Universidade de Aveiro
2020

HARITH OMAR MORGADINHO FAROOQ

AVANÇOS NA CONSERVAÇÃO DA BIODIVERSIDADE EM ÁFRICA SOB EVIDÊNCIA CIENTÍFICA

TOWARDS DATA-DRIVEN BIODIVERSITY CONSERVATION IN AFRICA

Universidade de
Aveiro
2020

HARITH OMAR MORGADINHO FAROOQ

AVANÇOS NA CONSERVAÇÃO DA BIODIVERSIDADE EM ÁFRICA SOB EVIDÊNCIA CIENTÍFICA

TOWARDS DATA-DRIVEN BIODIVERSITY CONSERVATION IN AFRICA

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Biologia e Ecologia das Alterações Globais, realizada sob a orientação científica do Doutor Amadeu Soares, Professor Catedrático do Departamento de Biologia da Universidade de Aveiro e do Doutor Alexandre Antonelli, Professor Catedrático da Universidade de Gotemburgo

Apoio financeiro da FCT
através da bolsa
PD/BD/128123/2016

o júri

presidente

Doutor Armando José Formoso de Pinho
Professor Catedrático, Universidade de Aveiro

Doutora Maria da Luz da Costa Pereira Mathias
Professora Catedrática, Universidade de Lisboa

Doutor Amadeu Mortágua Velho da Maia Soares (Orientador)
Professor Catedrático, Universidade de Aveiro

Doutor Fernando Manuel Raposo Morgado
Professor Associado Com Agregação, Universidade de Aveiro

Doutora Maria Ana Azeredo de Dornelas
Lecturer, University oi St. Andrews

Doutor Luca Santini
Investigador, Radboud University

agradecimentos

1 would like to start my acknowledgements by thanking the most important person in my life. My mom. She gave me everything she could and what she couldn't, just so nothing would be missing in arder for me to become my best self. Thank you. To have the privilege of doing my PhD between Portugal and Sweden, some people played major roles. 1 thank Francisco Noa for accepting my study leave from Unilúrio and Simon Loader, Fernando Morgado, Hugo Costa and of course Amadeu Soares for helping me with the project design. Amadeu was my supervisor in my bachelors, masters and now PhD. Thank you for believing in me. 1 thank EFN-WWF and WCS for helping me support my fieldwork and labwork, which allowed me to pursue my dream research. 1 thank Cristóvão, Josué, Francesco and Alex for joining me on my fieldwork in Mozambique and 1 thank Rui Rocha and Fernando Morgado for all the help with the logistics and for being good friends. 1 thank my colleagues in Gothenburg who helped me with R, such as Zizka, Josué and Dom, phylogenetics was with Tobi and *bash* with Daniel. 1 thank Anna for all the patience she had with me during labwork. 1 thank senior people such as Christine, Danielle and Bengt for all the high-level discussions 1 was fortunate to watch and 1 thank my other colleagues with whom 1 had so much fun at the department: Bea, Matheus, Viva, Igor, Juan, Ferran, Rob, Paola, Pavel, Mafe, Gytis, Helen, Hesam, Ntwae, Jonna and Johanna and many others 1 am probably forgetting. Some of you became friends outside work, which was really nice. 1 thank Rhian for proofreading all my papers and 1 thank Allison Perrigo for always being there for me. For the ups and downs of life, 1 couldn't have asked for a better friend in Gothenburg. Finally, for my supervisors in Sweden, Alexandre Antonelli and Seiren Faurby. 1 can't thank you enough for guiding me through this journey. It was a pleasure to work with you both and 1 look forward to the next projects where we can all collaborate again.

palavras-chave

África, amostragem de espécies, biodiversidade, conservação, endemismo, WEGE.

resumo

Em África podemos encontrar uma diversidade incrível de ecossistemas que alojam uma enorme diversidade de espécies. Infelizmente, de acordo com a literatura mais recente, as populações de mamíferos, aves, peixes, anfíbios e répteis sofreram um declínio de 65% nos últimos 50 anos. As principais causas incluem a desflorestação, o crescimento populacional, urbanização, a recolha de recursos naturais e o desmatamento para a implementação da agricultura. Com a previsão que a população humana em África duplique nos próximos 50 anos, espera-se um efeito exacerbado das causas que neste momento contribuem para o declínio da biodiversidade. O facto de um terço de África viver abaixo da linha de pobreza e dados os altos níveis de corrupção torna especialmente importante que os parcos recursos existentes destinados à preservação da biodiversidade possam ser transparente e efetivamente distribuídos pelas áreas de biodiversidade mais importantes.

Para além dos capítulos da Introdução e Discussão gerais, esta tese contém quatro capítulos, que representam artigos publicados ou submetidos. No capítulo 1, propõe-se uma fórmula que permite hierarquizar áreas de acordo com a sua importância em termos de biodiversidade. No capítulo 2 demonstra-se a falta de inventários de biodiversidade em África, revelando-se ainda uma tendência importante da amostragem no continente. São também feitas previsões em relação ao tempo e esforço necessário para amostrar biodiversidade em África. No capítulo 3, compila-se informação importante em anfíbios e répteis de uma área pouco amostrada em termos de biodiversidade no Norte de Moçambique, e no capítulo 4 mostra-se que os endemismos são muito dependentes da escala ou da taxonomia usada que é aplicada durante as análises.

keywords

África, biodiversity, conservation, endemisms, species sampling, WEGE.

abstract

In Africa we find incredibly diverse ecosystems which are home to a huge diversity of species. Unfortunately, as the latest literature suggests, the population sizes of mammals, birds, fish, amphibians and reptiles have declined by 65% in the last 50 years. The main drivers include deforestation, human population growth, urbanization, collection of natural resources and clearing of land for agriculture. With the African population expected to double in the next 50 years, it is expected an exacerbated effect of the drivers that currently contribute to biodiversity decline. The fact that a third of Africa lives under the poverty line together with high leveis of corruption makes it especially important that the sparse existing resources dedicated to biodiversity preservation can be transparently and effectively distributed throughout the most important areas for biodiversity.

Besides the chapters of general introduction and discussion, this thesis contains four chapters representing published or submitted papers. In chapter 1, 1 propose a new metric that allows the ranking of areas based on biodiversity importance. In chapter 2, 1 showcase the lack of biodiversity sampling in Africa unveiling an important sampling bias and making predictions on sampling time and effort to sample biodiversity in Africa. In chapter 3, 1 compile important baseline information on amphibians and reptiles from a poorly known area in terms of biodiversity in northern Mozambique and in chapter 4, we show that endemisms are very dependent on which scale or taxonomic treatment we decide to use in our analysis.

Overall, the work developed in this thesis is a contribution to the transparency and prioritisation procedures at biodiversity management globally and showcases the current situation of the African biodiversity sampling and how it may translate into the future. Furthermore, it provides an example on how biodiversity baseline information can be acquired and made available as well as highlighting the importance of taxonomy and scale when inferring patterns from spatial analysis.

CONTENTS

GENERAL INTRODUCTION

Biodiversity in Africa

The African continent is home to a great diversity of ecosystems, from the largest desert in the world – the Sahara, to the most endemic-rich – the Cape Flora, and also comprising species rich rain forests (Couvreur et al. 2020; Linder 2003; Linder et al. 2012).

The continent houses around a quarter of the global vertebrate diversity. Of all the 5899 mammals assessed by IUCN, 26% (1539 species), 17% of the amphibians (1185 out of 6892 species) and 24% (2662 out of 11147 species) of the birds occur in Africa (Fig. 1) (IUCN 2020).

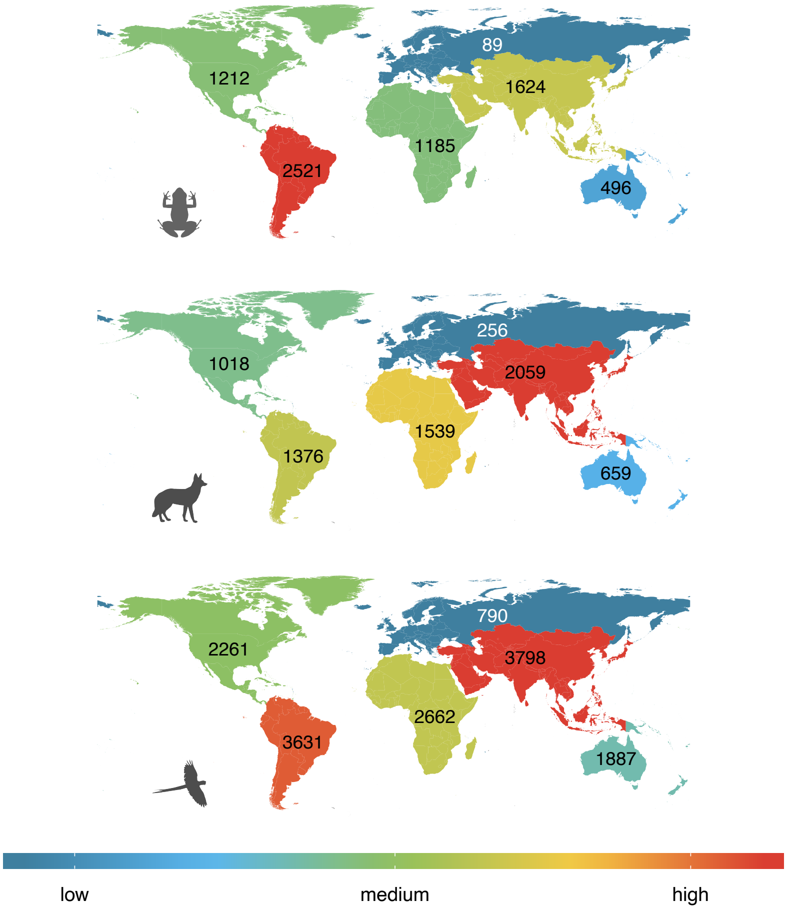


Figure 1: IUCN assessed species by continent. Top panel: Amphibians; Middle panel: Mammals; Bottom panel: Birds. Lower values are color-coded blue, middle values as yellow and high values as red (IUCN 2020).

The African mainland has approximately 44,830 species of angiosperms (Klopper et al. 2006), representing 15% of all angiosperms worldwide (295,383) (Christenhusz and Byng 2016) and despite only few studies on insect biodiversity have been conducted in the continent, 100,000 species have already been recorded, representing 9.4% of all insects worldwide (1,060,704) (Foottit and Adler 2017; Scholtz and Mansell 2017). These fractions may also be underestimates, since Africa is so poorly studied.

All these species are distributed across a number of regions that share similar habitats or vegetation types at a global scale – biomes. Nine out of 14 that exist globally, occur in Africa (Fig 2) (Olson et al. 2001).

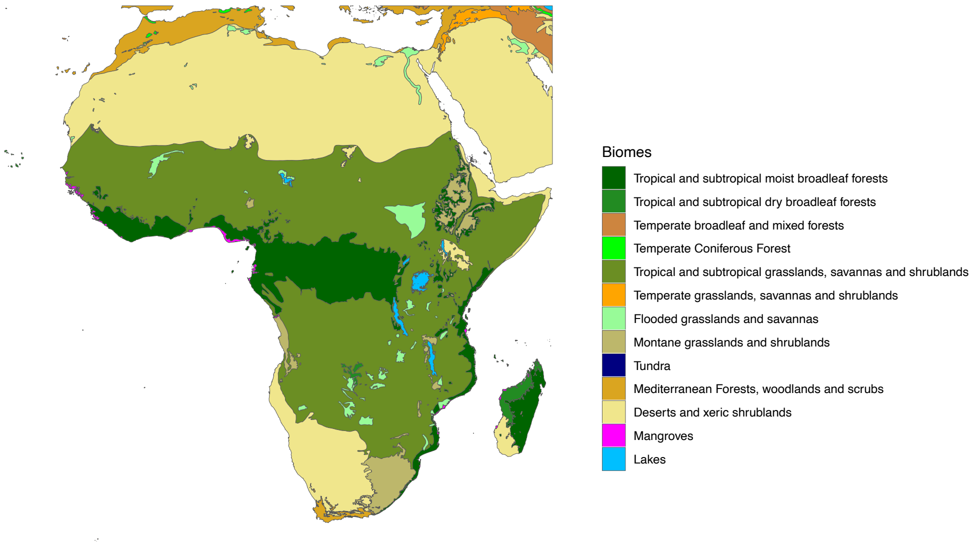


Figure 2: Biomes occurring in Africa (Olson et al. 2001). A great diversity of biomes occurs in the African continent.

Africa's biomes extend from mangroves to deserts, from Mediterranean to tropical forests, from temperate to sub-tropical and montane grasslands and savannahs, and even to ice-capped mountains (Olson et al. 2001).

Biodiversity is not uniformly distributed across the globe, and some areas accumulate more species than others. These areas with high biodiversity and high levels of threat have been mapped and named biodiversity hotspots (Myers et al. 2000). In Africa, there are nine of the world's 34 biodiversity hotspots. Such regions are defined by having at least 1 500 species of vascular plants (> 0.5 per cent of the world's total) as endemics and must have lost at least 70% of its original habitat (Myers et al. 2000). Therefore, biodiversity hotspots (Fig. 3) are designated on the basis of both existing biodiversity and the threats to that biodiversity with the intention of focusing protection efforts on these valuable areas. These hotspots are also home to some of the highest human densities in the world, therefore while providing vital ecosystem services such as water, pollination and climate regulation, they are also affected by human activity.

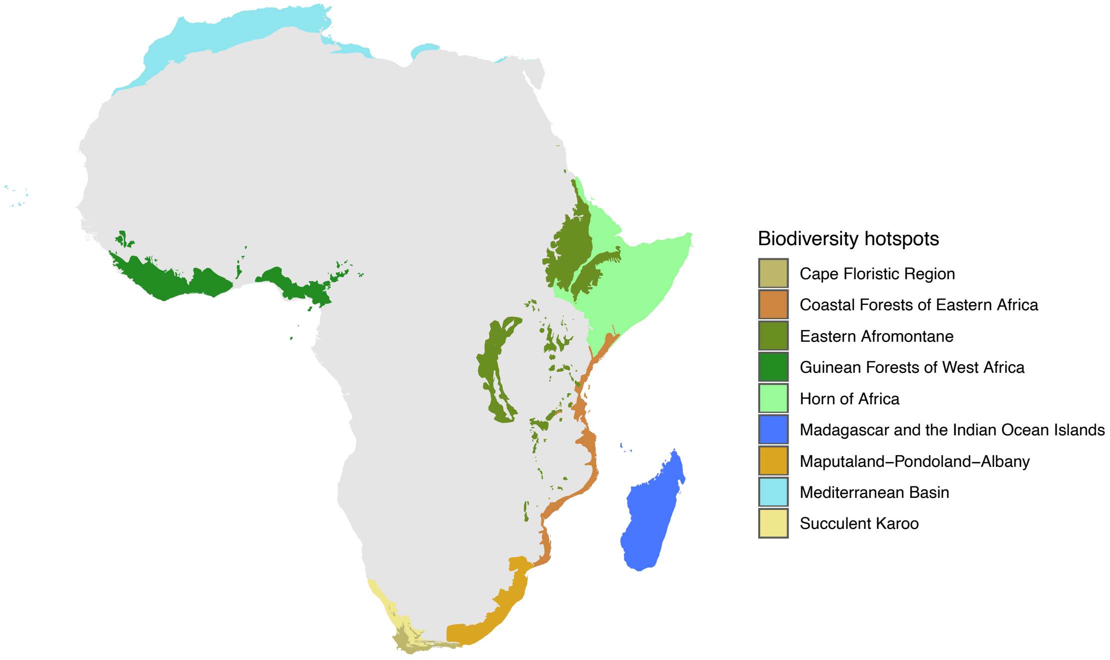


Figure 3: Biodiversity hotspots masked to Africa's mainland. Including the Mediterranean Basin, nine biodiversity hotspots occur in Africa (Noss et al. 2015).

Biodiversity is in decline and we lack time and resources to preserve species one by one (Ehrlich 1992). One way to maximize conservation investment are through large scale conservation planning initiatives focusing on selected areas such as ecoregions (Olson et al. 2001), biodiversity hotspots (Myers et al. 2000) and endemic bird areas (Stattersfield 1998). These initiatives, however, lack the capacity to identify targets at fine scale for the implementation of actual conservation initiatives. Site-specific conservation programs are crucial to preserve the natural habitats and prevent species extirpations (Bruner et al. 2001). It is therefore vital to map these sites, since the existing protected areas are seldomly created to preserve biodiversity systematically and often miss numerous species that need site-specific conservation action (Pressey et al. 1994). One way to achieve this is by using a similar approach to the IUCN Red List where quantitative and threshold-based criteria are used for for the assessment of extinction risk of species, but adapted to the concept of sites instead. Based on this idea, Eken et al. (2004) proposed a general framework and criteria for identifying key biodiversity areas (KBAs) (Fig. 4).

Figure 4: Key Biodiversity Areas in Africa. Polygons of all KBAs in Africa are represented in blue (BirdLife International 2020).

The KBA methodology uses a set of quantitative criteria that aims to be easily and consistently applied across the globe and for all taxonomic groups. The process should be developed at national or regional-levels in a bottom-up and interactive process involving local stakeholders to maximize the usefulness of such sites (Younge and Fowkes 2003).

As defined for Important Biodiversity Areas (Bibby et al. 1998), the Key Biodiversity Areas are sites that are either large enough or sufficiently interconnected to allow viable populations of the species that triggered the status of their site as a KBA. The categorization of areas is based on criteria such as presence and proportional inclusion of threatened species and ecosystems, species' distribution ranges, ecological integrity and irreplaceability. Using these criteria, together with discussions on the manageability of areas as biodiversity units, the KBA process assesses areas through a binary system, where a particular area will either trigger KBA status or not. The KBA criteria has been applied for birds, as IBAs, and for over 20 years there has been extensive research supporting the methodology as an effective framework to find important areas for site conservation.

Biodiversity research in Africa

The United Nations body for assessing the science related to climate change - The Intergovernmental Panel on Climate Change recognizes Africa as one of the least studied continents in terms of ecosystem dynamics and climate variability, despite its fast-growing human population and the potentially large impacts on natural resources and ecosystem conservation of climate change (IPCC, 2001; Washington et al. 2006).

In order to understand a system, it is crucial to collect information and test hypotheses. To understand how biodiversity is distributed and why, we need to record as much information as possible, preferable in the least biased way we can. Scientists have collected a huge amount of data throughout the years, which in many cases have been made available through the Global Biodiversity Information Facility (GBIF), a biodiversity data aggregator that allows anyone to download comprehensive lists of records from most of the world and for many taxa. However, there are data gaps and other biases that may affect scientist's analysis and hence our understanding of biodiversity globally. It is therefore crucial that we account for these gaps and biases. Some of the common biases in GBIF include taxonomic and spatial biases. The taxonomic biases are particularly evident by looking at vertebrates. Most records in GBIF are vertebrates (Di Marco et al. 2017; Troudet et al. 2017), while they only represent a small fraction of living organisms. The spatial biases are particularly noteworthy for Africa. Due to many and complex factors most of the continent remains under-sampled and in some cases even lacking any records (Scholtz and Mansell 2017).

To involve the whole biodiversity rather than solely more well-known groups or areas is crucial for the understanding of global biodiversity patterns and for the creation and implementation of effective conservation practices (Di Marco et al. 2017; Wilson 2000). The Afrotropical, despite being one of the most species rich regions in biodiversity, accounts for only 10% all biodiversity articles (Di Marco et al. 2017) (Fig. 5).

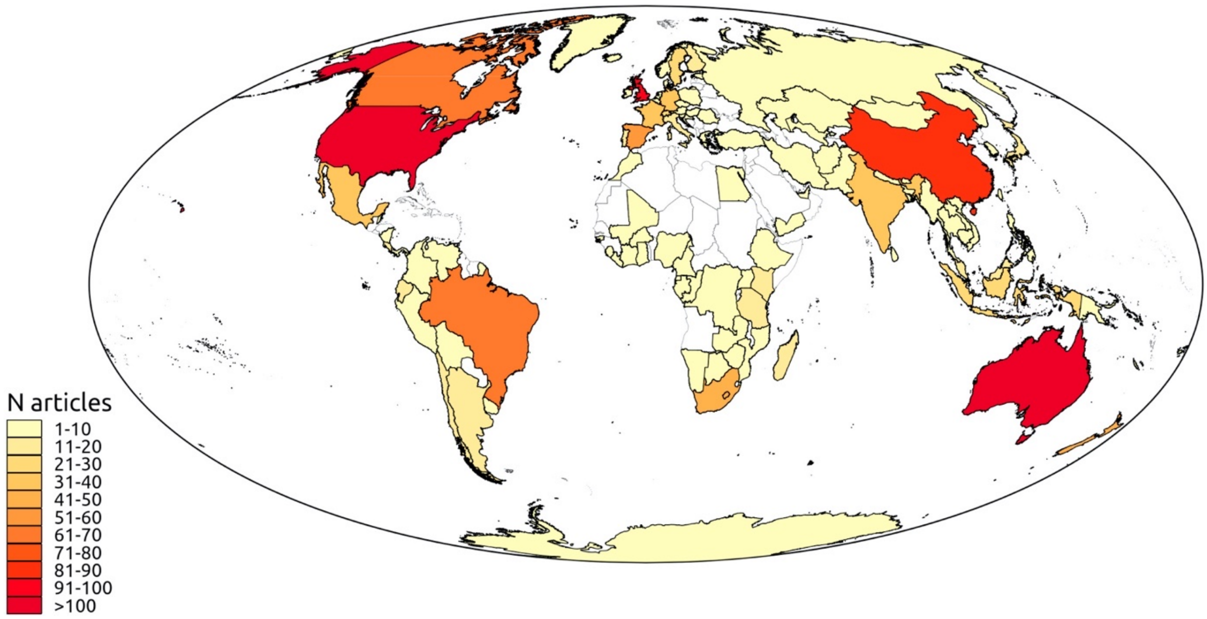


Figure 5: Number of articles by country. Africa produces considerably less research when compared to other continents (Di Marco et al. 2017).

Agreeing on a measure of biodiversity is fundamental to answering questions such as if biodiversity results in productivity, reflects stability, or if it is associated with disturbances in ecological or evolutionary history (Solbrig 1991). Quantifying through measurement is not only crucial to answer scientific questions, but can also reflect several issues at all levels of society, such as economic considerations, agriculture and pest management, pharmaceuticals, environmental applications, molecule-level benefits, sustainable development and environmental management and ecosystem function (Lovejoy 1994).

Finding such a measure can be used as a means to justify the channelling of limited funding for research and/or conservation to some areas in despite of others (Hawksworth 1995). To address that several indices that directly measure biodiversity have been proposed, such as species richness (SR: Colwell 2009), phylogenetic diversity (PD:Faith 1992), weighted endemism (WE:Crisp et al. 2001) and phylogenetic endemism (PE:Rosauer et al. 2009).

Although most biodiversity measures used in conservation relies on species richness, measures such as PD (Phylogenetic Diversity) and PE (Phylogenetic Endemism) add the evolutionary relations among species and minimize taxonomic conflicts. All these indices contribute to the understanding of how and where biodiversity is distributed in a continuous scale, and should allow the ranking of individual sites under consideration for conservation. However, their accuracy is highly dependent on the quality and availability of data, making poorly sampled areas particularly hard to evaluate (Faith 1992; Faith et al. 2004; Rosauer et al. 2009).

Although both the KBA and the raw biodiversity indices may be useful in various ways in conservation, none of them incorporate information on the threat status of the constituent species – the IUCN's Red List Assessment parameter. One exception is the Evolutionarily Distinct and Globally Endangered (EDGE) score (Isaac et al. 2007), which combines one biodiversity index – Evolutionary Distinctiveness (ED) – with the threat category of species.

ED is a measurement of the branch lengths divided by the number of species inside each clade. The EDGE score combines ED with values for species' extinction risk in order to generate a list of species that are both evolutionarily distinct and globally endangered ('EDGE species'). The EDGE score is however tailored to rank species rather than locations. Location scores may be computed as the sum of EDGE scores for all species in a site (Safi et al. 2013). However, this is not guaranteed to maximize conservation importance of individual sites, since the presence of widespread, critically endangered species produces higher EDGE scores than a vulnerable or endangered micro-endemic restricted to very few sites, which could rapidly go extinct if those sites are damaged. One example is the Atlantic bluefin tuna (*Thunnus thynnus*), which exists in great part of the Atlantic Ocean, but nevertheless is considered an endangered species (Collette et al. 2011).

The biodiversity of specific sites should arguably not be assessed by just summing the number of species existing in each location, but also taking into account other factors such as genetic diversity, distribution ranges or conservation status (Barthlott et al. 1999; Magurran 1988). Otherwise, the presence of many widespread species producing a high SR would mask the importance of vulnerable or endangered micro-endemic taxa (restricted to very few sites). The fact that SR and PD indices are known to be highly correlated with sampling effort (Bunge and Fitzpatrick 1993; Rodrigues et al. 2005; Rodrigues et al. 2011; Tucker and Cadotte 2013) further advocates against their use in inconsistently and poorly sampled regions, compared to dense sampling which will in most cases show higher species diversity. In addition, SR and PD completely disregard the information on species range in their score, which is a strong predictor of extinction risks for species (Purvis et al. 2000) and one of the fundamental aspects of conservation prioritization and management of natural resources (Anderson 1994; Myers et al. 2000; Roberts et al. 2002; Slatyer et al. 2007).

Metrics such as WE and PE are also expected to correlate with sampling effort, since new sets of records can only consolidate or increase the score but never decrease it (Lande 1996; Nipperess 2016), although this correlation seems to exist at lesser extent than in SR and PD (Oliveira et al. 2016; Soria-Auza and Kessler 2008). But besides the sampling effort issue, the use of WE and PE in conservation policies might encounter additional problems. A benefit of PE is that for two recently diverged taxa, a vast amount of their evolutionary history is shared and it therefore matters very little if they are treated as separate species or not. This is critical for groups with large genera, which often comprise both widespread and range-restricted species as a result of species radiations. Different biodiversity metrics are therefore expected to weight sites differently.

Conservation in Africa

Africa contains remarkable biodiversity, such as some of the most intact assemblages of large mammals on Earth (Morrison et al. 2007). However, habitats, species diversity and species abundance are in decline as threats to biodiversity increase. Despite close to 18% of terrestrial Africa (including Madagascar) belonging to the protected area network (WDPA 2020), the population sizes of mammals, birds, fish, amphibians and reptiles have declined by 65% in the last 50 years (WWF 2020).

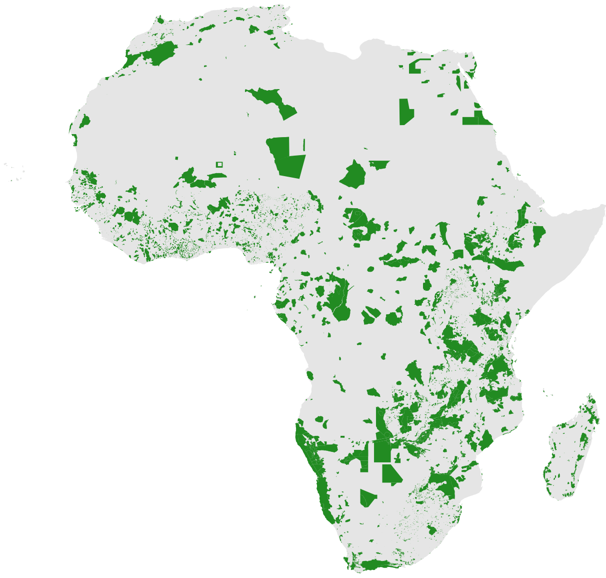


Figure 6: Protected area network in Africa. All protected areas in Africa are represented in green polygons (WDPA 2020).

In Africa, 3,444 animals and 5,131 plants are threatened with extinction on the IUCN Red List (IUCN 2020), including 21% of all freshwater species being threatened (Darwall et al. 2011) whereas population trends in smaller species are generally unknown (UNEP-WCMC 2016).

Three habitat types: 1) mangroves, 2) moist and seasonally dry forests and 3) wetlands have all declined significantly over the past twenty years, with the declines typically being in the range of one per cent loss per year (UNEP-WCMC 2016). Burgess et al. (2004), showed numerous habitats that were ‘Endangered’ or ‘Critically Endangered’ and although there has been no more recent continental scale analysis, this status is unlikely to have improved since then (UNEP-WCMC 2016).

The causes to such declines in habitats throughout Africa include climate change (Foden et al. 2007), habitat conversion (Gaston et al. 2003), over-harvesting (Grogan et al. 2010), poaching (Warchol 2004), pollution (Cohen et al. 1993) and introduction of invasive alien species (Richardson et al. 2005). Other indirect drivers include rapid population growth (United Nations 2017), urbanization (Saghir and Santoro 2018), inappropriate economic policies and technologies (Halderman 1985; Platteau 2000) and socio-political and cultural pressures (Muhumuza and Balkwill 2013). This loss of biodiversity will eventually result in the loss of nature’s contributions to people such as food, water, raw materials and energy with serious negative impacts on livelihoods (Costanza et al. 1997; IPBES 2019).

It is estimated that more than a fifth of tropical Africa has already been deforested (Aleman et al. 2018), a result of over three million hectares of natural habitat that is converted for other uses each year (UNEP 2012). In addition, the continent’s current population of 1.25 billion is likely to double by 2050, putting severe pressure on the continent’s biodiversity and nature’s contributions to people, unless appropriate policies and strategies are adopted and effectively implemented (United Nations 2017). Africa is also one of the most rapidly urbanizing continents (Saghir and Santoro 2018). Rapid and unplanned urbanization puts immense pressure on urban infrastructure and demand for services, including water supply, food supply, pollution control and waste management, as well as energy supply for households and industrial development (Annez et al. 2009).

The expected growth in African’s human population and consumption in the next decades will cause tremendous pressure on the remaining biodiversity and it is crucial that we develop solutions to safeguard nature. One way to promote biodiversity preservation is to restrict assess or exploitation in selected areas, such as creating protected areas (GBO-5 2020). Although, it is important to note that categorizing areas as protected areas will have little impact on the preservation of natural resources unless these areas are actively managed to ensure that no illegal activities are taking place. Failing to do so, these protected areas will be nothing more than paper-parks – areas that are protected in theory but not in practice. The most important causes of for failing to protect areas are lack of buy-in from local people, lack of financial investment, lack of qualified human resources and lack of local government support (Gill et al. 2017; McClanahan et al. 2016; Watson et al. 2014)

Africa being the poorest continent and lacking qualified workforce needs to be very strategic in order to optimize investment versus outcome in terms of biodiversity preservation. This is by no means an easy task since both biodiversity data and experiments on conservation procedures are both lacking in the continent.

Most of the conservation actions are experience based, based on personal experience or disseminated information instead of based on evidence or on effective proven scientific experimentation. Experience-based actions is less likely to be effective and aren't usually built in a way where knowledge can develop. But even if scientific evidence exists, there is no requirement upon which managers are requested to justify their management plans by citing scientific evidence(Pullin and Knight 2001; Sutherland et al. 2004). Therefore effective conservation will only be attained when a framework that ensures the use of conservation practice based on evidence are mandatory (Pullin and Knight 2001; Sutherland and Wordley 2017).

There is a significant gap between science and practice, resulting in numerous actions being taken with the intention to promote nature conservation but that have not yet been proved to be effective (Pullin and Knight 2001). The process of converting scientific knowledge into conservation practice (Pickett et al. 1997) and the lack of monitoring and evaluation of conservation actions such as management plans are two of the main issues revolving around the work of conservation. As conservation biologists, we must be concerned that our actions are supported by the best available evidence (Pullin and Knight 2001). In addition, conservation efforts typically rely on funding from taxpayers, businesses or charities, and it is essential that they can justify the investment made in them (Sutherland and Wordley 2017).

Much effort has gone into designating protected areas in Africa with the hope of saving these areas of crucial habitat. It is thus of great concern that conservation in Africa continues to face an enduring skills shortage (Wilson et al., 2016). Despite many examples of progress, conservation challenges and conflicts persist across Africa. Conservation failure and poverty are intrinsically associated (Hauenstein et al. 2019; Oldekop et al. 2016), making conservation in Africa specially challenging (Fig 6). Poverty drives millions of people to illegally exploit protected areas (Wilson and Primack 2019).

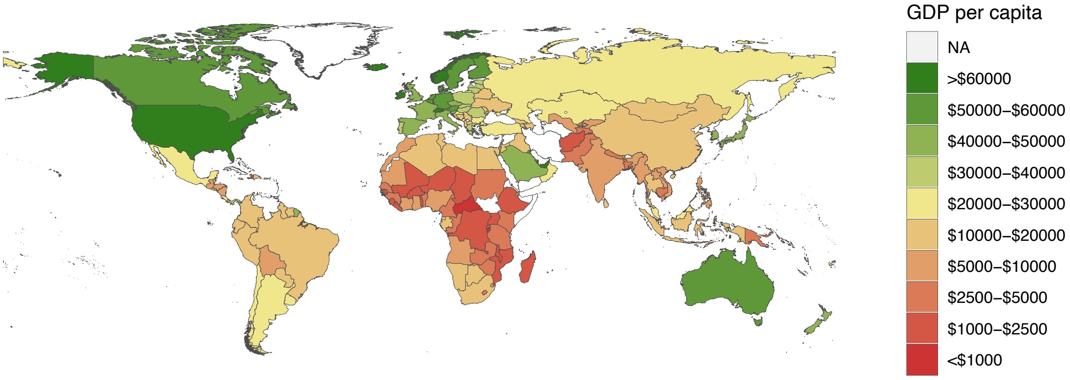


Figure 7: GDP per capita globally (World Bank, 2020). Africa is the continent with the lowest GDP per capita.

Structure of the thesis

In chapter 1 – **WEGE: A new metric for ranking locations for biodiversity conservation**, which has been published in Diversity and Distributions, I propose a new metric that is able to rank the biodiversity value of each area and may help to objectively identify Key Biodiversity Areas (KBA) in order to contribute to area prioritisation. In this paper we used IUCN data for all amphibians, mammals and birds to calculate in a 100km x 100km global grid which cells would trigger KBA status. We then calculated weighted endemism, the EDGE score, evolutionary distinctiveness, extinction risk and our new proposed metric: WEGE index, for each grid. We concluded that our proposed metric was better at ranking areas according to the KBA guidelines. I created an R package that calculates the WEGE index for any given area. https://cran.r-project.org/web/packages/WEGE/WEGE.pdf

In chapter 2 – **Mapping Africa's biodiversity: More of the same is just not good enough**, published in Systematic Biology, unveils a bias on biodiversity sampling in Africa based on analyses of vertebrate records from GBIF. I show that scientists tend to return to areas previously sampled rather than visiting new areas, and that it appears to be the existence of the knowledge itself that causes the scientists to return to these places. To do this I calculated the probability of visiting an area according to its inventory completeness using a Logistic model. Additionally, I predicted the required time to sample 90% of Africa, using survival analyses: where we treated unsampled cells as 'alive' and sampled cells as 'dead' and treated all cells still not visited in 2019 as censored. In this second analyses I found that previous sampling has been strongly influencing the resampling of areas, attracting repeated visits and that it may take between 172 and 274 years, depending on the taxonomic group, to achieve at least one sampling event per grid cell in the entire continent. Just one visit will however not be enough: in order to record ≥50% of the current diversity, it will require at least 12 sampling events for amphibians, 13 for mammals and 27 for birds.

In chapter 3 – **Shedding light on a biodiversity dark spot: Checklist of amphibians and reptiles of Pemba, Cabo Delgado province, Mozambique**, I compile all records from almost 20 years of amphibians and reptiles collected in Pemba, Northern Mozambique.

The manuscript has been submitted to the journal of Herpetological Conservation and Biology. In this study I report 53 species (19 amphibians and 34 reptiles), of which more than half had never previously been recorded from the area. Northern Mozambique is one of the least sampled areas in Sub-Saharan Africa and the results confirm the large under-sampling of the area and suggest that many other taxa remain unobserved and thereby lacking any form of active protection. I expect this data to be important for classifying and understanding species distributions in Africa.

In chapter 4 – **Endemism patterns are scale dependent**, which was published in Nature Communications. This is the only chapter in my thesis where I did not lead. My contribution in this study was, (1) together with the other co-authors, delineating the idea of the study, (2) during the analysis, contribute with the amphibian dataset and R code that was later included in the package bioregion, and (3) contribute to the editing of the manuscript. In this paper we showed that weighted endemism (WE) and phylogenetic endemism (PE) are dependent on grain sizes, spatial extents and taxonomic treatments. Variations in taxonomic opinions—whether species are treated by systematic 'lumping' or 'splitting'—can profoundly affect the allocation of WE hotspots. Global patterns of PE can provide insights into complex evolutionary processes but this congruence is lost at the continental to country extents. These findings are explained by environmental heterogeneity at coarser grains, and to a far lesser extent at finer resolutions. Regardless of scale, we found that widespread deficits of protection for endemism hotspots. Our study presents a framework for assessing areas for conservation that are robust to assumptions on taxonomy, spatial grain and extent.

CHAPTER 1

WEGE: A NEW METRIC FOR RANKING LOCATIONS FOR BIODIVERSITY CONSERVATION

Abstract

Aim

Effective policy making for biological conservation requires the identification and ranking of the most important areas for protection or management. One of the most frequently used systems for selecting priority areas is Key Biodiversity Areas (hereafter KBAs), developed by the International Union for Conservation of Nature (IUCN). However, KBAs cannot be used to rank areas, potentially limiting their use when limited funding is available. To tackle this shortcoming and facilitate spatial prioritization, here we develop and validate the "WEGE index" (Weighted Endemism including Global Endangerment index), consisting of an adaptation of the EDGE score (Evolutionarily Distinct and Globally Endangered). WEGE allows the ranking of any set of locations according to the KBA guidelines and on a continuous scale.

Location

Global.

Methods

We calculated the EDGE score, Weighted Endemism, Evolutionary distinctiveness, Extinction risk and our newly developed WEGE index for all terrestrial species of amphibians, mammals and birds accessed by IUCN. We then compared the performance of each of those five indices at prioritizing areas according to the KBA guidelines.

Results

We found that for all taxa surveyed, WEGE was consistently better at identifying areas that trigger KBA status.

Main conclusions

In our analyses, WEGE outperformed all other methods and metrics designed for similar purposes. It can serve as a robust evidence-based methodology to prioritize among otherwise equally qualified sites according to the KBA categories. WEGE can therefore support transparent, evidence-based and biologically meaningful decision-making for conservation priorities.

Farooq, H., Anderson, J., Belluardo, F., Nanvonamuquitxo, C., Bennett, D., Moat, J., Soares, A., Faurby, S. and Antonelli, A. WEGE: A new metric for ranking locations for biodiversity conservation. Divers Distrib. 2020; 00: 1– 11.
https://doi.org/10.1111/ddi.13148

Introduction

In order to protect biodiversity and promote conservation, the decision-making process should be based on scientific research and data, rather than expert judgement (Sutherland et al., 2004). Threats to biodiversity, such as habitat destruction, overexploitation, and invasive species, have the potential to completely extirpate species at local scales (Biofund, 2018; Mucova et al., 2018). Therefore, in recent years, instead of focusing on large extensions of low biodiversity land, there has been an increased awareness of the importance of protecting individual sites of high biological value (Butchart et al., 2012). Such shift of focus may ultimately determine whether species will survive or go extinct. Thus, conservation planning should not only encompass the concepts of global conservation prioritization (Myers et al., 2000), but also include a more local-scale approach more amenable to policymaking at national and finer political levels.

The Global Standards for the Identification of Key Biodiversity Areas (KBAs) is an attempt to gather a consensus of sites that contribute significantly to the global persistence of biodiversity (IUCN 2004). The criteria and methodology for identifying KBAs was created by the IUCN World Commission on Protected Areas (IUCN, 2016). KBAs can vary considerably in size, and the criteria aim to address aspects of biodiversity operating from regional to relatively local scales. The categorization of areas is based on criteria such as presence and proportional inclusion of threatened species and ecosystems, species' distribution ranges, ecological integrity and irreplaceability. Using these criteria, together with discussions on the manageability of areas as biodiversity units, the KBA process assesses areas through a binary system, where a particular area will either trigger KBA status or not. However, indices that directly measure biodiversity, such as species richness (SR: Colwell 2009), phylogenetic diversity (PD: Faith, 1992), weighted endemism (WE: Crisp et al., 2001) and phylogenetic endemism (PE: Rosauer et al., 2009) are, surprisingly, not included in the KBA methodology.

Although the four metrics cited above are useful in conservation, three of them fail to incorporate information on the threat status of the constituent species – the IUCN's Red List Assessment parameter. The exception is the Evolutionarily Distinct and Globally

Endangered (EDGE) score (Isaac et al., 2007), which combines one biodiversity index – Evolutionary Distinctiveness (ED) – with the threat category of species.

The EDGE score combines ED with values for species' extinction risk in order to generate a list of species that are both evolutionarily distinct and globally endangered ('EDGE species'). The EDGE score is thus tailored to rank species rather than locations. Location scores may be computed as the sum of EDGE scores for all species at a site (Safi et al., 2013). However, this is not guaranteed to maximize conservation importance of individual sites, since the presence of widespread, Critically Endangered (CR) species produces higher EDGE scores than a Vulnerable (VU) or Endangered (EN) micro-endemic restricted to very few sites, which could rapidly go extinct if those sites are damaged. One example is the Little Brown Bat (*Myotis lucifugus*), which occurs across a large portion of the North American continent, but is nevertheless considered an endangered species due to the invasive fungus *Pseudogymnoascus destructans* causing white-nose syndrome with very high mortality in the species (Solari 2018).

To tackle the shortcomings described above, in this study we propose the WEGE index (Weighted Endemism and Globally Endangered), a metric designed to rank any set of locations by their biodiversity importance. WEGE is an adaptation of the EDGE score (Mooers et al., 2008), but instead of incorporating the phylogenetic component it uses an endemism score.

We compared the ability of WEGE in ranking areas that trigger KBA status with four other metrics that can be used for similar purposes: WE, EDGE, ED and extinction risk (ER). Since the KBA methodology weights all species equally, irrespective of their evolutionary uniqueness, we did not include PD and PE in our benchmarking. We focused on the three vertebrate groups (amphibians, mammals and birds) where global range maps are available for all species. By using spatially coarse grids of 100 x 100km and 20 x 20km we compared the ability of WEGE in ranking areas that trigger KBA status with four other metrics that can be used for similar purposes: WE, EDGE, ED and extinction risk (ER). We found that WEGE outperforms all other metrics in the comparisons we made.

Methods

Using IUCN's range maps for all assessed terrestrial amphibians (6,615 species), mammals (5,610) and birds (10,549) (2019-3: IUCN 2019) we calculated EDGE, ED, ER, WE, WEGE and cells that trigger KBA status for criteria A1a, A1b, A1e and B1 in a global scale in resolutions of c. 100 x100 km (10,000 km^2) and c. 20 x 20 km (400 km^2) grids on a Berhmann projection map. We note that the sizes of the cells are area-true but involve some modifications of the shapes. The actual size is only 100 x100 km at 30 degrees North or South.

The species composition of each grid cell was obtained by intersecting IUCN range maps that met the following conditions: extant or probably extant, native or reintroduced, and resident or using the area as breeding site. For the 20 x 20 km grids we only assessed the indices and KBA status of the top left corner cell of a disaggregated 100 x100 km grid map, although all cells were included in the calculation of range size. This was done to reduce over-estimation of the number of KBAs due to neighbouring KBA-triggering cells.

Key Biodiversity Areas

The Global Standard for the Identification of KBAs (IUCN, 2016) has five main criteria: A. Threatened biodiversity; B. Geographically restricted biodiversity; C. Ecological integrity; D. Biological processes; and E. Irreplaceability through quantitative analysis. Of these, only criteria A and B can be applied to datasets consisting only of species range maps (Table 1).

Biodiversity indices

Table 1: Criteria and conditions used to produce the world KBA-gridded map

Criteria	Conditions in the KBA guidelines	Coded conditions to trigger KBA status
A1a)	The site regularly holds ≥0.5% of the global population size AND ≥5 reproductive units of a CR or EN species.	If at least one CR or EN species is present and has a range of 200 grid cells or fewer.
A1b)	The site regularly holds ≥1% of the global population size AND ≥10 reproductive units of a VU species.	If at least one VU species is present and has a range of 100 grid cells or fewer.
A1e)	Effectively the entire global population size of a CR or EN species.	If any CR or EN species with a range of 1 grid cell is present.
B1	The site regularly holds ≥10% of the global population size AND ≥10 reproductive units of a species.	If any species with a range of 10 grid cells or fewer is present.

To calculate WE, ED, ER and EDGE values for each grid cell, we summed the values of each metric for every species in a cell. To calculate weighted endemism (WE) for each species, we used one divided by the number of grid cells where a species occurs. The evolutionary distinctness (ED) scores were obtained from www.edgeofexistence.org (EDGE of Existence, 2019) and ER from (Davis et al., 2018). To accommodate taxonomic mismatching, we manually searched for synonyms. Failing to find them, we used the ED median value of the genus or family in case of missing the former. To calculate EDGE, we used the following formula from Mooers et al. (2008):

$$\mathrm{EDGE} = \ln(\mathrm{ED} * \mathrm{ER})$$

In contrast, the WEGE index uses weighted endemism (WE) instead of evolutionary distinctness (ED). Similar to EDGE, WEGE incorporates the probability of extinction (ER).

To calculate WEGE, we applied the formula:

$$\mathrm{WEGE} = \sum_{\mathrm{i}=1}^{\mathrm{SR}} \sqrt{\mathrm{WE_i}} * \mathrm{ER_i}$$

We calculated the WEGE index for each site as the sum across all local species of the square root of the partial weighted endemism value (WE_i) for each species (i) multiplied by its probability of extinction value (ER_i). The square root transformation was chosen to improve the normality of the weighted endemism data.

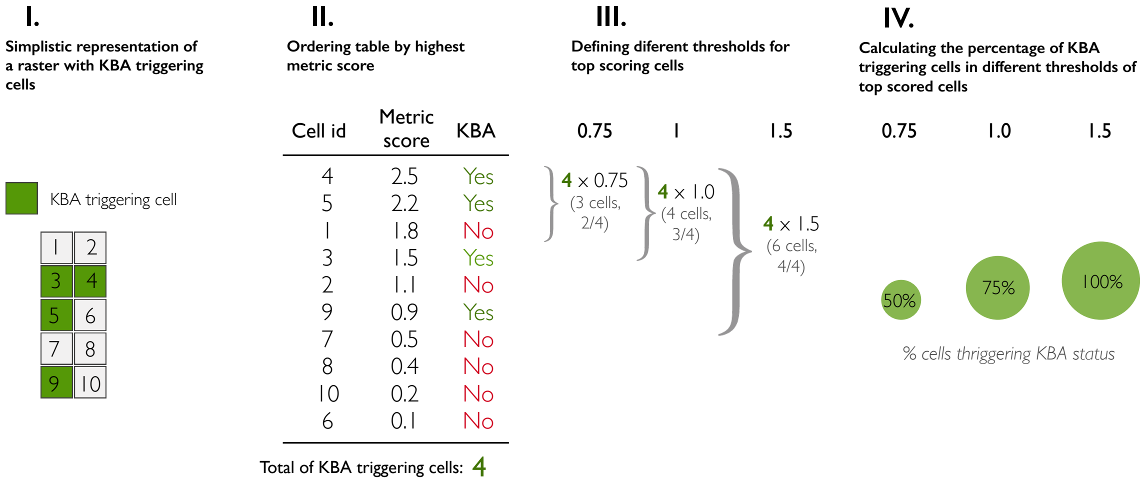
We used species range and conservation status in the WEGE formula since those parameters are also used in the KBA methodology, and since we aimed to rank areas that are also highlighted by the KBA methodology (rather than introducing other metrics such as phylogenetic diversity). In order to calculate WEGE values for a given spatial unit, we created a package in R – WEGE (https://CRAN.R-project.org/package=WEGE). We used the IUCN50 transformation for the ER as in (Davis et al., 2018), which scales the extinction risk over a 50-year period using the following extinction probabilities: LC = 0.0009, NT = 0.0071, VU = 0.0513, EN = 0.4276, CR = 0.9688. We assigned the extinction risk of the "DD" (Data Deficient) species with the VU probability following Bland et al. (2015).

To test our newly developed WEGE index (see above for equation), we built a global gridded map of cells that would trigger KBA status based on criterion A1a, A1b, A1e, or B1, hereafter referred to as "KBA-triggering cells", and calculated the percentage of cells triggering KBA status on the worldwide top-ranking cells for WE, ER, ED, EDGE and WEGE. We used hypothetical KBA triggered cells to avoid areas that trigger KBA status but are not yet considered KBAs, or areas that are considered KBAs due to other criteria not analysed in this study. We conducted separate analyses for amphibians, mammals, birds and one combining the three groups.

The KBA-triggering cells were the ones that conformed to the criteria A1a), A1b), A1e) and B1 of the KBA guidelines (Table 1). We assumed the presence of ≥10 reproductive units, as stated in the KBA guidelines, whenever a species range intersected with a grid cell. This extrapolation still holds for extremely demanding species such as the tiger (*Panthera tigris*), where a female tiger has a territory of ~20 km^2 (Carter et al., 2015).

Performance across indices

To test the performance of our newly developed index, we calculated the WEGE index, EDGE, ER, ED and WE globally and compared the percentage of KBA-triggering cells in the highest ranked cells of each metric. A score of 100% means perfect overlap between the highest-ranking cells and the KBA-triggering cells in their top scored cells, while a score of 50% means that in the top ranked cells for a particular index, only 50% are KBA-triggering cells. We first identified the number of KBA-triggering cells (K). Following this we identified how many of the KBA-triggering cells were contained within the K highest ranked cells for the other metrics. In this main analysis we used the threshold 1, and analysed the 1 × K highest ranked cells. We also tested five other thresholds (0.5, 0.75, 0.9, 1.2 and 1.5), where we e.g. in the 0.5 threshold tested how many KBA-triggering cells were found in the 0.5 × K highest ranked cells (Fig. 1). All thresholds produce identical patterns and the results can be found in the supplementary material. Our measure of performance thus means that the higher the percentage of KBA triggering cells in the top cells, the better the index performs at ranking KBAs.



Cell id	Metric score	KBA
4	2.5	Yes
5	2.2	Yes
1	1.8	No
3	1.5	Yes
2	1.1	No
9	0.9	Yes
7	0.5	No
8	0.4	No
10	0.2	No
6	0.1	No

Total of KBA triggering cells: 4

Figure 8: Schematic view of how metric performance is measured in this study. I. A simplistic representation of cells that trigger KBA status. II. A table ranking cells by their metric score. III. and IV. A representation of the calculation of the efficiency for different thresholds in ranking biodiversity according to the KBA criteria. The more overlap between potential KBA-triggering cells and top values for metrics, the higher the performance of a particular metric. The more overlap between KBA-triggering cells and top values for metrics, the higher the performance of a particular metric. The thresholds of 0.75, 1 and 1.5 represent the proportion of grid cells to be used in order to calculate the percentage of KBA triggering cells in the top scoring cells of each metric.

Results

In this study we used 6,615 species of amphibians, 10,549 of birds and 5,610 of mammals (Fig 2 A). Using IUCN's range polygons and a threshold of 50,000 km^2 for range-restricted taxa we grouped the different species in four main groups: (1) Wide-ranging threatened, (2) Range-restricted threatened, (3) Wide-ranging non-threatened, (4) Range-restricted non-threatened. We find that the vast majority of the threatened amphibians are range-restricted while a relatively larger fraction of threatened mammals or birds are wide-ranging. Due to different thresholds on the KBA criteria (Table 1), numerous threatened species were unable to trigger KBA status on our 10,000 km^2 grid cells.

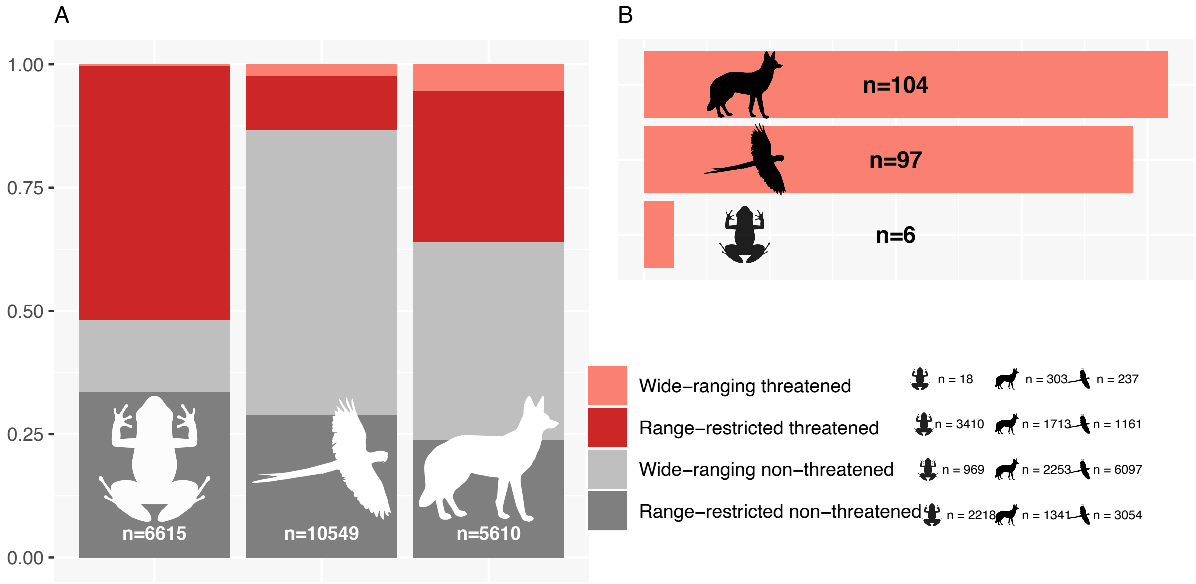


Figure 9: A. Species used in this study were classified into four groups by combinations of range size (range-restricted have ranges <50 000 km^2, wide-ranging have larger ranges) and threat status (CR, EN, VU and DD were classified as threatened, while NT (Near Threatened) and LC (Least Concern) were classified as non-threatened). B. Number of wide-ranging threatened species that trigger KBA status (Table 1). Most amphibians are range restricted, mostly threatened, while most birds are wide-ranging, and mammals have similar numbers of range-restricted and wide-ranging species. On a 10,000km^2 grid cell, only six wide-ranging threatened amphibians triggered KBA status, while this figure was ≈ 100 species for birds and mammals.

When assessing whether a grid cell would trigger KBA status, on a 100 x 100 km grid, we found 3,347 cells triggering KBA status for amphibians, 6,649 for mammals and 6,327 for birds. On a 20 x 20 km grid, we found 613 cells triggering KBA status for amphibians, 734 for mammals and 660 for birds. When combining all taxa, we found 9,228 cells on a 100 x 100 km grid, and 1,480 on a 20 x 20 km grid triggering KBA status, out of a pool of 17,283 terrestrial grid cells.

There was overall similarity in the spatial patterns between the three taxa but also some taxon- specific patterns. Most KBA-triggering cells were triggered by all three taxa (2,280), followed by grid cells triggered only by mammals (2,161) and only by birds (1,914) (Fig. 3).

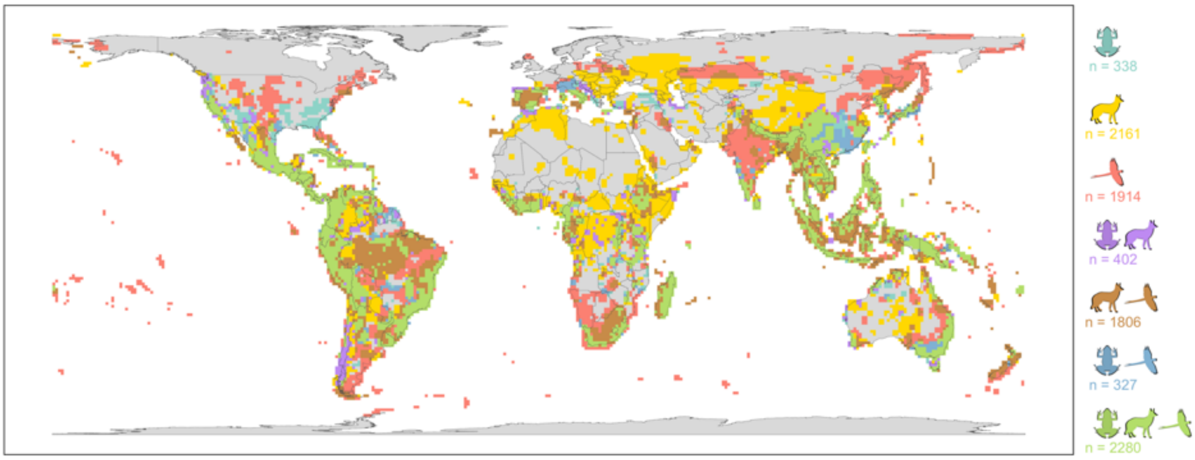


Figure 10: Plot showing KBA-triggering cells by different taxonomic group (amphibians, mammals, birds, and combinations of these; see legend for colour coding). Most KBA-triggering cells were triggered by all three taxa, followed by grid cells only triggered by mammals and then by birds. A small fraction of cells was triggered solely by amphibians.

We identified the number of cells that were classified as KBAs (K). We then looked at the K top-scoring cells for each metric and identified how many of these were KBAs (K values for the 100 x 100 km grid amounted to 3,347 cells for amphibians, 6,649 for mammals, 6,327 for birds and 9,228 for all taxa). By comparing the percentage of KBA-triggering cells on the top-scoring cells of each tested index, WEGE consistently outperforms ED, ER, EDGE and WE for all terrestrial amphibians, mammals and birds at both tested resolutions (Fig. 4). The second-best performer for all groups, as well as for the taxa combined, was WE, followed by ER (Fig. 4). The higher number of KBA triggering cells for mammals and birds compared to amphibians is explained by the existence of many wide-ranging threatened species in those two groups (Fig 2. B).

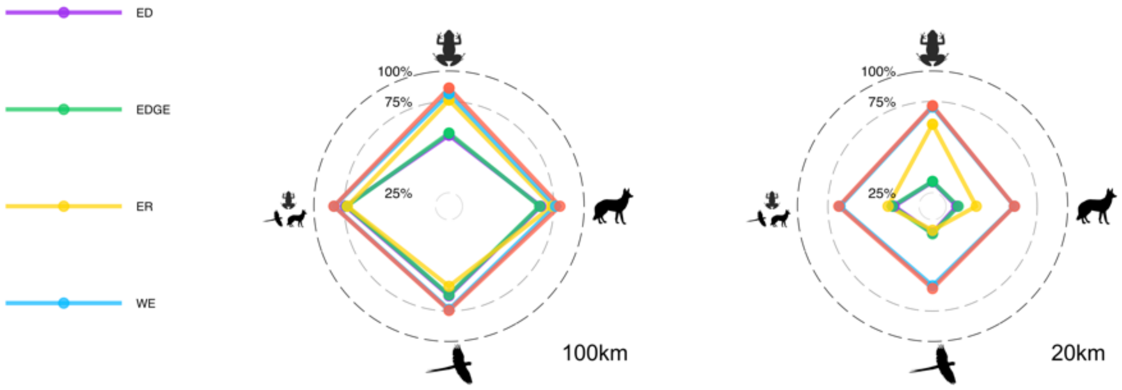


Figure 11: Web plot showing the percentage of presence of KBA-triggering cells in the top cells with resolutions of c. 100 x 100 km and c. 20 x 20 km, for ED, EDGE, ER, WE and WEGE. The ED and EDGE values were very similar, rendering the ED line almost indistinguishable (table of values can be found in the supplementary material). Our analyses show that WEGE outperforms all metrics consistently across all tested taxa and resolutions.

According to the global spatial distribution of the WEGE index, the most important WEGE hotspots of amphibians are concentrated in the western United States, Central America, Andes, West Africa, Eastern Arc Mountains in East Africa, eastern Madagascar, China and south-eastern Australia. The most important WEGE hotspots for mammals are found in Central America, Andes, Atlantic Forest, West and Central Africa, Madagascar and Southeast Asia. For birds, the WEGE hotspots comprise Central America, Andes, most of Brazil, India, Southeast Asia, New Zealand and most of sub-Saharan Africa (Fig. 5).

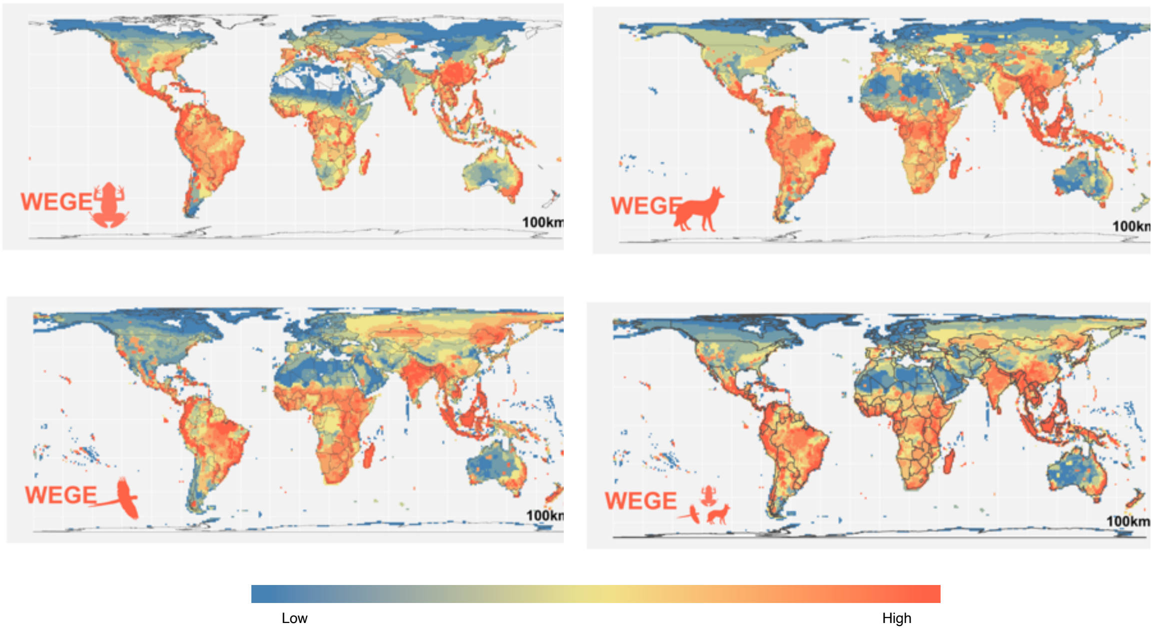


Figure 12: Global WEGE maps for all terrestrial amphibians, mammals and birds. Overall, WEGE hotspots were constant through taxa, where, Central America, the Andes, West Africa, East Africa, Eastern Madagascar, China and Eastern Australia registered higher values. For visualization purposes, the values in the cells were divided in 10 quantiles, where blue represent low quantile values, yellow medium, and red high. This means that all the same number of cells are shown in the same color for all four metrics irrespective of the skewness of the distribution of values for the metric.

Regional comparisons

To assess the behaviour of the WEGE index in relation to the other tested indices, we selected three different regions where differences between the indices are more visible. Global maps of all indices can be found in the supplementary material.

As the first example, we illustrate the pattern in amphibians in Africa (Fig. 6). Here, EDGE and ED were unable to highlight important grid cells for amphibians, such as the Eastern Arc Mountains, unlike WE, ER and WEGE. Amphibians in this hotspot are often both under threat and restricted to mountain tops, thus driving WE, ER and WEGE scores higher.

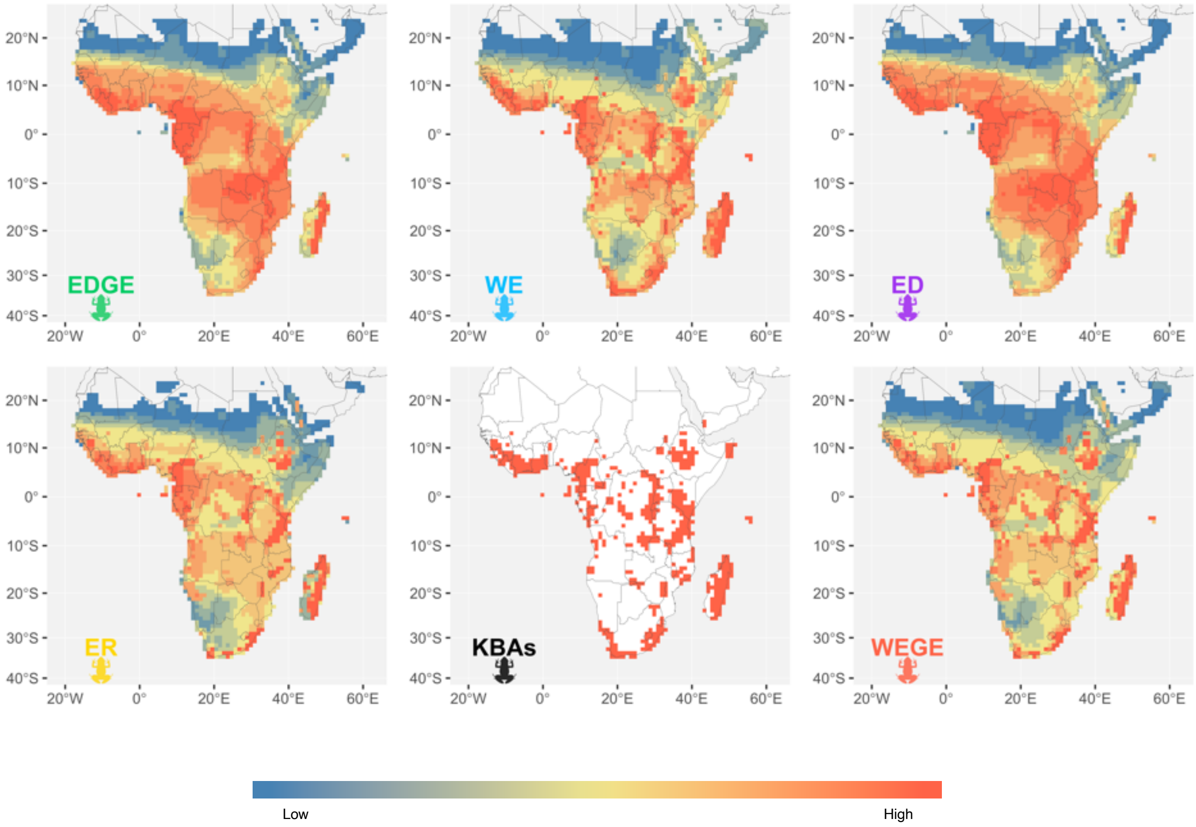


Figure 13: Plot illustrating the prioritization of cells for the different indices tested in this study for amphibians in Africa. EDGE and ED created similar gradients and were unable to highlight particular cells in East and Central Africa. WE, ER and WEGE highlighted the Eastern Arc Mountains as high priority grid-cells. Percentage of KBA triggering cells in each top metric: WEGE – 82%, EDGE – 48%, WE – 85%, ER – 71% and ED – 48%. For visualization purposes, the values in the cells were divided in 10 quantiles, where blue represent low quantile values, yellow medium, and red high. This means that all the same number of cells are shown in the same colour for all four metrics irrespective of the skewness of the distribution of values for the metric.

As the second example, we illustrate the pattern in mammals in the Americas. Unlike amphibians in Africa, mammals in North America (Fig. 7) show an overlap of high evolutionary distinctiveness and narrow ranges, that stretches from Central America to the West of North America. However, these areas house fewer threatened species when compared to the East side of the continent. The WEGE index was able to use information from both WE and ER and score grid cells in both West and East of the continent with high values.

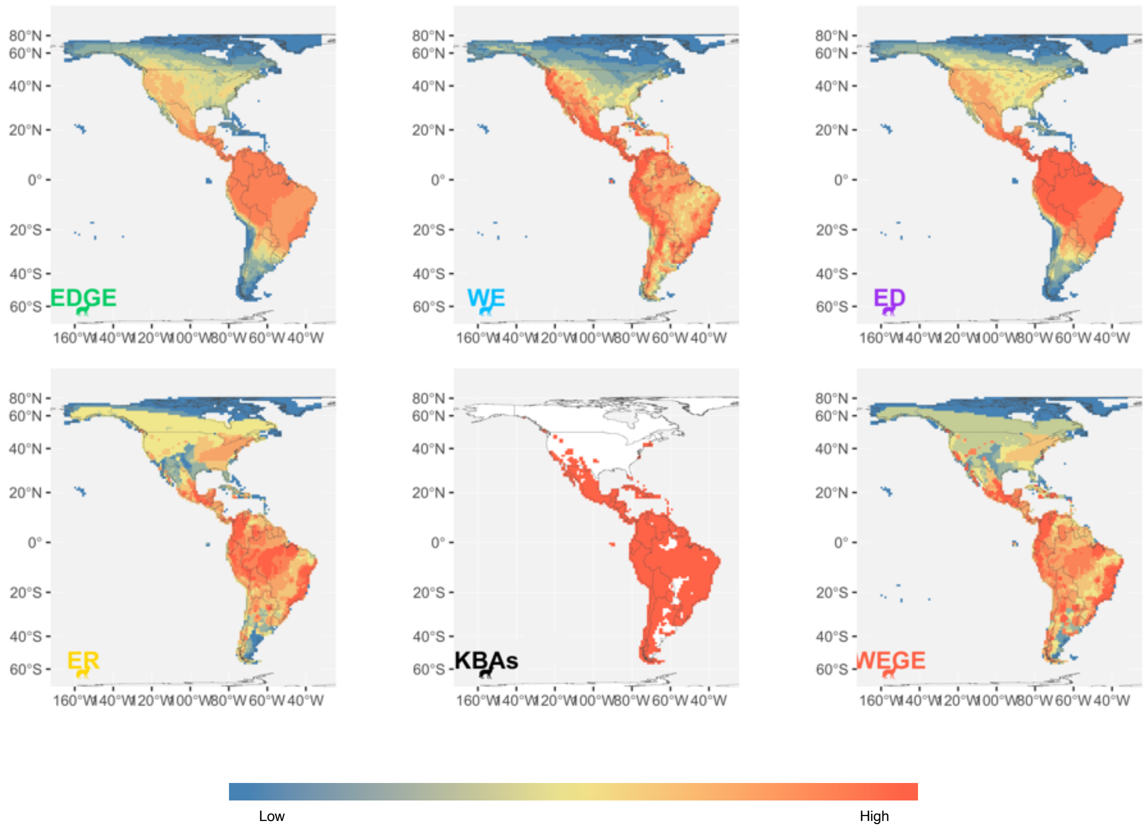


Figure 14: Plot illustrating the prioritization of cells for the different indices tested in this study for mammals in the Americas. KBAs were triggered primarily in the west but also in some eastern grids. EDGE and ED exhibit similar gradients and neither of them highlighted particular cells in the east. Both WE and ED scored eastern USA as a low priority area. In contrast, WEGE was able to highlight areas in both western and eastern North America, by also incorporating the ER information. Percentage of KBA triggering cells in each top metric: WEGE – 85%, EDGE – 74%, WE – 84%, ER – 76% and ED – 74%. For visualization purposes, the values in the cells were divided in 10 quantiles, where blue represent low quantile values, yellow medium, and red high. This means that all the same number of cells are shown in the same colour for all four metrics irrespective of the skewness of the distribution of values for the metric.

As the third and final example, we illustrate the pattern in birds in Australia. Most Australian birds have relatively narrow ranges particularly along the coast, but some of those areas particularly in the northwest are remote and the species occurring there are generally not threatened. Thus, WEGE ranked those cells lower and highlighted instead most of south-eastern and southwestern Australia as the most important areas, as well as a few grid cells in the north of Queensland which house threatened species (Fig. 8).

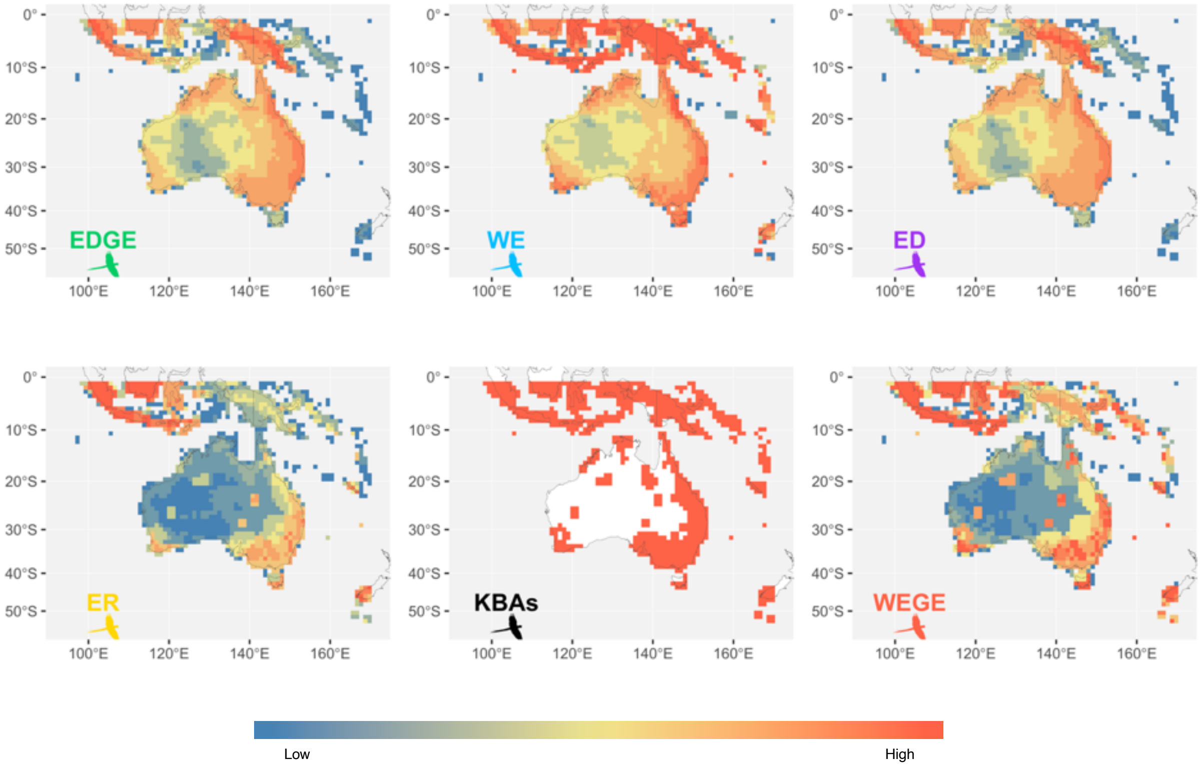


Figure 15: Plot illustrating the prioritization of cells for the different indices tested in this study for birds in Australia. EDGE and ED show similar gradients. The WEGE index highlights a few cells in the North of Australia, which are ranked low in terms of extinction risk but high on weighted endemism. Most of central and west Australia scored low on WEGE, despite WE ranking them high. Percentage of KBA triggering cells in each top metric: WEGE – 93%, EDGE – 70%, WE – 87%, ER – 88% and ED – 69%. For visualization purposes, the values in the cells were divided in 10 quantiles, where blue represent low quantile values, yellow medium, and red hig

Discussion

KBA-triggering cells

The IUCN's KBA system uses a set of guidelines to decide whether a particular site triggers conservation status, unlike biodiversity metrics which simply quantify different aspects of biodiversity and are therefore expected to weight sites differently. Out of the 9,228 KBA-triggering cells for the three taxonomic groups surveyed, most were triggered by mammals (6,649) and birds (6,327) and fewer by amphibians (3,347) (Fig. 2). We interpret this as a result of the higher numbers of wide-ranging threatened species of mammals (104) and birds (97) but whose ranges are nevertheless small enough to trigger KBA status (Table 1) when compared to amphibians (6) (Fig. 3). The Western Gorilla (*Gorilla gorilla*) and the Saiga (*Saiga tatarica*) are examples of such, by having a range of 104 and 106 grid cells and being CR species.

Despite spatially coarse grids having implications in our interpretations and should be reviewed in light of practical applications of WEGE, that would tend to use much higher resolution data inputs, the fact that most KBA cells were triggered by all three taxa (Fig. 2), suggests that important biodiversity areas tend to overlap between different vertebrate groups rather than being taxon-specific. Finally, since more than half of the terrestrial cells were triggered by at least one taxon, this may represent an impractical implementation of the KBA methodology, which may have also been overestimated due to the coarse scale resolution used in this study.

The WEGE index

WEGE is capable of ranking locations on a continuous scale and rank higher the areas that according to current criteria trigger KBA status. The WEGE index adds the component of conservation status of each species to the WE index in order to combine a conservation scoring of each species with a measure of the relative importance of the site in question for each species. This could also be achieved by combining a conservation score while incorporating evolutionary history such as PE rather than WE, but since KBAs by design weight all species equally (irrespective of their evolutionary uniqueness) we chose to select a measure with the same lack of taxonomic weighting. The main rationale for incorporating WE in the EDGE formula to create WEGE was to obtain an index in line with the widely used IUCN KBA framework, which none of the other metrics are able to achieve – WE, ED, ER, WE or EDGE.

Applicability of WEGE

WEGE can be used to rank areas or as a complementary tool in the KBA assessment process – by weighting the importance of biodiversity of sites so that they can be ranked objectively according to their biodiversity importance. The ranking of locations can bring important advantages when prioritizing efforts with limited resources. IUCN's criteria lack this aspect by attributing a binary system where one particular site either triggers KBA status or not. A policymaker may then focus conservation resource based on insufficiently supported or subjective decisions. WEGE therefore outperforms KBAs when dealing with the A1 and B criteria, by ranking sites within the same category and thereby facilitating the decision-making process with greater objectivity and transparency.

KBA sites which are triggered either by a single threatened or a single range-restricted species will change their status if re-assessed after the species goes extinct, becomes non-threatened or expands its range considerably. Consequently, lower scoring WEGE sites have higher odds of losing their KBA status. One example that illustrates this scenario could involve the Bramble Cay Melomys (*Melomys rubicola*), a rodent restricted to the island of Bramble Cay which was recently declared extinct (Waller et al. 2017). This species by itself would meet the requirements for the island to trigger KBA status, regardless of its IUCN status, since it was an accepted species entirely confined to a single small island, hence triggering criteria A1. e) (the whole range of a species in a site) and B1 (10% of the range of any species inside the site). Despite the fact the island may have the minimum requirements to trigger KBA status, the island would get a low WEGE value when compared to WEGE values of areas with numerous triggering species. Using only three vertebrate groups and 100x100 km grids, most grid-cells are classified as KBAs based on presence of endangered or range restricted species of at least one of these groups. If the KBA framework is expanded to large species-rich groups like angiosperms, we expect that the number of KBAs would further increase to a point where targeted conservation may not be realistically possible on all KBAs. Using WEGE, we are however able to quantify the importance of different KBAs, enabling a better prioritization of the limited available conservation funding.

Comparison between WEGE and other indices

For all tested taxa (amphibians, mammals and birds) and for both spatial resolutions (100 x 100 and 20 x 20 km), the WEGE index outperformed WE, ED, ER and EDGE by detecting a higher percentage of KBA-triggering cells among the its top-ranking cells. Under such test conditions, we therefore find that WEGE ranks biodiversity according to the KBA criteria better than all other tested metrics. The second-best metric was WE, followed by ER in the case of amphibians and mammals but not in birds. We interpret the low ER scores as due to the fact that IUCN assessments of amphibians and mammals (excluding bats) tend to reflect their range to a greater extent than in birds (Fig. 2A); therefore, threatened amphibians and mammals tend to trigger KBA status to a greater extent in comparison to birds.

EDGE and WEGE combine two clearly distinct metrics. EDGE combines phylogenetic information and threat status to highlight important species for biodiversity conservation (Isaac et al. 2007), while WEGE makes use of species distributions and IUCN conservation status, as does the IUCN's KBA criteria. Although range and threat status are not completely independent, since range size is one of the criteria for the IUCN status assessment, when we analysed species' ranges and threat statuses from the IUCN's polygons, we saw that 50% of all species of amphibians, 49% of mammals and 29% of birds are both range restricted and non-threatened. Therefore, by explicitly using range in its calculation, WEGE is better able to incorporate the high conservation value of species such as the widespread but endangered Saker Falcon (*Falco cherrug*), a species that occurs in a wide range across the Palearctic region from eastern Europe to western China (BirdLife International 2017), or species that are highly restricted but currently considered of Least Concern, such as the Broadley's Writhing Skink (*Mochlus lanceolatus*) (Conradie et al. 2019). As per our tests, we obtained a better measure combining WE and ER in one metric than solely relying on range or threat status to rank the biodiversity value of a grid cell in accordance to the KBA criteria.

Even though both EDGE and KBAs are aimed at the preservation of biodiversity (Isaac et al. 2007; IUCN 2016), according to our results they prioritize partially different areas. The use of EDGE scores to rank sites is only expected to be efficient when the threats are plausibly mitigated by the protection of a site. However, this may not always be feasible. Threatened species, for instance, may be very widespread under two different scenarios. Some species live in very low population densities, such as tigers (*Panthera tigris*), meaning that the protection of individual small areas may have little effect. Other species may be threatened by causes that are non-geographic in nature, such as in the case of the Tasmanian Devil (*Sarcophilus harrisii)* which is currently endangered by a sexually transmitted disease (Hawkins et al. 2006). Although the species in these two examples are both considered threatened, no single site will be as important for their protection as a site containing the majority of the range of a less threatened, micro-endemic species. This would be the case, for example, for the Near Threatened Mount Mabu Pygmy Chameleon (*Rhampholeon maspictus*), which has a predicted range of 108 km^2 (Tolley et al. 2019).

The importance of prioritisation

The importance of prioritisation among KBAs for better conservation policy has been previously highlighted (Ferrier et al. 2000; Plumptre et al. 2019; Pressey et al. 1994; Smith et al. 2019). Multiple metrics, including protection status, available funding, ‘irreplaceability’ (Plumptre et al. 2019) and systematic conservation planning (Smith et al. 2019) have been proposed to support the ranking of areas. This methodology, although providing a hierarchy among KBAs, still clusters them in different categories, rather than scoring individual sites as is the purpose of WEGE. In systematic conservation planning, practitioners must choose which conservation features should be used to represent biodiversity (Smith et al. 2019). WEGE represents a simple metric that encapsulates the biodiversity importance of a particular site, highlighting the same areas as the KBA criteria while adding the advantages of continuous scale. Therefore, WEGE may also be used as a feature in systematic conservation planning.

Limitations and challenges of the WEGE index

Despite ranking KBA-triggering cells more effectively than other metrics, WEGE uses a simpler methodology by employing only two (A1 and B1) out of seven criteria, since it only uses georeferenced species lists to rank KBAs. We therefore caveat that other criteria like Ecological Integrity (criteria C), Biological Processes (Criteria D) and Irreplaceability Through Quantitative Analysis (Criteria E) are not yet possible to include into our approach. Additional methodological developments would therefore be needed to fully develop a continuous metric fulfilling the full set of goals for the KBAs.

Proposing a particular site as a KBA requires an analysis of the manageability of the site with regard to its physical attributes, such as forest cover or the presence of rivers, and anthropogenic factors, such as roads and existence of human settlements, among other tasks. The WEGE index is not aimed at replacing this process, which we believe is of crucial importance and should be done case by case while involving local authorities and communities. The aim of the WEGE index is to highlight and rank sites, which should be further scrutinized at a local and practical level, as in the KBA process, or to rank already existing KBAs for more effective allocation of resources to maximize biodiversity outcomes.

The selection of sites as KBAs is important in multiple ways, including for conservation planning support and priority-setting at national and regional levels (IUCN 2016). The use of the WEGE index, allowing the ranking of KBAs, is expected to further support a transparent ranking of sites for evaluating conservation priorities.

Supporting Information

Methods used for calculating indices, r packages used, KBA guidelines and raw data (Appendix S1) are available online at Dryad: https://doi.org/10.5061/dryad.jsxksn06s

References

Anderson, S. (1994). Area and endemism. The quarterly review of biology, 69(4), 451-471.

Ball, I., & Possingham, H. (2000). MARXAN (V1. 8.2). Marine Reserve Design Using Spatially Explicit Annealing, a Manual.

Biofund. (2018). Retrieved from http://www.biofund.org.mz/

Bland, L. M., Collen, B., Orme, C. D. L., & Bielby, J. (2015). Predicting the conservation status of data-deficient species. Conservation Biology, 29(1), 250-259.

Butchart, S. H., et al. (2012). Protecting important sites for biodiversity contributes to meeting global conservation targets. PloS one, 7(3), e32529.

Carter, N., Levin, S., Barlow, A., & Grimm, V. (2015). Modeling tiger population and territory dynamics using an agent-based approach. Ecological Modelling, 312, 347-362.

Collette, B., et al. (2011). Thunnus thynnus. In IUCN 2012. IUCN Red List of Threatened Species.

Conradie, W., Verburgt, L., Tolley, K. A., Alexander, G. J., Farooq, H., Raimundo, A., & Sardinha, C. I. V. (2019). Mochlus lanceolatus. The IUCN Red List of Threatened Species 2019: e.T14668253A14668316. doi:https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T14668253A14668316.en.

Crisp, M. D., Laffan, S., Linder, H. P., & Monro, A. (2001). Endemism in the Australian flora. Journal of Biogeography, 28(2), 183-198.

Davis, M., Faurby, S., & Svenning, J.-C. (2018). Mammal diversity will take millions of years to recover from the current biodiversity crisis. Proceedings of the National Academy of Sciences, 115(44), 11262-11267.

EDGE of Existence. (2019). Retrieved from www.edgeofexistence.org

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biological Conservation, 61(1), 1-10.

Ferrier, S., Pressey, R. L., & Barrett, T. W. (2000). A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for further refinement. Biological Conservation, 93(3), 303-325.

Hawkins, C. E., et al. (2006). Emerging disease and population decline of an island endemic, the Tasmanian devil Sarcophilus harrisii. Biological Conservation, 131(2), 307-324.

Isaac, N. J., Turvey, S. T., Collen, B., Waterman, C., & Baillie, J. E. (2007). Mammals on the EDGE: conservation priorities based on threat and phylogeny. PloS one, 2(3), e296.

IUCN. (2016). A Global Standard for the Identification of Key Biodiversity Areas. Retrieved from Gland, Switzerland:

Mooers, A. Ø., Faith, D. P., & Maddison, W. P. (2008). Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. PloS one, 3(11), e3700.

Mucova, S. A. R., Leal Filho, W., Azeiteiro, U. M., & Pereira, M. J. (2018). Assessment of land use and land cover changes from 1979 to 2017 and biodiversity & land management approach in Quirimbas National Park, Northern Mozambique, Africa. Global Ecology Conservation, 16, e00447.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403(6772), 853.

Plumptre, A. J., et al. (2019). Conservation of vertebrates and plants in Uganda: Identifying Key Biodiversity Areas and other sites of national importance. Conservation Science and Practice, 1(2), e7.

Pressey, R., Johnson, I., & Wilson, P. (1994). Shades of irreplaceability: towards a measure of the contribution of sites to a reservation goal. Biodiversity & Conservation, 3(3), 242-262.

Purvis, A., Gittleman, J. L., Cowlishaw, G., & Mace, G. M. (2000). Predicting extinction risk in declining species. Proceedings of the Royal Society of London. Series B: Biological Sciences, 267(1456), 1947-1952.

Roberts, C. M., et al. (2002). Marine biodiversity hotspots and conservation priorities for tropical reefs. Science, 295(5558), 1280-1284.

Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C., & Cook, L. G. (2009). Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. Molecular ecology, 18(19), 4061-4072.

Safi, K., Armour-Marshall, K., Baillie, J. E., & Isaac, N. J. (2013). Global patterns of evolutionary distinct and globally endangered amphibians and mammals. PloS one, 8(5), e63582.

Slatyer, C., Rosauer, D., & Lemckert, F. (2007). An assessment of endemism and species richness patterns in the Australian Anura. Journal of Biogeography, 34(4), 583-596.

Smith, R. J., et al. (2019). Synergies between the key biodiversity area and systematic conservation planning approaches. Conservation Letters, 12(1), e12625.

Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. Trends in ecology evolution, 19(6), 305-308.

Tolley, K. A., Conradie, W., Farooq, H., Verburgt, L., Alexander, G. J., Sardinha, C. I. V., Raimundo, A., & Bayliss, J. (2019). Rhampholeon maspictus . The IUCN Red List of Threatened Species 2019: e.T61364057A149767110. doi:https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T61364057A149767110.en

Waller, N. L., Gynther, I. C., Freeman, A. B., Lavery, T. H., & Leung, L. K.-P. (2017). The Bramble Cay melomys Melomys rubicola (Rodentia: Muridae): a first mammalian extinction caused by human-induced climate change? Wildlife Research, 44(1), 9-21.

DATA AVAILABILITY STATEMENT

Data are available from the authors at Dryad: https://doi.org/10.5061/dryad.jsxksn06s

Author contributions: H.F, S.F and A.A conceptualized the project with input from J.A, F.B, C.N and J.M. H.F gathered the data and did the formal analysis with feedback input from S.F and D.B. H.F wrote the original draft H.F, constructively reviewed by S.F, A.A, J.A, F.B, D.B, J.M and A.S.

Supplementary material

Table SI 1: IUCN's Global Standard for the Identification of Key Biodiversity Areas Criteria Summary

KBA's Main criteria	Sub-criteria	Fits into article's scope	Criteria's target group
A. Threatened Biodiversity	A1. Threatened species	Yes	Proportion of the global population size of a species facing a high risk (CR, EN, VU) of extinction.
	A2. Threatened ecosystem types	No	Proportion of the global extent of an ecosystem type facing a high risk (CR, EN, VU) of collapse.
B. Geographically restricted biodiversity	B1: Individual geographically restricted species	Yes	Proportion of the global population size of a geographically restricted species.
	B2: Co-occurring geographically restricted species	Yes	Proportion of the global population size of multiple restricted-range species.
	B3: Geographically restricted assemblages	Yes	Sites that hold assemblages of species within a taxonomic group that are globally restricted
	B4: Geographically restricted ecosystem types	No	Proportion of the global extent of a geographically restricted ecosystem type
C. Ecological integrity		No	Sites that hold wholly intact ecological communities with supporting large-scale ecological processes.

D. Biological processes	D1: Demographic aggregations	Yes	Proportion of the global population size of a species during one or more life history stages or processes
	D2: Ecological refugia	Yes	Proportion of the global population size of a species during periods of environmental stress
	D3: Recruitment sources	Yes	Proportion of the global population size of a species is produced
E. Irreplaceability through quantitative analysis		No	Sites with high irreplaceability for the global persistence of biodiversity

The A1a) sub criteria states “*Site regularly holds ≥0.5% of the global population size AND ≥5 reproductive units of a CR or EN species*”;

The A1e) sub criteria states “*Site regularly holds effectively the entire global population size of a CR or EN species*”.

The B1 sub criteria states “*Site regularly holds ≥10% of the global population size AND ≥10 reproductive units of a species*”.

The B2 sub criteria states “*Site regularly holds ≥1% of the global population size of each of a number of restricted-range species in a taxonomic group, determined as either ≥2 species OR 0.02% of the global number of species in the taxonomic group, whichever is larger*.”

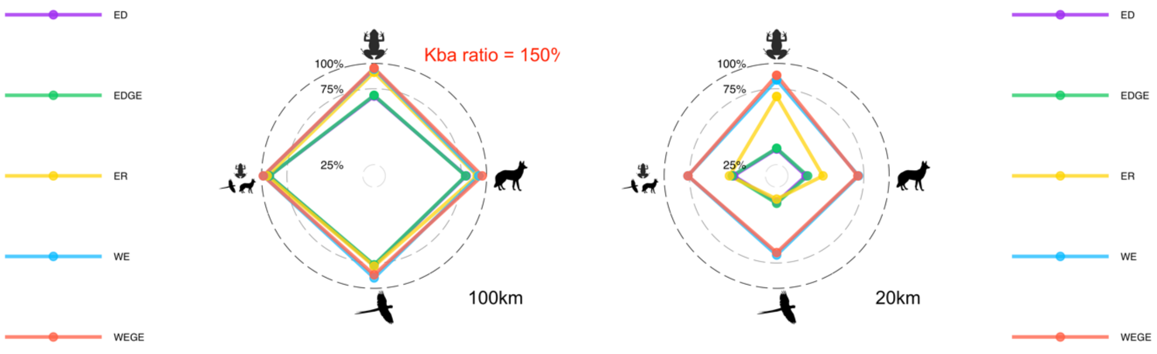


Figure S16: Web plot showing the percentage of presence of KBA-triggering cells in the top 1.5 x K cells with resolutions of c. 100 × 100 km and c. 20 × 20 km, for ED, EDGE, ER, WE and WEGE.

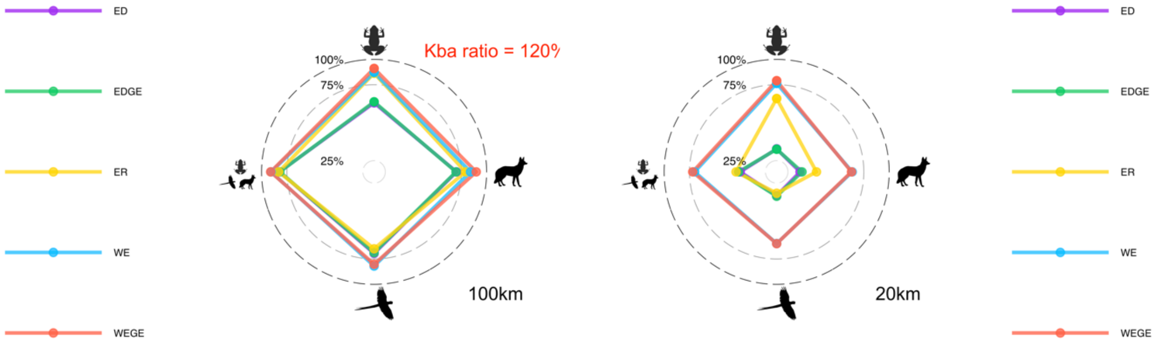


Figure S17: Web plot showing the percentage of presence of KBA-triggering cells in the top 1.2 x K cells with resolutions of c. 100 × 100 km and c. 20 × 20 km, for ED, EDGE, ER, WE and WEGE.

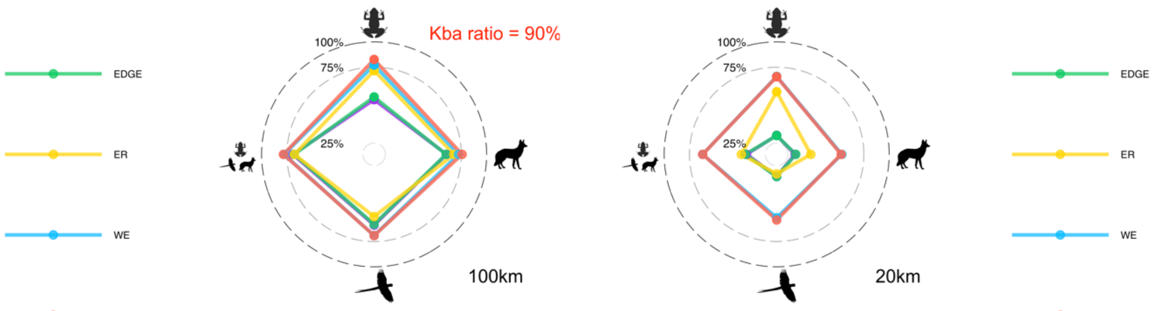


Figure 18: Web plot showing the percentage of presence of KBA-triggering cells in the top 0.9 x K cells with resolutions of c. 100 × 100 km and c. 20 × 20 km, for ED, EDGE, ER, WE and WEGE.

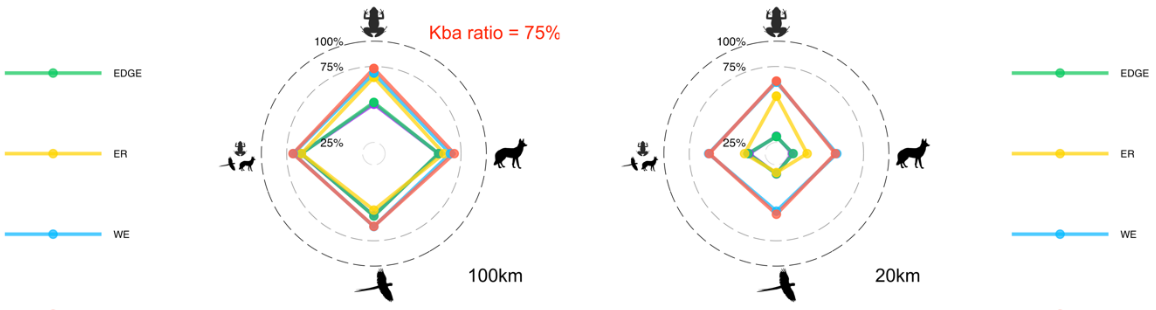


Figure 19: Web plot showing the percentage of presence of KBA-triggering cells in the top 0.75 x K cells with resolutions of c. 100 × 100 km and c. 20 × 20 km, for ED, EDGE, ER, WE and WEGE.

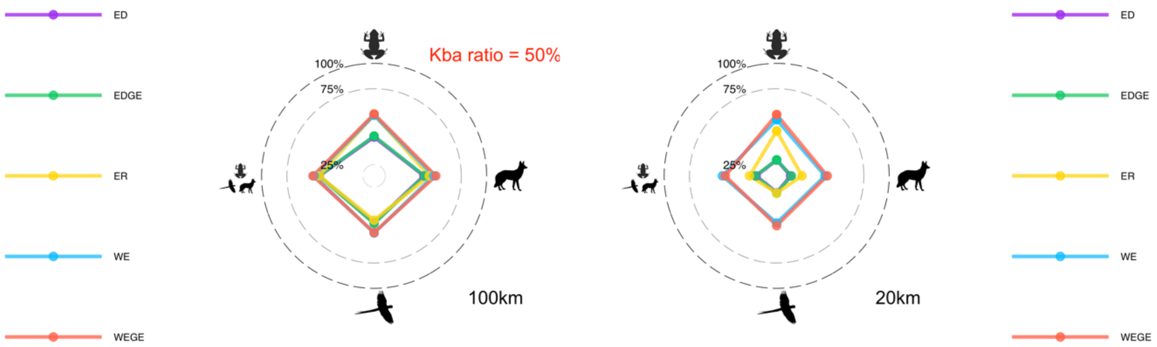


Figure 20: Web plot showing the percentage of presence of KBA-triggering cells in the top 0.5 x K cells with resolutions of c. 100 × 100 km and c. 20 × 20 km, for ED, EDGE, ER, WE and WEGE.

CHAPTER 2

MAPPING AFRICA'S BIODIVERSITY: MORE OF THE SAME IS JUST NOT GOOD ENOUGH

Abstract

Species distribution data are fundamental to the understanding of biodiversity patterns and processes. Yet, such data are strongly affected by sampling biases, mostly related to site accessibility. The understanding of these biases is therefore crucial in systematics, biogeography and conservation. Here we present a novel approach for quantifying sampling effort and its impact on biodiversity knowledge, focusing on Africa. In contrast to previous studies assessing sampling completeness (percentage of species recorded in relation to predicted), we investigate whether the lack of knowledge of a site attracts scientists to visit these areas and collect samples of species. We then estimate the time required to sample 90% of the continent under a Weibull distributed biodiversity sampling rate and the number of sampling events required to record ≥50% of the species. Using linear and spatial regression models, we show that previous sampling has been strongly influencing the resampling of areas, attracting repeated visits. This bias has existed for over two centuries, has increased in recent decades, and is most pronounced among mammals. It may take between 172 and 274 years, depending on the group, to achieve at least one sampling event per grid cell in the entire continent. Just one visit will however not be enough: in order to record ≥50% of the current diversity, it will require at least 12 sampling events for amphibians, 13 for mammals and 27 for birds. Our results demonstrate the importance of sampling areas that lack primary biodiversity data and the urgency with which this needs to be done. Current practice is insufficient to adequately classify and map African biodiversity; it can lead to incorrect conclusions being drawn from biogeographic analyses, and can result in misleading and self-reinforcing conservation priorities.

Farooq, H., Anderson, J., Soares, A., Antonelli, A. and Faurby, S. (2020). Mapping Africa's biodiversity: More of the same is just not good enough, *Systematic Biology*, , syaa090, https://doi.org/10.1093/sysbio/syaa090

Introduction

While the number of scientists, scientific organisations and publications are increasing worldwide (Stork & Astrin, 2014), our knowledge on the distribution of biodiversity – a cornerstone in our understanding of life on Earth – may not be expanding to the same extent (Bini et al., 2006; Boakes et al., 2010; Feeley & Silman, 2011; Stropp et al., 2016).

Georeferenced specimens in natural history collections and observations are fundamental for the classification and understanding of biodiversity patterns. Besides the increasing availability of observational data derived from citizen science initiatives, museum specimens are still the main source of information for taxonomic, systematic and ecological studies. (Shaffer et al., 1998; Graham et al., 2004). However, there remain important gaps of knowledge in the distribution of organisms – the "Wallacean shortfall" (Lomolino, 2004). Species distribution datasets are often strongly affected by temporal, spatial, and taxonomic biases (Meyer et al., 2016). Temporal biases can be influenced by intensive collecting periods or by seasonality (Ward, 2012). Spatial biases often relate to accessibility (Reddy & Dávalos, 2003), protected areas and particular habitats (Sánchez-Fernández et al., 2008) or climatic zones (Loiselle et al., 2008). Sampling biases are also known to be strongly affected by variables such as body size and taxonomic group (Schmidt-Lebuhn et al., 2013; Troudet et al., 2017). As an additional concern, there are differences between whether how well-studied groups are and how represented they are in taxonomic collections. One example is the reluctance of taking vouchers of supposedly "well-studied" groups such as birds (Bates et al., 2004; Schmitt et al., 2019)

Although it has been widely documented that certain parts of continents are visited and sampled more frequently than others (Meyer et al., 2015), the underlying causes for this unevenness may be attributed to several factors. These include language barriers (Harford, 2015), lack of basic resources, and poor infrastructure (Walker et al., 2006; Foster & Briceño-Garmendia, 2009; Beegle et al., 2016). Additional factors such as political regimes (Rydén et al., 2019), corruption (Mbaku, 2010; Bello-Schünemann & Moyer, 2018) dangerous tropical diseases (Hotez & Kamath, 2009; Amarasinghe et al., 2011; Bhatt et al., 2015), and expensive or burdensome permit requirements (Engel et al.,

2015) may further discourage work in particular countries and regions. Although many of these factors have been previously described in the literature and are well known to the systematic community, the influence of existing previous knowledge of the biodiversity of a site in attracting scientists remains unknown. Scientists may either preferentially visit areas that are more accessible or, alternatively, prefer to sample in well-known areas. Quantifying (Zizka et al., 2020) and creating awareness around sampling biases are crucial to the efficient implementation of conservation policies, and are essential to both scientists and decision makers.

Here we test whether the degree of previous knowledge of biodiversity within an area increases the likelihood that additional sampling will be done in the same cell. We then estimate the time and effort necessary to sample Africa's biodiversity under current practices. We perform our analyses on data from amphibians, mammals and birds, given their relatively high level of baseline knowledge derived from digitally available natural history collections, as compared to many other organisms.

We outline two possible scenarios. Under the first scenario, researchers may actively seek data-deficient areas because they offer an opportunity to find new species and fill gaps in biodiversity knowledge. Under this scenario, researchers may be more likely to sample accessible areas but the extent of knowledge of an area should reduce the desire to revisit it for additional sampling – a phenomenon previously documented for the collection of individual species (Steege et al., 2011). This is, for example, the case for certain biological surveys in the 16th and 17th centuries, when collected specimens were often treated as art pieces and hidden from competitors until their economic value was estimated (Ritterbush, 1969; Impey & MacGregor, 1985).

Under the second scenario, sampling planning may not be primarily driven by the attempt to fill in gaps, but rather by the likelihood of retrieving data, often under time constraints. Therefore, researchers may return to visited areas because finding a focal species to obtain appropriate tissue for molecular phylogenetic analyses may be easier, more certain, and more cost-efficient. One example of this scenario is the Mount Namuli in Mozambique, surveyed repeatedly in 1931–1932, 1998, 2007, 2011, 2014, and 2016, which rendered it considerably better sampled than any surrounding areas (Vincent, 1933;

Ryan et al., 1999; Timberlake et al., 2009; Portik et al., 2013; Farooq & Conradie, 2015; Conradie et al., 2016). This repeated sampling resulted in new range expansions (Farooq & Conradie, 2015), new species (Conradie et al., 2018), and a better understanding of the biogeography of the region through phylogenetic studies (Branch et al., 2014; Bittencourt-Silva et al., 2016).

Methods

Using a grid-cell size equivalent to 100 by 100 km (more specifically the cells were 100 by 100 km at 30 degrees North or South; cells at lower latitudes were wider and lower, while cells are higher latitude were thinner and higher), we tested whether knowledge of biodiversity within a cell changes the likelihood of additional sampling within it. One advantage of this approach is that it shows whether sampling is spatially restricted because some sites are easier to reach, or if the very existence of knowledge is causing scientists to revisit well-known areas.

For all analyses, we worked on a cylindrical equal-area Berhmann projection. We estimated the time it would take to sample at least once in 90% of the land area of Africa, and the number of sampling events required to record at least 50% of the species of an area of 10,000 km^2. Our estimation of the time to sample 90% of Africa was based on the assumption that the rate of biodiversity sampling since the 1800s can be adequately described by a Weibull distribution, while in the sampling effort analysis we removed the temporal aspect by randomizing the years 100 times.

All analyses were conducted using three groups: amphibians, mammals, and birds. Our species occurrence dataset consisted of records retrieved from the Global Biodiversity Information Facility (GBIF) for amphibians (https://doi.org/10.15468/dl.hyyea9), mammals (https://doi.org/10.15468/dl.gms3up) excluding bats, and birds (https://doi.org/10.15468/dl.unxn5u) recorded in Africa from 1801 until the end of 2019 (31st December, 2019). Bats and marine mammals were excluded from the analyses of mammals because they are generally sampled by different methods and researchers than for non-flying terrestrial mammals.

We focused on species occurrence records contributed by scientists, since these provide the primary source of information and material for the community of professional systematists. Citizen science observations, although important for popular engagement and data gathering, were therefore excluded due to their mixed systematic value (e.g., (Troudet et al., 2018). Although not all countries in Africa are formal participants of the GBIF Network, which could potentially lead to underestimation of the completeness of each grid cell, most species collections in Africa are housed by members of the GBIF network, such as in South Africa, western European countries and the United States. Other intrinsic limitations of GBIF are that it does not provide access to records collected and stored in non-participant countries or in scientific articles or reports where no voucher was collected, such as in photography-based inventories characteristic of Environmental impact assessments. This might also contribute to the underestimation of the true completeness rate.

To update and synonymize the taxonomy from GBIF, we used the R package RangeBuilder (version 1.4) (Rabosky et al., 2016) and removed the records of species not present in the IUCN's polygon list. We applied the R package CoordinateCleaner (Zizka et al., 2019) to exclude duplicates and records outside the IUCN range for each species.

We define a sampling event in each cell as sampling within a given calendar year.

To estimate a simple measure of sampling completeness, we followed a similar approach to Myers, et al (2015). We assume that the range polygons created by IUCN using version 2019-3 (IUCN, 2019) are accurate. We consider as unsampled those cells overlapped by a range polygon but without the respective species record, while any records outside the range polygons are assumed to be errors. While we acknowledge that none of these assumptions fully capture the complexity of species distributions and occurrences, we consider them sufficiently close to reality for the purpose of our analysis and unlikely to result in major systematic biases.

Effect of previous knowledge on sampling probability

We calculated the probability of visiting any grid cell according to its ratio of completeness using a Logistic model of the sampling events by completeness plus year. While few cells in non-desert parts of Africa have communities of amphibians, mammals and birds with fewer than 10 species, the same did not hold true for the threshold of at least 5 sampling events, resulting in the exclusion of a large portion of grid cells from our analysis

Cells with fewer sampling events are more likely to have more extreme effect sizes, resulting from the size of the difference between few data points. We therefore restricted our analyses to test only whether the effect was positive or negative. To distinguish between primarily colonial sampling and recent sampling, we conducted separate analyses for two time periods: 1801–1940, 1980–2019 and then for the total period: 1801–2019. Although most countries attained independence between 1950 and 1980, separating colonial times from independence dates can be misleading. We assumed the end of World War II as an igniter of the emancipation of African countries following Cooper (2019), and considered the period of 1940–1980 uncertain and impossible to assign across diverse African countries. The lower boundary of 1800 was arbitrarily chosen to encapsulate all subsequent colonial periods and to avoid any dubious earlier records. In summary, we considered pre-1940 colonial, post-1980 noncolonial, and 1940–1980 unknown and therefore excludable from our analyses. The term colonial is used here as a shorthand for the time period prior to the emergence of the independent African nations we see now, and that, depending on the part of Africa, this definition includes a period before the formal establishment of the colonial structure.

Spatial patterns might affect the degree at which sampling completeness increases the probability of researchers visiting a cell. To investigate this possibility, we carried out a multiple correlation analysis with variables that cover both environmental and social aspects. These included annual precipitation (Worldclim v 2: Fick & Hijmans, 2017), human influence (CIESIN, 2005), net primary production (Imhoff et al., 2004), protected areas coverage (UNEP-WCMC, 2019), elevation (Jarvis et al., 2008), Human Development Index (HDI), and Gross Domestic Product (GDP) per capita (Kummu et al., 2018).

To account for spatial autocorrelation, regressions were conducted through simultaneous autoregressive models with spatial error (i.e. SAR_{Err} models; (Haining & Haining, 2003). We tested 40 different neighbourhoods (10 with a fixed number of neighbours between one and ten, and 30 containing all combinations of cells up to 250, 500, 750, 1,000, 1,250 or 1,500 km away with either of the five default weighting schemes). The best neighbourhood was chosen as the one minimizing the corrected Akaike Information Criterion (AICc) by using the R package wiqid version 0.2.2 (Meredith, 2017). To obtain an estimate of how close our data are to the fitted regression line, we calculated the Nagelkerke pseudo R-squared using the function *summary.sarlm* from the R package spdep version 1.1-2 (Bivand et al., 2005).

To predict the required time to sample 90% of cells in Africa, we used survival analyses (Demetrius, 1978). We treated unsampled cells as 'alive' and sampled cells as 'dead', and treated all cells still not visited in 2019 as 'censored'. This treatment is usually done in survival analysis to describe unknown survival time. We then fitted our data with a Weibull distribution, which is able to accommodate the three basic survival curves with constant, monotonically increasing or monotonically decreasing mortality (Pinder et al., 1978). Mathematically, this assumes that the sampling rate r at time t can be modelled as

$$r(t) = \frac{\Phi}{\Psi}\left(\frac{t}{\Psi}\right)^{\Phi-1}$$

which for $\Phi=1$ means a constant sampling rate, whereas $\Phi>1$ means that sampling rates are increasing over time in a consistent manner.

Our assumption that the sampling rate of collection in Africa can be modelled by a single Weibull distribution constitutes a technical oversimplification, given the idiosyncrasies of each country and region. This rate is likely to have been affected by socio-political events (Rydén et al., 2019), and the implementation of regulations such as CITES, the Nagoya Protocol, and national and regional legislations, potentially leading to reduced fieldwork. The implementation of this assumption is however required by our modelling approach, which is limited by the requirement of a single rate for the temporal predictions.

To predict the number of sampling events necessary to find at least half of all the predicted species in a particular cell, we did a spatial regression analysis using the formula:

$$\text{SaEv50} = \text{SaEv} \sim \text{C},$$

where SaEv50 corresponds to the number of sampling events to record at least 50% of completeness, SaEv is the number of sampling events and C is the completeness value. To remove the effect of a particular year on our model, we randomized the year of visit in our dataset 100 times and used the median value.

Both the SARerr regressions and the linear regression were conducted in R 3.5.1 (R Core Team, 2018) using glm (R Core Team, 2018), errorsarlm from the package spdep (Bivand et al., 2005) and the R package survival ver 2.44-1.1 (Therneau & Lumley, 2015).

Additional R packages used for data manipulation and visualization were:

BBmisc version 1.11 (Bischl et al., 2017), Rphylopic version 0.2.0 (Chamberlain, 2018), Tidyverse version 1.2.1 (Hadley Wickham & Wickham, 2017), gridExtra version 2.3 (Auguie et al., 2017), raster version 2.9-5 (Hijmans et al., 2015), pbapply version 1.4-0 (Solymos, 2016), ggplot2 version 3.1.0 (H. Wickham, 2016), sf version 0.7-1 (Pebesma, 2018) and package rnaturalearth version 0.1.0 (South, 2017)

Results

By overlapping the map of Africa obtained from the R package rnaturalearth version 0.1.0 (South, 2017) with our grid, we obtained a total of 3,212 cells that include any portion of continental Africa. Among the groups surveyed, birds were the most well sampled in our study area, for which we retrieved 775,131 records distributed across 1,798 cells. Mammals had 63,521 records across 1,578 cells and amphibians only had 15,991 records distributed across 936 cells (Fig. 1). The lower number of cells for amphibians is due to the substantial areas of dryland habitats in Africa.

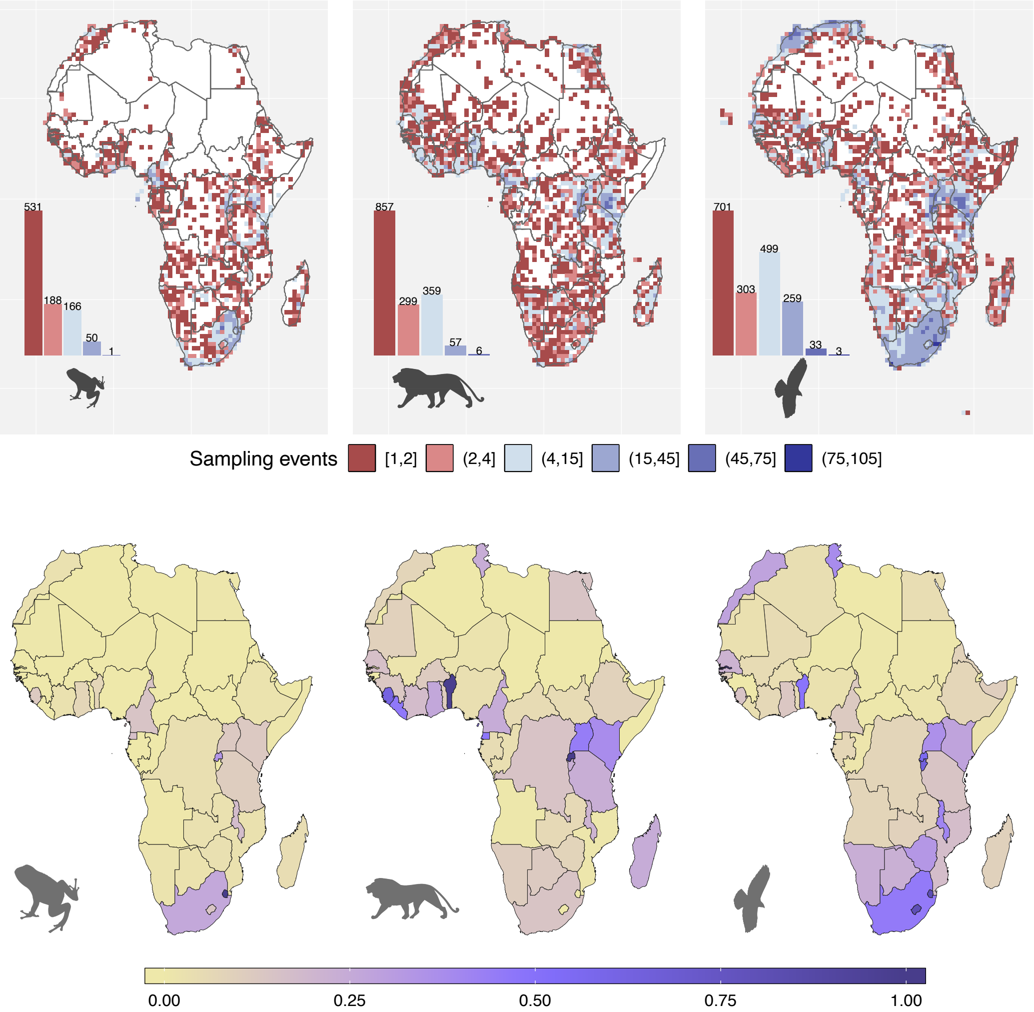


Figure 21: **Top:** Number of sampling events per grid cell for each taxon group, as estimated from the Global Biodiversity Information Facility. White areas represent areas without any records. Sampling events can be interpreted as the number of different years in which a particular cell has been visited. **Bottom:** Number of

sampling events by country. We divided the number of sampling events by country area and transformed the data by dividing each value by the maximum value of each taxon for visualization purposes. Much of Africa has rarely been visited. South Africa and Tanzania are in the top three with the greatest number of cells – more than five sampling events for all three taxa. Regarding the least-sampled countries – those lacking any cells with more than 5 sampling events – we identified 25 countries for amphibians, 16 for mammals, and 5 for birds. Comoros, Libya, and Djibouti have no cells with more than 5 sampling events for any of these groups. The full list of sampling events by country per taxa is provided in the Supplementary material (Table S1, S2, S3).

Effect of previous knowledge on the probability of visiting a cell

We assessed the effect of biodiversity knowledge available for a particular cell on the probability of sampling the cell, by using a logistic model between sampling completeness, sampling events, and year. Negative values correspond to cells in which the amount of existing knowledge decreases the probability of resampling a cell. In contrast, a positive value corresponds to a positive effect of existing knowledge on the probability of sampling.

By analyzing data that are readily available to the scientific community, we found that amphibians were seldom recorded between 1801 and 1940, when compared with mammals and birds (Fig 2, left panel). In our analyses, we explicitly accounted for overall temporal changes in sampling. Our results therefore investigate any patterns in sampling probability happening in addition to overall temporal changes. Across most of Africa, previous sampling in amphibians has strongly influenced the resampling of areas: between 1801 and 1940, 83% of the previously sampled cells had a positive effect. This is similar to the period between 1982 and 2019, where 78% of the previously sampled cells were positive (Fig. 2). Mammologists also show the strongest preference for revisiting areas previously sampled, with similar values of 79% before 1940 and 81% after 1982 (Fig. 2). The bird sampling shows that before 1940, there was an increased preference of discovery inferred by a higher sampling frequency in unsampled areas among researchers, when compared to after 1982. In both time periods, previous sampling decreased the likelihood of visits in the majority of cells, and the pattern became more pronounced in the later time period. This is demonstrated by an increase from 66% to 75% in the proportion of positive cells (Fig. 2).

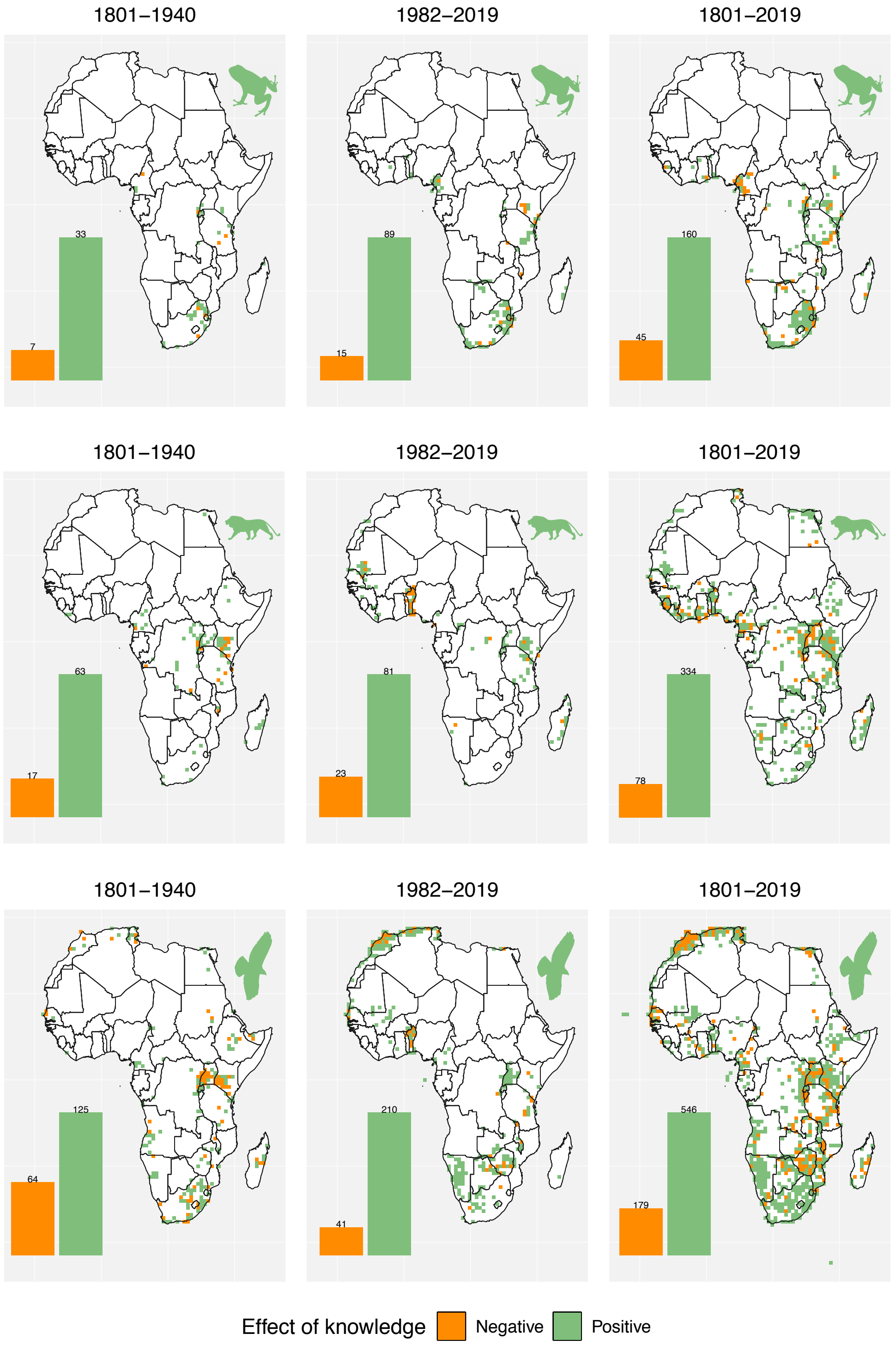


Figure 22: Effect of previous knowledge on sampling of amphibians, mammals, and birds. The spatial patterns were calculated through logistic regressions between completeness and sampling events in amphibians, mammals, and birds, and the sign of the effect mapped per cell. White grid cells were not used in this analysis because they either did not contain any records, had less than five visits, or had an expected richness lower than 10. The left panel corresponds to the filtered data available between 1801 and 1940; the middle panel corresponds to the period of 1982 to 2019; and the right panel includes data from 1801 to 2019.

Only cells with 5 or more sampling events were included to reduce the effects of overfitting. Bar values represent the number of cells corresponding to negative/positive values. For all three taxa analyzed, there is strong and pervasive evidence that previous knowledge leads to increased sampling. All taxa also show a temporal increase in this trend. Maps with only cells with p value < 0.05 can be found in the supplementary material (Fig S1, S2, S3).

Sampling events required to record 50% of species

Our analysis revealed small effects between the tested predictors and their effect on the attractiveness of an area based on pre-existing knowledge. The general pattern is that the existence of increased knowledge increases the likelihood of increased sampling (Table 1). We only found three instances of small significant effects. In amphibians, increased knowledge was slightly less likely to lead to increased sampling in areas with higher net primary productivity (NPP) (Table 1). In birds, increased knowledge was slightly less likely to lead to increased sampling in areas with more roads, and slightly more likely to lead to increased sampling in areas with higher precipitation (Table 1).

Table 2: Predictors for the effect of previous sampling on the probability of sampling for amphibians, mammals, and birds in Africa. Significance of the following predictors for the effect of previous sampling on the probability of sampling: IUCN predicted richness, protected area coverage, human influence, HDI, GDP per capita, NPP, precipitation, elevation, and road density. Human influence is a significant positive predictor for amphibians and negative for birds. Predicted richness is a significant negative predictor for amphibians.

	Amphibians	Mammals	Birds
	Estimate (std. error)	Estimate (std. error)	Estimate (std. error)
(Intercept)	0.777(0.023)***	0.806(0.011)***	0.758(0.031)***
IUCN richness	-0.089(0.038)*	-0.081(0.018)	-0.025(0.032)
Protected areas	-0.017(0.03)	0.009(0.019)	0.011(0.017)
Human influence	0.101(0.032)**	-0.036(0.021)	-0.069(0.023)**
HDI	-0.048(0.034)	-0.019(0.02)	-0.004(0.027)
GDP	-0.012(0.03)	0.02(0.021)	-0.03(0.018)
NPP	-0.036(0.04)	0.006(0.026)	-0.059(0.031)
Precipitation	-0.052(0.042)	-0.043(0.022)	0.053(0.03)

Elevation	-0.003(0.028)	0.02(0.017)	-0.019(0.019)
Road density	-0.002(0.028)	0.005(0.017)	-0.031(0.016)
AICc	214	390.4	774.8
Nagelkerke pseudo R-squared	0.141	0.073	0.104

* 0.05>p>0.01; ** 0.01>p>0.001; *** p<0.001

Time to sample at least once in 90% of Africa

The model predicted the time it would take, assuming a Weibull distributed biodiversity sampling rate, to sample at least once in 90% of Africa. For amphibians, we predicted the sampling coverage of Africa to be achieved between 2192 and 2233, for mammals between 2222 and 2257, and for birds between 2253 and 2294 (Fig. 3).

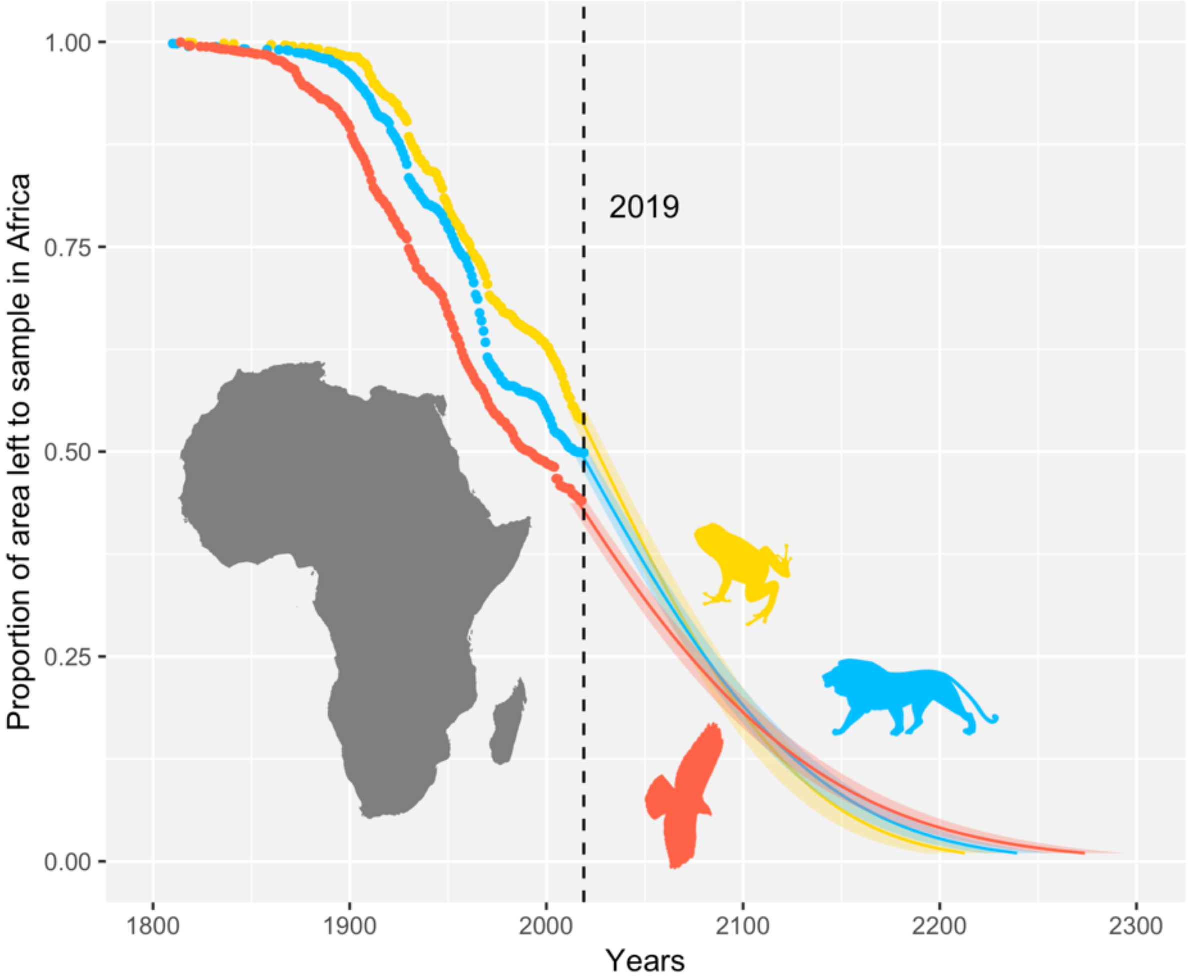


Figure 23: Proportion of grids (10,000 km^2) in Africa sampled at least once and containing at least 10 species. We used only the cells where the groups are expected to occur, meaning that amphibians, birds, and mammals had a different number of cells where they could be recorded. Ribbons indicate 95% confidence intervals. These analyses predict that, for birds, species occurrence data for 90% of Africa will only be achieved somewhere between years 2253 and 2294; for mammals between 2222 and 2257; and for amphibians between 2192 and 2233.

Sampling events required to achieve 50% inventory

Our spatial regression analysis showed that the number of sampling events required to record 50% of the species was mainly positively associated with Human Development Index and elevation of the grid cells for amphibians. For mammals, Human Development Index was positively associated with the number of sampling events. For birds, richness was positively associated with the number of sampling events, unlike NPP, which displayed a negative association. We found that it would take on average 11.5 visits for amphibians, 12.7 visits for mammals, and 27.0 visits for birds to recover 50% of all species within a cell (Supplementary Table S4).

Discussion

In this study, we unveil, quantify, and map a new bias in biodiversity data: that knowledge in itself is leading to an increase in sampling bias. For all three examined vertebrate groups, researchers tend to return to areas based on the existence of previous knowledge rather than visiting and sampling new areas.

The lack of biological specimens and tissues for most of the African continent is likely to impact the comprehensiveness of phylogenetic reconstructions, our understanding of taxonomic relationships, and the estimation of divergence times and biogeographic history (Hortal et al., 2015). These are just some of the expected effects caused by the exclusion of critical species, populations, and genetic diversity in phylogenetic analyses. Since there are few biological barriers across the African continent (Fjeldså, 1994), and political borders in Africa seldom follow natural features, it is essential that the systematics community acknowledges the need for wide spatial sampling in their research. In addition, conservation practitioners must also be aware of this bias. The need to reliably assess global biodiversity in an era where conservation relies increasingly on 'big data' (Arts et al., 2015) demands the use of estimation and extrapolation (Colwell, 2009). However, such procedures can be heavily compromised by uneven sampling across large land extensions (Reddy & Dávalos, 2003). The effect of sampling bias on diversity estimates may also obstruct solid inference on underlying drivers of biodiversity build-up as well as loss (Loiselle et al., 2008; Engemann et al., 2015). The reluctance to visit new areas may also affect the diversity pattern for microendemics, since such species may only be known from studied areas. For instance, species endemic to individual mountains or inselbergs in eastern Africa (Branch et al., 2014; Bittencourt-Silva et al., 2016) would only be known from sites visited by a specialist.

Since more than a third of Africa lacks digitally accessible information (Meyer, Kreft, Guralnick, & Jetz, 2015), we expected that researchers conducting inventories would be attracted to data deficient areas (Scenario 1 above). However, our results show the contrary: sampling events tend to occur where knowledge already exists (Scenario 2). We find a temporal increase in this trend, when comparing the sampling events from before 1940 with the sampling events after 1982. This pattern might be explained as a result of the spread of the reputation of a particular area for harbouring high diversity, as mentioned by Reddy and Dávalos (2003) when they recorded a disproportionate amount of sampling effort towards areas rich in biodiversity. It may also be that previous sampling indicates the area is accessible and so can be surveyed with fewer resources than a site that has not been sampled. There is also a higher risk of a non-surveyed site not containing the targeted species.

In order to exclude the influences of other predictors on the effect of previous knowledge in attracting more sampling events, we used a spatial regression analysis of completeness in relation to visits. We demonstrated that these effects cannot be attributed to any of the traditionally tested predictors (the predictors were not significant or had minimal effects). For amphibians, we observed a negative significant effect for predicted richness and a significant positive effect of human influence; however, in both cases those effects were quantitatively minimal and much smaller than the effect of previous knowledge (Table 1).

Completing the biodiversity inventory of Africa

Our analyses indicated that it could take between 172 and 274 years for the research community to carry out at least one sampling effort in 90% of all cells across Africa. These estimates are based on current and historic rates of biological exploration, and encompass only some of the most well-studied organism groups of all: birds, mammals, and amphibians. In addition, a single sampling effort is far from enough to correctly characterize the diversity of any site: in our estimates, between 12 and 27 events are required to record at least 50% of the existing species.

Our models showed significant positive effects of elevation and Human Development Index on the number of sampling events for amphibians; significant positive effects of Human Development Index on the number of sampling events for mammals; significant positive effects of IUCN predicted richness; and significant negative effects of protected areas and net primary production on the sampling of birds. Even though not all protected areas used in this analysis were officially established since the 19th century, many of them that eventually became protected were likely to have been under some form of protection earlier on, under classifications such as reserves or hunting concessions. The significant positive effect of elevation on the number of sampling events on amphibians might be a consequence of the high endemism within the group in mountains such as in the Eastern Arc (Burgess et al., 2007).

Analyzing data collected through long periods of time at a continental scale poses numerous challenges, one of them being the continuous process of human expansion and consequent habitat transformation. Some areas may experience the emergence of mining or even become urban centres. We expect this to be a problem at fine scales, but due to our grid-cell area of ~10 000 km^2, we expect the complete transformation of entire cells to be relatively rare. We further note that we only focus on species that, according to IUCN, currently occur in the cell. This means that species that are locally extirpated from much of Africa, such as lions, are not included even if there are records from the species prior to its local extirpation from any given area.

Conclusions and recommendations

This study conveys an urgent and crucial message: unless a radical and widespread change in research practice takes place, Africa's rich biodiversity will remain largely unknown. We cannot protect or understand what we do not know about, yet the data available for most of Africa to adequately identify and delimit species boundaries, understand spatial biodiversity patterns, or to effectively promote species conservation are insufficient.

Waiting more than a century to complete the biodiversity inventory of Africa is not a viable option. Africa is experiencing the highest population growth of any continent (Gerland et al., 2014), with an expected 209% increase between 2000 and 2050 (United Nations, 2017). Between 2015 and 2050, an additional 2.4 billion people are expected, which in combination with a rapidly changing climate will exert a tremendous pressure on natural ecosystems and their biodiversity. The recent Living Planet Report 2020 (WWF, 2020) shows that wild populations of African vertebrates have declined an alarming 65% over the last 50 years alone.

It is also important to note that our estimates are based only on birds, mammals, and amphibians – three well-studied groups. The knowledge bias and spatial patterns we report are likely to be considerably worse for other groups such as plants, fungi and insects (Stropp et al., 2016; Willis, 2017, 2018). Such diversity may hold important solutions to help achieve the Sustainable Development Goals (Antonelli et al., 2019), but will be largely lost if not effectively mapped and conserved.

To tackle the challenges outlined in this study we make the following four recommendations:

1. **Funding providers (agencies, companies, philanthropists) should actively promote projects aiming to sample areas that lack baseline biodiversity data.** We recognize the difficulties in weighing up the costs and benefits of allocating limited time and resources towards additional data collection (Grand et al., 2007). Although additional sampling is also needed in well-sampled areas (such as already

recognized biodiversity or endemism hotspots), for instance, to increase our full understanding of biodiversity and biotic interactions in those areas, it is essential to increase the focus on poorly sampled areas (Reddy & Dávalos, 2003).

2. **Researchers should, whenever possible, increase the taxonomic and methodological scope of their collection efforts.** Biodiversity inventories usually involve producing point-locality biological data and conducting basic taxonomy (da Fonseca et al., 2000). Given the logistic and legislative challenges of carrying out fieldwork across most of Africa, we urge scientists to collaborate with specialists in different institutions and with varied taxonomic expertise to responsibly sample the maximum possible number of taxa (in full or as tissue samples, especially for endangered or large species). Expeditions with multi-taxa foci may be especially valuable for creating baseline biodiversity data in numerous data-deficient areas, and should be encouraged by funding agencies and biodiversity institutions. It is also imperative to concentrate on physical specimens (whole specimens, tissue samples for DNA analyses, seeds for cultivation in the case of plants, among others), rather than only photographic evidence (Troudet et al., 2018), while preventing negative effects on the survival of threatened populations and species. Whenever possible, duplicates of samples should be deposited in multiple organisations, to increase their long-term safety and accessibility. The collection of rich metadata will increase the value of collections for many and as yet unforeseen uses (Bakker et al., 2019; Fernández et al., 2019).

3. **Engage globally and locally.** Biological sampling in Africa has to a large extent been carried out by European and North American institutions, with limited benefits returning to the countries of origin. Under the Access and Benefit-Sharing agreements of the Convention of Biological Diversity, it is crucial that future sampling activities are always done in close partnership with African institutions and researchers, for mutual benefits (Pearce et al., 2020)

4. **Clarity on processing research permits.** There is certain evidence that excessive in-country legislation that regulates research and collection permits can sometimes hinder research (Rydén et al., 2019; Williams et al., 2020). If possible, clarification on the process for sampling permits should be made transparent and available online for every country in the continent, to encourage and streamline biodiversity research.

References

Amarasinghe, A., Kuritsky, J. N., Letson, G. W., & Margolis, H. S. (2011). *Dengue virus infection in Africa*. Emerging infectious diseases, 17(8), 1349.

Antonelli, A., Smith, R. J., & Simmonds, M. S. (2019). *Unlocking the properties of plants and fungi for sustainable development*. Nature plants, 5(11), 1100-1102.

Arts, K., Van Der Wal, R., & Adams, W. M. (2015). *Digital technology and the conservation of nature*. Ambio, 44(4), 661-673.

Auguie, B., Antonov, A., & Auguie, M. B. (2017). *Package 'gridExtra'*. Miscellaneous Functions for "Grid" Graphics.

Bakker, F. T., et al. (2019). The Global Museum: natural history collections and the future of evolutionary biology and public education (2167-9843).

Bates, J. M., Bowie, R. C., Willard, D. E., Voelker, G., & Kahindo, C. (2004). *A need for continued collecting of avian voucher specimens in Africa: why blood is not enough*.

Beegle, K., Christiaensen, L., Dabalen, A., & Gaddis, I. (2016). Poverty in a rising Africa: The World Bank.

Bello-Schünemann, J., & Moyer, J. D. (2018). *Structural pressures and political instability-trajectories for sub-Saharan Africa*. ISS Africa Report, 2018(9), 1-32.

Bhatt, S., et al. (2015). *The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015*. Nature, 526(7572), 207-211.

Bini, L. M., Diniz-Filho, J. a. F., Rangel, T. F., Bastos, R. P., & Pinto, M. P. (2006). *Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot*. Diversity and distributions, 12(5), 475-482.

Bischl, B., Lang, M., Bossek, J., Horn, D., Richter, J., & Surmann, D. (2017). *BBmisc: Miscellaneous Helper Functions for B*. Bischl. R package version, 1.

Bittencourt-Silva, G. B., Conradie, W., Siu-Ting, K., Tolley, K. A., Channing, A., Cunningham, M., Farooq, H. M., Menegon, M., & Loader, S. P. (2016). *The phylogenetic position and diversity of the enigmatic mongrel frog Nothophryne Poynton, 1963 (Amphibia, Anura)*. Molecular phylogenetics evolution, 99, 89-102. doi:https://doi.org/10.1016/j.ympev.2016.03.021

Bivand, R., et al. (2005). *The spdep package*. Comprehensive R Archive Network, Version, 05-83.

Boakes, E. H., Mcgowan, P. J., Fuller, R. A., Chang-Qing, D., Clark, N. E., O'connor, K., & Mace, G. M. (2010). *Distorted views of biodiversity: spatial and temporal bias in species occurrence data*. PLoS biology, 8(6), e1000385.

Branch, W. R., Bayliss, J., & Tolley, K. A. (2014). *Pygmy chameleons of the Rhampholeon platyceps complex (Squamata: Chamaeleonidae): Description of four new species from isolated 'sky islands' of northern Mozambique*. Zootaxa, 3814(1), 1-36.

Burgess, N., et al. (2007). *The biological importance of the Eastern Arc Mountains of Tanzania and Kenya*. Biological Conservation, 134(2), 209-231.

Chamberlain, S. (2018). rphylopic: Get 'Silhouettes' of 'Organisms' from 'Phylopic'.

R package version 0.2.0. . Retrieved from https://CRAN.R-project.org/package=rphylopic

Colwell, R. K. (2009). *Biodiversity: concepts, patterns, and measurement*. The Princeton guide to ecology, 257-263.

Conradie, W., Bittencourt-Silva, G. B., Farooq, H. M., Loader, S. P., Menegon, M., & Tolley, K. A. (2018). *New species of Mongrel Frogs (Pyxicephalidae: Nothophryne) for northern Mozambique inselbergs*. African Journal of Herpetology, 1-25. doi:https://doi.org/10.1080/21564574.2017.1376714

Conradie, W., Bittencourt-Silva, G. B., Loader, S. P., Menegon, M., Nanvonamuquitxo, C., & Scott, M. (2016). *Exploration into the hidden world of Mozambique's sky island forests: new discoveries of reptiles and amphibians.* doi:https://doi.org/10.3897/zse.92.9948

Cooper, F. (2019). Africa since 1940: the past of the present (Vol. 13): Cambridge University Press.

Da Fonseca, G. A., et al. (2000). *…following Africa's lead in setting priorities*. Nature, 405(6785), 393.

Demetrius, L. (1978). *Adaptive value, entropy and survivorship curves*. Nature, 275(5677), 213-214.

Engel, U., Gebauer, C., & Hüncke, A. (2015). *Notes from Within and Without: Research Permits between Requirements and'Realities'*.

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., Peet, R. K., Violle, C., & Svenning, J. C. (2015). *Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot*. Ecology and Evolution, 5(3), 807-820.

Farooq, H. O. M., & Conradie, W. (2015). *Second record of a Scolecomorphus kirkii Boulenger, 1883 (Gymnophiona: Scolecomorphidae) for Mozambique*. Herpetology Notes, 8, 59-62.

Feeley, K. J., & Silman, M. R. (2011). *The data void in modeling current and future distributions of tropical species*. Global Change Biology, 17(1), 626-630.

Fernández, N., Guralnick, R., & Kissling, W. D. (2019). *A minimum set of Information Standards for Essential Biodiversity Variables*. Biodiversity Information Science and Standards.

Fick, S. E., & Hijmans, R. J. (2017). *WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas*. International journal of climatology, 37(12), 4302-4315.

Fjeldså, J. (1994). *Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities*. Biodiversity & Conservation, 3(3), 207-226.

Foster, V., & Briceño-Garmendia, C. M. (2009). Africa's infrastructure: a time for transformation: The World Bank.

Gerland, P., et al. (2014). *World population stabilization unlikely this century*. Science, 346(6206), 234-237.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). *New developments in museum-based informatics and applications in biodiversity analysis*. Trends in ecology & evolution, 19(9), 497-503.

Grand, J., Cummings, M. P., Rebelo, T. G., Ricketts, T. H., & Neel, M. C. (2007). *Biased data reduce efficiency and effectiveness of conservation reserve networks*. Ecology Letters, 10(5), 364-374.

Haining, R. P., & Haining, R. (2003). Spatial data analysis: theory and practice: Cambridge University Press.

Harford, J. B. (2015). *Barriers to overcome for effective cancer control in Africa*. The Lancet Oncology, 16(8), e385-e393.

Hijmans, R. J., et al. (2015). *Package 'raster'*. R package.

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.

Hotez, P. J., & Kamath, A. (2009). *Neglected tropical diseases in sub-Saharan Africa: review of their prevalence, distribution, and disease burden*. PLoS neglected tropical diseases, 3(8).

Imhoff, M., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R., & Lawrence, W. (2004). Global patterns in net primary productivity. Data distributed by the Socioeconomic Data and Applications Center (SEDAC). In.

Impey, O., & Macgregor, A. (1985). The Origins of Museums: The Cabinet of Curiosities in Sixteenth-and Seventeenth-Century Europe. In: Oxford: Clarendon Press.

Iucn. (2019). The IUCN Red List of Threatened Species. Version 2019-3. http://www.iucnredlist.org. Retrieved 18 July 2019

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database. In.

Kummu, M., Taka, M., & Guillaume, J. H. (2018). *Gridded global datasets for gross domestic product and Human Development Index over 1990–2015*. Scientific data, 5, 180004.

Loiselle, B. A., Jørgensen, P. M., Consiglio, T., Jiménez, I., Blake, J. G., Lohmann, L. G., & Montiel, O. M. (2008). *Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes?* Journal of Biogeography, 35(1), 105-116.

Lomolino, M. V. (2004). *Conservation biogeography*. Frontiers of Biogeography: new directions in the geography of nature, 293.

Mbaku, J. M. (2010). Corruption in Africa: Causes, consequences, and cleanups: Lexington Books.

Meredith, M. (2017). wiqid: Quick and Dirty Estimates for Wildlife Populations. R package version 0.2. 2. In.

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). *Global priorities for an effective information basis of biodiversity distributions*. Nature communications, 6, 8221.

Meyer, C., Weigelt, P., & Kreft, H. (2016). *Multidimensional biases, gaps and uncertainties in global plant occurrence information*. Ecology Letters, 19(8), 992-1006.

Pearce, T., et al. (2020). *International collaboration between collections-based institutes for halting biodiversity loss and unlocking the useful properties of plants and fungi*. Plants, People, Planet 2(5).

Pebesma, E. (2018). *Simple features for R: standardized support for spatial vector data*. The R Journal, 10(1), 439-446.

Pinder, J. E., Wiener, J. G., & Smith, M. H. (1978). *The Weibull distribution: a new method of summarizing survivorship data*. Ecology, 59(1), 175-179.

Portik, D. M., Mulungu, E., Sequeira, D., & Mcentee, J. (2013). *Herpetological surveys of the Serra Jeci and Namuli massifs, Mozambique, and an annotated checklist of the southern Afromontane archipelago*. Herpetological Review, 44(3), 394-406.

R Core Team. (2018). *R: A language and environment for statistical computing*.

Rabosky, A. R. D., Cox, C. L., Rabosky, D. L., Title, P. O., Holmes, I. A., Feldman, A., & Mcguire, J. A. (2016). *Coral snakes predict the evolution of mimicry across New World snakes*. Nature communications, 7, 11484.

Reddy, S., & Dávalos, L. M. (2003). *Geographical sampling bias and its implications for conservation priorities in Africa*. Journal of Biogeography, 30(11), 1719-1727.

Ritterbush, P. C. (1969). *Science Teaching and the Future*. The Science Teacher, 36(6), 32-39.

Ryan, P. G., Bento, C., Cohen, C., Graham, J., Parker, V., & Spottiswoode, C. (1999). *The avifauna and conservation status of the Namuli Massif, northern Mozambique*. Bird Conservation International, 9(4), 315-331.

Rydén, O., Zizka, A., Jagers, S. C., Lindberg, S. I., & Antonelli, A. (2019). *Linking democracy and biodiversity conservation: Empirical evidence and research gaps*. Ambio, 1-15.

Sánchez-Fernández, D., Lobo, J. M., Abellán, P., Ribera, I., & Millán, A. (2008). *Bias in freshwater biodiversity sampling: the case of Iberian water beetles*. Diversity and distributions, 14(5), 754-762.

Schmitt, C. J., Cook, J. A., Zamudio, K. R., & Edwards, S. V. (2019). *Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene*. Philosophical Transactions of the Royal Society B, 374(1763), 20170387.

Schmidt-Lebuhn, A. N., Knerr, N. J., & Kessler, M. (2013). *Non-geographic collecting biases in herbarium specimens of Australian daisies (Asteraceae)*. Biodiversity and Conservation, 22(4), 905-919.

Shaffer, H. B., Fisher, R. N., & Davidson, C. (1998). *The role of natural history collections in documenting species declines*. Trends in ecology & evolution, 13(1), 27-30.

Solymos, P. (2016). Zawadzki, Z. pbapply: Adding Progress Bar to '* apply'Functions. R package version 1.2-1. In.

South, A. (2017). R natural earth: World map data from natural earth. In: R package version 0.1. 0, URL https://CRAN. R-project. org/package ….

Steege, H. T., Haripersaud, P. P., Bánki, O. S., & Schieving, F. (2011). *A model of botanical collectors' behavior in the field: never the same species twice*. American Journal of Botany, 98(1), 31-37.

Stork, H., & Astrin, J. J. (2014). *Trends in biodiversity research—A bibliometric assessment*. Open Journal of Ecology, 4(07), 354.

Stropp, J., Ladle, R. J., M. Malhado, A. C., Hortal, J., Gaffuri, J., H. Temperley, W., Olav Skøien, J., & Mayaux, P. (2016). *Mapping ignorance: 300 years of collecting flowering plants in Africa*. Global Ecology and Biogeography, 25(9), 1085-1096.

Therneau, T. M., & Lumley, T. (2015). *Package 'survival'*. R Top Doc, 128.

Timberlake, J., et al. (2009). *Mt Namuli, Mozambique: biodiversity and conservation*. Report for Darwin Initiative Award, 15, 036.

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). *Taxonomic bias in biodiversity data and societal preferences*. Scientific reports, 7(1), 9132.

Troudet, J., Vignes-Lebbe, R., Grandcolas, P., & Legendre, F. (2018). *The increasing disconnection of primary biodiversity data from specimens: How does it happen and how to handle it?* Systematic biology, 67(6), 1110-1119.

Unep-Wcmc, I. (2019). Protected Planet: the World Database on Protected Areas (WDPA). Retrieved August, 28

United Nations. (2017). World population prospects 2017. In: United Nations New York.

Vincent, J. (1933). *The Namuli Mountains, Portuguese East Africa*. The Geographical Journal, 81(4), 314-327.

Walker, A., Bush, T., & Oduro, G. K. (2006). *New principals in Africa: preparation, induction and practice*. Journal of educational administration.

Ward, D. F. (2012). *More than just records: analysing natural history collections for biodiversity planning*. PloS one, 7(11), e50346.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. In. Springer-Verlag New York.

Wickham, H., & Wickham, M. H. (2017). *Package tidyverse*. Easily Install and Load the 'Tidyverse.

Wildlife Conservation Society - Wcs, & Center for International Earth Science Information Network - Ciesin - Columbia University. (2005). *Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic)*. Retrieved from: https://doi.org/10.7927/H4BP00QC

Williams, C., et al. (2020). *Conservation Policy: Helping or hindering science to unlock properties of plants and fungi*. Plants, People, Planet 2(5).

Willis, K. (2017). *State of the World's Plants 2017. Report*. R. Bot. Gard. Kew.

Willis, K. (2018). *State of the World's fungi 2018*. Royal Botanic Gardens, Kew.

Wwf. (2020). Living planet report-2020: Bending the curve of biodiversity loss. Retrieved from Gland, Switzerland:

Zizka, A., Antonelli, A., & Silvestro, D. (2020). *sampbias, a method for quantifying geographic sampling biases in species distribution data*. bioRxiv.

Zizka, A., et al. (2019). *CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases*. Methods in Ecology and Evolution, 10(5), 744-751.

Supplementary material

Table S1: Number of ~10,000 km^2 cells sampled for amphibians per country in the following groupings of sampling events (SEv): 0, 1-2, 3-4,5-15,16-45, 46-75,76-105,106-122 and sum of all cells with more than 5 SEv.

	Sampling events									
Country	0	1-2	3-4	5-15	16-45	46-75	76-105	106-122	Sum > 5	SEv/Km^2
Algeria	260	7	1	0	0	0	0	0	0	2308857.818
Angola	106	41	7	1	0	0	0	0	1	1244654.314
Benin	7	11	4	2	0	0	0	0	2	116113.309
Botswana	52	20	5	3	0	0	0	0	3	579029.2005
Burkina Faso	37	3	0	0	0	0	0	0	0	272769.3967
Burundi	3	3	3	0	0	0	0	0	0	27041.2663
Cabo Verde	8	0	0	0	0	0	0	0	0	3883.142708
Cameroon	34	20	3	9	7	0	0	0	16	464319.2783
Central African Republic	82	4	0	0	0	0	0	0	0	617984.24
Chad	156	1	0	0	0	0	0	0	0	1266282.336
Comoros	4	0	0	0	0	0	0	0	0	1672.226921
Democratic Republic of the Congo	179	65	20	16	3	0	0	0	19	2325240.424
Djibouti	6	2	0	0	0	0	0	0	0	21847.6141
Egypt	107	7	1	2	0	0	0	0	2	1001078.547
Equatorial Guinea	7	3	1	1	0	0	0	0	1	26671.69086
Eritrea	22	6	1	0	0	0	0	0	0	122537.7388
eSwatini	0	0	0	4	0	0	0	0	4	17111.84504
Ethiopia	102	22	12	5	0	0	0	0	5	1127375.98
Gabon	19	18	4	0	0	0	0	0	0	259968.4847
Gambia	1	2	1	0	0	0	0	0	0	10500.89704
Ghana	14	15	6	3	1	0	0	0	4	238668.677
Guinea	32	6	2	0	0	0	0	0	0	244301.9801
Guinea-Bissau	3	7	1	0	0	0	0	0	0	32829.62807
Ivory Coast	29	13	0	2	0	0	0	0	2	320677.1569
Kenya	31	20	10	13	4	0	0	0	17	585702.5844
Lesotho	1	4	2	1	0	0	0	0	1	30106.51807

Liberia	10	8	4	0	0	0	0	0	0	95298.15395
Libya	198	0	0	0	0	0	0	0	0	1623760.9
Madagascar	47	23	7	5	1	0	0	0	6	592983.8843
Malawi	13	5	3	5	0	0	0	0	5	119397.7665
Mali	161	4	0	0	0	0	0	0	0	1252723.657
Mauritania	121	5	6	1	0	0	0	0	1	1036391.824
Morocco	53	24	6	4	0	0	0	0	4	591745.2469
Mozambique	70	30	7	5	0	0	0	0	5	788448.5366
Namibia	51	37	17	4	0	0	0	0	4	822713.0502
Niger	143	4	1	0	0	0	0	0	0	1181300.892
Nigeria	98	11	5	0	0	0	0	0	0	907501.0844
Republic of the Congo	46	11	0	0	0	0	0	0	0	344888.6767
Rwanda	2	2	1	2	0	0	0	0	2	25305.05492
São Tomé and Principe	-1	0	0	2	1	0	0	0	3	1037.156137
Senegal	25	4	1	0	0	0	0	0	0	196224.2971
Sierra Leone	4	7	2	2	0	0	0	0	2	71611.71499
Somalia	59	8	0	1	0	0	0	0	1	471815.0034
Somaliland	23	4	0	0	0	0	0	0	0	167406.9577
South Africa	38	33	17	49	25	1	0	0	75	1219826.728
South Sudan	80	6	0	0	0	0	0	0	0	626861.8322
Sudan	221	1	0	0	0	0	0	0	0	1857637.067
Togo	9	4	0	0	0	0	0	0	0	56863.47734
Tunisia	25	4	1	0	0	0	0	0	0	156611.6936
Uganda	19	8	1	6	1	0	0	0	7	241853.6302
United Republic of Tanzania	52	33	13	22	2	0	0	0	24	941505.9871
Western Sahara	23	0	0	0	0	0	0	0	0	90494.88906
Zambia	66	30	4	5	0	0	0	0	5	751914.9431
Zimbabwe	32	13	8	3	0	0	0	0	3	389337.4811

Table S2: Number of ~10,000 km^2 cells sampled for mammals per country in the following groupings of sampling events (SEv): 0, 1-2, 3-4,5-15,16-45, 46-75,76-105,106-122 and sum of all cells with more than 5 SEv

	Sampling events									
Country	0	1-2	3-4	5-15	16-45	46-75	76-105	106-122	Sum > 5	SEv/Km^2
Algeria	234	32	2	0	0	0	0	0	0	2308857.818
Angola	111	40	3	1	0	0	0	0	1	1244654.314
Benin	1	2	3	13	5	0	0	0	18	116113.309
Botswana	30	30	9	11	0	0	0	0	11	579029.2005
Burkina Faso	16	16	3	5	0	0	0	0	5	272769.3967
Burundi	4	1	3	1	0	0	0	0	1	27041.2663
Cabo Verde	8	0	0	0	0	0	0	0	0	3883.142708
Cameroon	31	18	6	13	5	0	0	0	18	464319.2783
Central African Republic	69	10	2	4	1	0	0	0	5	617984.24
Chad	131	25	0	0	0	0	0	0	0	1266282.336
Comoros	2	1	1	0	0	0	0	0	0	1672.226921
Democratic Republic of the Congo	122	77	29	42	13	0	0	0	55	2325240.424
Djibouti	6	2	0	0	0	0	0	0	0	21847.6141
Egypt	41	39	14	20	3	0	0	0	23	1001078.547
Equatorial Guinea	4	5	1	2	0	0	0	0	2	26671.69086
Eritrea	25	4	0	0	0	0	0	0	0	122537.7388
eSwatini	3	1	0	0	0	0	0	0	0	17111.84504
Ethiopia	72	40	13	16	0	0	0	0	16	1127375.98
Gabon	21	12	6	2	0	0	0	0	2	259968.4847
Gambia	2	2	0	0	0	0	0	0	0	10500.89704
Ghana	10	15	4	10	0	0	0	0	10	238668.677
Guinea	13	17	5	5	0	0	0	0	5	244301.9801
Guinea-Bissau	8	3	0	0	0	0	0	0	0	32829.62807
Ivory Coast	6	22	7	7	2	0	0	0	9	320677.1569
Kenya	20	14	9	20	11	4	0	0	35	585702.5844
Lesotho	8	1	0	0	0	0	0	0	0	30106.51807
Liberia	6	3	6	7	0	0	0	0	7	95298.15395

Libya	172	25	1	0	0	0	0	0	0	1623760.9
Madagascar	23	23	14	22	1	0	0	0	23	592983.8843
Malawi	10	7	5	4	0	0	0	0	4	119397.7665
Mali	113	44	5	3	0	0	0	0	3	1252723.657
Mauritania	77	28	13	15	0	0	0	0	15	1036391.824
Morocco	25	37	18	7	0	0	0	0	7	591745.2469
Mozambique	68	39	4	1	0	0	0	0	1	788448.536
Namibia	26	53	16	14	0	0	0	0	14	822713.0502
Niger	114	29	1	4	0	0	0	0	4	1181300.892
Nigeria	75	25	9	5	0	0	0	0	5	907501.0844
Republic of the Congo	48	7	0	2	0	0	0	0	2	344888.6767
Rwanda	1	2	0	0	4	0	0	0	4	25305.05492
São Tomé and Principe	2	1	0	0	0	0	0	0	0	1037.156137
Senegal	7	12	6	5	0	0	0	0	5	196224.2971
Sierra Leone	4	2	2	6	1	0	0	0	7	71611.71499
Somalia	63	5	0	0	0	0	0	0	0	471815.0034
Somaliland	23	3	1	0	0	0	0	0	0	167406.9577
South Africa	40	58	37	25	3	0	0	0	28	1219826.728
South Sudan	55	19	7	5	0	0	0	0	5	626861.8322
Sudan	185	32	4	0	0	0	0	0	0	1857637.067
Togo	3	6	3	1	0	0	0	0	1	56863.47734
Tunisia	12	4	8	6	0	0	0	0	6	156611.6936
Uganda	3	7	8	12	5	0	0	0	17	241853.6302
United Republic of Tanzania	37	34	16	29	6	0	0	0	35	941505.9871
Western Sahara	23	0	0	0	0	0	0	0	0	90494.88906
Zambia	53	32	13	7	0	0	0	0	7	751914.9431
Zimbabwe	22	24	5	5	0	0	0	0	5	389337.4811

Table S3: Number of ~10,000 km^2 cells sampled for birds per country in the following groupings of sampling events (SEv): 0, 1-2, 3-4,5-15,16-45, 46-75,76-105,106-122 and sum of all cells with more than 5 SEv.

	Sampling events									
Country	0	1-2	3-4	5-15	16-45	46-75	76-105	106-122	Sum > 5	SEv/Km^2
Algeria	191	35	15	12	13	2	0	0	27	2.04E-05
Angola	63	40	25	25	2	0	0	0	27	3.78E-05
Benin	1	4	3	13	3	0	0	0	16	0.000137796
Botswana	18	16	8	35	3	0	0	0	38	0.000107076
Burkina Faso	23	14	0	3	0	0	0	0	3	4.03E-05
Burundi	3	0	1	4	1	0	0	0	5	0.000221883
Cabo Verde	0	4	2	2	0	0	0	0	2	0.002060187
Cameroon	34	19	4	11	5	0	0	0	16	6.25E-05
Central African Republic	68	15	2	1	0	0	0	0	1	3.24E-06
Chad	119	28	7	1	1	0	0	0	2	1.03E-05
Comoros	-1	3	2	0	0	0	0	0	0	0.00119601
Democratic Republic of the Congo	127	66	38	38	11	3	0	0	52	3.01E-05
Djibouti	6	2	0	0	0	0	0	0	0	0.000137315
Egypt	66	30	8	10	3	0	0	0	13	4.79E-05
Equatorial Guinea	7	3	2	0	0	0	0	0	0	0.000187465
Eritrea	21	4	1	3	0	0	0	0	3	4.08E-05
eSwatini	0	0	0	0	4	0	0	0	4	0.000233756
Ethiopia	64	46	8	21	2	0	0	0	23	6.21E-05
Gabon	21	10	5	5	0	0	0	0	5	6.92E-05
Gambia	0	1	0	3	0	0	0	0	3	0.00038092
Ghana	9	17	4	9	0	0	0	0	9	0.000100558
Guinea	24	12	4	0	0	0	0	0	0	3.27E-05
Guinea-Bissau	7	2	1	1	0	0	0	0	1	0.000121841
Ivory Coast	21	15	2	6	0	0	0	0	6	4.99E-05
Kenya	10	14	7	22	17	8	0	0	47	9.22E-05
Lesotho	0	1	0	6	1	0	0	0	7	0.000265723
Liberia	7	9	5	1	0	0	0	0	1	8.39E-05
Libya	187	10	1	0	0	0	0	0	0	3.70E-06
Madagascar	12	31	24	13	3	0	0	0	16	8.94E-05
Malawi	6	3	3	10	4	0	0	0	14	0.000167507

Mali	115	25	7	16	2	0	0	0	18	1.76E-05
Mauritania	88	28	8	9	0	0	0	0	9	1.74E-05
Morocco	22	7	10	20	23	5	0	0	48	0.000108155
Mozambique	28	26	19	32	6	1	0	0	39	6.09E-05
Namibia	7	25	21	50	6	0	0	0	56	0.000103317
Niger	122	23	1	2	0	0	0	0	2	5.08E-06
Nigeria	62	40	5	7	0	0	0	0	7	2.31E-05
Republic of the Congo	39	11	6	1	0	0	0	0	1	1.74E-05
Rwanda	0	2	1	0	4	0	0	0	4	0.000237107
São Tomé and Principe	-1	0	0	3	0	0	0	0	3	0.002892525
Senegal	5	9	4	7	5	0	0	0	12	0.000122309
Sierra Leone	3	4	5	3	0	0	0	0	3	0.000111714
Somalia	32	30	3	3	0	0	0	0	3	6.36E-06
Somaliland	12	9	1	5	0	0	0	0	5	2.99E-05
South Africa	-1	4	1	35	114	7	3	0	159	0.000133626
South Sudan	53	19	11	3	0	0	0	0	3	9.57E-06
Sudan	190	24	3	4	0	0	0	0	4	2.96E-05
Togo	6	4	2	1	0	0	0	0	1	0.000140688
Tunisia	8	3	2	9	7	1	0	0	17	0.000127704
Uganda	2	3	5	10	12	3	0	0	25	0.000111638
United Republic of Tanzania	35	32	16	29	10	0	0	0	39	7.75E-05
Western Sahara	17	3	2	1	0	0	0	0	1	1.11E-05
Zambia	41	30	14	19	1	0	0	0	20	6.65E-05
Zimbabwe	3	8	8	28	9	0	0	0	37	0.000123286

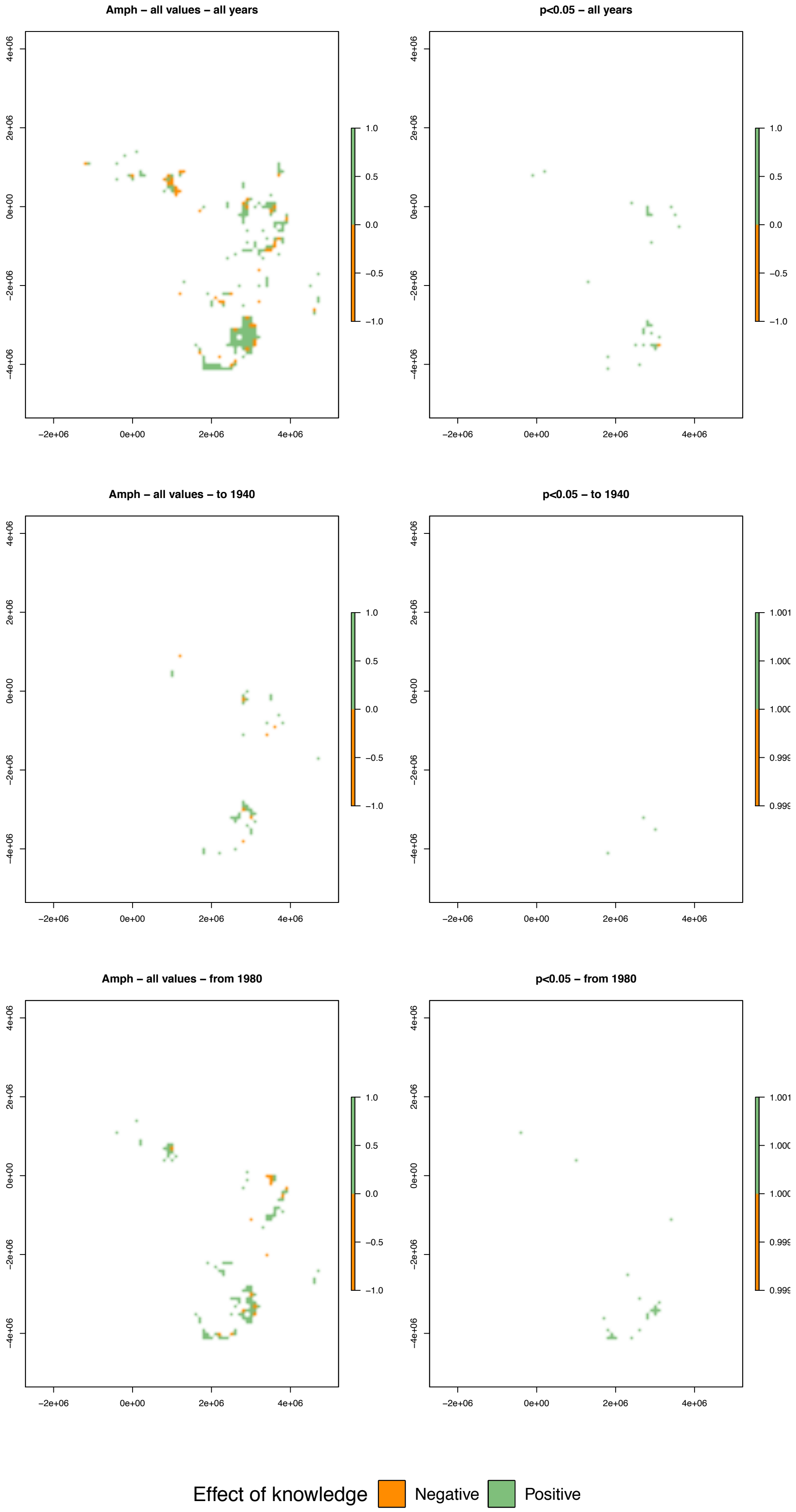

Amph – all values – all years
p<0.05 – all years
Amph – all values – to 1940
p<0.05 – to 1940
Amph – all values – from 1980
p<0.05 – from 1980
4e+06
2e+06
0e+00
−2e+06
−4e+06
−2e+06
0e+00
2e+06
4e+06
1.0
0.5
0.0
−0.5
−1.0
1.001
1.000
1.000
0.999
0.999
Effect of knowledge
Negative
Positive

Figure S1: Effect of previous knowledge on sampling of amphibians (p value < 0.05). The spatial patterns were calculated through logistic regressions between completeness and sampling events in amphibians, mammals and birds, and the sign of the effect mapped per cell. White grid cells were not used in this analysis because they either do not contain any records, have less than five visits or have an expected richness lower than 10. The top row corresponds to the filtered data available from 1801 to 1940; the middle row corresponds to the period of 1982 to 2019; and the bottom panel includes data from 1801 to 2019. Only cells with 5 or more sampling events were included to reduce the effects of overfitting. Bar values represent the number of cells corresponding to negative/positive values. For all three taxa analysed, there is strong and pervasive evidence that previous knowledge leads to increased sampling. All taxa also show a temporal increase in this trend.

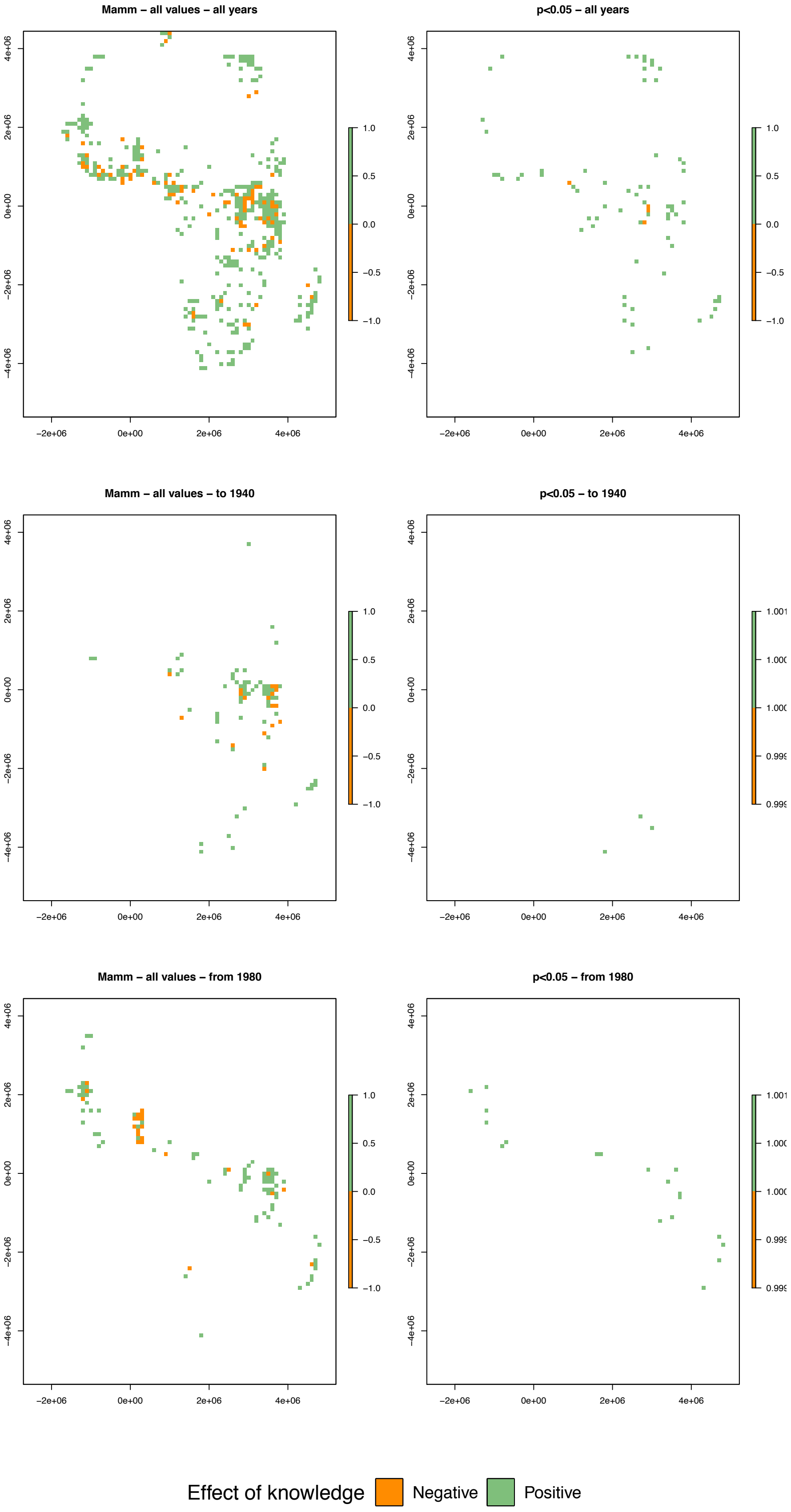

Mamm – all values – all years
p<0.05 – all years
Mamm – all values – to 1940
p<0.05 – to 1940
Mamm – all values – from 1980
p<0.05 – from 1980
4e+06
2e+06
0e+00
−2e+06
−4e+06
−2e+06
0e+00
2e+06
4e+06
1.0
0.5
0.0
−0.5
−1.0
1.001
1.000
1.000
0.999
0.999
Effect of knowledge
Negative
Positive

Figure S2: Effect of previous knowledge on sampling of mammals (p value < 0.05). The spatial patterns were calculated through logistic regressions between completeness and sampling events in amphibians, mammals and birds, and the sign of the effect mapped per cell. White grid cells were not used in this analysis because they either do not contain any records, have less than five visits or have an expected richness lower than 10. The top row corresponds to the filtered data available from 1801 to 1940; the middle row corresponds to the period of 1982 to 2019; and the bottom panel includes data from 1801 to 2019. Only cells with 5 or more sampling events were included to reduce the effects of overfitting. Bar values represent the number of cells corresponding to negative/positive values. For all three taxa analysed, there is strong and pervasive evidence that previous knowledge leads to increased sampling. All taxa also show a temporal increase in this trend.

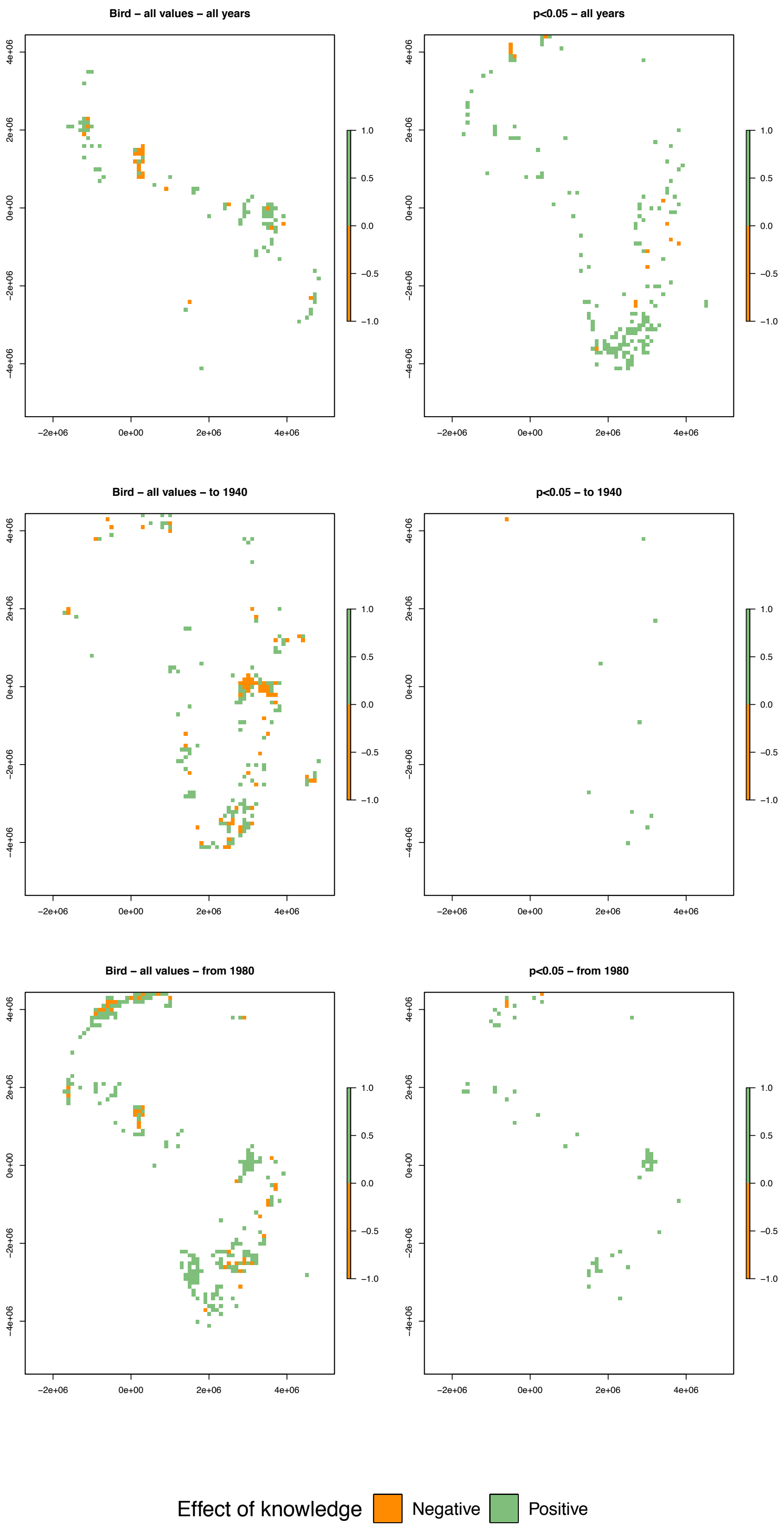


Figure S3: Effect of previous knowledge on sampling of birds. (p value < 0.05). The spatial patterns were calculated through logistic regressions between completeness and sampling events in amphibians, mammals and birds, and the sign of the effect mapped per cell. White grid cells were not used in this analysis because they either do not contain any records, have less than five visits or have an expected richness lower than 10. The top row corresponds to the filtered data available from 1801 to 1940; the middle row corresponds to the period of 1982 to 2019; and the bottom panel includes data from 1801 to 2019. Only cells with 5 or more sampling events were included to reduce the effects of overfitting. Bar values represent the number of cells corresponding to negative/positive values. For all three taxa analysed, there is strong and pervasive evidence that previous knowledge leads to increased sampling. All taxa also show a temporal increase in this trend.

Table S4: Spatial regression summary of the model predicting the number of required sampling events until at least 50% of the species of amphibians have been recorded in a particular 1-degree-grid, with the following predictors: IUCN richness, protected areas coverage, human influence, HDI, NPP, elevation.

	Amphibians	Mammals	Birds
	Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)
(Intercept)	11.456(0.708)***	12.174(1.206)***	27.002(2.297)***
IUCN richness	-	10.627(1.306)	7.354(2.968)*
Protected areas	-	-	-
Human influence	-	-	-
HDI	1.61(0.715)*	4.709(1.306)**	-
NPP	-	-	-7.158(2.534)**
Elevation	2.771(0.715)***	-	-
AICc	358.6	153	255.3
R^2 (a)	0.249	0.748	0.275

* 0.05>p>0.01; ** 0.05>p>0.001; *** p<0.001

(a) We used the adjusted R squared values for the linear model in the amphibians and mammals and the Nagelkerke pseudo R-squared for the SAR model in birds.

CHAPTER 3

SHEDDING LIGHT ON A BIODIVERSITY DARK SPOT:

CHECKLIST OF AMPHIBIANS AND REPTILES OF PEMBA, CABO DELGADO PROVINCE, MOZAMBIQUE

Abstract

The biodiversity of northern Mozambique is one of the least known in southern Africa and may be severely underestimated. The country is expected to have a relatively rich herpetofauna, due to the variety of different available habitat types as well as its large geographic extension. Here we summarize 17 years of reported collections and observations from the city of Pemba and surrounding areas in Cabo Delgado Province, in an attempt to better understand the diversity and distribution of amphibians and reptiles in coastal northern Mozambique. We report 53 species (19 amphibians and 34 reptiles), of which more than half had never previously been recorded from the area. Pemba has one endangered amphibian and two critically endangered reptiles. All amphibians recorded in this study have been assessed by IUCN, but only six of the 34 reptiles have conservation status. Five snakes are of medical importance, but 80% of them were recorded for the first time from Cabo Delgado Province.

Our results confirm the large under-sampling of the herpetofauna in northern Mozambique and suggest that many other taxa remain unobserved and thereby lacking any form of active protection.

Key Words. – Africa; distribution range; herpetofauna; inventory; Mozambique; Pemba.

Resumo. – A biodiversidade do norte de Moçambique é uma das menos conhecidas no sul de África, e corre o risco de estar amplamente subestimada. Supõe-se que o país tenha uma herpetofauna relativamente rica, devido à variedade de habitats diferentes disponíveis assim como à sua vasta extensão. Neste estudo nós compilamos dezassete anos de observações na cidade de Pemba e arredores em Cabo Delgado numa tentativa de melhor conhecer a diversidade e distribuição de anfíbios e répteis do norte de Moçambique. Nós reportamos 53 espécies (19 anfíbios e 34 répteis), dos quais mais de metade nunca tinham sido reportados da área. Todos os anfíbios neste estudo já tiveram o seu estado de conservação elaborado pela IUCN, mas apenas seis dos 34 répteis têm um estado de conservação. Cinco serpentes são de importância médica, sendo que 80% foram registados pela primeira vez na província de Cabo Delgado. Os nossos resultados testemunham a escassez de amostragem no Norte de Moçambique, e sugerem a existência de muitas espécies ainda por documentar e, portanto, sem qualquer tipo de protecção activa.

Palavras chave. – Africa; distribuição; herpetofauna; inventário; Moçambique; Pemba.

Submitted to: Journal of Herpetological Conservation
Authors: Harith Farooq, Cristóvão Nanvonamuquitxo, Bibiana Nassongole, Werner Conradie, Roger Bills, Amadeu Soares, Alexandre Antonell

Introduction

Northern Mozambique is one of the most poorly known areas in southern Africa in terms of biodiversity, especially herpetofauna (Poynton and Broadley 1991; Tolley et al. 2016). This has attracted the interest in recent years of international and local researchers resulting in several biodiversity surveys conducted in the northernmost provinces of Mozambique: Timberlake et al. (2009), Bayliss et al. (2010), Branch (2004), Branch et al. (2005), Portik et al. (2013a), Bayliss et al. (2014), Farooq et al. (2015), Conradie et al. (2016) and Jones et al. (2017; 2020). Most of these multidisciplinary expeditions were undertaken in the mountainous region of the interior parts of northern Mozambique: Mounts Chiperone, Mabu, Namuli, Inago, Ribaue, Mecula, and Njese, whereas, from the coastal Cabo Delgado Province, the only available published data were produced by Rasplus et al. (2008) which contain the results of multidisciplinary surveys of the coastal forests. A checklist of the amphibians derived from these surveys was subsequently published (Ohler and Frétey 2014). Together, these surveys resulted in the discovery of new species and in the range expansion of several species of crustaceans (Daniels and Bayliss 2012), butterflies (Bayliss et al. 2016; Bayliss et al. 2018; Van Velzen et al. 2016), amphibians (Conradie et al. 2018b; Farooq et al. 2015; Farooq and Conradie 2015), mammals (Monadjem et al. 2010; Taylor et al. 2012) and reptiles (Branch and Bayliss 2009; Branch et al. 2014; Branch et al. 2017; Branch et al. 2005; Branch and Tolley 2010; Branch et al. 2019; Broadley and Farooq 2013; Broadley and Measey 2016; Portik et al. 2013b; Verburgt et al. 2018).

Prior to 2005, the only surveys conducted in northern Mozambique were conducted by Barbour and Loveridge (1928) reaching the area of Lumbo, ca. 220 km south of Pemba and Blake (1965) who collected material in Quelimane all the way up to Ilha de Moçambique, also 220 km south of Pemba. The amphibians from this trip were published by Poynton (1966) and used in Amphibia Zambeziaca (Poynton and Broadley 1985a; 1987; 1988; Poynton and Broadley 1985b). Despite all these efforts and resources, the diversity of large areas in Mozambique –especially the northern coastal parts of the country – remains virtually unknown. It is very probable that the number and distribution of species is largely underestimated. This contrasts with the situation in neighboring countries, such as Malawi, South Africa and Tanzania, which are relatively well sampled. For example South Africa have approximately 458 species of reptiles (Bates et al. 2014; Bauer et al. 2019; Conradie et al. 2018a; Heinicke et al. 2017; Jacobsen et al. 2014; Travers et al. 2014; Weinell and Bauer 2018) and 133 species of amphibians (Du Preez and Carruthers 2017; Wilson and Channing 2019), whereas Tanzania has 343 species of reptiles (Spawls et al. 2018) and 206 species of amphibians (AmphibiaWeb 2020).

To date, there is no comprehensive and updated checklist of reptiles for Mozambique. A recent publication by (Ohler and Frétey 2014) attempted to provide a preliminary checklist of amphibians for the country, resulting in a list with 82 species (which excludes five recently described species mentioned above). In an alternative approach to amalgamating biodiversity surveys, an online checklist identified about ~211 species of reptiles (Uetz et al. 2020) and ~90 amphibian species (AmphibiaWeb 2020) which are expected to occur in the whole of Mozambique. Due to the variety of different habitat types available and the large size of the country, we therefore expect a considerably higher diversity than is currently reported, as has been speculated for decades (Atauri and de Lucio 2001; Pianka 1967).

For the region of Pemba, a book showcasing the most common herpetofauna around Pemba's bay is available (Farooq et al. 2014) but no formal checklist has yet been published. The only available records from the region are from the published dataset of the South African Institute for Aquatic Biodiversity (SAIAB) also available through the Global Biodiversity Information Facility (GBIF). They recorded the occurrence of 10 amphibian species collected in 2010 from the Pemba region.

In order to improve our understanding of the herpetofauna species richness of northern Mozambique, here we compile sporadic species records made by members of our team and collaborators since 2003 and merged with information in consulted museums and online databases. Our data sources include sightings, call presence, and collected specimens from the city of Pemba. Our compilation confirms the predicted distribution ranges of several species, and more than doubled the number of amphibians and reptiles known to occur in Pemba area.

Methods

Study site. – Pemba (-12.9732°, 40.5178°), is a port city in northern Mozambique. It is also the capital of the province of Cabo Delgado province and according to the latest census (2017) has a population of 201 846. The city lies on a peninsula in Pemba's Bay and its elevation is from sea level to a maximum of 110 m. The elevation is from sea level to a maximum of 110 m. The warmest month of the year is December, with an average temperature of 27.8 °C. July is the coldest month of the year with an average temperature of 23.9 °C. Regarding precipitation, there are two distinct seasons: the dry season and the wet season. The wet season spans from December to April, with the wettest month of the year typically being March, with an average of 199 mm total monthly precipitation. The dry season stretches from May to November, with the driest month of the year typically being September, with an average of 4 mm total monthly precipitation (https://en.climate-data.org/location/3995/).

In the city and surrounding areas, a diverse array of micro-habitats for amphibians and reptiles can be found. These include: coastal grasslands (Fig. 1B) with baobabs, mango and cashew trees; the coral rag around the bay (Fig. 1C); the urban areas (Fig. 1D); and sporadic freshwater wetlands formed usually due to the accumulation of rainwater (Fig. 1E).

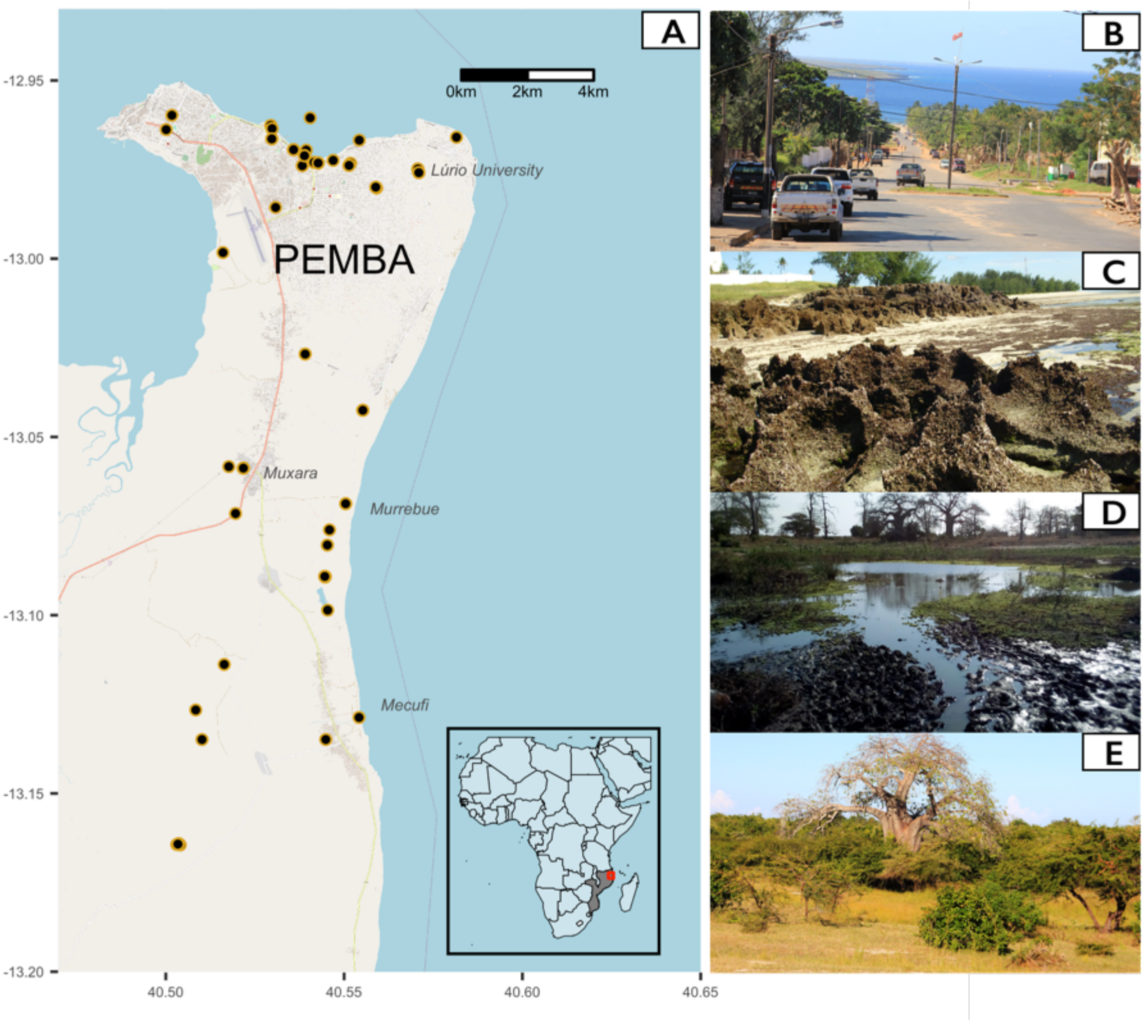


Figure 24: A: Map of Pemba showing the sites from where we obtained the records. B-E, A selection of characteristic habitats in Pemba: B. urban area, city center, C. coral rag at Maringanha's beach, D. wetland Chuiba's district, E. coastal grassland in Murrébué.

Data collection. – New records from Pemba were obtained between 2003 and 2020. The basis of records includes direct observations made by members of our team, the capture of specimens by hand or trapping and specimens collected and donated to us by third parties. All voucher specimens were first fixed in 10% formalin for a period of 10 to 30 days, after which they were immersed in water for a day in order to remove the formalin and finally transferred to a permanent storage solution of 70% ethanol deposited at the Lúrio University (Pemba campus), Port Elizabeth Museum (PEM) and the South African Aquatic Biodiversity Institute (SAIAB) (Table 1).

In order to produce the most up-to-date species list for Pemba, we compared the data compiled for this study with records sourced from other museums and online databases (e.g. iNaturalist; https://www.inaturalist.org and ReptileMap; http://vmus.adu.org.za). To the best of our knowledge, no other records from Pemba have been published. For this purpose, we downloaded all species records in GBIF using the search terms: amphibians; reptiles; and drawing a polygon around the city of Pemba (accessed on 10 April 2019; DOI: 10.15468/dl.zofsjg).

Results

In the current study we recorded a total of 51 species in Pemba region, representing 17 species of amphibians and 34 species of reptiles. Combined with other records from iNaturalist, the herpetofauna of Pemba now comprises a total of 53 species, of which 19 are species of amphibians and 34 of reptiles (Table 1, 2). Below we list the most relevant known species of amphibians and reptiles occurring in Pemba, organized by family and listing the source of each record, their main characteristic traits, distribution and general comments. We also noted whenever a species was of medical importance in accordance to the World Health Organization guidelines as in Longbottom et al. (2018). The remaining commented species can be found in the supplementary material.

AMPHIBIANS

Leptopelis mossambicus Poynton, 1985

Mozambique forest Tree Frog

Material: https://www.inaturalist.org/observations/19660822. **Identification**: A light brown colored *Leptopelis*, with a n-shaped dark brown band from near occipital to urostyle and with moderately developed digital toe discs. **Distribution**: This species occurs from southern Malawi, through the lowlands of central and southern Mozambique, southeastern Zimbabwe and Eswatini, into northern and eastern South Africa, south to Durban (Channing and Rödel 2019). In Mozambique it has been recorded from south and central parts of the country (Poynton and Broadley 1985b). **Comment**: This is the most northern record published of the species, extending the range 700 km northeast, although it has been found 25 km northwest of Pemba (https://www.inaturalist.org/observations/19794897).

Mertensophryne cf. *anotis* (Boulenger, 1907)

Chirinda Toad

Fig. 2J

Observations: https://www.inaturalist.org/observations/19535975; https://www.inaturalist.org/observations/21458586; https://www.inaturalist.org/observations/22156353; https://www.inaturalist.org/observations/22157714**.** **Identification**: Medium-sized toad with broad flattened parotid glands. The dorsum is light brown with irregular but paired darker patches and a small inverted V or pair of diverging patches in the post-occipital area. Fine light vertebral line running from post-occipital marking to above vent. The limbs are barred. **Distribution**: This species is only known from the Chirinda Forest in eastern Zimbabwe, the forest 24 km north of Dombe in adjacent Mozambique and from Taratibu in Quirimbas National Park in northern Mozambique (Farooq et al. 2015). **Comment:** The species was collected under the name of *M. loveridgei* by Pascal (2011) from Rovuma river and additional records recently emerged from Zambezi Province and coastal Mozambique from Pemba to Tanzanian border (Bittencourt-Silva and Conradie *pers. comm. 2019*). The taxonomic status of new populations in northern Mozambique is still unresolved. They either represent a continuous distribution of *M. anotis*, which can be assigned to Tanzanian *M. loveridgei*, or represent an undescribed species.

Mertensophryne lindneri (Mertens, 1955)

Lindner's Dwarf Toad

Fig. 2A

Material: FCN0548, -13.05875˚ 40.52175˚, Muxara (an unpaved area where water accumulates during the rainy season); https://www.inaturalist.org/observations/21264601. **Identification:** Has a gray-brown dorsum with small reddish-brown spots, long and flattened parotid glands and a distinct dark black ventral line from the throat to the vent. **Distribution:** This species is known to occur from northeastern Tanzania, south through eastern Tanzania to southeastern Malawi, crossing the Mozambican provinces of Niassa, Cabo Delgado, Nampula and Zambézia (Channing and Rödel 2019). **Comments:** (Poynton 1966) recorded its presence in the provinces of Niassa and Zambézia but no published records exist for its presence in Cabo Delgado Province. *Mertensophryne lindneri* collected in Pemba is 100 km east of IUCN's predicted range (IUCN SSC Amphibian Specialist Group 2013a).

Schismaderma carens (Smith, 1848)

Red Toad

Fig. 2B

Material: FCN0533, -12.97391˚ 40.53827˚, Josina Machel behind Shoprite; FCN0531, FCN0535, -12.97383˚ 40.55138˚, Bairro Eduardo Mondlane; FCN0549–52, -13.05875˚ 40.52175˚, Muxara (an unpaved area); FCN501, -12.97316˚ 40.54271˚, Wimbe beach. **Identification**: Easily identifiable bufonid, by having no parotid glands and a large and distinct tympanum. Has a prominent dorsolateral skin ridge and a mid-dorsum pair of dark markings. **Distribution**: This species ranges from north-western Tanzania and southern Kenya south to south-eastern Democratic Republic of Congo, eastern Angola and western Zambia, as far as south-eastern Botswana, southern Mozambique, Eswatini and the southern KwaZulu-Natal Province of South Africa (Channing and Rödel 2019). **Comment**: There are no published records for this species in Cabo Delgado Province before this study. The *Schismaderma carens* collected in Pemba is 250 km east of IUCNs predicted range (IUCN SSC Amphibian Specialist Group 2013b). This is one of the most common observed frogs in Pemba city.

REPTILES

Mochlus sundevalli (Smith, 1849)

Sundevall's Writhing Skink

Fig. 4B

Material: PEM R19737, PEM R19739–40, -12.96350˚ 40.52981˚, Pemba Beach Hotel garden; PEM R19850, PEM R19851, PEM R19860, PEM R19861, PEM R19862, -12.96639˚ 40.52972˚, Pemba Beach Hotel; PEM R05964, -12.97111˚ 40.53889˚, Pemba Beach Hotel. **Identification**: A medium to large, shiny-bodied, small limbed writhing skink with a short, flat, wedge-shaped snout and small eyes where the lower eyelid is scaly. The body is stout and cylindrical and the tail is thick, about half the total length. **Distribution**: This fossorial species is found in eastern and southern Africa, from Somalia in the north, through Kenya and Uganda southwest to Angola and Namibia, and south to northern South Africa. In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch et al. 2005; Pascal 2011; Timberlake et al. 2012). **Comment:** According to (Freitas et al. 2018), populations from Mozambique previously referred to as *M. afer* should be assigned to *M. sundevalli*, due to the absence of phylogenetic and morphological support for their differentiation. Common throughout the city.

Trachylepis margaritifera (Peters, 1854)

Rainbow Skink

Fig. 4C

Material: FCN0525, -12.97483˚ 40.57069˚, Eduardo Mondlane Campus; FCN0411, FCN0040, FCN0051, -12.97316 40.54271˚, Wimbe beach; PEM R19859, -12.96639˚ 40.52972˚, Pemba Beach Hotel. **Identification:** Females and juveniles are easily identifiable by having three conspicuous cream stripes on a black background with bright blue tails. Adult males are bronze with speckled white. **Distribution**: This skink has a wide distribution that spans from Kenya south through Tanzania, Mozambique to South Africa and Eswatini, inland from Uganda south through Rwanda, Burundi, DRC, Zambia, Malawi, Zimbabwe and Botswana (Spawls et al. 2018). There are no published records from Cabo Delgado Province but they have been recorded in neighboring provinces (Branch et al. 2005; Conradie et al. 2016; Farooq and Conradie 2015; Portik et al. 2013a; Timberlake et al. 2012). **Comment**: Common throughout the city, where they are very common on buildings.

Gerrhosaurus nigrolineatus Hallowell, 1857

Black-lined Plated Lizard

Fig. 4G

Material: FCN0412, -12.97585˚ 40.57096˚, University Lúrio campus; https://www.inaturalist.org/observations/5089046; https://www.inaturalist.org/observations/21264910; https://www.inaturalist.org/observations/27677035. **Identification:** Long stripped plated lizard. The ear opening is triangular and obvious. Brown above with two predominant yellow and black dorsolateral stripes, flanks irregularly speckled yellow on brown background. **Distribution**: This lizard is widely distributed across East Africa (Uetz et al. 2020). In Mozambique this species has been recorded from Cabo Delgado Province (Pascal 2011), Chimanimani (Broadley (1966) and Tete (Loveridge 1953). **Comment**: According to Bates et al. (2013), the eastern populations currently referred to *G. nigrolineatus* should now be referred to *G. intermedius* Lönnberg 1907. However, since Bates et al. (2013) did not provide a revised diagnosis of *G. intermedius* we retained the species as a synonymy of *G. nigrolineatus* Hallowell, 1857 (Uetz et al. 2020). It is probable that records of G. *flavigularis* from throughout the country should be assigned to *G. nigrolineatus* (=*intermedius*). We retrieved a single record for the lizard *Gerrhosaurus nigrolineatus* from GBIF, which was originally provided from iNaturalist.org (by Joachim on February 11th of 2017) and identified on the basis of a picture where a *Psammophis mossambicus* is seen swallowing the lizard. The lizard was originally identified as *Gerrhosaurus flavigularis*, but we refer to this as *G. nigrolineatus* according to the revised distribution of the genus (Bates et al. 2013).

Dalophia pistillum (Boettger, 1895)

Fig. 4I

Material: FCN1073, -12.97993˚ 40.55888˚, Eduardo Mondlane: **Identification:** A medium to large pink-white bodied specimen of *Dalophia* with a total size of 47.5 mm. **Distribution**: In Mozambique, *D. pistillum* has been recorded throughout most of the country with the most northern record from Lumbo (Loveridge 1920; 1941; Mertens 1922; 1967), **Comment**: The specimen has 287 body annuli, and 24 caudal annuli and was collected in a garden in an urban area of Pemba. The new material from Pemba conforms genetically to published *Dalophia pistillum* sequences. It represents a range extension of 220 km northwards and the northernmost record for the species to date.

Dendroaspis polylepis Günther, 1864

Black Mamba

Material: FCN0002, -12.97585˚ 40.57096˚, Lúrio University Pemba campus. **Identification:** A two meters long slender gray snake with a coffin shaped head with a fairly pronounced brow ridge. The inside of the mouth is black and the eye is medium sized with a round pupil. **Distribution**: This snake is regarded as common in sub-Saharan Africa and has been found as far north as Senegal and as far south as northeast South Africa (Spawls 2010). There were no records for this species in the interpreted distributions by IUCNfrom the northern provinces of Mozambique (Niassa, Cabo Delgado and Nampula) apart from Cabo Delgado Province islands and south of Angoche. This new record falls inside this gap and suggests that populations from Tanzania and southern Mozambique are continuous. **Comment**: This species is of medical importance (Longbottom et al. 2018).

Discussion

All 19 amphibian species recorded from Pemba have been assessed by IUCN and only one of them is under threat, the endangered *Mertensophryne anotis*. Four species were recorded considerably outside of their predicted range: *Schismaderma carens* (250 km east of its known range), *Mertensophryne lindneri* (100 km east of known range), *Mertensophryne anotis* (historically restricted to Chirinda Dorest and Dombe forest, 1000 km south-west, but see species account), and *Leptopelis mossambicus* (250 km northeast of its known range). These range extensions constitute evidence of the absence of records from the region and it is expected that more species will be added to this list as more interest is taken in the region. This study added five new records to the list of amphibians occurring in Cabo Delgado Province, bringing the provincial number up to 37 species (Ohler and Fretey 2014; Conradie et al. 2018, this study).

Of the 34 reptile species recorded from Pemba, only six have been assessed by IUCN (Table 2). Two of them are of conservation importance, being the critically endangered sea turtles, *Chelonia mydas* (Seminoff and (Southwest Fisheries Science Center 2004) and *Eretmochelys imbricate* (Mortimer et al. 2008). Five snakes are of medical importance: *Atractaspis bibronii, Bitis arietans, Dendroaspis angusticeps, Dendroaspis polylepis* and *Dispholidus typus*. All but *B. arietans* are recorded for the first time from Cabo Delgado Province. As there are few published herpetofauna studies from Cabo Delgado Province, most of our records are the first provincial records for this species. In total we added 19 new reptiles for the province, bringing the total reptile species for Cabo Delgado Province to 36 (see Supplementary Table).

Scolecoceps boulengeri was originally described from Lumbo, 220 km south of Pemba by Loveridge (1930). Laurent (1964) recorded additional material from Pemba collected in

September 1948. No precise locality details were provided in those publications, and the species could have been collected outside the official periphery of Pemba as considered in this study. Based on the recent discovery of a sister species, *Scolecoceps broadleyi* further north (Verburgt et al. 2018), these are very specialized species living in weak soil development in coastal woodland. We have failed to record any of the above-mentioned legless skinks from Pemba, but further surveys need to be conducted in suitable habitat.

In recent years three new endemic species of from northeastern coastal Mozambique have been described (e.g. *Hyperolius stictus* – Conradie et al. 2018, *Zygaspis maraisi* – Broadly and Measey 2016, *Scolecoseps broadleyi* – Verburgt et al. 2018). These new species descriptions and the growing species list for Cabo Delgado reinforce lack of surveys in Northern Mozambique and we expect these numbers to grow with results of more surveys.

Final remarks. – The data reported here, in combination with previous records, indicate that a total of at least 19 species of amphibians and 34 species of reptiles occur in Pemba and its immediate surroundings. The fact that the records presented here – reflecting 17 years of intermittent observations – increased the herpetofauna diversity for the province by five amphibians and 19 reptiles strongly indicates that substantial investments should be made in order to better document and protect northern Mozambique's rich but understudied biodiversity. This region, including Pemba, forms part of the larger Coastal Forest of the Eastern Africa Biodiversity Hotspot, but remains one of the most understudied regions in Africa. Our study sheds a small but bright light on this dark spot and emphasizes how biologically diverse this region is. We urge students and scientists around the world to consider contributing to this major challenge of surveying and documenting biodiversity in poorly documented areas.

Acknowledgements. – We thank Bernard Pfeil for constructive comments on this manuscript and Byron Grant for funding Roger Bill's trips to the region. Harith Farooq is supported thanks to financial support to CESAM (UIDP/50017/2020+UIDB/50017/2020) from FCT/MCTES through national funds, and the co-funding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020, by WCS Christensen Conservation Leaders Scholarship, the World Wildlife Foundation – Education for Nature scholarship and the Antonelli Lab. Alexandre Antonelli is supported by the Swedish Research Council (B0569601), the Swedish Foundation for Strategic Research, a Wallenberg Academy Fellowship and the Royal Botanic Gardens, Kew.

We also thank the students at Lúrio University and Pemba citizens for their donations of specimens to the natural history collection at Lúrio University.

References

AmphibiaWeb. https://amphibiaweb.org [accessed 13 Apr 2020.

Atauri JA, de Lucio JV (2001) The role of landscape structure in species richness distribution of birds, amphibians, reptiles and lepidopterans in Mediterranean landscapes. Landscape ecology 16: 147-159. doi:https://doi.org/10.1023/A:1011115921050

Barbour T, Loveridge A (1928) New skinks of the genus Scelotes from Mozambique and Madagascar. In: Proceedings of the New England Zoological Club. 63-65 pp.

Bates MF, Branch WR, Bauer AM, Burger M, Marais J, Alesander G, De Villiers M (2014) Atlas and red list of the reptiles of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, pp.

Bates MF, Tolley KA, Edwards S, Davids Z, Da Silva JM, Branch WR (2013) A molecular phylogeny of the African plated lizards, genus Gerrhosaurus Wiegmann, 1828 (Squamata: Gerrhosauridae), with the description of two new genera. Zootaxa 3750: 465-493

Bauer A, Branch W (1995) Geographic variation in western populations of the Pachydactylus punctatus complex (Reptilia Gekkonidae). Tropical Zoology 8: 69-84

Bauer AM, Childers JL, Broeckhoven C, Mouton PLN (2019) A new Nucras Gray, 1838 (Squamata: Lacertidae) from the Strandveld of the Western Cape, South Africa. Zootaxa 4560: 149-163

Bayliss J, Collins S, Congdon T (2016) A new species of Iolaus Hübner,[1819] subgenus Epamera Druce, 1891 (Lepidoptera: Lycaenidae: Theclinae) from Mts Namuli and Mabu, Northern Mozambique. Metamorphosis 27: 23-30

Bayliss J, Congdon T, Collins S (2018) A new subspecies of Papilio (Princeps) pelodurus Butler, 1896 (Lepidoptera: Papilionidae: Papilioninae) from several mountains in northern Mozambique. Metamorphosis, pp.

Bayliss J, Monteiro J, Fishpool L, Congdon C, Bampton I, Bruessow C, Matimele H, Banze A, Timberlake J (2010) Biodiversity and Conservation of Mount Inago, Mozambique. Report produced under Darwin Initiative Award 15: 036

Bayliss J, Timberlake J, Branch W, Bruessow C, Collins S, Congdon C, Curran M, De Sousa C, Dowsett R, Dowsett-Lemaire F (2014) The discovery, biodiversity and conservation of Mabu forest—the largest medium-altitude rainforest in southern Africa. Oryx 48: 177-185

Blake D (1965) The fourth Umtali museum expedition to Mozambique: November–December, 1964. The Journal of the Herpetological Association of Rhodesia 23: 31-46

Branch B (2012a) Tortoises, terrapins & turtles of Africa. Penguin Random House South Africa, pp.

Branch W (2004) Herpetological Survey of the Niassa Game Reserve. Northern Mozambique, Sociedade para a Gestão e Desenvolvimento da Reserva do Niassa:

Branch W (2012b) Biodiversity and Faunal Assessment, Niassa Green Resources, Lichinga District, Mozambique, Coastal and En- vironmental Services, Grahamstown.:

Branch WR, Bayliss J (2009) A new species of Atheris (Serpentes: Viperidae) from northern Mozambique. Zootaxa 2113: 41-54. doi:https://doi.org/10.5281/zenodo.274877

Branch WR, Bayliss J, Tolley KA (2014) Pygmy chameleons of the Rhampholeon platyceps complex (Squamata: Chamaeleonidae): Description of four new species from isolated 'sky islands' of northern Mozambique. Zootaxa 3814: 1-36

Branch WR, Guyton JA, Schmitz A, Barej MF, Naskrecki P, Farooq H, Verburgt L, Roedel M-O (2017) Description of a new flat gecko (Squamata: Gekkonidae: Afroedura) from Mount Gorongosa, Mozambique. Zootaxa 4324: 142-160

Branch WR, Rodel M, Marais J (2005) Herpetological survey of the Niassa Game Reserve, northern Mozambique-Part I: Reptiles. Salamandra 41: 195

Branch WR, Tolley KA (2010) A new species of chameleon (Sauria: Chamaeleonidae: Nadzikambia) from Mount Mabu, central Mozambique. African Journal of Herpetology 59: 157-172

Branch WR, Verburgt L, Bayliss J, Kucharzewski C, Rödel M, Conradie W (2019) New records of the Large-eyed Green Snake, Philothamnus macrops (Boulenger 1895), from Mozambique. Herpetology Notes 12: 19-29

Broadley D (1964) The second Umtali Museum Expedition to Mozambique 1963–4. The Journal of the Herpetological Association of Rhodesia 22: 13-16

Broadley D (1990) The herpetofaunas of the islands off the coast of south Moçambique. Arnoldia Zimbabwe 9: 469-493

Broadley D, Farooq H (2013) Geographical Distributions: Thelotornis usambaricus. African Herp News 59: 48-50

Broadley D, G (1992a) The savannah species of Kinixys (Testudinidae). The Journal of the Herpetological Association of Africa 40: 12-13

Broadley DG (1962) An expedition to Inhaca Island and South Mozambique. The Journal of the Herpetological Association of Rhodesia 19: 22-25

Broadley DG (1966) The herpetology of south-east Africa.

Broadley DG (1992b) Reptiles and amphibians from the Bazaruto Archipelago, Mozambique. Arnoldia Zimbabwe 9: 539-548

Broadley DG (2002) A review of the species of Psammophis Boie found south of Latitude 12 S (Serpentes: Psammophiinae). African Journal of Herpetology 51: 83-119

Broadley DG, Broadley S (1999) A Review of the African Worm Snakes from South of Latitude 12S° (Serpentes: Leptotyphlopidae). National Museums and Monuments of Zimbabwe, pp.

Broadley DG, Measey GJ (2016) A new species of Zygaspis (Reptilia: Squamata: Amphisbaenidae) from north-eastern Mozambique. African Journal of Herpetology 65: 115-122

Broadley DG, Wallach V (2009) A review of the eastern and southern African blind-snakes (Serpentes: Typhlopidae), excluding Letheobia Cope, with the description of two new genera and a new species. Zootaxa 2255: 1-100

Channing A, Rödel M-O (2019) Field Guide to the Frogs & Other Amphibians of Africa. Penguin Random House South Africa, pp.

Conradie W, Bittencourt-Silva GB, Loader SP, Menegon M, Nanvonamuquitxo C, Scott M (2016) Exploration into the hidden world of Mozambique's sky island forests: new discoveries of reptiles and amphibians. doi:https://doi.org/10.3897/zse.92.9948

Conradie W, Busschau T, Edwards S (2018a) Two new species of Acontias (Acontinae, Scincidae) from the Mpumalanga Highveld escarpment of South Africa. Zootaxa 4429: 89-106

Conradie W, Verburgt L, Portik DM, Ohler A, Bwong BA, Lawson LP (2018b) A new Reed Frog (Hyperoliidae: Hyperolius) from coastal northeastern Mozambique. Zootaxa 4379: 177-198. doi:https://doi.org/10.11646/zootaxa.4379.2.2

Daniels SR, Bayliss J (2012) Neglected refugia of biodiversity: mountainous regions in Mozambique and Malawi yield two novel freshwater crab species (Potamonautidae:

Potamonautes). Zoological Journal of the Linnean Society 164: 498-509. doi:https://doi.org/10.1111/j.1096-3642.2011.00773.x

Du Preez L, Carruthers V (2017) Frogs of Southern Africa. Cape Town: Penguin Random House:

Farooq H, Liedtke HC, Bittencourt-Silva GB, Conradie W, Loader SP (2015) The distribution of Mertensophryne cf. anotis with a new record in Northern Mozambique. Herpetology Notes 8: 305-307

Farooq H, Morgado F, Soares A (2014) Anfíbios e Répteis de Pemba - Amphibians and Reptiles of Pemba. Afrontamento, pp.

Farooq HOM, Conradie W (2015) Second record of a Scolecomorphus kirkii Boulenger, 1883 (Gymnophiona: Scolecomorphidae) for Mozambique. Herpetology Notes 8: 59-62

Freitas ES, Bauer AM, Siler CD, Broadley DG, Jackman TR (2018) Phylogenetic and morphological investigation of the Mochlus afer-sundevallii Species Complex (Squamata: Scincidae) across the arid corridor of sub-Saharan Africa. Molecular phylogenetics evolution:

Garnier J, Hill N, Guissamulo A, Silva I, Witt M, Godley B (2012) Status and community-based conservation of marine turtles in the northern Querimbas Islands (Mozambique). Oryx 46: 359-367

Heinicke MP, Turk D, Bauer AM (2017) Molecular phylogeny reveals strong biogeographic signal and two new species in a Cape Biodiversity Hotspot endemic mini-radiation, the pygmy geckos (Gekkonidae: Goggia). Zootaxa 4312: 449-470

Horner P, Adams M (2007) A molecular-systematic assessment of species boundaries in Australian Cryptoblepharus (Reptilia: Squamata: Scincidae): a case study for the combined use of allozymes and morphology to explore cryptic biodiversity. Museums and Art Galleries of the Northern Territory, pp.

Ihlow F, Farooq HM, Gvoždík V, Hofmeyr MD, Conradie W, Campbell PD, Harvey J, Verburgt L, Fritz U (2019) Geographic range extension of Speke's Hinge-back Tortoise Kinixys spekii Gray, 1863. Amphibian & Reptile Conservation

:

IUCN (2019) The IUCN Red List of Threatened Species. Version 2019-3. http://www.iucnredlist.org. pp.

IUCN SSC Amphibian Specialist Group (2013a) *Mertensophryne lindneri*. The IUCN Red List of Threatened Species 2013. doi:http://dx.doi.org/10.2305/IUCN.UK.2013-2.RLTS.T54692A18371654.en

IUCN SSC Amphibian Specialist Group (2013b) *Schismaderma carens*. The IUCN Red List of Threatened Species 2013. doi:http://dx.doi.org/10.2305/IUCN.UK.2013-2.RLTS.T54885A3021025.en

Jacobsen NH, Kuhn AL, Jackman TR, Bauer AM (2014) A phylogenetic analysis of the southern African gecko genus Afroedura Loveridge (Squamata: Gekkonidae), with the description of nine new species from Limpopo and Mpumalanga provinces of South Africa. Zootaxa 3846: 451-501

Jacobsen NH, Pietersen EW, Pietersen DW (2010) A preliminary herpetological survey of the Vilanculos Coastal Wildlife Sanctuary on the San Sebastian Peninsula, Vilankulo, Mozambique. Herpetology Notes 3: 181-193

Jones S, Clause J, Geeraert L, Jamie G, Sumbane E, Van Berkel T, Jocque M (2017) The Njesi Plateau expedition: a biological assessment of Mt Chitagal, Mt Sanga and the Njesi Plateau in Niassa Province, Mozambique. Biodiversity Inventory for Conservation, pp.

Jones SE, Jamie GA, Sumbane E, Jocque M (2020) The avifauna, conservation and biogeography of the Njesi Highlands in northern Mozambique, with a review of the country's Afromontane birdlife. Ostrich 91: 45-56

Kelly CM, Branch WR, Broadley DG, Barker NP, Villet MH (2011) Molecular systematics of the African snake family Lamprophiidae Fitzinger, 1843 (Serpentes: Elapoidea), with particular focus on the genera Lamprophis Fitzinger 1843 and Mehelya Csiki 1903. Molecular phylogenetics and evolution 58: 415-426

Longbottom J, Shearer FM, Devine M, Alcoba G, Chappuis F, Weiss DJ, Ray SE, Ray N, Warrell DA, de Castañeda RR (2018) Vulnerability to snakebite envenoming: a global mapping of hotspots. The Lancet 392: 673-684

Loveridge A (1920) 11. Notes On East African Lizards Collected 1915–1919, With Description Of A New Genus And Species Of Skink And New Subspecies Of Gecko. In: Proceedings of the Zoological Society of London. Wiley Online Library, 131-167 pp.

Loveridge A (1941) Revision of the African lizards of the family Amphisbaenidae. Museum of Comparative Zoology, pp.

Loveridge A (1953) Zoological Results of a Fifth Expedition to East Africa: Reptiles from Nyasaland and Tete. Museum of Comparative Zoology, pp.

Mertens R (1922) Verzeichnis der Typen in der herpetologischen Sammlung des Senckenbergischen Museums. pp.

Mertens R (1967) Die herpetologische Sektion des Natur-Museums und Forschungs-Institutes Senckenberg in Frankfurt aM nebst einem Verzeichnis ihrer Typen. Senckenberg. Naturforschende Ges., pp.

Monadjem A, Schoeman MC, Reside A, Pio DV, Stoffberg S, Bayliss J, Cotterill F, Curran M, Kopp M, Taylor PJ (2010) A recent inventory of the bats of Mozambique with documentation of seven new species for the country. Acta Chiropterologica 12: 371-391

Mortimer JA, Donnelly M, Group) ISMTS (2008) Eretmochelys imbricata. The IUCN Red List of Threatened Species 2008: e.T8005A12881238. https://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T8005A12881238.en.:

Nielsen SV, Daniels SR, Conradie W, Heinicke MP, Noonan BP (2018) Multilocus phylogenetics in a widespread African anuran lineage (Brevicipitidae: Breviceps) reveals patterns of diversity reflecting geoclimatic change. Journal of Biogeography 45: 2067-2079

Ohler A, Frétey T (2014) Going back to Rovuma: the frog fauna of a coastal dry forest, and a checklist of the amphibians of Mozambique. Journal of East African natural history 103: 73-124. doi:https://doi.org/10.2982/028.103.0203

Pascal O (2011) The Coastal Forests of Northern Mozambique, 2008-2009 expeditions. Our Planet Reviewed» Programme report:

Phelps T (2010) Old world vipers. A natural history of the Azemiopinae and Viperinae: 558

Pianka ER (1967) On lizard species diversity: North American flatland deserts. Ecology 48: 333-351. doi:https://doi.org/10.2307/1932670

Pietersen DW, Haacke WD (2013) First herpetological appraisal of the Parque Nacional de Banhine, Gaza Province, southern Mozambique. Annals of the Ditsong National Museum of Natural History 3: 153-163

Portik DM, Mulungu E, Sequeira D, McEntee J (2013a) Herpetological surveys of the Serra Jeci and Namuli massifs, Mozambique, and an annotated checklist of the southern Afromontane archipelago. Herpetological Review 44: 394-406

Portik DM, Travers SL, Bauer AM, Branch WR (2013b) A new species of Lygodactylus (Squamata: Gekkonidae) endemic to Mount Namuli, an isolated 'sky island'of northern Mozambique. Zootaxa 3710: 415-435. doi:https://doi.org/10.11646/zootaxa.3710.5.2

Poynton J (1966) Amphibia of northern Mozambique. Memoirs de Instituto de Investigação Científica de Moçambique Ser A Ciencas B 8: 13-34

Poynton J, Broadley D (1985a) Amphibia Zambesiaca 2. Ranidae. Annals of the Natal Museum 27: 115-181

Poynton J, Broadley D (1987) Amphibia Zambesiaca 3. Rhacophoridae and Hyperoliidae. Annals of the Natal Museum 28: 161-229

Poynton J, Broadley D (1988) Amphibia Zambesiaca 4. Bufonidae. Annals of the Natal Museum 29: 447-490

Poynton J, Broadley D (1991) Amphibia Zambeziaca 5. Zoogeography. Annals of the Natal Museum 32: 221-277

Poynton JC, Broadley DG (1985b) Amphibia Zambesiaca 1. Scolecomorphidae, Pipidae, Microhylidae, Hemisidae, Arthroleptidae. Annals of the Natal Museum 26: 503-553

Poynton JC, Loader SP, Conradie W, Roedel M-O, Liedtke HC (2016) Designation and description of a neotype of Sclerophrys maculata (Hallowell, 1854), and reinstatement of S. pusilla (Mertens, 1937)(Amphibia: Anura: Bufonidae). Zootaxa 4098: 73-94

Rasplus J, Martinez M, Madogolele N, Duplantier J (2008) Zoological Component. Expedition Report Reconnaissance of Coastal Forests in Cabo Delgado 22nd Nov–13th Dec: 31-41

Rhodin A, Iverson J, Bour R, Fritz U, Georges A, Shaffer H, Van Dijk P (2017) Turtles of the world: annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status. Conservation biology of freshwater turtles and tortoises: a compilation project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group Chelonian Research Monographs 7: 1-292

Seminoff JA, (Southwest Fisheries Science Center US (2004) Chelonia mydas. The IUCN Red List of Threatened Species 2004: e.T4615A11037468. https://dx.doi.org/10.2305/IUCN.UK.2004.RLTS.T4615A11037468.en.

Spawls S (2010) *Dendroaspis polylepis*. The IUCN Red List of Threatened Species 2010. doi:http://dx.doi.org/10.2305/IUCN.UK.2010-4.RLTS.T177584A7461853.en

Spawls S, Howell K, Drewes R, Ashe J (2004) A field guide to the reptiles of East Africa: Kenya, Tanzania, Uganda, Rwanda and Burundi. A & C Black London, pp.

Spawls S, Howell K, Hinkel H, Menegon M (2018) Field guide to East African reptiles. Bloomsbury Publishing, pp.

Taylor PJ, Stoffberg S, Monadjem A, Schoeman MC, Bayliss J, Cotterill FP (2012) Four new bat species (Rhinolophus hildebrandtii complex) reflect Plio-Pleistocene divergence of dwarfs and giants across an Afromontane archipelago. PloS one 7: e41744. doi:https://doi.org/10.1371/journal.pone.0041744

Timberlake J, Bayliss J, Dowsett-Lemaire F, Congdon C, Branch B, Collins S, Curran M, Dowsett RJ, Fishpool L, Francisco J (2012) Mt Mabu, Mozambique: biodiversity and conservation. Report produced under the Darwin Initiative Award 15: 036

Timberlake J, Dowsett-Lemaire F, Bayliss J, Alves T, Baena S, Bento C, Cook K, Francisco J, Harris T, Smith P (2009) Mt Namuli, Mozambique: biodiversity and conservation. Report for Darwin Initiative Award 15: 036

Tolley KA, Alexander GJ, Branch WR, Bowles P, Maritz B (2016) Conservation status and threats for African reptiles. Biological Conservation 204: 63-71

Trape J-F, Mediannikov O (2016) Cinq serpents nouveaux du genre Boaedon Duméril, Bibron & Duméril, 1854 (Serpentes: Lamprophiidae) en Afrique centrale. Bulletin de la Société Herpétologique de France 159: 61-111

Travers SL, Jackman TR, Bauer AM (2014) A molecular phylogeny of Afromontane dwarf geckos (Lygodactylus) reveals a single radiation and increased species diversity in a South African montane center of endemism. Molecular phylogenetics and evolution 80: 31-42

The Reptile Database. http://www.reptile-database.org [accessed 20 Apr 2020.

Vamberger M, Hofmeyr M, Cook C, Netherlands E, Fritz U (2019) Phylogeography of the East African Serrated Hinged Terrapin Pelusios sinuatus (Smith, 1838) and resurrection of Sternothaerus bottegi Boulenger, 1895 as a subspecies of P. sinuatus. Amphibian & Reptile Conservation 13: 42-56

Van Velzen R, Collins S, Brattstrom O, Congdon C (2016) Description of a new Cymothoe Hübner, 1819 from northern Mozambique (Lepidoptera: Nymphalidae: Limenitidinae). Metamorphosis 27: 34-41

Verburgt L, Verburgt UK, Branch WR (2018) A new species of Scolecoseps (Reptilia: Scincidae) from coastal north-eastern Mozambique. African Journal of Herpetology 67: 86-98. doi:https://doi.org/10.1080/21564574.2017.1413014

Weinell JL, Bauer AM (2018) Systematics and phylogeography of the widely distributed African skink Trachylepis varia species complex. Molecular phylogenetics and evolution 120: 103-117

Wilson L, Channing A (2019) A new sand frog from Namaqualand, South Africa (Pyxicephalidae: Tomopterna). Zootaxa 4609: zootaxa. 4609.4602. 4602-zootaxa. 4609.4602. 4602

Table 1. Checklist of amphibians from Pemba, Cabo Delgado Province, Mozambique.

Taxon	**New records from this study (plus SAIAB and PEM)**	**Other Records (eg. iNaturalist)**	**IUCN status**	**New record for Cabo Delgado**
AMPHIBIANS				
Arthroleptidae				
Arthroleptis stenodactylus	X	X	LC	-
Leptopelis mossambicus		X	LC	X
Brevicipitidae				
Breviceps mossambicus	X	X	LC	-
Bufonidae				
Sclerophrys pusilla	X	-	LC	-
Mertensophryne anotis		X	EN	-
Mertensophryne lindneri	X	X	LC	X
Schismaderma carens	X	-	LC	X
Microhylidae				
Phrynomantis bifasciatus	X	X	LC	X
Hemisotidae				
Hemisus marmoratus	X	-	LC	-
Hyperoliidae				
Afrixalus delicatus	X	X	LC	-
Afrixalus fornasini	X	-	LC	-
Hyperolius argus	X	-	LC	X
Hyperolius tuberilinguis	X	X	LC	-
Phlyctimantis maculata	X	-	LC	-
Phrynobatrachidae				
Phrynobatrachus acridoides	X	-	LC	-
Ptychadenidae				

Ptychadena anchietae	X	-	LC	-
Ptychadena mossambica	X	X	LC	-
Pyxicephalidae				
Pyxicephalus edulis	X	-	LC	-
Rhacophoridae				
Chiromantis xerampelina	X	-	LC	-

Table 2. Checklist of reptiles from Pemba, Cabo Delgado Province, Mozambique.

Taxon	**Current study (plus SAIAB and PEM records)**	**Other Records (eg. iNaturalist)**	**IUCN status**	**New record for Cabo Delgado**
REPTILES				
Amphisbaenidae				
Dalophia pistillum	X	-	NE	X
Gekkonidae				
Lygodactylus grotei	X	X	NE	-
Hemidactylus mabouia	X	X	NE	-
Pachydactylus punctatus	X		NE	X
Chondrodactylus turneri	X	-	NE	X
Chamaeleonidae				
Chamaeleo dilepis	X	X	LC	-
Agamidae				
Agama mossambica	X	X	NE	-
Scincidae				
Mochlus sundevalli	X	X	LC	-
Trachylepis margaritifera	X	-	LC	X
Trachylepis striata	X	-	NE	-
Trachylepis varia	X	X	NE	-
Cryptoblepharus africanus	X	-	NE	X
Gerrhosauridae				
Gerrhosaurus nigrolineatus	X	X	NE	-
Varanidae				
Varanus albigularis	X	-	NE	-
Typhlopidae				
Afrotyphlops mucruso	X	X	NE	X
Leptotyphlopidae				

Myriopholis longicauda	X	-	NE	X
Leptotyphlops scutifrons	X	-	NE	X
Pythonidae				
Python natalensis	X	-	NE	X
Atractaspididae				
Atractaspis bibronii	X	-	NE	X
Lamprophiidae				
Boaedon fulginosus-capensis complex	X	-	NE	X
Psammophis angolensis	X	-	NE	-
Psammophis mossambicus	X	-	NE	X
Psammophis orientalis	X	-	NE	-
Rhamphiophis rostratus	X	-	NE	X
Telescopus semiannulatus	X	-	NE	-
Dispholidus typus	X	-	NE	X
Elapidae				
Naja mossambica	X	X	NE	X
Dendroaspis polylepis	X	-	LC	X
Dendroaspis angusticeps	X	-	NE	X
Viperidae				
Bitis arietans	X	-	NE	-
Testudinidae				
Kinixys spekii	X	-	NE	X
Pelomedusidae				
Pelusius sinuatus	X	-	NE	X
Cheloniidae				
Chelonia mydas	X	-	CR	-
Eretmochelys imbricata	X	-	CR	-

Figures:

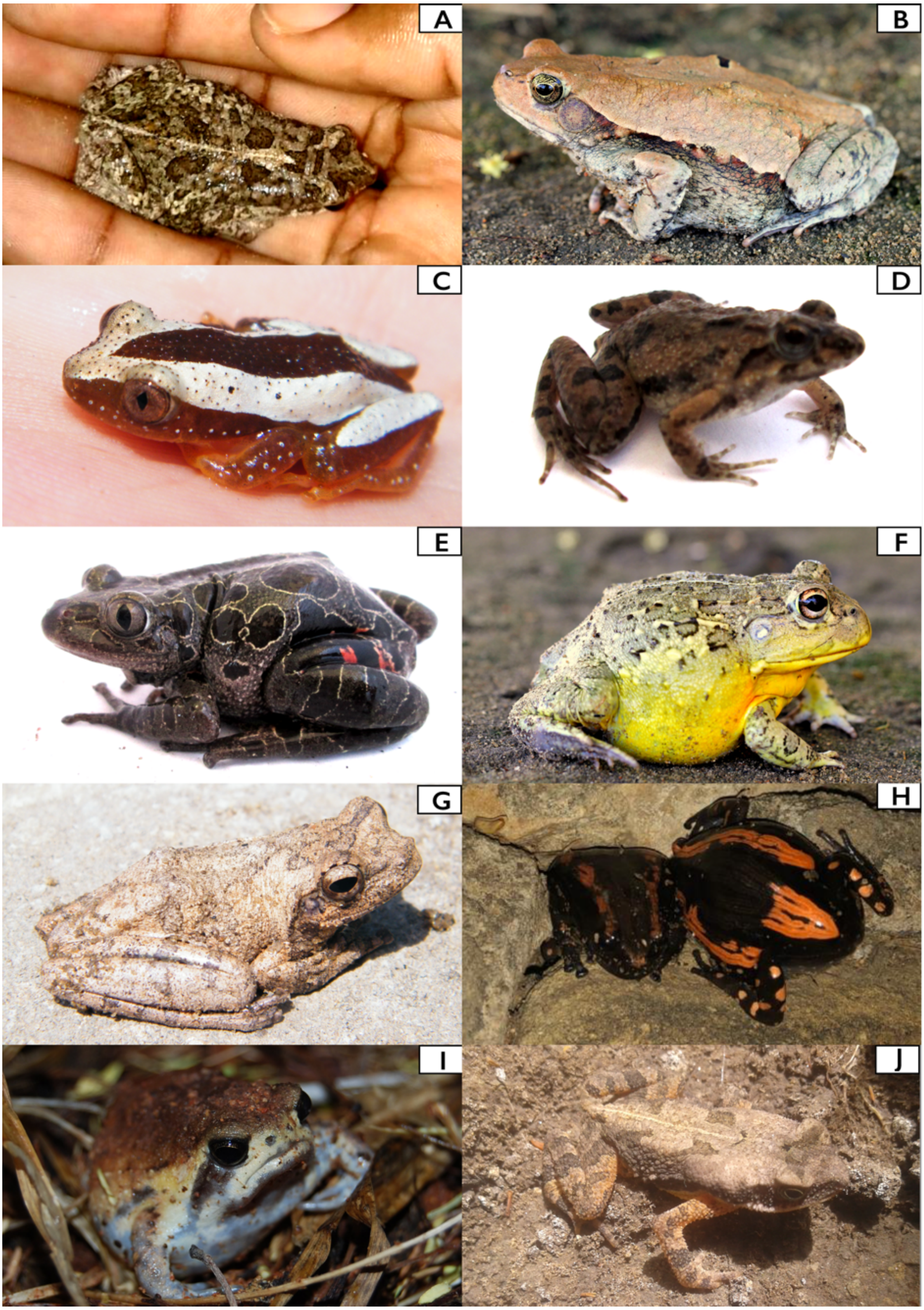


Figure 25: A–Mertensophryne lindneri; B–Schismaderma carens; C–Afrixalus fornasinii; D–Phrynobatrachus acridoides; E–Phlyctimantis maculata; F–Pyxicephalus edulis, G – Chiromantis xerampelina; H–Phrynomantis bifasciatus,I – Breviceps mossambicus; J – Mertensophryne cf anotis.

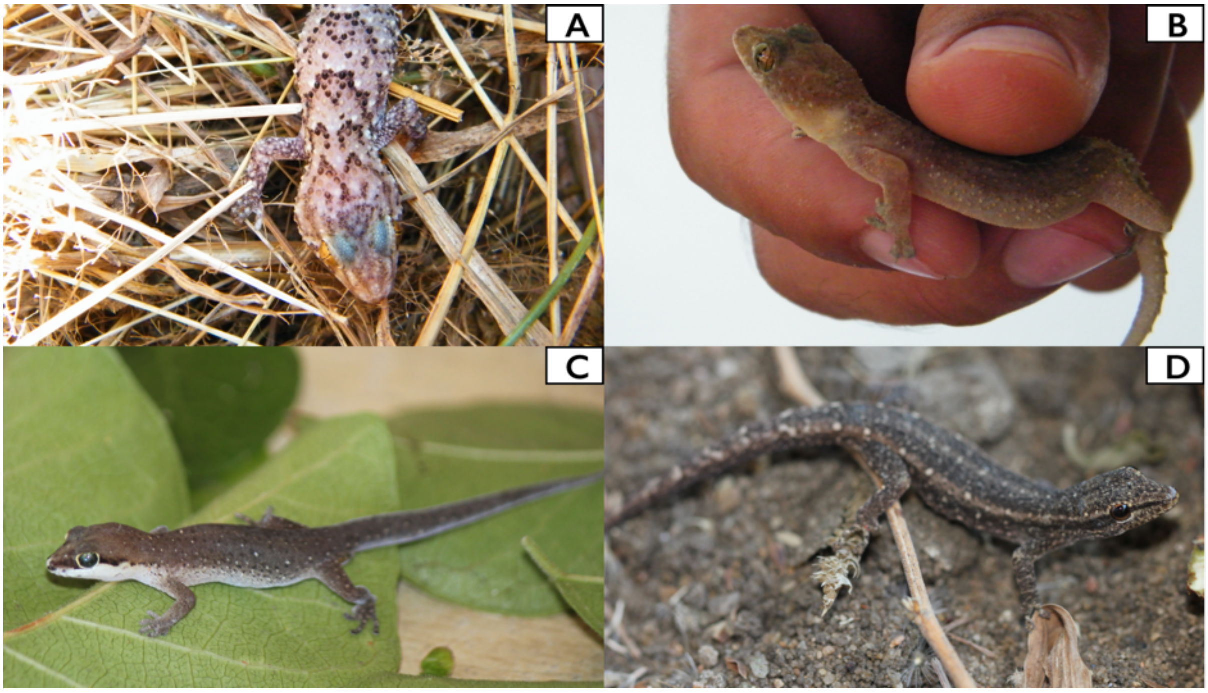


Figure 26: A–Chondrodactylus turnerii; B– Hemidactylus mabouia; C–Pachydactylus puntactus; D– Lygodactylus grotei.

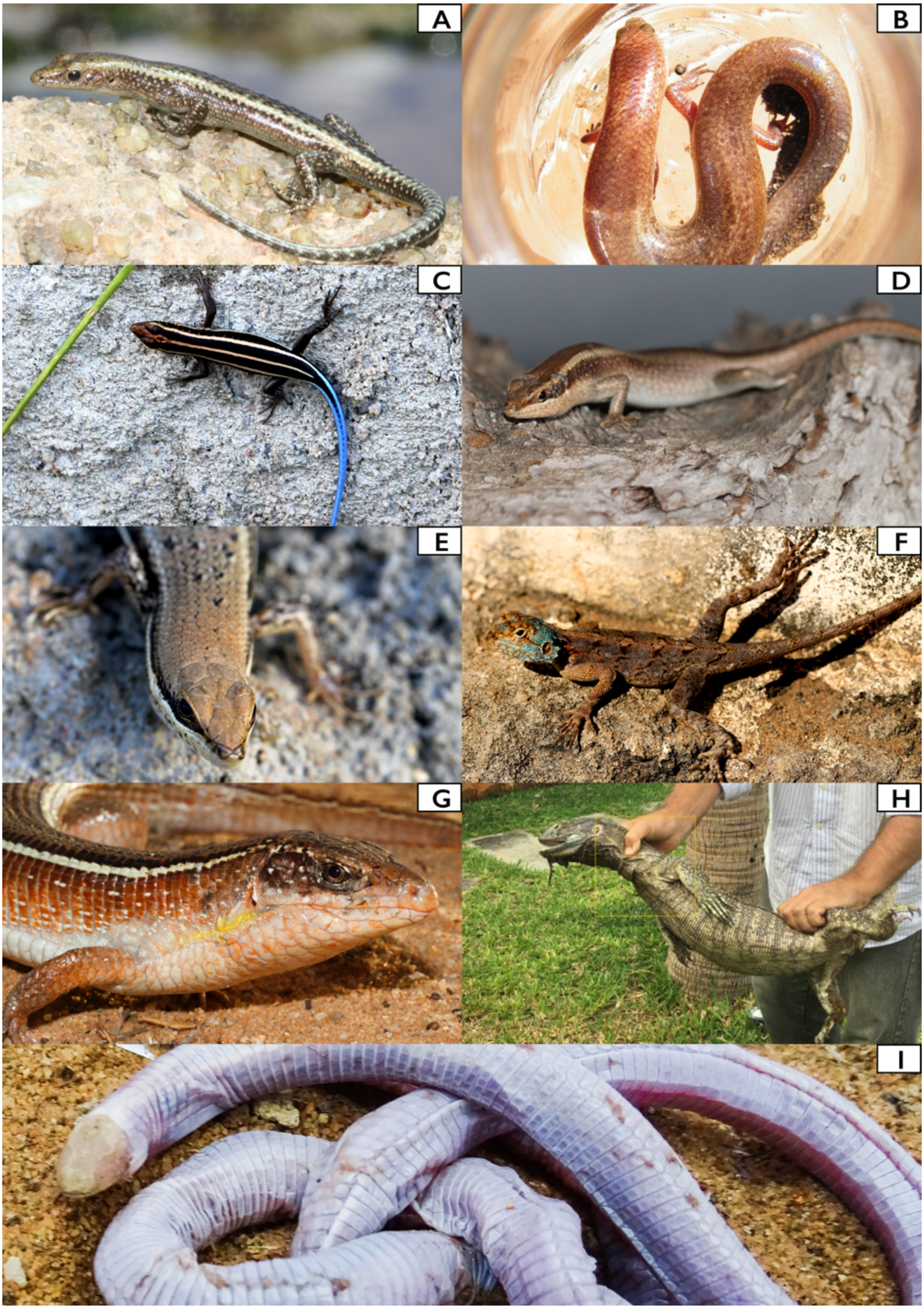

A
B
C
D
E
F
G
H
I

Figure 27: A– Cryptoblepharus africanus; B– Mochlus sundevalli; C– Trachylepis margaritifera; D– Trachylepis striata; E– Trachylepis varia; F– Agama mossambica; G– Gherrosaurus nigrolineatus; H– Varanus albigularis; I– Dalophia pistillum.


A
B
C
D
E
F
G
H
I
J

Figure 28: A– Myriapholis longicaudus; B– Python natalensis; C–Atractaspis bibronii; D– Boedon capensis-fulginosus; E–Psammophis orientalis; F– Ramphiophis rostratus. G– Telescopus semiannulatus, H– Dispholidus typus; I–Naja mossambica; J– Bitis arietans

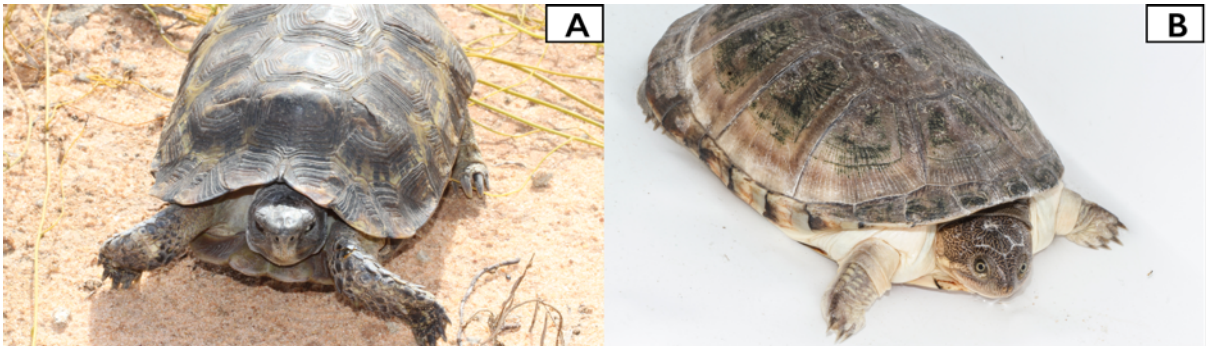


Figure 29: A– Kinixys spekii; B– Pelusius sinuatus.

Supplementary material

Appendix 1: Species accounts

AMPHIBIANS

Arthroleptidae

Arthroleptis stenodactylus Pfeffer, 1893

Common Squeaker

Material: FCN0540, -13.05875˚ 40.52175˚, Muxara (an unpaved area); https://www.inaturalist.org/observations/21458924**. Identification:** Easily identifiable from other *Arthroleptis spp*. due to the presence of a large inner metatarsal tubercle on the heel and darker diamond-shaped hourglass pattern on the back. **Distribution:** This species ranges from coastal Kenya and eastern and southern Tanzania, through Mozambique, Zambia and southern Democratic Republic of the Congo to western Angola, northern Botswana, Zimbabwe and northeastern South Africa (Channing and Rödel 2019). In Mozambique it has been recorded from central and northern parts of the country (Branch 2004; Conradie et al. 2016; Ohler and Frétey 2014; Portik et al. 2013a).

Brevicipitidae

Breviceps mossambicus Peters, 1854

Mozambique Rain Frog

Fig. 2I

Material: SAIAB 88176 (8 specimens), SAIAB 88179 (4 specimens), SAIAB 201111 (9 specimens), -12.96261˚ 40.52944˚, Pemba Beach Hotel (garden). **Identification**: All specimens exhibit the typical coloration of *B. mossambicus* in that it has the conspicuous facial mask with no paravertebral or dorsolateral blotches and with uniform beige dorsum. The large outer metatarsal tubercle is separated from the inner metatarsal by a large clef. **Distribution:** Ranges from northeastern Tanzania south to the KwaZulu-Natal Province of South Africa, through southeastern Democratic Republic of Congo, Zambia, Malawi, Mozambique, Zimbabwe and South Africa (Channing and Rödel 2019). In Mozambique it has been recorded throughout most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1985a). **Comment**: Some of this material was used in a recently phylogenetic study (Nielsen et al. 2018) that assigns this material to the nominal form and documents cryptic diversity further south.

Bufonidae

Sclerophrys pusilla (Mertens, 1937)

Merten's Striped Toad

Material: FCN0104–6, -12.97316° 40.54271°, Wimbe beach, Praia do Wimbe. **Identification:** Typical toad with rough dorsal skin with rounded, wart-like elevations on the back, flattened parotoid glands and granular underside. No red infusions between the thighs. **Distribution:** Used to be part of *S. maculata*, but a recent split restricted the nominal form to West Africa and *S. pussila* for central, eastern and southern Africa (Poynton et al. (2016). In Mozambique it has been recorded from central and northern Mozambique (Branch (2004); Portik et al. (2013a); Conradie et al. (2016); Ohler and Frétey (2014). **Comment:** Common throughout Pemba.

Microhylidae

Phrynomantis bifasciatus (Smith, 1847)

Banded Rubber Frog

Fig. 2H

Material: FCN1464, -12.96954˚ 40.53938˚, Praia do Wimbe; SAIAB 88567 (6 specimens), -12.96942 40.53583, Pemba Beach Hotel. **Identification:** Unmistakable species with a shiny black skin with continuous vivid orange to red bands extending from the snout over the eyelids to the back of the body. **Distribution**: Known from eastern South Africa north through Zimbabwe and Mozambique to southern DRC, Tanzania, southern Kenya and adjacent Uganda (Channing and Rödel 2019). It is expected to occur across the country and has been recorded from most provinces (Poynton and Broadley 1985b) but no records were available from Cabo Delgado Province and this represent the first record for the area. **Comment**: Seen both near wetlands and in urban areas.

Hemisotidae

Hemisus marmoratus (Peters, 1854)

Marbled Shovel-nosed Frog

Material: SAIAB 88556 (1 tadpole), -13.04252˚ 40.55525˚, behind Pemba Marine Institute. **Identification:** Almost spherical body with bloated appearance. Narrow head, with hardened pointed chisel-shaped snout used for digging. Transverse skinfold across top of the head. Back marbled with yellow upper lip and no spots. **Distribution**: Widespread in the savannah zone of sub-Saharan Africa, from Senegal and Gambia, east to western Eritrea, western Ethiopia, and southern Somalia, thence south to the northern and northeastern parts of South Africa, eastern and northern Botswana, northeastern Namibia, and Angola (Channing and Rödel 2019) and has been recorded from throughout the country (Ohler and Frétey 2014; Portik et al. 2013a; Poynton and Broadley 1985b). **Comment**: Juveniles were collected from pitfall traps near the beach in a sandy grassland, but not retained as vouchers.

Hyperoliidae

Afrixalus delicatus Pickersgill, 1984

Delicate Leaf-folding Frog

Material: FCN0532, -12.97383˚ 40.55138˚, Bairro Eduardo Mondlane; SAIAB 88563 (6 specimens), -12.96942˚ 40.53583˚, Pemba Beach Hotel. **Identification:** Small *Afrixalus* with a prominent broad, dark brown, vertebral line and a band across full width of the tibia. **Distribution**: This species ranges from KwaZulu-Natal (eastern South Africa) northwards through the coastal belt of Mozambique, the lower altitude of Malawi to Tanzania, southeastern Kenya, and southern Somalia (AmphibiaWeb 2020). Poynton and Broadley (1987), Portik et al. (2013a) and Conradie et al. (2016) have recorded the species from central and northern Mozambique. It has been collected in Cabo Delgado Province under the name of *Afrixalus brachynemis* (Ohler and Frétey 2014). **Comment**: Records of *Afrixalus brachynemis* from northern Mozambique according to Pickersgill (2005) should be ascribed to *Afrixalus delicatus.*

Afrixalus fornasini (Bianconi, 1849)

Great Leaf-folding Frog

Fig. 2C

Material: FCN0541, -13.05875˚ 40.5217˚; SAIAB 88562 (5 specimens), -12.96942˚ 40.53583˚, Pemba Beach Hotel; SAIAB 88556 (10 tadpoles), -13.04252˚ 40.55525˚, behind Pemba Marine Institute. https://www.inaturalist.org/observations/19756407**.** **Identification:** Large *Afrixalus*, with body covered with asperities. Has a prominent broad, dark brown, vertebral band terminating in a point between eyes. **Distribution**: This species ranges from coastal Kenya southward through eastern and southern Tanzania, Malawi, Mozambique and eastern Zimbabwe to coastal KwaZulu-Natal in South Africa (IUCN 2019). Portik et al. (2013a), Conradie et al. (2016) and Ohler and Frétey (2014) have recorded the species from central and northern Mozambique.

Hyperolius argus Peters, 1854

Argus Reed Frog

Material: FCN1030, -12.97322˚ 40.5518˚, Bairro Eduardo Mondlane. **Identification:** A large *Hyperolius*, exhibiting large round light spots on a brown/dark red background. **Distribution**: This species is known from the eastern side of the African continent ranging from extreme southern Somalia, south through Kenya, Tanzania and Mozambique to KwaZulu-Natal Province in South Africa, and inland to southern Malawi and extreme south-eastern Zimbabwe (Channing and Rödel 2019). In Mozambique, Poynton and Broadley (1987) recorded the species from central Mozambique in the Nampula region. **Comment**: This is the first published record for Cabo Delgado Province.

Hyperolius tuberilinguis Smith, 1849

Tinker Reed Frog

Material: SAIAB 88565 (1 specimens), -12.96942˚ 40.53583˚, Pemba Beach Hotel; SAIAB 88559 (1 tadpole), -13.04252˚ 40.55525˚, behind Pemba Marine Institute. **Identification**: Large uniform green to yellow *Hyperlolius* with the inner thighs with red. **Distribution**: This species ranges from south-central Kenya, through eastern and southern Tanzania, Malawi, Mozambique and eastern Zimbabwe, to Eswatini to KwaZulu-Natal Province in South Africa (Channing and Rödel 2019). In Mozambique it has been recorded from most of the country including Cabo Delgado (Ohler and Frétey 2014).

Phlyctimantis maculata (Duméril, 1853)

Red-legged Kassina

Fig. 2E

Material: FCN0542, -13.05875˚ 40.52175˚, Muxara; FCN0528, -12.97383˚ 40.55138˚, Bairro Eduardo Mondlane; SAIAB 88564 (2 specimens), -12.96942˚ 40.53583˚, Pemba Beach Hotel; SAIAB 185923, -13.07606˚ 40.54589˚, Cidade de Pemba. **Identification:** A large *Kassina* with large dark spots with a pale outline on an olive background. It has red markings in the groin and armpits. **Distribution**: This species ranges from coastal Kenya south through Tanzania and Mozambique, and inland to southern Malawi, eastern Zimbabwe, to KwaZulu-Natal Province South Africa and eastern Eswatini (Channing and Rödel 2019). In Mozambique it has been recorded from most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1987). **Comment**: Collected from water holes in the city.

Phrynobatrachidae

Phrynobatrachus acridoides (Cope, 1867)

Eastern Puddle Frog

Fig. 2D

Material: FCN0529, FCN0543–7, -13.05875˚ 40.52175˚, Muxara; SAIAB 88557 (3 specimens), -13.04252˚ 40.55525˚, behind Pemba Marine Institute; SAIAB 88156 (1 specimens), SAIAB 186322 (1 specimen) , -13.06872˚ 40.55044, near Pemba Marine Institute; SAIAB 88160, -13.07147˚ 40.51956˚, roadside pool on N1; SAIAB 88572 (6 specimens),-13.08914˚ 40.54458˚, wetland near Atolo; SAIAB 88148 (4 specimens), -13.08917˚ 40.54472˚, wetland near Atolo. **Identification:** Easily identifiable by a pair of longitudinal chevron shaped ridges from behind the eye over the shoulder region. The fingers and toes lack webbing and have enlarged tips. **Distribution**: This species ranges from southern Somalia and Kenya, south to eastern Zimbabwe, Mozambique and northern KwaZulu-Natal Province in South Africa (Channing and Rödel 2019). In Mozambique it has been recorded throughout most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1985a). **Comment**: Common near water wholes throughout the city

Ptychadenidae

Ptychadena anchietae (Bocage, 1868)

Plain Grass Frog

Material: SAIAB 88570 (2 specimens) and SAIAB 88574 (4 specimens), -13.08914˚ 40.54458˚, wetland near Atolo. **Identification:** A medium sized *Ptychadena* identifiable by a lighter top of snout, forming a clearly defined triangular patch. **Distribution**: This species ranges from Eritrea, Ethiopia, Djibouti and Somalia, south through East Africa to South Africa and Botswana, and west to Angola and southern Congo (Channing and Rödel 2019). In Mozambique it has been recorded throughout most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1985a). **Comment**: Seen calling at the airport parking area.

Ptychadena mossambica (Peters, 1854)

Mozambique ridged frog

Material: SAIAB 88566 (3 specimens), -12.96942˚ 40.53583˚, Pemba Beach Hotel; https://www.inaturalist.org/observations/43584403; https://www.inaturalist.org/observations/19756360. **Identification:** Medium to large sized *Ptychadena* with a light dorsal band from snout tip to vent present. **Distribution:** This species ranges from southern Kenya, through Tanzania, to Malawi, Zambia, the Zambezi Strip of Namibia, northern and eastern Botswana, Zimbabwe, Mozambique, northern and northeastern South Africa, and Eswatini (Channing and Rödel 2019). In Mozambique it has been recorded throughout most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1985a).

Pyxicephalidae

Pyxicephalus edulis Peters, 1854

Edible Bullfrog

Fig. 2F

Material: FCN0534, -12.97392˚ 40.53827˚, Josina Machel behind Shoprite; FCN0530, FCN0536–39, -12.97383˚ 40.55138˚, Eduardo Mondlane; FCN0046, FCN0090, FCN0103, FCN505, -12.97316˚ 40.54271˚, Wimbe beach; SAIAB 88568 (1 tadpole) and SAIAB 88037 (4 specimens), -12.96942˚ 40.53583˚, Pemba Beach Hotel; SAIAB 88561 (19 tadpoles), -12.97241˚ 40. 54694˚, pools behind Pemba Beach Hotel; SAIAB 185947 (49 tadpoles), -13.06872˚ 40.55044˚, near Atolo; SAIAB 88150 (7 tadpoles), -13.08033˚ 40.54530˚, near Atolo.; SAIAB 88158 (11 tadpoles); https://www.inaturalist.org/observations/19756319. **Identification:** Large sized frog, easily identified by the presence of a white elliptical shaped spot on tympanum. **Distribution**: This species ranges from southern Somalia, Kenya and Tanzania, south to Mozambique, southern Malawi, southern Zambia, Zimbabwe, Botswana, northern and eastern South Africa and Eswatini (IUCN 2019). In Mozambique it has been recorded throughout most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1985a). **Comment:** One of the most common frogs in Pemba, mostly near Wimbe beach where they hatch in a wetland and cross the road and end up on the beach in the thousands.

Rhacophoridae

Chiromantis xerampelina (Peters, 1854)

Foam Nest Frog

Fig. 2G

Material: FCN0091, FCN0098, -12.97316˚ 40.54271˚, Wimbe beach; SAIAB 88149 (5 specimens), -13.08033˚ 40.54530˚, near Atolo; SAIAB 88569 (1 specimen) and SAIAB

88573 (46 tadpoles), -13.08914˚ 40.54458˚, near Atolo; SAIAB 88028 (~100 tadpoles), -13.08917˚ 40.54472˚, near Atolo; https://www.inaturalist.org/observations/30640966; https://www.inaturalist.org/observations/21264464**.** **Identification:** Unmistakable gray frog, easily identifiable by having the fingers in opposing pairs with the outer pair better developed. **Distribution**: This species ranges from coastal Kenya south to northern KwaZulu-Natal Province in South Africa, and inland as far west as eastern Botswana and west-central Angola. It is present on the island of Zanzibar in Tanzania (Channing and Rödel 2019). In Mozambique it has been recorded throughout most of the country including Cabo Delgado Province (Ohler and Frétey 2014; Poynton and Broadley 1985a). **Comment**: Very common in Pemba throughout the city, where they lay their foam nests above water level in overhanging vegetation. Their skin become darker or lighter according to light exposure.

REPTILES

Gekkonidae

Chondrodactylus turneri (Gray, 1864)

Turner's Thick-toed Gecko

Fig. 3A

Material: FCN0045, -12.97585˚ 40.57096˚, Lúrio University Pemba Campus. **Identification:** Rough gecko with a rounded snout. The nostrils are pointed upwards and has a prominent eye with golden-brown iris and a vertical pupil. The toe tips are dilated. **Distribution**: Widely distributed across Southern Africa, extending southwards from Kenya to South Africa and westwards to Namibia and Angola. In Mozambique there are no published records from Cabo Delgado, but the species has been recorded from neighboring provinces (Branch et al. 2005; Conradie et al. 2016).

Hemidactylus mabouia (Moreau de Jonnès, 1818)

Tropical House Gecko

Fig. 3B

Material: FCN0069, -12.97585° 40.57096°, Lúrio University Campus; FCN0208–29, -12.97316° 40.54271°, Wimbe beach; PEM R19735, -12.96350° 40.52981°, Pemba Beach Hotel; PEM R19855–6, -12.96639° 40.52972°, Pemba Beach Hotel. **Identification:** Common and easily identifiable genus by the presence of claws. Very similar to *H. platycephalus* but can be distinguished due to smaller size and absence of enlarged tubercles on upper part of legs. **Distribution**: Has a wide distribution, occurring throughout most of the African continent and are also present in the Americas as an introduced species (Uetz et al. 2020). In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch 2012b; Pascal 2011; Portik et al. 2013a). **Comment**: One of the most common geckos in Pemba found in most households.

Lygodactylus grotei (Sternfeld, 1911)

Common Dwarf Gecko

Fig 3. D

Material: FCN1452, FCN1454, FCN1455, -12.95977° 40.50172°, Ingonhane, PEM R19736, -12.9635° 40.52981°, Pemba Beach Hotel, PEM R21943–4, -13.02675° 40.53908°, Kaia Village Hotel. **Identification**: Common and unmistakable small gray-brown dwarf gecko. **Distribution**: Northern Mozambique to Tanzania. In Mozambique it has been recorded from Cabo Delgado Province (Broadley and Farooq 2013; Pascal 2011), Zambézia Province (Portik et al. 2013a), Nampula Province (Conradie et al. 2016) and Niassa Province (Branch et al. 2005). **Comment:** One of the most common geckos in Pemba and present throughout the city.

Pachydactylus punctatus Peters, 1854

Speckled-lipped Thick-toed Gecko

Fig. 3C

Material: FCN1465, -12.98018˚, 40.55912˚, Eduardo Mondlane. **Identification:** Small gecko, with small dark punctations on the dorsum and granular scales, which become more imbricate on the tail. Easily identifiable by their prominent eyes with a blue iris. **Distribution**: Wide distribution, occurring in most of Southern Africa. In Mozambique there are no published records from Cabo Delgado Province but they have been recorded in neighboring provinces (Bauer and Branch 1995; Branch et al. 2005). **Comment**: The species was initially described from Tete and Sena on the Zambezi River in central Mozambique (Peters, 1854). This represent the most northeastern record of the species.

Chamaeleonidae

Chamaeleo dilepis Leach, 1819

Flap-necked Chameleon

Material: FCN0057, -12.97585˚ 40.57096˚, Lúrio University Campus. **Identification:** Big chameleon, easily identifiable by its large moveable ear flaps. Common throughout the less urbanized parts of the city. **Distribution**: This species is widely distributed throughout southern and eastern Africa. In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch 2012b; Branch et al. 2005; Conradie et al. 2016; Farooq and Conradie 2015; Pascal 2011; Portik et al. 2013a). **Comment:** There is a local myth that this species is highly venomous which may lead to it being actively killed by local people.

Agamidae

Agama mossambica Peters, 1854

Mozambican Agama

Fig. 4F

Material: FCN0032, -12.96379˚ 40.50011˚, Cimento; FCN0207, FCN0031, -12.97316˚ 40.54271˚, Wimbe beach. https://www.inaturalist.org/observations/30637436. **Identification:** Large and common cosmopolitan agama with flattened body, head is short and pointed and eye fairly large and high on the head. **Distribution**: This species occurs from Tanzania south through Malawi, eastern Zambia, eastern Zimbabwe and Mozambique (Uetz et al. 2020). In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch 2012b; Farooq and Conradie 2015; Pascal 2011; Timberlake et al. 2012). **Comment**: Common throughout the city.

Scincidae

Trachylepis striata (Peters, 1844)

Striped Skink

Fig. 4D

Material: FCN0526, -12.97383˚ 40.55139˚: **Identification:** A robust striped skink with head and body slightly depressed. The body is brown with two cream dorsolateral stripes. **Distribution**: Widely distributed species occurring throughout Sub-Saharan Africa (Uetz et al. 2020). In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch 2012b; Branch et al. 2005; Farooq et al. 2015; Pascal 2011; Portik et al. 2013a; Timberlake et al. 2009). **Comment**: Common throughout the city.

Trachylepis varia (Peters, 1867)

Variable Skink

Fig. 4E

Material: FCN0037, FCN0041, FCN0050, FCN0055, -12.97585˚ 40.57096˚, Lúrio University Campus; FCN0557, -13.05833˚ 40.51767˚, Muxara, PEM R21950, -13.026750˚ 40.53908˚, Kaia Village Hotel. **Identification:** A medium sized skink, with a slightly depressed body. The back is speckled brown, with white flank stripe extending from the upper lip to the hind limbs. **Distribution**: According to Weinell and Bauer (2018), *T. varia* occurs in eastern South Africa, eastern and southern Zimbabwe, Malawi, and Mozambique; possibly southeastern Botswana. In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch 2012b; Branch et al. 2005; Conradie et al. 2016; Farooq and Conradie 2015; Pascal 2011; Portik et al. 2013a; Timberlake et al. 2009). **Comment**: Common throughout the city.

Cryptoblepharus africanus Mertens, 1928

Coral Rag Skink

Fig. 4A

Material: FCN0044, -12.96675˚ 40.55425˚, Eduardo Mondlane (beach), PEM R05958, PEM R05960, PEM R16230, PEM R16231, PEM R16232, PEM R16233, PEM R16234–8, -12.97111˚ 40.5389˚, coral rags at Pemba Beach Hotel. **Identification:** A small skink with well-developed limbs, five long clawed toes and immoveable eyelids, easily identifiable by having a blackish-bronze dorsum, with two speckled gold dorsolateral stripes and occurring on the coral rag on the beach. **Distribution**: Occurring along the coastline from Somalia south through Kenya, Tanzania, Mafia Island, Mozambique to KwaZulu-Natal, South Africa (Uetz et al. 2020). There are no published records from Cabo Delgado Province but they have been collected from elsewhere in Mozambique (Blake 1965; Horner and Adams 2007). **Comment**: Common on the coral rag on the beach.

Varanidae

Varanus albigularis Daudin, 1802

Rock Monitor

Fig. 4F

Material: Sight records only. **Identification**: A big heavily built gray monitor lizard. Easily identifiable by the presence of a bulbous snout with a big nostril. **Distribution**: Southern Africa, northwards to Angola, Zambia and Mozambique. In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch et al. 2005; Broadley 1990; Pascal 2011). **Comment:** Not a commonly observed species in Pemba. Our records are based on a video of a *Varanus albigularis* underwater next to the mangroves at Pemba Bush Camp; a picture of an adult in Chuiba; and a newborn caught and released at the Lúrio University campus.

Typhlopidae

Afrotyphlops mucruso (Peters, 1854)

Zambezi Blind Snake

Material: FCN0406, FCN0520, -13.05875˚ 40.52175˚, Muxara (an unpaved area of the city). https://www.inaturalist.org/observations/19756612**.** **Identification:** Dark gray blind snake with a prominent snout that is obtusely angled and keratinized edge to the rostrals. **Distribution**: The species has a wide distribution, occurring in most of southern and central Africa (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but they have been recorded in neighboring provinces (Branch et al. 2005; Broadley and Wallach 2009). **Comment:** This is one of the largest species of blind snakes known and was previously known as *Megatyphlops* (Broadley and Wallach 2009)

Leptotyphlipidae

Myriopholis longicauda (Peters, 1854)

Long-tailed Thread Snake

Fig. 5A

Material: Photo. **Identification:** A thin pink cylindrical worm snake with a tail with a conical terminal spine and the eye is a tiny black dot. **Distribution**: Known from Somalia, Mozambique, South Africa, Zimbabwe, Botswana and Zambia (Uetz et al. 2020). In Mozambique the species has been collected throughout the country (Branch 2012b; Pietersen and Haacke 2013; Verburgt et al. 2018). **Comments:** This represents the northernmost record of this species for Mozambique.

Leptotyphlops scutifrons (Peters, 1854)

Peter's Thread Snake

Material: No voucher; https://www.inaturalist.org/observations/25668764**.** **Identification:** A thin black worm snake with a cylindrical body and ending with a short terminal spine. **Distribution**: Widely distribution across most of southern Africa (Uetz et al. 2020**).** In Mozambique all records are restricted to south of the Zambezi River (Broadley and Broadley 1999). **Comments:** Common after the rains when sandy soils get saturated with water and individuals of this species are forced to the surface.

Pythonidae

Python natalensis Smith, 1833

Southern African Rock Python

Fig. 5B

Material: Seen at Murrebue beach being sold by locals (-13.12869° 40.55419°). **Identification**: Huge snake with sub triangular head and rounded snout. The body is well patterned with a mix of brown, tan, yellow and gray blotches. **Distribution**: Wide distribution across southern, central and eastern Africa (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but they have been recorded throughout the country (Branch et al. 2005; Broadley 1990; Jacobsen et al. 2010). **Comments**: This is a protected species in Mozambique and listed under CITES appendix II.

Atractaspididae

Atractaspis bibronii Smith, 1849

Bibron’s Stiletto Snake

Fig. 5C

Material: FCN0024, -12.96593˚ 40.58151˚, Maringanha; FCN0523, -12.97483˚ 40.57069˚, Lúrio University Campus. **Identification:** A slim dark snake with a prominent snout and with a cylindrical body with gray with purplish sheen on scales. **Distribution**: Wide distribution across southern, central and eastern Africa (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but it has been collected throughout most of the country (Broadley 1964; 1990; Broadley 1962; 1992b; Jacobsen et al. 2010). **Comment**: Unmistakable distinctive response when disturbed by shaking the head sideways in an attempt to stab the attacker with its protruding fangs. This species is of medical importance (Longbottom et al. 2018).

Lamprophiidae

Boaedon fuliginosus-capensis complex

Brown House Snake

Fig. 5D

Material: FCN0004, -12.97296˚ 40.54156˚, Wimbe beach; FCN0013, FCN0016–17, FCN0025, FCN0233, FCN0236, -12.97585˚ 40.57096˚, Lúrio University Campus; https://www.inaturalist.org/observations/22155469. **Identification:** Has a subtriangular head and vertical pupil with a yellow iris. The body is brown with a pair of pale lines each side of the head, one through the eye and one trough the cheek. **Distribution**: Occurring through most of western, southern and eastern Africa (Uetz et al. 2020). In Mozambique the only published records from Cabo Delgado Province were recorded in Vamizi Island (Broadley and Farooq 2013) and in the Rovuma river (Pascal 2011). They have also been recorded in neighboring provinces (Branch et al. 2005; Conradie et al. 2016; Farooq and Conradie 2015; Portik et al. 2013a). **Comments**: The taxonomy of the *Boaedon fuliginosus-capensis* complex is still unresolved. Recent studies have restricted true *Boaedon fuliginosus* to West Africa, but most of the rest of range comprise a complex of cryptic species with complicated taxonomic history (Kelly et al. 2011; Trape and Mediannikov 2016). One of the most common snakes in the city.

Psammophis angolensis (Bocage, 1872)

Dwarf Sand Snake

Material: University Lúrio campus (sight record only). **Identification**: Striped sand snake. The top of the head is black with three fine white crossbars between the eye and the nape. Broad dark brown, finely yellow-edged vertebral stripe. **Distribution**: Wide distribution across southern, central and eastern Africa (Uetz et al. 2020). No published records exist for Cabo Delgado Province, but it has been collected throughout the country (Broadley 2002).

Psammophis mossambicus Peters, 1882

Olive Grass Snake

Material: FCN0522, -12.97483˚ 40.57069˚, Lúrio University Campus; https://www.inaturalist.org/observations/5089048. **Identification:** A brown sand snake with a cylindrical and muscular body and a rounded deep snout. **Distribution**: Wide distribution extending from Tanzania southwards South Africa and inland reaching Congo, Namibia and Angola (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but they have been recorded throughout the country (Branch et al. 2005; Broadley 2002; Portik et al. 2013a). **Comment:** Common in the areas of Eduardo Mondlane and Chuiba.

Psammophis orientalis Broadley, 1977

Eastern Stripe-bellied Sand Snake

Fig. 5E

Material: FCN0286, -12.97585˚ 40.57096˚, Lúrio University Campus; FCN0524, -12.97483˚ 40.57069˚, Lúrio University Campus. **Identification**: A medium sized brown snake with pointed snout. The belly has two broad black lines with yellow between them and white on the outer edges. **Distribution**: East African species occurring from Kenya, southwards to Mozambique and inland reaching Malawi, and Zimbabwe (Uetz et al. 2020). In Mozambique the only published records from Cabo Delgado Province were recorded in Vamizi Island (Broadley and Farooq 2013) and in the Rovuma river (Pascal 2011). They have also been recorded throughout the country (Branch et al. 2005; Broadley 2002; Conradie et al. 2016; Farooq and Conradie 2015; Portik et al. 2013a). **Comments**: Common in the areas of Eduardo Mondlane and Chuiba. This is a very fast-moving snake and is often very hard to catch.

Rhamphiophis rostratus Peters, 1854

Rufous Beaked Snake

Fig. 5F

Material: FCN0239, FCN0224, FCN0230, -12.97585˚ 40.57096˚, Lúrio University Campus. **Identification:** The body is white-gray with dark speckling. The head is short with a pointed snout and a large eye with a broad dark line through it. **Distribution**: Wide distribution in Africa, mostly in eastern and southern Africa, but also occurring in Namibia and DRC (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but they have been recorded in neighboring provinces such as Niassa (Branch et al. 2005). **Comment**: One of the most observed snakes at the Lurio University Pemba campus.

Colubridae

Telescopus semiannulatus Smith, 1849

Common Tiger Snake

Fig. 5G

Material: FCN006, FCN0021, -13.05875˚ 40.52175˚, near Pemba Airport. **Identification:** Unmistakable snake with a pale orange dorsum with black cross bars along the back and tail. **Distribution**: Wide distribution across southern, central and eastern Africa (Uetz et al. 2020). In Mozambique it has been recorded in Cabo Delgado and neighboring provinces (Branch 2012b; Pascal 2011).

Dispholidus typus (Smith, 1828)

Boomslang

Fig. 5H

Material: FCN0521, -12.97483˚ 40.57069˚, Lúrio University campus. **Identification:** Pupil is almost round but elongated at the front. **Distribution**: The species occurs in most of Sub-Saharan Africa (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but they have been recorded in neighboring provinces (Branch et al. 2005; Conradie et al. 2016). **Comment:** Eimermacher (2013) have shown deep genetic divergence among *Dispholidus* species and treated northern populations as *Dispholidus typus viridis*. This species is of medical importance (Longbottom et al. 2018). It often inflates its neck when disturbed.

Elapidae

Naja mossambica Peters, 1854

Mozambique Spitting Cobra

Fig. 5I

Material: FCN0001, -12.97316˚ 40.54271˚, Wimbe beach, garden near Wimbe-beach; https://www.inaturalist.org/observations/37406859. **Identification:** A cobra with a brown back and an underside pinkish. The throat and anterior third of the belly have a mixture of bars and blotches in black. **Distribution**: Wide distribution occurring in most of southern, central and eastern Africa (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province but they have been recorded in neighboring provinces (Branch et al. 2005; Conradie et al. 2016). **Comment**: This species is of medical importance (Longbottom et al. 2018).

Dendroaspis angusticeps (Smith, 1849)

Green Mamba

Material: PEM R5964, -12.97111˚ 40.53889˚, Pemba Beach Hotel. **Identification:** Bright green snake with coffin shape head and the inside of the mouth is black. **Distribution**:

Common in most of East Africa, southwards from Kenya to South Africa and west to Angola and DRC (Uetz et al. 2020). In Mozambique there are no published records from Cabo Delgado Province nor in surrounding provinces. **Comment:** This species is of medical importance (Longbottom et al. 2018).

Viperidae

Bitis arietans Merrem, 1820

Puff Adder

Fig. 5J

Material: FCN0003, -12.97585˚ 40.57096˚, Lúrio University Pembs campus. **Identification:** A large brown viper with a serious of dark v-shapes chevrons along the back. It has a thin neck, the body is fat and depressed with a fairly short tail. Has a broad flat triangular head with a small eye and a vertical pupil. **Distribution**: Extends almost continuously throughout sub-Saharan Africa, from the southern border of the Sahara southwards to the South African coast, excluding equatorial rainforest regions of Central and West Africa, hyper-arid regions of the Namib Desert and some high-altitude zones, such as the Eastern Arc Mountains in Tanzania and the Drakensberg Mountains in South Africa (Phelps 2010; Spawls et al. 2004). In Mozambique it has been recorded in Cabo Delgado Province and neighboring provinces (Branch et al. 2005; Conradie et al. 2016; Farooq and Conradie 2015; Pascal 2011). **Comment**: This species is of medical importance (Longbottom et al. 2018).

Testudinidae

*Kinixys spekii (*Gray, 1863)

Speke's Hinge-back Tortoise

Fig. 6A

Material: Photograph taken in Chuiba area. **Identification**: Small hinged tortoise with a flattened shell with yellow scales with darker margins. The head is rounded and the beak is a single cusp. The plastron is tan, with black edging. **Distribution**: Widely distributed range across sub-Saharan Africa (Broadley 1992a; Rhodin et al. 2017). **Comment**: Recently, Ihlow et al. (2019) through molecular analysis enlarged the known geographical distribution range of *K. spekii* to coastal northern Mozambique, but also demonstrated that the species' distribution range is still not properly known.

Pelomedusidae

Pelusios sinuatus (Smith, 1838)

Serrated Hinged Terrapin

Fig. 6B

Material: PEM R19857–8, -13.08931˚, 40.54467˚, near Atolo; PEM R19741, -13.09858˚ 40.54539˚, PEM R19742–4, -13.08914˚ 40.54458˚, near Atolo. **Distribution**: It occurs in East Africa from southern Ethiopia and Somalia southwards to Eswatini (formerly Eswatini) and northeastern South Africa (Branch 2012a). There are no published records from Cabo Delgado Province but it is known from throughout the country (Vamberger et al. 2019). **Comment.** According to Vamberger et al. (2019), *P. sinuatus* have two major clades distributed along East Africa. One clade occurring in the northern and central parts of the distribution range (Tanzania, Mozambique, and Botswana), and another in the south (Botswana and South Africa). Close to the border region of Botswana, Zimbabwe, and South Africa the two clades overlap.

Cheloniidae

Chelonia mydas (Linnaeus, 1758)

Green Sea Turtle

Material: Sight record (-12.96049°, 40.54048°). **Identification:** Hard-shelled, deep-bodied turtle with a small blunt head, the beak is not hooked and has no cusps. Divers commonly record the presence of this species in Pemba's bay. **Distribution**: The Green Turtle has a circumglobally distribution, occurring throughout tropical and, to a lesser extent, subtropical waters of the Atlantic Ocean, Indian Ocean, and Pacific Ocean (Seminoff, JA. 2004). **Comment**: The species as been recorded in Cabo Delgado Province in the Island of Vamizi (Garnier et al. 2012) and has been assessed as Critically Endangered by the IUCN. Even though they are regularly seen in Pemba's bay, there is no evidence that they use the beaches as breeding sites.

Eretmochelys imbricata (Linnaeus, 1766)

Hawksbill Sea Turtle

Material: Sight record (-12.96049°, 40.54048°). **Identification:** Small sea turtle with bird-like beaked jaws and oval carapace. Divers commonly record the presence of this species in Pemba's bay. **Distribution**: The Hawksbill has a circumglobally distribution throughout tropical and, to a lesser extent, subtropical waters of the Atlantic Ocean, Indian Ocean, and Pacific Ocean (Mortimer et al. 2008). **Comment:** The species as been recorded in Cabo Delgado in the Island of Vamizi (Garnier et al. 2012) and has been assessed as Critically Endangered by the IUCN. Even though individuals of this species are regularly seen in Pemba's bay, there is no evidence that they use the beaches as breeding sites.

CHAPTER 4

ENDEMISM PATTERNS ARE SCALE DEPENDENT

Abstract

Areas of endemism, where species with small ranges concentrate, represent important units for postulating hypotheses in biogeography and are priority targets for conservation action because they capture facets of biodiversity not represented elsewhere. However, the spatial scales at which areas of endemism are relevant to research and conservation are poorly known. Here, we calculated weighted endemism (WE) and phylogenetic endemism (PE) separately for all species of birds and amphibians across the globe. We show that scale dependence is widespread for both indices and manifests across grain sizes, spatial extents and taxonomic treatments. Variations in taxonomic opinions – whether species are treated by systematic 'lumping' or 'splitting' – can profoundly affect the allocation of WE hotspots. Global patterns of PE reflect diversification processes but are lost at the regional to country scales. This is because species distributions span socio-political and regional borders, and patterns of PE missed in one part of the range can compromise patterns of endemism for all countries in which a species occurs. Our models indicate that these findings are overwhelmingly explained by environmental heterogeneity at finer grains (temperature, precipitation, elevation, productivity), and to a far lesser extent at coarser resolutions. We illustrate that regardless of grain size, spatial extent or taxonomic treatment, only 22 to 29% of avian endemism hotspots, and 24 to 25% of amphibian hotspots, meet a minimum target of 10% potential coverage by the global system of protected areas. Our study presents a framework for assessing areas for conservation that are robust to particular assumptions on taxonomy, spatial grain and extent.

Daru, B. H., Farooq, H., Antonelli, A., & Faurby, S. (2020). Endemism patterns are scale dependent. Nature Communications, 11(1), 1-11. https://doi.org/10.1038/s41467-020-15921-6

Introduction

Biodiversity patterns and their underlying mechanistic processes are inherently scale dependent[1–4]. Patterns and processes predicted at one spatial scale may not be predictive at another scale. Information can be lost at coarser spatial scales[5,6], while on the other hand emerging properties, for instance, caused by speciation dynamics may only be present at larger scales[7]. Several studies have indicated that scale dependence may be pervasive in patterns of species richness[3,8–12], density dependence[13–15], extinction risk[16], ratios of native/exotic species[17] or migration and colonization rates[18,19]. However, the effects of scale for patterns of endemism has not been thoroughly explored.

Two important spatial metrics of endemism are weighted endemism and phylogenetic endemism. Weighted endemism is species richness inversely weighted by species ranges[20,21]. Phylogenetic endemism is the phylogenetic equivalent of species endemism, and measures the degree to which phylogenetic diversity is restricted to any given area[22]. Just as the two metrics capture different facets of endemism and are increasingly considered crucial for conservation prioritization[23–25], the effect of scale is also expected to vary differently among them. This is because both metrics depend on spatial grain (i.e. resolution), extent[10] and/or taxonomic treatment[26]. Weighted endemism can be sensitive to changes in taxonomic opinion because small-ranged species are weighted equally. Advances in taxonomic knowledge lead to continuous changes in the number and delineation of species, either through lumping several species into one or splitting single species into several[27,28]. For instance, over the past 110 years, bird species have witnessed varying estimates in their numbers: 18,939 species in 1909[29], 8590 in 1951[30] and 10,738 species today[31]. Such changes in taxonomic concepts can influence estimates of WE, and by consequence bias, undermine or obscure any underlying evolutionary mechanisms[32].

In contrast, phylogenetic endemism offers a potential solution to deal with new taxonomic knowledge in conservation strategies because some lineages and areas harbor far more endemic diversity than would be expected from species ranges alone[22,33,34]. Patterns of phylogenetic endemism tend to manifest at large global scales, but phylogenetic endemism can be severely influenced in a regional setting[24,35]. For example, the Galápagos penguin (*Spheniscus mendiculus*) is the only penguin occurring naturally outside the Southern Hemisphere, endemic to the Galápagos Islands north of the equator[36]. Assuming all else is equal, its phylogenetic endemism is expected to be higher at a continental scale north of the equator, but less likely in a global setting because its closest relatives comprise a group of about 20 species exclusive to the Southern Hemisphere. Incomplete sampling (i.e., missing taxa) or randomly added taxa on the phylogeny (as is often done in macroecological studies) could potentially inflate estimates of phylogenetic endemism[37], in which case taxonomical effects could potentially accumulate. We therefore predict that phylogenetic endemism should vary strongly with spatial extent, whereas weighted endemism should vary depending on taxonomic conclusions i.e. whether the group has been subject to primarily splitting or lumping.

Because different groups of organisms differ in their dispersal abilities and home ranges, they are likely to differ greatly in their utilization of habitat at different spatial grains and extents[38]. Species with wide dispersal capabilities might reflect large geographic range sizes[39], whereas narrow-ranged species may correlate with fine-grained habitat richness[12]. For example, birds have diversified to occupy various habitats and functional roles across most terrestrial and aquatic ecosystems. They show distinct geographic variation in phylogenetic diversity[40,41], and concentrations of spatially restricted phylogenetic diversity have been identified for some clades[25]. On the other hand, amphibians are poor dispersers and possess reduced geographic ranges compared to birds[42], and thus we predict the effect of scale on endemism in amphibians to be strongest at finer grains.

Here, we test the hypothesis that spatial scale influences the observed patterns of weighted endemism and phylogenetic endemism using comprehensive datasets on the phylogenetic relationships and geographic distributions for *c.* 10,000 species of birds and 6000 species of amphibians across the globe. Specifically, we assess the effects of variations in spatial grain (50, 100, 200, 400, and 800 km), extent (global, continental and national), and taxonomic treatment (based on species' divergence times from 1, 2, 3 to 5 Ma) with respect to the identification of hotspots of weighted and phylogenetic endemism. We ask three questions: i) how do patterns of weighted and phylogenetic endemism of different vertebrate clades vary across spatial scales?, ii) at what spatial scales does heterogeneity in environmental factors influence patterns of endemism?, and iii) how effective are the global systems of protected areas in representing hotspots of endemism across grain sizes, spatial extent and taxonomic treatment?

Methods

Species distribution data

The geographic distributional data for birds were obtained from BirdLife International[67], a comprehensive global geographic database for all land and non-pelagic species (n = 10,079 species) available as range map polygons. Range maps for all amphibians were obtained from the IUCN Red List database (https://www.iucnredlist.org/resources/spatial-data-download) of the native extent-of-occurrence of all amphibian species (n = 6337 species). Both sets of maps represent the extent-of-occurrence of the breeding ranges based on museum specimens, direct field observations which have been validated by experts. We matched the range maps to standardized taxonomic authorities including Frost[68] and data from the American Museum of Natural History (AMNH; http://research.amnh.org/vz/herpetology/amphibia/index.php) for amphibians and ref.[31] for birds.

Phylogenetic data

Phylogenetic data for birds comprised of a phylogeny for all extant bird species representing 10,079 species which was based on a distribution of 10,000 possible tree topologies from ref.[40]. The amphibian phylogeny comprised of a phylogeny of 7238 species (94% of all extant amphibians) based on 15 genes on a distribution of 10,000 possible tree topologies from ref.[69]. To account for phylogenetic uncertainty in our analyses for both birds and amphibians, we drew 100 trees at random from a posterior distribution of fully resolved trees generated in ref.[40] for birds and ref.[69] for amphibians.

Degree of *in situ* protection

We quantified the extent at which the global network of protected areas represent hotspots of endemism across grain sizes, spatial extent and taxonomic treatment using the World

Database on Protected Areas (http://protectedplanet.net/)[70]. Our analysis was done on the basis of all terrestrial protected areas classified as IUCN categories I to VI as having sufficient protection status that increases the likelihood that species are well-protected. For each hotspot cell, we quantified the amount of polygon area and examined the proportion of cell overlapping with global system of protected areas. We adopted a 10% cut-off spatial coverage by protected areas corresponding to a conservative coverage target for effective biodiversity protection[62,63].

Data analysis

We constructed a binary presence-absence matrix by overlapping the extent-of-occurrence range map of each bird species with equal-area grid cells. These grid cells were mapped at five consecutive grain sizes following the Behrmann equal area projection system: 50×50 km^2, 100×100 km^2, 200×200 km^2, 400×400 km^2 and 800×800 km^2. At each grain size, we calculated species weighted endemism and phylogenetic endemism.

We used a variant of Laffan & Crisp's[21] weighted endemism metric, defined as the sum of the number of species present in each cell in a local neighborhood, weighting each by the fraction of the area they inhabit[21]. Weighted endemism (WE) is expressed as:

$$WE = \sum_{\{t \in T\}} \frac{r_t}{R_t}$$

where R_t represents the full geographic range of taxon t, and r_t is the local range of taxon t, with the range of a taxon counted in units of number of grid cells in which it is found. We estimated changing spatial scale in weighted endemism for both birds and amphibians under two scenarios of taxonomic treatment: splitters and lumpers. We quantified a lumper’s taxonomic treatment, by successively slicing the phylogenetic tree at various time depths (from 1, 2, 3, to 5 million years ago (Ma)), collapsed nodes and ranges that originated at each time depth, and generated new maps of endemism. It is not possible based on available data to investigate the effects of increased splitting on endemism patterns but it is tempting to assume that some of the hotspots we identify as sensitive to taxonomy may be so in both directions. We used the function *getClusters* in our new R package bioregion[71] to manipulate the phylogenetic tree and collapse nodes and ranges at varying time depths. Weighted endemism was calculated using the function *weighted.endemism(x)* also in our new R package bioregion[71], where *x* is a community matrix or data frame. Our results were integrated across variations of tree topologies and branch lengths for both birds and amphibians by repeating the weighted endemism calculation for each 100 trees from the posterior distribution of trees and taking the median across grid cells.

Phylogenetic endemism was measured as the total phylogenetic branch length spanned by species in an area, dividing each branch length by the global range size of its descendant clade measured in Myr/km^2 ref.[22]. Phylogenetic endemism was calculated using the function *phyloendemism(x, tree, weighted = TRUE)* in the R package PDcalc[72], where *x* is a community matrix or data frame, and *tree* is a phylogenetic tree object. Phylogenetic endemism (PE) is expressed as follows:

$$PE = \sum_{\{i \in I\}} \frac{L_i}{R_i}$$

where $\{I\}$ represents the set of branches connecting species to the root of a phylogenetic tree, L_i is the length of branch *i*, expressed as proportion of the total length of the tree and R_i is the range size of the clade. Because we assumed that PE would vary strongly with spatial extent, we varied the PE analysis across successive spatial extents (global, continental and national) and mapping the hotspots at each spatial extent (see explanatory Figure M1). At the continental or country levels, PE was only calculated based on the species present in that particular inference space. We integrated our result across variations of tree topologies and branch lengths for both birds and amphibians by repeating the PE calculation for each 100 trees from the posterior distribution and taking the median across grid cells for further

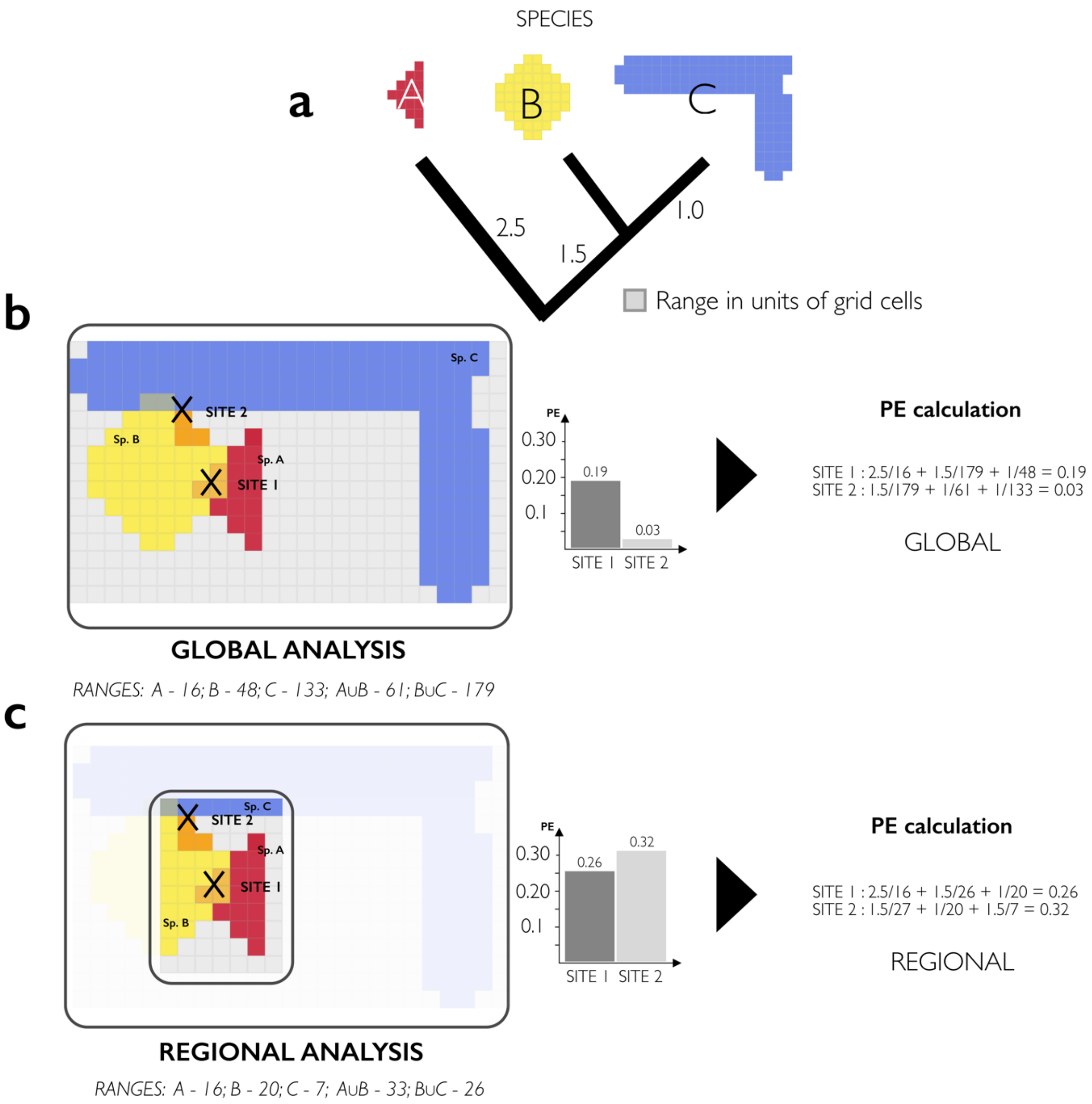


Figure M1 | Hypothetical example showing the variation of phylogenetic endemism with spatial extent (global vs regional scales). **a** Species phylogenetic relationships for three species with varying geographic ranges. b Spatial distribution of phylogenetic endemism across a global scale. At a global scale, PE is calculated accounting for the full geographic range of the species. **c** Distribution of phylogenetic endemism (PE) at a regional scale (continent or country). When species ranges span socio-political borders such that PE is calculated regionally (within a continent or country) without consideration of a species' full range, an inflation of phylogenetic endemism results.

We chose biodiversity hotspots as the basis for quantifying scale dependence of endemism because hotspots can guide allocation of limited conservation resources (e.g., ref.[73]) and endemism lies at the core of understanding the variation of biodiversity across space[64]. Hotspots are areas with significant species richness and endemism. Hotspots for each metric were defined as the 2.5% of grid cells with the highest values of WE and PE[65,66]. To assess uncertainty in the results, we re-ran all analyses by increasing the threshold percentage to 5% to examine if a different threshold percentage value altered the areas identified as hotspots in our analyses. Hotspots were calculated using the function *hotspots(x, values, prob = 2.5)* in our new R package bioregion[71], where *x* is a data frame, *values* the variable in the data frame on which to compute hotspots analysis, and *prob* the threshold quantile for representing the highest proportion of biodiversity in an area. By default, the threshold is set to *prob* = 2.5%.

Environmental heterogeneity

We selected key environmental factors that are commonly used to examine biodiversity-environment associations. These variables included mean annual temperature (MAT), mean annual precipitation (MAP), annual net primary productivity (NPP) and elevation (ALT). MAT, MAP and ALT were downloaded from the WorldClim database[74] at a resolution of 2.5'. NPP was downloaded from NASA Moderate Resolution Imaging Spectroradiometer (MODIS) at a resolution of 1 km and calculated using the MOD17 algorithm. At each grain size, we assessed the disparity between the actual climate of all points within a grid cell and estimated cell endemism values. Specifically, we extracted all the climate points occurring in every grid cell and calculated the standard deviation. We then performed a spatial autocovariate regression to correct for potential spatial autocorrelation in the data[75,76], and a linear mixed model to regress endemism (PE or WE) against the standard deviation of environmental variables and grid cell identities as random covariates. For each focal cell, we varied the weighting function and neighborhood sizes using the next one to two cell/county neighbors to remove spatial autocorrelation.

Results and Discussion

Using five grid resolutions (50, 100, 200, 400, and 800 km), we evaluated changing spatial scale in bird and amphibian weighted endemism based on different extents of taxonomic lumping. To approximate the effect of lumping we successively sliced the phylogenetic tree at various time depths (from 1, 2, 3, to 5 million years ago (Ma)), collapsed nodes and ranges that originated at each time depth, and generated new hotspot maps of endemism which we compared to the original data. As species are treated by taxonomic lumping based on their divergence times at varying time depths, our results show that grid cells identified as being among the richest 2.5% in weighted endemism values (i.e. “hotspots”, henceforth) successively decline with increasing spatial grain (**Fig. 1**), because species lumping collapses smaller ranges into few larger bits. This trend of declining weighted endemism across grain sizes was less steep at finer grain sizes (e.g. 50 km) but became more pronounced at coarser grain sizes such as 800 km (Figure S1). This suggests that the more a taxon has been subjected to systematic lumping based on phylogenetic results, the larger is the reduction. This effect highlights a major property of weighted endemism: all species are weighted equally because weighted endemism does not encapsulate phylogenetic relationships[20]. By increasing spatial grain, we may downweigh the effect of true micro-endemics and lose hotspots of endemism in areas such as small oceanic islands or mountain tops[43]. This effect is due to the assumption of larger species ranges and hide key biogeographical processes such as the influence of geographical barriers such as rivers and mountains.

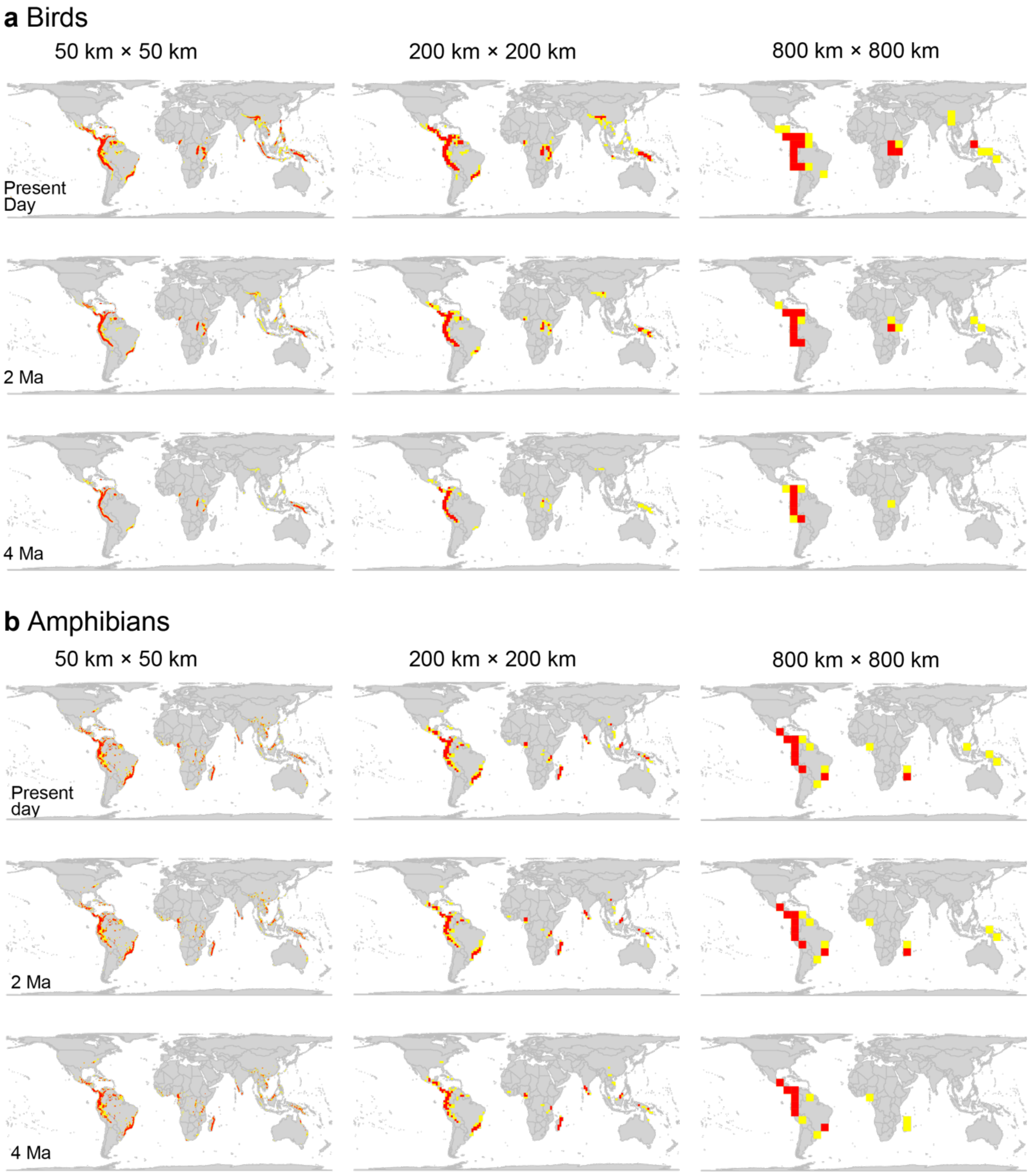


Fig. 1 | Scale dependence of spatial grain and taxonomic treatment of hotspots of weighted endemism, for **a** birds (n = 10,018 species), and **b** amphibians (n = 5872 species) across grains and taxonomic treatment. Hotspots are defined as the grid cells with the highest 2.5% of weighted endemism (indicated in red), and 5% hotspots in yellow. Variations of taxonomic treatments of presented results are based on species' divergence times at varying time depths from present-day, 2, and 4 million years ago (Ma) – whether species are treated by systematic 'lumping' or 'splitting'. Taxonomic lumping results in hotspots successively declining with increasing spatial grain. Analysis of clade collapse based on a randomly selected subset of 100 trees from a posterior distribution of 600 trees for birds and 100 trees from a posterior distribution of 10,000 trees for amphibians. The maps are in Behrmann projection. See Supplementary Figure 2 for full variation of weighted endemism across grid cells at 50, 100, 200, 400, and 800 km and at varying taxonomic treatment based on species' divergence times from 1, 2, 3 to 5 million years ago (Ma).

Across taxa, entire avian hotspots of weighted endemism – e.g. Hawaii, Brazil, West Africa, Sri Lanka, Hengduan-Himalaya, and Southeast Asia – disappeared at both higher spatial resolutions and under severe taxonomic lumping, i.e. when splits that originated around 2 Ma or higher were collapsed (**Fig. 1a,** see also Figure S2a for full variation of weighted endemism across grain sizes at 50, 100, 200, 400, and 800 km). Similarly, hotspots of amphibian weighted endemism saw great declines at higher grain sizes and under taxonomic lumping, resulting in a greater loss or shrinking of amphibian hotspots that affects geographic regions such as Appalachia and Texas in the US, South Africa, West Africa, Hengduan-Himalaya and Australia (**Fig. 1b,** Figure S2b). On the one hand, coarser grain might capture other evolutionary patterns at large scales, such as allopatric speciation and diversification[7]. On the other hand, while inconsistent taxonomy creates challenges in conservation[32,44], we show here that even if the same taxonomic principle (a standardized cut-off at particular evolutionary depths) is used consistently across the phylogeny it influences the results. Our results show that using the biological or phylogenetic species concepts can produce different results and might influence conservation prioritization differently.

Hotspots of phylogenetic endemism are influenced strongly by spatial extent, varying along global, regional and local scales at country level (**Fig. 2,** Figure M1 in Methods). Phylogenetic endemism captures the degree to which the phylogeny is restricted to a single area, highlighting the irreplaceability of these areas for the preservation of deep branches of the tree of life[24,45,46]. For both birds and amphibians at the global scale, well known biodiversity hotspots in the tropics corresponding to Mesoamerica, the Andes, Africa, Madagascar, Papua New Guinea, and South-Central China, plus an additional few in the temperate regions (for amphibians) including Appalachia and the region around Portland Oregon in the US, Southern Chile, Southern Africa, and Queensland Australia, emerged as priority regions at fine to intermediate grains but were absent at coarser grain sizes (**Fig. 2**; Figure S3).

Global patterns of phylogenetic endemism can provide insights into complex evolutionary processes such as dispersal, speciation and extinction shaping large-scale biodiversity patterns[47,48] and may influence the latitudinal diversity gradient, where higher richness and endemism are observed at lower latitudes for most taxonomic groups[49–51]. However, these effects are lost at the regional to country scales. At the continental scale for example, hotspots of phylogenetic endemism are less spatially clumped and more dispersed into new locations outside the tropics including southern Europe (Spain, Portugal, Italy, Greece), Georgia, Azerbaijan and Antarctica for birds (**Fig. 2a**); and Southern Europe, Tasmania and Perth in Australia and New Zealand for amphibians (**Fig. 2b**). In parallel, some regions that emerged as hotspots at the global scale including the Atlantic Forest of Brazil, Hawaii, New Zealand and the Oceanic Islands disappeared at the continental scale (**Fig. 2**). At the national scale, spatial patterns of phylogenetic endemism became more widespread across countries, clustering more at the socio-political borders of countries and decreasing toward coarser grain sizes (**Fig. 2**).

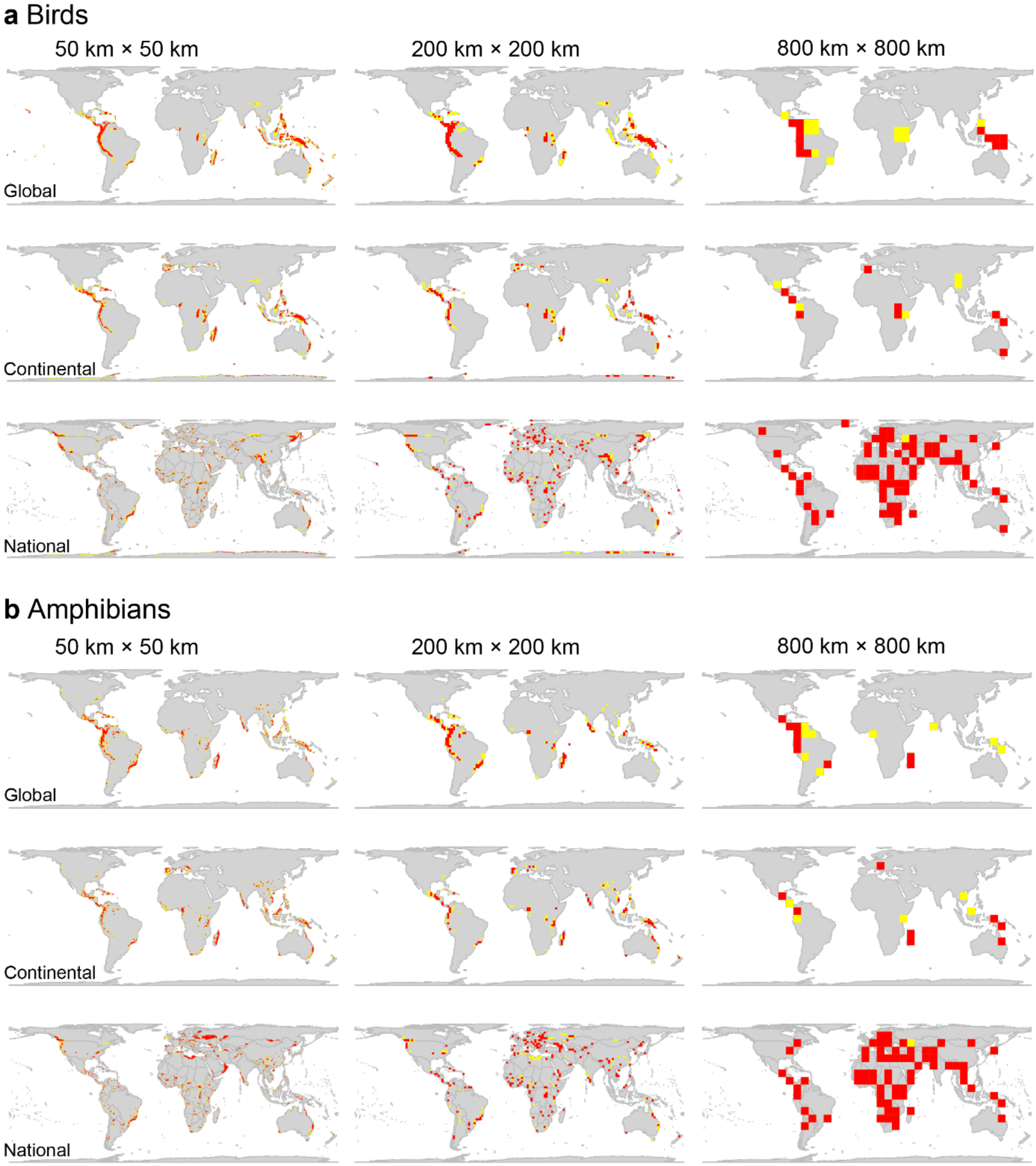


Fig. 2 | Hotspots of phylogenetic endemism are influenced strongly by spatial extent, varying along global, continental and local scales at country level for **a** birds (n = 10,018 species), and **b** amphibians (n = 5872 species) of the world across three levels of spatial extents (national, continental, and global). Global patterns of phylogenetic endemism reflect diversification processes but are lost at the regional to country scales. Hotspots are defined as the grid cells with the highest 2.5% of phylogenetic endemism (indicated in red), and 5% hotspots in yellow. The maps are in Behrmann projection. See Supplementary Figure 3 for full variation of phylogenetic endemism across grid cells at 50, 100, 200, 400, and 800 km.

Socio-political borders serve no ecological roles because they rarely coincide with ecological boundaries, reflected by the fact that most species ranges span political and regional borders. Patterns of endemism missed in one part of a species range can compromise endemism on either side of the border that the species overlaps[52]. Biodiversity components on each side of the border are therefore often subject to conflicting management practices[53]. This means that too little or too high attention can be made to border regions for conservation purposes. For example, South Texas in the United States is well-known for its high concentrations of species richness and endemism of birds (including the Green jay *Cyanocorax luxuosus* and the ringed kingfisher *Megaceryle torquata*) and amphibians (e.g. mole and lungless salamanders)[25]. However, most species in this region are mobile and migratory, posing challenges for assessing endemism or extinction risk because their status under conservation legislation can change radically across borders[54,55]. On the other hand, analyses conducted just at the country level can overestimate endemism levels for species barely reaching into a country (see Figure M1 in Methods). This is the case for the Red-billed Pigeon (*Patagioenas flavirostris*). This species is widespread in Mexico and central America but has a small breeding population in southern United States close to Rio Grande in Texas. Unless managers on both sides of a socio-political border adopt compatible management strategies, conservation actions on only one side are likely to lead to suboptimal solutions. Thus, the spatial scale of the habitats supporting species should match the scale of management strategies designed to protect the species through international collaborations[56].

The question of which scale is ideal for analyzing areas of endemism will depend on the objectives of the study. Rahbek[10] suggests the use of a grain size as small as the smallest range sizes among the species in the study area. Increased scale may reduce biases associated with sampling artefacts because small grain can represent well-known rather than diverse areas. The extent-of-occurrence maps commonly used in biogeographical analyses are drawn by experts to depict the maximum geographical extent of a species and might be compromised by false presences if analyzed at too fine a scale. It is therefore generally recommended to interpret these analyses with grain smaller than 1° or 2° latitude/longitude (~110–220 km around the equator) with caution[57–59]. Our results suggest that endemism should be analyzed at as high a resolution as the data can allow (which generally is at intermediate grains of 100 km to 200 km).

To test whether environmental heterogeneity influences patterns of weighted and phylogenetic endemism we used four commonly used environmental predictors: elevation, mean annual temperature (temperature henceforth), mean annual precipitation (precipitation henceforth) and net primary productivity (productivity henceforth). We performed these analyses across grain sizes using spatial autocovariate regression models (**Fig. 3**). Across clades, our results indicated that, in general, the explanatory power of environmental factors decreased with increasing spatial grain for weighted endemism and was particularly strong at the two finest scales of 50 and 100 km (**Fig. 3**). For instance, at 50 km, heterogeneity (i.e. standard deviation) in elevation, temperature and productivity offered strong predictions of avian weighted endemism (elevation: beta = 0.12, $p < 0.001$; temperature: beta = 0.10, $p = 0.048$; and productivity: beta = 0.029, $p < 0.0001$). The opposite was found for phylogenetic endemism, showing stable relationship at fine to intermediate grains (50 to 200 km) and is lowest for coarse-grained assemblages (**Fig. 3**), with strong relationships of temperature and precipitation for avian phylogenetic endemism and precipitation for amphibian weighted and phylogenetic endemism. Environmental heterogeneity is assumed to promote dispersal barriers that may decrease species diversity leading to increase speciation rates[7]. As grain size increases, climatic variables are often assigned a summary value for the grid cell (which can be the centroid value, mean, or median), and can directly bias the importance of regional climate heterogeneity or the locality from where species actually occur, thus leading to spurious conclusions[60,61]. Our results highlight the limitation of comparing endemism-environment relationships at single grains, such that our approach can help in locating hotspots of endemism that are more meaningfully associated to the environmental features of the region. Thus, a multi-grain approach to endemism-environment relationships should be considered in model testing and conservation planning.

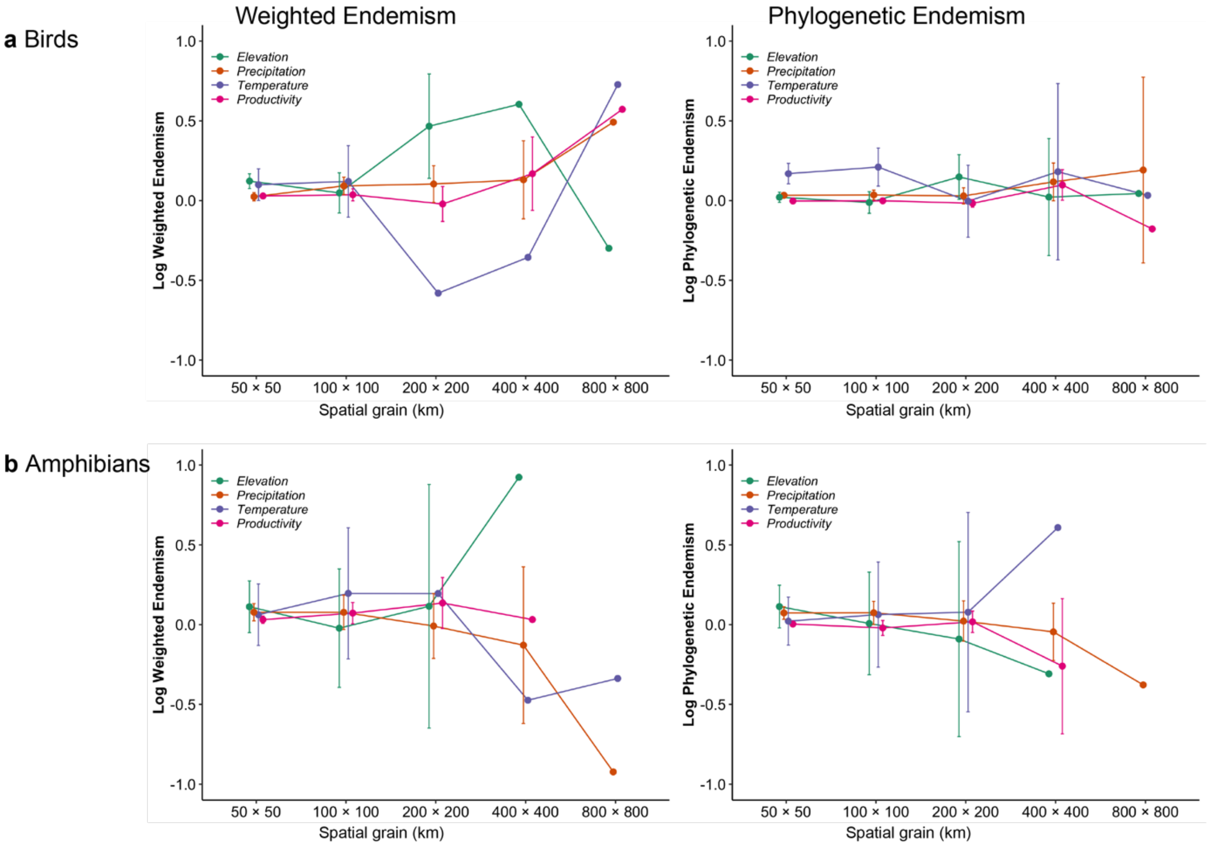


Fig. 3 | Changes in patterns of weighted endemism and phylogenetic endemism in relation to heterogeneity in environmental variables at different spatial scales, for **a** birds and **b** amphibians. Statistical analysis is based on a mixed effects model using spatial autocovariate regression models using standard deviation of environmental heterogeneity for each grid cell. These models indicate that our findings are overwhelmingly explained by environmental heterogeneity at finer grains (temperature, precipitation, elevation, productivity), and to a far lesser extent at coarser resolutions. Error bars represent lower and upper confidence intervals. The standard errors at the 800 km × 800 km resolution go outside the scale.

To highlight the critical gaps in protecting areas of endemism across scales, we mapped hotspots of weighted and phylogenetic endemism for birds and amphibians of the world across grain sizes, spatial extents and taxonomic treatments. We then assessed the scale to which areas of endemism are captured in at least 10% by the current network of protected areas. The 10% threshold is a conservative target of coverage by protected areas advocated for safeguarding biodiversity[62,63]. Overall, we reveal that only 22 to 29% of avian endemism hotspots, and 24 to 25% of amphibian hotspots, meet a minimum target of merely 10% potential coverage by the global system of protected areas (**Fig. 4**). Across scales, the situation is even more alarming. Hotspots of weighted endemism for both birds and amphibians are more protected at finer to intermediate grain sizes (50-200 km) with up to 28-33% coverage by protected areas for birds (**Fig. 4a**) and 26% for amphibians (**Fig. 4b**). This pattern is insensitive to the extent of taxonomic over-lumping. Importantly, all hotspots of phylogenetic endemism – regardless of grain size or spatial extent – fall below the critical 10% coverage target for protection, with the exception of hotspots of amphibian phylogenetic endemism analyzed at the continental scale that meet the minimum protection threshold of 10% by protected areas (**Fig. 4**). Overall, we found widespread deficits of protection for endemism hotspots regardless of grain size, spatial extent or taxonomic treatment.

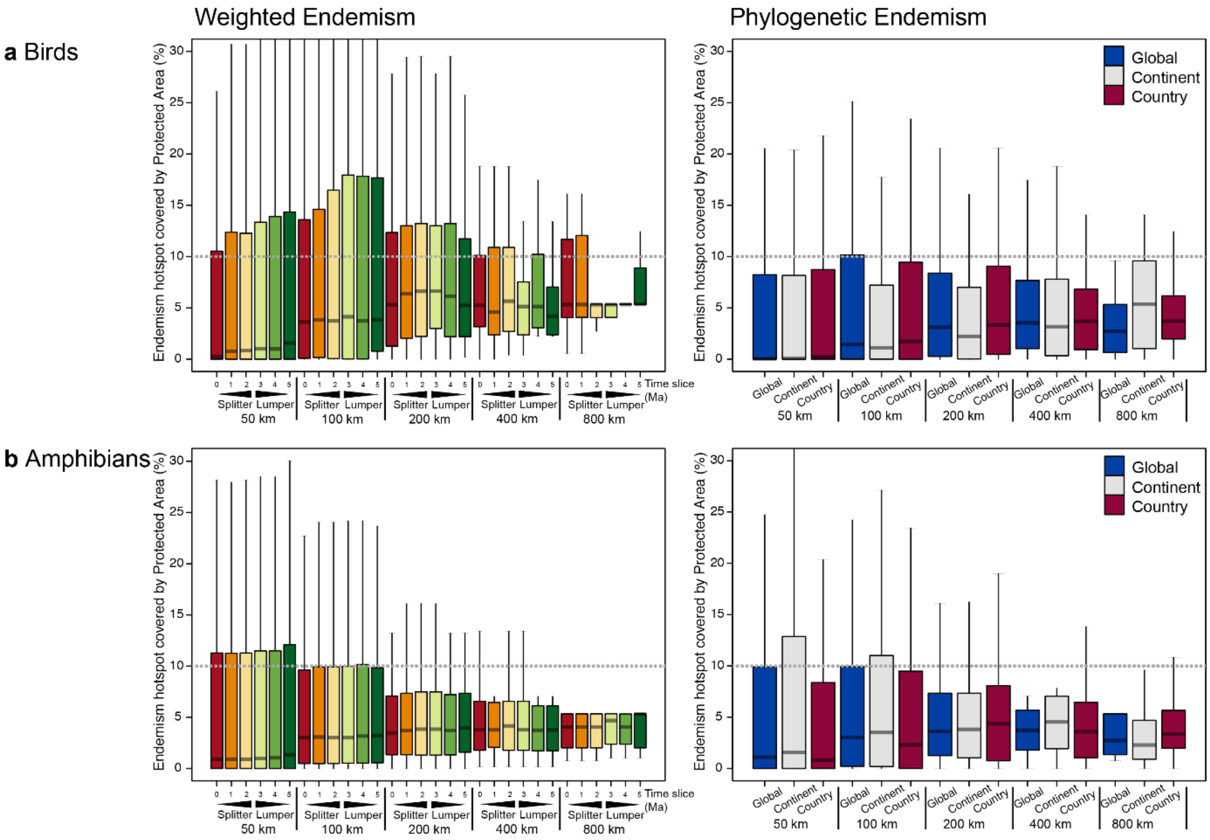


Fig. 4 | Relationship between variation in grain size and the proportion of endemism hotspots covered by the global systems of Protected Areas. **a** Bird weighted endemism (left) and phylogenetic endemism (right), and **b** Amphibian weighted endemism (left) and phylogenetic endemism (right). The dotted lines represent the 10% threshold corresponding to the minimum representation target for sustaining species persistence. These findings demonstrate widespread deficits of protection for endemism hotspots regardless of grain size, spatial extent or taxonomic treatment.

In conclusion, our study shows that changes in taxonomic treatments and spatial grain strongly influence patterns of endemism. Lumping can be detrimental to conservation if species are delisted as a result, with subsequent cessation in monitoring and policy efforts for their protection[44]. Conversely, splitting can lead to suboptimal conservation solutions and management problems because populations are managed as distinct units without any assisted gene flow, potentially causing inbreeding issues. Splitting can also lead to potentially wrong spatial conservation prioritization which may limit the number of biological species that are saved. Coarse spatial grain misinterpretations can overlook important areas of endemism which may become vulnerable and degraded due to anthropogenic disturbance. Endemism represents one important but overlooked component in conservation[64–66]. Our study showcases how to assess important areas for conservation that are robust to particular assumptions on taxonomy, spatial grain and extent.

References

1. Wiens, J. A., Rotenberry, J. T. & Van Horne, B. Habitat occupancy patterns of North American shrubsteppe birds: the effect of spatial scale. *Oikos* **48**, 132–147 (1987).
2. Wiens, J. A. Spatial scaling in ecology. *Funct. Ecol.* **3**, 385–397 (1989).
3. Levin, S. A. The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967 (1992).
4. Schneider, D. C. *Quantitative ecology: spatial and temporal scaling* (Academic Press, San Diego, CA, 1994).
5. Henderson-Sellers, A., Wilson, M. F. & Thomas, G. The effect of spatial resolution on archives of land cover type. *Clim. Change* **7**, 391–402 (1985).
6. Meentemeyer, V. & Box, E. O. in *Landscape heterogeneity and disturbance* (New York: Springer-Verlag, 1987).
7. Quintero, I. & Jetz, W. Global elevational diversity and diversification of birds. *Nature* **555**, 246–250 (2018).
8. Rahbek, C. & Graves, G. R. Multiscale assessment of patterns of avian species richness. *Proc. Natl. Acad. Sci. USA* **98**, 4534–4539 (2001).
9. Willis, K. J. & Whittaker, R. J. Species diversity—scale matters. *Science* **295**, 1245–1248 (2002).
10. Rahbek, C. The role of spatial scale and the perception of large-scale species-richness patterns. *Ecol. Lett.* **8**, 224–239 (2005).
11. Mykrä, H., Heino, J. & Muotka, T. Scale-related patterns in the spatial and environmental components of stream macroinvertebrate assemblage variation. *Glob. Ecol. Biogeogr.* **16**, 149–159 (2007).
12. Belmaker, J. & Jetz, W. Cross-scale variation in species richness–environment associations. *Glob. Ecol. Biogeogr.* **20**, 464–474 (2011).
13. Jarosik, V. & Lapchin, L. An experimental investigation of patterns of parasitism at three spatial scales in an aphid-parasitoid system (Hymenoptera: Aphidiidae). *Eur. J. Entomol.* **98**, 295–299 (2001).
14. Gunton, R. M. & Kunin, W. E. Density effects at multiple scales in an experimental plant population. *J. Ecol*. **95**, 435–445 (2007).

15. Gunton, R. M. & Kunin, W. E. Density-dependence at multiple scales in experimental and natural plant populations. *J. Ecol*. **97**, 567–580 (2009).
16. Hartley, S. & Kunin, W. E. Scale dependency of rarity, extinction risk, and conservation priority. *Conserv. Biol.* **17**, 1559–1570 (2003).
17. Davies, K. F., Chesson, P., Harrison, S., Inouye, B. D., Melbourne, B. A. & Rice, K. J. Spatial heterogeneity explains the scale dependence of the native-exotic diversity relationship. *Ecology* **86**, 1602–1610 (2005).
18. Menendez, R. & Thomas, C. D. Metapopulation structure depends on spatial scale in the host-specific moth *Wheeleria spilodactylus* (Lepidoptera: Pterophoridae). *J. Anim. Ecol*. **69**, 935–951 (2000).
19. Englund, G. & Hamback, P. A. Scale dependence of immigration rates: models, metrics and data. *J. Anim. Ecol*. **76**, 30–35 (2007).
20. Crisp, M. D., Laffan, S., Linder, H. P. & Monro, A. Endemism in the Australian flora. *J. Biogeogr.* **28**, 183–198 (2001).
21. Laffan, S. W., & Crisp, M. D. Assessing endemism at multiple spatial scales, with an example from the Australian vascular flora. *J. Biogeogr*. **30**, 511–520 (2003).
22. Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, C. & Cook, L. G. Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* **18**, 4061–4072 (2009).
23. Laity, T. et al. Phylodiversity to inform conservation policy: An Australian example. *Sci. Total Environ.* **534**, 131–143 (2015).
24. Rosauer, D. F. & Jetz, W. Phylogenetic endemism in terrestrial mammals. *Glob. Ecol. Biogeogr*. **24**, 168–179 (2015).
25. Daru, B. H., le Roux, P. C., Gopalraj, J., Park, D. S., Holt, B. G. & Greve, M. Spatial overlaps between the global protected areas network and terrestrial hotspots of evolutionary diversity. *Glob. Ecol. Biogeogr*. **28**, 757–766 (2019).
26. Faurby, S., Eiserhardt, W. L. & Svenning, J. Strong effects of variation in taxonomic opinion on diversification analyses. *Methods Ecol. Evol*. **7**, 4-13 (2016).
27. Collen, B. et al. Investing in evolutionary history: implementing a phylogenetic approach for mammal conservation. *Phil. Trans. R Soc. B* **366**, 2611–2622 (2011).
28. Burgin, C. J., Colella, J. P., Kahn, P. L. & Upham, N. S. How many species of mammals are there? *J. Mammal.* **99**, 1–14 (2018).

29. Sharpe, R. B. *A hand-list of the genera and species of birds, Vol. 5* (British Museum, London, UK, 1909).
30. Mayr, E. & Amadon, D. A classification of recent birds. *Am. Mus. Novit.* **1496**, 1–42 (1951).
31. Gill, F. & Donsker, D. *IOC World Bird List v9.1* (doi :10.14344/IOC.ML.9.1., 2019).
32. Robuchon, M. et al. Species splitting increases estimates of evolutionary history at risk. *Biol. Conserv.* **235**, 27–35 (2019).
33. Sechrest, W. et al. Hotspots and the conservation of evolutionary history. *Proc. Natl Acad. Sci. USA* **99**, 2067–2071 (2002).
34. Faith, D. P., Reid, C. A. M. & Hunter, J. Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. *Conserv. Biol.* **18**, 255–261 (2004).
35. Jarzyna, M. A. & Jetz, W. Taxonomic and functional diversity change is scale dependent. *Nat. Commun.* **9**, 2565 (2018).
36. Boersma, P. D. An ecological and behavioral study of the Galapagos penguin. *Living Bird* **15**, 43–93 (1976).
37. Park, D. S., Worthington, S., Xi, Z. Taxon sampling effects on the quantification and comparison of community phylogenetic diversity. *Mol. Ecol.* **27**, 1296–1308 (2018).
38. Kotliar, N. B. & Wiens, J. A. Multiple scales of patchiness and patch structure – a hierarchical framework for the study of heterogeneity. *Oikos* **59**, 253–260 (1990).
39. Lester, S. E., Ruttenberg, B. I., Gaines, S. D. & Kinlan, B. P. The relationship between dispersal ability and geographic range size. *Ecol. Lett.* **10**, 745–758 (2007).
40. Jetz, W., Thomas, G. H., Joy, J. B. & Mooers, A. Ø. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
41. Jetz, W., Thomas, G. H., Joy, J. B., Redding, D. W., Hartmann, K. & Mooers, A. Ø. Global distribution and conservation of evolutionary distinctness in birds. *Curr. Biol.* **24**, 919–930 (2014).
42. Hendriks, A. J., Willers, W. J. C., Lenders, H. J. R. & Leuven, R. S. E. W. Towards a coherent allometric framework for individual home ranges, key population patches and geographic ranges. *Ecography* **32**, 929–942 (2009).
43. Rabinowitz, D. in *The biological aspects of rare plant conservation* (New York: John Wiley & Sons Ltd, 1981).

44. Morrison, W. R. et al. The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant? *Biol. Conserv.* **142**, 3201–3206 (2009).
45. Faith, D. P. Threatened species and the potential loss of phylogenetic diversity: conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conserv. Biol*. **22**, 1461–1470 (2008).
46. Cadotte, M. W. & Davies, T. J. Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales. *Divers. Distrib.* **16**, 376–385 (2010).
47. Antonelli, A. Biogeography: Drivers of bioregionalization. *Nat. Ecol. Evol.* **1**, 0114 (2017).
48. Jetz, W. et al. Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol*. **3**, 539–551 (2019).
49. Singh, J. J. *Study on the development of transboundary natural resource management areas in southern Africa—Global review; lessons learned* (Washington: Biodiversity Support Program, 1999).
67. BirdLife International and NatureServe. *Bird species distribution maps of the world* (Cambridge, UK: BirdLife International, 2014).
68. Frost, D. R. *Amphibian species of the world: An online reference* (Version 5.3. Retrieved from http://research.amnh.org/vz/herpetology/amphibia/index.php, 2009).
69. Jetz, W. & R. Pyron, R. A. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nat. Ecol. Evol.* **2**, 850–858 (2018).
70. IUCN and UNEP-WCMC. *The World Database on Protected Areas* (WDPA) (Cambridge, UK: UNEP-WCMC. Retrieved from www.protectedplanet.net, 2015).
71. Daru B. H. & Schliep, K. *bioregion: an R package for mapping biodiversity and conservation* (R package version 0.1.0. https://github.com/darunabas/bioregion, 2019).
72. Nipperess, D. & Wilson, P. PDcalc: An implementation of the Phylogenetic Diversity (PD) calculus in R (R package version 0.3.2.9000, 2019).
73. Forest, F. et al. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760 (2007).

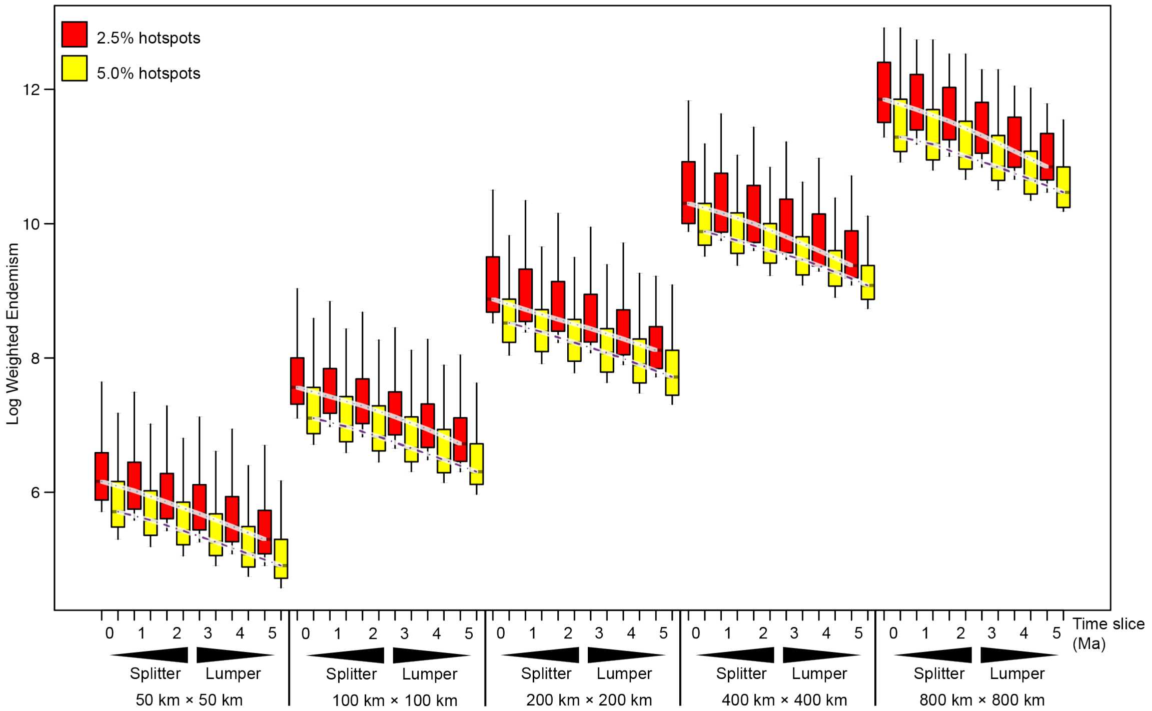
74. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol*. **25**, 1965–1978 (2005).

75. Augustin, N. H., Mugglestone, M. A. & Buckland, S. T. An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol*. **33**, 339–347 (1996).

76. Gelfand, A. E. & Vounatsou, P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**, 11–25 (2003).

Supplementary Material

a Birds



b Amphibians

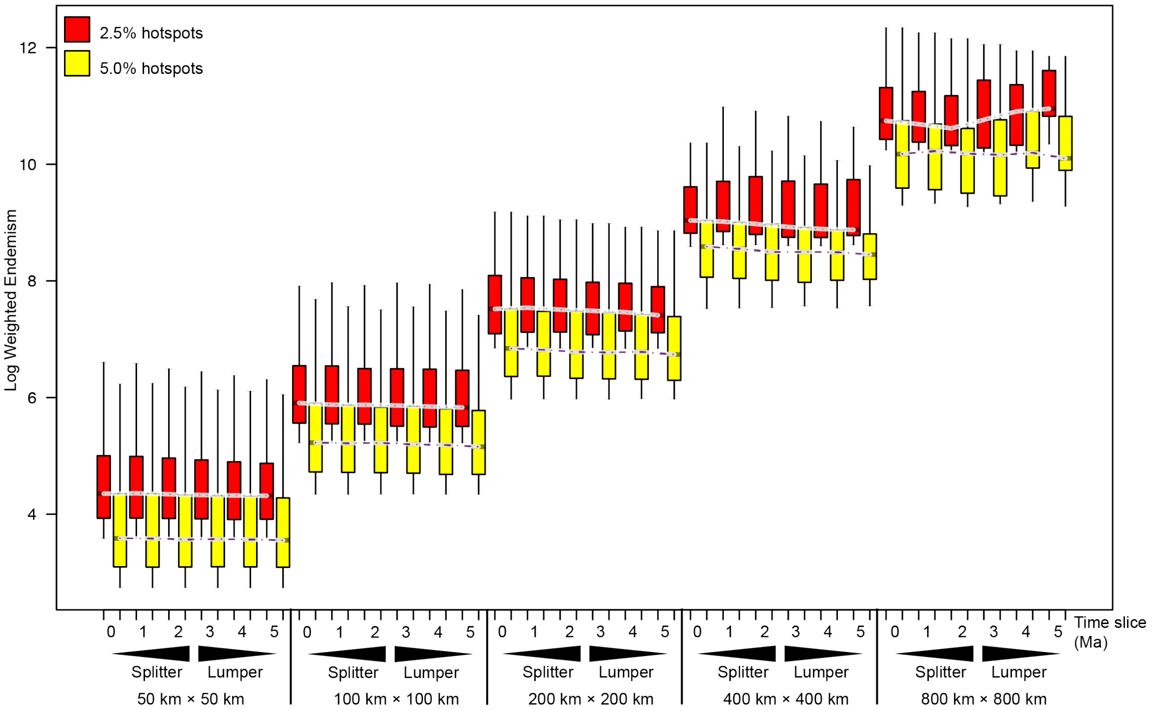


Figure S1 | Changes in weighted endemism in relation to scale (50 × 50, 100 × 100, 200 × 200, 400 × 400, and 800km × 800 km) for **a** bird species (n = 10,018 species), and **b** amphibians (n = 5872 species). The effect of spatial grain is evident in the fitted slopes between species endemism and taxonomic treatment (splitters vs lumpers across varying time slices). Box plots represent hotspot cells with the highest weighted endemism values (top 2.5% endemism in red and 5.0% endemism in yellow). Lines within the boxes represent the 50th percentile (median), and whiskers represent 2.5th and 97.5th percentiles. Analysis was based on phylogenetic data derived from a random draw of 100 trees from a Bayesian posterior distribution of 600 trees for birds and 10,000 trees for amphibians. We integrated our results across variations of tree topologies and branch lengths for both birds and amphibians by calculating WE for each 100 trees from the posterior distribution and computing the median across grid cells.

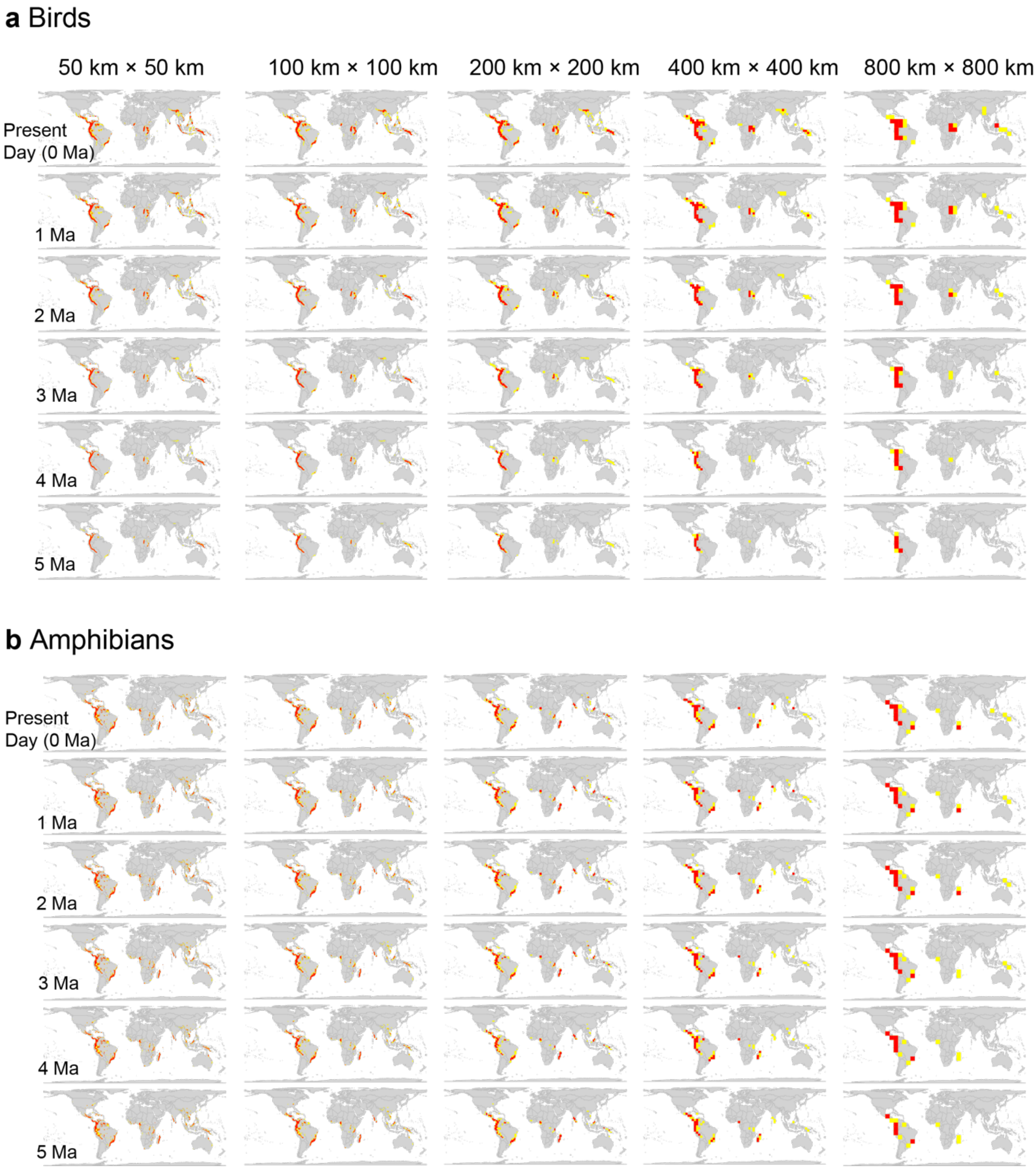


Figure S2 | Scale dependence of spatial grain and taxonomic treatment of hotspots of weighted endemism, for **a** birds (n = 10,018 species), and **b** amphibians (n = 5872 species) across grains and taxonomic treatment. Hotspots are defined as the grid cells with the highest 2.5% of weighted endemism (indicated in red), and 5% hotspots in yellow. Variations of taxonomic treatments of presented results are based on species' divergence times at varying time depths from present-day, 1, 2, 3, 4 and 5 million years ago (Ma) – whether species are treated by systematic 'lumping' or 'splitting' – and can profoundly affect the allocation of weighted endemism hotspots. Analysis of clade collapse based on a randomly selected subset of

100 trees from a posterior distribution of 600 trees for birds and 100 trees from a posterior distribution of 10,000 trees for amphibians. The maps are in Behrmann projection.

Figure S3 | Hotspots of phylogenetic endemism are influenced strongly by spatial extent, varying along global, continental and local scales at country level for **a** birds (n = 10,018 species), and **b** amphibians (n = 5872 species) of the world across three levels of spatial extents (national, continental, and global) and grid sizes at 50, 100, 200, 400, and 800 km. Hotspots are defined as the grid cells with the highest 2.5% of phylogenetic endemism (indicated in red), and 5% hotspots in yellow. The maps are in Behrmann projection.

Code availability. All scripts and code necessary to repeat the analyses described here have been made available in the new R package bioregion[71].

Data availability. All data necessary to repeat the analyses described here have been made available through the Dryad digital data repository (https://doi.org/10.5061/dryad.wh70rxwhs).

DISCUSSION AND CONCLUDING REMARKS

This thesis contributes to biodiversity conservation by both presenting a new tool to rank important areas of biodiversity and by highlighting the current lack of baseline biodiversity knowledge in the African continent. I unveil important behavior of scientists when sampling biodiversity in Africa and make predictions on time and effort to sample biodiversity in the continent. In this section I synthesize the main findings of the thesis, explore some of the challenges on biodiversity research and conservation in the African continent and outline some of the potential future research I plan to dedicate to in the future.

Challenges for biodiversity research in Africa

As seen in the recent Living Planet Report 2020 (WWF 2020), Africa lost 65% of its biodiversity in the last 50 years. In Chapter 2 we show that at the current rate it will take between 172 and 274 years for the most well-known studied animals to be sampled at least once in each 100 km by 100 km grid cell in Africa. However, as also predicted in chapter 2, it may take between 12 and 27 sampling events depending on the group to record at least 50% of the species occurring in a single cell. These numbers showcase the fact that unless a radical and widespread change in research practice takes place, Africa's rich biodiversity will remain largely unknown. Sampling in Africa is urgently needed to delimit species boundaries and understand spatial biodiversity patterns. Work presented in chapter 3 where I compile years of collections of amphibians and reptiles from a poor sampled area is one example of studies that can considerably contribute to understand species distributions in the continent. Most of the continent is under-sampled and vertebrates are over-represented. These biases have implications in the understanding of biodiversity threats and consequently impact the allocation of funding to address these threats (Donaldson et al. 2016). Only with baseline information we will be able to promote effective species conservation.

As mentioned in chapter 2, funding providers such as agencies, companies, and/or philanthropists should pay special attention to projects that aim to generate baseline information in areas and groups where no data is available and make sure the collected information becomes publicly and easily accessible on platforms such as GBIF.

One way to attain basic and widespread knowledge on biodiversity in Africa is by working with local people and institutions. Much of the biological sampling in Africa has been carried out by European and North American institutions which may be constrained due to expensive trips, duration of fieldwork and communication with local authorities such as in the process of acquiring research permits.

Conservation in Africa – not an easy task

In an ideal world, the implementation of effective conservation practices would rely on a complete understanding of species relationships and distributions backed up by an array of extensive experiments conducted in order to assess which practices are more efficient. If we place all continents in a gradient going from data availability on the right to lack of data to the left, Africa would, without a doubt, be placed far left together with South America.

Having some of the most biodiversity rich ecosystems in our planet, nature conservation in the continent is crucial for the preservation of the world's biodiversity. However, Africa is also the poorest continent, where the health and education levels of human populations are still at precarious levels and nations with very limited budgets will understandably lean towards prioritizing the improvement of the economic situation and human well-being in their states, leaving conservation to be mostly funded by foreign and international bodies. Funding may be the biggest constraint when implementing conservation. Resources are needed to hire and keep staff, vehicles, other equipment and to conduct research. Therefore, developing tools that enable a transparent ranking of the importance of different areas is crucial to better distribute resources. Not only these tools need to be developed but also make them easily available to decision makers to be easily applicable. By developing a tool that enables the ranking the biodiversity importance of areas I expect to contribute to the process of prioritisation of conservation resources in order to optimize the effort on preserving biodiversity.

The importance of prioritisation in conservation has already been discussed (Ferrier et al. 2000; Plumptre et al. 2019; Pressey et al. 1994; Smith et al. 2019) and multiple metrics, including protection status, available funding, 'irreplaceability' (Plumptre et al. 2019) and systematic conservation planning (Smith et al. 2019) have been proposed to rank areas. This methodology, although providing a hierarchy among KBAs, still clusters them in different categories, rather than scoring individual sites as is the purpose of the metric I developed in chapter 1 – WEGE. This metric encapsulates the biodiversity importance of a particular site, highlighting the same areas as the KBA criteria while adding the advantages of a continuous ranking system.

Species range is one of the most important aspects of species. They may determine how common and resilient the species are. Hence endemism being an important component of the KBA assessment and of the WEGE metric. In chapter 4, we show that taxonomic treatments and spatial grain strongly influence patterns of endemism. The results are especially relevant since the choice of scale or taxonomy will determine whether some areas are considered hotspots of endemic species or not. For this reason, agreement on species taxonomy and scale at which analyses are conducting may contribute to a more transparent conservation assessment process.

Future research

In Chapter 1, where we developed a tool that spatially ranks the biodiversity importance of an area along the lines of the Key Biodiversity Areas – the WEGE index may be seen as a first step on addressing a bigger question, in which instead of trying to comply to the Key Biodiversity Areas guidelines, we make use of species phylogenetic relationships as well as species role in the environment in the form of functional diversity in order to in a more holistic approach spatially rank biodiversity importance. In the introduction of this thesis I discuss the importance of evidence-based conservation and in chapter 2, I showcase the lack of biodiversity research conducted in Africa. Therefore, I would like to work on making the analysis conducted specially on chapter 1 easily accessible to decision makers. This will include a comprehensive analysis of the WEGE index for all KBAs globally and for all species assessed by IUCN and rank them by country. This information may also be available on a dedicated website where WEGE scores will be available for all Key Biodiversity Areas as well as for all Protected Areas.

In chapter 3 I unveil the biodiversity of amphibians and reptiles of Pemba, a city in Northern Mozambique. Unfortunately, numerous expeditions undertaken during my PhD didn’t make it in this thesis due to time constraints. These include several range extensions as well as new species descriptions that I plan to work on after my PhD. Filling gaps of species distributions with records as seen in chapter 2 is crucial to systematics and to the understanding of global biodiversity patterns.

Concluding remarks

Evidence-based conservation is an important step towards effective conservation policies, however baseline information for most of Africa either doesn't exist or is not publicly accessible. At current rates, sufficient biodiversity knowledge in Africa to allow effective conservation decisions will only exist in hundreds of years, but in a continent that already lost 65% of its biodiversity in its last 50 years and is expected to have its population double by 2050, waiting hundreds of years to have baseline information to guide conservation practice is not a viable option.

In a continent where a third of the population live under the line of poverty and where levels of corruption are some of the highest worldwide, transparent methodologies and prioritisation of conservation efforts are likely to contribute to the subsistence of biodiversity.

REFERENCES

Aleman JC, Jarzyna MA, Staver AC (2018) Forest extent and deforestation in tropical Africa since 1900. Nature ecology & evolution 2: 26-33

Anderson S (1994) Area and endemism. The quarterly review of biology 69: 451-471

Annez PC, Buckley RM, Spence M (2009) Urbanization and growth. Commission on Growth and Development: World Bank, pp.

Barthlott W, Biedinger N, Braun G, Feig F, Kier G, Mutke J (1999) Terminological and methodological aspects of the mapping and analysis of global biodiversity. 103-110 pp.

Bibby CJ, Jones M, Marsden S (1998) Bird surveys. Expedition Advisory Centre London, pp.

BirdLife International (2020) Digital boundaries of Important Bird and Biodiversity Areas from the World Database of Key Biodiversity Areas. . In: BirdLife International (Ed), www.keybiodiversityareas.org, pp.

Bruner AG, Gullison RE, Rice RE, Da Fonseca GA (2001) Effectiveness of parks in protecting tropical biodiversity. Science 291: 125-128

Bunge J, Fitzpatrick M (1993) Estimating the number of species: a review. Journal of the American Statistical Association 88: 364-373

Burgess N, Hales JA, Underwood E, Dinerstein E, Olson D, Itoua I, Schipper J, Ricketts T, Newman K (2004) Terrestrial ecoregions of Africa and Madagascar: a conservation assessment. Island Press, pp.

Christenhusz MJ, Byng JW (2016) The number of known plants species in the world and its annual increase. Phytotaxa 261: 201-217

Cohen AS, Bills R, Cocquyt CZ, Caljon A (1993) The impact of sediment pollution on biodiversity in Lake Tanganyika. Conservation Biology 7: 667-677

Collette B, Amorim A, Boustany A, Carpenter K, de Oliveira Leite Jr N, Di Natale A, Die D, Fox W, Fredou F, Graves J (2011) Thunnus thynnus. IUCN 2012 IUCN Red List of Threatened Species.

Colwell RK (2009) Biodiversity: concepts, patterns, and measurement. The Princeton guide to ecology 663: 257-263

Costanza R, d'Arge R, De Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O'neill RV, Paruelo J (1997) The value of the world's ecosystem services and natural capital. Nature 387: 253-260

Couvreur TLP, Dauby G, Blach-Overgaard A, Deblauwe V, Dessein S, Droissart V, Hardy OJ, Harris DJ, Janssens SB, Ley AC, Mackinder BA, Sonké B, Sosef MSM, Stévart T, Svenning J-C, Wieringa JJ, Faye A, Missoup AD, Tolley KA, Nicolas V, Ntie S, Fluteau F, Robin C, Guillocheau F, Barboni D, Sepulchre P (2020) Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna. Biological Reviews n/a. doi:10.1111/brv.12644

Crisp MD, Laffan S, Linder HP, Monro A (2001) Endemism in the Australian flora. Journal of Biogeography 28: 183-198

Darwall W, Smith K, Allen D, Holland R, Harrison I, Brooks E (2011) The diversity of life in African freshwaters: underwater, under threat: an analysis of the status and distribution of freshwater species throughout mainland Africa. The diversity of life in African freshwaters: underwater, under threat: an analysis of the status and distribution of freshwater species throughout mainland Africa:

Di Marco M, Chapman S, Althor G, Kearney S, Besancon C, Butt N, Maina JM, Possingham HP, von Bieberstein KR, Venter O (2017) Changing trends and persisting biases in three decades of conservation science. Global Ecology and Conservation 10: 32-42

Donaldson MR, Burnett NJ, Braun DC, Suski CD, Hinch SG, Cooke SJ, Kerr JT (2016) Taxonomic bias and international biodiversity conservation research. Canadian Science Publishing 65 Auriga Drive, Suite 203, Ottawa, ON K2E 7W6, pp.

Ehrlich PR (1992) Population biology of checkerspot butterflies and the preservation of global biodiversity. Oikos: 6-12

Eken G, Bennun L, Brooks TM, Darwall W, Fishpool LD, Foster M, Knox D, Langhammer P, Matiku P, Radford E (2004) Key biodiversity areas as site conservation targets. BioScience 54: 1110-1118

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation 61: 1-10

Faith DP, Reid C, Hunter J (2004) Integrating phylogenetic diversity, complementarity, and endemism for conservation assessment. Conservation Biology 18: 255-261

Ferrier S, Pressey RL, Barrett TW (2000) A new predictor of the irreplaceability of areas for achieving a conservation goal, its application to real-world planning, and a research agenda for further refinement. Biological Conservation 93: 303-325

Foden W, Midgley GF, Hughes G, Bond WJ, Thuiller W, Hoffman MT, Kaleme P, Underhill LG, Rebelo A, Hannah L (2007) A changing climate is eroding the geographical range of the Namib Desert tree Aloe through population declines and dispersal lags. Diversity and distributions 13: 645-653

Foottit RG, Adler PH (2017) Introduction. Insect Biodiversity. 1-7. doi:10.1002/9781118945568.ch1

Gaston KJ, Blackburn TM, Goldewijk KK (2003) Habitat conversion and global avian biodiversity loss. Proceedings of the Royal Society of London Series B: Biological Sciences 270: 1293-1300

Gill DA, Mascia MB, Ahmadia GN, Glew L, Lester SE, Barnes M, Craigie I, Darling ES, Free CM, Geldmann J (2017) Capacity shortfalls hinder the performance of marine protected areas globally. Nature 543: 665-669

Grogan J, Blundell AG, Landis RM, Youatt A, Gullison RE, Martinez M, Kometter R, Lentini M, Rice RE (2010) Over-harvesting driven by consumer demand leads to population decline: Big-leaf mahogany in South America. Conservation Letters 3: 12-20

Halderman JM (1985) Problems of pastoral development in Eastern Africa. Agricultural Administration 18: 199-216

Hauenstein S, Kshatriya M, Blanc J, Dormann CF, Beale CM (2019) African elephant poaching rates correlate with local poverty, national corruption and global ivory price. Nature communications 10: 1-9

Hawksworth DL (1995) Biodiversity: measurement and estimation. Springer Science & Business Media, pp.

Intergovernmental Panel On Climate Change (2001) Climate change 2007: Impacts, adaptation and vulnerability. Genebra, Suíça:

IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. In: Díaz S, Settele J, Brondízio E, Ngo H, Guèze M, Agard J, Arneth A, Balvanera P, Brauman K, Butchart S (Eds), pp.

Isaac NJ, Turvey ST, Collen B, Waterman C, Baillie JE (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. PloS one 2: e296

IUCN (2020) The IUCN Red List of Threatened Species. Version 2020-2. http://www.iucnredlist.org. pp.

Klopper RR, Smith GF, Chatelain C, Gautier L, Spichiger R (2006) Inventory of the African flora: a world first for the forgotten continent: news & views. South African journal of science 102: 185-186

Lande R (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos: 5-13

Linder HP (2003) The radiation of the Cape flora, southern Africa. Biological Reviews 78: 597-638

Linder HP, de Klerk HM, Born J, Burgess ND, Fjeldså J, Rahbek C (2012) The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa. Journal of Biogeography 39: 1189-1205

Lovejoy TE (1994) The quantification of biodiversity: an esoteric quest or a vital component of sustainable development? Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 345: 81-87

Magurran AE (1988) Why diversity? Ecological diversity and its measurement. Springer, 1-5

McClanahan TR, Maina JM, Graham NA, Jones KR (2016) Modeling reef fish biomass, recovery potential, and management priorities in the Western Indian Ocean. PloS one 11: e0154585

Morrison JC, Sechrest W, Dinerstein E, Wilcove DS, Lamoreux JF (2007) Persistence of large mammal faunas as indicators of global human impacts. Journal of mammalogy 88: 1363-1380

Muhumuza M, Balkwill K (2013) Factors affecting the success of conserving biodiversity in national parks: A review of case studies from Africa. International Journal of Biodiversity 2013:

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853

Nipperess DA (2016) The rarefaction of phylogenetic diversity: formulation, extension and application. Biodiversity Conservation and Phylogenetic Systematics 14: 197-217

Noss RF, Platt WJ, Sorrie BA, Weakley AS, Means DB, Costanza J, Peet RK (2015) How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. Diversity and distributions 21: 236-244

Oldekop JA, Holmes G, Harris WE, Evans KL (2016) A global assessment of the social and conservation outcomes of protected areas. Conservation Biology 30: 133-141

Oliveira U, Paglia AP, Brescovit AD, de Carvalho CJ, Silva DP, Rezende DT, Leite FSF, Batista JAN, Barbosa JPPP, Stehmann JR (2016) The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Diversity and distributions 22: 1232-1244

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GV, Underwood EC, D'amico JA, Itoua I, Strand HE, Morrison JC (2001) Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51: 933-938

Pickett ST, Burch WR, Dalton SE, Foresman TW, Grove JM, Rowntree R (1997) A conceptual framework for the study of human ecosystems in urban areas. Urban ecosystems 1: 185-199

Platteau J-P (2000) Does Africa need land reform? Evolving land rights, policy and tenure in Africa: 51-74

Plumptre AJ, Ayebare S, Behangana M, Forrest TG, Hatanga P, Kabuye C, Kirunda B, Kityo R, Mugabe H, Namaganda M (2019) Conservation of vertebrates and plants in Uganda: Identifying Key Biodiversity Areas and other sites of national importance. Conservation Science and Practice 1: e7

Pressey R, Johnson I, Wilson P (1994) Shades of irreplaceability: towards a measure of the contribution of sites to a reservation goal. Biodiversity & Conservation 3: 242-262

Pullin AS, Knight TM (2001) Effectiveness in Conservation Practice: Pointers from Medicine and Public Health. Conservation Biology 15: 50-54. doi:10.1111/j.1523-1739.2001.99499.x

Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000) Predicting extinction risk in declining species. Proceedings of the Royal Society of London Series B: Biological Sciences 267: 1947-1952

Richardson DM, Rouget M, Ralston SJ, Cowling RM, Van Rensburg BJ, Thuiller W (2005) Species richness of alien plants in South Africa: environmental correlates and the relationship with indigenous plant species richness. Ecoscience 12: 391-402

Roberts CM, McClean CJ, Veron JE, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295: 1280-1284

Rodrigues A, Brooks TM, Gaston KJ (2005) Integrating phylogenetic diversity in the selection of priority areas for conservation: does it make a difference. Phylogeny Conservation 8: 101-119

Rodrigues AS, Grenyer R, Baillie JE, Bininda-Emonds OR, Gittlemann JL, Hoffmann M, Safi K, Schipper J, Stuart SN, Brooks T (2011) Complete, accurate, mammalian phylogenies aid conservation planning, but not much. Philosophical Transactions of the Royal Society of London B: Biological Sciences 366: 2652-2660

Rosauer D, Laffan SW, Crisp MD, Donnellan SC, Cook LG (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. Molecular ecology 18: 4061-4072

Safi K, Armour-Marshall K, Baillie JE, Isaac NJ (2013) Global patterns of evolutionary distinct and globally endangered amphibians and mammals. PloS one 8: e63582

Saghir J, Santoro J (2018) Urbanization in Sub-Saharan Africa. Center for Strategic & International Studies Report, Washington, DC, USA www csis org:

Scholtz CH, Mansell MW (2017) Insect Biodiversity in the Afrotropical Region. Insect Biodiversity. 93-109. doi:10.1002/9781118945568.ch5

Secretariat of the Convention on Biological Diversity (2020) Global Biodiversity Outlook 5 – Summary for

Policy Makers., Montréal, pp.

Slatyer C, Rosauer D, Lemckert F (2007) An assessment of endemism and species richness patterns in the Australian Anura. Journal of Biogeography 34: 583-596

Smith RJ, Bennun L, Brooks TM, Butchart SH, Cuttelod A, Di Marco M, Ferrier S, Fishpool LD, Joppa L, Juffe-Bignoli D (2019) Synergies between the key biodiversity area and systematic conservation planning approaches. Conservation Letters 12: e12625

Solbrig OT (1991) The origin and function of biodiversity. Environment: Science and Policy for Sustainable Development 33: 16-38

Soria-Auza RW, Kessler M (2008) The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. Diversity and distributions 14: 123-130

Stattersfield AJ (1998) Endemic bird areas of the world-Priorities for biodiversity conservation. Bird Life International:

Sutherland WJ, Pullin AS, Dolman PM, Knight TM (2004) The need for evidence-based conservation. Trends in ecology evolution 19: 305-308

Sutherland WJ, Wordley CF (2017) Evidence complacency hampers conservation. Nature ecology & evolution 1: 1215-1216

Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. Scientific reports 7: 9132

Tucker C, Cadotte M (2013) Unifying measures of biodiversity: Understanding when richness and phylogenetic diversity should be congruent. 845-854 pp. doi:10.1111/ddi.12087

UNEP (2012) UNEP annual report 2012.

UNEP-WCMC (2016) The state of biodiversity in Africa: A mid-term review of progress towards the Aichi biodiversity targets.

United Nations (2017) World population prospects 2017. United Nations New York, pp.

Warchol GL (2004) The transnational illegal wildlife trade. Criminal justice studies 17: 57-73

Washington R, Harrison M, Conway D, Black E, Challinor A, Grimes D, Jones R, Morse A, Kay G, Todd M (2006) African climate change: taking the shorter route. Bulletin of the American Meteorological Society 87: 1355-1366

Watson JE, Dudley N, Segan DB, Hockings M (2014) The performance and potential of protected areas. Nature 515: 67-73

The world database on protected areas (WDPA). [accessed

Wilson EO (2000) A global biodiversity map. Science 289: 2279-2279

Wilson WJ, Primack RB (2019) Conservation Biology in Sub-Saharan Africa. Open Book Publishers, pp.

World Bank, Indicators WD (2020) GDP per capita (current US$). pp.

WWF (2020) Living planet report-2020: Bending the curve of biodiversity loss. In: Almond REA, Grooten M, Petersen T (Eds), Gland, Switzerland, pp.

Younge A, Fowkes S (2003) The Cape Action Plan for the Environment: overview of an ecoregional planning process. Biological Conservation 112: 15-28