

SUPPLEMENTARY INFORMATION

Mapping ignorance to uncover shortfalls in the knowledge on global Orthoptera distribution

Rodrigo Antônio Castro-Souza^{1*}; Geiziane Tessarolo²; Juliana Stropp³; José Alexandre Diniz-Filho²; Richard J. Ladle⁴; Neucir Szinwelski⁵; Joaquín Hortal^{2,6} Thadeu Sobral-Souza¹

(1) Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso, 78060-900, Cuiabá, MT, Brazil

(2) Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil

(3) Department of Biogeography, University of Trier, Trier, Germany

(4) Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, Alagoas, Brazil

(5) Laboratório de Orthoptera, Centro de Ciências Biológicas e da Saúde, Universidade Estadual do Oeste do Paraná, Brazil

(6) Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

corresponding author: rodrigodesouzaac@gmail.com

Supplementary Note 1

For most digitally accessible sources, such as: Global Biodiversity Information Facility (GBIF); Integrated Digitized Biocollections (IDigBio); National Biodiversity Network Atlas (NBN atlas); Sistema da Informação sobre a Biodiversidade Brasileira (SiBBR); Atlas of Living Australia (ALA); Monarch; Canadensys; VertNet; Portal da Biodiversidade – ICMbio (PortalBio); Ecdysis; Natural History Museum Data Portal (NHM Data Portal); Smithsonian Tropical Research Institute (STRI); speciesLink; Grasshoppers, Crickets and Allied Insects (Orthoptera) of Ireland; Madrean Discovery Expedition Flora & Fauna Database; Berkeley Ecoinformatics Engine (Ecoengine); NEON Biorepository; Ocean Biodiversity Information System (OBIS); Biodiversidad de Guatemala Data Portal; InvertEBase), searches were conducted for all available primary records for the order Orthoptera, in the format of Darwin Core spreadsheets (when available)¹, or in the format commonly used by the aforementioned databases (see Table 1 in the Manuscript).

For the citizen science platform iNaturalist, standardized searches for primary records were conducted at the "verifiable data" level for each Orthoptera family and, in some cases, for each category below the family level. A single search for the entire group was not performed due to the extraction limit on iNaturalist, which is 200,000 observations. In the case of Orthoptera, more than

900,000 records were found during the extraction period² (see Table 1 in the Manuscript). Therefore, we followed the family nomenclature proposed by the Catalogue Orthoptera Species File (OSF)³, available on the Catalogue of Life platform (<https://www.catalogueoflife.org/>), to guide each of the searches.

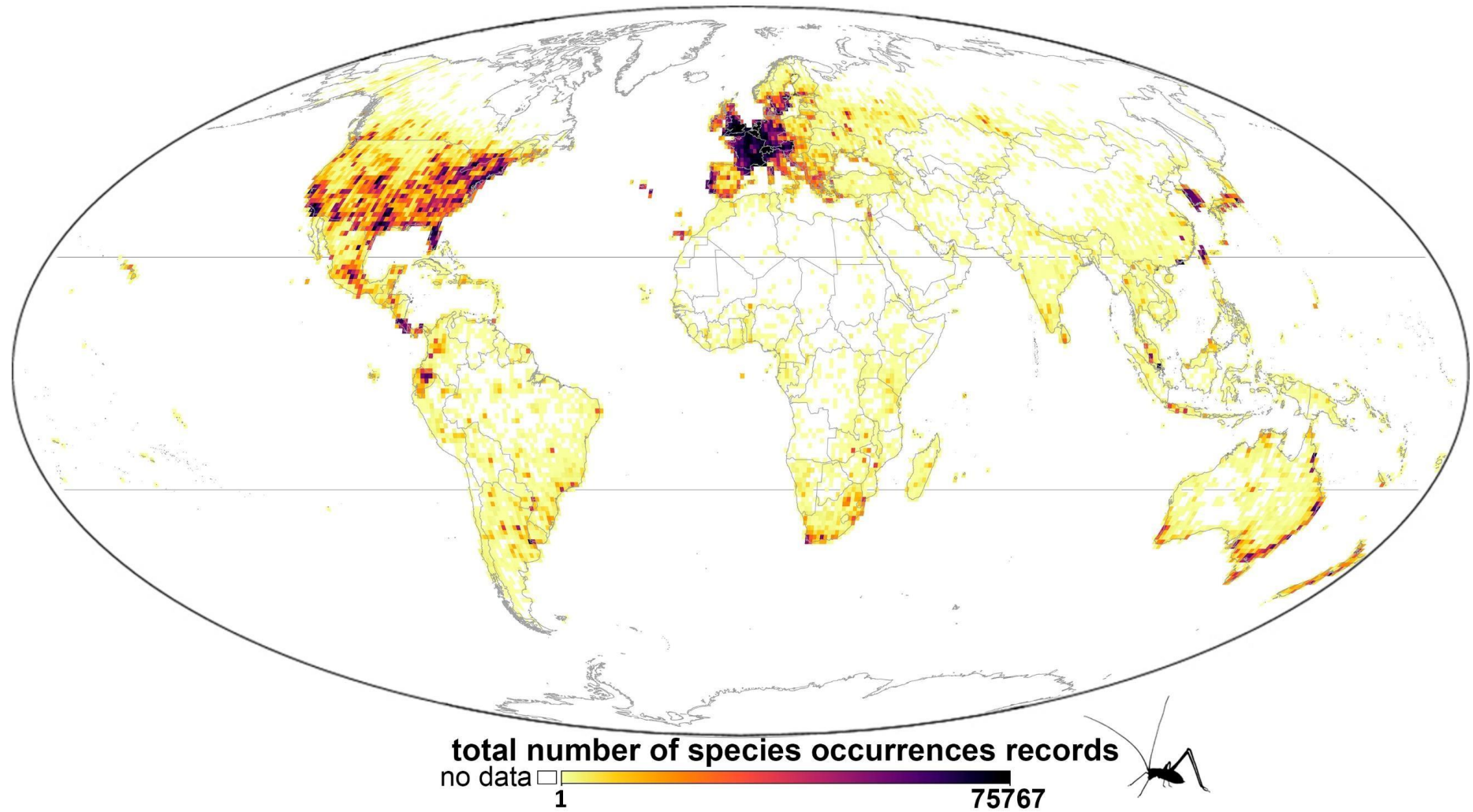
For the taxonomic catalog Orthoptera Species File (OSF), which uses the Species File Software (SFS) program for structuring and storing record information in a Microsoft SQL Server relational database management system⁴, we adopted a web scraping approach with the help of the RSelenium package in the R environment. This approach allowed us to collect the necessary data to create a matrix that combines taxonomic, geographic, and temporal information for cataloged and digitized orthopterans. We chose this approach because it was not feasible to extract standardized lists with all biogeographical information using the tools available in OSF. However, currently, this information scanning is already almost possible, as the OSF has been migrating to a Darwin Core-based system, the TaxonWorks platform (see <https://orthoptera.speciesfile.org/>).

Additionally, for the citizen science platform EcoRegistros, we conducted manual searches for all primary Orthoptera records available up to the extraction date. This was due to the site's structural limitations, which do not allow for automated downloading of all records (see <https://www.ecoregistros.org/site/index.php>). Furthermore, the platform had a relatively small number of records, making manual extraction a viable option.

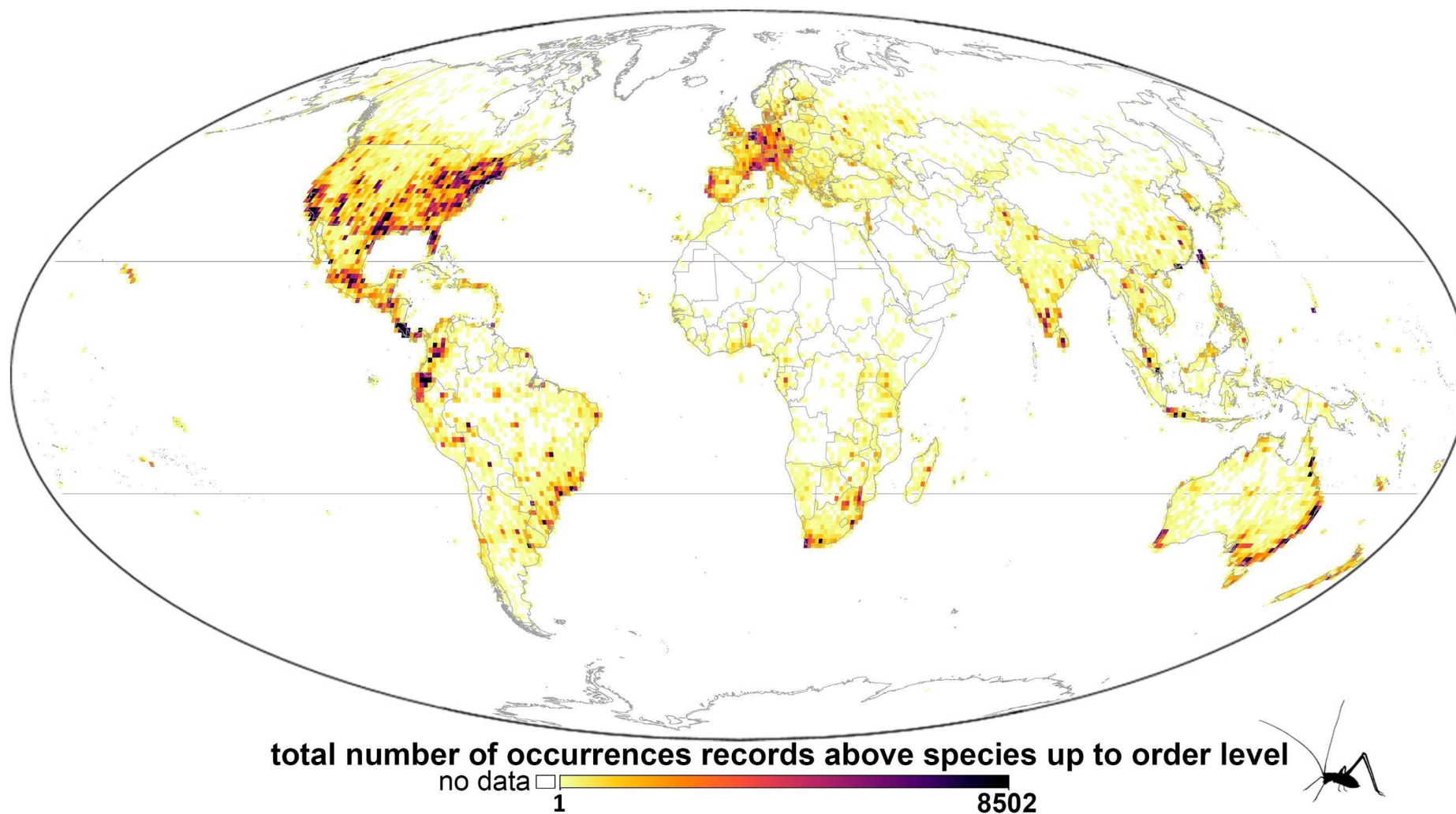
References

1. Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS One* **7**, e29715 (2012).
2. iNaturalist. <https://www.inaturalist.org> (2022).
3. Cigliano, M. M. *et al.* Catalogue of Life Checklist. <https://doi.org/10.48580/d4sw-388> (2022).
4. Song, H. Grasshopper systematics: past, present and future. *J Orthoptera Res* **19**, 57–68 (2010).

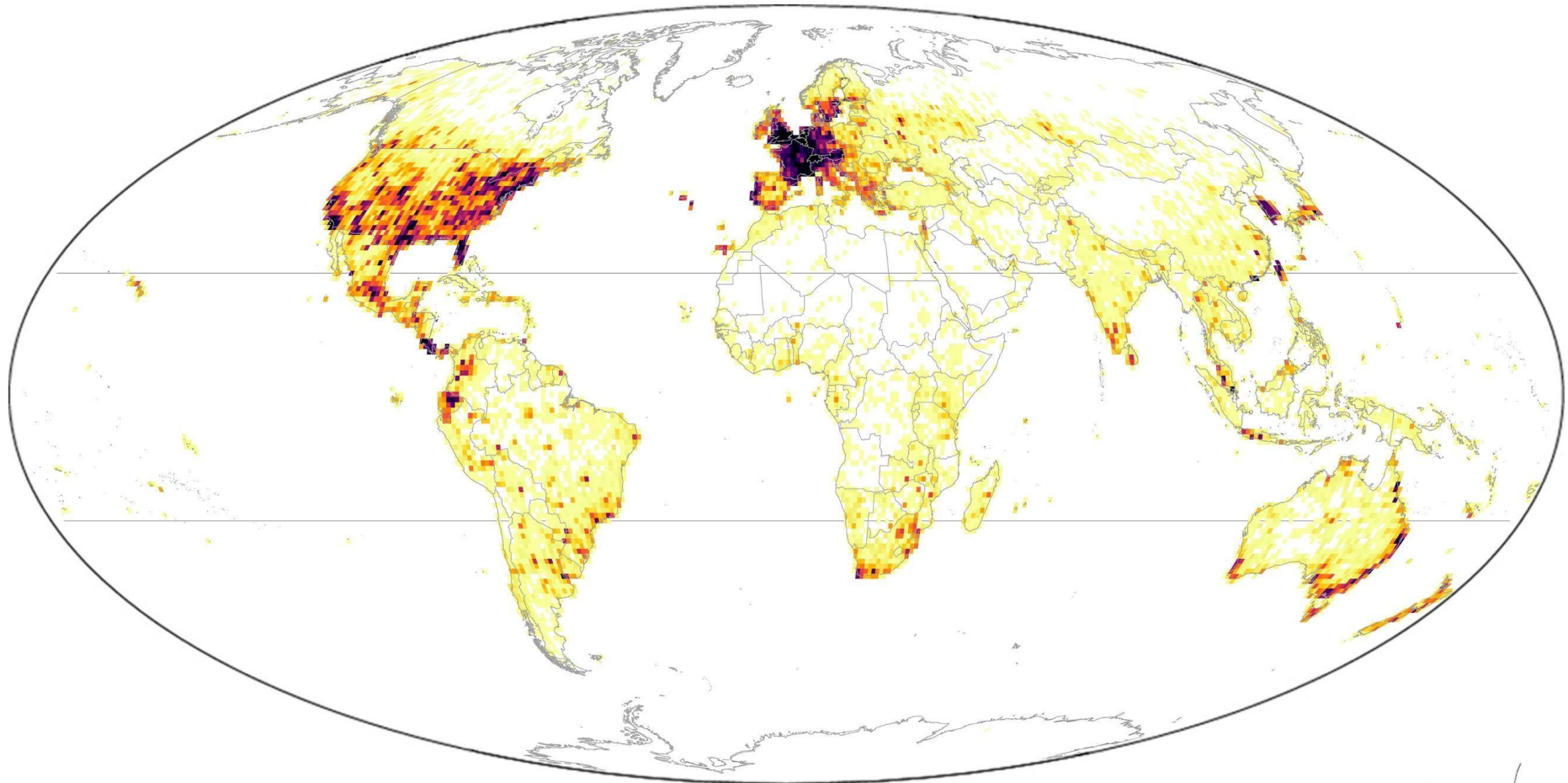
Supplementary Figures



Supplementary Figure 1. Total number of species occurrences records after taxonomic, spatial, and temporal filtering processes (see methods).



Supplementary Figure 2. Total number of occurrences records above species up to order level after spatial and temporal filtering processes (see methods).

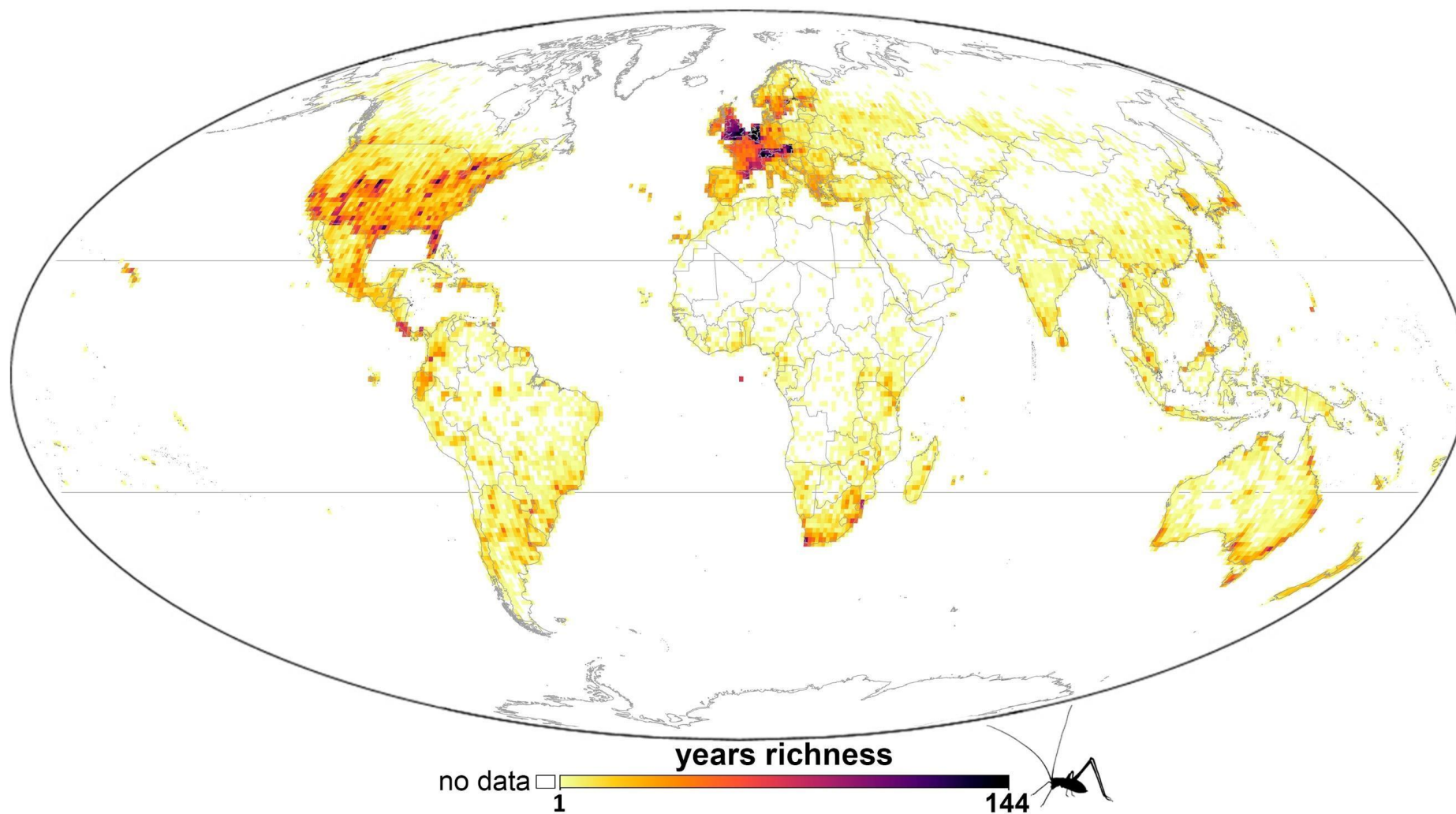


total number of occurrences records = species level + above species up to order level

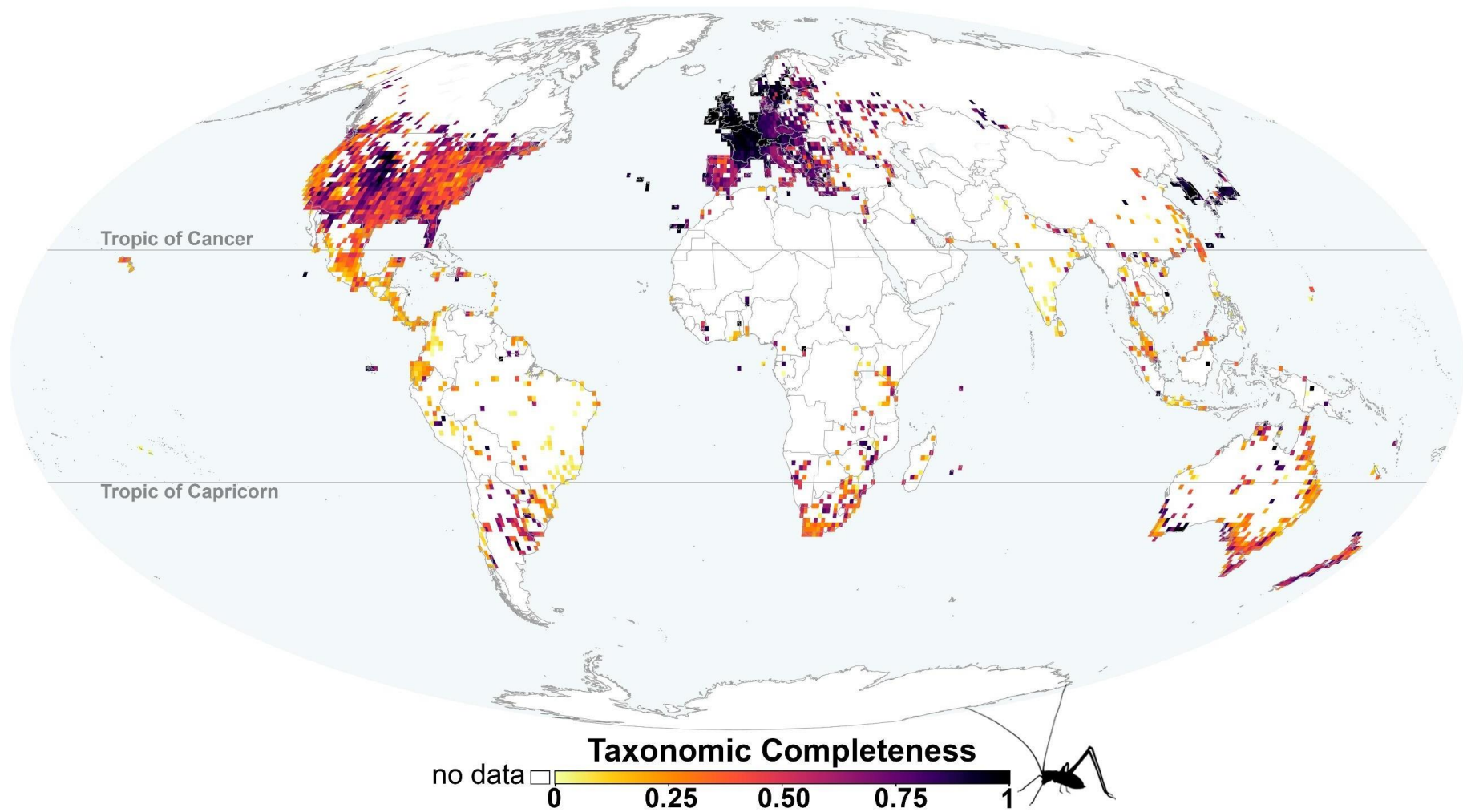
no data  1 76273



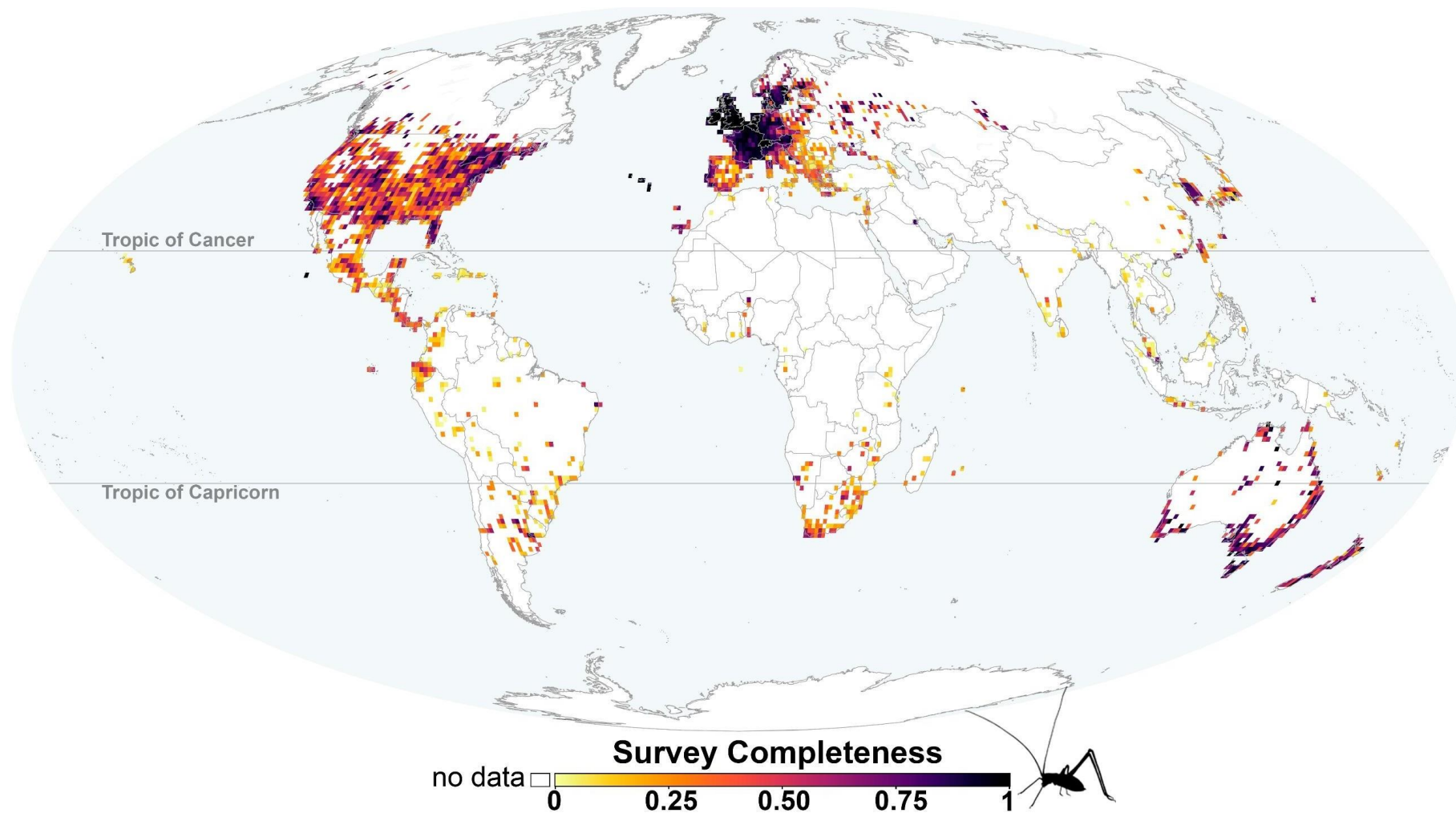
Supplementary Figure 3. Total number of occurrences records resulting from the sum of the total number of species occurrences records and the total number of occurrences records above the species level up to the order level per raster cell (see Figures 1 and 2 mentioned above).



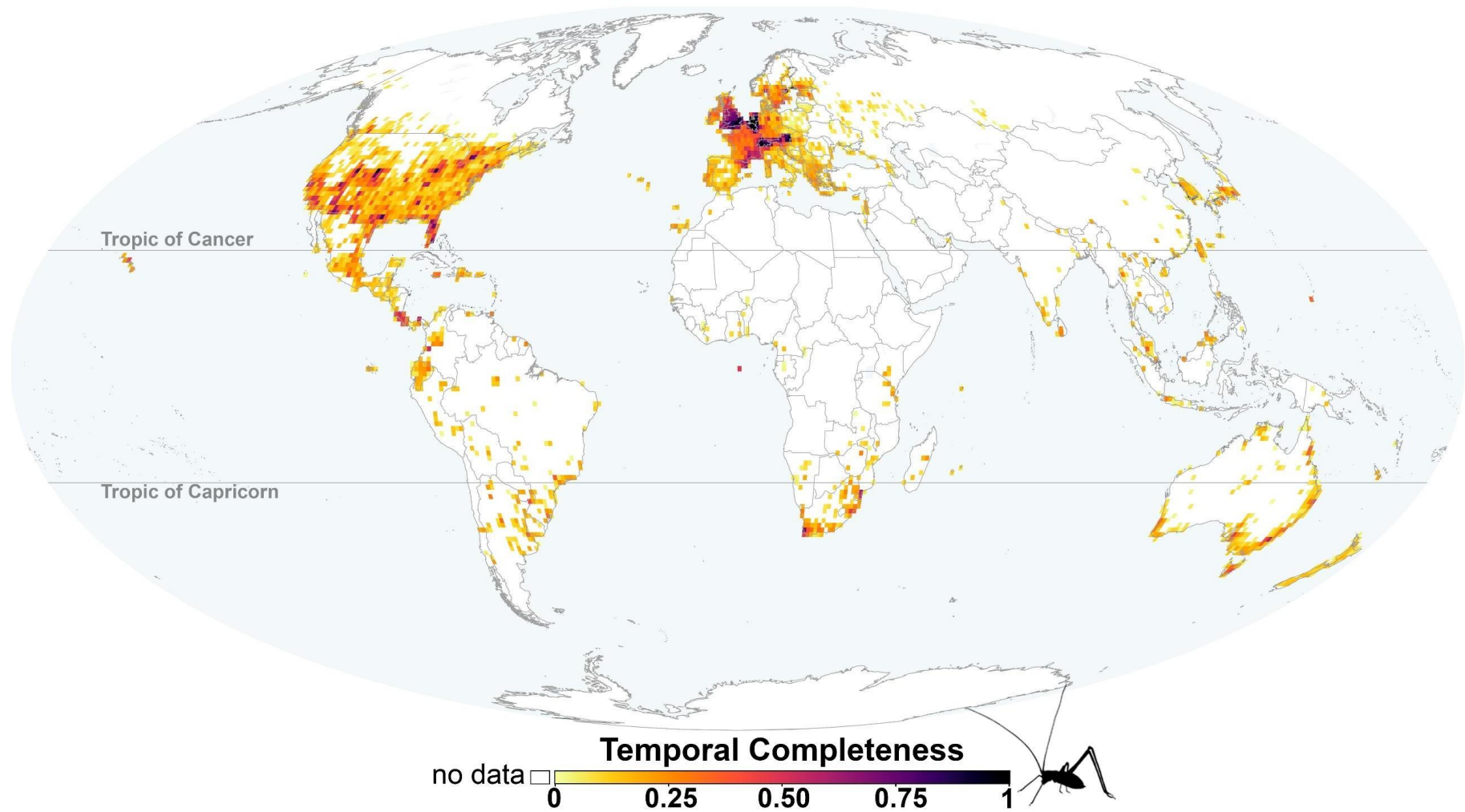
Supplementary Figure 4. Year richness calculated from digitally accessible data for the order Orthoptera for each raster cell.



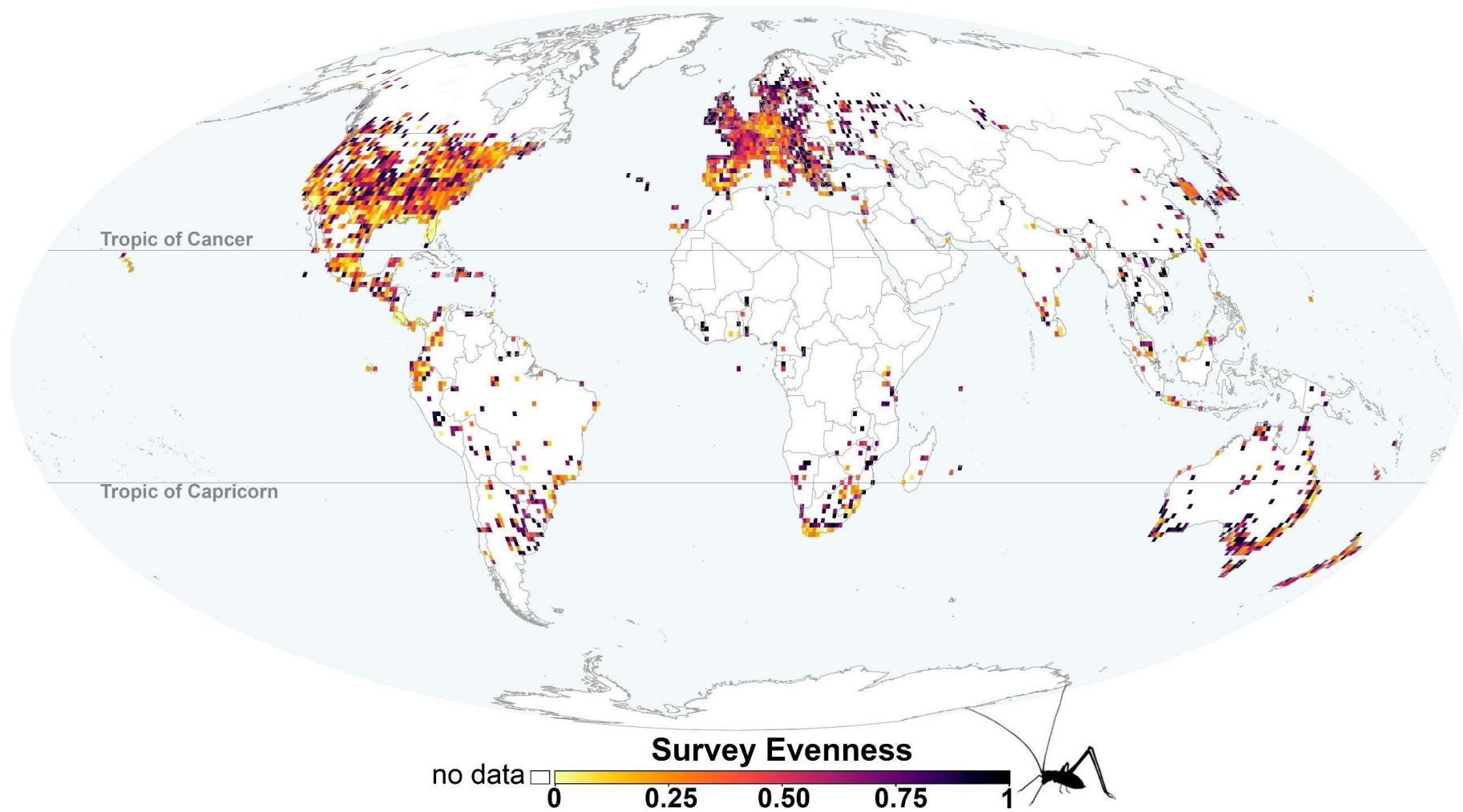
Supplementary Figure 5. Taxonomic Completeness (TaxC), values close to 1 indicate that most records are identified to the species taxonomic level.



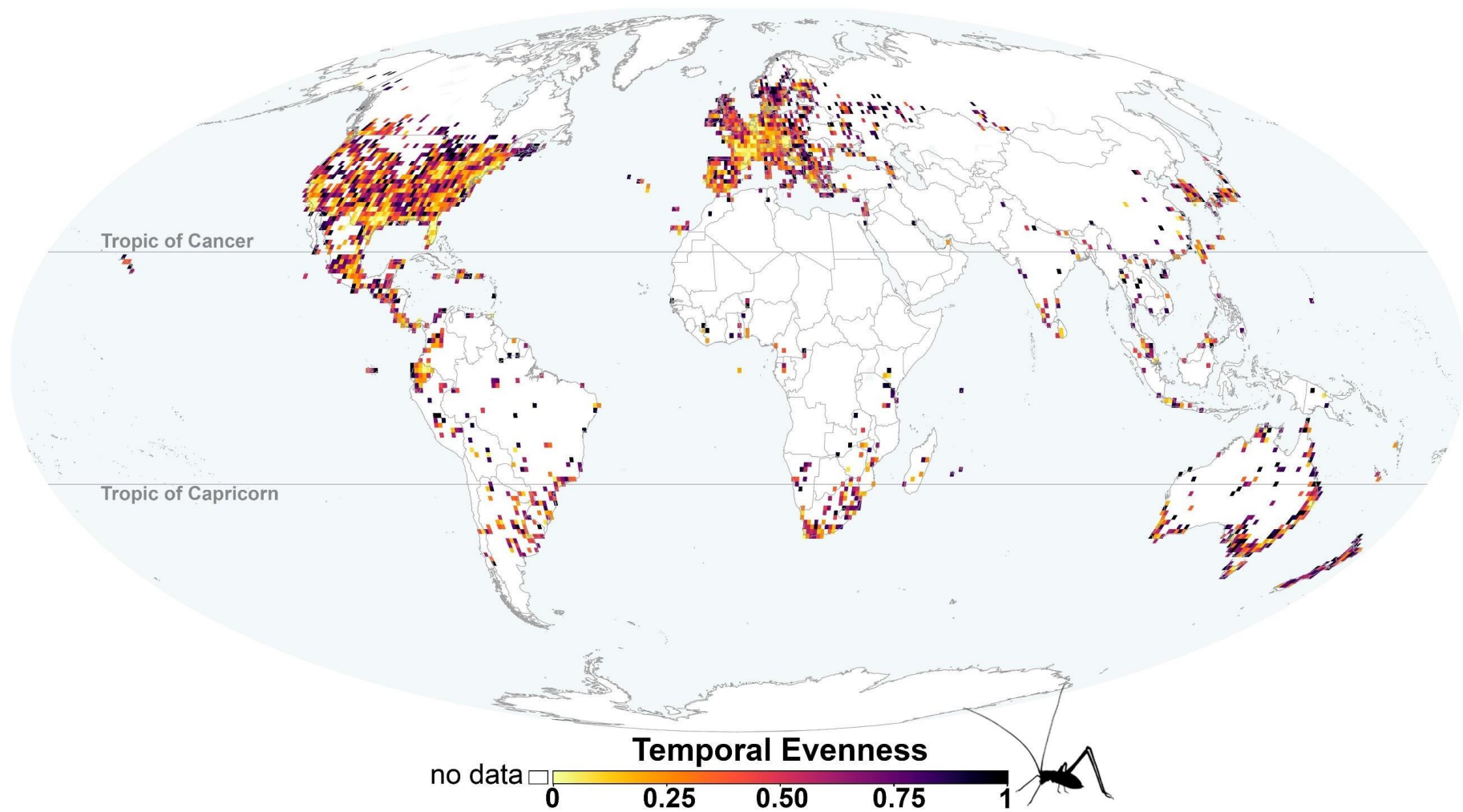
Supplementary Figure 6. Survey Completeness (SC), values close to 1 indicate an almost complete inventory.



Supplementary Figure 7. Temporal Completeness (TE), values close to 1 indicate maximum temporal coverage.



Supplementary Figure 8. Survey Evenness (SE), values close to 1 indicate no local bias in the distribution of the numbers of records per species.



Supplementary Figure 9. Temporal Evenness (TE), values close to 1 indicate no local bias in the distribution of samples per year.