schoenoxiphium* Nees, and *Uncinia* Pers). Later phylogenetic and phylogenomic studies confirmed this (Martín-Bravo et al. 2019; Villaverde et al. 2020) and led to a recently revised infrageneric arrangement of subgenera, sections, and informal groups (Global Carex Group et al. 2021). This process of a major rearrangement of *Carex* is far from complete. Phylogenetic studies focusing on unexplored groups will lead to taxonomic revisions and the description of new species. Examples include the section *Rarae* C.B. Clarke (Oda et al. 2019), the subgenera *Psyllophorae* (Degl.) Peterm. (Benítez-Benítez, Otero, et al. 2021), the Macaronesian section *Rhynchocystis* Dumort (Míguez et al. 2021), the *Carex macroglossa* group (Takahashi et al. 2021), and the subgenera *Uncinia* (Pers.) Peterm. (García-Moro et al. 2022). In this context, the Global *Carex* Group (GCG), an international consortium of *Carex* experts, was established in the mid-2010s to foster collaboration and develop a revised systematic framework. Since its founding, the GCG has contributed to several collective publications aimed at improving the evolutionary knowledge and systematic ordination of the genus (Global Carex Group 2015; Global Carex Group et al. 2016; Martín-Bravo et al. 2019; Villaverde et al. 2020; Global Carex Group et al. 2021).

Because of its complexity, using raw *Carex* occurrence data may lead to inaccurate conclusions (e.g., Oleas et al. 2019). In May 2020, GBIF contained 5,738,145 *Carex* occurrence records, corresponding to 2225 species names. With the initial intention of using this dataset to develop a global ecological model of the genus, we identified three main issues: (1) nomenclatural incongruences with the Plants of the World Online database (POWO 2025; 2083 species in 2024), which is the nomenclatural reference the GCG has agreed to utilise; (2) misplaced occurrences for several widely distributed species, likely due to misidentifications, the lack of distinction between native and introduced occurrences, or use of names as collective taxonomic identifiers (e.g., taxonomic confusion of *Carex cespitosa* with the closely related *C. elata* and *C. nigra*; Jiménez-Mejías et al. 2014); and (3) large areas, with considerable *Carex* diversity, exhibited an alarmingly low number of occurrences in the dataset (e.g., Mexico, Central and South America, or E Asia).

Given these issues, this work aims to: (1) compile a master database for all *Carex* occurrences hosted in GBIF until 2020; (2) incorporate additional occurrences from complementary, publicly available databases to cover sampling gaps; (3) conduct a comprehensive curation of the dataset, including standard filtering procedures and expert curation tasks (tagging problematic occurrences and identifying putative introduced ones); (4) manually incorporate occurrences from Mexico, South America, and East Asia to enhance coverage in key geographical areas; and (5) consolidate the resulting information from these multiple resources into a single dataset, facilitating its usability in future studies.

2 | Methods

2.1 | Data Collection

We define occurrences as unique combinations of species and their geolocation. Primary occurrences were obtained from GBIF and Tropicos (https://tropicos.org/). Most of the geographic occurrences of *Carex* species were downloaded from these repositories during May 2020 using the R package *rgbif* (R Core Team 2021; Chamberlain et al. 2022). For 19 *Carex* species exceeding 100,000 records, data had to be downloaded manually. In Tropicos, we searched all the *Carex* taxa via the quick search engine, and species information was retrieved manually, excluding records without coordinates. We checked the "Specimens" tab to identify any available specimens and added all relevant specimen information into the dataset, maintaining all Tropicos columns. All records were manually reviewed to prevent duplicate entries. To complement the GBIF and Tropicos, records from the following nine online databases, already expert curated, were also manually incorporated: the Atlas of Living Australia (ALA, http://www.ala.org.au), the Botanical Research and Herbarium Management System (BRAHMS, https://herbaria.plants.ox.ac.uk/bol/topo), Flora Polski (https://atlas-roslin.pl/), FloraFaunaAltoAdige (https://www.florafauna.it/), iDigBio (https://www.idigbio.org), Database of the Czech Flora and Vegetation (https://pladias.cz/en/), RAINBIO (Dauby et al. 2016), SEINet (https://swbiodiversity.org/seinet), and Waarnemingen.be/Observation.org. For Waarnemingen.be/Observation.org, data quality was ensured by using only occurrence records for which the identification was either approved by the expert plant validators or the automatic image recognition software, or which were done by plant observers considered trustworthy. Then, duplicated occurrences were removed.

Specimen information from specialised literature (see Appendix S1) and herbarium specimens were manually collected and georeferenced. We processed not only specimens with coordinates on their labels, but also older specimens (from 1850 onwards) whose geographic location needed to be placed manually. In these cases, after evaluating satellite images from Google Maps (https://www.google.com/maps), we retained only those with confidently determined coordinates within a 5 km radius. We included revised *Carex* collections from the herbaria listed in Table S1, with emphasis on critical specimens and collections from South America. All georeferenced specimens were identified by at least one coauthor (see Author contributions).

Specimen lists from systematic studies on *Carex* (see repository files) were also included. These datasets were particularly valuable since the specimens listed already had a confident expert identification. For groups prone to misidentification (e.g., *Carex appressa*/*C. virgata*, sects. *Fecundae*, *Junciformes*, *Phacocystis*, *Racemosae*, *Rhynchocystis, Schiedeanae,* and *Schoenoxiphium*), the datasets from these systematic works replaced occurrences from online repositories to improve data reliability.

Finally, we also incorporated data from field research expeditions to South America conducted by several collaborators of this work (see Author contributions). This fieldwork focused on collecting *Carex* and has taken place in the last 10 years, and the specimens are deposited in the Pablo de Olavide University herbarium (UPOS).

2.2 | Data Curation

POWO (POWO 2025) was used as reference for accepted names, and record names were updated when they did not reflect the most recent nomenclature of the species.

Downloaded GBIF-T raw records (from herein GBIF-T raw dataset) were submitted to a quality control procedure based on Coca-de-la-Iglesia et al. (2023). Data were filtered with the R package *dplyr* (Wickham et al. 2022). The following records were removed: (1) erroneous coordinates (e.g., longitude and latitude both equal to 0); (2) records with less than three decimals to make the dataset easier to handle; (3) occurrences prior to 1950 to minimise potentially inexact geolocation; (4) duplicated records (those with the identical coordinates and species ID); and (5) records not corresponding to preserved specimens. The latter were mostly retrieved from iNaturalist, a source of observation-based data that provides georeferenced photographs of living specimens. While most specimens of the genus *Carex* are identifiable at genus or sectional rank, the platform's records often need independent curation due to low image quality and high uncertainty in taxonomic validation. These specific problems require semi-automated pipelines that extend beyond the scope of this work. Collector and herbarium numbers were visually checked to ensure that no duplicates remained due to typos or typeset differences. Geographic coordinates were formatted and homogenised to decimal degrees (WGS84 as a Coordinates Reference System) to unify the information from all different sources using the packages *tidyr* (Wickham and Girlich 2022) and *biogeo* (Robertson et al. 2016). Although we discarded those occurrences with less than three decimals of precision, we accepted curated data with two decimals for a few critical species. These are indicated in the "incidence" column of the final dataset. Eventually, to avoid oversampling in some areas and facilitate the manageability of the resulting dataset, data thinning was applied to retain only those occurrences separated at least by 1 km using the *spThin* package (Aiello-Lammens et al. 2015).

After quality control, datasets were merged with the other additional sources of data (additional databases, georeferenced herbarium specimens, specimen lists from literature, and fieldwork). All occurrences were plotted on species/subspecies maps. To allow for visual curation at the regional scale, these maps were circumscribed to nine larger geographic areas: North America (north of Mexico), Mexico and Central America, South America, Western Palearctic, Tropical Africa, Temperate Asia, Tropical Asia, Australia, and New Zealand. These maps were sent to the different co-authors of this paper for expert curation of regional areas and key taxonomic groups (see Author contributions section). These co-authors provided feedback in the form of comments (pointing to potentially erroneous data and introduced populations) or additional data whenever significant gaps needed to be filled. Records added or marked by the co-authors were not discarded but flagged in our dataset to allow comparison with future raw downloads of data from the repositories. The resulting expert-curated dataset (ECD herein) was integrated manually into a master table.

2.3 | Technical Validation

Following the authors' efforts to remove inaccurate records and identify introduced populations, we performed three alternative cross-dataset comparisons. First, a visual comparison of species richness was conducted by plotting occurrences from the GBIF-T raw dataset, after cleaning and filtering, and the ECD. This allowed us to identify areas where expert curation had significantly improved data quality. Secondly, we generated species richness maps to assess differences in species distribution across datasets. Species richness was estimated by tabulating species presence in each Level 3 botanical country, as defined by the International Working Group on Taxonomic Databases for Plant Sciences (Brummitt et al. 2001; mostly matching political countries but representing sub-national entities in large countries as USA or China). Maps were generated using the R package *ggplot2* (Wickham 2016) with the "Zissou1" palette from the package *wesanderson* (Ram and Wickham 2018). Richness values based on the POWO reference taxon lists, GBIF-T raw data, and the expert-curated dataset were plotted. We performed three pairwise comparisons: (1) GBIF-T raw data versus ECD, (2) POWO versus ECD, and (3) POWO versus GBIF-T raw data.

Finally, to verify that, for certain ecological analyses, curated data provide noticeable differences compared to raw data, we conducted Species Distribution Models (SDMs) for the following species: *Carex pendula*, *C. acutata*, and *C. spartea*. The former has a wide distribution area, whereas the latter two are more geographically restricted. For each species, we built two distribution models. One based on the GBIF-T raw dataset, while the other used the ECD to extract the geographic distribution per species. Records flagged with "introduced" and "erroneous" in the "curation_change" column of the ECD were removed. A data thinning was applied to the records of *C. pendula* using a minimum distance of 10 km. The total number of records for each species and dataset is shown in Table 1. We retrieved 19 bioclimatic variables (resolution of 2.5 min) from WorldClim (Fick and Hijmans 2017) for current conditions. To avoid using highly correlated variables, we calculated a correlation matrix with absolute values for each species. Subsequently, this was converted into a distance matrix, which was visualised as a dendrogram. We selected one variable per clade when the branch length < 0.5. Additionally, using the *HH* package (Heiberger 2024) in R (R Core Team 2021), we calculated the variance inflation factor (VIF), a measure that estimates the severity of the effect of multicollinearity in a model (Guisan et al. 2017). The climatic variables with a VIF < 0.5 were selected, which made biological sense based on the study species. Thus, for *C. pendula*, Bio1, Bio7, Bio10, and Bio12 were selected. For *C. acutata*, the chosen variables were Bio3, Bio5, and Bio15; and for *C. spartea*, they were Bio3, Bio9, Bio13, Bio15. The selected climatic variables were cropped to a specific extent based on the geographic distribution of the species, except for *Carex pendula*. The extent used for South America was −100, −20, −60, 15, and the extent for Africa was −25, 55, −40, 38. To compute the potential distribution, we implemented the R package *Biomod2* (Thuiller et al. 2024) using four different algorithms: Generalised Additive Model (GAM), Generalised Boosted Regression Model (GBM), Generalised Linear Model (GLM), and Random Forest (RF). Thereafter, we performed an ensembled model including all algorithms to build more accurate projections. We randomly built a set of 10,000 pseudo-absences from each area and generated a data splitting (70% training data and 30% testing data) to assess the models by cross-validation with 10 replicates. The Area Under the Curve (AUC; Swets 1988) was used as a metric evaluation for building models with a threshold > 0.8 (Guisan et al. 2017).

TABLE 1 | Records from both GBIF-T raw and ECD datasets for conducting the species distribution modelling for *Carex pendula*, *C. acutata* and *C. spartea*, and their current distribution range.

Species	Records from GBIF-T raw dataset	Records from ECD	Distribution range
Carex pendula	393	3430	Europe and introduced in Australia, New Zealand, and United States
Carex acutata	12	27	South America
Carex spartea	113	158	South Africa

3 | Results

The expert curation led to notable improvements in species richness. The regions showing the greatest additions include South America (Figure 1a), Southeast Asia (Figure 1b), and the Czech Republic in Europe (Figure 1c).

Furthermore, maps showing the number of species per country for each dataset are displayed in Figure 2a–c. The highest species richness in the three maps is concentrated in the Northern Hemisphere, with minor differences mainly related to the taxonomic ranks used (e.g., many infraspecific taxa in Oregon, USA). However, there are no substantial differences in the number of species across these three maps. The three comparisons across datasets (Figure 2d–f) reveal important patterns. Maps comparing POWO and GBIF-T raw data (Figure 2f) showed that China, Poland, and Romania exhibit the highest values, indicating that there are more species listed in POWO than in the raw primary downloads. This highlights a limitation of the raw dataset in terms of the taxonomic diversity of *Carex* in these countries. In contrast, the curated dataset (ECD) contains more species in China, Peru, and Chile than the GBIF-T raw dataset (Figure 2d) validating our approach to covering sampling gaps. In the case of the state of Oregon (USA), which showed low values when comparing ECD with GBIF-T raw data and POWO with GBIF-T raw dataset (Figure 2d,f), this is again due to the presence of infraspecific taxa in the raw dataset that are not considered in POWO and curated datasets. When examining the differences between POWO and ECD (Figure 2e), the most diverse areas of the Northern Hemisphere have a higher richness in the POWO dataset than in the curated one, but the effect is less noticeable in the Southern Hemisphere. This pattern may be influenced by the combined effect of (1) rare/introduced species lacking records in some areas, and (2) certain areas having a particularly low number of occurrences (e.g., Romania and South Korea). Both cases depict that further sampling efforts may be needed in these areas. A noticeable exception is the Iberian Peninsula, which shows lower values compared to the rest of Europe. This is a consequence of the large number of authors of this work based in Spain, resulting in an increased level of knowledge for the region.

Finally, the SDMs retrieved AUC values between 0.80 and 0.95, which indicate a good predictive ability. The inferred potential distribution for *C. pendula* was slightly similar in both models (Figure 3a), but with fewer potential areas retrieved in the ECD-based model, especially regarding areas peripheral to the taxon's native range in Europe (e.g., Eastern and Southern Europe), as well as in other areas of the world on other continents. As

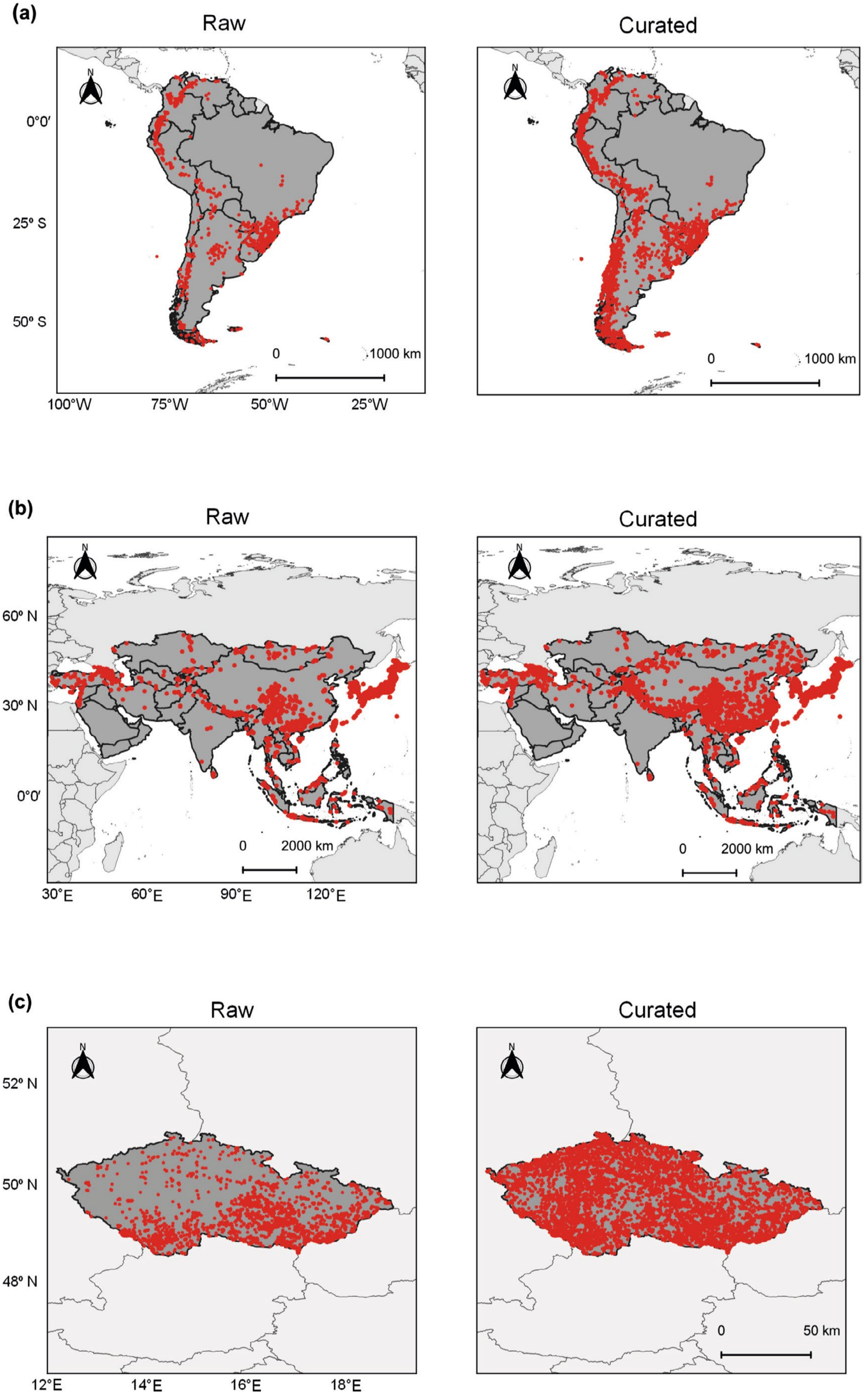

FIGURE 1 | Spatial distribution of the GBIF-T raw data and expert-curated dataset for *Carex* occurrences in (a) South America, (b) Southeast Asia, and (c) Czech Republic.

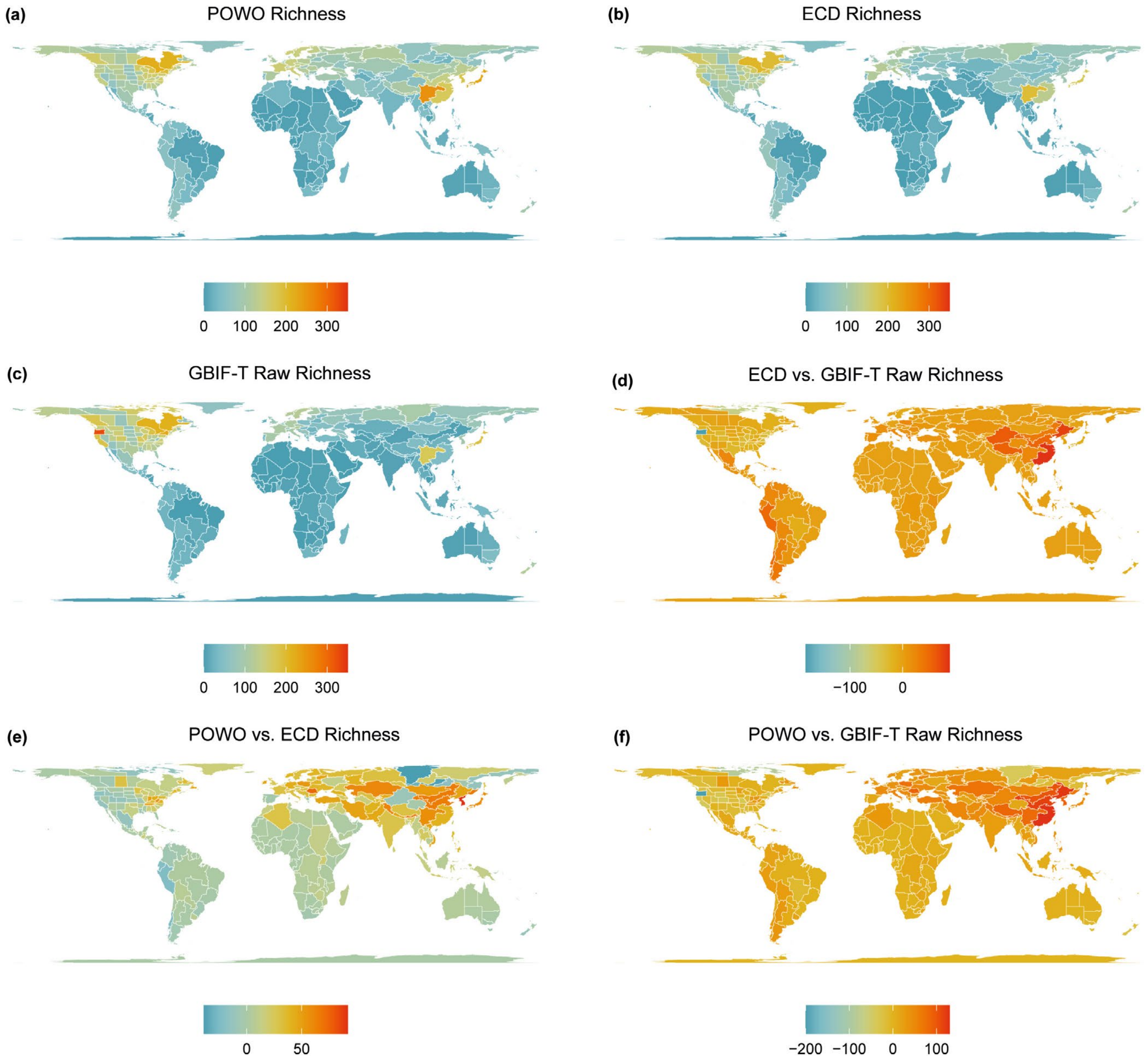

FIGURE 2 | Species richness maps for: (a) POWO, (b) ECD, and (c) GBIF-T raw data; and the richness maps resulting from pairwise comparisons between (d) ECD versus GBIF-T raw data, (e) POWO versus ECD, and (f) POWO versus GBIF-T raw data.

a practical application of our ECD, *C. pendula* is a species of concern with invasive potential (CAL-IPC 2024; NZPCN 2024), for which the use of raw data may confound the potential invasiveness of the species. Likewise, when comparing models using the GBIF-T raw dataset to those using the ECD, projections for *Carex acutata* (Figure 3b) and *C. spartea* (Figure 3c) revealed different potential areas at a continental scale, particularly regarding the areas with the highest suitability. This illustrates how, in species with a relatively low number of occurrences, any misidentification may potentially confound the results.

4 | Usage Notes

The dataset described here can be used for bioclimatic and (macro-)evolutionary analyses (e.g., niche evolution, species distribution models) that require highly accurate geographic information. Even though the GBIF and Tropicos records have already been filtered, we recommend data thinning to avoid spatial correlations following Coca-de-la-Iglesia et al. (2023) on the curated dataset to conduct bioclimatic niche studies. The thinning procedure keeps the most valuable spatial information while eliminating some records to reduce the effects of sampling bias (Aiello-Lammens et al. 2015). Alternatively, if a specific analysis requires a minimal number of records to be conducted, we suggest retaining records that have more than two decimal places in their coordinates, ensuring that these occurrences are as accurate as possible and correspond to the distribution of the studied species. The performed SDMs illustrate how the ECD could display more accurate results, especially in cases where the species distribution is restricted.

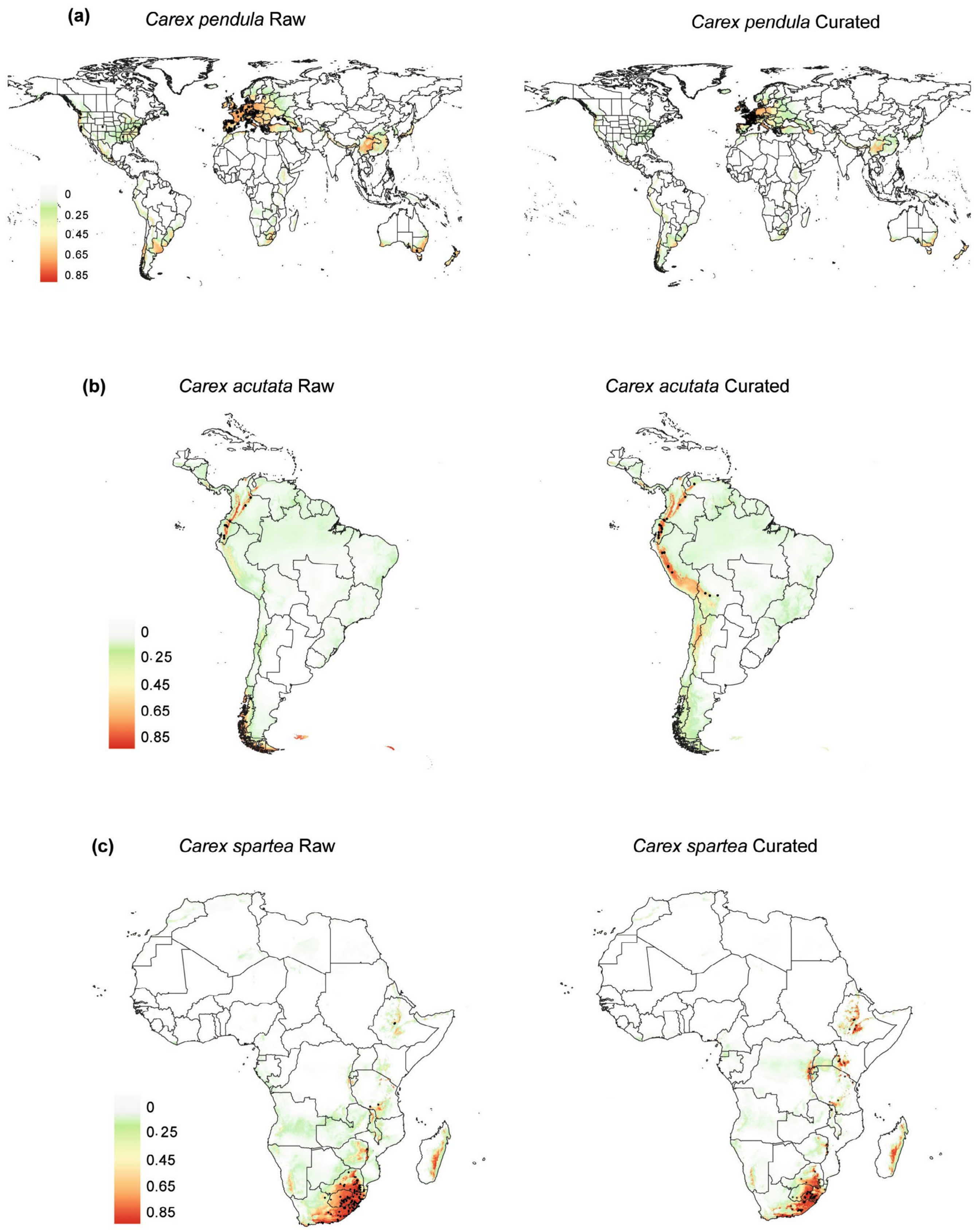

FIGURE 3 | Climatic suitability and potential distributions predicted by biomod at present for GBIF-T raw dataset and ECD for (a) *Carex pendula*, (b) *C. acutata*, and (c) *C. spartea*. Records used for the distribution modelling, representing the current distribution area, are shown as black dots in each panel. The probability scale is the same for all projections.

The "curation_change" column in ECD dataset allows the user to select or discard particular occurrences that were flagged during the expert curation process as doubtful or introduced. We recommend using these flagged occurrences to evaluate the certainty or native status of additional records added to GBIF after our dataset is downloaded. By plotting unflagged, flagged, and new records using *curatoR* (Bradley 2023), users can apply their own criteria based on the needs of their study.

5 | Discussion

The quality control procedure described here addressed several common issues found in online repositories, such as duplicate records and erroneous occurrences (e.g., records in the ocean or at country centroids). By implementing our expert curation approach, we achieved two major improvements over primary occurrence data from repositories. First, the contributors identified potentially erroneous data and flagged introduced populations when appropriate. Second, expert taxonomists verified the identities of additional records from revised herbarium material and field collections. Furthermore, fieldwork and taxonomic revision efforts by the authors helped fill several main sampling gaps in regions with previously limited data, such as Mexico (Reznicek et al. 2021), Central and South America (Villaverde et al., 2015; Jiménez-Mejías and Roalson 2016; Jiménez-Mejías and Escudero 2016; Jiménez-Mejías et al. 2016; Jiménez-Mejías and Dorr 2018; Jiménez-Mejías et al. 2018; Jiménez-Mejías et al. 2020; Jiménez-Mejías, Fabbroni and Haigh 2020; Jiménez-Mejías et al. 2021; Jiménez-Mejías and Reznicek 2018; Jiménez-Mejías et al. 2021; Jiménez-Mejías et al. 2023; Jiménez-Mejías et al. 2023; Lois et al. 2023; [subg. Uncinia]; Muñoz-Schüler et al. 2023), China (Zhou and Jin 2014; Lu and Jin 2022), Czech Republic (https://pladias.cz/en/), and Greece (https://portal.cybertaxonomy.org/flora-greece).

The whole quality control procedure and resulting dataset can serve as an example of how high-quality taxonomic revisionary work can provide a significant improvement of occurrence databases, and a solid foundation on which other data-based analyses can be grounded. However, we recognise that future additions to GBIF and other repositories would potentially face the same issues we tried to address in this work. To mitigate this, our curated dataset (ECD) can serve as a reference for identifying occurrences needing further attention. By comparing our dataset with other databases or updated versions of GBIF, occurrences not associated with confirmed presences can be easily flagged for revision or treated with caution. This approach not only strengthens occurrence-based studies but also has the potential to improve the taxonomic curation of outlier records, ensuring more accurate SDM and reliable ecological conclusions.

Despite the improvements, limitations remain, particularly in boreal regions and Southeast Asia. In boreal regions, complex sedge groups are often labelled with collective names, masking species diversity. These groups are often the focus of revisions that shed light on the taxonomy, as seen in cases like *Carex bigelowii* sect. *Phacocystis* (Westergaard, Kyrkjeeide and Brandrud 2021), and *Carex rotundata* sect. *Vesicariae* (Pedersen et al. 2016). For such groups comprising multiple species, it is crucial to prioritise updated and verified datasets, replacing raw occurrences with more reliable and accurate information. In SE Asia, records are underrepresented compared to other parts of the world. Species from this area are rarely collected in their native ranges, and the few available herbarium materials are often outdated (usually prior to 1980s) and lack precise geographical information. Greater sampling and recording efforts are needed to cover these knowledge gaps, as has been achieved in this paper for Mexican and South American sedges.

Our study highlights the importance of expert curation to address common biases in large occurrence databases, offering robust quality-control procedures for biogeography, ecology, and conservation analyses. Continued collaboration with taxonomists, alongside increasing sampling efforts in underrepresented regions, will be crucial for improving future biodiversity databases.

Author Contributions

P.J.-M. conceived the project idea and supervised the study. M.S.-A. homogenised the data. P.J.-M., P.G.-M., and M.S.-A. drafted the manuscript. C.B.-B., M.C.-de-la-I., A.D., A.G.-N., K.K.S., D.S., M.S.-A., and P.V. downloaded, filtered, and prepared the data. P.J.-M., S.D., M.E., A.M.-A., M.L., P.M.-S., and R.F.C.N. revised herbaria specimens; also, P.G.-M., J.I.M.-C., R.L., A.M.-A., and M.S.-A. manually georeferenced these specimens. C.B., F.B., K.A.F., G.G., S.G., M.S.G.-E., D.H., M.H.H., P.J.-M., X.-F.J., J.K., B.L., Y.-F.L., M.L., A.M., A.M.-A., P.M.-S., R.F.C.N., R.R., A.A.R., M.S.-A., A.S., F.V., K.L.W., O.Y., and S.Z. curated the occurrences. Furthermore, S.G., M.S.G.-E., M.H.H., X.-F.J., B.L., A.M.-A., P.B., R.R., and S.Z. contributed to the dataset with additional data. C.B.-B., A.C., M.F., P.G.-M., P.J.-M., X.-F.J., Y.-F.L., M.L., J.I.M.-C., S.M.-B., M.M., A.M.-A., P.M.-S., A.M.M., N.O., L.P.-S., A.A.R., E.S., and M.S.-A. conducted field trips. D.S. and M.S.-A. analysed the data and prepared the figures. For more details about the individual contributions of each co-author see Table S2 in the Supporting Information section.

Affiliations

[1]Área de Botánica, Departamento de Biología Molecular e Ingeniería Bioquímica, Universidad Pablo de Olavide, Seville, Spain | [2]Biodiversity and Conservation Area, Department of Biology and Geology, Physics and Inorganic Chemistry, IICG-Universidad Rey Juan Carlos, Madrid, Spain | [3]Departamento de Biología Vegetal y Ecología, IICG-Universidad de Sevilla, Seville, Spain | [4]Departamento de Biología, Universidad Autónoma de Madrid, Madrid, Spain | [5]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA | [6]Centro Ricerche Floristiche Dell'appennino, Università di Camerino—Parco Nazionale del Gran Sasso e Monti Della Laga, L'Aquila, Italy | [7]University of Chinese Academy of Sciences, Beijing, China | [8]State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany, Chinese Academy of Sciences, Beijing, China | [9]China National Botanical Garden, Beijing, China | [10]Department of Ecology and Conservation Biology, Texas A&M University, Texas, USA | [11]Universidad Nacional Mayor de San Marcos, Museo de Historia Natural y Facultad de Ciencias Biológicas, Lima, Peru | [12]Independent Researcher, Natagora Asbl, Namur, Belgium | [13]Instituto de Botánica Darwinion (CONICET-ANCEFN), Buenos Aires, Argentina | [14]Cátedra de Plantas Vasculares, Facultad de Ciencias Naturales, Universidad Nacional de Salta, Salta, Argentina | [15]Allan Herbarium (CHR), Manaaki Whenua, Landcare Research, Lincoln, New Zealand | [16]Sezione di Botanica, Museo di Storia Naturale di Milano, Milano, Italy | [17]Senckenberg Institute for Plant Form and Function (SIP), Jena, Germany | [18]Plant Biodiversity Group, Institute of Biodiversity, Ecology and Evolution, Friedrich Schiller University, Jena, Germany | [19]Herbarium Haussknecht (JE), Jena, Germany | [20]CIIDIR, Durango, Instituto Politécnico Nacional, Durango, Dgo., México | [21]Independent Researcher, Lavau-sur-Loire,

France | [22]Martin Luther University Halle-Wittenberg, Geobotany and Botanical Garden, Botanical Garden, Halle, Germany | [23]School of Forestry and Bio-Technology, Zhejiang Agriculture and Forestry University, Zhejiang, China | [24]Independent Researcher, Choszczno, Poland | [25]Departamento de Biodiversidad y Gestión Ambiental (Botánica), Facultad de Ciencias Biológicas y Ambientales, Universidad de León, León, Spain | [26]College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China | [27]Independent Researcher, Celldömölk, Hungary | [28]The Bolus Herbarium, Department of Biological Sciences, University of Cape Town, Rhodes Gift, Cape Town, South Africa | [29]Herbario CONC, Departamento de Botánica, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Casilla 160-C, Concepción, Chile | [30]New York Botanical Garden, Bronx, New York, USA | [31]Centro de Investigación de la Biodiversidad y Cambio Climático (BioCamb) y Facultad de Ciencias de Medio Ambiente, Universidad Indoamérica, Quito, Ecuador | [32]Central Laboratory for Electron Microscopy, Prorectorate for Research and Innovation, Universidade Federal de Santa Catarina, Campus Trindade, Florianópolis, Santa Catarina, Brazil | [33]Department of Forest Botany, Dendrology and Geobiocoenology, Mendelova Univerzita v Brně, Brno, Jihomoravský, Czech Republic | [34]University of Michigan Herbarium, Ann Arbor, Michigan, USA | [35]Independent Researcher, Ørbæk, Denmark | [36]Natuurpunt Studie Vzw, Mechelen, Belgium | [37]Meise Botanic Garden, Meise, Belgium | [38]National Herbarium of New South Wales, Botanic Gardens of Sidney, Australian Botanic Garden, Mount Annan, New South Wales, Australia | [39]Faculty of Biosphere-Geosphere Science, Okayama University of Science, Okayama-shi, Okayama, Japan

Acknowledgements

The authors thank all curators and staff of the following herbaria A, AHH, B, BAA, BR, COL, CONC, CPUN, DUKE, E, ECON, F, GH, GZU, HAL, HSP, HUSA, HUT, IRVC, ISC, K, LGO, LIL, LL, LP, LPB, M, MA, MICH, MO, MOL, MOLF, MSB, NHA, NY, P, PRC, QCA, QCNE, SI, UEC, UMBS, UPS, UPOS, US, USM, VEN, WIS, and WS for granting them access to their collections and for assistance. We would also like to acknowledge all contributors for providing new data and necessary corrections. This work has been carried out with financial support by the project PID2020-113897GB-I00 (DANZ, to P.J.-M. and S.M.-B.) and a Ramón y Cajal postdoctoral contract (RYC2021-031238-I) for P.J.-M, both from the Spanish Ministry of Science and Innovation. Also, we have been funded by the Regional Government of Madrid, Spain (Macondo, SI1/PIJ/2019-00333), the Smithsonian Postdoctoral Fellowship program, and a Youth Guarantee contract towards P.G.-M. (PEJ-2020-AI/AMB-18719, Regional Government of Madrid). C.B.-B., M.S.-A., and P. G.-M. were supported by a Predoctoral Fellowship grant from the Ministry of Science, Innovation and Universities (FPU16/01257, PREP2023-C1FPI2024-02) and from the Ministry of Science and Innovation (PRE2021-100655), respectively. A. M.-A. was financed by the Bentham-Moxon Trust grant 2021 (Kew Royal Botanic Gardens, UK), and M.E. was sponsored by the "Beca Iberoamérica. Jóvenes Profesores e Investigadores del Programa Santander Universidades".

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data are available on Zenodo at https://doi.org/10.5281/zenodo.14998163 (version 1). Updated versions will be uploaded annually to the same repository, each with a unique DOI; the general DOI https://doi.org/10.5281/zenodo.14998162 always resolves to the latest release. When using the dataset, please cite both this article and the specific version employed. Updates and additional resources are also available on GitHub, https://github.com/msanzarnal/carex_occurrence_dataset.

References

Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015. "spThin: An R Package for Spatial Thinning of Species Occurrences Records for Use in Ecological Niche Models." *Ecography* 38, no. 5: 541–545. https://doi.org/10.1111/ecog.01132.

Benítez-Benítez, C., S. Martín-Bravo, C. S. Bjorå, et al. 2021. "Geographical vs. Ecological Diversification in *Carex* Section Phacocystis (Cyperaceae): Patterns Hidden Behind a Twisted Taxonomy." *Journal of Systematics and Evolution* 59, no. 4: 642–667. https://doi.org/10.1111/jse.12731.

Benítez-Benítez, C., A. Otero, K. A. Ford, et al. 2021. "An Evolutionary Study of *Carex* Subg. *Psyllophorae* (Cyperaceae) Sheds Light on a Strikingly Disjunct Distribution in the Southern Hemisphere, With Emphasis on Its Patagonian Diversification." *Frontiers in Plant Science* 12: 735302. https://doi.org/10.3389/fpls.2021.735302.

Benítez-Benítez, C., M. Sanz-Arnal, M. Urbani, P. Jiménez-Mejías, and S. Martín-Bravo. 2022. "Dramatic Impact of Future Climate Change on the Genetic Diversity and Distribution of Ecologically Relevant Western Mediterranean *Carex* (Cyperaceae)." *PeerJ* 10: e13464. https://doi.org/10.7717/peerj.13464.

Bradley, M. 2023. "Curator: Curate Biological Datasets." https://github.com/matthewbradley22/curator.

Brummitt, R. K., F. Pando, S. Hollis, and N. Brummitt. 2001. *World Geographical Scheme for Recording Plant Distributions*. 2nd ed. Hunt Institute for Botanical Documentation.

CAL-IPC. 2024. *Carex pendula*. California Invasive Plant Council. https://www.cal-ipc.org/plants/profile/carex-pendula-profile/.

Chamberlain, S., D. Oldoni, V. Barve, et al. 2022. *Rgbif: Interface to the Global Biodiversity Information Facility API*. https://doi.org/10.32614/CRAN.package.rgbif.

Coca-de-la-Iglesia, M., N. G. Medina, J. Wen, and V. Valcárcel. 2022. "Evaluation of Tropical–Temperate Transitions: An Example of Climatic Characterization in the Asian Palmate Group of Araliaceae." *American Journal of Botany* 109, no. 9: 1488–1507. https://doi.org/10.1002/ajb2.16059.

Coca-de-la-Iglesia, M., V. Valcárcel, and N. G. Medina. 2023. "A Protocol to Retrieve and Curate Spatial Data From Online Biodiversity Databases Using R." *Bio-Protocol* 13, no. 20: e4847. https://doi.org/10.21769/bioprotoc.4847.

Eckert, S., A. Hamad, C. J. Kilawe, et al. 2020. "Niche Change Analysis as a Tool to Inform Management of Two Invasive Species in Eastern Africa." *Ecosphere* 11, no. 2: e02987. https://doi.org/10.1002/ecs2.2987.

Fick, S. E., and R. J. Hijmans. 2017. "WorldCim 2: New 1 Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37, no. 12: 4302–4315. https://doi.org/10.1002/joc.5086.

Fourcade, Y. 2016. "Comparing Species Distributions Modelled From Occurrence Data and From Expert-Based Range Maps. Implication for Predicting Range Shifts With Climate Change." *Ecological Informatics* 36: 8–14. https://doi.org/10.1016/j.ecoinf.2016.09.002.

García-Moro, P., A. Otero, C. Benítez-Benítez, et al. 2022. "Biogeography and Systematics of Carex Subgenus Uncinia (Cyperaceae): A Unique Radiation for the Genus Carex in the Southern Hemisphere." *Taxon* 71, no. 3: 587–607. https://doi.org/10.1002/tax.12678.

GBIF. 2023. "The Global Biodiversity Information Facility." https://www.gbif.org/.

Global Carex Group. 2015. "Making *Carex* Monophyletic (Cyperaceae, Tribe Cariceae): A New Broader Circumscription." *Botanical Journal of the Linnean Society* 179, no. 1: 1–42. https://doi.org/10.1111/boj.12298.

Global Carex Group, P. Jiménez-Mejías, M. Hahn, et al. 2016. "Megaphylogenetic Specimen-Level Approaches to the *Carex* (Cyperaceae) Phylogeny Using ITS, ETS, and matK Sequences: Implications for Classification." *Systematic Botany* 41, no. 3: 500–518. https://doi.org/10.1600/036364416X692497.

Global Carex Group, E. H. Roalson, P. Jiménez-Mejías, A. L. Hipp, C. Benítez-Benítez, and L. P. Bruederle. 2021. "A Framework Infrageneric Classification of *Carex* (Cyperaceae) and Its Organizing Principles." *Journal of Systematics and Evolution* 59, no. 4: 726–762. https://doi.org/10.1111/jse.12722.

Govaerts, R., P. Jiménez-Mejías, J. Koopman, et al. 2022. "Cyperaceae in POWO. Plants of the World Online." http://www.plantsoftheworldonline.org/.

Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press.

Heiberger, R. M. 2024. *HH: Statistical Analysis and Data Display: Heirberger and Holland*. https://doi.org/10.32614/CRAN.package.HH.

Hendrichs, M., S. Michalski, D. Begerow, F. Oberwinkler, and F. H. Hellwig. 2004. "Phylogenetic Relationships in *Carex*, Subgenus *Vignea* (Cyperaceae), based on ITS Sequences." *Plant Systematics and Evolution* 246, no. 1: 109–125. https://doi.org/10.1007/s00606-004-0127-1.

Hendrichs, M., F. Oberwinkler, D. Begerow, and R. Bauer. 2004. "*Carex*, Subgenus *Carex* (Cyperaceae)—A Phylogenetic Approach Using ITS Sequences." *Plant Systematics and Evolution* 246: 89–107. https://doi.org/10.1007/s00606-004-0128-0.

Hortal, J., J. M. Lobo, and A. Jiménez-Valverde. 2007. "Limitations of Biodiversity Databases: Case Study on Seed-Plant Diversity in Tenerife, Canary Islands." *Conservation Biology* 21, no. 3: 853–863. https://doi.org/10.1111/j.1523-1739.2007.00686.x.

Jiménez-Mejías, P., A. Hilpold, B. Frajman, et al. 2014. "*Carex cespitosa*: Reappraisal of Its Distribution in Europe." *Willdenowia* 44, no. 3: 327–343. https://doi.org/10.3372/wi.44.44303.

Martín-Bravo, S., P. Jiménez-Mejías, T. Villaverde, M. Escudero, M. Hahn, and D. Spalink. 2019. "A Tale of Worldwide Success: Behind the Scenes of Carex (Cyperaceae) Biogeography and Diversification." *Journal of Systematics and Evolution* 57, no. 6: 695–718. https://doi.org/10.1111/jse.12549.

Mejía, O., N. Martínez-Méndez, F. Pérez-Miranda, and W. A. Matamoros. 2022. "Climatic Niche Evolution of a Widely Distributed Neotropical Freshwater Fish Clade." *Biological Journal of the Linnean Society* 135, no. 4: 839–855. https://doi.org/10.1093/biolinnean/blab153.

Míguez, M., P. Jiménez-Mejías, C. Benítez-Benítez, H. Schaefer, and S. Martín-Bravo. 2021. "Systematics of the Giant Sedges of Carex Sect. Rhynchocystis (Cyperaceae) in Macaronesia With Description of Two New Species." *Systematic Botany* 46, no. 2: 304–320. https://doi.org/10.1600/036364421X16231782047442.

Moreira, A., J. Boavida-Portugal, P. R. Almeida, S. Silva, and C. M. Alexandre. 2022. "Macro-Habitat Suitability for Threespine Stickleback (*Gasterosteus aculeatus* L.) Near the Southern Limit of Its Global Distribution: Implications for Species Management and Conservation." *Fishes* 7, no. 5: 271. https://doi.org/10.3390/fishes7050271.

Naczi, R. F. C., T. W. Barger, D. D. Spaulding, M. R. Naczi, J. E. Dorey, and J. K. Triplett. 2020. "Revealing a Significant Center of Sedge Diversity: *Carex* (Cyperaceae) of Jackson County, Alabama, U.S.A." *American Midland Naturalist* 184, no. 1: 17–47. https://doi.org/10.1637/0003-0031-184.1.17.

NZPCN 2024. *Carex pendula*. "New Zealand Plant Conservation Network." https://www.nzpcn.org.nz/flora/species/carex-pendula/.

Oda, J., S. Fuse, J. Yamashita, and M. N. Tamura. 2019. "Phylogeny and Taxonomy of *Carex* (Cyperaceae) in Japan I. C. Sect. Rarae." *Acta Phytotaxonomica et Geobotanica* 70, no. 2: 69–85. https://doi.org/10.18942/apg.201821.

Oleas, N. H., K. J. Feeley, J. Fajardo, A. W. Meerow, and J. Francisco-Ortega. 2019. "Muddy Boots Beget Wisdom: Implications for Rare or Endangered Plant Species Distribution Models." *Diversity* 11, no. 1: 10. https://doi.org/10.3390/d11010010.

Palacio, R. D., P. J. Negret, J. Velásquez-Tibatá, and A. P. Jacobson. 2021. "A Data-Driven Geospatial Workflow to Map Species Distributions for Conservation Assessments." *Diversity and Distributions* 27, no. 12: 2559–2570. https://doi.org/10.1111/ddi.13424.

Pedersen, A. T. M., M. D. Nowak, A. K. Brysting, R. Elven, and C. S. Bjorå. 2016. "Hybrid Origins of *Carex rostrata* var. *borealis* and *C. stenolepis*, Two Problematic Taxa in *Carex* Section *Vesicariae* (Cyperaceae)." *PLoS One* 11, no. 10: e0165430. https://doi.org/10.1371/journal.pone.0165430.

POWO. 2025. "Facilitated by the Royal Botanic Gardens, Kew." https://powo.science.kew.org/.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Ram, K., and H. Wickham. 2018. *Wesanderson: A Wes Anderson Palette Generator*. https://doi.org/10.32614/CRAN.package.wesanderson.

Roalson, E. H., J. T. Columbus, and E. A. Friar. 2001. "Phylogenetic Relationships in Cariceae (Cyperaceae) Based on ITS (nrDNA) and TrnT-L-F (cpDNA) Region Sequences: Assessment of Subgeneric and Sectional Relationships in *Carex* With Emphasis on Section *Acrocystis*." *Systematic Botany* 26, no. 2: 318–341.

Robertson, M. P., V. Visser, and C. Hui. 2016. "Biogeo: An R Package for Assessing and Improving Data Quality of Occurrence Record Datasets." *Ecography* 39, no. 4: 394–401. https://doi.org/10.1111/ecog.02118.

Sanbonmatsu, K. K., and D. Spalink. 2022. "A Global Analysis of Mosses Reveals Low Phylogenetic Endemism and Highlights the Importance of Long-Distance Dispersal." *Journal of Biogeography* 49, no. 4: 654–667. https://doi.org/10.1111/jbi.14333.

Santamarina, S., R. G. Mateo, E. Alfaro-Saiz, and C. Acedo. 2023. "On the Importance of Invasive Species Niche Dynamics in Plant Conservation Management at Large and Local Scale." *Frontiers in Ecology and Evolution* 10. https://doi.org/10.3389/fevo.2022.1049142.

Spalink, D., R. Kriebel, P. Li, et al. 2018. "Spatial Phylogenetics Reveals Evolutionary Constraints on the Assembly of a Large Regional Flora." *American Journal of Botany* 105, no. 11: 1938–1950. https://doi.org/10.1002/ajb2.1191.

Starr, J. R., R. J. Bayer, and B. A. Ford. 1999. "The Phylogenetic Position of *Carex* Section *Phyllostachys* and Its Implications for Phylogeny and Subgeneric Circumscription in *Carex* (Cyperaceae)." *American Journal of Botany* 86, no. 4: 563–577.

Swets, J. A. 1988. "Measuring the Accuracy of Diagnostic Systems." *Science* 240: 1285–1293.

Takahashi, K. T., J. Oda, S. Fuse, and M. N. Tamura. 2021. "Biosystematic Studies of *Carex* (Cyperaceae) I. Molecular Phylogenetic Analysis of the *C. Macroglossa* Complex With Reference to Variation in Morphology, Chromosomal Features and Species Delimitation." *Acta Phytotaxonomica et Geobotanica* 72, no. 2: 81–92. https://doi.org/10.18942/apg.202019.

Thuiller, W., D. Georges, M. Gueguen, R. Engler, F. Breiner, and B. Lafourcade. 2024. *biomod2: Ensemble Platform for Species Distribution Modeling version 4.2–5-2*. https://doi.org/10.32614/CRAN.package.biomod2.

Villaverde, T., P. Jiménez-Mejías, M. Luceño, et al. 2020. "A New Classification of Carex (Cyperaceae) Subgenera Supported by a HybSeq Backbone Phylogenetic Tree." *Botanical Journal of the Linnean Society* 194, no. 2: 141–163. https://doi.org/10.1093/botlinnean/boaa042.

Westergaard, K. B., M. O. Kyrkjeeide, and M. K. Brandrud. 2021. "Using Genomics to Guide Seed-Sourcing at the Right Taxonomical Level for Ecological Restoration Projects: The Complex Case of *Carex bigelowii* s.lat. in Norway." *Ecology and Evolution* 11, no. 23: 17117–17131. https://doi.org/10.1002/ece3.8350.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. https://doi.org/10.32614/CRAN.package.ggplot2.

Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughan. 2022. *dplyr: A Grammar of Data Manipulation*. https://doi.org/10.32614/CRAN.package.dplyr.

Wickham, H., and M. Girlich. 2022. *Tidyr: Tidy Messy Data*. https://doi.org/10.32614/CRAN.package.tidyr.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** Specialised literature on *Carex* taxonomy reviewed during georeferencing. **Table S1:** List of herbaria surveyed for *Carex* specimen revision. **Table S2:** Summary of individual author contributions to the study.

Author Biography

María Sanz-Arnal previously worked as a research assistant in a research group focused on evolutionary biology, biogeography, and taxonomy of sedges (Cyperaceae), with particular emphasis on the genus *Carex* L. She is currently a PhD student at Universidad Rey Juan Carlos (URJC), investigating the origin and evolution of C_4 photosynthesis in the genus *Cyperus* L. Her research combines molecular techniques, genomic analyses, and ecological niche modelling to study species niche evolution, relying on datasets like the one presented in this work.