Supplementary Materials for

Advances and shortfalls in knowledge of Antarctic terrestrial and freshwater biodiversity

L. R. Pertierra *et al*.

Corresponding author: L. R. Pertierra, luis.rpertierra@mncn.csic.es

The PDF file includes:

Materials and Methods
Figs. S1 to S4

SUPPLEMENTARY MATERIALS AND METHODS

Species discovery rates and spatial completeness of biodiversity inventories.

Much of the subsequent analyses are operated with a consolidated checklist of terrestrial species for Antarctica. This checklist was generated as a supporting data paper (17) explicitly to guide the analyses of the current work. All the methodological details regarding the checklist compilation and curation are related in the supporting data paper and were scrutinized by both peer reviewers and the journal's technical reviewers. Datasets are both available in GBIF and the Spanish CSIC data server (lodged at https://digital.csic.es/handle/10261/307449.). The list of species comprises organisms with taxonomic classification from where all indicator data is compiled and analysed. The corresponding input data collated is lodged in (16): SUPPORTING FILE 1: BIODIVERSITY CHECKLIST (.xlsx).

***Species accumulation curves (SACs*) (Fig. 1a).** For the generation of the species accumulation curves we retrieved the first year of description in Antarctica for all the species listed. The estimate was restricted to a temporal range of data between 1850 and 2022. To do this, we first downloaded all the records available in the Global Biodiversity Information Facility (GBIF) of the checklist of species. We also extracted the information from the Australian Antarctic Database Centre (121) as curated in (26). As a result, GBIF in combination with (121) provides the best coverage of Antarctic occurrence records available at the moment to our knowledge, hence their integration and usage. We then classified each species based on their biogeographic location as:

A) 'Endemic': those species with occurrences records located exclusively between -60º and -90º Latitude or with no coordinates but assigned originally and exclusively to the Antarctic continent.

B) 'Cosmopolitan': those species with occurrence records placed not only in the Antarctic region but above -60º of Latitude.

C) 'UncertainAntarctic': those species referenced in the Antarctic region whose occurrences records are only located outside of the study area.

D) 'UncertainGlobally': species with no georeferenced information in either GBIF or the Australian list.

Based on the biogeographical status of the information we assigned the first year of the Antarctic description as: 'Endemic': year of species description extracted from the scientific name; or 'Cosmopolitan': eldest year of collection of a record in the Antarctic region extracted from the occurrence records compilation.

The R code written for the Fig 1a. calculations and visualization is lodged in (16): SUPPORTING FILE 2: SACs R CODE (R.)

Spatial inventory completeness analysis (Fig. 1b)

We calculated the spatial completeness of species inventories in the continent using the R package KnowBR (53). The inventory of completeness was calculated using two sources of geographical information. Occurrences records with coordinates placed in the study area uploaded in GBIF.org (18) (accession: 06 March 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.jm8nze together with extra spatial information from an updated version of the spatial occurrence records of (26) listed in (121). Coordinates of the records were rounded to 2 decimals to harmonize the potential lack of precision in the information and to detect potential duplicates. A total of >200,000 records were compiled. We defined cells at a spatial resolution of 50x50 km from a grid that covered all the emerged lands of the Antarctica. Results display the slope of the inventory discovery rates as an indicator of completeness. We used the 'rational' method for the estimations and set the abundance to 1 for each observation. Main results were masked to the cells containing ice-free areas and visually displayed with the licensed GIS software ArcMap 10.3.

The corresponding output data generated is lodged in (16): SUPPORTING FILE 3: ANTARCTIC INVENTORIES SPATIAL COMPLETENESS (.csv).

The R code written for the Fig 1b. calculations and visualization is lodged in (16): SUPPORTING FILE 4: SPATIAL COMPLETENESS R CODE (.R)

Diagrams of biological data mobilization to centralized meta-repositories (Fig. 2)

We depict the completeness of ecological data mobilization towards centralized public meta-repositories. This ‘data mobilization’ reflects a measure of how easily available is to obtain the ‘topic’ information by being present in broad repositories instead of requiring accession to the source publications on a one-by-one basis. It represents a form of mitigation of ‘unknown knowns’. To this end, we use a set of indicators (specifically in the form of measurable parameters of ecological data availability in repositories). For each category we retrieved:

- Whether the species had any entry in a repository. Summarized as the number of species with data per realm (animal, plant, fungi or SAR/protist).
- The fraction of species of a biodiversity subgroup (see categories below) with at least one entry in a repository.
- The average number of entries per biodiversity subgroup in a repository.

One set of representative data mobilization indicators was selected per shortfall discipline. However, this assessment excluded Linnean knowledge levels as we use the realized inventories as backbone references. And also Prestonian population dynamics since there is barely meta-population information for vertebrates and vasculars in IUCN website (the originally selected Prestonian indicator databank), using as reference the population estimates of the IUCN red list). Data is lodged in (16): SUPPORTING FILE 5: INDICATOR DATABASE (.xlsx’). It must be noted that these indicators do not provide a general view of the status for each shortfall as they cover many elements, but rather provide an example of a widespread metric typically used in some elements of the shortfall dimension. These are:

1. BIOGEOGRAPHICAL SPATIAL DATA. 'Wallacean' data mobilization was retrieved using the indicator “species with spatial records in GBIF repository”. Input data was obtained from Figure 1b (S3-4).
2. ABIOTIC TOLERANCE DATA. 'Hutchinsonian' was evaluated with the indicator “species with thermal tolerance limits in Globtherm compendium” (22). However, in this case we updated the compendium with additional entries of Antarctic species with thermal profiles from Antarctic research literature. Data compilation was searched with literature retrieval queries of “Antarctic* AND tolerance OR thermal limit” at Web of Science. The WoS query code written for the searches is lodged in (16): SUPPORTING FILE 6: BIOTIC INTERACTION DATA RETRIEVAL WOS CODE (.txt).
3. EVOLUTIONARY DATA. 'Darwinian' was evaluated with the indicator: species with genetic sequences in GenBank repository (19) by retrieving the number of sequences associated to each (17) taxon. See supporting PEARL code for data retrieval in Genbank The Pearl code written for the calculations and visualization is lodged in (16): SUPPORTING FILE 7: GENETIC DATA RETRIEVAL PEARL CODE (.txt)

4. TRAIT DATA. 'Raunkieran' was evaluated with the indicator "species. with trait data in centralized repositories". For the case of vertebrates it was Pantheria. For lichen and vascular it was TRY database. For bryophytes it was an extension of TRY named BryoForTraits (largely biased towards forest species, but the most complete compendium available). Fungi data was complemented from FungalTraits repository. For other groups we could not find centralized trait repositories and so were all scored as 0. TRY data accession queries were requested via https://www.try-db.org/TryWeb/dp.php (21). We obtained all fungal functional traits available for our curated species inventory using the fungaltraits v0.0.3 R package. We recorded which fungal species had functional traits present as well as counting how many individual records of functional traits have been recorded for each species on our inventory.
5. BIOTIC INTERACTION AND NETWORK DATA. 'Eltonian' was evaluated with the binary indicator "species with reported interactions listed in GLOBI repository (20). The R code written for the calculations and visualization is lodged in (16): SUPPORTING FILE 8: BIOTIC INTERACTION DATA RETRIEVAL R CODE (.R).

Figure 2's Venn diagrams of data mobilization and histograms of fractional coverage/average number of entries to global online repositories were generated from all (or some) subclades for eight biodiversity groups (2 per realm in sensu lato):

1. VERTEBRATES: Seals and birds.
2. INVERTEBRATES: free-living arthropods, free-living non-arthropod invertebrates, and parasites (arthropod and non-arthropod). See breakdown in supplementary.
3. HIGHER PLANTAE: Vascular plants.
4. BRYOPHYTES: Mosses, hornworts and liverworts.
5. LICHEN: Lichenized fungi.
6. FUNGI: Free-living fungi.
7. ALGAE: Green algae plus diatoms.
8. PROTISTS: Heterotrophic SAR/Chromista.

Note: no diagrams could be made for microorganism groups due to their sheer large numbers and absence of consolidated lists of individual species. This comes as a limitation due to the uncertainty levels on their diversity. Precisely because of such operational limitations the elicitation of Figure 3 complements such caveats by looking at the overall paucity of progression seen by experts. Thus, we restricted here the analysis to the species of the eukaryotic checklist of Antarctic biodiversity (17). All indicator information per species can be seen in the supporting dataset file 3 – sheet 1 in the corresponding columns of the same name as a binary character (1 = presence, 0 = absence). Also, the raw number of entries per species per repository is also supplied (additional sheets).

For all analyses we used a curated checklist of eukaryotic species, available in (17). The list of species comprises about 2000 organisms with taxonomic classification from where all indicator data is compiled and analysed.

The R code written for the calculations and visualization is lodged in (16): SUPPORTING FILE 9: DATA MOBILIZATION VENN DIAGRAMS R CODE (.R)

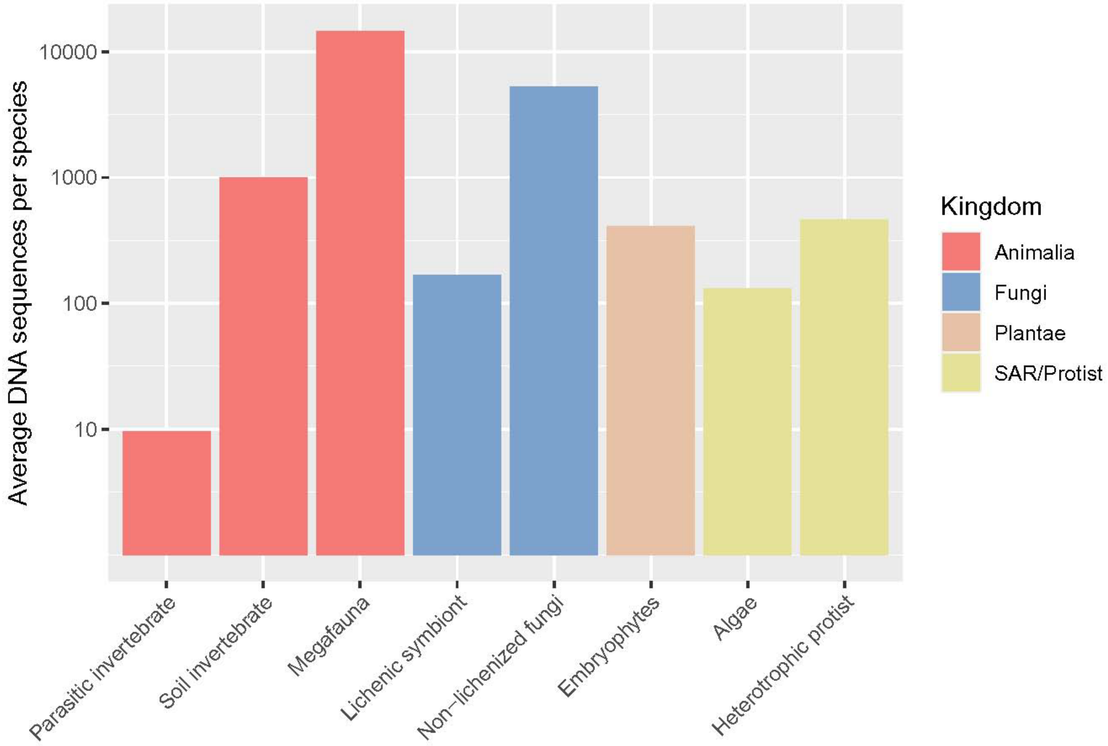
Expert elicitations of ecological knowledge shortfalls by biodiversity group (Fig. 3). A complementary screening and systematic revision of the state of knowledge under the framework of shortfalls was performed under a structured IDEA (Investigate, Discuss, Estimate and Aggregate) elicitation process (24) paired with a literature narrative synthesis. A diverse group with expertise in different aspects of ecological research were approached to jointly assess the global state of shortfalls in knowledge of Antarctic biodiversity and ecosystem functioning. The group of experts were selected to confer diversity (age, sex and nationality). All experts were biologists with specialization in different taxa and ecological disciplines (more information on their background is provided in the supplementary methods). The admixture of experts was conceived to enrich the discussions from multiple specialization perspectives. All potential knowledge gaps were systematically arranged per discipline, then the experts examined the relative knowledge progression of each group. Different taxonomic groups / ecological disciplines had a variable number of experts based on their expertise. Experts screened relevant literature first, splitting all biodiversity groups to ensure no groups were left untouched. To support this available Antarctic publications listed in Web of Knowledge with basic keywords retrieved per group were scrutinized (17). In order to assist with the evaluations, several experts were asked to do an exercise of classification of a list of papers per ecological discipline (see keyword queries in 17). Also, as part of the elicitation, three rounds of online group meetings were arranged for joint discussions allowing for revisited scoring. Sessions were guided by 3 moderators and also split into fauna, flora and microorganisms. Individual scores were always kept separate and private, with the group averages merged for display following the IDEA formula (24). Individual scorings were also guided by the collated indicators, with the experts allowed to revisit their ponderings all along the elicitation process. A final online meeting was held in March 2024 to offer a last opportunity for reconsideration. Data lodged in (16): SUPPORTING DATA FILE 10: EXPERT ELICITATION ON ANTARCTIC ECOLOGICAL KNOWLEDGE (.csv).

Estimates were provided in a frequency completion format (%). These experts were asked to provide insight on the levels of our understanding of Antarctic biodiversity patterns and processes, identify knowledge gaps and unknowns, discuss the implications of the current state of knowledge for both theoretical understanding and practical conservation, and propose targeted mitigation efforts. Evaluations were organized into four groups of terrestrial, semi-terrestrial and/or freshwater species: vertebrates (breeding birds and seals), invertebrates (arthropods and other microinvertebrates, free-living and parasitic), flora (vascular plants, bryophytes, lichen and fungi), algae and protists, and microorganisms (bacteria, archaea and viruses). Experts were questioned about the status of the different groups of biota present in Antarctic terrestrial & freshwater ecosystems following a four-step expert elicitation protocol (IDEA). We used standardized credible intervals (80%) together with levels of confidence given by the participants following the Hemmings et al. (2018) formula (24). In cases where the adjusted intervals fell beyond reasonable bounds (such as [0, 100%] for probabilities), we truncated distributions at their extremes. Experts were encouraged to revise their estimates in Round 2 if the extrapolation did not represent their true belief. Main summary results are then summarized in 8 categories (2 per realm in sensu lato). All the provided responses (coded in the form of anonymous numerical scores) from the opinions of all the participant experts were stored in a summary table (. These materials were generated with the following methodological procedures and used for the below calculations and visualizations.

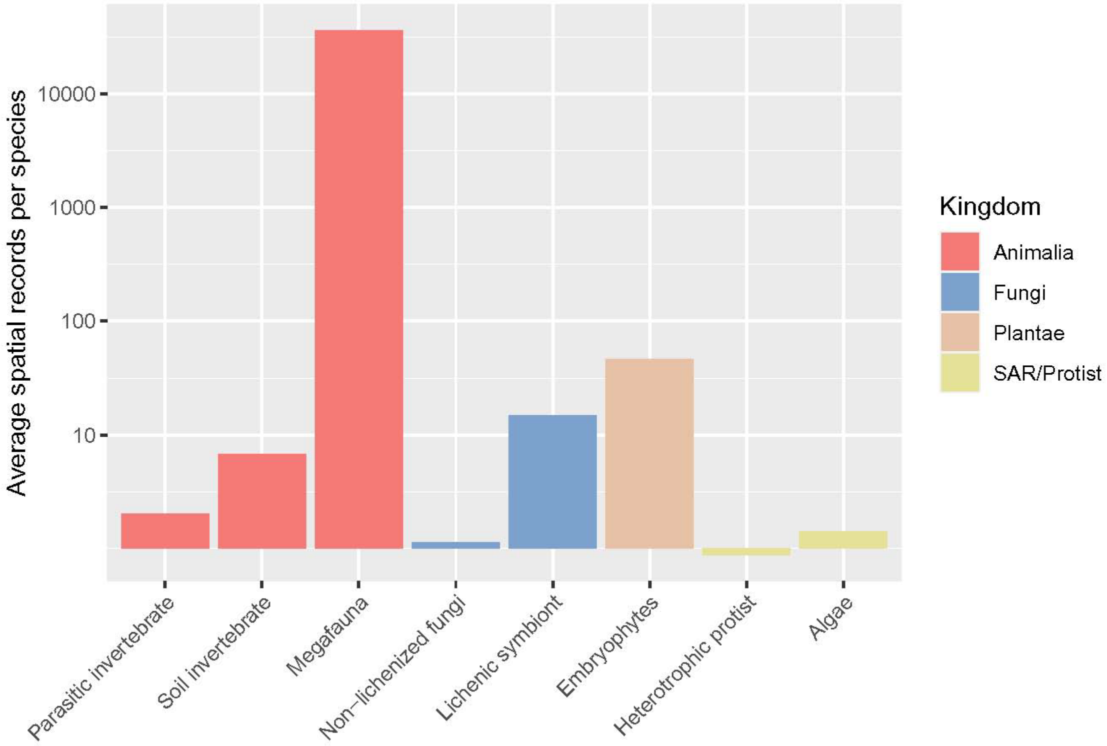
Conservation value. This score was calculated as the integration of the depth of the knowledge gap set by Antarctic biodiversity experts, and the joint general value for conservation practices (set as climate change action, biosecurity, pollution control, living resources management and area protection) set by 5 shortfall science experts. Blue colors indicate high filling importance, and red colors indicate comparatively low one.

The R code written for the calculations and visualization of Fig. 3 is lodged in (16): SUPPORTING FILE 11: EXPERT ELICITATION R CODE (.R)

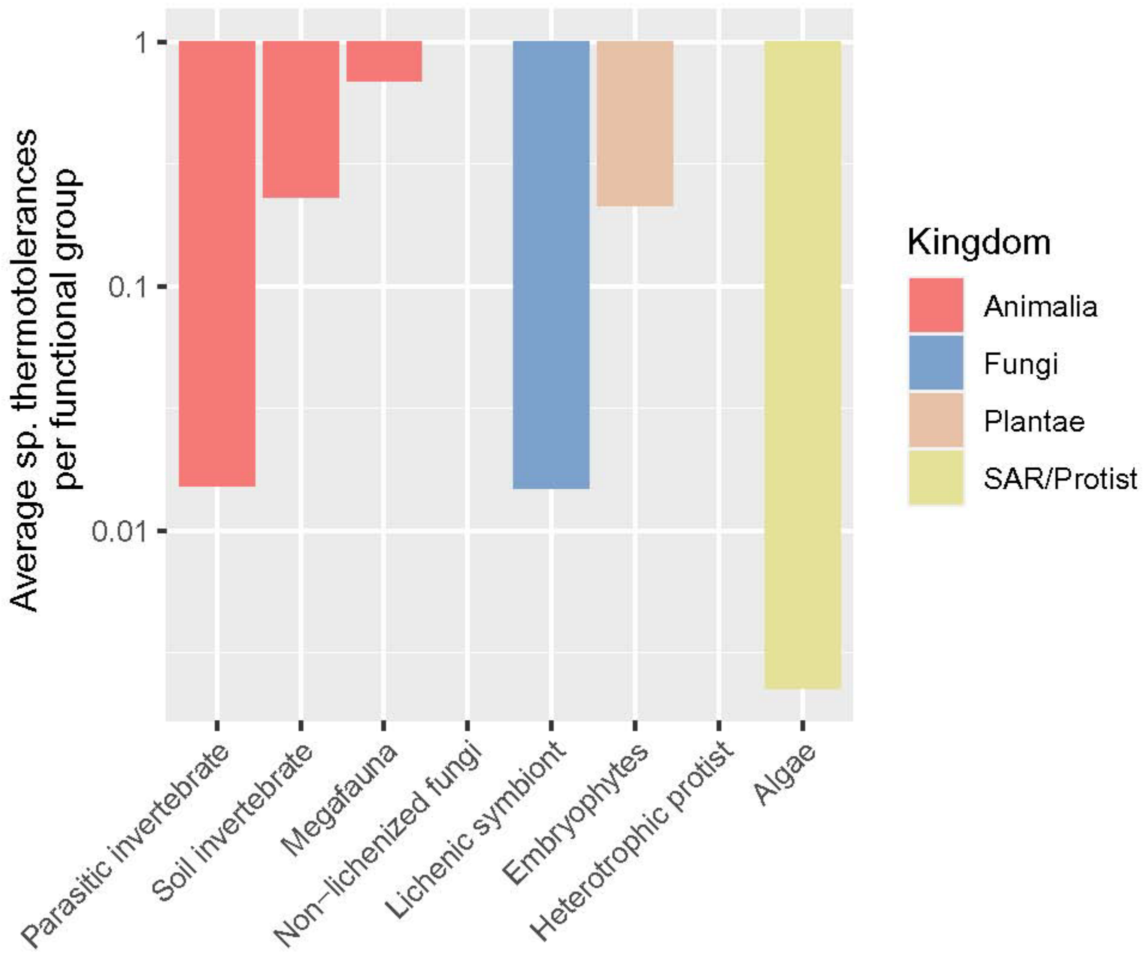
SUPPLEMENTARY FIGURES



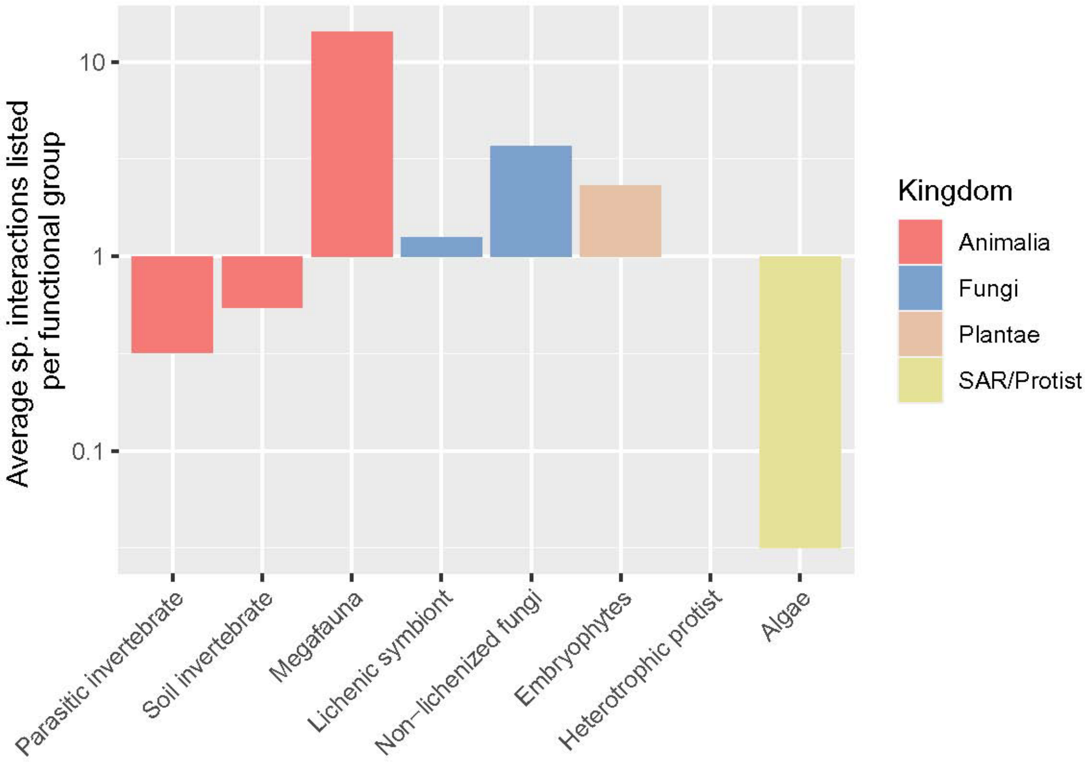
S1. Average DNA sequences lodged in Genbank databank per Antarctic species in each biodiversity subgroup.



S2. Average occurrence records lodged in GBIF databank per Antarctic species in each biodiversity subgroup.



S3. Average thermal tolerances experimentally measured lodged in Globthem compilation per Antarctic species in each biodiversity subgroup.



S4. Average species interactions lodged in GLOBI databank per Antarctic species in each biodiversity subgroup.