A standard protocol for harvesting biodiversity data from Facebook

Authors and Affiliations

Shawan Chowdhury*[1,2,3,4], Sultan Ahmed[5], Shofiul Alam[5], Corey T Callaghan[6], Priyanka Das[5], Moreno Di Marco[7], Enrico Di Minin[8,9,10], Ivan Jarić[11,12], Mahzabin Muzahid Labi[5], Md. Rokonuzzaman[5], Uri Roll[13], Valerio Sbragaglia[14], Asma Siddika[5], Aletta Bonn[1,2,3]

* = corresponding author; contact: dr.shawanchowdhury@gmail.com

[1]Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Straße 159, 07743 Jena, Germany

[2]Helmholtz Centre for Environmental Research - UFZ, Department of Ecosystem Services, Permoserstr. 15, 04318 Leipzig, Germany

[3]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr. 4, 04103 Leipzig, Germany

[4]School of Biological Sciences, The University of Queensland, 4072, Australia

[5]Department of Zoology, University of Dhaka, Dhaka 1000, Bangladesh

[6]Department of Wildlife Ecology and Conservation, Fort Lauderdale Research and Education Center, University of Florida, Davie, FL 33314-7719

[7]Department of Biology and Biotechnologies Charles Darwin, Sapienza University of Rome, viale dell'Università 32, I-00185 Rome, Italy

[8]Department of Geosciences and Geography, University of Helsinki, FI-00014 Helsinki, Finland

[9]Helsinki Institute of Sustainability Science, University of Helsinki, FI-00014 Helsinki, Finland

[10]School of Life Sciences, University of KwaZulu-Natal, Durban 4041, South Africa

[11]Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique Evolution, Gif-sur-Yvette, France

[12]Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic

[13]Mitrani Department of Desert Ecology, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion 8499000, Israel

[14]Department of Marine Renewable Resources, Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain

Keywords

Bangladesh, citizen science, crowdsourcing, Facebook, iEcology, social media, tropics, megadiverse countries, Wallacean shortfall

Author contributions

SC conceptualised the idea; SC developed the method; SC, SuA, ShA, PD, MML, MR, and AS collected the data; SC did the analysis; everyone contributed to the analysis; SC wrote the paper; everyone contributed to the writing.

Acknowledgements

SC and AB gratefully acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) and the sMon project funded by the German Research Foundation (DFG-FZT 118, 202548816).

Data availability

Our extracted Facebook data are publicly available (Chowdhury et al., 2022b).

Abstract

1. The expanding use of citizen science platforms has led to an exponential increase in biodiversity data in global repositories. Yet, our understanding of species distribution remains patchy for most of the world. Social media data has the potential to reduce the global biodiversity knowledge gap. However, practical guidelines and standardised pipelines to harvest such data sources are still missing.

2. Here, we provide a standardised framework to extract species distribution records from Facebook groups that allow access to their data following data privacy and protection safeguards. Some countries actively use and moderate Facebook groups to share species records. We present how to structure keywords, search for species photographs, and georeference localities for such records. We further highlight some challenges users might face when extracting species distribution data from Facebook and suggest potential solutions.

3. Following our proposed framework, we present a case study on Bangladesh's biodiversity – a tropical megadiverse South Asian country. We scraped nearly 45,000 unique locality data for 967 species, with a median of 27 records per species. About 12% of the distribution data were for threatened species, which represent 27% of all species. We also obtained data for 56 Data Deficient species.

4. If carefully harvested, social media data can significantly reduce global biodiversity knowledge gaps. Consequently, developing an automated tool to extract and interpret social media biodiversity data is an essential research priority.

Introduction

Amid the sixth mass extinction, many species worldwide are dramatically declining – 28% of all assessed species in the IUCN Red List are threatened with extinction (Dirzo et al., 2014; Pimm et al., 2014; Murali et al., 2023). A recent study that analysed population trends for all IUCN Red List assessments revealed that 48% of species are declining, 49% are in stable condition, and only 3% of assessed species are increasing (Finn et al., 2023). The Living Planet Index report reveals an average 69% decrease in monitored wildlife populations since 1970 (WWF, 2022). However, global assessments are highly biased towards certain taxa or regions (Miqueleiz et al., 2020). For example, 85% of described reptile species have been assessed by the IUCN Red List (Meiri et al., 2023) *vs* ca. 1% of arthropod species (IUCN, 2023). This severe discrepancy reflects a long-known bias in research interests (Clark & May, 2002; Cardoso et al., 2011; Di Marco et al., 2017) and can be partially attributed to missing and inadequate distribution data that is fundamental for species threat assessments (Beck et al., 2014; Hughes et al., 2021). For example, a taxonomic revision and distribution sample of beetle species found that 53% of the 186 species were only known from a single locality, and 13% from a single specimen (Stork, 1997). These sorts of biases are known as Linnean shortfalls (taxonomic knowledge gaps) and Wallacean shortfalls (distribution knowledge gaps; Hortal et al., 2015; Diniz-Filho et al., 2023). Such poor representation of species records is also prominent in the most extensive biodiversity repository - Global Biodiversity Information Facility (GBIF) – which contains locality data only for 10% of described insect species (Chowdhury et al., 2023a).

To bridge this knowledge gap, many initiatives are using knowledge from the general public (Hochkirch et al., 2015). These data collection initiatives are commonly known as citizen science or community science, in which people share their species observation records through different, often online, applications (Conrad & Hilchey, 2011; Bela et al., 2016; Chandler et al., 2017; Pocock et al., 2019; Callaghan et al., 2021). These observations, in many cases, are eventually deposited in large online repositories (e.g. GBIF; Roy et al., 2018; Callaghan et al., 2022). There are hundreds of such citizen science applications globally that have greatly improved our understanding of biodiversity patterns in recent years. Since 2007, there has been a 12-fold increase in biodiversity data in GBIF (Heberling et al., 2021; GBIF contains > 2.3 billion species occurrence records as of 9 June 2023). However, despite increases in the volume of biodiversity data available, geographic and taxonomic knowledge gaps on species distributions remain (Hughes et al., 2021). Most biodiversity observation records are from Europe and North America, resulting in major sampling and observation biases (Ramírez et al., 2022). While most species are found in tropical forests, our knowledge of the biodiversity of these regions is extremely limited (Collen et al., 2008; Hortal et al., 2015; Meyer et al., 2015; Hughes et al., 2021; Chowdhury et al., 2023b). Though it comprises less than 2% of the Earth's area, nearly 36% of the biodiversity data in GBIF come from the United States. Conversely, Brazil, which has a similar land area to the USA and is the most biodiverse country on Earth, is represented by only 0.8% of the records in GBIF (accessed on 22 May 2023). Therefore, new approaches and methods are needed to overcome the Linnean and Wallacean shortfalls.

With the increasing popularity of social media and the growing availability of digital phones and fast internet, many people post biodiversity observations on different social media platforms (e.g., Facebook, Twitter; Di Minin et al., 2015; Andrachuk et al., 2019; Toivonen et al., 2019) that don't necessarily make it to the existing pipelines to GBIF. Among these platforms, Facebook has become the most popular social media network (Anderson et al., 2012). There are thousands of biodiversity observation groups on Facebook globally, with a wealth of species distribution information, often with more in-depth data than available in global biodiversity repositories (Chowdhury et al., 2023b). For example, by scraping a single Facebook group for the butterflies of Bangladesh, Chowdhury et al. (2021) obtained about 35 times more distribution records from Facebook than were deposited in GBIF at the time. Moreover, Facebook also contains data regarding many unique species that are absent from GBIF altogether (Chowdhury et al., 2023b). Such data have also been shown to be key for improved spatial conservation prioritisation (Chowdhury et al., 2023c). The utility of Facebook as a biodiversity repository is possible due to volunteer contributions of moderators and administrators of Facebook groups that help users identify species (Chowdhury et al., 2023b; Marceno et al., 2021). Moreover, many scientists are unaware of the great biodiversity potential found in social media data or are unfamiliar with pipelines to extract such data (Chowdhury et al., 2023c).

Several studies explored the importance of using Facebook data to fill the global biodiversity data shortfall (e.g., Chamberlain, 2018; Marcenò et al., 2021; Chowdhury et al., 2023b; O'Neill et al., 2023). While having a standard data collection protocol is essential to improve conservation assessments, there is no such protocol that researchers can follow to collate species distribution data from Facebook. Here, we provide a complete pipeline to harvest quality biodiversity data from Facebook groups, which were accessed by receiving permission from the groups' administrators, for Bangladesh as a case study. We discuss i) the keyword formulation process, ii) how to search inside Facebook groups and extract and filter species photographs, and iii) how to georeference the location information. Finally, we point out the potential of species distribution data obtained from Facebook and some challenges that users might face when extracting species distribution data from Facebook.

Methods

Data extraction framework

Extracting species locality information from Facebook contains three steps: group selection, data extraction, and georeferencing the record location (Figure 1a).

Step 1 - Group selection

Before starting the data extraction process, it is important to filter the relevant groups using a systematic search. Search keywords could contain a combination of taxon and country names. For example, if there is an interest in extracting data for 'birds' from 'Bangladesh', the keywords can be: 'Bird Bangladesh'. Here, singular/plural form or capitalisation of words do not make any substantial difference - except in their order of appearance. We recommend using different combinations of keywords (including region names) when listing relevant Facebook groups. After formulating keywords, the search can be conducted using the Facebook search function to identify relevant groups (Figure 1).

Step 2 - Data extraction

When extracting data, the authors should carefully consider the privacy of the group: whether it is public, private or secret and maintain the data usage policy (Di Minin et al., 2021). Authors should also prepare a complete list of species that need manual search.

When a relevant group is selected, the next step is to extract species information. This can be performed by searching for species locality records inside individual groups. Searching within Facebook groups is similar to searching in engines like Google Scholar: species names can be searched with the search function of the selected Facebook group ('magnifying glass' icon inside the Facebook group), which includes options to restrict the search (e.g., year; Figure 1).

During data extraction (from posts, but not photographs), some issues may be encountered. For example: i) species misidentification in the caption, ii) inconsistencies with species names (some users use the scientific name, while others use the English or local name), iii) Facebook becomes too slow when the keyword searches produce too many search results, and iv) search results also include erroneous species due to partial keyword overlap (e.g., butterfly named 'common pierrot' could appear when searching for 'common jay'). To handle misidentification issues, we recommended double-checking species identifications before extracting data. In active Facebook groups, moderators, administrators, or other users often verify individual photographs and provide suggestions when required. This is particularly relevant when they help users identify photographs and confirm locality information. Such information can often be found in the comment section of each photograph. To control naming inconsistencies, we recommend searching by scientific name, English name, and local names for each species in each group. Another issue is that the same photographs might appear in different searches due to the partial keyword overlap. To control this issue, researchers could use a unique identifier for the users and then remove duplicate information. To expedite the search process during an individual search, we recommend restricting the search by year using the 'Date Posted' tab on the left-hand side of the screen. Here, individual years (e.g., 2021) can be filtered for each search, and then the process should be repeated until there are no further search results. Finally, duplicate records can be removed.

Step 3 - Georeferencing

Facebook lacks an automated georeferencing system. Using the location information from species photographs (Figure 1), any mapping software (e.g., Google Map, Google Earth, ArcGIS) can be used to extract the latitude and longitude information. Conveniently, Google Place ID API (https://developers.google.com/maps/documentation/places/web-service/place-id) enables automatic georeferencing for many locations. It can also be done using the ‘geocode’ function of the ‘ggmap’ R package (Kahle and Wickham, 2013).

Case study on Bangladesh

Species list

To illustrate the application of our proposed framework, we performed a case study for extracting biodiversity data from Bangladesh's dedicated Facebook groups. We compiled a checklist of the animals of Bangladesh with a known conservation status (1,619 species) from the most recent national Red List (IUCN Bangladesh, 2015). While this is the most comprehensive checklist for the biodiversity of Bangladesh, the list is already eight years old, and several other undocumented species remain. From the Red List database, we extracted group names (e.g., Birds, Butterflies), Order, Family, scientific name, English name, local name, and the Red List assessment status in Bangladesh. While cross-checking the data, we noticed that the Family information for butterflies was old and did not match the Family names found in GBIF. We updated family names following the GBIF taxonomy to handle this issue. Specifically, we updated the information for four Families (Acraeidae, Amathusiidae, Danaidae, and Satyridae) and moved them under the Family Nymphalidae. Finally, we removed regionally extinct species (assessed as RE in the national Red List) from the species list.

Step 1 - Group selection

First, we searched for each taxonomic group (e.g., Bird) and added the country name (Bangladesh) at the end (e.g., Bird Bangladesh). From our previous experiences, we were aware that some Facebook groups contain the term 'biodiversity'. Consequently, we also searched Facebook using the keywords 'biodiversity Bangladesh'. Altogether our search keywords included the following seven combinations: 'Amphibian Bangladesh', 'Bird Bangladesh', 'Butterfly Bangladesh', 'Crustacean Bangladesh', 'Fish Bangladesh', 'Mammal Bangladesh', and 'Reptile Bangladesh'. Based on the search results, we filtered the most popular Facebook groups, for each taxon, based on i) moderation activity (e.g., whether group moderators help users with species identification), ii) group rules (if the group has strict rules about the location and date of the photographs), and iii) group activities (if members post every day).

Step 2 - Data extraction

We followed a range of approaches when extracting species distribution data from Facebook groups. Before starting the data extraction process, we conducted a test search with some common and rare species. To control naming inconsistencies, we searched each species with its scientific name, English name, and local name in Bengali, which led to three rounds of search for each species in each group. To expedite the search process, we restricted searches by year with the 'Date Posted' tab. Here, we filtered individual years (e.g., 2021) for each search and extracted all relevant results for that year. We repeated this process until there were no more new search results. Finally, we removed duplicate records from the compiled datasheet (see above).

From each post, we extracted the following information: species name (search keyword), life stage (e.g., adult, egg), date (day/month/year), location information, and the name of the photographer. For quality control, we double-checked if the species identification was correct before extracting its information. We also skimmed through the comments of individual photographs. We excluded photographs if i) the photograph was not from Bangladesh, ii) the photograph was unclear, iii) the species was not identified up to the species level, iv) the locality information was missing, and v) if the area of the location was over 100 km^2.

Step 3 - Georeferencing

As stated above, Facebook lacks an automated georeferencing system. Using the location information from each post, we searched it in Google Map (https://www.google.com/maps), selected a random point within that area, and extracted the latitude and longitude (in decimals). Considering that we discarded photographs if the specified location was an area over 100 km^2, the precision uncertainty of these extracted distribution records was within 10 km.

Results

Data harvested from Bangladesh

Overall, we identified 42 Facebook groups for different animals in Bangladesh (supplementary Table S1). While 15 of these groups are on birds, only one is on odonatans (Figure 2). There are only four groups with > 50,000 members. Two groups are focused on multiple taxa, one on birds and one on reptiles (Figure 2; supplementary Table S1).

Of the 42 Facebook groups, we chose the seven most popular Facebook groups: Birds Bangladesh (https://www.facebook.com/groups/2403154788); Deep Ecology And Snake Rescue Foundation (https://www.facebook.com/groups/959896627527624); Biodiversity of Bangladesh; (https://www.facebook.com/groups/249240636186853); Butterfly Bangladesh; (https://www.facebook.com/groups/488719627817749); Mammals of Bangladesh; (https://www.facebook.com/groups/647662968655338); Amphibians and Reptiles of Bangladesh; (https://www.facebook.com/groups/560709511527645); Biodiversity of Greater Kushtia (https://www.facebook.com/groups/244807066739477).

We collated 44,726 occurrence records for 967 species, ranging between 1-719 records per species. These data included 45 amphibian species, 494 bird species, 265 butterfly species, 72 mammal species, and 91 reptile species (Figures 3A, 4). We could not locate any species occurrence records for fishes or crustaceans.

While the median occurrence records per species were 27, it varied substantially amongst taxa: the median species occurrence records were 54 for birds, 33 for butterflies, 5 for mammals, and 3 for reptiles. There were 196 species with 80 or more records, of which only seven were butterflies (no threatened species), and the rest were birds (one threatened species, *Threskiornis melanocephalus*, Vulnerable). The following were the most popular species from each group: *Alcedo atthis* (birds, 719 records), *Danaus chrysippus* (butterflies, 107 records), *Prionailurus viverrinus* (mammals, 66 records), *Xenochrophis piscator* (reptiles, 39 records), *Duttaphrynus melanostictus* and *Polypedates leucomystax* (amphibians, 14 records).

The number of occurrence records grew substantially with time, with some random fluctuations (Figure 3B). Although Facebook started in the early 2000s, we obtained many records before that date. For birds, there were records available from 1978. For other taxa, the records started in 1992 for reptiles, 1998 for mammals, 2004 for butterflies and 2005 for amphibians. For all taxa, the number of species occurrence records increased markedly from the start date to early 2020; however, the numbers rapidly declined afterwards, possibly due to the COVID-19 pandemic (Figure 3B).

We obtained distribution data for 260 threatened species (27% of species), 651 non-threatened species (67% of species), and 56 Data Deficient species (6% of species) (see the supplementary Table S2). While 12% of the species distribution data was for threatened species, it varied substantially across taxa. For mammals, 70% of the occurrence records were for threatened species, 47% for butterflies, 14% for reptiles, 6% for amphibians, and 3% for birds (Figure 4A). The pattern was different when considering the proportion of species numbers within taxa recorded, both for threatened and non-threatened species. For butterflies, we obtained records for twice more species that were classed as threatened (167) than non-threatened (84). For other groups, threatened species contributed to 46% of all mammal species, 20% for reptiles, 18% for amphibians, and 7% for birds (Figure 4B).

Discussion

The increasing number of citizen science applications are contributing to a sharp increase in species distribution records (Heberling et al., 2021). However, despite this increase, a substantial bias remains in our understanding of global biodiversity – the distribution of tropical species remains overlooked (Hortal et al., 2015; Di Marco et al., 2017; Kühl et al., 2020; Hughes et al., 2021). Here, we present a standardised protocol for Facebook data extraction, and we demonstrate its use on Facebook biodiversity records from Bangladesh, to assess how data obtained from this platform can help reduce the global bias in biodiversity knowledge. We obtained nearly 45,000 records for 967 species from Bangladesh, of which 27% are nationally threatened, and many are Data Deficient. We found that over time, data increased sharply for all taxa; however, there was a substantial decline in the amount of data in 2021 during the COVID-19 pandemic. We further provide step-by-step guidelines to extract species locality data, which could be helpful for future researchers to obtain local and global biodiversity data that could aid in conservation assessments.

Facebook data has the potential to improve our biodiversity knowledge. However, there are a few issues that users should consider (Di Minin et al., 2015; Toivonen et al., 2019; Chowdhury et al., 2023b). First, extracting data from Facebook is a time-consuming and multi-step process. On average, it took us nearly 33 minutes to complete the data extraction process for a single species (i.e. ~532 hours in total). Second, Facebook photographs do not contain precise or automatic geolocation functions, which results in coordinate uncertainties. Third, maintaining the quality of posts in Facebook groups requires active moderation activities, which include a high level of taxonomic expertise, and spending time reviewing every photograph. While this is typical for Bangladesh, where administrators, moderators, and users help maintain every photograph's quality, this might not be the case for many countries. We present a standardised protocol to extract biodiversity data from Facebook. While automatic extraction is not (yet) possible, developing such tools to expedite the process would be useful. Likewise, developing tools that automatically deposit species distribution records from Facebook to the global biodiversity repositories could be instrumental in harvesting this source for biodiversity records (Jarić et al., 2020; Correia et al., 2021). If more resources are allocated for this endeavour, people can be appointed to check Facebook posts regularly, help administrators and moderators maintain group quality, extract data regularly, and create a local database that will eventually be deposited into global biodiversity repositories such as GBIF.

When using social media data for research purposes, inherent risks must be addressed to protect individuals from potential harm, whether intentional or unintentional (Di Minin et al., 2021). To mitigate these risks and ensure the users' safety, it is important to adopt practices such as data minimisation, anonymisation, and strict data management protocols. Employing risk-based approaches, such as conducting data privacy impact assessments, can aid in identifying and minimising privacy risks for social media users. Besides, when sharing the observation records, authors should carefully consider whether they should share threatened species' locations (Sbragaglia et al., 2021). This not only showcases accountability but also ensures compliance with data protection laws, to safeguard the privacy of individuals involved in the research process (Di Minin et al., 2021).

Conclusion

While scientists worldwide are formulating innovative approaches to improve our understanding of species distributions, there remain substantial taxonomic and regional gaps and biases. Several studies have shown Facebook's importance in extracting biodiversity records (e.g., Chowdhury et al., 2021, 2023b,c); however, there is no standard pipeline that researchers can follow. Here, we provide detailed guidelines on preparing a list of relevant Facebook groups and recommend approaches to search efficiently. While the entire process still requires time, as we could not automate this yet, it is worth the effort, given we obtained thousands of species occurrence records for a tropical biodiverse country, Bangladesh. We recommend a concerted effort to create a global database containing biodiversity groups on Facebook.

References

Andrachuk, M., Marschke, M., Hings, C., & Armitage, D. (2019). Smartphone technologies supporting community-based environmental monitoring and implementation: a systematic scoping review. *Biological Conservation*, *237*, 430-442.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, *19*, 10-15.

Bela, G., Peltola, T., Young, J. C., Balázs, B., Arpin, I., Pataki, G., ... & Bonn, A. (2016). Learning and the transformative potential of citizen science. *Conservation Biology*, *30*(5), 990-999.

Callaghan, C. T., Mesaglio, T., Ascher, J. S., Brooks, T. M., Cabras, A. A., Chandler, M., ... & Young, A. N. (2022). The benefits of contributing to the citizen science platform iNaturalist as an identifier. *PLoS biology*, *20*(11), e3001843.

Callaghan, C. T., Poore, A. G., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., ... & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, *71*(1), 55-63.

Cardoso, P., Erwin, T. L., Borges, P. A., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, *144*, 2647-2655.

Chamberlain, J. (2018). Using social media for biomonitoring: how Facebook, Twitter, Flickr and other social networking platforms can provide large-scale biodiversity data. In *Advances in Ecological Research* (Vol. 59, pp. 133-168). Academic Press.

Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., ... & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280-294.

Chowdhury, S., Zalucki, M. P., Hanson, J. O., Tiatragul, S., Green, D., Watson, J. E., & Fuller, R. A. (2023a). Three-quarters of insect species are insufficiently represented by protected areas. *One Earth*, *6*(2), 139-146.

Chowdhury S, Aich U, Rokonuzzaman M, Alam S, Das P, Siddika A, … & Callaghan CT. (2023b). Increasing biodiversity knowledge through social media: a case study from tropical Bangladesh. *BioScience*.

Chowdhury S, Fuller R, Ahmed S, Alam S, Callaghan C, Das P, … & Bonn A. (2023c). Using social media records to inform conservation planning. *Conservation Biology*. (in press) https://doi.org/10.1111/cobi.14161.

Chowdhury, S., Jennions, M. D., Zalucki, M. P., Maron, M., Watson, J. E., & Fuller, R. A. (2022a). Protected areas and the future of insect conservation. *Trends in Ecology & Evolution*, *38*, 85-95.

Chowdhury, S., Aich, U., Rokonuzzaman, M., Alam, S., Das, P., Siddika, A., ... & Fuller, R. A. (2022b). Spatial occurrence data for the animals of Bangladesh derived from Facebook. *Pangaea*. DOI: https://doi.pangaea.de/10.1594/PANGAEA.948104.

Chowdhury, S., Alam, S., Chowdhury, S. U., Rokonuzzaman, M., Shahriar, S. A., Shome, A. R., & Fuller, R. A. (2021). Butterflies are weakly protected in a mega-populated country, Bangladesh. *Global Ecology and Conservation*, *26*, e01484.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science*, *1*(2), 75-88.

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment*, *176*, 273-291.

Correia, R. A., Ladle, R., Jarić, I., Malhado, A. C., Mittermeier, J. C., Roll, U., ... & Di Minin, E. (2021). Digital data sources and methods for conservation culturomics. *Conservation Biology*, *35*(2), 398-411.

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., ... & Watson, J. E. M. (2017). Changing trends and persisting biases in three decades of conservation science. *Global Ecology and Conservation*, *10*, 32-42.

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*(2), 437-446.

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*, 63.

Diniz Filho, J. A. F., Jardim, L., Guedes, J. J., Meyer, L., Stropp, J., Frateles, L. E. F., ... & Hortal, J. (2023). Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Frontiers of Biogeography*. https://doi.org/10.21425/F5FBG59566.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., & Collen, B. (2014). Defaunation in the Anthropocene. *Science*, *345*(6195), 401-406.

Finn, C., Grattarola, F., & Pincheira-Donoso, D. (2023). More losers than winners: investigating Anthropocene defaunation through the diversity of population trends. *Biological Reviews*. https://doi.org/10.1111/brv.12974.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, *118*(6), e2018093118.

Hochkirch, A., Samways, M. J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P., ... & Dijkstra, K. D. B. (2021). A strategy for the next decade to address data deficiency in neglected biodiversity. *Conservation Biology*, *35*(2), 502-509.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 523-549.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., ... & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*(9), 1259-1269.

IUCN Bangladesh. 2015. Red list of Bangladesh: a brief on assessment result 2015, IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh.

Jarić, I., Correia, R. A., Brook, B. W., Buettel, J. C., Courchamp, F., Di Minin, E., ... & Roll, U. (2020). iEcology: harnessing large online resources to generate ecological insights. *Trends in Ecology & Evolution*, *35*(7), 630-639.

Kahle, D. & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144-161.

Kühl, H.S., Bowler, D.E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, D., … & Bonn, A. (2020) Effective biodiversity monitoring needs a culture of integration. *One Earth*, *3*, 462-474.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A., ... & Guarino, R. (2021). Facebook groups as citizen science tools for plant species monitoring. *Journal of Applied Ecology*, *58*(10), 2018-2028.

Meiri, S., Chapple, D. G., Tolley, K. A., Mitchell, N., Laniado, T., Cox, N., ... & Roll, U. (2023). Done but not dusted: Reflections on the first global reptile assessment and priorities for the second. *Biological Conservation*, *278*, 109879.

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, *6*(1), 1-8.

Miqueleiz, I., Bohm, M., Ariño, A. H., & Miranda, R. (2020). Assessment gaps and biases in knowledge of conservation status of fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *30*(2), 225-236.

Murali, G., Iwamura, T., Meiri, S., & Roll, U. (2023). Future temperature extremes threaten land vertebrates. *Nature*, *615*(7952), 461-467.

O'Neill, D., Häkkinen, H., Neumann, J., Shaffrey, L., Cheffings, C., Norris, K., & Pettorelli, N. (2023). Investigating the potential of social media and citizen science data to track changes in species' distributions. *Ecology and Evolution*, *13*(5), e10063.

Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., ... & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*(6187), 1246752.

Pocock, M.J., Roy, H.E., August, T., Kuria, A., Barasa, F., Bett, J., Githiru, M., Kairo, J., Kimani, J. & Kinuthia, W. (2019). Developing the global potential of citizen science: Assessing opportunities that benefit people, society and the environment in East Africa. *Journal of Applied Ecology*, *56*, 274-281.

Ramirez, F., Sbragaglia, V., Soacha, K., Coll, M., & Piera, J. (2022). Challenges for marine ecological assessments: completeness of findable, accessible, interoperable, and reusable biodiversity data in European seas. *Frontiers in Marine Science*, *8*, 802235.

Roy, H., Groom, Q., Adriaens, T., Agnello, G., Antic, M., Archambeau, A. S., ... & Cardoso, A. C. (2018). Increasing understanding of alien species through citizen science (Alien-CSI). *Research Ideas and Outcomes*, *4*, e31412.

Stork, N. E. (1997). Measuring global biodiversity and its decline. *Biodiversity II: understanding and protecting our biological resources*, *41*, 41-68.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., ... & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298-315.

WWF. (2022) Living Planet Report 2022 – Building a naturepositive society. Almond, R.E.A., Grooten, M., Juffe Bignoli, D. & Petersen, T. (Eds). WWF, Gland, Switzerland.

List of Figures

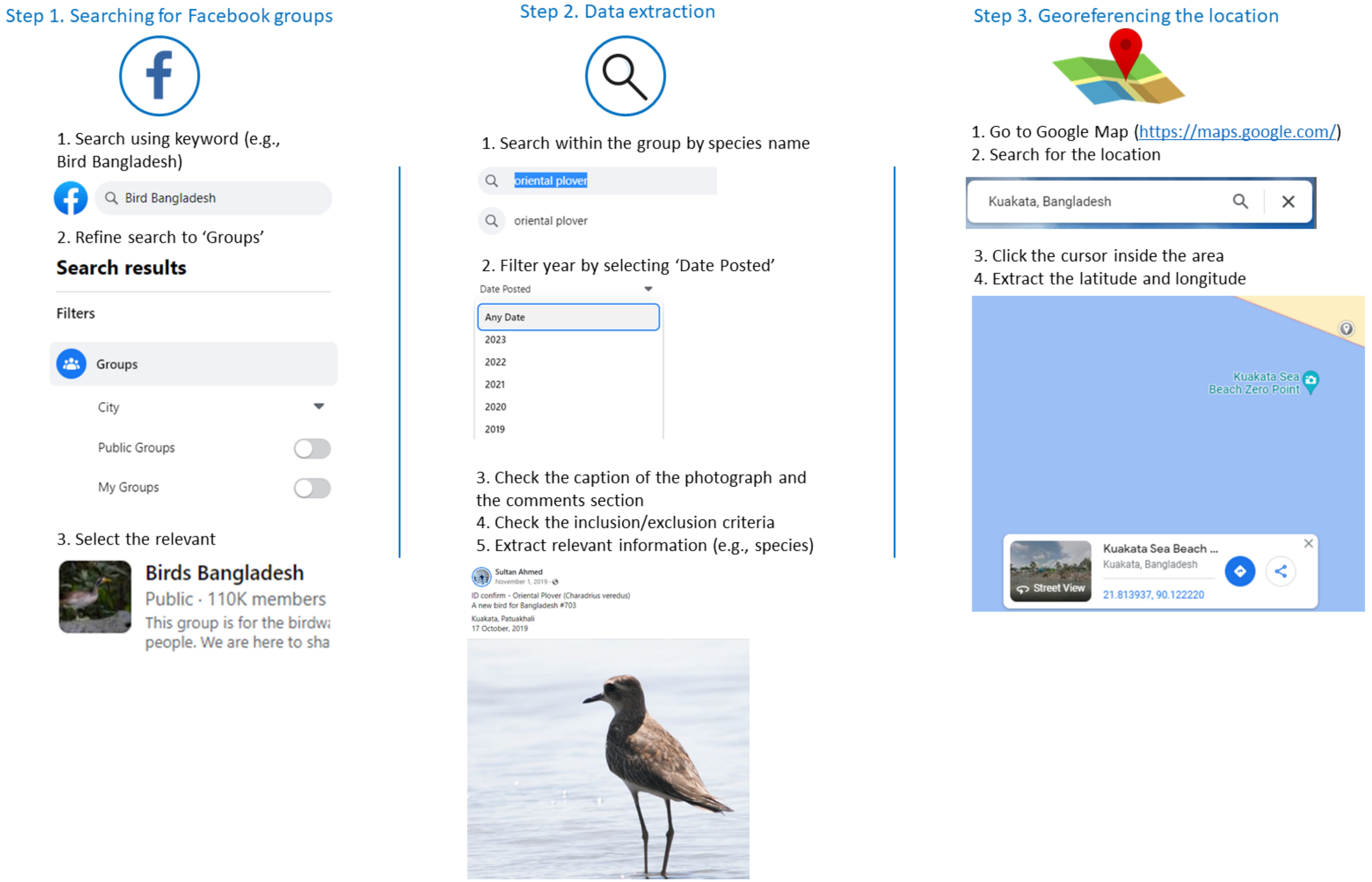
Figure 1. (A) The data extraction pipeline for obtaining species distribution records from Facebook, and (B) the inclusion/exclusion criteria when filtering species photographs for different animal groups of Bangladesh.

Figure 2. The number and membership details of the available Facebook groups for Bangladeshi animals. Here, ‘Herpetofauna’ is ‘Amphibians’ and ‘Reptiles’, and we named ‘Multiple taxa’ when the Facebook group was focused on > 2 taxa.

Figure 3. (A) The density (at 1 km^2) and (B) yearly growth of species occurrence records, obtained from Facebook for different animal groups of Bangladesh.

Figure 4. The relative proportion of threatened and non-threatened species in occurrence records (A) and species records (B), obtained from Facebook for different animal groups in Bangladesh, namely amphibians, birds, butterflies, mammals and reptiles. Here the black dots represent the percentages of species that are nationally threatened.

Figure 1.



Step 1. Searching for Facebook groups
1. Search using keyword (e.g., Bird Bangladesh)
Bird Bangladesh
2. Refine search to 'Groups'
Search results
Filters
Groups
City
Public Groups
My Groups
3. Select the relevant
Birds Bangladesh
Public · 110K members
Step 2. Data extraction
1. Search within the group by species name
oriental plover
oriental plover
2. Filter year by selecting 'Date Posted'
Date Posted
Any Date
2023
2022
2021
2020
2019
3. Check the caption of the photograph and the comments section
4. Check the inclusion/exclusion criteria
5. Extract relevant information (e.g., species)
Sultan Ahmed
ID confirm - Oriental Plover (Charadrius veredus)
A new bird for Bangladesh #703
Kuakata, Patuakhali
17 October, 2019
Step 3. Georeferencing the location
1. Go to Google Map (https://maps.google.com/)
2. Search for the location
Kuakata, Bangladesh
3. Click the cursor inside the area
4. Extract the latitude and longitude
Kuakata Sea Beach Zero Point
Kuakata Sea Beach ...
Kuakata, Bangladesh
Street View
21.813937, 90.122220

Figure 2.

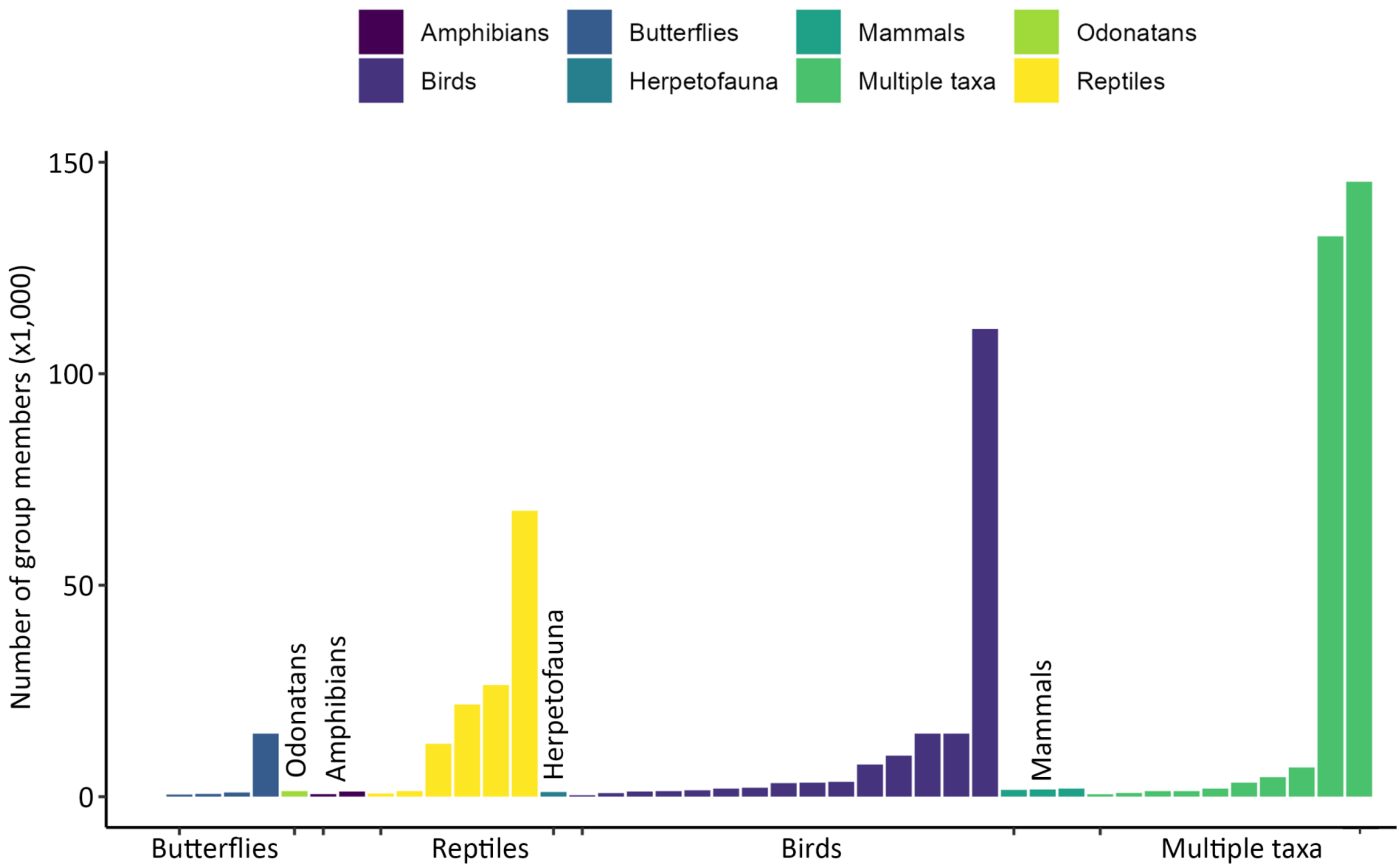
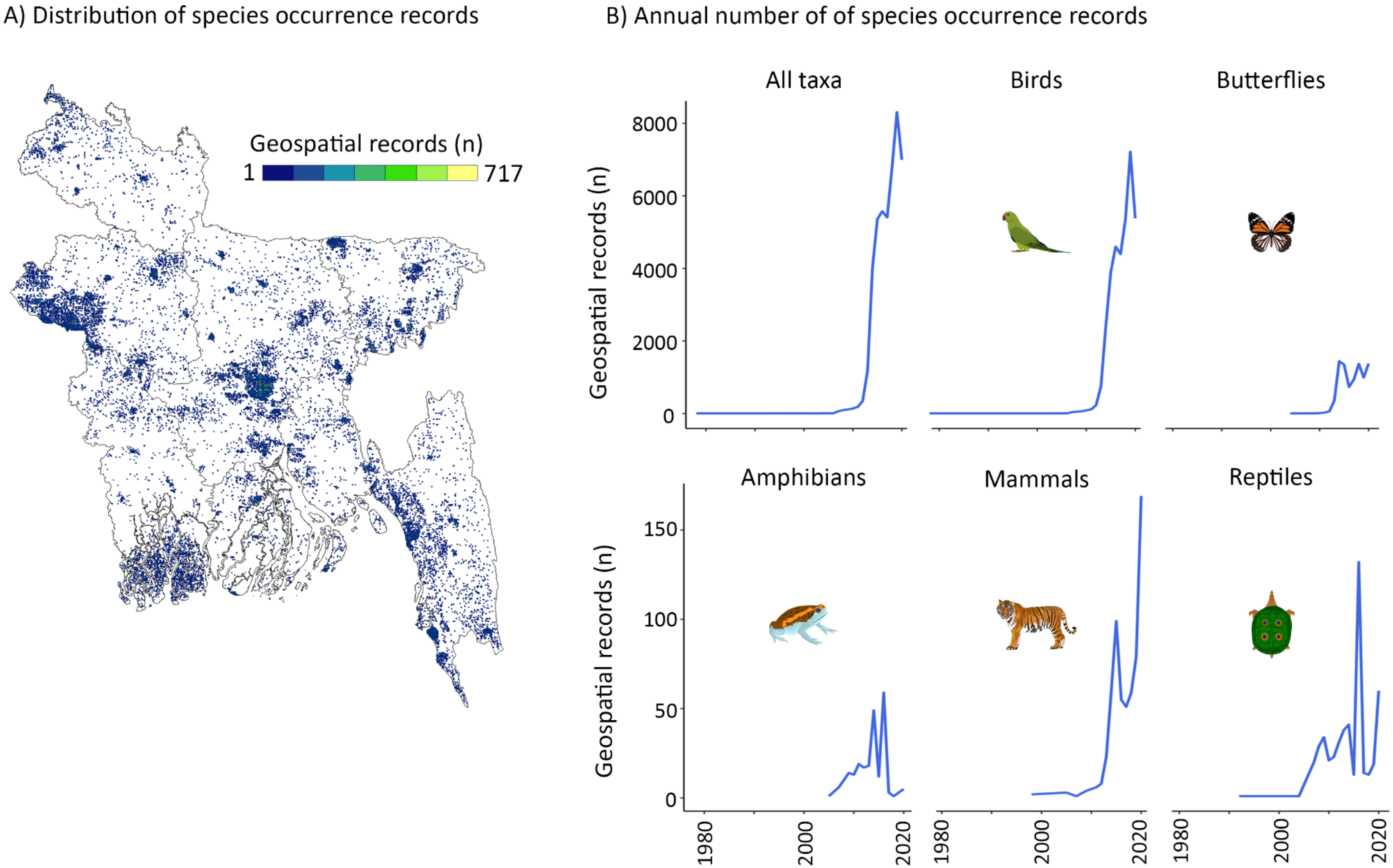
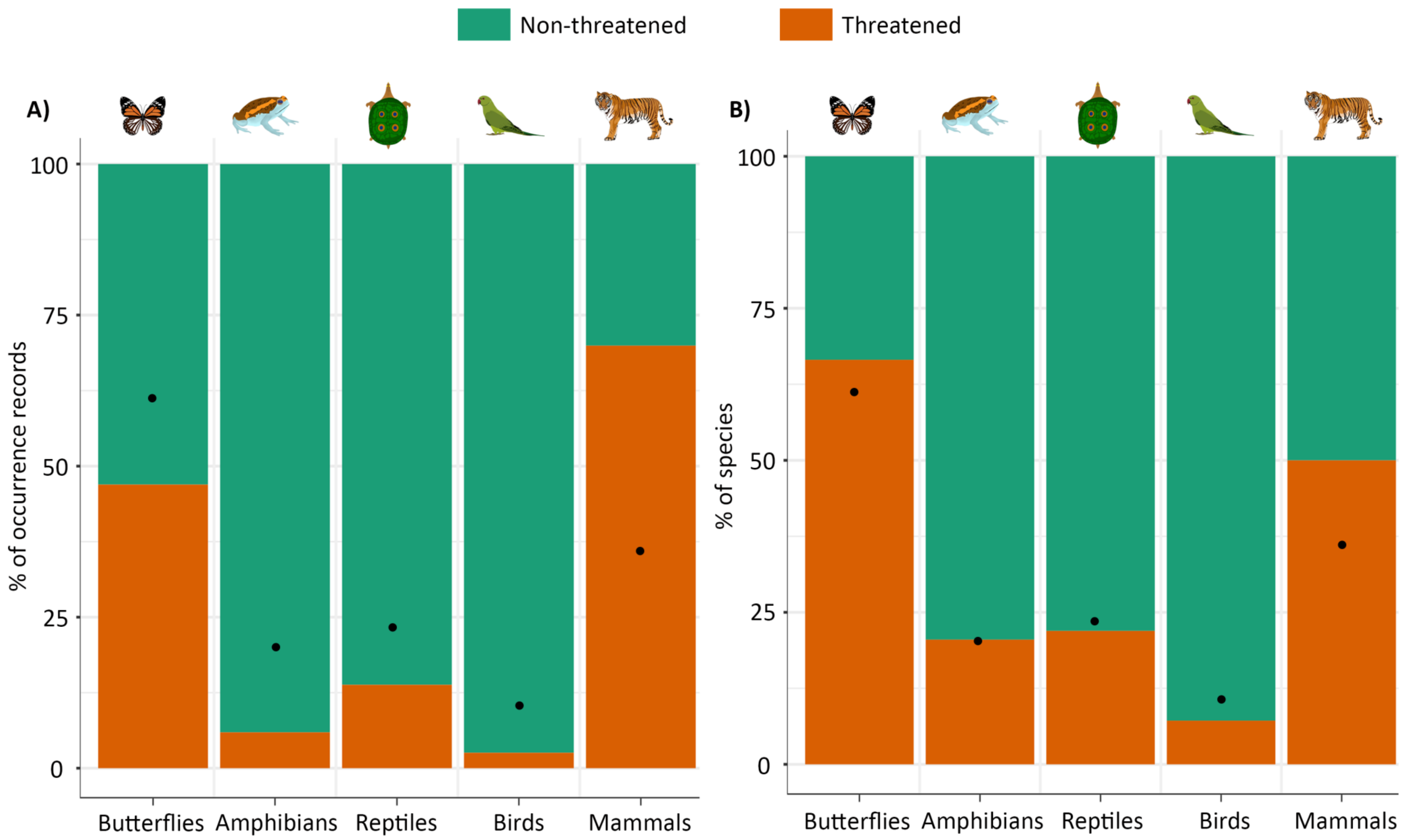

Amphibians
Birds
Butterflies
Herpetofauna
Mammals
Multiple taxa
Odonatans
Reptiles
Number of group members (x1,000)
150
100
50
0
Odonatans
Amphibians
Herpetofauna
Mammals
Butterflies
Reptiles
Birds
Multiple taxa

Figure 3.


A) Distribution of species occurrence records
Geospatial records (n)
1
717
B) Annual number of of species occurrence records
All taxa
Birds
Butterflies
Amphibians
Mammals
Reptiles
Geospatial records (n)
8000
6000
4000
2000
0
Geospatial records (n)
150
100
50
0
1980
2000
2020
1980
2000
2020
1980
2000
2020

496 **Figure 4.**



497