Where Are the New Species?: Taxonomic Discoveries in 2022

BSc Honours Marine & Freshwater Biology

BIOL4246P Investigative Honours Project

Student ID: 2433429

Supervisor: Prof. Roderic Page

Date of Submission: 15/03/2023

Word count: 4,993

Table of Contents

Abstract

The future of taxonomy is relatively unknown, with many worried that it is a dying science. The demise of taxonomy is mainly due to insufficient funding and a loss of expertise. Taxonomy is incredibly important for many forms of biology, such as evolution and conservation, where describing and classifying species is essentially the base of their studies. In the 250 year history of taxonomy, certain taxa and geographic areas have been more heavily studied than others. Journal paywalls have meant that access to taxonomic research is financially restricted This limits the access to research for institutions which are smaller or from developing countries. This study aims to investigate the taxonomic discoveries of 2022 in terms of taxonomic, spatial, researcher, and funding distributions, as well as accessibility. The dataset was obtained from BioRSS, which monitors journal websites and search engines such as Google Scholar for articles describing new species and groups them taxonomically and geographically. Where publications had Digital Object Identifiers (DOIs) services from CrossRef, ORCID, and Unpaywall were used to obtain information on authors, funders, and open access status of those publications. The top discoveries were Arthropods and Tracheophytes, although the proportion of discoveries were not as expected given the number of known species per phylum. China was a front runner in taxonomic discoveries of 2022 producing 25% of the discoveries and therefore had the highest number of funding agencies. There were 66% of researchers residing in the country of species origin, with the top three including Brazil, China, and India. Nearly half of the articles were open access. Direct comparison of data has been difficult due a lack of information available. In this ever evolving world of biology going online and an increase in accessibility, a website which shows the world of taxonomy could be a useful tool.

Introduction

As one of the oldest disciplines in biology, taxonomy discovers, describes, names and organises species into classifications (Guerra García, et al., 2008). Taxonomy dates back approximately 250 years to Carl Linnaeus (1707-1778) who is considered the founding father of modern taxonomy (Tahseen, 2014). Linnaeus introduced the binomial scheme and principles of biological classification which have since been adapted to the nomenclature system now used in taxonomy (Tahseen, 2014). Traditionally taxonomy comprises of three levels; alpha, beta and gamma. Alpha taxonomy is at the surface level of the species whereby it is described and named. Beta taxonomy identifies species within natural groups and biological classes. Gamma taxonomy analyses the evolutionary processes and patterns of the species (Tahseen, 2014). This delineation of the world's species has become a crucial component to many other biological studies concerning biodiversity such as evolution, ecology, conservation and biogeography (Ebach, et al., 2011). There would be many losses without the study of taxonomy. Many species losses would be unknown, and there would be a lack of focus within conservational planning. Further, there would be many opportunities for medicinal discoveries that would be missed, and it would be difficult to track disease outbreaks (Agnarsson and Kunter, 2007; Grieneisen, et al., 2014). Currently there are over 2 million species known on Earth, however, many scientists have estimated the actual number of species on Earth and these estimates can ranges from 5 million to 30 million or more (Fontaine, et al., 2012). Therefore, discoveries made by taxonomists so far are potentially only a fraction of species which could be discovered in this world.

Over the last decade there has been increasing worry among taxonomist for the future of taxonomy. With the millions of species to potentially discover, at the current rate of species discovery it will take centuries to reach these estimates (Fontaine, et al., 2012). Due to the current major biodiversity loss, which has resulted from habitat destruction, pollution and overexploitation, many taxonomists feel that it is now a race against time. Taxonomy is facing two driving forces against its decline: (1) insufficient funding along with (2) a lack of taxonomic experts (Britz, 2020). Funding for taxonomy has been on the decline while other scientific professions have become more of a priority. Technological advance in disciplines such as molecular biology has shifted the monetary focus away from traditional taxonomy (Tahseen, 2014). However, there are some projects, such as the NSF-funded Planetary Biodiversity Inventories and Partnerships for Enhancing Expertise in Taxonomy, which are trying to fight this taxonomic impediment (Britz, et al., 2020).

In 2014, the world's taxonomic workforce was believed to be around 4000 to 6000 people and there is concern that this number is dropping, particularly in Europe and North America (Grieneisen, et al., 2014). The lack of taxonomic experts is partly due to the decline of young scientists in Europe becoming taxonomists to fill the gap of ones which are retiring (Tahseen, 2014). Additionally, there is an imbalance in taxonomists from developing countries which are known to be biodiversity rich. Historically, taxonomists from Europe and North America were involved with most of the species discovery around the world (Liu, et al., 2023). Even though there has been a loss of taxonomists in the western society, there has in contrast been an increase in resident taxonomists in countries such as Brazil, China and Colombia (Liu, et al., 2023). This shift away from the colonial history of taxonomy could show a push in taxonomic discoveries in these developing countries which are known to be biodiversity rich (Fontaine, et al., 2012).

As previously mentioned, taxonomy is highly important for conservation, however, the knowledge of species taxonomically and spatially is extremely patchy due to taxonomic bias (dos Santos, et al., 2020). Generally, species which are highly researched tend to be large and charismatic, mainly plants and vertebrates. Despite this deep understanding of some species, it is not useful on its own for efficient conservational planning. Species which are rare, small and uncharismatic often play key roles in the functioning of the ecosystem (Troudet, et al., 2017). Additionally, the knowledge of many species in minimal due to the feasibility of the research, particularly when requiring fieldwork. The species could be located in an area which is extremely difficult and costly to get to, as well as being hard to observe. For example, the species could present nocturnal activity patterns or have cryptic coloration (dos Santos, et al., 2020). In an attempt to combat the taxonomic bias, the UN Convention of Biological Diversity created the Aichi Biodiversity Targets which includes 20 targets although multiple researchers have questioned the effectiveness of these targets (McRae, et al., 2017).

In the process of reducing taxonomic bias, being aware of the taxonomic discoveries could highlight taxa which are being forgotten. Unfortunately accessing scientific articles can be difficult. Over the last decade it has been a pressing topic that an estimate of 75% of scientific literature has limited access due to paywalls (Himmelstein, et al., 2018). This has put increasing pressure on institutions, such as universities, research centres, and medical centres, which need to pay subscriptions for journals. The price of journal subscriptions has increased at faster rate than inflation (Himmelstein, et al., 2018). High prices for journal subscriptions has resulted in many libraries lacking in particular areas of literature, potentially depriving their users of vital information, as they can not afford to buy all the journals. Furthermore, often institutes which are smaller or in developing countries have poor access to the literature hidden behind these paywalls (Himmelstein, et al., 2018). Due to this imbalance and lack of accessibility to important resources, a movement towards "open access" has arisen. Open access comes in many different forms; gold, hybrid, and green. Within gold, journals can give free access to all of their articles, which is termed as direct, or give free access to their articles after a period of time, which is termed as delayed (Lewis, 2012). For some journals, hybrid open access occurs when authors pay to make there article freely available (Lewis, 2012). Finally, green is the most restricted version of open access, where a summarised version of the article is freely available, often on the website of the author or institution (Lewis, 2012). From the occurrence of open access, it can now be seen that over 50% of scientific articles can be found on the internet for free (Archambault, et al., 2014). The use of open access has shown benefits for outreach in the scientific world as these articles were cited more than ones which required payment (Archambault, et al., 2014). Even though open access is on the rise, many individuals have taken it into their own hands through piracy of scientific articles. An example of this is which has been running since 2011 is Sci-Hub (Himmelstein, et al., 2018). Sci-hub is a pirate site which provides free access to articles which would normally be restricted. It was founded by a graduate student called Alexandra Elbakyan who had the goal to provide free and unrestricted access to knowledge (Elbakyan, 2016). However, Sci-hub has had issues previously due to copyright licensing as the brand does not only collate openly licensed content (Himmelstein, et al., 2018).

This study aims to explore the taxonomic discoveries of 2022 in terms of taxonomic, spatial, researcher, and funding distributions, as well as accessibility. These discoveries will be dissected to investigate if the discovery rate for each taxa is on the same trajectory as previously shown and whether there is still taxonomic bias occurring. To display where taxonomists are in the world nowadays, if these developing countries which are biodiversity rich have a high representation in taxonomy. Furthermore, to show which countries are the current front runners in terms of funding for taxonomy.

Finally, as there has been a rise in open access papers, this study will investigate the accessibility of papers within 2022 to find the percentage of open access.

Investigative Methods

In this study, a year’s worth of data from 2022 was obtained from the BioRSS project (BioRSS, 2022). Created by Roderic Page, inspired by uBioRSS (Leary, et al., 2007) the BioRSS project regularly harvested papers, which published new species descriptions, from journals or Google Scholar using RSS feeds. RSS is used on websites where content changes, such as news, papers and podcasts, keeping subscribers up to date. This is a commonly used method to get updated articles from scientific journals (Leary, et al., 2007). The RSS feeds can be set to the subscribers preference (Holvoet, 2006), for example, BioRSS was set to search for terms such as “new species” and “n.sp.”. BioRSS also had the ability to geolocate papers and find the taxonomic name of the species. The country of species was found based on the title and abstract of a paper using the “Glasgow geoparser” (Page, 2021). Geoparsing uses name-entity recognition (NER) which identifies named entities, such as place names, and then a gazetteer is used to match geographic names found by NER (Alex, et al., 2016). For the taxonomic names, taxon finder (*taxonfinder.org | Find scientific names in text)* is used to find the names which are matched with taxonomic names in GBIF (GBIF: The Global Biodiversity Information Facility, 2023). For 2022 BioRSS provides data on the number of articles it harvested, whether those articles has a DOI, what (if any country) was mentioned in the title or abstract of the article, and what (if any) taxa were mentioned (figure 1).

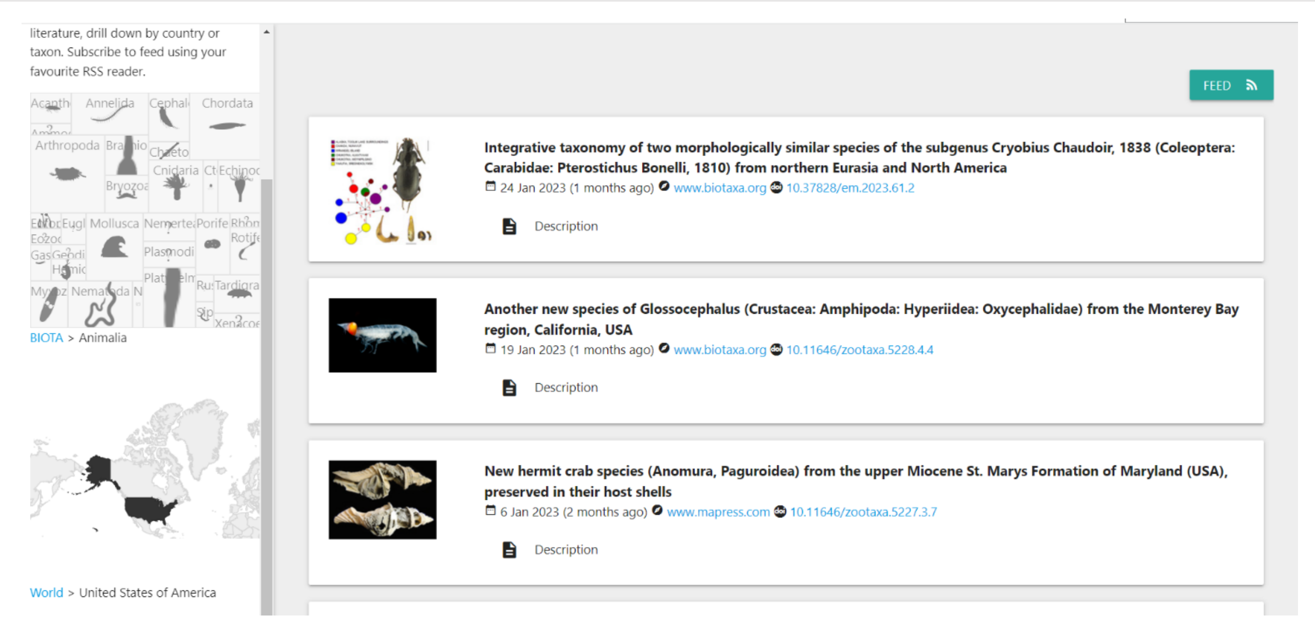


Figure 1. Image of BioRSS where a country (America) and kingdom (Animalia) have been selected.

With the data gathered by BioRSS, it was compared against the taxonomic distribution of all known species on Earth as well as the total number of occurrences for each country. To get this, Catalogue of Life was used to get the total number of known species as well as the distribution between phyla

(Bánki, et al., 2023).. GBIF was used to get the total number of occurrences in each country (GBIF: The Global Biodiversity Information Facility, 2023).

To investigate the background of the researchers involved in the discovery of species in 2022, the digital object identifiers (DOI) of the articles were extracted. For each DOI the CrossRef web services were used to retrieve the ORCID IDs for authors, where those were declared by the authors (Hendricks). If an author had an ORCID ID that was used to access the author's profile in ORCID and that profile often included information on where the author was based (*ORCID*). Additionally, CrossRef often includes information on the agencies that funded the research in a given paper. These agencies are identified by a DOI, which can in turn be used to query Wikidata for further information about that funder such as the country that they are based in (*Wikidata*). This enabled a comparison between the location of the paper, authors, and funding agency.

Finally, to determine the accessibility of the paper, the DOI for each paper was run through a service called Unpaywall to assess whether the paper was open access or not (*Unpaywall: An open database of 20 million free scholarly articles*). The Unpaywall project attempts to legally find a free to read version of scientific articles. This gave an understanding to the extent papers in 2022 were open access as well as a comparison of the accessibility across countries.

Results

Taxonomic Distribution

During 2022, the BioRSS project harvested 8,567 articles in 2022, of which 1,766 lacked a species name or mention of a geographic place name. Of the species which were described in the dataset, the taxonomic distribution displayed a dominance of the Animalia kingdom (figure 2), in particular the phylum Arthropoda representing 35% of the dataset. Animalia is also prominently represented by Chordata (12%) along with Mollusca (2.5%). Taxonomists further made discoveries in the Plantae kingdom in which 93% of the species in Plantae were classified as Tracheophytes. Overall the Tracheophytes represent 20% of the species discovered in 2022. Fungi and Bacteria were discovered at similar rates with Fungi representing 9% of species and Bacteria representing 8%.

Currently all species on Earth which have been discovered equates to 2,088,949 (Bánki, et al., 2023). Similarly to 2022, the Animalia kingdom dominates this Earth with 55% of species classified as Arthropods (figure 3). Furthermore, comparable to 2022 (figure 2), the Molluscs (6%) and Chordates (3.5%) are next in line within Animalia, although, in the opposite order (figure 3). Tracheophytes also summate a large proportion of the worlds diversity with 17% of species. In terms of Fungi and Bacteria, there are 7% fungi and only 0.5% are bacteria.

To compare the taxonomic distribution between 2022 and overall, the proportions for each phylum were calculated and compared (figure 4). Even though Arthropoda represented 35% of the 2022 dataset, it was found that the number of species articles were less than expected given the number of known Arthropods. This was a similar outcome for the Molluscs, although a considerably less difference. Phyla which displayed an increase in discoveries over 2022 included Chordata, Tracheophyta, Proteobacteria, Basidiomycota, Annelida, and Ascomycota. Most phyla only displayed a slight increase, however, the Chordates displayed the greatest increase overall.

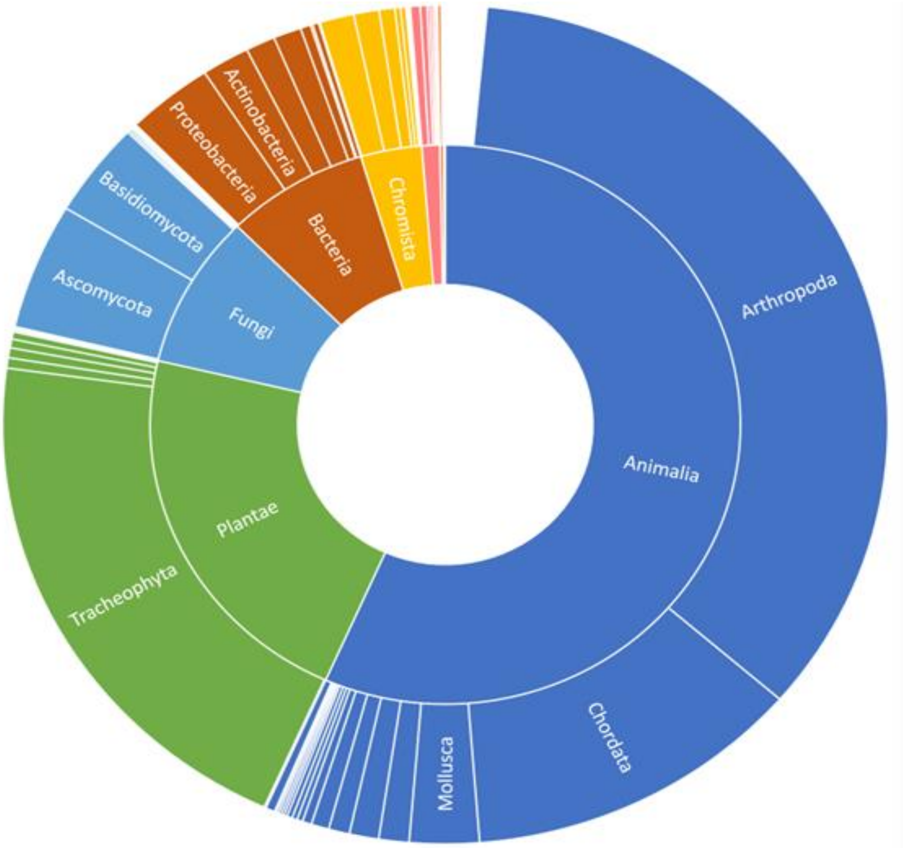


Figure 2. Sunburst displaying the taxonomic distribution of new species discovered in 2022. Includes the kingdom and phylum levels to which the species are categorised in.

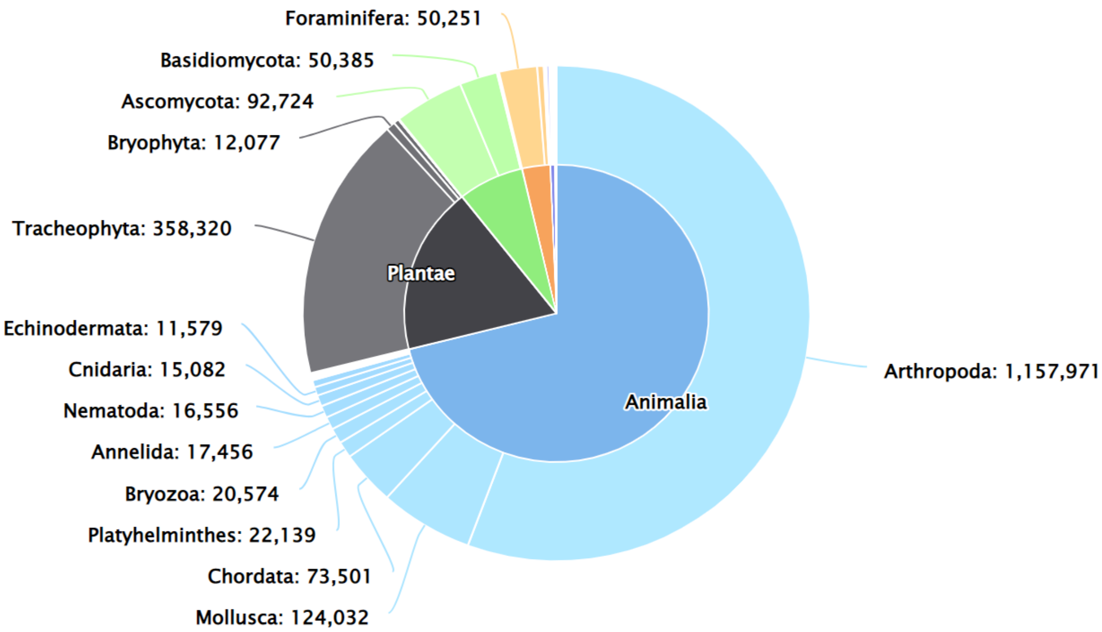


Figure 3. Sunburst displaying the taxonomic distribution of all species on Earth. Includes the kingdom and phylum levels to which the species are categorised in (Bánki, et al., 2023).

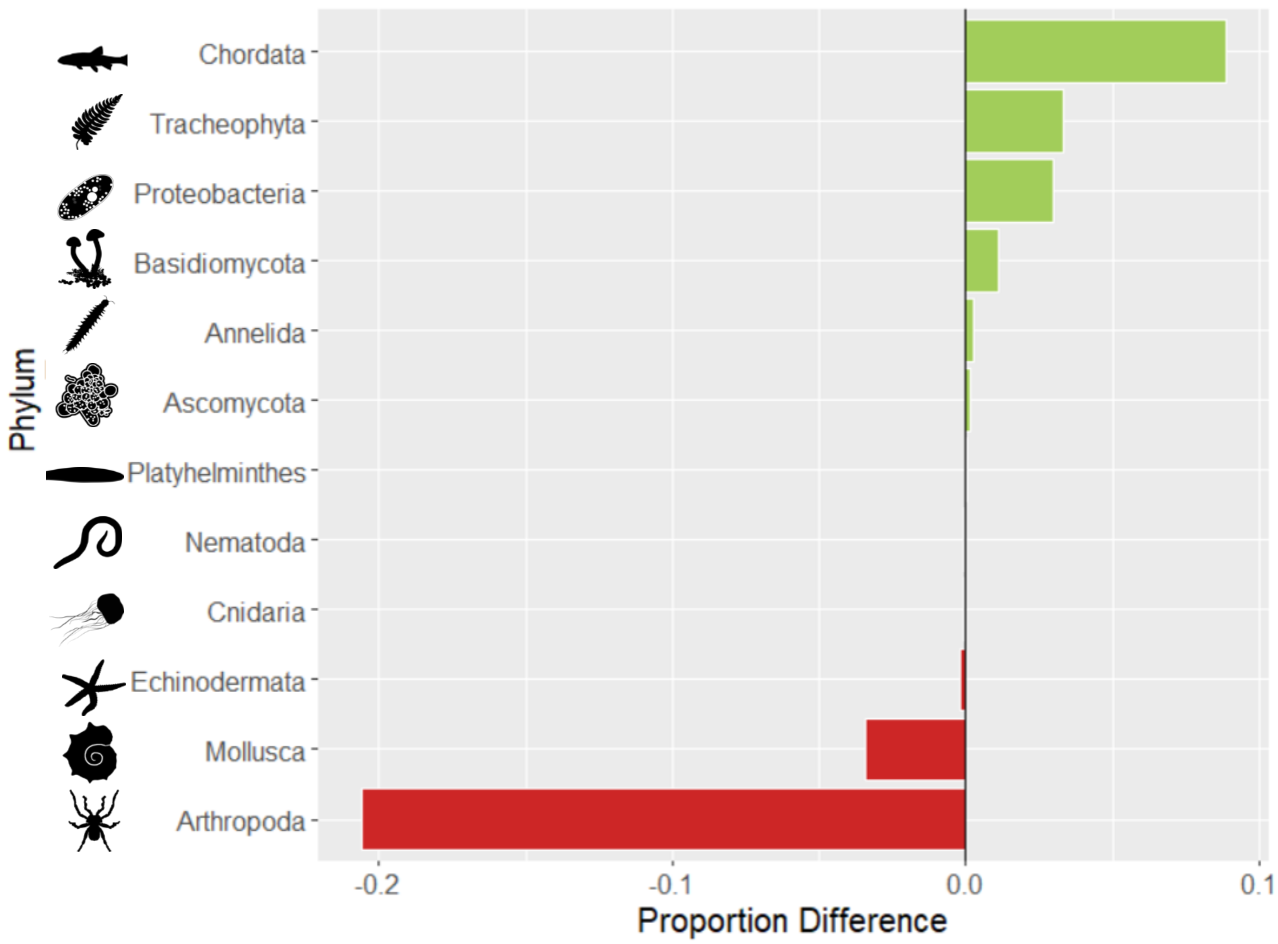


Figure 4. Bar graph displaying the proportional difference between number of articles in 2022 and that expected based on Phyla numbers in Catalogue of Life. The vertical line at x = 0 depicts the expected proportions from Catalogue of Life (Bánki, et al., 2023). Green (right) and red (left) bars show whether the phylum exhibited a higher or lower proportion in the 2022 dataset.

Spatial Distribution

Through BioRSS, 8,208 papers were able to be linked to 155 different countries indicating the origin of the species discovery. Although the world's oceans cover over 70% of the world, BioRSS was limited by country names and so the marine environment was not accounted for. The front runner for species discovery in 2022 was China, with 2,086 papers released (figure 5). The rest of the top five countries were as follows; India (783), Brazil, (560), America (299), and Mexico (288). Within the 155 countries, 141 of them released under 100 papers. Furthermore, 66 of these countries released under 10 papers.

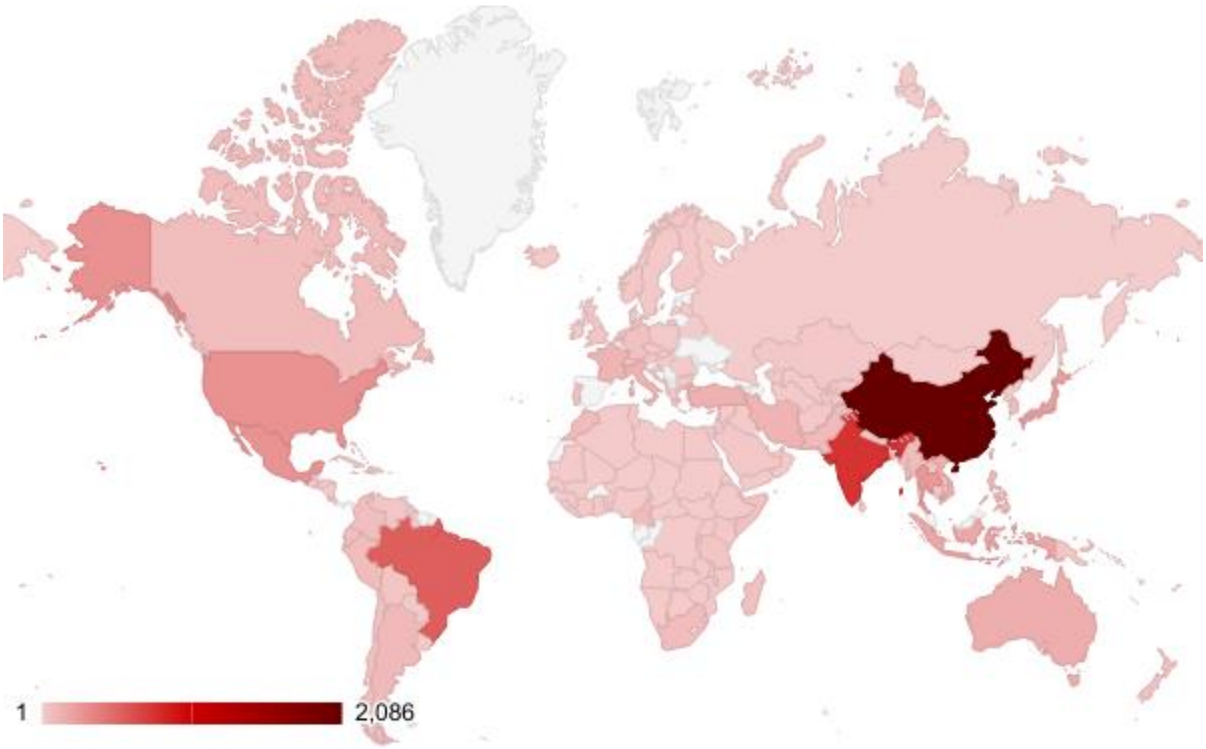


Figure 5. Geographic distribution of new species discovered in 2022. Scale represents the number of papers about a new species released, ranging from 1 to 2086. The locations are represented by the territories for each country currently recognized by the United Nations.

From the databases of GBIF, species occurrences were taking for each country to be used as a proxy for activity in each country. These occurrences can range from 1,961 to 810,355,779 (figure 6). According to GBIF, the occurrences are highest in America with 819,355,779. Following on from that there is France (135,661,904), Canada (133,960,951), Sweden (119,665,437), and Australia (112,339,324). In terms of countries found in the continents of Asia, Africa, and South America, there is relatively low occurrences seen.

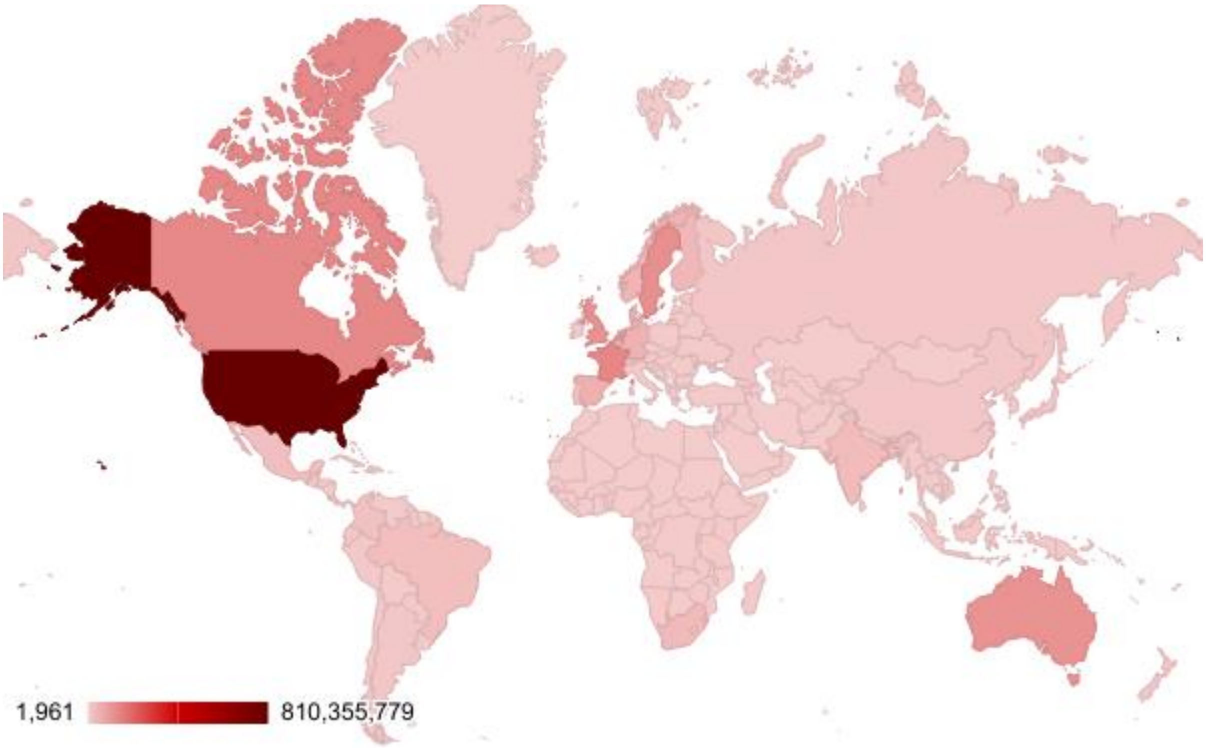


Figure 6. Geographic distribution of species occurrences per country according to GBIF (GBIF: The Global Biodiversity Information Facility, 2023). Scale represents the number of species occurrences, ranging from 1,961 to 810,355,779. The locations are represented by the territories for each country currently recognized by the United Nations.

Researcher Distribution

From the researchers which obtained an ORCID ID, there was a subset of 1658 researchers in which the country they came from was known. The distribution of researchers is varied. Within Africa, the distribution of researchers is quite sparse with the majority of countries who do display a researcher housing only 1(figure 7). An exception to this is South Africa which has 23 researchers. The countries which show the highest number of researchers include Brazil (255), China (247), India (153), Mexico (64), and America (63).

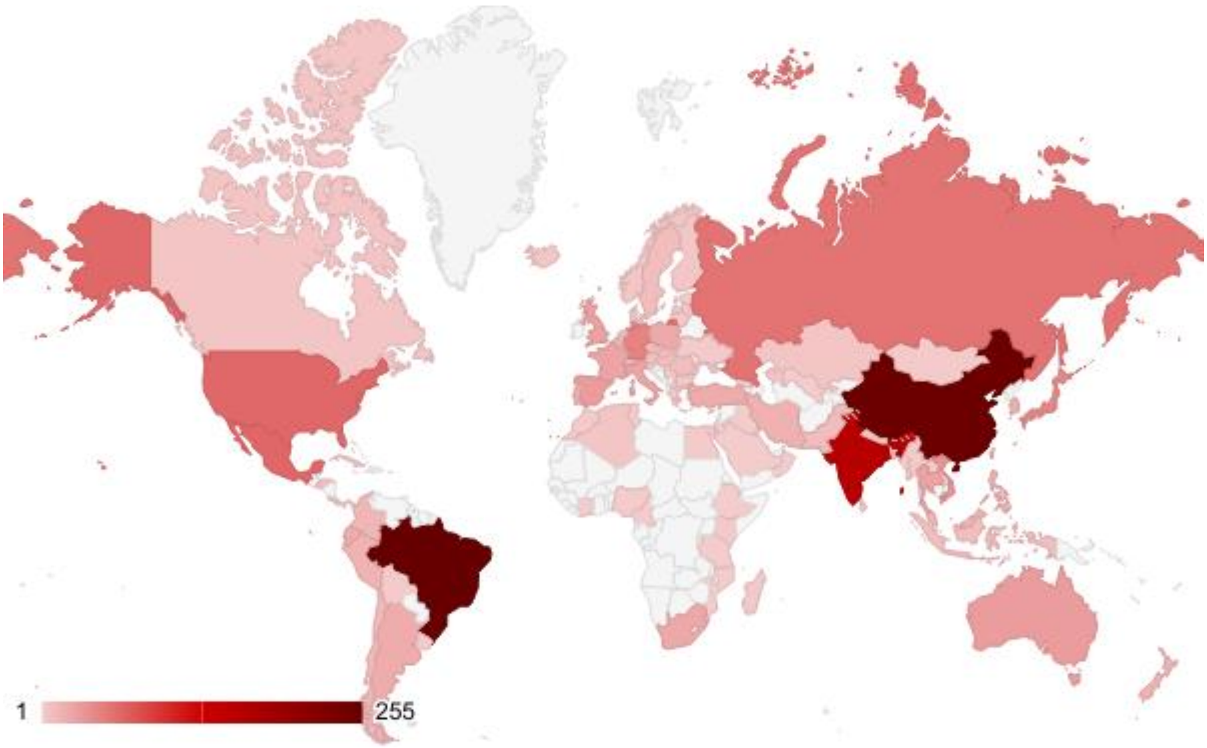
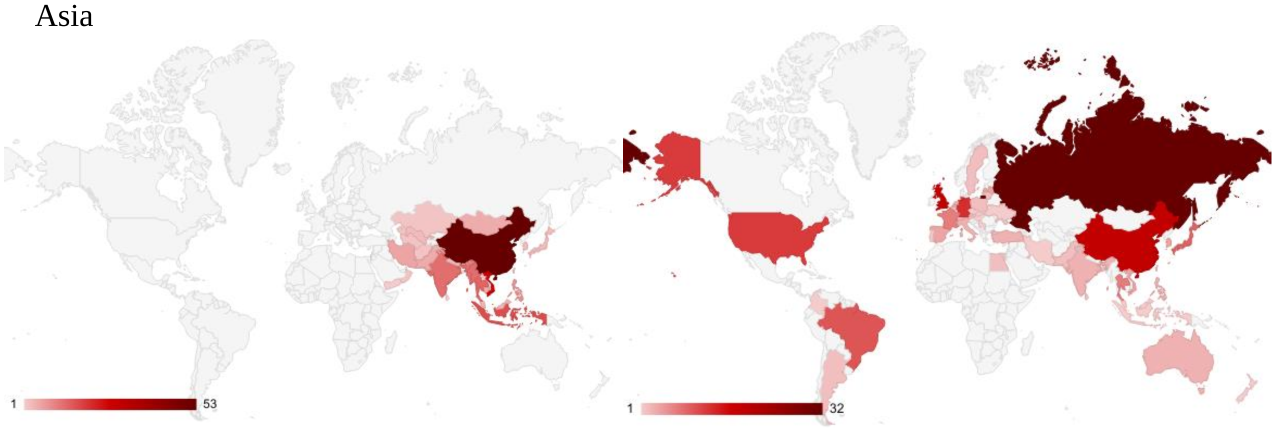


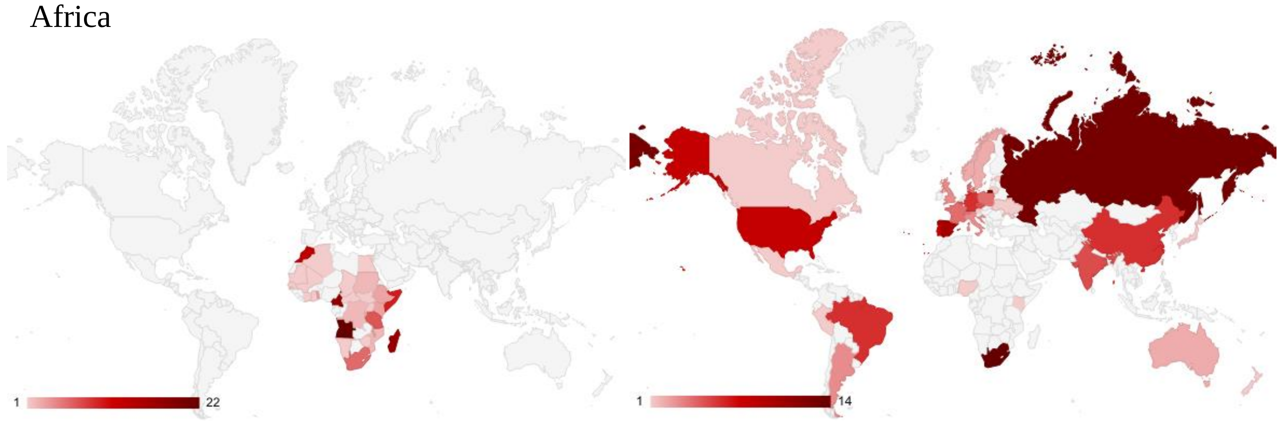
Figure 7. Geographic distribution of researchers involved in new species discoveries in 2022. Scale represents the number of researchers, ranging from 1 to 255. The locations are represented by the territories for each country currently recognized by the United Nations.

Linking distribution of researchers with the spatial distribution of the species, the results revealed that 66% of the researchers were residents of the same country as species origin. The remaining 34% of researchers were compared against the species origin which showed a dominance in researchers coming from countries in Europe no matter where the species originated (figure 8). Species from Asia showed a high variation in researcher distribution across all Asia and Europe. Overall there was little representation of African researchers, in and out of Africa. When it came to species discovered in South America, there was a higher number of researchers from South America. For species in Oceania, there was a fairly even spread of researchers around the world, excluding Africa.

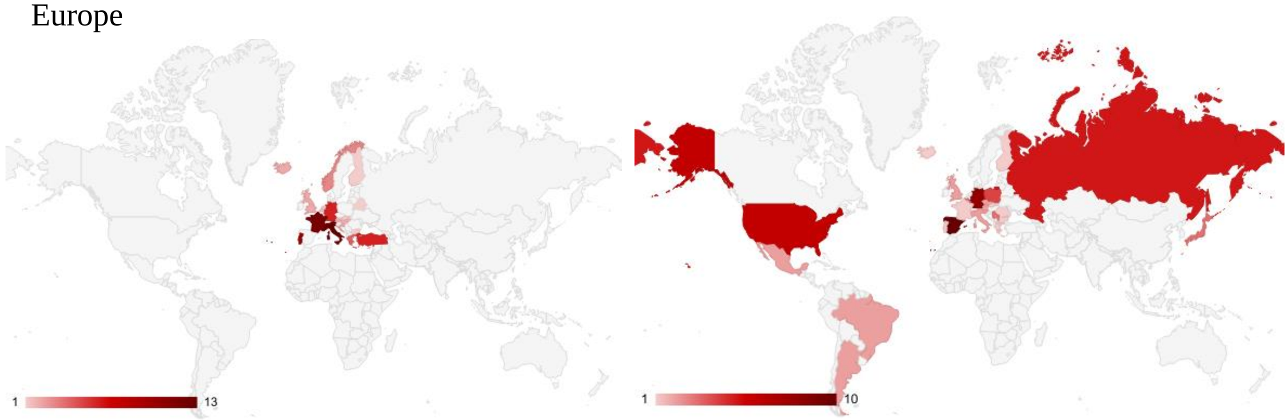
Asia



Africa



Europe



North America

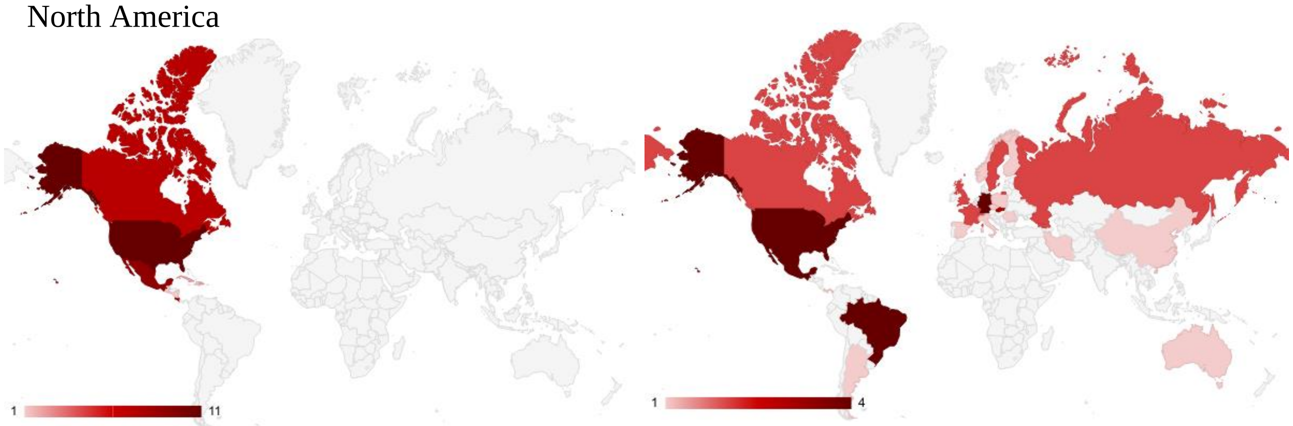
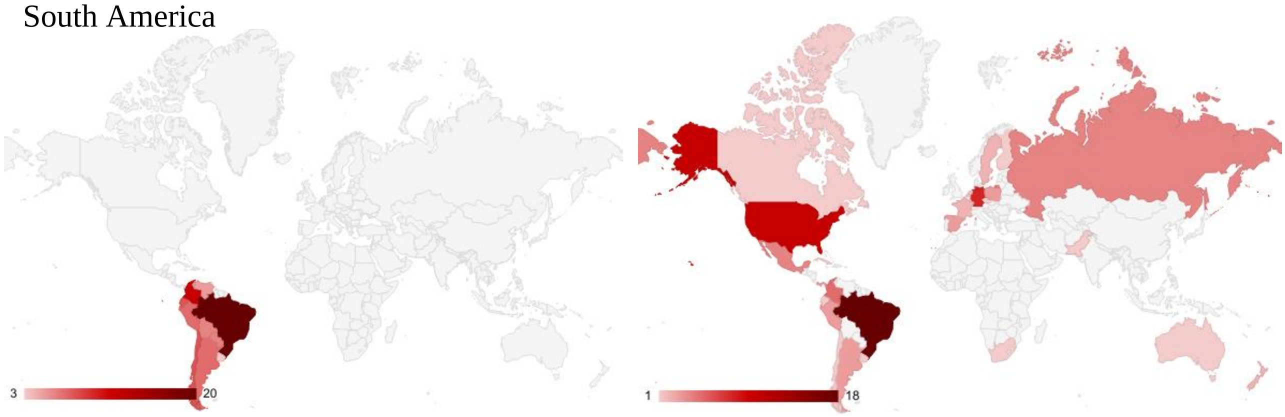


Figure 8 cont.

South America



Oceania

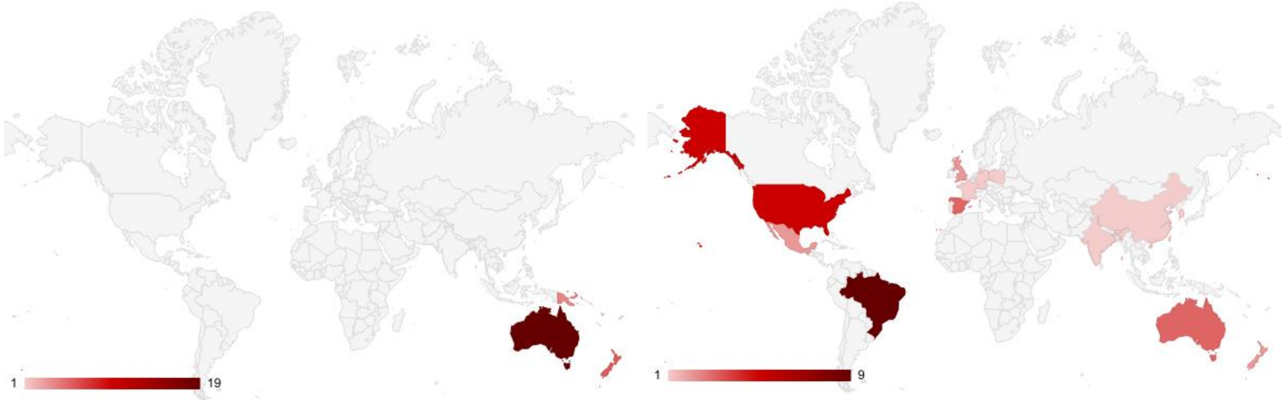


Figure 8. Comparison of the geographic distribution of species and researchers. The distribution of species is highlighted on the left maps and is split up continentally: Asia, Africa, Europe, North America, South America, and Oceania. The distribution of researchers globally for each continental species subset is highlighted on the right maps.

Funding

From the data gathered in this study, information surrounding funding was lacking in most papers resulting in a remarkably smaller subset of 882 papers. Out of the 882 papers, 69% were funded by a funding agency that was based in the same country as species origin. Geographically, the distribution of funding agencies span the globe however excludes countries in Africa, apart from South Africa (figure 9). China has considerably more funding agencies, at 337, than any other country. With around a third of the number of funding agencies of China, Brazil has 105, followed on by America (89), South Korea (40), and Japan (33).

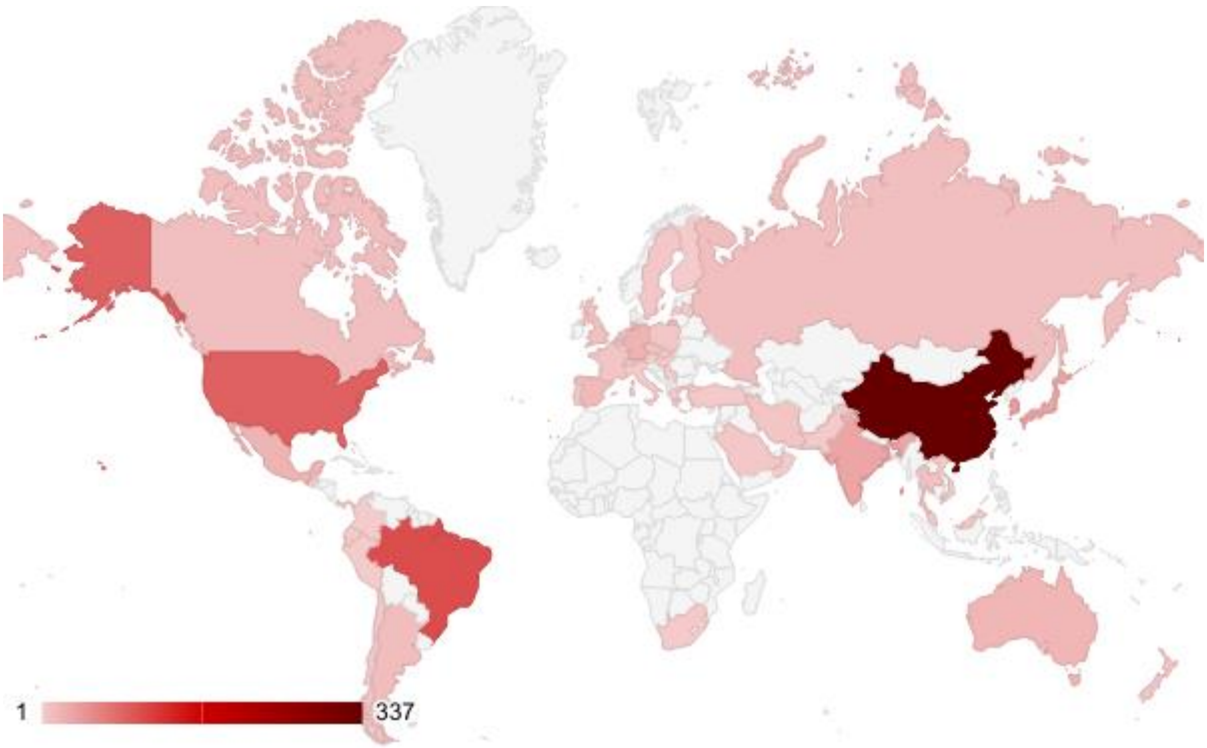


Figure 9. Geographic distribution of funding agencies involved in funding the papers on new species discoveries in 2022. Scale represents the number of funding agencies, ranging from 1 to 377. The locations are represented by the territories for each country currently recognized by the United Nations.

Accessibility

For 2022, it was found that overall 49% of the new species discoveries were open access. This shows that whether the paper will be open access or not is 50:50 chance. The accessibility of papers is varied depending on where the species is discovered (figure 10), although it is worth noting that the numbers of papers does vary between countries and so can play a part in these percentages. Some of the countries which display 0% open access in papers include Russia, the Bahamas, and Nicaragua, however they all display only 1 or 2 papers. This scenario is the same for the opposite end of 100% open access where the number of papers range from 1 to 9. Countries which released more papers have a higher reliability in their open access percentage, such as Brazil (264) was 30%, China (839) was 47%, and India (344) was 56%.

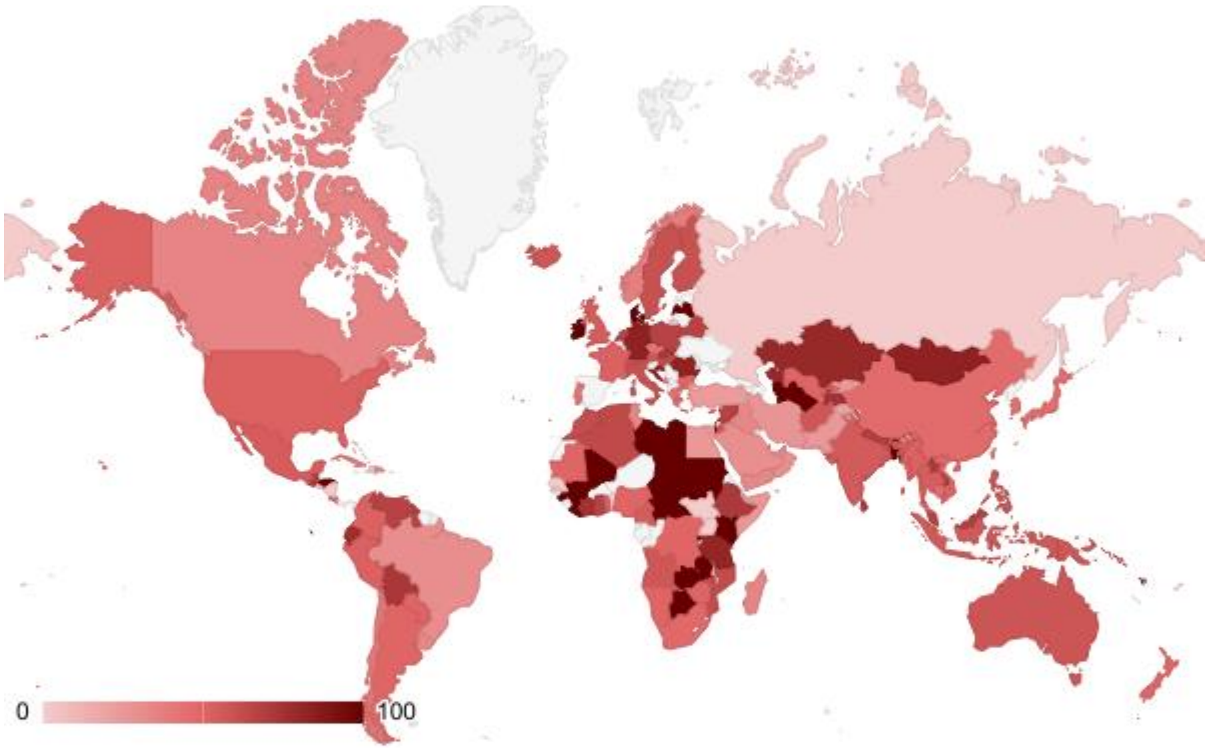


Figure 10. Percentage of papers about new species discovery in 2022 which were open access represented geographically. Scale represents the percentage of papers which are open access. The locations are represented by the territories for each country currently recognized by the United Nations. Countries which are closer to 100 have a majority of open access papers, whereas as countries closer to 0 have a majority of non-open access papers.

When looking into the percentage of researchers making their work open access, this can have a better representation of how accessible scientific literature is for each country (figure 11). The majority of countries in Europe and North America have a higher percentage of open access papers. In Europe, most of the countries are in the percentage range of 50% to 80%. Again the imbalance of researchers in countries has an effect on the percentage as the countries with the highest number of researchers tend to be around the 50% mark of open access. For example, Brazil (407) was 40% accessible and China (332) was 49% accessible.

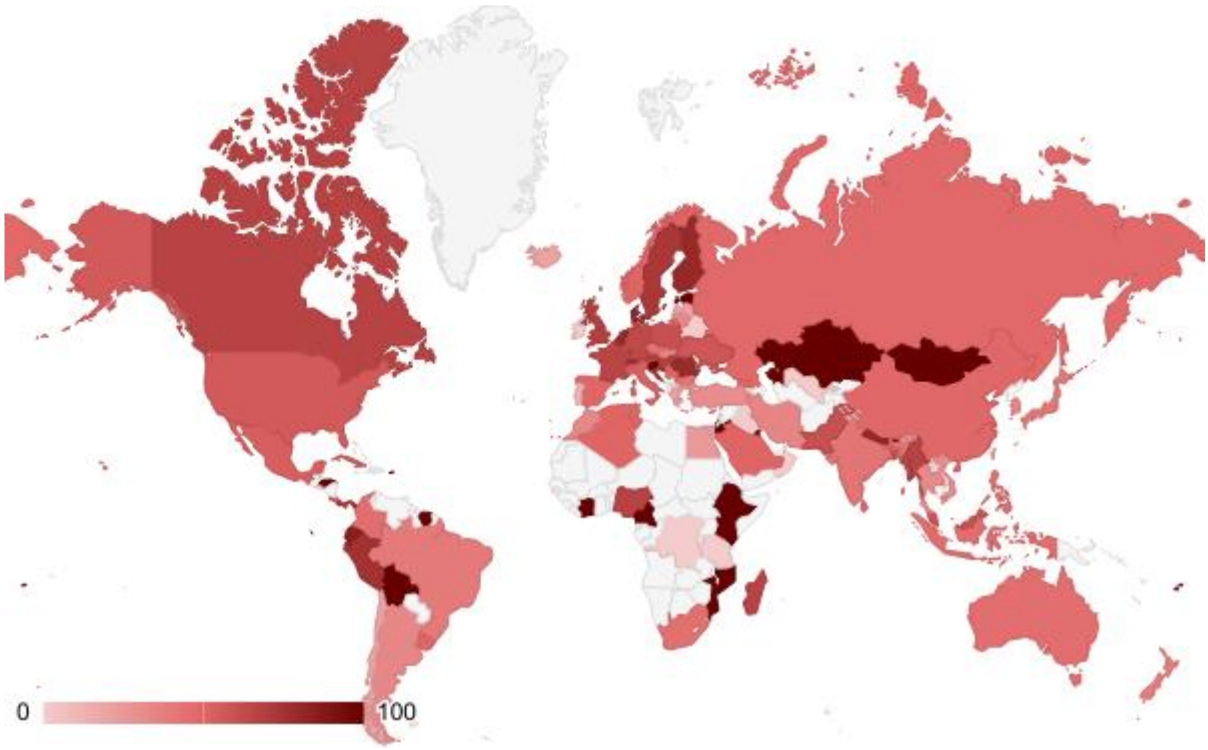


Figure 11. Percentage of researchers who have released papers about new species discovery in 2022 which were open access represented geographically. Scale represents the percentage of papers which are open access. The locations are represented by the territories for each country currently recognized by the United Nations. Countries which are closer to 100 have a majority of open access papers, whereas as countries closer to 0 have a majority of non-open access papers.

Discussion

The current biodiversity crisis is highly reliant on the increasing knowledge from taxonomic research. Based on the annual rate of species description of approximately 20,000, centuries of work will be required to discover the millions of species not yet known (Grieneisen, et al., 2014). As advances in science have improved the efficiency of taxonomy through the use of discrimination techniques, such as DNA sequencing, there is still the ever pressing reality that the number of taxonomists are decreasing (Fontain, et al., 2012). Historically, Europe has been the geographic concentration of taxonomic expertise predominantly due to the colonial origin of the science, much like many other scientific fields (Fontaine, et al., 2012). The colonial trajectory which taxonomy was built upon involved European scholars sourcing species from overseas in which they "discovered" and then named the new species (Eichhorn, et al., 2019). One issue here was that many of these species which were "discovered" often were already well known and named by locals and indigenous groups. This showed a complete disregard for the expertise and long-lasting knowledge from local and indigenous people. Recently, there has been a push towards decolonising taxonomy and therefore increasing importance of resident taxonomists. In 2022, 66% of the taxonomists in new species discovery resided in the country of species origin. Furthermore the countries which obtained the highest number of taxonomists were Brazil, China and India (figure 7). Costello et al (2013) explained that there has been an increasing number of taxonomists based in South America and Asia. This has shown that the taxonomic world that is seen in 2022 is a different picture to how taxonomy historically used to be, however there is still more push towards resident taxonomists to be done. The distribution for researchers in Africa was particularly low (figure 7), although it was shown that the researchers which were there either stayed in their own country or researched in another country within Africa (figure 8). When analysing the remaining 34% of researchers, Europeans would research all over the world (figure 8). Researchers from the America's also played a big role in species discovery outwith their country. The high number of taxonomists within countries which are known to be species rich is promising for the future of taxonomy as geographic bias could be less prevalent.

Geographic bias has been long standing in taxonomy due to the dominance of European and North American taxonomists. The diversity of species is known to be at its highest within the tropics and considerably less within more temperate areas. This phenomenon can be termed as the latitudinal diversity gradient (LDG) (Freeman and Pennell, 2021). Contrary to LDG, the geographic bias creates a taxonomic debt in the tropics producing a latitudinal taxonomic gradient (Freeman and Pennell, 2021). In 2022, China dominated the species discovery producing 25% of the year's discoveries, followed by India and Brazil (figure 5). This indicated that much of the discoveries of 2022 occurred within the tropic and sub-tropic areas of the world. In comparison, the occurrences of species from GBIF (figure 6) presented a dominance of many western societies, including America, France, Canada, Sweden, and Australia. However, the high occurrences of species within these countries could be more due to the rate at which these countries add to online databases, instead of their species diversity. Although European countries have been heavily investigated for new species in the past, there is still new species being found within Europe and at a constant rate (Fontaine, et al., 2012). The species discovery in Europe is yet to plateau and so even though there should be push towards a larger taxonomic workforce within biodiversity rich countries found in the tropics, efforts in Europe should not be fully abandoned.

Taxonomy does not only experience bias geographically but also taxonomically, with some taxa researched considerably more than others (Troudet, et al., 2017). A class which is particularly well researched is the Aves. In 2022, Arthropoda represented 35% of the new species, along with Tracheophyta representing 20% and Chordata representing 12% (figure 2). When compared to the expected proportion from all known species in the catalogue of life, a different story was told. Arthropoda experienced less discoveries than would have been expected considering 55% of known species are classified as Arthropods (figure 3, figure 4). In contrast, the Tracheophytes and Chordates displayed a higher discovery than expected (figure 4). This is conclusive with the taxonomic bias which has continued to dominate in taxonomy. The Chordates includes the well known class Aves, which has plateaued in discoveries due to the extensiveness they have already been researched (Fontaine, et al., 2012). The rate at which species within certain taxa are discovered varies, some on the steady rise, others going through an explosion of discovery, and some becoming rarer to find. However, taxa which are having decreasing rates of discovery is not necessarily plateauing in their discovery potential but a representation of the decrease in taxonomic workforce available (Fontaine, et al., 2012).

The reduction in the taxonomic workforce can also be attributed to the insufficient funding available to the industry. Unsurprisingly, China had the most funding agencies for taxonomy in which the efforts can be seen by the high number of species discovery in 2022 (figure 5, figure 9). Brazil and America, which also showed high species discovery also showed a high number of funding agencies.

Excluding South Africa, there were no funding agencies found in Africa. This demonstrates the need for funding in taxonomy as wherever the money is the research will follow. There are many grant programs in place, such as PEET, that emphasises the importance of international collaboration which engages with host country scientists as well as training locals while collecting the specimens (Grieneisen, et al., 2014). Another way in which taxonomic funding is impacted is the influence of end users, which includes professions such as conservationists and field ecologists, determining the importance and significance of research through their demand attracting funds (Ebach, et al., 2011). These end users rely heavily on taxonomy for identifications and accurate descriptions of species to carry out their own work.

Ebach et al (2011) believes that the funding impediment involving end users can be solved through increasing the access of taxonomic information. There are many projects already running which increase the taxonomic information available online such as GBIF and Catalogue of Life. Furthermore, the push for open access to scientific literature has created many different electronic libraries, such as the Biodiversity Heritage Library, which ensures data storage that the end users are able to use (Ebach, et al., 2011). For new species discovery in 2022, nearly half (49%) of the papers were open access indicating that there has potentially been a plateau in the increase of open access articles in taxonomy. In 2014, more than 50% of scientific articles were freely accessible (Archambault, et al., 2014). An additional way to bolster taxonomic research is the involvement of citizen science. One example of this is a website called iNaturalist (*iNaturalist*). iNaturalist is a popular platform which the public record their observations of species and from there can get help with the identification process. As a collection, iNaturalist can then provide geographical distribution of taxa as well as interactions and activity patterns (Fritz and Ihlow, 2022). From this, iNaturalist can be a great tool in investigations. Fritz and Ihlow (2022) used iNaturalist in their study which investigated the geographic variation of coloration and pattern in the common grass snake. Here they used the photographic records of the common grass snake and linked it with the geographic location described for each photo to come to a conclusion. From Fritz and Ihlow's study, it was shown that citizen science was a great source of information.

Overall this study has highlighted many areas in which there is a lack of information available, particularly about the researchers themselves. Comparisons between data is restricted due to the differing size of datasets available. Illustrating this point is the difference between species (8208) and researcher (1658) distribution. The fear of taxonomy dying out has not been met with the funding to be able to rebuild and train the taxonomic workforce. For the future of taxonomy, I propose that all this data could be stored in a website, similarly to GBIF. This would include how many species there are, what the species are to indicate the taxonomic distribution, when the discoveries are made, who is finding these species, who is funding the projects, and the accessibility of the papers. Here, gaps in taxonomy can be addressed and acted upon therefore avoiding taxonomic and geographic bias. Comparisons between years could be made, documenting changes and patterns, allowing questions to be asked such as "how much of taxonomy has been decolonised?". Additionally, advancements in technology, particularly AI could be incorporated within the website to promote machine learning to be able to determine whether an article is truly about a new species or not. This could potentially avoid confusion between the new species and new fossils found. Even though there is still a requirement to increase the number of taxonomic experts, this would be a useful tool to know what the world of taxonomy is currently looking like to aid planning for taxonomy's future.

Limitations

Through the use of BioRSS and online data in general, there is limitation due to the reliance for the data to be online. BioRSS relies on scientific articles which are online, particularly in journals which have an RSS feed or can be searched through Google Scholar. Approximately 16,000 to 20,000 species are discovered per year (Fontaine, et al., 2012; Grieneisen, et al., 2014), however, BioRSS only detected 8,567. This difference could be due to BioRSS finding articles on new species, it does not adapt to the number of new species mentioned in the paper. This makes it difficult to measure the effectiveness of BioRSS. Furthermore, many new species found are potentially fossils and therefore not a new species currently living. As BioRSS is also built to find the geolocation of each species, the country may not of been identified if the country name was not mentioned. For example, the article stated "Indian" instead of "India". Additionally, the system is exclusionary for the location of marine species due to the inability to detect areas in the ocean. Finally, to connect papers to their authors the system requires the authors to not only have an ORCID ID but for the paper to publish the ORCID ID's as well. To connect to the funders, it is required that papers list the funders using proper identifiers. Both of these aspects require the dots to be joined to further understand the field of taxonomy.

Acknowledgements

I wish to thank my supervisor Roderic Page who guided me through this project as well as supplying the data for me from his website BioRSS. Also to the researchers who made gathering the data easier through making the information that I required accessible.

References

Agnarsson, I. and Kuntner, M., 2007. Taxonomy in a changing world: seeking solutions for a science in crisis. *Systematic biology*, *56*(3), pp.531-539.

Alex, B., Llewellyn, C., Grover, C., Oberlander, J. and Tobin, R., 2016, May. Homing in on Twitter users: Evaluating an enhanced geoparser for user profile locations. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 3936-3944).

Archambault, É., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, L. and Roberge, G., 2014. Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013.

Bánki, O., Roskov, Y., Döring, M., Ower, G., Vandepitte, L., Hobern, D., Remsen, D., Schalk, P., DeWalt, R. E., Keping, M., Miller, J., Orrell, T., Aalbu, R., Abbott, J., Adlard, R., Adriaenssens, E. M., Aedo, C., Aescht, E., Akkari, N., et al. 2023. Catalogue of Life Checklist (Version 2023-02-07). *Catalogue of Life*. Available at: https://doi.org/10.48580/dfrq

BioRSS. 2022. *Herokuapp.com*. Available at: https://biorss.herokuapp.com/?feed=Y291bnRyeT1XT1JMRCZwYXRoPSU1QiUyMkJJT1RBJTIyJTVE

Britz, R., Hundsdörfer, A. and Fritz, U., 2020. Funding, training, permits—the three big challenges of taxonomy. *Megataxa*, *1*(1), pp.49-52.

Costello, M.J., May, R.M. and Stork, N.E., 2013. Can we name Earth's species before they go extinct?. *science*, *339*(6118), pp.413-416.

dos Santos, J.W., Correia, R.A., Malhado, A.C., Campos‐Silva, J.V., Teles, D., Jepson, P. and Ladle, R.J., 2020. Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. *Animal Conservation*, *23*(6), pp.679-688.

Ebach, M.C., Valdecasas, A.G. and Wheeler, Q.D., 2011. Impediments to taxonomy and users of taxonomy: accessibility and impact evaluation. *Cladistics*, *27*(5), pp.550-557.

Eichhorn, M.P., Baker, K. and Griffiths, M., 2019. Steps towards decolonising biogeography. *Frontiers of Biogeography*, *12*(1), pp.1-7.

Elbakyan, A., 2016. Sci-Hub is a goal, changing the system is a method. *Engineuring [Internet]*.

Freeman, B.G. and Pennell, M.W., 2021. The latitudinal taxonomy gradient. *Trends in Ecology & Evolution*, *36*(9), pp.778-786.

Fritz, U. and Ihlow, F., 2022. Citizen Science, taxonomy and grass snakes: iNaturalist helps to clarify variation of coloration and pattern in Natrix natrix subspecies. *Vertebrate Zoology*, *72*, pp.533-549.2

Fontaine, B., van Achterberg, K., Alonso-Zarazaga, M.A., Araujo, R., Asche, M., Aspöck, H., Aspöck, U., Audisio, P., Aukema, B., Bailly, N. and Balsamo, M., 2012. New species in the Old World: Europe as a frontier in biodiversity exploration, a test bed for 21st century taxonomy. *PLoS One*, *7*(5), p.e36881.

GBIF: The Global Biodiversity Information Facility. 2023. *What is GBIF?*. Available at: https://www.gbif.org/what-is-gbif

Grieneisen, M.L., Zhan, Y.U., Potter, D. and Zhang, M., 2014. Biodiversity, taxonomic infrastructure, international collaboration, and new species discovery. *BioScience*, *64*(4), pp.322-332.

Guerra García, J.M., Espinosa Torre, F. and García Gómez, J.C., 2008. Trends in taxonomy today: an overview about the main topics in taxonomy. *Zoológica baetica*, *19, 15-49*.

Hendricks, G. *You are Crossref*, *Crossref*. Available at: https://www.crossref.org/ (Accessed: 14 March 2023).

Himmelstein, D.S., Romero, A.R., Levernier, J.G., Munro, T.A., McLaughlin, S.R., Greshake Tzovaras, B. and Greene, C.S., 2018. Sci-Hub provides access to nearly all scholarly literature. *ELife*, *7*, p.e32822.

Holvoet, K., 2006. What is RSS and how can libraries use it to improve patron service?. *Library Hi Tech News*, *23*(8), pp.32-33.

iNaturalist (no date) *iNaturalist*. Available at: https://www.inaturalist.org/ (Accessed: 14 March 2023).

Leary, P.R., Remsen, D.P., Norton, C.N., Patterson, D.J. and Sarkar, I.N., 2007. uBioRSS: tracking taxonomic literature using RSS. *Bioinformatics*, *23*(11), pp.1434-1436.

Lewis, D.W., 2012. The inevitability of open access. *College & research libraries*, *73*(5), pp.493-506.

Liu, J., Jin, X., Yao, S., Wang, Y., Lu, Y., Chen, Q., Wu, C., Slik, F. and Lindenmayer, D., 2023. Who will name new plant species? Temporal change in the origins of taxonomists in China. *Proceedings of the Royal Society B*, *290*(1992), p.20221954.

McCann, S.R., 2020. Taxonomy. *Bone Marrow Transplantation*, *55*(12), pp.2221-2223.

ORCID. Available at: https://orcid.org/ (Accessed: 14 March 2023).

Page, R. 2021. 'Glasgow Geoparser'. Available at: https://github.com/rdmpage/glasgow-geoparser

Tahseen, Q., 2014. Taxonomy-The crucial yet misunderstood and disregarded tool for studying biodiversity. *Journal of Biodiversity & Endangered Species*, *2*, p.128.

taxonfinder.org | Find scientific names in text. Available at: https://taxonfinder.org/ (Accessed: 14 March 2023).

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. and Legendre, F., 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific reports*, *7*(1), pp.1-14.

Unpaywall: An open database of 20 million free scholarly articles. Available at: http://unpaywall.org/ (Accessed: 14 March 2023).

Wikidata. Available at: https://www.wikidata.org/wiki/Wikidata:Main_Page (Accessed: 14 March 2023).