UNIVERSIDADE FEDERAL DE ALAGOAS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
Programa de Pós-Graduação em Diversidade Biológica e Conservação nos Trópicos

THAINÁ LESSA PONTES SILVA

LIDANDO COM A INCERTEZA E IGNORÂNCIA NOS DADOS DA BIODIVERSIDADE:

uma perspectiva taxonômica e espacial

MACEIÓ - ALAGOAS
Março/2024

THAINÁ LESSA PONTES SILVA

LIDANDO COM A INCERTEZA E IGNORÂNCIA NOS DADOS DA BIODIVERSIDADE:

uma perspectiva taxonômica e espacial

Tese apresentada ao Programa de Pós-Graduação em Diversidade Biológica e Conservação nos Trópicos, Instituto de Ciências Biológicas e da Saúde. Universidade Federal de Alagoas, como requisito para obtenção do título de Doutor em CIÊNCIAS BIOLÓGICAS, área de concentração em Conservação da Biodiversidade Tropical.

Orientador: Prof. Dr. Richard James Ladle

MACEIÓ - ALAGOAS
Março/2024

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária: Girlaine da Silva Santos – CRB-4 – 1127

S586l Silva, Thainá Lessa Pontes
Lidando com a incerteza e ignorância nos dados da biodiversidade: uma perspectiva taxonômica e espacial / Thainá Lessa Pontes Silva. – 2024.
104 f.: il.: color.

Orientador: Richard James Ladle.

Tese (Doutorado em Ciências Biológicas) - Universidade Federal de Alagoas. Instituto de Ciências Biológicas e da Saúde. Programa de Pós-Graduação em Diversidade Biológica e Conservação nos trópicos, Maceió, 2024.

Inclui bibliografias.

1. Biodiversidade. 2. Lacuna Linneana. 3. Taxonomia. 4. Lacuna Wallaceana. 5. Distribuição geográfica. 5. Biogeografia. I. Título.

CDU: 574

Agradecimentos

Eu sempre gostei de pensar que os resultados dos projetos que a gente desenvolve, seja no TCC ou no doutorado, nos levam a contar uma história. Houve momentos que eu achava que a história desse meu doutorado não fazia sentido, e a prof. Ana Malhado, sempre otimista, dizia que ia dar certo, que perto do final essa história ficava encaixada. Bom, no final realmente deu certo. Eu me sinto muito orgulhosa do que foi feito até aqui, da história que vivi e contei nesses 4 anos.

Enfrentamos uma pandemia, vi meus orientadores atravessarem o oceano, sai da casa da minha mãe, fui morar com meu companheiro, trabalhei muito, viajei muito, fui à África e à Portugal 2x, passei por uma cirurgia, tive três sobrinhos lindos, e, é claro, comecei a terapia. Entre inseguranças e felicidades. Entre frustrações e realizações. Reconhecendo minhas qualidades e limitações. Consegui. Mas consegui porque tive muito apoio e colaboração de gente querida.

Então gostaria de agradecer aos meus orientadores, Richard Ladle e Ana Malhado, que tanto acreditaram, me apoiaram e me proporcionaram oportunidades incríveis nesses últimos anos. Estamos juntos no Lacos21 há mais de 10 anos, eles sempre me tirando da zona de conforto, mas eu não consigo dizer não.

Agradeço também ao meu grupinho SOS (Evelynne, Cacá, Carol e Ludmilla), minhas colegas de doutorado e meus amigos de laboratório, que dividimos tantos momentos, entre cafés e cervejas, nesse laboratório só tem pessoas/profissionais que tanto admiro e me inspiro. Eu sempre digo que sou formada em Lacos21.

Agradeço aos colaboradores dos artigos, em especial a Ju Stropp e Joaquín Hortal que estiveram comigo no primeiro capítulo; a Karol com K que segurou minha mão e me deu forças no segundo capítulo; e a Fernanda e o Javí que tanto me ajudaram e apoiaram no terceiro capítulo. Aprendi demais e me senti muito honrada em trabalhar com cada um de vocês.

Agradeço a Capes e ao Tropibio pelo financiamento para que essa pesquisa fosse realizada, apoiando com a bolsa de estudo, com as viagens e publicações dos artigos.

Agradeço a todos os membros de bancas que tive no decorrer deste processo. Em especial a banca de defesa, Prof. Robson, Prof. Guilherme, Dra. Geiziane e Dra. Filipa, que foi escolhida com muito carinho na certeza de que teríamos uma troca muito respeitosa e engrandecedora.

Agradeço ao meu companheiro, Gaspar, que é meu Porto Seguro, meu ombro amigo e meu fã nº 1, ele é a pessoa que mais me apoia e impulsiona, ainda por cima dividiu o peso de um doutorado junto comigo.

Agradeço minha família, minha mãe, meu pai, padrasto, irmãos, tios, primos, avó que mesmo sem entender muito bem o que eu faço, me apoiam e se orgulham de mim. Minha mãe, Stella, sempre falou para estudar e não ficar atrás do balcão. Mainha, acho que já tá bom de estudar, não é?

Agradeço aos meus sobrinhos (Gabriel, Caetano e Bernardo), e suas mães, minha irmã gêmea Thais, minha prima-irmã Juliane e minha amiga-irmã Yalli. Foram meu fôlego e ânimo nos momentos mais delicados. Cada sorrisinho deles, palavra de apoio delas, foi muito importante para mim.

E por fim, e não menos especial, eu agradeço as minhas amigas e meus amigos, sem citar nomes, pois tenho medo de esquecer de alguém. Mas se eu fechar os olhos, eu me lembro de cada um que entendeu meu cansaço e minha ausência, quando estava viajando ou quando neguei saídas para trabalhar, ou quando estava longe, quando deveria estar perto. Obrigada pelo apoio, colo e motivação. Atravessei esse rio, que foi uma vida, e estive muito bem acompanhada.

Resumo

Esta tese, composta por uma revisão da literatura e três capítulos em formato de manuscrito, oferece uma valiosa contribuição para a compreensão dos desafios relacionados à ignorância e incerteza nos dados taxonômicos e espaciais da biodiversidade. A revisão da literatura traz uma visão geral acerca das lacunas Linneana e Wallaceana. No primeiro capítulo, exploramos os múltiplos conceitos e métricas associadas à definição de espécies, ressaltando a complexidade e as incertezas taxonômicas resultantes. Enfatizamos a influência das mudanças taxonômicas nas estimativas de riqueza de espécies, destacando a necessidade de incluir informações sobre estas mudanças para melhorar a precisão das estimativas de biodiversidade. No segundo capítulo, a pesquisa ainda se concentra na lacuna Linneana relacionada à nomenclatura de aves globais, apresentando uma métrica de incerteza taxonômica, e avaliando associações entre características biológicas e ecológicas das aves à incerteza. O terceiro capítulo aborda a lacuna Wallaceana, avaliando as lacunas temporais e espaciais no esforço amostral da biodiversidade na Namíbia, utilizando dados do GBIF, e analisando qual a influência de variáveis sociogeográficas no esforço amostral. Concluímos evidenciando os desafios enfrentados por cientistas e coletadores em várias dimensões do conhecimento da biodiversidade, desde as definições de espécies à coleta e compartilhamento de dados. Ressaltamos a importância de investir em ciência de base para preencher as lacunas e garantir a ampla disseminação do conhecimento taxonômico e espacial das espécies.

Palavras-chave: lacunas do conhecimento, biodiversidade, taxonomia, distribuição, conservação.

Abstract

This thesis consists of a literature review and three chapters in manuscript format. We offer a valuable contribution to understanding the challenges related to ignorance and uncertainty in taxonomic and spatial biodiversity data. The literature review provides an overview of the Linnean and Wallacean shortfalls. In the first chapter, we explore the multiple concepts and metrics associated with species definition, highlighting the complexity and taxonomic uncertainties. We emphasize the influence of taxonomic changes on species richness estimates, pointing out the need to include information on these changes to improve the precision of biodiversity estimates. In the second chapter, the research still focuses on the Linnean shortfalls related to the nomenclature of global birds, presenting a metric of taxonomic uncertainty, and evaluating associations between biological and ecological characteristics of birds and taxonomic uncertainty. The third chapter addresses the Wallacean shortfalls, assessing the temporal and spatial gaps in the biodiversity sampling effort in Namibia, using GBIF data, and analyzing the influence of sociogeographic variables on the sampling effort. We conclude this thesis by arguing the challenges faced by scientists and collectors in several dimensions of biodiversity knowledge, from species definitions to data collection and sharing. We emphasize the importance of investing in basic science to fill the gaps and ensure the broad dissemination of taxonomic and spatial knowledge of species.

Key-word: knowledge shortfalls, biodiversity, taxonomy, distribution, conservation.

Lista de Figuras

Revisão da Literatura

Capítulo 1: How taxonomic change influences forecasts of the Linnean Shortfall (and what we can do about it)

Capítulo 2: Do biological and ecological variables influence nomenclatural uncertainty in birds?

Capítulo 3: Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa

Lista de Tabelas

Revisão da Literatura

Capítulo 2: Do biological and ecological variables influence nomenclatural uncertainty in birds?

Capítulo 3: Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa

Sumário

Apresentação

Espécie é a unidade taxonômica primordial para investigar os demais componentes da biodiversidade. A partir do reconhecimento, descrição e nomeação das espécies, que cientistas podem então conhecer sobre dados populacionais, taxas de endemismo, status de conservação, quais as relações evolutivas, qual nicho e papel ecológico daquela peça na complexidade da natureza. Faz parte da cultura humana classificar a vida, e por isto ainda não existe um consenso universal sobre o que são espécies e como nomeá-las. Muitos conceitos de espécies foram lançados na literatura, sejam mais tradicionais baseando-se em caracteres morfológicos, ou mais sofisticados, incluindo diversas fontes de informação. Como consequências, Espécies são consideradas hipóteses, que ao serem reveladas para comunidade científica podem ou não ser aceitas e adotadas.

Estas diversas fontes de informação, conceitos e métricas, podem - e acarretam - em incertezas taxonômicas, que, por conseguinte, afetam as estimativas de riqueza de espécies. Além disso, os taxonomistas, isto é, os cientistas que têm o papel de descrever e nomear as espécies se baseiam em códigos internacionais de nomenclatura, que por sua vez, permite liberdade na escolha dos nomes científicos. Isto também gera casos de sinonímia, quando a mesma espécie foi nomeada com diferentes nomes científicos, ou ainda processos de mudanças taxonômicas, quando as espécies são trocadas, agrupadas ou divididas. A partir dos processos de revisões taxonômicas que acontece o intercâmbio taxonômico (swap), ou seja, quando a espécie muda completamente de táxon, o agrupamento (lumping) que é quando várias espécies (ou subespécies) são consideradas parte da mesma espécie, ou o processo inverso, ocorrendo à divisão (splitting) de uma espécie em várias diferentes.

Para a elaboração de planos e ações de conservação e para a avaliação de como a biodiversidade está se adaptando às mudanças evolutivas e antrópicas na natureza, é essencial que os dados e informações da biodiversidade sejam disponíveis e que estes apresentem qualidade e confiabilidade suficientemente aplicável. Iniciativas nacionais e internacionais têm aumentado o compartilhamento de dados em todo o mundo, diminuindo assim fronteiras para o desenvolvimento de pesquisas. Atualmente, uma iniciativa de grande relevância por disponibilizar mais de 2.6 bilhões de dados, tanto de coleções museológicas quando de ciência cidadã, é o Global Biodiversity Information Facility (GBIF). Entretanto, estes dados não são distribuídos de forma homogênea, e existem muitas tendências e vieses, seja na distribuição espacial, taxonômica e temporal. Diante dos desafios apontados, a presente tese foca nas lacunas Linneana e Wallaceana, isto é, nos respectivos conhecimentos (ou escassezes) da taxonomia e distribuição geográfica da biodiversidade.

A tese é dividida em duas partes, sendo a primeira parte uma revisão da literatura sobre a lacuna Linneana, abrangendo os diversos conceitos de espécies e a problemática por trás disto, e a lacuna Wallaceana, com foco nos vieses de recolha de dados biológicos. A segunda parte trata-se de três capítulos em formato de manuscrito.

O primeiro capítulo é um manuscrito intitulado *"How taxonomic change influences forecasts of the Linnean Shortfall (and what we can do about it)"* que foi publicado na Journal of Biogeography. Elaboramos uma perspectiva sobre os impedimentos que cientistas enfrentam para estimar a lacuna Linneana, ou seja, desvendar o número de espécies existentes e descritas no mundo. Além disso, discutimos que as previsões de riqueza de espécies são influenciadas tanto pelas taxas de exploração e amostragem, como também podem ser superestimadas ou subestimadas pelas mudanças taxonômicas, especialmente, as divisões e agrupamentos taxonômicos. Finalmente, abordamos a importância de incluir estas informações nas métricas de estimativas da biodiversidade para torná-las mais precisas e confiáveis, porém reconhecendo que este tipo de conhecimento não é de amplo acesso nos bancos de dados ou são mal documentados para maioria dos táxons.

No segundo capítulo, o manuscrito intitulado *"Do biological and ecological variables influence nomenclatural uncertainty in birds?"*, abordamos a problemática da lacuna Linneana quanto à nomenclatura das espécies. Precisamente, nós avaliamos as várias listas de aves globais, que além de apresentarem números divergentes de espécies, constam divergências de nomenclatura, isto é, espécies que apresentam nomes científicos diferentes, seja nome genérico, nome específico e/ou nome completo. Criamos e calculamos para 11.140 espécies de aves, uma métrica para avaliar a incerteza de nomenclatura, utilizando a proporção de discordância (e ausência) de nomes científicos entre as listas globais de aves. Posteriormente, nós avaliamos se e quais características biológicas e ecológicas (massa corporal, tamanho da área de distribuição, densidade do habitat, estilo de vida, status da IUCN e distinção evolutiva) das aves estão associadas com a incerteza de nomenclatura.

O terceiro e último capítulo, possui o título *"Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa"* e foi publicado na Ecological

Indicators. Tratamos sobre a lacuna Wallaceana no sentido de avaliar as tendências e lacunas temporais e espaciais no esforço amostral da biodiversidade. Como estudo de caso, utilizamos a biodiversidade terrestre (aves, mamíferos, répteis, anfíbios e insetos) da Namíbia por meio de dados do GBIF como estudo de caso, e quantificamos as lacunas espaciais pela métrica de "Ignorance Score". A biodiversidade da Namíbia é extremamente diversificada e adaptada às condições áridas da região. Apesar da baixa densidade populacional, a Namíbia apresenta um forte sistema de áreas protegidas, com cerca de 40% do seu território protegido. Avaliamos quais variáveis sociogeográficas (densidade rodoviária, densidade populacional humana, distância de instituições de pesquisa, distância de área protegida e cobertura vegetal) influenciam as taxas de ignorância espacial da biodiversidade. Finalizamos reforçando sobre usabilidade e confiabilidade das recentes abordagens para avaliar qualidade de esforço amostral e incertezas espacial, temporal e taxonômica.

Revisão da Literatura

Sistemas de Classificação dos Organismos

Caracterizar os organismos e descrever processos e padrões do mundo natural faz parte da característica humana há milênios. Um notável exemplo foram os feitos dos filósofos gregos Aristóteles (384 a.C. – 322 a.C.) e seu discípulo Teofrasto (372-287 a.C.). Aristóteles criou uma classificação dos organismos vivos, separando-os pela presença de sangue e similaridades morfológicas. Ele delimitou alguns grupos faunísticos, como mamíferos, peixes, aves e insetos, embora tenha realizado alguns enquadramentos errôneos, como não ter incluído as baleias aos mamíferos, devido à diferença no formato corporal (KLEPKA; CORAZZA, 2018). Aristóteles acreditava que os organismos deveriam ser classificados pela ordem crescente de complexidade. Já Teofrasto elaborou uma classificação botânica, agrupando plantas de acordo com seus habitats, periodicidade e porte, dividindo assim as herbáceas, os arbustos e as árvores, e as estações anuais, bienais e perenes (GOMES-DA-COSTA, 2010). Estas ideias foram perduradas por mais de 2.000 anos, até ser superada pela classificação sistematizada proposta por Carl von Linné (Lineu) (1707 – 1778).

Na classificação de Lineu existiam cinco seções: classe, ordem, gênero, espécie e variedade. Na seção gênero eram agrupadas as características morfológicas entre as espécies, que eram obtidas em três diferentes caracteres: artificial, essencial e/ou natural (KLEPKA; CORAZZA, 2018). O termo "características essenciais" era empregado para descrever os atributos necessários que diferenciava uma espécie da outra. Seu sistema não reflete necessariamente as relações entre os organismos e seu grau de parentesco. Entretanto, sua trajetória foi marcada pela criação do sistema de binômios latinos para os nomes das espécies de plantas e animais. Mais de 250 anos após Lineu iniciar seu inventário sistemático das espécies do mundo, ainda há debates consideráveis sobre quantas espécies realmente existem e quantas delas já foram documentadas. Isso acontece devido à biodiversidade, complexidade e dinamismo da natureza, adicionado à baixa capacidade humana de coletar e identificar as espécies (LADLE; HORTAL, 2013). Deste modo, alcançar uma compreensão abrangente sobre qualquer aspecto da biodiversidade permanece em grande parte inalcançável.

O número estimado de espécies viventes em todo o mundo varia entre 3 e 100 milhões (MAY, 2010), baseado em opinião de especialistas, entretanto este número é divergente ao depender da métrica utilizada (Tabela 1). Apenas uma pequena fração de toda a biodiversidade foi formalmente descrita, cerca de 2 milhões de espécies (BÁNKI et al., 2022), e muitas foram extintas antes mesmo de serem descobertas (COSTELLO; MAY; STORK, 2013). Da mesma maneira, o conhecimento sobre a distribuição das espécies é espacialmente heterogêneo, com muitas áreas e ambientes mais conhecidos que outros. Logo, por que algumas espécies e regiões receberam devida atenção e empenho para uma avaliação mais completa do que outras? São características intrínsecas ou extrínsecas às espécies que norteiam nosso conhecimento da vida? São ambientes mais acessíveis que aumentam o conhecimento da sua biodiversidade? Tais classificações refletem os objetivos de quem os fazem, ou seja, são agregados aspectos culturais, temporais e espaciais, servindo como ilustrações para representar a realidade e produzir conhecimento científico (ROSEN, 1996). Sendo assim, a mensuração da biodiversidade (e seus aspectos taxonômicos, espaciais e temporais) constitui apenas uma das múltiplas abordagens para enquadrar a diversidade da vida.

Lacunas do Conhecimento da Biodiversidade

Atualmente, a ausência ou carência de conhecimento sobre os aspectos da biodiversidade são divididos em sete categorias de lacunas (BROWN; LOMOLINO, 1998; HORTAL et al., 2015). Estas lacunas representam importantes áreas da biologia, e onde a pesquisa-ação é particularmente necessária. Resumidamente, são elas: a) Lacuna Linneana (taxonomia): muitas espécies ainda não foram descritas ou catalogadas; b) Lacuna Wallaceana (distribuição geográfica): falta de informações sobre a distribuição de espécies em diferentes regiões geográficas; c) Lacuna Prestoniana (ecologia de populações): falta conhecimento sobre os tamanhos populacionais, dinâmicas e abundância das espécies; c) Lacuna Darwiniana (evolução): falta de conhecimento sobre a evolução e árvore da vida das espécies; d) Lacuna Raunkiareana (traços funcionais): desconhecimento sobre as funções e características ecológicas das espécies; e) Lacuna Hutchinsoniana (tolerâncias abióticas): falta de conhecimento sobre como as espécies se adaptam e toleram as mudanças ambientais; f) Lacuna Eltoniana (interações ecológicas): falta de compreensão sobre como as espécies interagem e sobrevivem.

Estas lacunas do conhecimento da biodiversidade significam que os cientistas estão trabalhando muitas vezes com dados incompletos, incertos e não representativos e sobre um número limitado de organismos e suas características (LADLE; HORTAL, 2013). Isto compromete, por exemplo, a capacidade de descrever a biodiversidade existente, fazer previsões precisas sobre como os organismos estão sendo impactados por mudanças ambientais e como podem se adaptar, e utilizar de forma ineficiente de recursos para conservação. Todas as lacunas do conhecimento estão interconectadas em graus variados, de acordo com a escala e a cobertura espacial, temporal e taxonômica (HORTAL et al., 2015). Logo, reconhecer e quantificar as lacunas no conhecimento sobre da biodiversidade global é imperativo (CARDOSO et al., 2011).

A Lacuna Linneana

A lacuna Linneana afeta criticamente todas as demais lacunas, pois a falta de informação sobre a identidade das espécies impede necessariamente a descrição de qualquer outra característica atribuída e relação envolvida (WHITTAKER et al., 2005). Antes mesmo de existir qualquer estudo acerca da biologia, ecologia e/ou evolução da biodiversidade, é fundamental reconhecer o grupo taxonômico em questão. A taxonomia é a disciplina da biologia com o objetivo de identificar, descrever, classificar e nomear os organismos vivos e extintos. Espécie é comumente a hierarquia avaliada para atingir metas de conservação, como por exemplo, a proteção de espécies endêmicas, emblemáticas e ameaçadas de extinção e os ambientes em que habitam (AGAPOW et al., 2004; BARROWCLOUGH et al., 2016; LADLE; WHITTAKER, 2014).

Mesmo no século XXI, muitas espécies continuam sendo descobertas, especialmente devido aos esforços de taxonomistas, avanços nas análises genéticas e reavaliações de acervos museológicos (LADLE; WHITTAKER, 2014). Em um mundo perfeito, cada espécie recém-descrita teria sido coletada, meticulosamente documentada, identificada de forma inequívoca, nomeada e alocada permanentemente um ramo único na árvore da vida, baseados em métodos idênticos e em um conceito universalmente aplicado do que constitui uma espécie (STROPP et al., 2022). Entretanto, no mundo real, os taxonomistas têm usado uma variedade de conceitos de espécies e abordagens (relevantes, porém limitantes) para identificar novas espécies e reavaliar espécies já documentadas (Tabela 1) (KITCHENER et al., 2022; ZACHOS, 2016).

Tabela 1: Principais abordagens para estimar a lacuna Linneana (modificado de MORA et al., 2011); as limitações foram descritas a partir da análise das abordagens pela autora desta tese e seu orientador.

Abordagem	**Descrição e referência**	**Limitação**
Índices de diversidade	Multiplicação das proporções relativas de espécies que são conhecidas de ocorrer em um habitat para estimar os números prováveis de ocorrer em outros habitats (ERWIN, 1982; GARCÍA-ROBLEDO et al., 2020; HAMILTON et al., 2010)	A estimativa não é confiável para grupos taxonômicos hiperdiversos e pouco estudados.
Curvas de descoberta	Extrapolando a partir de taxas históricas de descoberta de espécies (BEBBER et al., 2007)	As variações temporais no processo de descoberta tornam as curvas pouco informativas.
Opinião de especialistas	Suposições fundamentadas ou pesquisas de opinião de especialistas, como o método Delphi (FISHER et al., 2015)	Há grande incerteza sobre as estimativas, sem protocolos, variando para mais ou menos. Interessante apenas para táxons novos.
Extrapolação de área bem amostrada	Extrapolação de proporções de espécies não descritas de uma área intensamente estudada para a extensão total dessa área (SHEN; CHAO; LIN, 2003)	O desempenho do estimador deteriora-se à medida que o tamanho da amostra prevista aumenta.
Extrapolação de táxons bem conhecidos	Extrapolação a partir de proporções de números de espécies em categoria taxonômica alta para prever os números em áreas onde apenas alguns táxons foram bem descritos (MORA et al., 2011)	As estimativas são confiáveis para mudanças taxonômicas em táxons superiores. A interpretação da extrapolação no táxon inferior deve ser feita com cautela.
Relação tamanho corporal	Extrapolação das relações entre tamanhos corporais e números de espécies (ZAPATA; ROSS ROBERTSON, 2007)	As estimativas utilizam apenas espécies conhecidas. Incomparável entre grupos taxonômicos.
Relação espécie-área	Estimando a riqueza de espécies de uma grande área a partir de pesquisas locais dispersas dentro dela (KUNIN et al., 2018)	Em regiões pouco exploradas, muitos táxons são consequentemente mal amostrados, tornando inviável o acúmulo de informações.
Taxas de Sinonímia	Extrapolação da razão entre espécies aceitas e sinônimos observados para famílias de plantas individuais (GOVAERTS, 2001; SCOTLAND; WORTLEY, 2003)	Muitas fontes de erros estão associadas a dados taxonômicos, tornando-os pouco confiáveis. Não utiliza dados históricos.
Estimativas de probabilidades de descoberta em nível de espécie	Usando atributos biológicos, ambientais e sociológicos em nível de espécie para avaliar a probabilidade de descoberta de espécies (MOURA; JETZ, 2021)	Existem lacunas de dados multivariados para muitos táxons. O modelo não é capaz de distinguir definições de espécies válidas e mudanças nas práticas taxonômicas.
Extrapolações de níveis de amostragem molecular	Usando *machine learning* para criar um modelo preditivo que identifica espécies nomeadas que provavelmente contêm diversidade oculta com base nos níveis de cobertura de dados genéticos (PARSONS et al., 2022)	Não existe um banco de dados de sequências genéticas para muitos táxons. Ainda é necessário revisar muitas sequências problemáticas.

São dois grandes desafios que a taxonomia moderna enfrenta: o primeiro é encontrar um consenso sobre a definição de espécies, e o segundo é delimitar e descrever o número de espécies existente em todo o mundo (PADIAL et al., 2010). A verdade é que espécies são hipóteses, e frequentemente seu conceito é baseado em padrões de similaridade, ancestralidade, evolução e filogenia (PANTE et al., 2015). Em 1997, Mayden realizou um levantamento sobre os conceitos teóricos de espécies (MAYDEN, 1997), e encontrou 24 conceitos, cujas incompatibilidades de definição poderiam gerar conclusões divergentes quanto ao número e limites do que é espécie.

Os conceitos de espécies mais tradicionais na literatura são: a) Conceito biológico: espécie é um grupo de organismos que são capazes de se reproduzir entre si, produzindo descendentes férteis, e que estão reprodutivamente isolados de outros grupos. Isso significa que os membros de uma espécie podem se acasalar e gerar descendentes viáveis, enquanto são isolados de outros grupos reprodutivamente (MAYR; PROVINE, 1980); b) Conceito ecológico: espécie é definida como uma linhagem que ocupa uma zona adaptativa (ou seja, determinadas pelos recursos explorados e hábitats ocupados) divergente de outras linhagens, na sua área de distribuição, e que evolui separadamente de todas as linhagens (VAN VALEN, 1976); c) Conceito evolutivo: uma espécie é vista como uma linhagem única de organismos que compartilham características derivadas de seu ancestral comum e que mantém suas características separada de outras linhagens, como adaptações específicas ao ambiente, modificações genéticas ou outras características que evoluíram ao longo do tempo (WILEY, 1978); d) Confeito Filogenético: espécie é considerada como o menor grupo de um conjunto de organismos que compartilham um padrão parental de ancestralidade e descendência, isto é, um ancestral comum exclusivo (monofilético) (CRACRAFT, 1983).

Na prática os diferentes conceitos de espécies têm consequências importantes na biologia e conservação (DE QUEIROZ, 2007). Por exemplo, podem alterar as estimativas de riqueza de espécies. O conceito filogenético de espécies, baseado em ancestralidade, tem sido apontado por reconhecer um maior número de espécies quando comparado ao conceito biológico, baseado em morfologia. Um estudo realizado em 2005 avaliou as implicações da aplicação de diferentes conceitos de espécies nas estimativas de riqueza de aves da África subsaariana, e revelou que houve um aumento de 33% no número de espécies quando utilizado o conceito de espécies filogenéticas (n= 2.098) em relação ao conceito de espécies biológicas (n= 1.572) (DILLON;

FJELDSÅ, 2005). O mesmo padrão foi observado para aves do México (NAVARRO-SIGÜENZA; PETERSON, 2004; PETERSON; NAVARRO-SIGÜENZA, 1999). Além disso, os diferentes conceitos de espécies podem afetar estudos de história natural, padrões de fluxo gênico e avaliações de áreas geográficas, como a delimitação de áreas de endemismos, áreas de alta biodiversidade e área prioritárias para conservação (BATES; DEMOS, 2001; MEIJAARD; NIJMAN, 2003; PETERSON; NAVARRO-SIGÜENZA, 1999). Portanto, identificações erradas de espécies nos levam a respostas erradas ou inconsistentes para questões da biologia e conservação.

A taxonomia tradicional, baseada em morfoespécies, não leva em consideração a biodiversidade críptica, isto é, espécies (ou complexo de espécies) que são extremamente semelhantes na morfologia, porém apresentam distinções genéticas e/ou ecológicas (SHIN; ALLMON, 2023). Apesar de não existir um consenso taxonômico universal, e taxonomistas de diferentes áreas usam diversas práticas para descrever espécies, já está sendo notado que o estudo da especiação é crucial para a taxonomia. As espécies serão mais precisamente delimitadas quando entendermos as causas de sua origem e as influências em suas trajetórias evolutivas, portanto, é necessário adotar uma perspectiva múltipla e complementar neste processo (PADIAL et al, 2010).

Nas últimas décadas, houve um aumento expressivo nas taxas globais de descrição para muitos grupos de organismos, especialmente devido aos avanços de novas técnicas computacionais e ao acesso de dados museológicos, geográficos e genéticos, (TROUDET et al., 2017). A taxonomia integrativa surge assim como arcabouço para resolver os conflitos entre delimitações de espécies e revelar a biodiversidade oculta de forma mais rápida e confiável do que os métodos tradicionais (PADIAL et al., 2010). Nesta abordagem são utilizadas múltiplas fontes de dados complementares (morfológicos, moleculares, ecológicos, comportamentais e geográficos) para descrever espécies, buscando uma compreensão mais holística das relações filogenéticas e taxonômicas (DAYRAT, 2005). A taxonomia integrativa é particularmente útil para resolver problemas taxonômicos complexos, como a delimitação de espécies crípticas (espécies que são morfologicamente semelhantes, mas geneticamente distintas), a identificação de populações em processo de especiação e a revisão de grupos taxonômicos mal definidos.

A nomenclatura é o sistema pelo qual os taxonomistas se baseiam para nomear e classificar as espécies/táxons. Para isto, os taxonomistas de diversas áreas utilizam as regras estabelecidas pelos seus respectivos Códigos internacionais de nomenclatura, que por sua vez, permitem liberdade quanto à escolha dos critérios e atributos, aumentando incertezas e alterações taxonômicas (RHEINDT et al., 2023). A incerteza taxonômica refere-se à falta de clareza ou confiança na classificação de organismos em categorias taxonômicas específicas, como gênero, espécie e/ou família. À medida que novas evidências surgem, especialmente através de estudos genéticos e filogenéticos, a classificação taxonômica de certos grupos de organismos pode ser revisada. Isso pode levar a mudanças na taxonomia que geram incerteza sobre a classificação anterior (LESSA et al., 2024).

A mobilização (ou a falta) de dados também é um dos fatores que afeta a incerteza taxonômica. Para alguns grupos taxonômicos, pode haver uma escassez de dados disponíveis, sejam morfológicos, moleculares ou ecológicos, o que torna difícil uma classificação precisa. Além disso, uma proporção das espécies que foram formalmente descritas ainda não foi incluída em monografias taxonômicas ou listas de verificação com curadoria (HORTAL et al., 2019). As listas de verificação de espécies aceitas, cujo principal exemplo é o Catálogo da Vida (HOBERN et al., 2021), procuram fornecer um consenso sobre todas as espécies válidas (e outras categorias de classificação taxonômica superior e inferior). A descrição de uma espécie em periódico especializado pode gerar um processo de discussão e revisão por parte da comunidade taxonômica que inevitavelmente gera um lapso de tempo antes de sua aceitação geral e da inclusão nessas listas (Figura 1).

Além disso, pode gerar potenciais divergências se o consenso diferir entre diferentes comunidades científicas que utilizam conceitos de espécie diferentes e, assim, mantêm listas de verificação diferentes para o mesmo táxon (NEKOLA; HORSÁK, 2022). Por exemplo, mesmo em grupos taxonômicos bem estudados, como as aves, as lacunas e incertezas taxonômicas são proeminentes. Como reflexo, existem múltiplas listas de aves globais, com base em múltiplos critérios taxonômicos, e múltiplos propósitos científicos e conservacionistas (McCLURE, 2020; NEATE-CLEGG, 2021). Incertezas taxonômicas podem levar a estimativas equivocadas da riqueza de espécies. Um estudo recente descobriu que 68% de quatro listas de aves apresentam algum grau de discordância, seja por omissões taxonômicas, uso de nome científico diferente (nome completo, epíteto e/ou gênero) ou tratamento em nível taxonômico (espécie e subespécie) (McCLURE, 2020). Portanto, lidar com este desafio é crucial e requer compromissos com a conceituação das espécies.

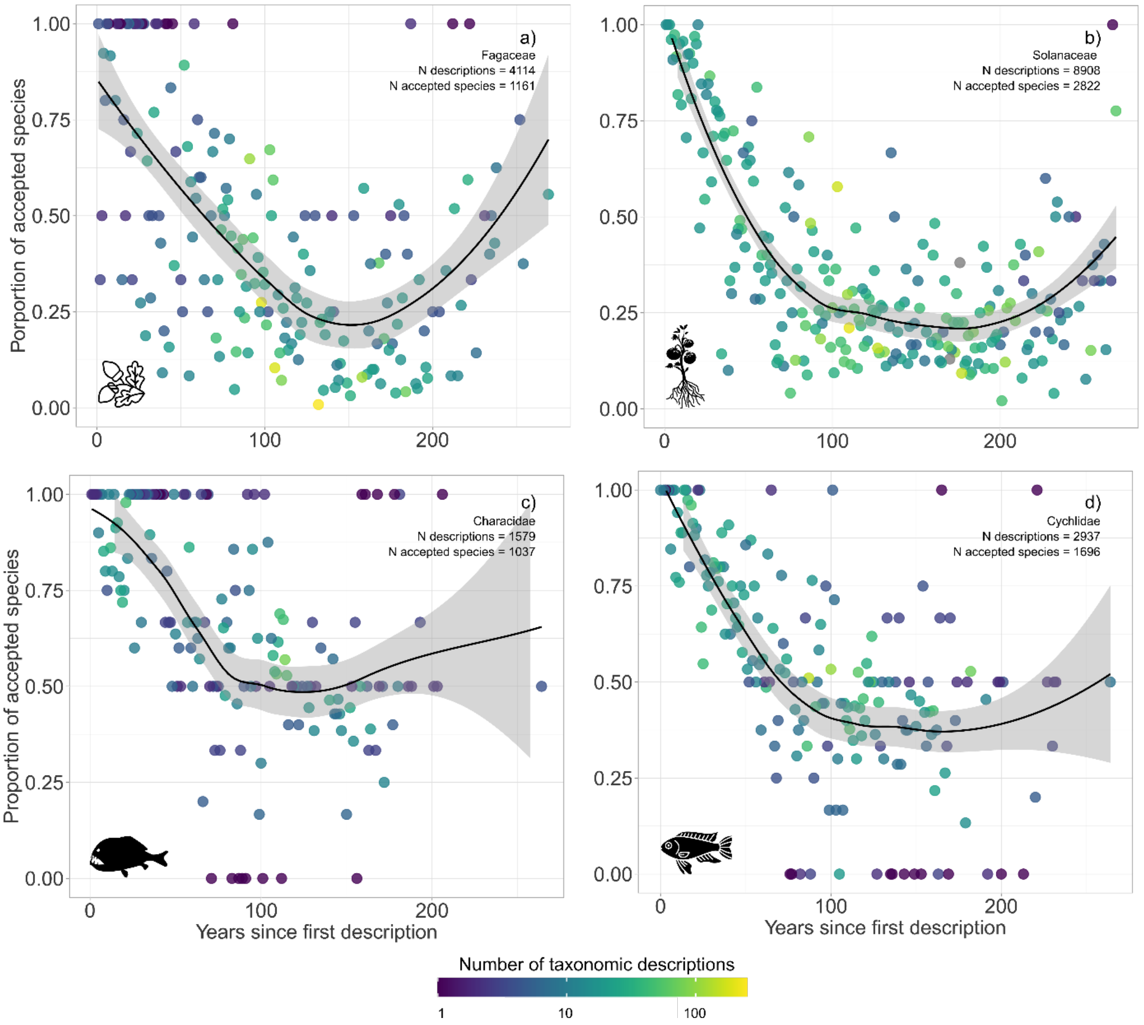


Figura 1: Proporção de espécies atualmente aceitas em relação ao total de descrições taxonômicas feitas a cada ano; as tonalidades das cores indicam o número de descrições taxonômicas feitas a cada ano em escala logarítmica, sendo que as tonalidades mais claras indicam maior número de descrições. Os painéis a) e b) representam dados de duas famílias de plantas, Fagaceae e Solanaceae, enquanto os painéis c) e d) o fazem para duas famílias de peixes, Cychlidae e Characidae. Fontes de dados: A lista de nomes de táxons, status taxonômico e ano de publicação foi recuperada de Govaerts (2022) para as famílias de plantas, e de Froese et al. (2022) para as famílias de peixes. Créditos da imagem: silhuetas foram baixadas em https://thenounproject.com; Fagaceae (carvalho de Eucalyp); Solanaceae (tomateiro de Michael Zick Doherty); Cychlidea (Cichlid por Ametyst Studio); Characidae (piranha de Agne Alesiute). Figura elaborada pela Dra. Juliana Stropp, compartilhada com permissão. Os dados e o código R estão disponíveis em https://github.com/justropp.

Existem pelo menos duas dimensões relacionadas à incerteza taxonômica: a) o erro taxonômico: onde a identificação da espécie foi feita utilizando múltiplas taxonomias e/ou múltiplos conceitos de espécies, espécies que passaram por revisões taxonômicas, espécies em táxons mal resolvidos (complexo de espécies); b) o erro de identificação: o identificador/taxonomista fez uma identificação incorreta quanto ao nome científico, ou seja, pode depender da experiência do identificador (amador, especialista ou taxonomista; TESSAROLO et al., 2021), do tipo de dado (fotografias, observações humanas ou espécimes de museu), pode ocorrer em espécies simpátricas ou ainda táxons mais recentes.

As incertezas taxonômicas mais comuns são relacionadas à sinonímia, isto é, quando o mesmo táxon (por exemplo, espécie) possui nomes científicos diferentes (MORA et al., 2011). Isso ocorre quando: i) taxonomistas de regiões diferentes estão trabalhando de forma independente e publicam novas espécies sem o conhecimento prévio; e ii) espécies estão "ocultas" dentro de outras espécies e são reveladas após de revisões taxonômicas (DAYRAT, 2005). Neste último contexto, as revisões taxonômicas podem gerar divisões (splitting), quando uma espécie é separada em mais de uma espécie ou subespécies, ou o oposto, gerando agrupamentos (lumping), quando várias espécies que eram reconhecidas como distintas são na verdade a mesma espécie. Estes processos de revisões e alterações taxonômicas podem levar a inflação ou deflação taxonômica, isto é, o aumento ou diminuição do número de espécies. Tem havido intensos debates sobre se a inflação/deflação taxonômica é real, se é simplesmente um reflexo do progresso taxonômico, e se influenciará significativamente as percepções dos padrões de biodiversidade (GARNETT; CHRISTIDIS, 2017; HARRIS; FROUFE, 2005; SANGSTER, 2009; WHEELER, 2020). Por exemplo, as estimativas do número de espécies de palmeiras neotropicais (gênero *Attalea*) eram muito maiores na década de 1970, antes que muitas espécies anteriormente válidas fossem agrupadas (HENDERSON, 2020) (Figura 2).

Está cada vez mais evidente que o preenchimento desses déficits é extremamente desafiador devido à existência de vários impedimentos, como a escassez de taxonomistas para muitos táxons (ENGEL et al., 2021), e uma série de outros obstáculos técnicos à coleta, compilação e análise de múltiplas formas de dados sobre a biodiversidade. Em particular, os dados de ocorrência de espécies derivados de espécimes armazenados em coleções museológicas sofrem de qualidade incerta e vieses espaciais e temporais (STROPP et al., 2016). Além disso, os dados sobre alterações taxonômicas não estão atualmente facilmente disponíveis para a grande maioria das espécies, e quando disponíveis é preciso muito tempo para compilação e processamento (KONSTANTINOV; NAMYATOVA, 2019), o que torna a análise das tendências taxonômicas históricas altamente desafiadoras e acrescentando incerteza a muitos métodos de previsão para avaliar a Lacuna Linneana (STROPP et al., 2022). Mesmo que todos os taxonomistas concordassem em usar exclusivamente um único conceito de espécie, ainda haveria divergências sobre a delimitação das espécies, uma vez que existe um elemento de julgamento humano ao decidir se uma ou mais populações merecem ser designadas como espécie (ZACHOS, 2018).

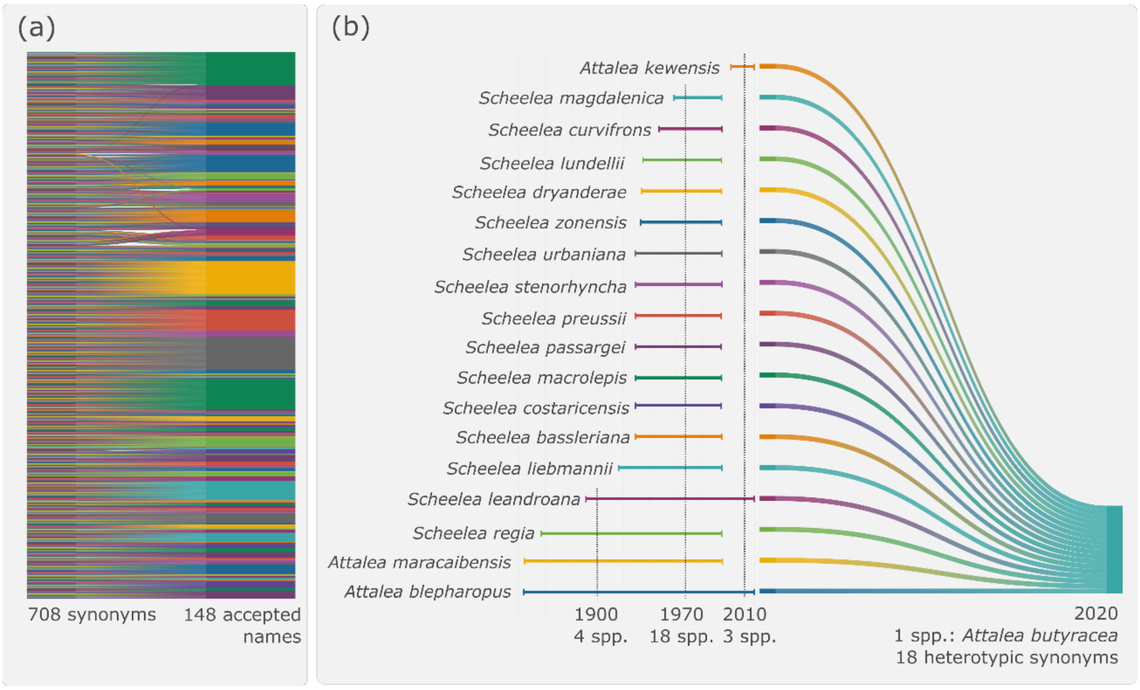


Figura 2: Mudança taxonômica em palmeiras amazônicas (Arecaceae; Palmae). O painel (a) mostra a reclassificação taxonômica de 708 sinônimos em 148 nomes aceitos; cada cor representa um nome de espécie. O painel (b) descreve detalhadamente a ligação entre sinônimos heterotípicos e nome aceito em uma linha do tempo de agrupamento taxonômico (lumping) para um desses 148 nomes aceitos, Attalea butyracea; linhas coloridas horizontais marcam o ano de descrição de 18 sinônimos heterotípicos e o ano de sinonimização, sendo que cada cor representa um sinônimo atual; as linhas verticais indicam a contagem de nomes de espécies aceites num determinado ano; linhas curvas representam o agrupamento de 18 sinônimos heterotípicos em A. butyracea. Fontes de dados: para o painel (a) a lista de nomes aceitos de palmeiras amazônicas foi extraída de (Cardoso et al., 2017; ter Steege et al., 2019), enquanto a lista de sinônimos foi obtida de (Govaerts et al., 2022); e a lista de sinônimos heterotípicos mostrada no painel (b) foi obtida de Henderson (2020). Figura publicada em short communication no Journal of Biogeography (https://doi.org/10.1111/jbi.14463), compartilhada com permissão dos autores.

A ciência já te avançado para preencher as lacunas Linneana e reduzir as incertezas taxonômicas. Como por exemplo, debatendo sobre os conceitos e definições de espécies, utilizando diversas fontes de dados para fortalecer a hipótese de espécie, como também, compartilhando em banco de dados digitais informações taxonômicas, como descrições e listas de espécies e revisões taxonômicas. Entretanto, ainda há muito trabalho e esforços a serem enfrentados. As estimativas atuais demonstram que estamos 400 anos atrasados em pesquisas taxonômicas, para obtenção de um inventário completo (PADIAL et al., 2010). Considerando que as intensas atividades e pressões humanas sobre a natureza estão levando à extinção de espécies, é urgente reconhecer a biodiversidade existente, para revelar seus valores. Portanto, para melhor a qualidade e previsões taxonômicas é necessário um esforço conjunto de taxonomistas e macroecologistas para: i) documentar e descrever a história taxonômica das espécies, incluindo informações sobre divisões, agrupamentos e revisões; ii) incorporar estas informações em modelos de estimativas de riqueza de espécies; e iii) prever como estas mudanças taxonômicas poderão remodelar os padrões de riqueza de espécies (STROPP et al., 2022).

A Lacuna Wallaceana

O conhecimento sobre a taxonomia e distribuição geográfica das espécies é estritamente conectado e considerado de fundamental importância para estudos da biodiversidade (BINI et al., 2006; HORTAL et al., 2015). A falta de conhecimento sobre a distribuição geográfica das espécies é conhecida como lacuna Wallaceana (HORTAL et al., 2015; LOMOLINO, 2004). Esta lacuna é impulsionada pela variação temporal e espacial no esforço de amostragem. A variação espacial no esforço amostral é heterogênea resultando em áreas – sejam países, regiões ou ecossistemas – mal representadas em coleções científicas e em bancos de dados de biodiversidade (LOBO, 2008). Isto ocorre especialmente em áreas remotas, como densas florestas tropicais ou zonas abissais, áreas montanhosas ou regiões de clima extremamente árido, que são consideradas historicamente negligenciadas por taxonomistas e biogeógrafos (LADLE; WHITTAKER, 2014; LESSA et al., 2019).

Além disso, existe viés nos dados de distribuição de espécies em termos de conveniência/comportamento dos cientistas (e coletadores) e das tendências históricas de colonização e inventariação (BINI et al., 2006; MEYER et al., 2015; SASTRE; LOBO, 2009). Estes vieses espaciais podem ocorrer dentro de unidades políticas, com tendência de uma maior concentração de registros de ocorrências de espécimes em áreas mais acessíveis, localizados perto das vias de acesso, como estradas e rios navegáveis, em áreas conhecidas por serem ricas em espécies, como as áreas protegidas, áreas próximas às instituições de pesquisas e em cidades com melhor infraestrutura (SASTRE; LOBO, 2009). Por exemplo, dentro do continente africano há uma disparidade na distribuição espacial de registro de ocorrência de plantas com flores, sendo a África do Sul apresentando trinta vezes mais dados quando comparado com países vizinhos, como a Namíbia (STROPP et al., 2016). Em resumo, áreas remotas, inacessíveis ou politicamente sensíveis podem não ser adequadamente conhecidas quanto sua biodiversidade. Esta desigualdade no esforço amostral da biodiversidade pode fazer com que a interpretação de mapas de riqueza de espécies seja similar aos mapas de esforço amostral, um padrão que é visualmente marcante na África Subsaariana (LESSA et al., 2024; STROPP et al., 2016).

A variação espacial no esforço amostral está, por sua vez, relacionada com a variação na disponibilidade de recursos humanos e financeiros, e da capacidade científica da área de estudo (RUETE, 2015). A lacuna Wallaceana é mais proeminente no sul global, devido à sua dependência das tendências históricas em capacidade científica e recursos humanos (BECK et al., 2014; HORTAL et al., 2015; JETZ; MCPHERSON; GURALNICK, 2012). Coleções museológicas localizadas em cidades desenvolvidas tendem a receber mais material biológico de diversos lugares (PENN, CAFFERTY, & CARINE, 2018), ou ainda pesquisadores encontram melhores infraestruturas (universidades e laboratórios) e recursos financeiros em centros urbanos (CORREIA et al., 2019; KADMON; FARBER; DANIN, 2004; LESSA et al., 2019; LOBO, 2008; OLIVEIRA et al., 2016). A coleta de dados sobre a distribuição de espécies muitas vezes requer recursos significativos, incluindo financiamento, tempo de atividade de campo e pesquisadores qualificados. Em algumas áreas, especialmente em países em desenvolvimento, pode haver falta de capacidade técnica para realizar levantamentos biológicos abrangentes, falta de acesso a tecnologias e métodos eficazes de coleta de dados, o que contribui para a falta de informações detalhadas sobre a distribuição de espécies (LESSA et al., 2024).

O cenário passa a ser mais crítico devido à rápida transformação dos ambientes naturais em detrimento do uso humano (como desmatamentos para agricultura, mineração, produção de energia, transporte e construção civil), às alterações climáticas globais, desastres ambientais, e até mesmo perdas físicas (tal como incêndios em coleções museológicas), que têm afetado a qualidade dos dados de distribuições de espécies ao longo do tempo, tornando-os imprecisos e desatualizados (LADLE; HORTAL, 2013; LADLE; WHITTAKER, 2011; NABOUT et al., 2016; TESSAROLO et al., 2017; VEACH; MOILANEN; MININ, 2017). Em respostas às mudanças ambientais as espécies podem alterar suas dinâmicas espaciais em curta escala temporal, diminuindo a qualidade e a acurácia dos dados de distribuição mais antigos (HUGHES et al., 2012; STROPP et al., 2020). Isto pode resultar na perda de conhecimento das populações das espécies antes mesmo de serem documentadas. Como demonstrado no estudo publicado em 2017 que avaliou como as taxas de desmatamento na Amazônia brasileira afetariam os dados de distribuição de espécies de árvores. Os autores descobriram que até o ano de 2017, 12% da Amazônia brasileira foi desmatada, sem que houvesse um único exemplar de árvore registrada em coleções museológicas, e 37% das áreas apontadas como bem amostradas teriam sido desmatadas (STROPP et al., 2020). Portanto, áreas desmatadas em ambientes inexplorados tem o poder de dizimar a biodiversidade, bem como prejudicar a oportunidade de investigação e recolha de dados em campo.

A desigualdade dos dados de distribuição de espécies apresenta consequências inestimáveis para a conservação das espécies, uma vez que estes dados são usados para identificação e priorização de áreas para conservação (KUJALA; MOILANEN; GORDON, 2018) e para previsão de impactos de futuras mudanças ambientais (SINGER et al., 2016; THUILLER et al., 2008). A lacuna Wallaceana tem um impacto importante nas estimativas dos status de ameaça e risco de extinção, pois para elaboração de planos e ações de conservação são necessárias informações acuradas sobre o tamanho da área de distribuição das espécies-alvo (HORTAL et al., 2015). A União Internacional para Conservação da Natureza (IUCN) utiliza esta informação para definir e classificar o status de ameaça das espécies, sendo consideradas prioritárias as espécies com menor distribuição geográfica (RIDDLE et al., 2011; RODRIGUES et al., 2006). A falta de conhecimento sobre a distribuição das espécies também afetar os parâmetros utilizados para quantificar a biodiversidade, como estimativas de riqueza de espécies, delimitar o nicho das espécies, identificar o grau de endemismo, entre outras (COSTELLO; WILSON, 2011; MORA et al., 2011; RIDDLE et al., 2011; ROCCHINI et al., 2011; WHITTAKER et al., 2005).

Todos estes vieses espaciais e problemas relacionados corroboram para um aumento da incerteza e redução da qualidade dos dados de distribuições de espécies. Uma forma de mitigar o efeito da lacuna Wallaceana, é tornar os dados de distribuição de espécies intactos e acessíveis em todo o mundo, para isso é fundamental digitalizar e manter estes dados em plataformas digitais (HEDRICK et al., 2020). Em 2001, foi implementada uma rede internacional de compartilhamento de dados de espécimes de coleções de história natural e de ciência cidadã, o Global Biodiversity Information Facility (GBIF) (BECK et al., 2013; CHANDLER et al., 2017; GAIJI et al., 2013; NELSON; ELLIS, 2019). Atualmente o GBIF é a maior base de dados da biodiversidade em todo o mundo, abrigando mais de 2.6 bilhões de registros de espécies de vários grupos taxonômicos, desde indivíduos unicelulares a vertebrados e plantas (GBIF, 2023). Desta forma, investigadores de qualquer lugar do mundo podem acessar informações disponíveis na internet, ultrapassando barreiras geográficas para elaboração de pesquisas acerca da biodiversidade.

Entretanto, mesmo grandes bancos de dados da biodiversidade, como o GBIF, sofrem de incompletude, isto é, faltam dados completos ou confiáveis, seja da taxonomia, do ano de coleta ou da distribuição espacial (DE ARAUJO; QUARESMA; RAMOS, 2022; FREEMAN; PETERSON, 2019; ROCHA-ORTEGA; RODRIGUEZ; CÓRDOBA-AGUILAR, 2021). Isso acontece, muitas vezes, porque os dados de ocorrência de espécies possuem diversas fontes de informação. O primeiro passo para obter um conhecimento de qualidade da biodiversidade é reconhecer os limites do conhecimento atual, identificando o quanto não sabemos e as lacunas relacionadas. Uma vez que tenhamos descrito suficientemente o conhecimento da biodiversidade e seus vieses e limitações, a próxima tarefa é melhorar o inventário da pesquisa global sobre a biodiversidade de forma a maximizar a cobertura espacial e dispor de maneira mais eficaz os recursos limitados para pesquisa e conservação. A criação de “mapas da ignorância” é uma das estratégias utilizadas para fornecer uma medida completa da confiabilidade dos dados de ocorrência de espécies (CORREIA et al., 2019; RUETE, 2015; STROPP et al., 2016; TESSAROLO et al., 2021).

Existem várias abordagens que avaliam as lacunas e qualidade do conhecimento da biodiversidade. Uma das formas de fazer isso é através da criação de “mapas de ignorância” que distinguem áreas com amostragem intensiva daquelas com amostragem insuficiente (ROCCHINI et al., 2011). Uma medida para criar tais mapas é através do “Inventory Completeness”, que calcula a integridade dos inventários de espécies a partir de repositórios digitais em diferentes unidades espaciais (SOUSA-BAENA; GARCIA; PETERSON, 2014; STROPP et al., 2016), estimando curvas de acumulação de espécies em cada unidade amostral com base no número de registros. No entanto, esta abordagem não funciona bem para inventários de espécies não padronizados, e é muito sensível a números baixos de registos, particularmente em regiões com elevada biodiversidade (CORREIA et al., 2019; RUETE, 2015).

Outra abordagem ainda mais robusta é a criação de ‘Maps of biogeographical ignorance’ (MoBIs), que utiliza diversas fontes de informação, tal como o “Inventory Completeness”, a qualidade taxonômica, e o decaimento temporal e espacial dos dados da biodiversidade (HORTAL et al., 2022; TESSAROLO et al., 2021). Por ser mais elaborada, essa métrica também é sensível ao baixo número de registros de ocorrência e a baixa abundância de espécies em coleções de história natural (MEYER ET AL., 2016, STEEGE ET AL., 2011, STROPP ET AL., 2016). Uma abordagem alternativa mais simples que as anteriores é o "Ignorance Score". As pontuações de ignorância da biodiversidade têm a vantagem de ser simples de calcular, não utiliza estimativas de riqueza de espécies (como as curvas de acumulação), e baseia-se apenas em dados brutos de ocorrências de espécies (CORREIA et al., 2019; LESSA et al., 2024; MAIR; RUETE, 2016; RUETE, 2015), sendo ideal para ser aplicados em áreas com pouco conhecimento disponível. Como resultado são fornecidas informações sobre a cobertura e confiabilidade da amostragem, e relata explicitamente a distribuição espacial do viés e a falta de esforço de amostragem em uma região de estudo.

Independente da abordagem utilizada é necessário reconhecer que a qualidade dos dados é crucial para resultados fidedignos da distribuição espaciais, não só da biodiversidade, mas dos esforços de cientistas e coletadores em explorar a natureza. O preenchimento da lacuna Wallaceana contribui significativamente para o entendimento da ecologia e da biogeografia, mostrando como fatores históricos e ecológicos moldam a distribuição e a diversidade das espécies ao longo do tempo.

Referências

AGAPOW, P. et al. The Impact of Species Concept on Biodiversity Studies. The Quarterly Review of Biology, v. 79, n. 2, p. 161–179, jun. 2004.

BÁNKI, O. et al. Catalogue of life checklist (Version 2022-01-14).

BARROWCLOUGH, G. F. et al. How many kinds of birds are there and why does it matter? PLoS One, v. 11, n. 11, p. e0166307, 2016.

BATES, J. M.; DEMOS, T. C. Do we need to devalue Amazonia and other large tropical forests? Diversity and Distributions, v. 7, n. 5, p. 249–255, 20 set. 2001.

BEBBER, D. P. et al. Predicting unknown species numbers using discovery curves. Proceedings of the Royal Society B: Biological Sciences, v. 274, n. 1618, p. 1651–1658, 2007.

BECK, J. et al. Online solutions and the "Wallacean shortfall": What does GBIF contribute to our knowledge of species' ranges? Diversity and Distributions, v. 19, n. 8, p. 1043–1050, 2013.

BECK, J. et al. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecological Informatics, v. 19, p. 10–15, 2014.

BINI, L. M. et al. Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. Diversity and Distributions, v. 12, n. 5, p. 475–482, 2006.

BROWN, J. H.; LOMOLINO, M. V. Biogeography. Sunderland, MA: [s.n.].

CARDOSO, P. et al. The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation, v. 144, n. 11, p. 2647–2655, nov. 2011.

CHANDLER, M. et al. Contribution of citizen science towards international biodiversity monitoring. Biological Conservation, v. 213, p. 280–294, set. 2017.

CORREIA, R. A. et al. Using ignorance scores to explore biodiversity recording effort for multiple taxa in the Caatinga. Ecological Indicators, v. 106, n. June 2019, p. 105539, 2019.

COSTELLO, M. J.; MAY, R. M.; STORK, N. E. Can We Name Earth's Species Before They Go Extinct? Science, v. 339, n. 6118, p. 413–416, 25 jan. 2013.

COSTELLO, M. J.; WILSON, S. P. Predicting the number of known and unknown species in European seas using. Global Ecology and Biogeography, n. 20, p. 319–330, 2011.

CRACRAFT, J. Species concepts and speciation analysis. Current ornithology. Vol. 1, p. 159–187, 1983.

DAYRAT, B. Towards integrative taxonomy. Biological Journal of the Linnean Society, v. 85, n. 3, p. 407–415, 24 jun. 2005.

DE ARAUJO, M. L.; QUARESMA, A. C.; RAMOS, F. N. GBIF information is not enough: national database improves the inventory completeness of Amazonian epiphytes. Biodiversity and Conservation, v. 31, n. 11, p. 2797–2815, 2022.

DE QUEIROZ, K. Species Concepts and Species Delimitation. Systematic Biology, v. 56, n. 6, p. 879–886, 1 dez. 2007.

DILLON, S.; FJELDSÅ, J. The implications of different species concepts for describing biodiversity patterns and assessing conservation needs for African birds. Ecography, v. 28, n. 5, p. 682–692, 20 out. 2005.

ENGEL, M. S. et al. The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. Zoological Journal of the Linnean Society, v. 193, n. 2, p. 381–387, 24 set. 2021.

ERWIN, T. L. Tropical forests: Their richness in Coleoptera and other arthropod species. The Coleopterists Bulletin, v. 36, n. 1, p. 74–75, 1982.

FISHER, R. et al. Species richness on coral reefs and the pursuit of convergent global estimates. Current Biology, v. 25, n. 4, p. 500–505, 2015.

FREEMAN, B.; PETERSON, A. T. Completeness of digital accessible knowledge of the birds of western Africa: Priorities for survey. Condor, v. 121, n. 3, 2019.

GAIJI, S. et al. Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. Biodiversity Informatics, v. 8, n. 2, 9 jul. 2013.

GARCÍA-ROBLEDO, C. et al. The Erwin equation of biodiversity: From little steps to quantum leaps in the discovery of tropical insect diversity. Biotropica, v. 52, n. 4, p. 590–597, 2020.

GARNETT, S. T.; CHRISTIDIS, L. Taxonomy anarchy hampers conservation. Nature News, v. 546, n. 7656, p. 25, 2017.

GBIF. GBIF Home Page. Disponível em: https://www.gbif.org/grscicoll/institution/search>. Acesso em: 7 jan. 2023.

GOMES-DA-COSTA, S. M. Sistemas de Classificação em Reinos da Diversidade Biológica. In: Organização dos Seres Vivos. Maringá: Eduem, 2010. p. 55–78.

GOVAERTS, R. How Many Species of Seed Plants Are There? Taxon, v. 50, n. 4, p. 1085–1090, 2001.

HAMILTON, A. J. et al. Quantifying Uncertainty in Estimation of Tropical Arthropod Species Richness. The American Naturalist, v. 176, n. 1, p. 90–95, jul. 2010.

HARRIS, D. J.; FROUFE, E. Taxonomic inflation: species concept or historical geopolitical bias? Trends in Ecology & Evolution, v. 20, n. 1, p. 6–7, 2005.

HEDRICK, B. P. et al. Digitization and the Future of Natural History Collections. BioScience, v. 70, n. 3, p. 243–251, 1 mar. 2020.

HENDERSON, A. J. A revision of Attalea (Arecaceae, Arecoideae, Cocoseae, Attaleinae). Phytotaxa, v. 444, p. 1–76, maio 2020.

HOBERN, D. et al. Towards a global list of accepted species VI: The Catalogue of Life checklist. Organisms Diversity & Evolution, p. 1–14, 2021.

HORTAL, J. et al. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst., v. 46, p. 523–552, 2015.

HORTAL, J. et al. Accounting for biogeographical ignorance within biodiversity modelling. Research Outreach, n. 129, 4 abr. 2022.

HUGHES, A. C. et al. The projected effects of climatic and vegetation changes on the distribution and diversity of Southeast Asian bats. Global Change Biology, v. 18, n. 6, p. 1854–1865, 2012.

JETZ, W.; MCPHERSON, J. M.; GURALNICK, R. P. Integrating biodiversity distribution knowledge: Toward a global map of life. Trends in Ecology and Evolution, v. 27, n. 3, p. 151–159, 2012.

KADMON, R.; FARBER, O.; DANIN, A. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications, v. 14, n. 2, p. 401–413, 2004.

KITCHENER, A. C. et al. A system for designating taxonomic certainty in mammals and other taxa. Mammalian Biology, p. 1–11, 2022.

KLEPKA, V.; CORAZZA, M. J. O essencialismo na classificação de Lineu e a repercussão dessa controvérsia na Biologia. História da Ciência e Ensino: construindo interfaces, v. 18, p. 73–110, 2018.

KONSTANTINOV, F. V.; NAMYATOVA, A. A. Taxonomic Revisions and Specimen Databases in the Internet Age: Dealing with a Species Rich Insect Taxon. Entomological Review, v. 99, n. 3, p. 340–361, 30 jun. 2019.

KUJALA, H.; MOILANEN, A.; GORDON, A. Spatial characteristics of species distributions as drivers in conservation prioritization. Methods in Ecology and Evolution, v. 9, n. 4, p. 1121–1132, 11 abr. 2018.

KUNIN, W. E. et al. Upscaling biodiversity: estimating the species–area relationship from small samples. Ecological Monographs, v. 88, n. 2, p. 170–187, 2018.

LADLE, R.; HORTAL, J. Mapping species distributions: living with uncertainty. Frontiers of Biogeography, v. 5, n. 1, 4 abr. 2013.

LADLE, R. J.; WHITTAKER, R. J. Conservation Biogeography. Oxford University Press, Oxford., 2011.

LADLE, R. J.; WHITTAKER, R. J. Biogeografia e preservação ambiental. 1. ed. Oxford: Wiley-Blackwell, 2014.

LESSA, T. et al. Known unknowns: Filling the gaps in scientific knowledge production in the Caatinga. PLoS ONE, v. 14, n. 7, p. 1–12, 2019.

LESSA, T. et al. Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa. Ecological Indicators, v. 158, p. 111490, jan. 2024.

LESSA, T. et al. How taxonomic change influences forecasts of the Linnean shortfall (and what we can do about it)? Journal of Biogeography, mar 2024.

LOBO, J. M. Database records as a surrogate for sampling effort provide higher species richness estimations. Biodiversity and Conservation, v. 17, n. 4, p. 873–881, 2008.

LOMOLINO, M. V. Conservation biogeography. Frontiers of Biogeography: new directions in the geography of nature, v. 293, 2004.

MAIR, L.; RUETE, A. Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. PLOS ONE, v. 11, n. 1, p. e0147796, 28 jan. 2016.

MAY, R. M. Tropical Arthropod Species, More or Less? Science, v. 329, n. 5987, p. 41–42, 2 jul. 2010.

MAYDEN, R. L. A hierarchy of species concepts: the denouement in the saga of the species problem. In: Species: The units of diversity. [s.l.] Chapman & Hall, 1997. p. 381–423.

MAYR, E.; PROVINE, W. B. The Evolutionary Synthesis. [s.l.] Harvard University Press, 1980.

MEIJAARD, E.; NIJMAN, V. Primate Hotspots on Borneo: Predictive Value for General Biodiversity and the Effects of Taxonomy. Conservation Biology, v. 17, n. 3, p. 725–732, 29 jun. 2003.

MEYER, C. et al. Global priorities for an effective information basis of biodiversity distributions. Nature Communications, v. 6, n. 1, p. 8221, 8 set. 2015.

MORA, C. et al. How Many Species Are There on Earth and in the Ocean? PLoS Biology, v. 9, n. 8, p. e1001127, 23 ago. 2011.

MOURA, M. R.; JETZ, W. Shortfalls and opportunities in terrestrial vertebrate species discovery. Nature Ecology & Evolution, v. 5, n. 5, p. 631–639, 2021.

NABOUT, J. C. et al. The Impact of Global Climate Change on the Geographic Distribution and Sustainable Harvest of Hancornia speciosa Gomes (Apocynaceae) in Brazil. Environmental Management, v. 57, n. 4, p. 814–821, 2016.

NAVARRO-SIGÜENZA, A. G.; PETERSON, A. T. An alternative species taxonomy of the birds of Mexico. Biota Neotropica, v. 4, n. 2, p. 1–32, 2004.

NEKOLA, J. C.; HORSÁK, M. The impact of empirically unverified taxonomic concepts on ecological assemblage patterns across multiple spatial scales. Ecography, v. 2022, n. 5, p. 1–13, 2022.

NELSON, G.; ELLIS, S. The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B: Biological Sciences, v. 374, n. 1763, p. 20170391, 7 jan. 2019.

OLIVEIRA, U. et al. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Diversity and Distributions, p. 1–13, 2016.

PADIAL, J. M. et al. The integrative future of taxonomy. Frontiers in Zoology, v. 7, n. 1, p. 16, 2010.

PANTE, E. et al. Species are hypotheses: avoid connectivity assessments based on pillars of sand. Molecular Ecology, v. 24, n. 3, p. 525–544, 2015.

PARSONS, D. J. et al. Analysis of biodiversity data suggests that mammal species are hidden in predictable places. Pnas, v. 119, n. 14, p. 1–7, 2022.

PETERSON, A. T.; NAVARRO-SIGÜENZA, A. G. Alternative species concepts as bases for determining priority conservation areas. Conservation Biology, v. 13, p. 427–431, 1999.

RHEINDT, F. E. et al. Tightening the requirements for species diagnoses would help integrate DNA-based descriptions in taxonomic practice. PLOS Biology, v. 21, n. 8, p. e3002251, 22 ago. 2023.

RIDDLE, B. R. et al. Basic Biogeography: Estimating Biodiversity and Mapping Nature. In: Conservation Biogeography. [s.l: s.n.].

ROCCHINI, D. et al. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Progress in Physical Geography, v. 35, n. 2, p. 211–226, 2011.

ROCHA-ORTEGA, M.; RODRIGUEZ, P.; CÓRDOBA-AGUILAR, A. Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. Ecological Entomology, v. 46, n. 4, p. 718–728, 23 ago. 2021.

RODRIGUES, A. S. L. et al. The value of the IUCN Red List for conservation. Trends in Ecology and Evolution, v. 21, n. 2, 2006.

ROSEN, R. On the limitations of scientific knowledge. In: CASTI, J.; KARLQVIST, A. (Eds.). . oundaries and barriers: On the limits to scientific knowledge. [s.l.] MA: Perseus Books, 1996. p. 199–214.

RUETE, A. Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. Biodiversity Data Journal, v. 3, n. 1, 2015.

SANGSTER, G. Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. Proceedings of the Royal Society B: Biological Sciences, v. 276, n. 1670, p. 3185–3191, 2009.

SASTRE, P.; LOBO, J. M. Taxonomist survey biases and the unveiling of biodiversity patterns. Biological Conservation, v. 142, n. 2, p. 462–467, fev. 2009.

SCOTLAND, R. W.; WORTLEY, A. H. How Many Species of Seed Plants Are There? Taxon, v. 52, n. 1, p. 101–104, 2003.

SHEN, T. J.; CHAO, A.; LIN, C. F. Predicting the number of new species in further taxonomic sampling. Ecology, v. 84, n. 3, p. 798–804, 2003.

SHIN, C. P.; ALLMON, W. D. How we study cryptic species and their biological implications: A case study from marine shelled gastropods. Ecology and Evolution, v. 13, n. 9, 5 set. 2023.

SINGER, A. et al. Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? Ecological Modelling, v. 326, p. 63–74, abr. 2016.

SOUSA-BAENA, M. S.; GARCIA, L. C.; PETERSON, A. T. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions, v. 20, n. 4, p. 369–381, 2014.

STROPP, J. et al. Mapping ignorance: 300 years of collecting flowering plants in Africa. Global Ecology and Biogeography, v. 25, n. 9, p. 1085–1096, 2016.

STROPP, J. et al. The ghosts of forests past and future: deforestation and botanical sampling in the Brazilian Amazon. Ecography, v. 43, n. 7, p. 979–989, jul. 2020.

STROPP, J. et al. Taxonomic uncertainty and the challenge of estimating global species richness. Journal of Biogeography, v. 49, n. 9, p. 1654–1656, 28 set. 2022.

TESSAROLO, G. et al. Temporal degradation of data limits biodiversity research. Ecology and Evolution, v. 7, n. 17, p. 6863–6870, set. 2017.

TESSAROLO, G. et al. Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. Ecography, v. 44, n. 12, p. 1743–1755, 2021.

THUILLER, W. et al. Predicting global change impacts on plant species' distributions: Future challenges. Perspectives in Plant Ecology, Evolution and Systematics, v. 9, n. 3–4, p. 137–152, mar. 2008.

TROUDET, J. et al. Taxonomic bias in biodiversity data and societal preferences. Scientific Reports, v. 7, n. 9132, p. 1–14, 2017.

VAN VALEN, L. Ecological species, multspecies, and Oaks. Taxon, v. 25, n. 2–3, p. 233–239, 26 maio 1976.

VEACH, V.; MOILANEN, A.; MININ, E. DI. Threats from urban expansion, agricultural transformation and forest loss on global conservation priority areas. PLoS ONE, v. 12, n. 11, p. 1–14, 2017.

WHEELER, Q. A taxonomic renaissance in three acts. Megataxa, v. 1, n. 1, p. 4–8, 2020.

WHITTAKER, R. J. et al. Conservation Biogeography: Assessment and Prospect. Diversity and Distributions, 2005.

WILEY, E. O. The Evolutionary Species Concept Reconsidered. Systematic Zoology, v. 27, n. 1, p. 17, mar. 1978.

ZACHOS, F. E. Species Concepts in Biology. Switzerland: Springer International Publishing, 2016.

ZACHOS, F. E. (New) Species concepts, species delimitation and the inherent limitations of taxonomy. Journal of genetics, v. 97, n. 4, p. 811–815, 2018.

ZAPATA, F. A.; ROSS ROBERTSON, D. How many species of shore fishes are there in the Tropical Eastern Pacific? Journal of Biogeography, v. 34, n. 1, p. 38–51, 2007.

Objetivos

Objetivo Geral

O objetivo geral desta tese foi avaliar, quantificar e discutir as incertezas e ignorância nos dados disponíveis da biodiversidade sobre a taxonomia (lacuna Linneana) e distribuição geográfica (lacuna Wallaceana).

Objetivos Específicos

Esta tese é dividida em três capítulos em formato de manuscrito, cujos objetivos específicos foram:

Capítulo 1: How taxonomic change influences forecasts of the Linnean Shortfall (and what we can do about it)

1. Discutir como a dinâmica do processo taxonômico afeta as estimativas de espécies conhecidas e desconhecidas.

Capítulo 2: Do biological and ecological variables influence nomenclatural uncertainty in birds?

1. Criar uma métrica para avaliar a incerteza nomenclatural das espécies de aves globais.
2. Explorar as tendências da incerteza nomenclatural entre as espécies e Ordens de aves globais.
3. Analisar associações entre variáveis biológicas e ecológicas das espécies de aves globais e o escore de incerteza nomenclatural.

Capítulo 3: Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa

1. Aplicar a abordagem de escores de ignorância ("Ignorance Score") para avaliar as lacunas e vieses temporais, espaciais e taxonômicos nos registros de ocorrência de espécies disponíveis no GBIF para Namíbia, África.
2. Analisar a influência das variáveis sociogeográficas na distribuição do esforço

de registros de ocorrência, a partir dos escores de ignorância.

Capítulo 1
Revista: Journal of Biogeography
Status: Aceito para publicação em 29/02/2024

How taxonomic change influences forecasts of the Linnean Shortfall (and what we can do about it)

Running title: Taxonomic change and the Linnean Shortfall

Thainá Lessa[1], Juliana Stropp[2,3], Joaquín Hortal[2,4], Richard J. Ladle[1,5,6,*]

Institutional Affiliation:

[1]Institute of Biological and Health Sciences, Federal University of Alagoas, Maceió, AL, Brazil
[2]Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain
[3]Department of Biogeography, Trier University, Trier, Germany.
[4]Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás (UFG), Goiânia, Brazil
[5]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal
[6]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal

*Corresponding author: Richard J. Ladle; CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.

Email: richardjamesladle@gmail.com

Abstract

The gap between the number of described species and the number of species that actually exist is known as the Linnean shortfall and is of fundamental importance for biogeography and conservation. Unsurprisingly, there have been many attempts to quantify its extent for different taxa and regions. In this Perspective we argue that such forecasts remain highly problematic because the extent of the shortfall does not only depend on the rates of exploration (sampling undescribed taxa) on which estimates have been commonly based, but also on rates of taxonomic change (lumping and splitting). These changes highly depend on the species concepts adopted and the information and methods used to delimit species. Commonly used methods of estimating the number of unknown species (e.g., discovery curves, taxon ratios) can underestimate or overestimate the Linnean shortfall if they don't effectively account for trends and rates of taxonomic change. A further complication is that the history of taxonomic change is not well documented for most taxa and is not typically available in biodiversity databases. Moreover, wide geographic and taxonomic variation in the adoption of species concepts and delimitation methods mean that comparison of estimates of the Linnean shortfall between taxa and even for the same taxon between regions may be unreliable. Given the high likelihood of future taxonomic changes for most major taxa, we propose two main strategies to consider the influence of taxonomic change on estimates of unknown species; i) a highly conservative approach to estimating the Linnean shortfall, restricting analysis to groups and regions where taxonomies are relatively stable; and ii) explicitly incorporating metrics of taxonomic change into biodiversity models and estimates. In short, relevant estimates of the number of known and unknown species will only be achieved by accounting for the dynamic nature of the taxonomic process itself.

Keywords: species descriptions, taxonomic fluctuations, taxonomic lumping and splitting, global number of species, unknown species, shortfalls.

Introduction

Species are arguably the most important units for measuring biodiversity, and rates of species loss (extirpation and extinction) are key statistics for tracking human impacts on the environment (Ladle & Whittaker, 2011). Nevertheless, over 250 years after Carl von Linné (1707-1778) began his systematic inventory of the World's species, there is still considerable debate about how many species actually exist and how many of these we have already documented. The difference between the former and the latter is known as the Linnean shortfall (Brown & Lomolino, 1998), and is of fundamental importance for ecology, biogeography and the conservation of the Earth's remaining biological resources (Hortal et al., 2015; Whittaker et al., 2005). For example, unrecognized variation between regions in the proportion of unknown species could lead to misidentification of biodiversity patterns with knock-on effects for conservation prioritization and the inference of ecological and evolutionary processes (Diniz Filho et al., 2023; Edie et al., 2017; Stropp et al., 2022).

The challenge for biogeographers is that the numbers of known species are affected by the discovery of new species and the taxonomic reorganization of already known taxa. This leads to different types of unknown species. First, those species that are yet to be sampled. These are probably most common in the few remaining large, under-surveyed regions of the World such as the tropical moist broadleaf forests (Moura & Jetz, 2021), forests of southwest Amazonia (Hopkins, 2007, 2019) or the Brazilian Caatinga dry forest (Lessa et al., 2019), and poorly studied ecosystems such as the deep sea (Danovaro et al., 2010) or the upper canopies of rainforests (Ellwood & Foster, 2004). The number of these 'yet to be sampled species' is probably quite small for well-known groups of vertebrates, and much larger for less well-sampled taxa such as many invertebrates (Cardoso et al. 2011) and many largely unknown microbial groups.

Second, those species that have already been sampled but are yet to be formally described. These include historical specimens in museums and other natural history collections that have never been properly evaluated (Bebber et al., 2010). The number of such species may conceivably run into the hundreds of thousands, many of which may already be extinct (Alberch, 1993). In addition, there are unknown species that are currently "hidden" within another species, but which will be upgraded to accepted species status (i.e. taxonomic splitting) after a taxonomic re-evaluation (see Parsons et al., 2022). Finally, there are many currently valid (accepted) species that may be aggregated into a single species (i.e. taxonomic lumping) in the future due to taxonomic revision.

These different types of unknown species and taxonomic reorganizations have direct consequences for the estimation of the Linnean shortfall. Specifically, estimates of unknown species typically only account for new discoveries, thus failing to account for the fluctuations in species numbers that emerge purely from taxonomic reorganizations, such as new species created by splitting a taxon into two or more during taxonomic revision, or the synonymization of two or more taxa which are lumped into a single one. This is not a trivial problem, since changes in species designations are inherent to the taxonomic process (Hobern et al., 2021; Thiele et al., 2021) and may significantly outstrip new species discoveries in many taxa (Simkins et al., 2020). In this Perspective we discuss how we might account for the dynamics associated with the taxonomic process to improve our estimates of known and unknown species.

Extrapolating from uneven foundations

Many methods have been proposed to measure the Linnean shortfall. Most of these methods, even those based on expert opinion, are ultimately linked to counts of currently valid species, be that extrapolations of species discovery trends or from well-known taxa or intensively studied areas. In an ideal world, each newly described species would be meticulously documented, unambiguously identified, named, and definitively and permanently allocated a unique branch on the tree of life using identical methods based on a single, universally applied concept of what constitutes a species (Stropp et al, 2022). In the real world, taxonomy is built on an uneven foundation of different species concepts and delimitation methods (Zachos, 2016, 2018a). Furthermore, the rules established by the International Codes of Nomenclature allow freedom in the choice of criteria for species comparison and diagnoses, which increase the chances of new taxonomic changes and, concomitantly, decrease taxonomic stability (Rheindt et al. 2023).

The application of different species concepts and delimitation methods inevitably leads to multiple propositions for the number of species within a taxon or geographic region – even for well-known taxa such as birds, multiple global species lists are still in operation (Neate-Clegg et al., 2021). The difficulties associated with producing a single, universally accepted list of species have recently been discussed in a series of articles (Conix et al., 2021; Hobern et al., 2021; Lien et al., 2021; Pyle et al., 2021; Thiele et al., 2021; Thomson et al., 2021). Specifically, they highlight that for any such list to be widely adopted, mechanisms would have to be developed to ensure: i) the accuracy and consistency of additions to the list; ii) that the list is regularly updated and maintained; iii) that obscure and newly described taxa are consistently dealt with, and; iv) there are robust mechanisms for arbitrating (the inevitable) disputes or alternative taxonomic viewpoints. Details concerning how a global checklist can be accessed, how it will be maintained, and the way in which the list and its contents are properly cited still need to be determined. Some of these issues have been considered, and sometimes solved, by the Catalogue of Life (Bánki et al., 2023), which included 2,121,194 species as of 28th February 2024 (https://www.catalogueoflife.org).

Taxonomy is a dynamic discipline with great variation in practices over time, in different geographic regions and often between taxonomists working on different (or even the same) groups of organisms (Nekola & Horsák, 2022; Pyle et al., 2021). For example, arthropods are known to have the majority undescribed species, either due to incipient exploration or because there are species hidden in others (Costello et al., 2013). However, even for well-known groups, such as mammals and birds, these gaps are not completely filled (Mora et al., 2011; Parsons et al., 2022). In short, the Linnean shortfall is generally more severe for organisms that are smaller in size, niche width, or distributional range and which are less complex or phenotypically conspicuous, with this pattern holding both between and within taxonomic groups (Zapata & Robertson, 2006; Riddle et al., 2011; dos Santos et al., 2020; Guedes et al., 2023; Poulin et al., 2023). Recent research suggests that scientists have still only robustly sampled 6.74% of the Earth, with the tropics, high elevations and deep seas especially poorly covered (Hughes et al., 2021). It is possible that in the future we will achieve a high degree of taxonomic synthesis for some taxa (Nekola & Horsák, 2022), but we are not there yet. Alternatively, we may adopt a different way of accounting for nature such as the recently proposed multilevel organismal diversity concept (Martynov & Korshunova, 2022), the widespread adoption of which has the potential to radically change biological nomenclature and perceptions of biodiversity patterns.

Lumping versus splitting: a zero-sum game?

The number of known species can either increase or decrease, even in the absence of new discoveries (Figure 1). Increases occur when existing subspecies/races/populations are raised to species level, 'splitting' a formerly-recognized species into two or more due to changes in the application of species concepts or species delimitation methods, especially in recently described species delimited based on genetic information (Isaac et al., 2004). Such increases are frequently a sign of taxonomic progress (Korshunova et al., 2023) and have been identified in many taxa, but are particularly prevalent in well-known vertebrate groups such as birds (Simkins et al., 2020) and mammals (Gippoliti & Groves, 2012; Zachos, 2018b). For example, a recent analysis of temporal trends in known mammal species numbers found an increase of 1,079 species over 13 years, nearly all due to taxonomic revisions (Burgin et al., 2018), a trend that is forecasted to continue (Parsons et al., 2022). Moreover, taxonomic changes are ongoing and it is highly probable that, for example, many intraspecific taxa (e.g., subspecies) of birds will be formally recognized as valid species over the forthcoming decades (Barrowclough et al., 2016). There have been intense debates about whether such increases (often termed 'taxonomic inflation') are justified (Garnett & Christidis, 2017; Harris & Froufe, 2005; Padial & De la Riva, 2006; Sangster, 2009). From the perspective of the Linnean shortfall, high frequencies of future taxonomic splits will lead to current underestimates of the total number of species in a taxon using standard forecasting methods such as extrapolation of discovery curves (Figure 1).

The number and proportion of species that will be created by taxonomic splits, as opposed to undiscovered species, is difficult to evaluate for all but the best known taxa (e.g., Parsons et al., 2022). For birds, estimates of the number of species in the world remained remarkably consistent over the second half of the 20th century. In 1946, Ernst Mayr, the first proponent of the biological species concept, estimated there to be approximately 8,600 species (Mayr, 1946). This figure had only slightly changed by the end of the century; Sibley & Monroe (1990) suggested there may be as many as 9,700 bird species. With the introduction of more sophisticated molecular techniques and widespread changes in taxonomic practice, these estimates have continued to increase along with the number of officially recognized species (notwithstanding that there are multiple global checklists). For example, the current list of the International Ornithological Committee (IOC) recognizes 10,928 extant species (Gill et al., 2022), whereas the list given by the Handbook of the Birds of the World (HBW) recognizes 10,824 extant species (del Hoyo et al., 2013). It is an ongoing question of how many of these species might be revised (either split or lumped, or both) in the future. When a diagnostic, evolutionary species concept was applied to a morphological and distributional data set from 200 species, it was estimated that there could be as many as 18,043 species of birds worldwide (95% confidence interval of 15,845 to 20,470; Barrowclough et al., 2016). This figure is close to the current number (19,883) of recognized sub-species on the IOC list. How many of these sub-species will eventually be upgraded to full species status? This is almost impossible to answer, since it depends on continued changes in taxonomic theory and practice. Whatever the final figure is, recent history (Simkins et al. 2020) suggests it will be significantly greater than the proportion of newly discovered bird species in the field.

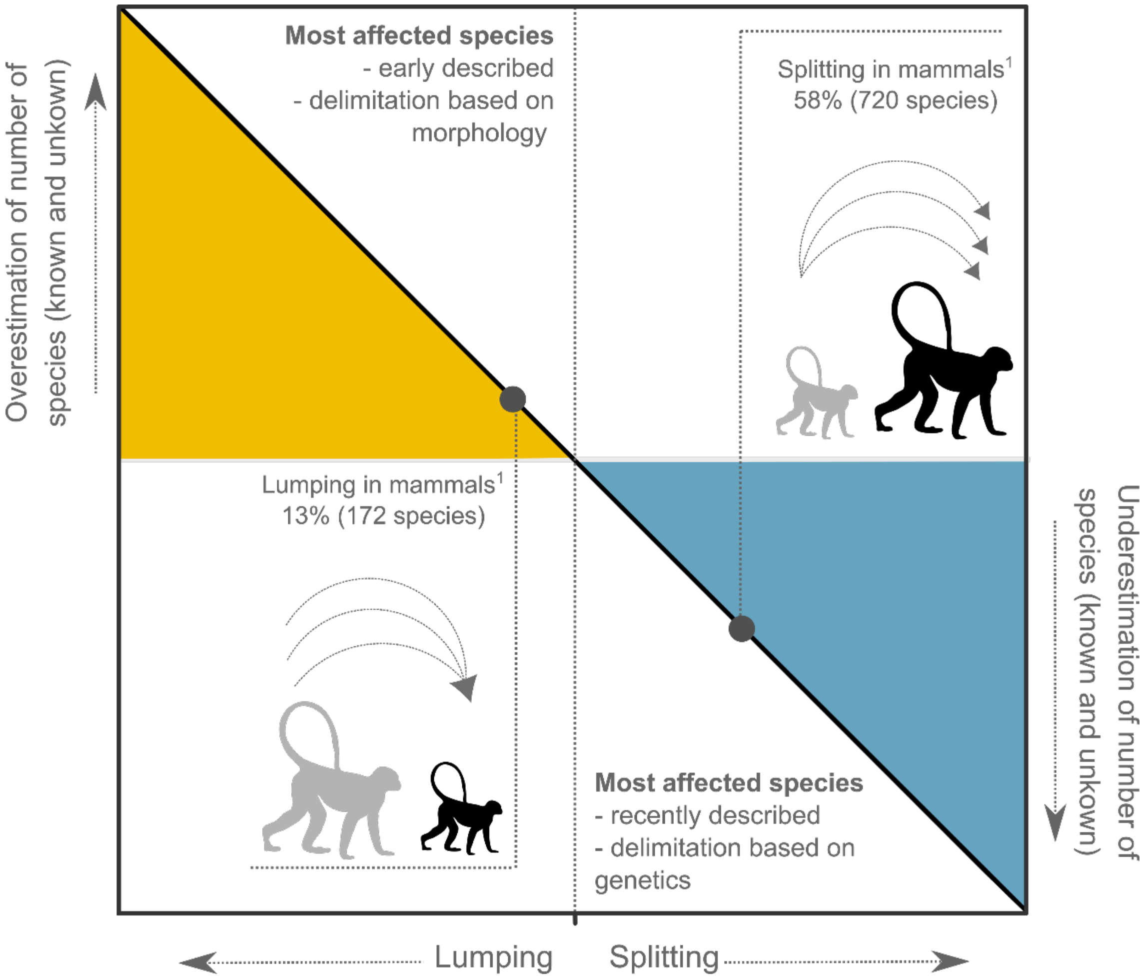


Figure 1: Schematic representation of the impact of lumping and splitting on estimates of the number of known and unknown species (Linnean Shortfall). The centre of the graph represents taxonomic stability (no further splitting or lumping of species). A scenario of more splitting than lumping leads to underestimates of the shortfall (right lower quadrant, blue), while more lumping than splitting leads to overestimates of the shortfall (left upper quadrant, yellow). For example, there was relatively more splitting than lumping of mammals between 2005 and 2017 (figures from Burgin et al. 2018), meaning pre-2005 estimates of the total number of mammal species (unknown + known) were almost certainly underestimates. Size of animal silhouettes is proportional to the number of valid species, with grey indicating pre-2005 and black post-2005.

The taxonomic history of birds and mammals is not necessarily reflective of other taxa (Knapp et al., 2005), and taxonomic revisions can also decrease the number of valid species. A proportion of currently recognized species will, at some point in the future, be relegated to synonyms due to 'lumping'. These lumping events are most common in early described species where its delimitation was based on morphological traits. Bouchet (2006) suggested that as many as a fifth of all recently described species names may become synonyms. In the case of amphibians, Hillis (2019) predicts that the future is likely to see high levels of lumping as more careful analysis corrects widespread 'over-splitting' that occurred when taxonomists began to introduce molecular methods and adopt the phylogenetic species concept.

In some taxa, such as Neotropical palms, the number of synonyms can outstrip the number of valid/accepted species (Stropp et al., 2022). This way, taxonomic change leads to a reduction in the number of valid/accepted species. For example, the 27 species that currently comprise the palm genus Attalea (Henderson et al., 1995) have previously been described as 16 different genera and associated with 167 species at different points over the past two hundred years (Henderson, 2020). The most recent revision of the Attalea proposes 30 species, which are associated with 93 heterotypic synonyms (data from Henderson 2020). It is important to remember that heterotypic synonyms produced by taxonomic lumping were once considered as valid species and consequently, if the Linnean shortfall had been calculated before heterotypic synonymizations were proposed, the scientists of the time would have vastly overestimated the number of unknown species using forecasting methods such as extrapolating from discovery curves (e.g., Bebber et al., 2007; Stropp et al., 2022).

There is no reason to assume that, at the level of higher taxa, new species created by splitting will be broadly 'compensated' by losses due to lumping. Rather, historical, current and future levels of splitting and lumping will vary according to biocultural factors, including the number of taxonomists working on a particular taxon in a particular region (Freeman & Pennell, 2021), levels of cryptic biodiversity (Beheregaray & Caccone, 2007), geographical biases in taxonomic practices and the completeness of biological collections (Harris & Froufe, 2005; Lavoie, 2013), among others. In other words, splitting and lumping cannot be assumed to be a 'zero-sum game' and taxonomic trends may vary enormously even between closely related taxa (Williams, 2022). As we include more diverse taxa in our analysis, systematic biases between taxonomic groups caused by splitting and lumping practices may become even more pronounced because of different states of taxonomic knowledge and continued progress in taxonomy (Lughadha et al., 2016; Troudet et al., 2017).

Is it possible to account for the taxonomic changes?

There are eleven general approaches for estimating the number of undiscovered species, each with its own assumptions. It is certainly possible to make broad predictions about how many species remain unknown (either undiscovered or unrecognized) based on these approaches. However, in our opinion there is a largely unrecognized caveat that future taxonomic revisions have the potential to significantly increase or decrease forecasts. Even if this caveat is widely known, the vast majority of published forecasts for the Linnean shortfall fail to account for, or simply ignore, the impacts of taxonomic change on their estimates (e.g., Costello et al., 2015; Costello & Wilson, 2011; Gatti et al., 2022; Legros et al., 2020; Moura & Jetz, 2021).

Dealing with the effects of taxonomic change on forecasts of unknown species is by no means simple, but this does not justify continuing to ignore this ‘elephant in the room’ of biogeographical and macroecological research. In our opinion there are two main approaches that can be adopted to deal with taxonomic change, though both have significant limitations. Firstly, we could be cautious and only make extrapolations for groups with a fairly stable taxonomy or within geographic areas with uniform taxonomic practice (Freeman & Pennell, 2021). Smaller taxonomic and geographic scales would certainly improve forecasts and could potentially be aggregated. Nevertheless, as taxonomy is currently undergoing a period of intense change in practices (Wheeler, 2020).

A second approach is to better account for taxonomic change in models and forecasts. This could potentially be achieved by identifying and quantifying how lumping, splitting, or new discoveries have influenced past biodiversity estimates. Unfortunately, gathering data on historical taxonomic change is difficult and very labour intensive because most, taxonomic databases only provide accounts of currently accepted and not accepted names, and do not give the full history of changes associated with these names (Franz & Peet, 2010). It would be useful to know for how long and when each name was associated with a valid species. One potential barrier to obtaining historical taxonomic information is the current difficulty of accessing and collecting relevant data for a wide range of taxa. Data on when names cease to be valid are often scattered in the taxonomic literature and not easily accessible via the generally-available databases and checklists. Indeed, most taxonomic databases typically codify only when a name was proposed and/or reinstated. Some notable examples could be used as references for other taxa, such as Amphibian Species of the World, The reptile database and International Ornithological Community, organizations that collate and share taxonomic information for amphibians, reptiles and birds, respectively. However, this is further complicated by the fact that species are hypotheses (Pante et al., 2015), and proposals for changes in taxonomic status are not necessarily accepted or adopted by the taxonomic community. Furthermore, acceptance of a proposed change does not occur immediately. A partial solution to this could be to use culturomics analysis (Ladle et al., 2016). Here, analyses of changes in name frequency in the academic literature could help identify when species 'hypotheses' were adopted or rejected, though a formal method to achieve this has not yet been proposed (but see Newberry & Plotkin, 2022).

Taxonomic databases such as the Catalogue of Life (Bánki et al., 2023) provide information on synonyms, but not when and for how long these were considered as valid species. The reality is that reconstructing the history of taxonomic change often requires painstaking research on the taxonomic literature proposing heterotypic synonyms (Creighton et al., 2022; Vaidya et al., 2018). This makes analysis of historical taxonomic trends for a large number of taxa or large geographic regions highly challenging, especially for non-taxonomists (such as many biogeographers and macroecologists). Advances in text mining technologies based on machine learning (Farrell et al., 2022) and the widespread digitalization of taxonomic monographs will certainly help reduce this data gap. Once these data are available we could model the probability of taxonomic change associated with each accepted species name. In turn, the results of these models can be used to identify taxon - or geographically- specific scenarios of taxonomic change and to incorporate these into forecasting models (Alroy, 2002; Edie et al., 2017).

Conclusions: Revisiting the Linnean Shortfall

The discrepancy between the number of species recognized to exist now and how many actually exist (if we had access to all the relevant data) is highly dependent on how we define and evaluate species. Moreover, taxonomic tools and practices are rapidly evolving and are likely to do so for many years to come (Gill, 2014; Padial & De la Riva, 2021; Wheeler, 2020). Much of these changes are due to advances in molecular and computational methods, combined with a large number of competing species concepts jostling for dominance (Kitchener et al., 2022; Stankowski & Ravinet, 2021; Zachos, 2018b). The culture of taxonomy may also be changing, with a greater willingness among taxonomists to designate new species from allopatric or parapatric populations that were previously part of polytypic species (Meiri & Mace, 2007). The discrepancies in taxonomic practices (Harris & Froufe, 2005; Stankowski & Ravinet, 2021), coupled with uneven biological knowledge for different taxa and regions (e.g., Meyer et al., 2016), mean that estimates of the Linnean shortfall at scale are, and will remain, immensely challenging.

Biogeographers and macroecologists need to be aware that the Linnean shortfall is a biocultural phenomenon whose characteristics depend on the complex interplay of various factors, including: i) history of where we have explored; ii) how much effort we have expended searching for different taxa; iii) how easy or difficult a species is to observe/collect; iv) how we define and delimit species; v) which is the level of digitalization and data mobilisation of the taxonomic work already devoted to each taxa; vi) how long and when each name is associated with an accepted species; and vii) how we use the data on how many species exist within a defined area to predict how many species remain undiscovered in that area. Both the numbers of known and unknown species are therefore subject to considerable fluctuations due to changes in taxonomic practice and, with them, to a significant degree of taxonomic uncertainty, thus calling for considerable caution when conducting and interpreting global estimates. This caveat may be addressed by designing strategies to account for taxonomic change when estimating current diversity or for forecasting the numbers of unknown species. These new estimates should account for the species that are yet to be discovered in the field, those that one day will be split from currently recognized taxa, and those that will be aggregated due to taxonomic lumping. Only when the resulting fluctuations in species number stemming from these three processes are integrated into the biogeographers' analytical toolbox will our models of global biodiversity patterns and processes become truly robust. This necessity is increasingly being acknowledged by the ecology and biogeography community, as evidenced by recent proposals to tighten requirements for species comparisons and diagnoses (Rheindt et al. 2023) and to create a globally integrated structure for taxonomy (Sandall et al. 2023). Such advances need necessarily to go a long way towards creating a more stable taxonomy that provides a solid basis for extrapolations of the Linnean Shortfall.

Acknowledgements: This work was funded by the European Union's Horizon 2020 research and innovation programme (grant agreement #854248). JS was also supported by EU H2020 Marie Skłodowska-Curie Action TAXON-TIME (grant agreement #843234), and JH by project SCENIC (grant PID2019-106840GB-C21, funded by Spanish MCIN/AEI/ 10.13039/501100011033. RJL is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 854248. TL was funded by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" - Brazil (CAPES) - Finance Code 001. Additional support has been provided by the "Instituto Nacional de Ciência e Tecnologia em Ecologia, Evolução e Conservação da Biodiversidade" (INCT-EECBio) (MCTIC/CNPq 465610/2014-5; FAPEG 201810267000023). No fieldwork was conducted so no collection permits were required for this work.

References

Alberch, P. (1993). Museums, collections and biodiversity inventories. Trends in Ecology & Evolution, 8(10), 372–375.

Alroy, J. (2002). How many named species are valid? Pnas, 99(6), 3706–3711. https://doi.org/https://doi.org/10.1073/pnas.062691099

Bánki, O., Roskov, Y., Döring, M., Ower, G., Hernández Robles, D. R., Plata Corredor, C. A., Stjernegaard Jeppesen, T., Örn, A., Vandepitte, L., Hobern, D., Schalk, P., DeWalt, R. E., Ma, K., Miller, J., Orrell, T., Aalbu, R., Abbott, J., Adlard, R., Adriaenssens, E. M., et al. (2023). Catalogue of Life Checklist (Version 2023-08-17). Catalogue of Life. https://doi.org/10.48580/dft7

Barrowclough, G. F., Cracraft, J., Klicka, J., & Zink, R. M. (2016). How many kinds of birds are there and why does it matter? PLoS One, 11(11), e0166307.

Bebber, D. P., Carine, M. A., Wood, J. R. I., Wortley, A. H., Harris, D. J., Prance, G. T., Davidse, G., Paige, J., Pennington, T. D., Robson, N. K. B., & Scotland, R. W. (2010). Herbaria are a major frontier for species discovery. Pnas, 107(51), 22169–22171. https://doi.org/10.1073/pnas.1011841108

Bebber, D. P., Marriott, F. H. C., Gaston, K. J., Harris, S. A., & Scotland, R. W. (2007). Predicting unknown species numbers using discovery curves. Proceedings of the Royal Society B: Biological Sciences, 274(1618), 1651–1658.

Beheregaray, L. B., & Caccone, A. (2007). Cryptic biodiversity in a changing world. Journal of Biology, 6(4), 1–5.

Bouchet, P. (2006). The magnitude of marine biodiversity. The Exploration of Marine Biodiversity: Scientific and Technological Challenges, 31–62.

Brown, J. H., & Lomolino, M. V. (1998). Biogeography.

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? Journal of Mammalogy, 99(1), 1–14.

Conix, S., Garnett, S. T., Thiele, K. R., Christidis, L., van Dijk, P. P., Bánki, O. S., Barik, S. K., Buckeridge, J. S., Costello, M. J., & Hobern, D. (2021). Towards a global list of accepted species III. Independence and stakeholder inclusion. Organisms Diversity & Evolution, 21(4), 631–643.

Costello, M.J., May, R. M., Stork, N. E. (2013). Can we name Earth's species before they go extinct? Science, 339, 413–416. https://doi.org/10.1126/science.1230318

Costello, M. J., Lane, M., Wilson, S., & Houlding, B. (2015). Factors influencing when species are first named and estimating global species richness. Global Ecology and Conservation, 4, 243–254. https://doi.org/10.1016/j.gecco.2015.07.001

Costello, M. J., & Wilson, S. P. (2011). Predicting the number of known and unknown species in European seas using. Global Ecology and Biogeography, 20, 319–330. https://doi.org/10.1111/j.1466-8238.2010.00603.x

Creighton, M. J. A., Luo, A. Q., Reader, S. M., & Mooers, A. Ø. (2022). Predictors of taxonomic splitting and its role in primate conservation. Animal Conservation. https://doi.org/https://doi.org/10.1111/acv.12825

Danovaro, R., Company, J. B., Corinaldesi, C., D'Onghia, G., Galil, B., Gambi, C., Gooday, A. J., Lampadariou, N., Luna, G. M., & Morigi, C. (2010). Deep-sea biodiversity in the Mediterranean Sea: the known, the unknown, and the unknowable. PloS One, 5(8), e11832.

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A., del Hoyo, J., Elliott, A., & Sargatal, J. (2013). Handbook of the birds of the world. Lynx edicions.

Diniz Filho, J. A. F., Jardim, L., Guedes, J. J. M., Meyer, L., Stropp, J., Frateles, L. E. F., Pinto, R. B., Lohmann, L. G., Tessarolo, G., & de Carvalho, C. J. B. (2023). Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. Frontiers of Biogeography.

dos Santos, J.W., Correia, R.A., Malhado, A.C.M., Campos-Silva, J.V., Teles, D., Jepson, P., Ladle, R.J. (2020). Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. Anim. Conserv., 23, 679-688. https://doi.org/10.1111/acv.12586

Edie, S. M., Smits, P. D., & Jablonski, D. (2017). Probabilistic models of species discovery and biodiversity comparisons. Pnas, 114(14), 3666–3671. https://doi.org/10.1073/pnas.1616355114

Ellwood, M. D. F., & Foster, W. A. (2004). Doubling the estimate of invertebrate biomass in a rainforest canopy. Nature, 429(6991), 549–551.

Farrell, M. J., Brierley, L., Willoughby, A., Yates, A., & Mideo, and N. (2022). Past and future uses of text mining in ecology and evolution. Proceedings of the Royal Society B, 289(1975), 1–9. https://doi.org/https://doi.org/10.1098/rspb.2021.2721

Franz, N. M., & Peet, R. K. (2010). Perspectives: Towards a language for mapping relationships among taxonomic concepts. Systematics and Biodiversity, 7(1), 5–20. https://doi.org/https://doi.org/10.1017/S147720000800282X

Freeman, B. G., & Pennell, M. W. (2021). The latitudinal taxonomy gradient. Trends in Ecology & Evolution, 1, 1–9. https://doi.org/10.1016/j.tree.2021.05.003

Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. Nature News, 546(7656), 25.

Gatti, R. C., Reich, P. B., Gamarra, J. G. P., Crowther, T., Hui, C., Morera, A., Bastin, J.-F., de-Miguel, S., Nabuurs, G.-J., & Svenning, J.-C. (2022). The number of tree species on Earth. Proceedings of the National Academy of Sciences, 119(6).

Gill, Donsker, & Rasmussen. (2022). IOC World Bird List. https://www.worldbirdnames.org/new/

Gill, F. B. (2014). Species taxonomy of birds: which null hypothesis? The Auk: Ornithological Advances, 131(2), 150–161.

Gippoliti, S., & Groves, C. P. (2012). "Taxonomic inflation" in the historical context of mammalogy and conservation. Hystrix, 23(2), 8.

Guedes, J. J. M., Moura, M. R., Diniz-Filho, J. A. F. (2023). Species out of sight: elucidating the determinants of research effort in global reptiles. Ecography. e06491. https://doi.org/10.1111/ecog.06491

Harris, D. J., & Froufe, E. (2005). Taxonomic inflation: species concept or historical geopolitical bias? Trends in Ecology & Evolution, 20(1), 6–7.

Henderson, A. J. (2020). A revision of Attalea (Arecaceae, Arecoideae, Cocoseae, Attaleinae). Phytotaxa, 444, 1–76. https://doi.org/10.11646/phytotaxa.444.1.1

Henderson, A. J., Galeano, G., & Bernal, R. (1995). A field guide to the palms of the Americas. Princeton University Press. https://doi.org/10.2307/j.ctvcszzzd

Hillis, D. M. (2019). Species delimitation in herpetology. Journal of Herpetology, 53(1), 3–12.

Hobern, D., Barik, S. K., Christidis, L., T Garnett, S., Kirk, P., Orrell, T. M., Pape, T., Pyle, R. L., Thiele, K. R., & Zachos, F. E. (2021). Towards a global list of accepted species VI: The Catalogue of Life checklist. Organisms Diversity & Evolution, 1–14.

Hopkins, M. J. G. (2007). Modelling the known and unknown plant biodiversity of the Amazon Basin. Journal of Biogeography, 34, 1400–1411. https://doi.org/10.1111/j.1365-2699.2007.01737.x

Hopkins, M. J. G. (2019). Are we close to knowing the plant diversity of the Amazon? Anais Da Academia Brasileira de Ciências, 91.

Hortal, J., Bello, F. de, Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst., 46, 523–552. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. Ecography.

Isaac, N. J. B., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution, 19(9), 464–469.

Kitchener, A. C., Hoffmann, M., Yamaguchi, N., Breitenmoser-Würsten, C., & Wilting, A. (2022). A system for designating taxonomic certainty in mammals and other taxa. Mammalian Biology, 1–11.

Knapp, S., Lughadha, E. N. N., & Paton, A. (2005). Taxonomic inflation, species concepts and global species lists. Trends in Ecology & Evolution, 1(20), 7–8. https://doi.org/10.1016/j.tree.2004.10.012

Korshunova, T., Lundin, K., Malmberg, K., & Martynov, A. (2023). Narrowly defined taxa on a global scale: The phylogeny and taxonomy of the genera Catriona and Tenellia (Nudibranchia, Trinchesiidae) favours fine-scale taxonomic differentiation and dissolution of the “lumpers & splitters” dilemma. Evolutionary Applications, 16(2), 428-460.

Ladle, R. J., Correia, R. A., Do, Y., Joo, G., Roberge, J., & Jepson, P. (2016). Conservation culturomics. Frontiers in Ecology and the Environment, 14, 269–275. https://doi.org/10.1002/fee.1260

Ladle, R. J., & Whittaker, R. J. (2011). Conservation Biogeography. In Conservation Biogeography. https://doi.org/10.1002/9781444390001

Legros, V., Rochat, J., Reynaud, B., & Strasberg, D. (2020). Known and unknown terrestrial arthropod fauna of La Réunion Island, Indian Ocean. Journal of Insect Conservation, 24, 199–217.

Lessa, T., Dos Santos, J. W., Correia, R. A., Ladle, R. J., & Malhado, A. C. M. (2019). Known unknowns: Filling the gaps in scientific knowledge production in the Caatinga. PLoS ONE, 14(7), 1–12. https://doi.org/10.1371/journal.pone.0219359

Lien, A. M., Conix, S., Zachos, F. E., Christidis, L., van Dijk, P. P., Bánki, O. S., Barik, S. K., Buckeridge, J. S., Costello, M. J., & Hobern, D. (2021). Towards a global list of accepted species IV: Overcoming fragmentation in the governance of taxonomic lists. Organisms Diversity & Evolution, 21(4), 645–655.

Lughadha, E. N., Govaerts, R., Belyaeva, I., Black, N., Lindon, H., Allkin, R., Magill, R. E., & Nicolson, N. (2016). Counting counts: revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. Phytotaxa, 272(1), 82. https://doi.org/10.11646/phytotaxa.272.1.5

Martynov, A.V. & Korshunova, T.A. (2022). Multilevel Organismal Diversity in an ontogenetic framework as a solution for the species concept. In: Cryptic Species Morphological Stasis, Circumscription, and Hidden Diversity. Cambridge University Press. p. 78–129.

Mayr, E. (1946). The number of species of birds. The Auk, 63(1), 64–69.

Meiri, S., & Mace, G. M. (2007). New taxonomy and the origin of species. Public Library of Science San Francisco, USA.

Meyer, C., Jetz, W., Guralnick, R. P., Fritz, S. A., & Kreft, H. (2016). Range geometry and socio-economics dominate species-level biases in occurrence information. Global Ecology and Biogeography, 25(10), 1181–1193.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? PLoS Biology, 9(8), e1001127. https://doi.org/10.1371/journal.pbio.1001127

Moura, M. R. R., & Jetz, W. (2020). Shortfalls and opportunities in terrestrial vertebrate species discovery. BioRxiv, 5(5), 631–639. https://doi.org/10.1038/s41559-021-01411-5

Neate-Clegg, M. H. C., Blount, J. D., & Şekercioğlu, Ç. H. (2021). Ecological and biogeographical predictors of taxonomic discord across the world's birds. Global Ecology and Biogeography, 30(6), 1258–1270. https://doi.org/10.1111/geb.13300

Nekola, J. C., & Horsák, M. (2022). The impact of empirically unverified taxonomic concepts on ecological assemblage patterns across multiple spatial scales. Ecography, 2022(5), 1–13.

Newberry, M. G., & Plotkin, J. B. (2022). Measuring frequency-dependent selection in culture. Nature Human Behaviour, 6, 1048–1055. https://doi.org/s41562-022-01342-6

Padial, J. M., & De la Riva, I. (2006). Taxonomic inflation and the stability of species lists: the perils of ostrich's behavior. Systematic Biology, 55(5), 859–867.

Padial, J. M., & De la Riva, I. (2021). A paradigm shift in our view of species drives current trends in biological classification. Biological Reviews, 96(2), 731–751.

Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A., Destombe, C., Forcioli, D., Valero, M., Viard, F., Samadi, S., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A., Destombe, C., Forcioli, D., & Valero, M. (2015). Species are hypotheses: avoid connectivity assessments based on pillars of sand. Molecular Ecology, 24(3), 525–544. https://doi.org/https://doi.org/10.1111/mec.13048

Parsons, D. J. J., Pelletier, T. A., Wieringa, J. G. G., Duckett, D. J., Carstens, B. C. C., Pelletierc, T. A., Wieringa, J. G. G., Ducketta, D. J., & Carstens, B. C. C. (2022). Analysis of biodiversity data suggests that mammal species are hidden in predictable places. Proceedings of the National Academy of Sciences, 119(14), e2103400119. https://doi.org/10.1073/pnas.2103400119

Poulin R., Presswell B., Bennett J., Dutra, D. A., Salloum, P. M. (2023). Biases in parasite biodiversity research: why some helminth species attract more research than others, International Journal for Parasitology: Parasites and Wildlife, 21, 89-98, https://doi.org/10.1016/j.ijppaw.2023.04.010.

Pyle, R. L., Barik, S. K., Christidis, L., Conix, S., Costello, M. J., van Dijk, P. P., Garnett, S. T., Hobern, D., Kirk, P. M., & Lien, A. M. (2021). Towards a global list of accepted species V. The devil is in the detail. Organisms Diversity & Evolution, 1–19.

Rheindt F.E., Bouchard P., Pyle R.L., Welter-Schultes F.., Aescht E., Ahyong S.T., et al. (2023). Tightening the requirements for species diagnoses would help integrate DNA-based descriptions in taxonomic practice. PLoS Biol 21(8): e3002251. https://doi.org/10.1371/journal.pbio.3002251

Riddle, B. R., Ladle, R. J., Lourie, S. A., & Whittaker, R. J. (2011). Basic Biogeography: Estimating Biodiversity and Mapping Nature. Conservation Biogeography. https://doi.org/10.1002/9781444390001.ch4

Sandall, E. L., Maureaud, A. A., Guralnick, R., McGeoch, M. A., Sica, Y. V., Rogan, M. S., Booher, D. B., Edwards, R., Franz, N., Ingenloff, K., Lucas, M., Marsh, C. J., McGowan, J., Pinkert, S., Ranipeta, A., Uetz, P., Wieczorek, J., & Jetz, W. (2023). A globally integrated structure of taxonomy to support biodiversity science and conservation. Trends in Ecology & Evolution. https://doi.org/10.1016/j.tree.2023.08.004

Sangster, G. (2009). Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. Proceedings of the Royal Society B: Biological Sciences, 276(1670), 3185–3191.

Sibley, C. G., & Monroe, B. L. (1990). Distribution and Taxonomy of Birds of the World. Yale University Press.

Simkins, A. T., Buchanan, G. M., Davies, R. G., & Donald, P. F. (2020). The implications for conservation of a major taxonomic revision of the world's birds. Animal Conservation, 23(4), 345–352.

Stankowski, S., & Ravinet, M. (2021). Quantifying the use of species concepts. Current Biology, 31(9), R428–R429.

Stropp, J., Ladle, R. J., Emilio, T., Lessa, T., & Hortal, J. (2022). Taxonomic uncertainty and the challenge of estimating global species richness. Journal of Biogeography, 49(9), 1654–1656. https://doi.org/10.1111/jbi.14463

Thiele, K. R., Conix, S., Pyle, R. L., Barik, S. K., Christidis, L., Costello, M. J., van Dijk, P. P., Kirk, P., Lien, A., & Thomson, S. A. (2021). Towards a global list of accepted species I. Why taxonomists sometimes disagree, and why this matters. Organisms Diversity & Evolution, 21(4), 615–622.

Thomson, S. A., Thiele, K., Conix, S., Christidis, L., Costello, M. J., Hobern, D., Nikolaeva, S., Pyle, R. L., van Dijk, P. P., & Weaver, H. (2021). Towards a global list of accepted species II. Consequences of inadequate taxonomic list governance. Organisms Diversity & Evolution, 21(4), 623–630.

Troudet, J., Grandcolas, P., Blin, A., & Vignes-lebbe, R. (2017). Taxonomic bias in biodiversity data and societal preferences. Scientific Reports, 7(9132), 1–14. https://doi.org/10.1038/s41598-017-09084-6

Vaidya, G., Lepage, D., & Guralnick, R. (2018). The tempo and mode of the taxonomic correction process: How taxonomists have corrected and recorrected North American bird species over the last 127 years. PLoS ONE, 13(4), 1–19. https://doi.org/https://doi.org/10.1371/journal.pone.0195736

Wheeler, Q. (2020). A taxonomic renaissance in three acts. Megataxa, 1(1), 4–8.

Whittaker, R. J., Ara, xfa, jo, M. B., Jepson, P., Ladle, R. J., James, E. M. W., Willis, K. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation Biogeography: Assessment and Prospect. Diversity and Distributions. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Williams, P. H. (2022). Novel splitting/lumping index reflects the history of species concepts applied to bumblebees (Insecta: Apidae). Zoological Journal of the Linnean Society, 196(2), 704–719. https://doi.org/https://doi.org/10.1093/zoolinnean/zlab123

Zachos, F. E. (2016). Species Concepts in Biology. Springer International Publishing. https://doi.org/10.1007/978-3-319-44966-1

Zachos, F. E. (2018a). (New) Species concepts, species delimitation and the inherent limitations of taxonomy. Journal of Genetics, 97(4), 811–815.

Zachos, F. E. (2018b). Mammals and meaningful taxonomic units: the debate about species concepts and conservation. Mammal Review, 48(3), 153–159.

Zapata, F. A., & Ross Robertson, D. (2006). How many species of shore fishes are there in the Tropical Eastern Pacific? Journal of Biogeography, 34(1), 38–51. https://doi.org/10.1111/j.1365-2699.2006.01586.x

Capítulo 2
Revista: Global ecology and biogeography
Status: Finalizado para submissão

Do biological and ecological variables influence nomenclatural uncertainty in birds?

Thainá Lessa[a*], Fernanda Alves-Martins[b,c], Javier Martinez Arribas[b,c], Karoline Azevedo[a], Ana C.M. Malhado[a], Juliana Stropp[d,e], Richard J. Ladle[a]

[a]Institute of Biological and Health Sciences, Federal University of Alagoas, 57072-970 Maceió, Alagoas, Brazil.
[b]CIBIO-InBIO, Research Centre in Biodiversity and Genetic Resources, University of Porto, Campus de Vairão, 4485-661 Vairão, Portugal.
[c]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[d]Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain
[e]Department of Biogeography, Trier University, Trier, Germany.

*Corresponding author: Thainá Lessa;
Email: thainalessaps@gmail.com

Abstract

Aim: Taxonomic nomenclature is the system by which scientists standardise and communicate unambiguously about organisms. However, even well-studied taxa such as birds there are taxonomic inconsistencies, due to divergences in the concept of species. As a result, there are multiple lists of birds, with different estimates of species number. Information on which characteristics of birds influence the degree of taxonomic uncertainty is incipient. Here we create a metric to calculate nomenclatural uncertainty exploring biases in bird species and Orders. Further, we analyse associations between our nomenclatural uncertainty metric and biological and ecological variables of bird species.

Location: Global.

Time period: Present.

Major taxa studied: Birds.

Methods: We used the International Ornithological Community (IOC) World Bird List which compiles and compares among other global bird lists the scientific names of species around the world. We created a nomenclatural uncertainty score for 11,140 bird species, using the proportion of disagreement (and absence) of scientific names between nine world bird lists. We assess drivers of variation in our nomenclatural uncertainty metric from bird' biological and ecological variables: body mass, range size, habitat density, lifestyle, IUCN status, and evolutionary distinctiveness.

Results: More than 50% of global bird species presented some degree of nomenclatural uncertainty, and for bird Orders the percentage was greater than 80%. Birds with smaller body size, smaller geographic range, classified by the IUCN as threatened and deficient data, and with low evolutionary distinctiveness were the most affected by nomenclatural uncertainty.

Main conclusions: Biological-ecological variables such as body mass and range size are strictly related to the accessibility and convenience of taxonomists to collect and describe species. Larger, more widely distributed, and more common species are easier to observe and collect. Species with characteristics contradictory to these should be priorities for assessments of taxonomic stability.

Key-words: taxonomy, nomenclature, linnean shortfall, conservation, ornithology

Introduction

Taxonomy is the first step of most systematic biology and natural history studies, and underpins our understanding of biodiversity (Agnarsson & Kuntner, 2007). As the fundamental and most frequently used taxonomic unit, 'species' are universally used in studies of biodiversity, ecology, evolution, biogeography and conservation (Agapow et al., 2004; Barrowclough et al., 2016; De Queiroz, 2007; Isaac et al., 2004; Ladle & Whittaker, 2011). However, in a dynamic and complex nature world achieving comprehensive knowledge about any aspect of species remains largely impracticable (Ladle & Hortal, 2013). Disagreements over the taxonomic classification and identity of species can have significant impacts on fundamental characteristics such as the number of known and unknown species. The gap in our knowledge of species identities is referred to as Linnean shortfall (Brown & Lomolino, 1998) and encompasses both species that have not yet been discovered in nature, and species that have been collected but not formally named (Hortal et al., 2015). The latter case includes specimens in museum collections that have never been adequately assessed (Bebber et al., 2010; Hortal et al., 2015), but also species currently included within another species – often recognized as subspecies or another taxonomic category – that will be upgraded to full species in the future through re-evaluation using new approaches (Parsons et al., 2022). This is further complex by the fact that species are hypotheses (Pante et al., 2014), and proposed changes in taxonomic status are not necessarily accepted or adopted by the scientific community.

It would be recommended that organisms collected be unambiguously described and named prior to further analysis (Riddle et al., 2011). However, the scientific consensus about how to identify and delimit species has changed dramatically over the last decades. The current taxonomic landscape is characterised by several distinct (and viable) concepts to define what species are (Kitchener et al., 2022; Zachos, 2016, 2018), and a large number of methodologies and criteria that are used to delimit it (Carstens et al., 2013; Hillis et al., 2021; Rannala, 2015). Modern approaches stress the importance of integrative taxonomy, combining genetic data with phenotypic, behavioral, and ecological traits to identify significant discontinuities and reproductively isolated populations that merit species status (Cicero et al., 2021). Such an approach requires high scientific capacity and technology, and extensive geographic sampling that substantially captures variations in population structure including putative contact zones. In summary, there remains a large amount of both conceptual and practical uncertainty around species data that increases with spatial and temporal scale.

Even in well-studied taxa such as birds there is a constant state of flux with knock on impacts for estimates of total species richness (known + unknown species) (Jetz et al., 2012; Neate-Clegg et al., 2021). Taxonomic inconsistency is reflected by the multiple bird lists worldwide, based on divergent taxonomic criteria (McClure et al., 2020; Neate-Clegg et al., 2021). These multiple bird lists attend to distinct purposes, have a singular importance for science and conservation, and are periodically reviewed by groups of experts based on a series of species definitions (Garnett & Christidis, 2007; Hey et al., 2003; Thomson et al., 2018), a methodology that can lead to discordant estimates of species richness and taxonomic uncertainty. A recent study found that 68% of four bird lists exhibit some degree of discordance, whether through taxonomic omissions, the use of different scientific name (full name, epithet, and/or genus), or treatment at the taxonomic level (species and subspecies) (McClure et al., 2020). Dealing with this challenge is crucial and requires commitments to the original conceptualization of species. The need for greater standardisation of data and methodologies to address how species are identified and catalogued is well recognized, and there are recent proposals for how this might eventually be dealt with (e.g., Cicero et al., 2021; Garnett et al., 2020; Orr et al., 2022). Until this happens, it is important that scientists better understand the causes and consequences of taxonomic uncertainty so that it might be incorporated into biogeographic and conservation models (Tessarolo et al., 2021).

One approach to this is to investigate the biocultural characteristics of taxonomically disputed/uncertain species. Here, it is important to distinguish between uncertainty about the identity of an observation or specimen (e.g. species A is mistakenly identified and recorded as species B) and uncertainty about the validity of a taxonomic identification. The latter can occur when the taxonomic status of a species is disputed (e.g. some scientists consider it a sub-species, others consider it a valid species) or when there are multiple unresolved taxonomies, leading to synonyms. Studying the drivers of taxonomic uncertainty thus faces the considerable challenge of distinguishing between and quantifying degrees of certainty and consensus for species level biodiversity data. All things being equal the probability of mistaken identity increases with decreasing levels of taxonomic/natural history expertise, and level of experience of the observer has been used as a proxy of taxonomic certainty in biogeographic studies (Tessarolo et al., 2021). Lower taxonomic categories (e.g. genera and species) are more problematic and have typically been evaluated through measures of nomenclatural stability. For example, degree of consensus among global species checklists, each one reflecting the taxonomic judgement of a subset of the global scientific community (Hobern et al., 2021).

Recognizing and quantifying gaps in knowledge about the species' identity is imperative, as distorted data can mislead our understanding of biodiversity descriptions and their ecological and evolutionary processes (Cardoso et al., 2011; Hortal et al., 2015; Stropp et al., 2022). A striking example is highlighted in global assessments of bird species numbers, which range from 10,000 to 18,000 depending on the application of species concept (Barrowclough et al., 2016; Mayr, 1946; Neate-Clegg et al., 2021). This disparity shows the importance of addressing taxonomic uncertainties to enhance the accuracy and reliability of biodiversity estimates. The first step is to assess these official lists of species in an attempt to make them a unique global reference (Lien et al., 2021) evaluate whether the intrinsic characteristics of birds influence uncertainties. Here, we compare scientific name disagreement between world bird lists using the International Ornithological Community (IOC) World Bird List, version 13.1 as a reference list. We create a nomenclatural uncertainty score using the proportion of disagreement (and absence) of scientific names between the world bird lists, and we explore biases in bird species and Orders. Finally, we analyse statistical associations between biological and ecological variables of bird species and our nomenclatural uncertainty score.

Methods

World bird lists

We downloaded the International Ornithological Community (IOC) World Bird List version 13.1 (Gill et al., 2023) (hereafter IOC 2023.1) in November 2023. The IOC 2023.1 is an open access database that compiles the scientific names of birds species and subspecies globally (living and extinct) based on evolutionary classification (Gill et al., 2024), and compares it with two previous versions (IOC 2022.1 and IOC 2022.2) and other eight bird lists: eBird/Clements (2022), HBW and BirdLife International (2022), Jimmy Gaudin (2021-2022), John H. Boyd (2019-2021), Howard and Moore (2014), Sibley and Monroe (1993); del Hoyo et al. (1992-2013), Peters et al. (1931-1986).

The IOC 2023.1 contains 30,971 bird species and subspecies, with 1,031 bird species mentioned in other lists not recognized by IOC 2023.1. As a taxonomic delimitation, we filtered only ‘species’ classification, resulting in 11,140 bird species. We made two types of comparisons: a) between the numbers of species among 11 world bird lists: we conducted pairwise comparisons between the bird lists to assess the percentage of agreement in the number of bird species (i.e. those species that had the same scientific names) across these 11 world bird lists, using IOC 2023.1 as reference list. In disagreement cases, we evaluated where it occurred in the scientific name, for example, discordances in the generic name (genus), specific name (epithet) or complete name (Supplementary material 1); b) between the agreement/disagreement/absence of scientific names filled in nine world bird lists; this last database was used for nomenclature uncertainty calculations (see next section).

Taxonomic uncertainty score

We compute scientific name disagreement (and absence) of birds among world bird lists as a proxy for nomenclatural uncertainty. As our preliminary results showed that IOC 2022.1 and IOC 2022.2 are extremely similar to IOC 2023.1 (>99%), we removed them from the calculation to avoid underestimation. We performed nomenclatural uncertainty calculations in Microsoft Excel and did it for all 11,140 bird species listed in IOC 2023.1. We made the calculation by comparing whether the scientific name listed in the IOC 2023.1 was the same, different or if was not mentioned in the other eight lists (mentioned in the section above) (Table 1). For this, we first counted the number of filled cells with a scientific name in the range of the nine lists evaluated (each cell was a list: IOC 2023.1 and eight bird lists), using the scientific name listed in IOC 2023.1 as a counting criteria. The outcome was zero when the cell (list) was not filled with a scientific name. The results of these counts were summed and then divided by the difference of the total number of lists evaluated squared and the multiplication of the count of the number of lists evaluated and the number of lists not filled with a scientific name. Finally, the final values were subtracted by one to obtain nomenclatural uncertainty. We transformed the final values into percentages for a more intuitive representation (see Equation 1):

$$Tax_{un} = \left(1 - \frac{\sum_{i=n}^{n} \left|A_i\right|}{n^2 - \left(n \times n_0\right)}\right) x\ 100$$

Where $\sum_{i=n}^{n} \left|A_i\right|$ is the sum of the number of lists with a scientific name, obtained by counting from each one of the nine bird lists with a present (agreed or disagreed) or absent scientific names based on the IOC 2023.1 list; n is the number of lists evaluated (in our case this was nine); n0 is the number of lists not filled with a scientific name. This equation presents a comprehensive methodology for evaluating nomenclatural uncertainty, encompassing considerations of both unfilled list occurrences and the variety of scientific names within each list. Once we recognized that there are two contexts with different outcomes: scientific names absent in the lists and scientific names in disagreement based on a reference list. The final expression yields a percentage measure of nomenclatural uncertainty, facilitating interpretation and comparison across different taxonomic scenarios.

Table 1: Examples of nomenclatural uncertainty among scientific names on world bird lists.

IOC 2023.1	**eBird/ Clements (2022)**	**HBW and BirdLife Intern. (2022)**	**Howard and Moore (2014)**	**del Hoyo et al. (1992-2013)**	**Peters et al . (1931-1986)**	**John H. (2019-2021)**	**Jimmy Gaudin (2021-2022)**	**Sibley and Monroe (1993)**
Struthio camelus	*Struthio camelus*	*Struthio camelus*	*Struthio camelus*	*Struthio camelus*	*Struthio camelus*	*Struthio camelus*	*Struthio camelus*	*Struthio camelus*
Pardirallus nigricans	*Pardirallus nigricans*	*Pardirallus nigricans*	*Pardirallus nigricans*	*Pardirallus nigricans*	*Ortygonax nigricans*	*Pardirallus nigricans*	*Pardirallus nigricans*	*Pardirallus nigricans*
Ardea alba	*Ardea alba*	*Ardea alba*	*Ardea alba*	*Egretta alba*	*Ardea alba*	*Casmerodius albus*	*Casmerodius albus*	*Casmerodius albus*
Nesotriccus tumbezanus	*Nesotriccus tumbezana*	*Phaeomyias tumbezana*				*Phyllomyias tumbezanus*	*Nesotriccus tumbezanus*	

To illustrate the outputs, if the species has 100% of agreement between the nine bird lists, consequently has 0% of nomenclatural uncertainty. Therefore, nomenclatural uncertainty was considered null when there was complete agreement (uncertainty = 0%) and varying when there was disagreement or absence in scientific name between the lists (uncertainty ≠ 0%). For example, *Struthio camelus* had the same scientific name in all nine lists (including IOC 2023.1), so its nomenclatural uncertainty was 0%; In one of nine list, *Pardirallus nigricans* had a discordant scientific name, so this species had 20% of taxonomic uncertainty; *Ardea alba* had a discordant scientific name in four lists then there was 57% uncertainty associated with its scientific name; and *Nesotriccus tumbezanus* had its scientific name equal in one list, absent scientific name in four lists and discordant scientific name in three lists, resulting in 84% of nomenclatural uncertainty (Table 1; see detailed list in Supplementary material 2). In addition to assessing nomenclatural uncertainty across bird species, we also examined the distribution of nomenclatural uncertainty across bird Orders.

Biological and ecological variables of birds

To understand what drives the variation in nomenclatural uncertainty, we collected biological and ecological variables (Table 2) based on IOC 2023.1 bird list. Firstly, we cross-referenced the IOC 2023.1 list with the AVONET list - a database published in 2022 that promotes detailed curation on global bird's traits, ecology and biogeography (Tobias et al., 2022). AVONET database presents separate information from three bird lists: BirdLife International, eBird and BirdTree. We used data from BirdLife International as it contains the largest number of species (n=10,999). We selected information on body mass, range size, habitat density, lifestyle and IUCN conservation status. We group IUCN' status into three categories: i) Not threatened (NTR): encompasses both Least Concern (LC) and Near Threatened (NT); ii) Threatened (TR): comprises Vulnerable (VU), Endangered (EN) and Critically Endangered (CR); and iii) Data Deficient (DD). Nomenclatural uncertainty calculations were carried out for Extinct (EX) or Extinct in the Wild (EW) bird species, but they were not included in the analysis model. The variable habitat density has three categories: dense, semi-open and open habitats. The lifestyle variable has five categories: Aerial, Aquatic, Insessorial, Terrestrial and Generalist. As a result, we obtained data for 9,994 bird species listed in IOC 2023.1. Secondly, we collected data on the evolutionary distinctiveness (ED) of bird species from EDGE of Existence, a Zoological Society of London's conservation a program that supports conservation and research for species both evolutionarily distinct and globally threatened (Faith, 2019). This database provides data on 10,954 species, the ED Scores were used. Combining data from IOC 2023.1, AVONET and EDGE of Existence, was returned complete information on 9,864 bird species.

We model these biological and ecological variables as a function of nomenclatural uncertainty. Given that our response variable (nomenclatural uncertainty) represents a proportion, we conducted a beta regression using the 'betareg' package (Cribari-Neto & Zeileis, 2010). Prior to the analysis, we added 10e-06 to nomenclatural uncertainty scores to prevent zeros, as the beta distribution requires values greater than zero (for a similar approach see Lessa et al., 2024).

Table 2: Biological and ecological variables used to explain nomenclatural uncertainty among bird species. The table provides a brief assumption of why variables were included into the model and the source of data collected.

Variables	Justification	Format	Source
Body Mass	Small species are more difficult to collect (Hey et al., 2003), consequently present greater nomenclatural uncertainty.	Continuous	Avonet
Range Size	Species with restricted distribution area are more challenging and expensive to research (Moura & Jetz, 2021), so have greater taxonomic uncertainty.	Continuous	Avonet
Habitat Density	Species from dense habitats are harder to study (Neate-Clegg et al., 2021), thus have greater nomenclatural uncertainty.	Categoric	Avonet
Lifestyle	Generalist species are more common in urban environments and easier to observe (Callaghan et al., 2019), therefore they present lower taxonomic uncertainty.	Categoric	Avonet
IUCN	Threatened species attract less research interests (Jarić et al., 2019), therefore have greater nomenclatural uncertainty	Categoric	Avonet
Evolutionary distinctiveness	Species with greater evolutionary distinctiveness are more studied (dos Santos et al., 2020), thus have lower nomenclatural uncertainty.	Continuous	EDGE

Results

We analysed the number of species described in world bird lists and the degree of agreement between their scientific names in comparison to the 11,140 bird species from the reference list, IOC 2023.1. The lists that showed the highest percentage between the numbers of bird species with agreed scientific names by IOC 2023.1 were its previous versions IOC 2022.2 and IOC 2022.1, both showing 99% of agreement in scientific names (Figure 1). Following, the most recent lists with the largest number of species were also those that achieved the highest percentage of agreement in scientific names: eBird/Clements (2022) (95%) and HBW and BirdLife International (2022) (93%). On the other hand, the lists with the greatest disagreements in scientific names were the oldest with the lowest number of species: Peters (1931-1986) with more than 50% of disagreements in relation to IOC 2023.1, and Sibley & Monroe (1993) with 65% (Figure 1).

The majority of sources of disagreement in the scientific names in all lists were divergence in the generic name and full name. However, there was also discordance in the specific name, especially in the oldest lists: Peters (1931-1986), Sibley & Monroe (1993) and del Hoyo (1992-2013). For example, from 46% of bird species with scientific names in disagreement in Peters (1931-1986) list, 1,812 species had a generic name disagreed, 424 species had a specific name disagreed and 388 species had a scientific name completely disagreed. In Sibley & Monroe (1993) list 35% of bird species had disagreement in scientific names, being 1,718 species with disagreement in generic name, 323 species disagreement in specific name, and 326 species disagreement in full scientific name. In previous versions of IOC 2023.1, IOC 2022.1 had six species with a completely different scientific name, 33 species that had a generic name disagreed and three species with a specific name disagreed; and IOC 2022.2 also had six species with a completely scientific name disagreed and 17 species had their generic name disagreed (Figure 1).

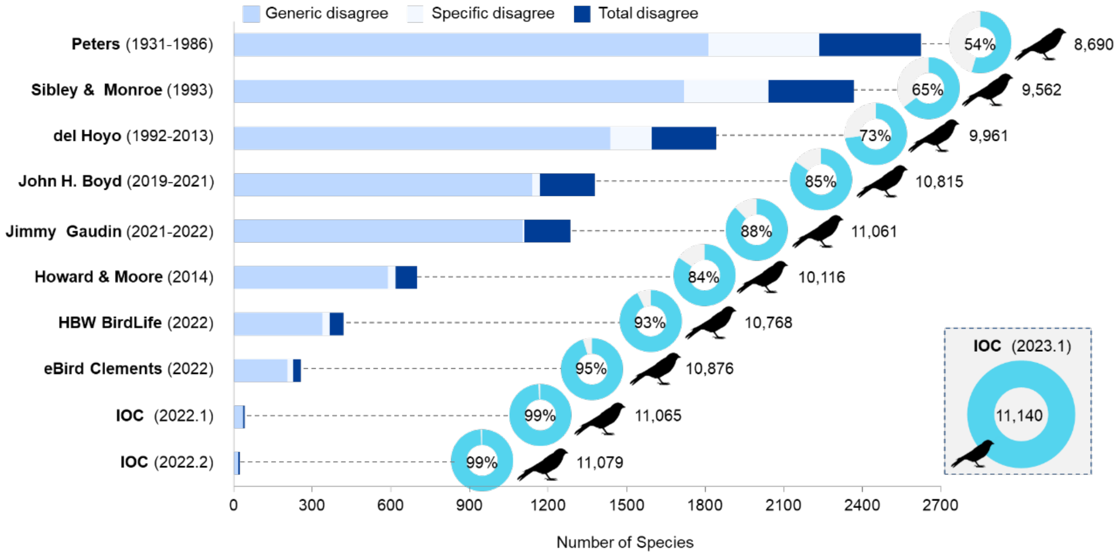
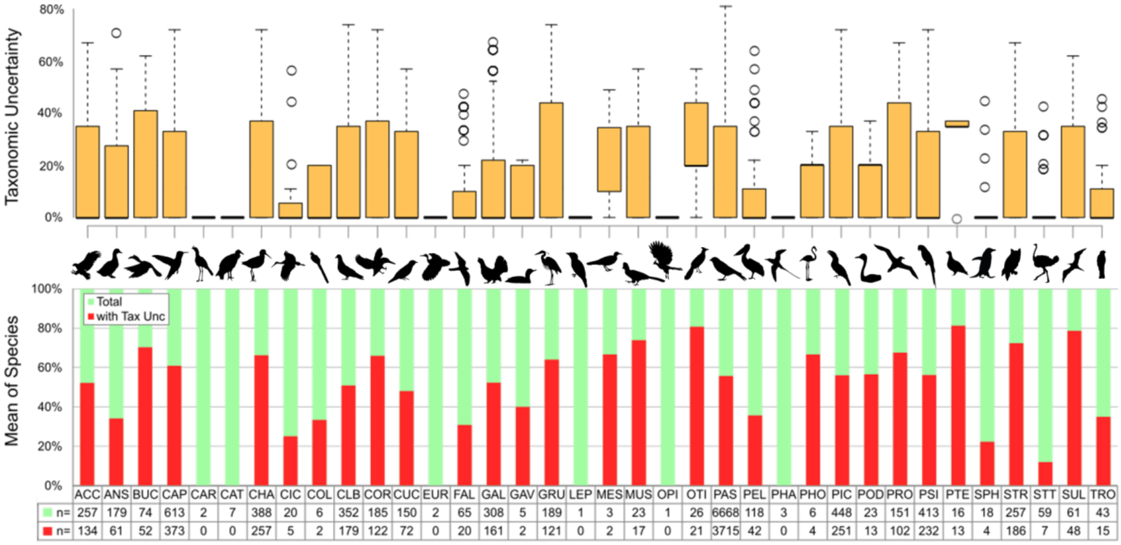


Figure 1: Bar chart representing the nomenclatural agreement of extant world bird lists in comparison with a reference list (IOC 2023.1). Dark blue bars represent the number of completely discordant species (genus and species different), medium blue represents change in genus (medium blue) and light blue represents a change in species name. The number next to the bird silhouette refers to the total number of bird species in each list. The percentages in the graph have been rounded to the nearest whole number. For full values, see Supplementary material 1.

From our nomenclatural uncertainty calculation, 36 bird Orders were evaluated (Figure 2). Only six bird Orders did not have nomenclatural uncertainty, i.e. all bird species had scientific names that agreed among all nine lists used in the metric. This bird Orders were also those had the smallest number of species: Cariamiformes (n= 2), Cathartiformes (n= 7), Eurypygiformes (n= 2), Leptosomiformes (n= 1), Opisthocomiformes (n= 1) and Phaethontiformes (n= 3). The Orders that had the highest mean of number of species with nomenclatural uncertainty were: Pterocliformes (n= 29) with 81% of the birds in this Order presenting nomenclatural uncertainty that ranged between 35% to 37% (mean of 36%); Otidiformes (n= 47) with 80% of the birds in this Order had nomenclatural uncertainty ranging between 20% to 69% (mean of 37%); Suliformes (n= 109) with 78% of the birds in this Order showed nomenclatural uncertainty varying between 20% to 78% (mean of 47%); Musophagiformes (n= 40) about 74% of the birds in this Order presented nomenclatural uncertainty that ranged between 11% to 64% (mean of 36%); Strigiformes (n= 443) almost 72% of the birds in this Order exhibited nomenclatural uncertainty ranging between 11% to 89% (mean of 37%); and Bucerotiformes (n= 126) with 70% of the birds in this Order presenting nomenclatural uncertainty that ranged between 11% to 67% (mean of 34%).



	ACC	ANS	BUC	CAP	CAR	CAT	CHA	CIC	COL	CLB	COR	CUC	EUR	FAL	GAL	GAV	GRU	LEP	MES	MUS	OPI	OTI	PAS	PEL	PHA	PHO	PIC	POD	PRO	PSI	PTE	SPH	STR	STT	SUL	TRO
🟩 n=	257	179	74	613	2	7	388	20	6	352	185	150	2	65	308	5	189	1	3	23	1	26	6668	118	3	6	448	23	151	413	16	18	257	59	61	43
🟥 n=	134	61	52	373	0	0	257	5	2	179	122	72	0	20	161	2	121	0	2	17	0	21	3715	42	0	4	251	13	102	232	13	4	186	7	48	15

Figure 2: Column graph representing the percentage of the number of species with nomenclatural uncertainty in bird Orders. The table below is the total number of species (green) and number of species with nomenclatural uncertainty (red) per Order. Boxplot (orange) shows the distribution of nomenclatural uncertainty in the Orders. Accipitriformes (ACC), Anseriformes (ANS), Bucerotiformes (BUC), Caprimulgiformes (CAP), Cariamiformes (CAR), Cathartiformes (CAT), Charadriiformes (CHA), Ciconiiformes (CIC), Coliiformes (COL), Columbiformes (CLB), Coraciiformes (COR), Cuculiformes (CUC), Eurypygiformes (EUR), Falconiformes (FAL), Galliformes (GAL), Gaviiformes (GAV), Gruiformes (GRU), Leptosomiformes (LEP), Mesitornithiformes (MES), Musophagiformes (MUS), Opisthocomiformes (OPI), Otidiformes (OTI), Passeriformes (PAS), Pelecaniformes (PEL), Phaethontiformes (PHA), Phoenicopteriformes (PHO), Piciformes (PIC), Podicipediformes (POD), Procellariiformes (PRO), Psittaciformes (PSI), Pterocliformes (PTE), Sphenisciformes (SPH), Strigiformes (STR), Struthioniformes (STT), Suliformes (SUL), Trogoniformes (TRO).

Further about our nomenclatural uncertainty calculation for the 11,140 bird species evaluated, over half species (56%) had some degree of nomenclatural uncertainty, varying from 11% to 89% (see Supplementary material 2). More than 16%

(n=1,812) of bird species had a nomenclatural uncertainty score greater than 50%. The birds species with the highest nomenclatural uncertainty score (89%) were: Centropus burchellii (Cuculiformes), Cinnyris whytei (Passeriformes) and Ramosomyia wagneri (Caprimulgiformes) which were only present on the IOC 2023.1 list and on another list with a scientific name in disagreement, and Loriculus bonapartei (Psittaciformes), Otus bikegila (Strigiformes), Phylloscopus floresianus (Passeriformes), Setophaga aestiva (Passeriformes) and Zosterops paruhbesar (Passeriformes) which were only present on the IOC 2023.1 list. Other bird species with a high nomenclatural uncertainty score (>81%) were: Nesotriccus tumbezanus (Passeriformes), Ptilinopus gularis (Columbiformes), Arborophila diversa (Galliformes), Melanerpes santacruzi (Piciformes), Psittacara brevipes (Psittaciformes) and Ramosomyia viridifrons (Caprimulgiformes).

The results of our model to assess the relationship between nomenclatural uncertainty score and biological and ecological variables of birds (Table 3) revealed that the body mass was associated with nomenclatural uncertainty, with a trend to the nomenclatural uncertainty being lower in larger bird species. The species distribution area was also associated with nomenclatural uncertainty score, which nomenclatural uncertainty was lower in bird species with a wider range size. Evolutionary distinctiveness, i.e. species from an unusual group with a remarkable evolutionary history, likewise showed an association with nomenclatural uncertainty, with uncertainty being lower in more distinctive bird species. Finally, IUCN conservation status was associated with nomenclatural uncertainty, and was greater in bird species categorised as Threatened (VU, EN and CR). The habitat density and lifestyle variables were not significant.

Table 3: Significant results from the Beta regression model analysing the association between nomenclatural uncertainty and biological and ecological variables of bird species. Full results with non-significant associations are available in Supplementary material 3.

Variables	**Coefficient estimate**	**z value**	**Pr(>|z|)**
Body Mass	-1.123e-05	-2.031	0.042208 *
Range Size	-5.549e-03	-3.779	0.000158 ***
Evol.Distinctiveness	-9.964e-03	-4.434	9.23e-06 ***
IUCN Threatened	1.915e-01	5.485	4.13e-08 ***

Discussion

Birds are one of the most studied classes of vertebrate fauna worldwide and recognized to show a taxonomic stability (Barrowclough et al., 2016; Neate-Clegg et al., 2021). However, there are still major gaps and deficiencies regarding the classification and nomenclature of this group. Our results revealed that there are great divergences in the number of species in each bird list evaluated, with the International Ornithological Community (IOC) lists presenting the highest species count. The IOC carries out biannual updates of taxonomic reviews, and the newest update published in 2024 already counts an increase in the number of bird species (55 new species) (Gill et al., 2024). On the other hand, older bird checklists such as Peters (1931-1986), Sibley & Monroe (1993) and del Hoyo (1992-2013) set fewer species numbers, which is not surprising due to technological limitations to describe and name species formerly. Nonetheless, the existence of these checklists demonstrates the importance of pioneering studies for the improvement of more complete studies/lists, such as the creation of the eBird/Clements and the Howard and Moore lists that were based on Peters list (1931-1986) (Lepage & Warnier, 2014).

The checklist from the citizen science platform eBird (eBird/Clements (2022)) and the list used by the International Union for Conservation of Nature (IUCN) (HBW and BirdLife International (2022)) also exhibit a high number of species and similarity of scientific names in comparison to the IOC 2023.1 list, with 95% and 93% of agreement, respectively. These discrepancies, to a large or small degree, are most likely related to the species concept applied by each list. For example, the IOC list uses the evolutionary species concept, while the eBird and, HBW and BirdLife International lists applies the biological species concept (McClure et al., 2020). There is already evidence that when different species concepts are applied occur changes in estimates of species richness, and even in patterns of endemism. For sub-Saharan African birds there was a 33% increase in the number of species using the phylogenetic species concept (n= 2,098) in relation to the biological species concept (n= 1,572) (Dillon & Fjeldså, 2005). The same standard was observed for birds from Mexico, which after literature and museum specimens review, 135 biological species were split into 323 phylogenetic/evolutionary species, increasing 125% in the number of species endemic to Mexico (Navarro-Sigüenza & Peterson, 2004; Peterson & Navarro-Sigüenza, 1999).

In our study, more than half (56%) of global bird species showed some degree of nomenclatural uncertainty associated with differences in the scientific name among nine global bird lists. The scenario becomes more critical in higher taxonomic levels, with more than 80% of bird Orders holding nomenclatural uncertainty. The factors of disagreement in scientific names were especially linked to processes of splitting, lumping and taxonomic swap. These processes, especially splitters, can result in 'taxonomic inflations and deflations', i.e., increase and decrease in the number of species by taxonomic changes and delimitations (Isaac et al., 2004). Debates are intense about whether taxonomic inflation/deflation is real or it is simply a reflection of taxonomic progress (Garnett & Christidis, 2017; Harris & Froufe, 2005; Sangster, 2009; Wheeler, 2020). Perhaps this widespread concern among ornithologists could be better addressed by understanding and appropriately applying integrative taxonomy – that is to include multiple lines of evidence that are useful when inferring reproductive isolation to determine species limits (Mandiwana-Neudani et al., 2021; Winker & Rasmussen, 2021). In any case, taxonomic changes need to be recognized and incorporated into biodiversity metrics (Stropp et al., 2022)).

The most marked divergences between the scientific names across bird lists evaluated here were in the generic name (genus), full scientific name and specific name (epithet), although the latter was less prominent. This trend should be also a reflection of taxonomic progress that began to include molecular data and phylogenetic relationships in species descriptions, leading to an increase in the number of genera, as observed in ferns (Christenhusz & Chase, 2018). Another explanation is that taxonomists may prefer to split genera into smaller groups to facilitate and detail ecological and evolutionary knowledge of species (Gasper, 2016). Although international codes of nomenclature stipulate rules, there are no objective criteria of what constitutes a category (e.g., genera or family) in biological classification. Taxonomists are free to choose scientific names, based on available data, and propose it for acceptance (or not) to the scientific community (Rheindt et al., 2023). Perhaps taxonomy will never reach a definitive consensus due to the dynamics and constant evolution of nature. But rather also for human characteristics, since the designation of what a species is, is part of human judgement and culture. Even if there was a single species concept and scientists worked with the same dataset, there would be disagreements (Conix et al., 2023).

Some trends explain why many species are awaiting to be discovered while others have already undergone several taxonomic reviews (Alroy, 2003; May, 1988; Moura & Jetz, 2021). As revealed in our model, the bird species most affected by nomenclatural uncertainty were those that had smaller body size, smaller geographic distribution (i.e. restricted or endemic to a region), species classified by the IUCN as threatened and deficient data, and species with low evolutionary distinctiveness. For example, Otus bikegila (Strigiformes) is an owl critically endangered and endemic to Principe Island in São Tomé and Príncipe, and Zosterops paruhbesar (Passeriformes) is a small passerine listed as endangered and restricted to Wangi-wangi Island, Indonesia, both species were recently described after careful review using a huge of evidences (Irham et al., 2023; Melo et al., 2022).

Furthermore, several species with the highest nomenclatural uncertainty scores were considered subspecies among the lists, such as Phylloscopus floresianus (Passeriformes) which is restricted to Timor-Leste and is considered a subspecies (Phylloscopus presbytes floresianus and Phylloscopus presbytes floris) in other lists (see Supplementary material 2). The debate surrounding the utility of subspecies has persisted for decades, with the controversial principle that they are critical stages in evolution, forming part of a continuum from limited differentiation among populations to reproductive isolation (De Queiroz, 2007). However, many subspecies often require re-examination to assess their validity, as they were described decades ago based on few specimens and characters, with nomenclature applied to populations that are poorly diagnosable (Cicero et al., 2021; Winker & Haig, 2010), which could be resulting in these higher rates of nomenclatural uncertainty.

Two biological and ecological variables are strictly related to the accessibility and convenience of taxonomists in collecting and describing species, such as body size and geographic range (Clark & May, 2002; Ladle & Whittaker, 2011; Moura & Jetz, 2021). For example, for Australian birds, more scientific publications are found for those with larger body size, wide geographic distribution and abundance (Yarwood et al., 2019). Furthermore, a similar study evaluating levels of agreement between four bird lists and ecological variables demonstrated that agreement was greater in larger species (Neate-Clegg et al., 2021). The premise behind this is that larger species are easier to observe and collect, but are also correlated with lower diversity, consequently presenting lower taxonomic uncertainties (Alroy, 2003; dos Santos et al., 2020; Hey et al., 2003). Similarly, species with wide geographic range are also more likely to be collected and studied, as their distribution often overlaps with the distribution of research groups/taxonomists (dos Santos et al., 2020; Meyer et al., 2015; Yarwood et al., 2019), and therefore, are less likely to have nomenclatural uncertainty.

A similar bias was observed for bird Orders, although we did not evaluate it statistically. Birds from Orders that presented the greatest number of species with nomenclatural uncertainty share some biological and ecological characteristics, which may be hindering field collection and taxonomic investigation. For example, species belonging to Pterocliformes (sandgrouse) are small in size and range, and have cryptic plumage. Otidiformes (bustards) species also have cryptic plumage, solitary behaviour, and live in savannas and/or arid landscapes. And Strigiformes species, such as owls, although they have a wide geographic range, have solitary and nocturnal behaviour.

Our nomenclatural uncertainty metric has a number of unavoidable limitations. Firstly, it is important to acknowledge that taxonomy is a dynamic, continually evolving science and that our analysis is only a 'snapshot' of the current level of nomenclatural discordance. As the taxonomic system evolves further discordance will be generated and, at the same time, older species lists will fall into obsolescence. Alternatively, the current taxonomic and nomenclatural system may be revised and a single, authoritative global species list may emerge (Lien et al., 2021; Pyle et al., 2021) and become universally adopted. Until that time, duplicated efforts and confusion in the scientific use of taxonomic information will continue to breed nomenclatural uncertainty along with its negative consequences for ecologically and conservation research (Thomson et al., 2021). It is also important to note that the status of some taxonomic units will always be contentious since, even when faced with the same data, experts may disagree about the taxonomic status of groups in the process of speciation (Conix et al., 2023).

Previous studies have already carried out investigations into taxonomic quality, such as the binary metric proposed by Neate-Clegg et al., (2021) that assesses agreement between taxonomic authorities for global birds; and the indicator of identifiers' expertise (taxonomists, experts and amateurs) on the quality of taxonomic identifications for Scarabaeidae dung beetle species (Tessarolo et al., 2021). However, the aforementioned metrics are limited by not detailing the degree of taxonomic/nomenclatural quality or/and needs to prior know regarding the types of identifiers in the database, which may be impractical in global assessments. Our nomenclatural uncertainty metric has the advantage of evaluating the degree of nomenclatural uncertainty, in a broader and finer resolution, making bird species and Order comparable, and using the available taxonomic data. In addition, it was possible to assess the motivations behind the uncertainty, whether divergences in scientific names and the biological and ecological characteristics of the species. Nevertheless, the creation of our metric was only allowed due to the hard work carried out by researchers and institutions, which compiled and made taxonomic information accessible.

Despite the immeasurable relevance of taxonomy as a basic science for biodiversity studies, there are vast deficiencies that have led to a global crisis. On the one hand, scientists are increasingly striving against the clock to collect and describe species before they become threatened due to current environmental impacts worldwide (Stropp et al., 2020). On the other hand, taxonomy is no longer seen with the 'glamour' it once was, with there being less interest among young scientists, as well as there being a shortage of taxonomists for many taxa (Engel et al., 2021). Especially due to the scarce investments in the area and the bureaucracy involved from the description to the publication of a new species (Britz et al., 2020). The requirement for additional information beyond morphology increases the impediments to publishing articles on review and descriptions of new species. As we saw here, approximately 12% of bird species (not counting the subspecies) had no available data about biology and ecology. Therefore, science will only fill the remaining empty boxes of knowledge if the Linnean shortfalls are first filled (Hortal et al., 2015). Only when we join efforts and financing actions that protect both ‘species’ in danger of extinction, will we move towards true taxonomic progress (Wägele et al., 2011).

References

Agapow, P., Bininda-Emonds, O. R. P., Crandall, K. A., Gittleman, J. L., Mace, G. M., Marshall, J. C., & Purvis, A. (2004). The Impact of Species Concept on Biodiversity Studies. The Quarterly Review of Biology, 79(2), 161–179. https://doi.org/10.1086/383542

Agnarsson, I., & Kuntner, M. (2007). Taxonomy in a Changing World: Seeking Solutions for a Science in Crisis. Systematic Biology, 56(3), 531–539. https://doi.org/10.1080/10635150701424546

Alroy, J. (2003). Taxonomic inflation and body mass distributions in north american fossil mammals. Journal of Mammalogy, 84(2), 431–443. https://doi.org/10.1644/1545-1542(2003)084<0431:TIABMD>2.0.CO;2

Barrowclough, G. F., Cracraft, J., Klicka, J., & Zink, R. M. (2016). How many kinds of birds are there and why does it matter? PLoS One, 11(11), e0166307.

Bebber, D. P., Carine, M. A., Wood, J. R. I., Wortley, A. H., Harris, D. J., Prance, G. T., Davidse, G., Paige, J., Pennington, T. D., Robson, N. K. B., & Scotland, R. W. (2010). Herbaria are a major frontier for species discovery. Pnas, 107(51), 22169–22171. https://doi.org/10.1073/pnas.1011841108

Britz, R., Hundsdorfer, A., & Fritz, U. (2020). Funding, training, permits—the three big challenges of taxonomy. Megataxa, 1(1). https://doi.org/10.11646/megataxa.1.1.10

Brown, J. H., & Lomolino, M. V. (1998). Biogeography.

Callaghan, C. T., Major, R. E., Wilshire, J. H., Martin, J. M., Kingsford, R. T., & Cornwell, W. K. (2019). Generalists are the most urban-tolerant of birds: a phylogenetically controlled analysis of ecological and life history traits using a novel continuous measure of bird responses to urbanization. Oikos, 128(6), 845–858. https://doi.org/10.1111/oik.06158

Cardoso, P., Erwin, T. L., Borges, P. A. V., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation, 144(11), 2647–2655. https://doi.org/10.1016/j.biocon.2011.07.024

Carstens, B. C., Pelletier, T. A., Reid, N. M., & Satler, J. D. (2013). How to fail at species delimitation. Molecular Ecology, 22(17), 4369–4383. https://doi.org/10.1111/mec.12413

Christenhusz, M. J. M., & Chase, M. W. (2018). PPG recognises too many fern genera. TAXON, 67(3), 481–487. https://doi.org/10.12705/673.2

Cicero, C., Mason, N. A., Jiménez, R. A., Wait, D. R., Wang-Claypool, C. Y., & Bowie, R. C. K. (2021). Integrative taxonomy and geographic sampling underlie successful species delimitation. Ornithology, 138(2). https://doi.org/10.1093/ornithology/ukab009

Clark, J. A., & May, R. M. (2002). Taxonomic Bias in Conservation Research. Science, 297(5579), 191–192. https://doi.org/10.1126/science.297.5579.191b

Conix, S., Cuypers, V., Zachos, F. E., & De Block, A. (2023). Taxonomic disagreement about ranks in gray-area taxa: A vignette study. BioScience, 73(10), 728–737. https://doi.org/10.1093/biosci/biad081

Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. Journal of Statistical Software, 34(2). https://doi.org/10.18637/jss.v034.i02

De Queiroz, K. (2007). Species Concepts and Species Delimitation. Systematic Biology, 56(6), 879–886. https://doi.org/10.1080/10635150701701083

Dillon, S., & Fjeldså, J. (2005). The implications of different species concepts for describing biodiversity patterns and assessing conservation needs for African birds. Ecography, 28(5), 682–692. https://doi.org/10.1111/j.2005.0906-7590.04344.x

dos Santos, J. W., Correia, R. A., Malhado, A. C. M., Campos-Silva, J. V., Teles, D., Jepson, P., & Ladle, R. J. (2020). Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. Animal Conservation, 23(6), 679–688. https://doi.org/10.1111/acv.12586

Engel, M. S., Ceríaco, L. M. P., Daniel, G. M., Dellapé, P. M., Löbl, I., Marinov, M., Reis, R. E., Young, M. T., Dubois, A., Agarwal, I., Lehmann A., P., Alvarado, M., Alvarez, N., Andreone, F., Araujo-Vieira, K., Ascher, J. S., Baêta, D., Baldo, D., Bandeira, S. A., … Zacharie, C. K. (2021). The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. Zoological Journal of the Linnean Society, 193(2), 381–387. https://doi.org/10.1093/zoolinnean/zlab072

Faith, D. P. (2019). EDGE of existence and phylogenetic diversity. Animal Conservation, 22(6), 537–538. https://doi.org/10.1111/acv.12552

Garnett, S. T., & Christidis, L. (2007). Implications of changing species definitions for conservation purposes. Bird Conservation International, 17(3), 187–195. https://doi.org/10.1017/S0959270907000809

Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. Nature News, 546(7656), 25.

Garnett, S. T., Christidis, L., Conix, S., Costello, M. J., Zachos, F. E., Bánki, O. S., Bao, Y., Barik, S. K., Buckeridge, J. S., Hobern, D., Lien, A., Montgomery, N., Nikolaeva, S., Pyle, R. L., Thomson, S. A., van Dijk, P. P., Whalen, A., Zhang, Z.-Q., & Thiele, K. R. (2020). Principles for creating a single authoritative list of the world's species. PLOS Biology, 18(7), e3000736. https://doi.org/10.1371/journal.pbio.3000736

Gasper, A. L. de. (2016). The fern family Blechnaceae: old and new genera re-evaluated, using molecular data. Universidade Federal de Minas Gerais.

Gill, Donsker, & Rasmussen. (2024). IOC World Bird List. https://www.worldbirdnames.org/new/

Gill, F., Donsker, D., & Rasmussen, P. (2023). IOC World Bird List (v13.1). https://doi.org/10.14344/IOC.ML.13.1.

Harris, D. J., & Froufe, E. (2005). Taxonomic inflation: species concept or historical geopolitical bias? Trends in Ecology & Evolution, 20(1), 6–7.

Hey, J., Waples, R. S., Arnold, M. L., Butlin, R. K., & Harrison, R. G. (2003). Understanding and confronting species uncertainty in biology and conservation. Trends in Ecology & Evolution, 18(11), 597–603. https://doi.org/10.1016/j.tree.2003.08.014

Hillis, D. M., Chambers, E. A., & Devitt, T. J. (2021). Contemporary Methods and Evidence for Species Delimitation. Ichthyology & Herpetology, 109(3). https://doi.org/10.1643/h2021082

Hobern, D., Barik, S. K., Christidis, L., T Garnett, S., Kirk, P., Orrell, T. M., Pape, T., Pyle, R. L., Thiele, K. R., & Zachos, F. E. (2021). Towards a global list of accepted species VI: The Catalogue of Life checklist. Organisms Diversity & Evolution, 1–14.

Hortal, J., Bello, F. de, Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst., 46, 523–552. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Irham, M., Prawiradilaga, D. M., Menner, J. K., O'Connell, D. P., Kelly, D. J., Analuddin, K., Karya, A., Meads, M., Marples, N. M., & Rheindt, F. E. (2023). A distinct new species of Zosterops white-eye from the Sulawesi region, Indonesia. Ibis, 165(3), 808–816. https://doi.org/10.1111/ibi.13148

Isaac, N. J. B., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution, 19(9), 464–469.

Jarić, I., Correia, R. A., Roberts, D. L., Gessner, J., Meinard, Y., & Courchamp, F. (2019). On the overlap between scientific and societal taxonomic attentions — Insights for conservation. Science of The Total Environment, 648, 772–778. https://doi.org/10.1016/j.scitotenv.2018.08.198

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. Nature, 491(7424), 444–448. https://doi.org/10.1038/nature11631

Kitchener, A. C., Hoffmann, M., Yamaguchi, N., Breitenmoser-Würsten, C., & Wilting, A. (2022). A system for designating taxonomic certainty in mammals and other taxa. Mammalian Biology, 1–11.

Ladle, R. J., & Hortal, J. (2013). Mapping species distributions: living with uncertainty. Frontiers of Biogeography, 5(1). https://doi.org/10.21425/F55112942

Ladle, R. J., & Whittaker, R. J. (2011). Conservation Biogeography. In Conservation Biogeography. https://doi.org/10.1002/9781444390001

Lepage, J., & Warnier, J. (2014). The Peters' Check-list of the Birds of the World (1931-1987) Database. https://avibase.bsc-eoc.org/peterschecklist.jsp

Lessa, T., Alves-Martins, F., Martinez-Arribas, J., Correia, R. A., Mendelsohn, J., Chimbioputo Fabiano, E., Angombe, S. T., Malhado, A. C. M., & Ladle, R. J. (2024). Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa. Ecological Indicators, 158, 111490. https://doi.org/10.1016/j.ecolind.2023.111490

Lien, A. M., Conix, S., Zachos, F. E., Christidis, L., van Dijk, P. P., Bánki, O. S., Barik, S. K., Buckeridge, J. S., Costello, M. J., & Hobern, D. (2021). Towards a global list of accepted species IV: Overcoming fragmentation in the governance of taxonomic lists. Organisms Diversity & Evolution, 21(4), 645–655.

Mandiwana-Neudani, T. G., Crowe, T. M., Little, R. M., & Bowie, R. C. (2021). The importance of adopting an integrative taxonomy framework in species delimitation: Response to Hunter et al. (2021). Ostrich, 92(2), 162–167. https://doi.org/10.2989/00306525.2021.1898170

May, R. M. (1988). How Many Species Are There on Earth? Science, 241(4872), 1441–1449. https://doi.org/10.1126/science.241.4872.1441

Mayr, E. (1946). The number of species of birds. The Auk, 63(1), 64–69.

McClure, C. J. W., Lepage, D., Dunn, L., Anderson, D. L., Schulwitz, S. E., Camacho, L., Robinson, B. W., Christidis, L., Schulenberg, T. S., Iliff, M. J., Rasmussen, P. C., & Johnson, J. (2020). Towards reconciliation of the four world bird lists: hotspots of disagreement in taxonomy of raptors. Proceedings of the Royal Society B: Biological Sciences, 287(1929). https://doi.org/10.1098/rspb.2020.0683

Melo, M., Freitas, B., Verbelen, P., da Costa, S. R., Pereira, H., Fuchs, J., Sangster, G., Correia, M. N., de Lima, R. F., & Crottini, A. (2022). A new species of scops-owl (Aves, Strigiformes, Strigidae, Otus) from Príncipe Island (Gulf of Guinea, Africa) and novel insights into the systematic affinities within Otus. ZooKeys, 1126, 1–54. https://doi.org/10.3897/zookeys.1126.87635

Meyer, C. F. J., Aguiar, L. M. S., Aguirre, L. F., Baumgarten, J., Clarke, F. M., Cosson, J.-F., Estrada Villegas, S., Fahr, J., Faria, D., Furey, N., Weise, C. D., & Kalko, E. K. V. (2015). Species undersampling in tropical bat surveys: Effects on emerging biodiversity patterns. Journal of Animal Ecology, 84(1), 113–123. https://doi.org/10.1111/1365-2656.12261

Moura, M. R., & Jetz, W. (2021). Shortfalls and opportunities in terrestrial vertebrate species discovery. Nature Ecology & Evolution, 5(5), 631–639. https://doi.org/10.1038/s41559-021-01411-5

Navarro-Sigüenza, A. G., & Peterson, A. T. (2004). An alternative species taxonomy of the birds of Mexico. Biota Neotropica, 4(2), 1–32. https://doi.org/10.1590/s1676-06032004000200013

Neate-Clegg, M. H. C., Blount, J. D., & Şekercioğlu, Ç. H. (2021). Ecological and biogeographical predictors of taxonomic discord across the world's birds. Global Ecology and Biogeography, 30(6), 1258–1270. https://doi.org/10.1111/geb.13300

Orr, M. C., Feijó, A., Chesters, D., Vogler, A. P., Bossert, S., Ferrari, R. R., Costello, M. J., Hughes, A. C., Krogmann, L., Ascher, J. S., Zhou, X., Li, D.-Z., Bai, M., Chen, J., Ge, D., Luo, A., Qiao, G., Williams, P. H., Zhang, A., … Zhu, C.-D. (2022). Six steps for building a technological knowledge base for future taxonomic work. National Science Review, 9(12). https://doi.org/10.1093/nsr/nwac284

Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A., Destombe, C., Forcioli, D., Valero, M., Viard, F., & Samadi, S. (2014). Species are hypotheses: avoid connectivity assessments based on pillars of sand. Molecular Ecology, 24(3), 525–544. https://doi.org/https://doi.org/10.1111/mec.13048

Parsons, D. J., Pelletierc, T. A., Wieringa, J. G., Ducketta, D. J., & Carstens, B. C. (2022). Analysis of biodiversity data suggests that mammal species are hidden in predictable places. Pnas, 119(14), 1–7. https://doi.org/10.1073/pnas.2103400119

Peterson, A. T., & Navarro-Sigüenza, A. G. (1999). Alternative species concepts as bases for determining priority conservation areas. Conservation Biology, 13, 427–431.

Pyle, R. L., Barik, S. K., Christidis, L., Conix, S., Costello, M. J., van Dijk, P. P., Garnett, S. T., Hobern, D., Kirk, P. M., & Lien, A. M. (2021). Towards a global list of accepted species V. The devil is in the detail. Organisms Diversity & Evolution, 1–19.

Rannala, B. (2015). The art and science of species delimitation. Current Zoology, 61(5), 846–853. https://doi.org/10.1093/czoolo/61.5.846

Rheindt, F. E., Bouchard, P., Pyle, R. L., Welter-Schultes, F., Aescht, E., Ahyong, S. T., Ballerio, A., Bourgoin, T., Ceríaco, L. M. P., Dmitriev, D., Evenhuis, N., Grygier, M. J., Harvey, M. S., Kottelat, M., Kluge, N., Krell, F.-T., Kojima, J., Kullander, S. O., Lucinda, P., … Pape, T. (2023). Tightening the requirements for species diagnoses would help integrate DNA-based descriptions in taxonomic practice. PLOS Biology, 21(8), e3002251. https://doi.org/10.1371/journal.pbio.3002251

Riddle, B. R., Ladle, R. J., Lourie, S. A., & Whittaker, R. J. (2011). Basic Biogeography: Estimating Biodiversity and Mapping Nature. In Conservation Biogeography. https://doi.org/10.1002/9781444390001.ch4

Sangster, G. (2009). Increasing numbers of bird species result from taxonomic progress, not taxonomic inflation. Proceedings of the Royal Society B: Biological Sciences, 276(1670), 3185–3191.

Stropp, J., Ladle, R. J., Emilio, T., Lessa, T., & Hortal, J. (2022). Taxonomic uncertainty and the challenge of estimating global species richness. Journal of Biogeography, 49(9), 1654–1656. https://doi.org/10.1111/jbi.14463

Stropp, J., Umbelino, B., Correia, R. A., Campos-Silva, J. V, Ladle, R. J., & Malhado, A. C. M. (2020). The ghosts of forests past and future: deforestation and botanical sampling in the Brazilian Amazon. Ecography, 43(7), 979–989. https://doi.org/https://doi.org/10.1111/ecog.05026

Tessarolo, G., Ladle, R., Lobo, J. M., Rangel, T. F., & Hortal, J. (2021). Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. Ecography, 44(12), 1743–1755. https://doi.org/10.1111/ecog.05793

Thomson, S. A., Pyle, R. L., Ahyong, S. T., Alonso-Zarazaga, M., Ammirati, J., Araya, J. F., Ascher, J. S., Audisio, T. L., Azevedo-Santos, V. M., Bailly, N., Baker, W. J., Balke, M., Barclay, M. V. L., Barrett, R. L., Benine, R. C., Bickerstaff, J. R. M., Bouchard, P., Bour, R., Bourgoin, T., … Zhou, H.-Z. (2018). Taxonomy based on science is necessary for global conservation. PLOS Biology, 16(3), e2005075. https://doi.org/10.1371/journal.pbio.2005075

Thomson, S. A., Thiele, K., Conix, S., Christidis, L., Costello, M. J., Hobern, D., Nikolaeva, S., Pyle, R. L., van Dijk, P. P., & Weaver, H. (2021). Towards a global list of accepted species II. Consequences of inadequate taxonomic list governance. Organisms Diversity & Evolution, 21(4), 623–630.

Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J. M., Yang, J., Sayol, F., Neate-Clegg, M. H. C., Alioravainen, N., Weeks, T. L., Barber, R. A., Thomas, G. H., & Schleuning, M. (2022). AVONET: morphological, ecological and geographical data for all birds. Ecology Letters, 25(3), 581–597. https://doi.org/10.1111/ele.13898

Wägele, H., Klussmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A., & Wägele, J. W. (2011). The taxonomist - an endangered race. A practical proposal for its survival. Frontiers in Zoology, 8(1), 25. https://doi.org/10.1186/1742-9994-8-25

Wheeler, Q. (2020). A taxonomic renaissance in three acts. Megataxa, 1(1), 4–8.

Winker, K., & Haig, S. M. (2010). Avian subspecies. In American Ornithologists' Union (Ed.), Ornithological Monographs (67th ed.).

Winker, K., & Rasmussen, P. C. (2021). Species limits and taxonomy in birds. Ornithology, 138(2). https://doi.org/10.1093/ornithology/ukab017

Yarwood, M. R., Weston, M. A., & Symonds, M. R. E. (2019). Biological determinants of research effort on Australian birds: a comparative analysis. Emu - Austral Ornithology, 119(1), 38–44. https://doi.org/10.1080/01584197.2018.1501274

Zachos, F. E. (2016). Species Concepts in Biology. Springer International Publishing. https://doi.org/10.1007/978-3-319-44966-1

Zachos, F. E. (2018). (New) Species concepts, species delimitation and the inherent limitations of taxonomy. Journal of Genetics, 97(4), 811–815.

Capítulo 3
Revista: Ecological Indicators
Status: Publicado (doi.org/10.1016/j.ecolind.2023.111490)

Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa

Thainá Lessa[a,b*,§], Fernanda Alves-Martins[b,c,§], Javier Martinez-Arribas[b,c], Ricardo A. Correia[d,e,f], John Mendelsohn[g], Ezequiel Chimbioputo Fabiano[h], Simon T. Angombe[i], Ana C.M. Malhado[a,b,c], Richard J. Ladle[a,b,c]

[a]Institute of Biological and Health Sciences, Federal University of Alagoas, 57072-970 Maceió, Alagoas, Brazil.
[b]CIBIO-InBIO, Research Centre in Biodiversity and Genetic Resources, University of Porto, Campus de Vairão, 4485-661 Vairão, Portugal.
[c]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[d]Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland.
[e]Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, 00014 Helsinki, Finland.
[f]DBIO & CESAM, Centre for Environmental and Marine Studies, University of Aveiro, 3810-193 Aveiro, Portugal.
[g]Ongava Research Centre, Ombika, Namibia.
[h]Department of Wildlife Management & Tourism Studies, Katima Mulilo Campus, University of Namibia, Windhoek, Namibia.
[i]School of Agriculture and Fisheries Sciences, Neudamm Campus, University of Namibia, Windhoek, Namibia

*Corresponding author: Thainá Lessa and Fernanda Alves-Martins.
Emails: thainalessaps@gmail.com, ferfealvesmartins@gmail.com.
§ These authors contributed equally to this work.

Abstract

Effective conservation efforts and predictions of future impacts on biodiversity depend heavily on publicly available information about species distributions. However, data on species distributions is often patchy, especially in many countries of the Global South where resources for biological surveys have been historically limited. In this study, we use biodiversity ignorance scores to quantify and visualize gaps and biases in biodiversity data for Namibia, with a focus on five terrestrial taxa at a spatial scale of 10 x 10 km. We model the relationship between ignorance scores and socio-geographical variables using generalized additive models for location, scale and shape (GAMLSS). Our findings demonstrate that despite a high volume of occurrence records available on the Global Biodiversity Information Facility (GBIF), publicly available knowledge of Namibia's terrestrial biodiversity remains very limited, with large areas contributing few or no records for key taxa. The exception is birds that have benefitted from a massive influx of data from the citizen science platform eBird. Our study also highlights the importance of citizen science initiatives for biodiversity knowledge and reinforces the usefulness of ignorance scores as a simple intuitive indicator of the relative availability and distribution of species occurrence records. However, further research, biological surveys, and renewed efforts to make existing data held by museums and other institutions widely available are still necessary to enhance biodiversity data coverage in countries with patchy data.

Key-words: GBIF, occurrence records, survey effort, ignorance scores, species distributions, vertebrates, Southern Africa.

Introduction

The development of new surveying tools and national and international biodiversity information systems is making existing species records available to researchers worldwide via the internet (Hedrick et al., 2020). The most important of these endeavours is the Global Biodiversity Information Facility (GBIF) that was started in 2001 (Edwards, 2004). The GBIF takes advantage of long-term, coordinated and ongoing efforts to digitize specimens from world's natural history collections (Gaijy et al., 2013; Nelson & Ellis, 2019) and, more recently, from some citizen science databases (Chandler et al., 2017). By most measures the GBIF has been a remarkable success, and currently hosts over two and half billion species occurrence records from over two thousand institutions and open access data repositories (GBIF, 2023). However, even enormous databases such as the GBIF are incomplete and uncertain, with a considerable amount of biodiversity data subject to errors, often due to the low accuracy, low precision and lack of standardization from multiple data sources (Barve & Otegui, 2016; Cobos et al., 2018; D'Antraccoli et al., 2022; Ladle & Hortal, 2013).

Accepting that biases and gaps in biodiversity data often cannot be avoided, especially in the more unevenly sampled countries and regions of the world (Danovaro et al., 2010; Hopkins, 2019; Lessa et al., 2019), it becomes important to quantify and understand the limits of our biodiversity knowledge. There are several ways to evaluate biodiversity knowledge gaps and data quality. One recent proposal is through the creation of 'maps of biogeographical ignorance' (MoBIs) that distinguish data quality degree and intensively sampled areas from poorly sampled ones (Rocchini et al., 2011; Tessarolo et al., 2021). MoBIs are typically based on a combination of: i) completeness of species inventories for defined sampling units (e.g., Stropp et al., 2016); ii) estimates of taxonomic quality, and; iii) temporal and spatial decay in data (Ladle & Hortal, 2013; Tessarolo et al., 2021). Unfortunately, MOBIs are not appropriate to use in data-poor regions because they are very sensitive to low record numbers and to non-natural relative abundances of species in natural history collections (Meyer et al., 2016; Steege et al., 2011; Stropp et al., 2016); for example, the overrepresentation of rare species in museum collections relative to their true abundances (Gotelli et al., 2021).

A simpler alternative to MoBIs is to quantify and map the absence of data (i.e., ignorance) in biodiversity databases through the 'ignorance scores' approach (Ruete, 2015). These are useful to rapidly assess and visualize biases and shortfalls related with taxonomy, temporal and spatial data. Ignorance scores can also be used to characterize the degree of biodiversity knowledge based on the effort (or weakness of it) to record species occurrences (Correia et al., 2019; Mair & Ruete, 2016; Ruete, 2015). The approach has the added advantage of being simple to calculate, does not involve prediction or estimation of the total number of species in a given area, has a very limited number of assumptions, and relies solely on raw data of species occurrences, i.e., presence-only data (Correia et al., 2019). The score provides information on recording coverage and reliability that can be used to measure the spatial distribution of recording effort across a study area, and thus to identify undersampled and priorities areas for data collection (Mair & Ruete, 2016).

Multiple factors have been identified as potentially influencing recording effort. For example, reasons associated with how accessible and/or practical it is to sample a given area, such as road density, human population density, or proximity to universities (Meyer et al., 2015; Sastre & Lobo, 2009), and public and/or scientific interest are known to positively influence the site selection for recording biodiversity (Millar et al., 2019; Oliveira et al., 2016). Collectors (biodiversity researchers and citizen scientists) often prefer to sample sites perceived as being poorly studied, ecologically unique, more diverse or well-preserved, such as formally protected areas (PA) or sites with pristine native vegetation (Boakes et al., 2010; Rocha-Ortega et al., 2021; Yang et al., 2014). Additionally, collectors frequently prefer to sample areas near research centres (Ribeiro et al., 2016; Carvalho et al., 2023), which are typically located in economically more developed regions (Meyer, 2016). Species distribution data thus tend to vary more among political than ecological units, reflecting historical patterns of collecting, collating and digitalizing biogeographical data (Hortal et al., 2015; Stropp et al., 2016). In unevenly sampled regions this can lead to maps of species richness that closely resemble maps of survey effort (Hortal et al., 2015), a pattern that is particularly striking in sub-Saharan Africa (Stropp et al., 2016).

Namibia is a large, arid southwest African country with high levels of endemism and low human population density (Atlas of Namibia Team, 2022; Simmons et al., 1998). It has a strong system of protected areas, but Namibia's species occurrence records are very patchy on publicly available platforms such as GBIF (GBIF, 2023). These characteristics make the country an ideal political unit to evaluate the spatial patterns of biodiversity ignorance through ignorance scores (Lessa et al., 2019). Therefore, we applied the ignorance score approach to evaluate variation in species occurrence records available from GBIF across multiple terrestrial taxa for Namibia.

Specifically, we used species records collected from GBIF to: i) characterize temporal and taxonomic biases in recording efforts; ii) evaluate and map spatial shortfalls in recent recording effort, iii) analyse the influence of multiple socio-geographical variables on the distribution of recent recording effort, and iv) discuss the usability of ignorance score approach to evaluate quality in publicly available biodiversity data.

Methods

Study area

Namibia is a southwest African country with a terrestrial area of approximately 824,000 km2 (Figure 1). Its geomorphology is dominated by the great escarpment along the western side of the country, forming a transition between the narrow coastal desert and a flat inland plateau dominated by aeolian sand. Namibia is the most arid sub-Saharan country (Gargallo, 2020), with the Namib desert in the southwest of the country receiving an annual average precipitation of less than 50 mm. Moreover, rainfall is very variable, mostly falling over short, intense periods (Atlas of Namibia Team, 2022).

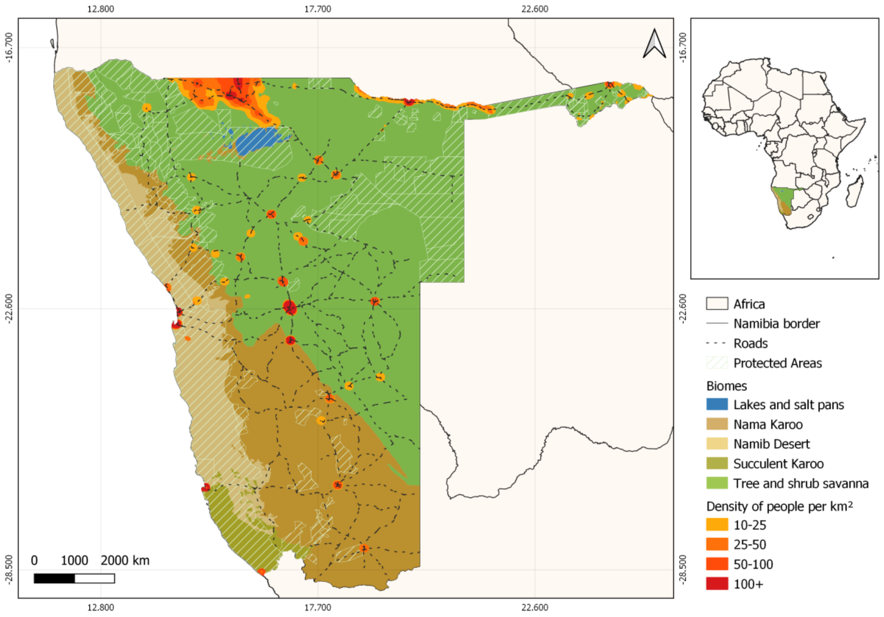


Figure 1: Map of Namibia highlighting socio-geographical variables used in our analysis, for example, roads, protected areas, vegetation cover and density of people.

There are few permanent rivers; the Kunene and Okavango Rivers form the northern border with Angola, the Kwando, Linyanti, Zambezi and Chobe Rivers form the borders with Botswana and Zambia, and the Orange River borders South Africa in the south. The vegetation of Namibia can be broadly classified in deserts (16% of the country), savannas (64%) and woodlands (20%) (Giess, 1971) with both summer and winter rainfall zones. Over 70% of the country is classified as arid or semi-arid (Simmons et al., 1998). The great variability in rainfall means that the amount of standing herbaceous vegetation varies considerably from year to year (Wardell-Johnson, 2000).

Species occurrence records and filtering processes

We collected all species occurrence records (hereafter just ‘records’) available for Namibia in GBIF (https://www.gbif.org/). We chose to collect data from GBIF because it has an international mandate to compile global species records and is one of the most commonly used sources of data for biodiversity studies globally. We first collected data from records of Namibia (1,656,016 records, GBIF, 2021). The following methods were used to exclude records: i) suspicious geographical coordinates - these are records with geographical coordinates assigned to the centroid of a municipality, state, country or falling in the ocean; ii) invalid, unlikely, mismatched or absent collection dates; iii) absent taxonomic identification at species level, or; iv) uncertain taxonomic data at species level - taxonomic data that does not match any known species or where matches can only be obtained through fuzzy matching. Records were collected with the ‘rgbif’ library for R software (Chamberlain & Boettiger, 2017). The precision of geographical coordinates was examined with the ‘CoordinateCleaner’ library for R software (Zizka et al., 2019). Given that GBIF includes datasets with varying coordinate precision, we accepted geographical coordinates with varying levels of precision, allowing for a certain degree of rounding. However, these coordinates were accepted only if they kept a precision greater than the spatial unit used in the study, which was the 10 km cell (Zizka et al., 2020). The validation of scientific names was performed with the ‘taxize’ library for R software (Chamberlain et al., 2020) and manually. The cleaning process returned 1,139,786 records. Then, we filtered only records on five reference taxonomic groups that would be evaluated (as described below), which resulted in a dataset with 1,075,916 records collected from 1783 to 2021 (full dataset).

We arranged records into subcategories of reference taxonomic groups, which refer to a group of species that can be studied or collected across similar methodologies. All species belonging to a reference taxonomic group probably share the record bias analogously. In these cases it is standard practice to use occurrence counts of species from that taxonomic group as a substitute for its recording effort (Phillips et al., 2009;

Ponder et al., 2001). The assumption underlying this statement is that the absence of records for any species from a reference taxonomic group (e.g. mammals) in a particular area is likely due to a lack of a specialist, rather than the total absence of the reference group in the area. Similarly, if there are many records of a reference taxonomic group in a given place, it is likely that the lack of records of a particular species in that place is due to true absence (Phillips et al., 2009; Ruete, 2015). We considered five taxonomic groups for calculating ignorance scores: birds (Aves), mammals (Mammalia), amphibians (Amphibia), reptiles (Reptilia) and insects (Insecta).

After the data cleaning process, we carried out a temporal filtering process keeping only records collected from 2000 onwards for making ignorance scores and maps, and spatial analysis. The chosen time window reduces the probability of changes in collection behaviours, thereby minimizing recording biases (Ponder et al., 2001; Ruete, 2015) and ensuring that the period of time the records were collected is congruent with the socio-geographic variables used to explore recording biases (details in section 2.4). The temporal filtering dataset (from 2000-2021; hereafter, recent dataset) returned 1,010,800 records which were analysed to create ignorance scores and maps (section 2.3) and explore factors influencing the spatial distribution of records (section 2.4).

Ignorance scores and maps

We calculated ignorance scores for each reference taxonomic group over the entire Namibian territory by defining recording units (SUs) of 10 x 10 km, using the recent dataset (data from 2000-2021). For this, we generated a raster grid using an equal-area (Eckert IV) projection, which returns 8,567 grid cells. We chose square grid because it is the most commonly used polygon shape in spatial analysis by ecologists, it is simple for calculations, transformations and comparisons, and is frequently used in Geographic Information Systems rasters (Birch et al., 2007). The 10 x 10 km spatial resolution was chosen because it was considered an adequate size to be reasonably sampled during a recording visit (Correia et al., 2019). We then converted the grid to WGS84 to match species records projection. Ignorance scores were calculated using the 'Log-Normalization approach' suggested by Ruete (2015) – ignorance is equal to one minus the normalization of the natural logarithm of the data – and defined by the following equation: Ignorance score = $1 - (\ln(Ni + 1)/\ln(Nm+1))$.

Where Ni is the number of records in a grid cell i and Nm, the maximum number of records in the cell with the highest number of records. For example, in the case of Birds the highest number of records (N = 29,914) was found in a cell near Windhoek. Therefore, the 'Log-Normalization approach' considered the maximum value of 29,914 when calculating the ignorance score for birds. The ‘Log-Normalization approach’ transforms records counts into a 0-1 scale of ignorance, with a score of one indicating complete ignorance, i.e., no single record available for the cell, and a score of 0 indicating the best available knowledge, i.e., the maximum number of records (Nm). This approach is the most suitable when there are large differences in the minimum and maximum number of records for a given reference taxonomic group, which is our case (Birds = 1-29,914; Mammals = 1-594; Reptiles = 1-79; Amphibians = 1-21; Insects = 1-540) and allows comparisons among the distinct reference taxonomic groups.

Environmental and socio-geographical drivers of species recording effort

We collected data on five socio-geographical variables that may drive the spatial distribution of recording effort based on perceptions of site accessibility or biological value: 1) road density; 2) human population density; 3) minimum distance to a university; 4) minimum distance to a protected area; and 5) average vegetation cover.

Road density was estimated as the total length of roads (in km) in an area of 100 km2 (10 x 10 km grid cells) covering the Namibian territory based on data from the OpenStreetMap database (see Correia et al. (2019) for a similar approach). Human population estimates for the years 2000, 2005, 2010, 2015, and 2020 were obtained at 1 km resolution from the Center for International Earth Science Information Network – CIESIN – Columbia University (2018), and aggregated for the grid cells resolution (10 x 10 km) by summing cells’ values, so that population density refers to the total count of people in cells of 100 km2. Vegetation cover at 2000 was obtained at 30 meters’ resolution from (Hansen et al., 2013) and aggregated for the grid cells resolution by the mean value of cells. Minimum distance to universities was calculated for each grid cell based on the location of universities and other higher education institutions (e.g., colleges) inside the country. The geographical location of each higher education institution was obtained from OpenStreetMap database. Grid cells containing at least one higher education institution were assigned a distance of zero. For grid cells without any higher education institution, the distance of the cell centroid to the nearest higher education institution was estimated. Minimum distance to Protected Areas (PAs) was calculated for each cell based on the location of PAs in the region. Maps of PAs were obtained from the World Database on Protected Areas on November 2021 and include national parks, private reserves, communal conservancies among other categories of protected areas (available from https://www.protectedplanet.net). Cells covered by protected areas were assigned a distance of zero, otherwise the distance from the cell centroid to the nearest boundary of a protected area was calculated.

Spatial analyses were performed on QGIS 3.20. We used Spearman's correlation to assess pairwise correlation among variables and observed a weak correlation (rs < 0.3 for all variable pairs), except for population density and vegetation cover, which exhibited a correlation coefficient of 0.53. The exclusion of population density during the model selection process (see below and Supplementary Material 1) mitigates any multicollinearity issues arising from this.

Data analysis

Initially we used the full dataset (cleaned records from 1783-2021) to characterize temporal and taxonomic biases in recording efforts. To do this, we created bar and spider graphs using R software. Afterwards, we used the recent dataset (cleaned records from 2000-2021) to create ignorance scores and ignorance maps for the five reference taxonomic groups, and to perform statistical analysis. When exploring ignorance scores and maps, we found a large proportion of grid cells that had ignorance scores of 1 (i.e., without any record). Based on this evaluation, we used Generalized Additive Models for Location, Scale and Shape (GAMLSS) (Rigby & Stasinopoulos, 2005) to explore the relationship between ignorance scores and the multiple environmental and socio-geographical variables outlined for Namibia.

GAMLSS was chosen because our response variable, ignorance scores, follow a one-inflated beta distribution ("BEINF1"; Ospina & Ferrari, 2010), with values ranging between 0 and 1 (0<Ignorance score<=1) and containing a large proportion of ignorance scores of 1. This distribution is suitable when there is an excess of ones in the data compared to what would be expected from a standard beta distribution, and cannot be modelled using a classical Generalized Linear Model (GLM) approach. In addition to allowing the use of a wide range of statistical distributions, GAMLSS can deal with heterogeneous, highly skewed and kurtotic data, such as the left-skewed distribution of the ignorance scores. We converted ignorance scores equal to zero to 10e-06, as the log-Normalization approach returns ignorance scores equal to zero for cells with the maximum number of records, and the one-inflated beta distribution only accepts values greater than zero.

GAMLSS models assume that the response variable is described by a density function defined by up to 4 parameters (μ, σ, ν, τ) that determine its location μ (i.e., mean), scale σ (i.e., standard deviation) and shape (i.e., skewness ν and kurtosis τ) (Stasinopoulos & Rigby, 2007). We examined the relationship between ignorance scores and socio-geographical factors by assessing how these factors affect the location (i.e., the mean), skewness and kurtosis (i.e., the shape of the relationship). To capture non-linear relationships, we applied a smoothing function (P-splines). Finally, we used a model selection approach based on Generalized Akaike Information Criterion (GAIC) scores to select the most informative socio-geographical variables for each reference taxa model. GAIC is an extension of AIC (Akaike Information Criterion), which takes into account the additional complexity of GAMLSS models, which have more parameters than traditional GLM models, and therefore include a higher penalty for the number of parameters in the model. In general, the smaller the GAIC value, the better the model fit (Stasinopoulos et al., 2017).

We ran GAMLSS models for the reference taxonomic groups. GAMLSS models were calculated independently. All model results, including the relative explanatory power of each model, are reported in Supplementary Material 2. Statistical analyses were carried out in R statistical software 4.2.0 (R Team Core, 2017) using the ‘gamlss’ package (Rigby & Stasinopoulos, 2005). Models were implemented with the ‘gamlss’ function and pseudo R-squared values for each model were obtained with function ‘Rsq’ using option ‘Cragg Uhler’ (Stasinopoulos & Rigby, 2007).

Results

Based on records incorporated into GBIF, some clear temporal biases were observed in recording effort over the nearly 240 years (the full temporal window - 1783-2021) of recording biodiversity in Namibia. Specifically, there are very few records available before the 1990s, representing only 3.2% of all data. The highest peaks of records occurred from the 1990s onwards. The first peak occurred between 1993-2007 holding 14.8% of the full dataset, and the second peak between 2008-2019 with the highest number of records (71.7%) - five times more recording effort than the initial period. The year with the highest volume of records was 2019 (Figure 2, grey line).

The temporal biases in recording effort for each taxonomic group followed a similar pattern to the overall dataset for birds (Figure 2, blue line). For mammals, the largest influx of records into GBIF was after 2000s, however a significant influx of records was noted in 1970s (red line). For reptiles and amphibians, the greatest recording efforts were made between the 1970s and 1990s, with few records collected and/or available after the 2000s (orange and purple line, respectively). Finally, the pattern of data influx for insects was more uniform when compared to the other taxonomic groups, showing peaks in the number of records collected in the 1920s, 1970s and 2000s (green line).

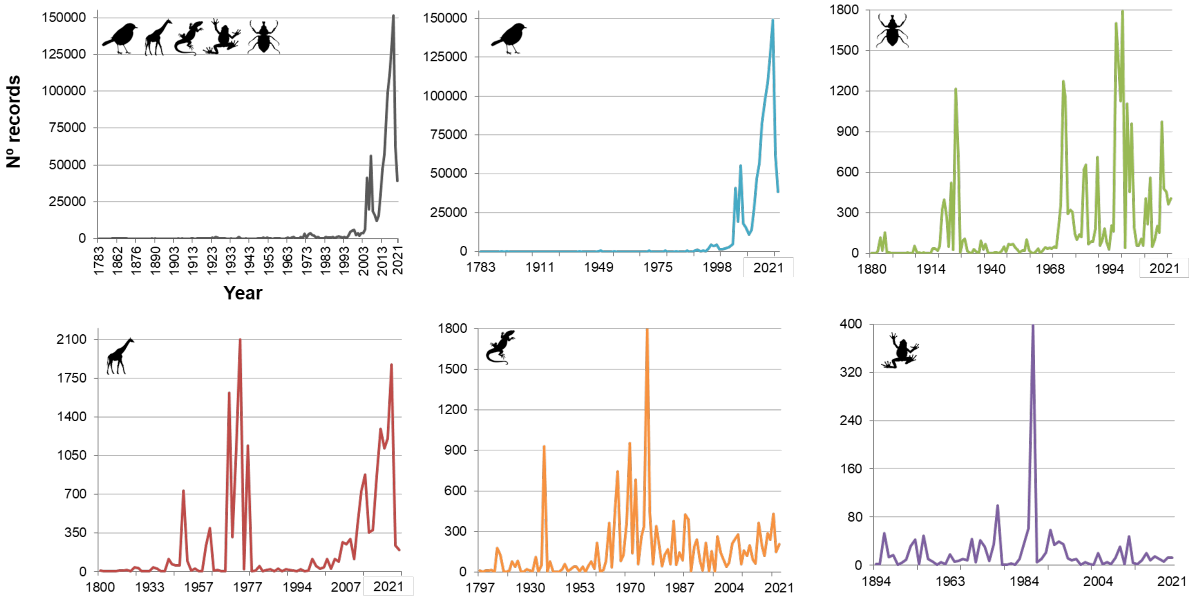


Figure 2: Historical progression of the number of occurrence records for Namibia's biodiversity publicly available on GBIF platform (full dataset). Number of records of all taxa (grey line) and separately, according to the fauna silhouette.

Despite the relatively large volume of data on Namibia’s biodiversity available in the GBIF from 1783-2021, our analysis still revealed strong biases in terms of taxonomic groups and in the characteristics of the records. Approximately 94% (n = 1,011,197) of records in the full dataset refers to birds, and 99.6% of these records were from human observations rather than specimens. Birds had the lowest rate (2.5%) of data loss after the temporal filtering process (2000-2021; recent dataset) (Figure 3). The second most representative reference taxonomic group was insects, with 26,707 records in the full dataset, though 86.6% of these records came from preserved specimens rather than observations. However, 65.4% of records in the full dataset were lost after the temporal filtering process (2000-2021). The records of mammals represented only 1.8%

of the full database, and after the year 2000 there was a decrease of 44.5% in the number of records. About 55% of mammals' records came from observations. The most critical shortfall was in Herpetofauna records. With 1,540 records, amphibians showed the lowest number of records in the full dataset (0.14%), and the highest rate (85%) of record loss after 2000. Furthermore the low number of records of amphibians implies that most species in our dataset are represented by only one or few records (i.e., singletons, doubletons, etc.). Reptiles showed the second lowest number of records in the full dataset (16,916 records) and 76% was lost after 2000s. Over 85% of herpetofauna records are based on preserved specimens (Figure 3).

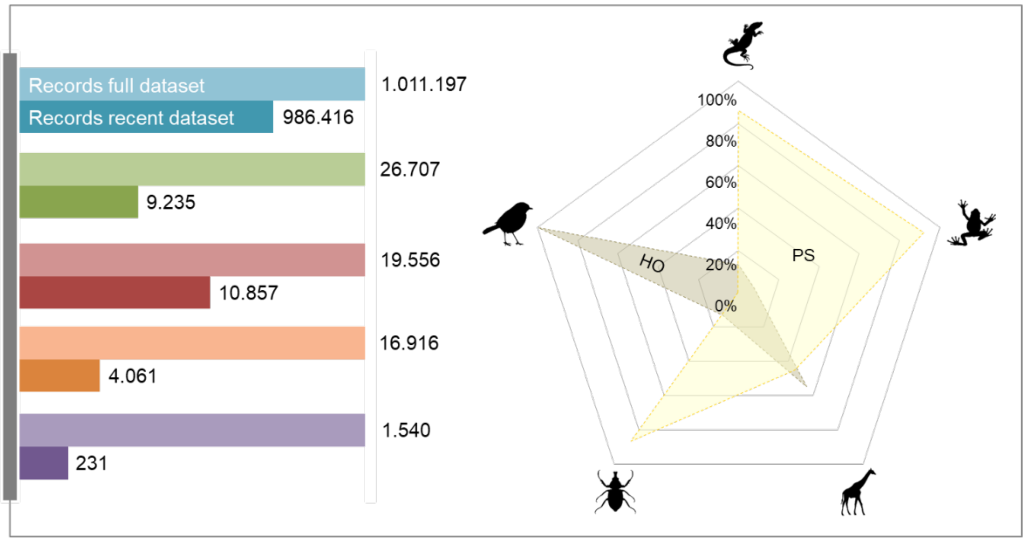


Figure 3: Bar graphic indicates the number of records in full dataset (1783-2021) and recent dataset (2000-2021). Radar chart illustrates the percentages of GBIF's basis of records using full dataset, with HO=Human Observation and PS=Preserved Specimen.

Notable gaps and biases for all reference taxonomic groups were observed in spatial distribution of ignorance scores in Namibia. A temporal decline in ignorance scores was noted when accumulating records (see temporal figure in Supplementary material 3). In more populated areas, such as the capital Windhoek, the coastal cities Swakopmund and Walvis Bay, the central cities Otjiwarongo and Okahandja, Rundu (northeast) and Keetmanshoop (southern), were characterized by high recording effort, and consequently, low ignorance scores. The coastal zone, where the Namib Desert is located, had low ignorance scores for mammals, birds and insects. The Succulent Karoo region showed low ignorance only for birds and insects. For almost all taxonomic groups (except birds), the eastern portion of Namibia – a scrub savanna region of aeolian sands bordering Botswana and South Africa – showed lower recording efforts and thus, high rates of ignorance scores (Figure 4).

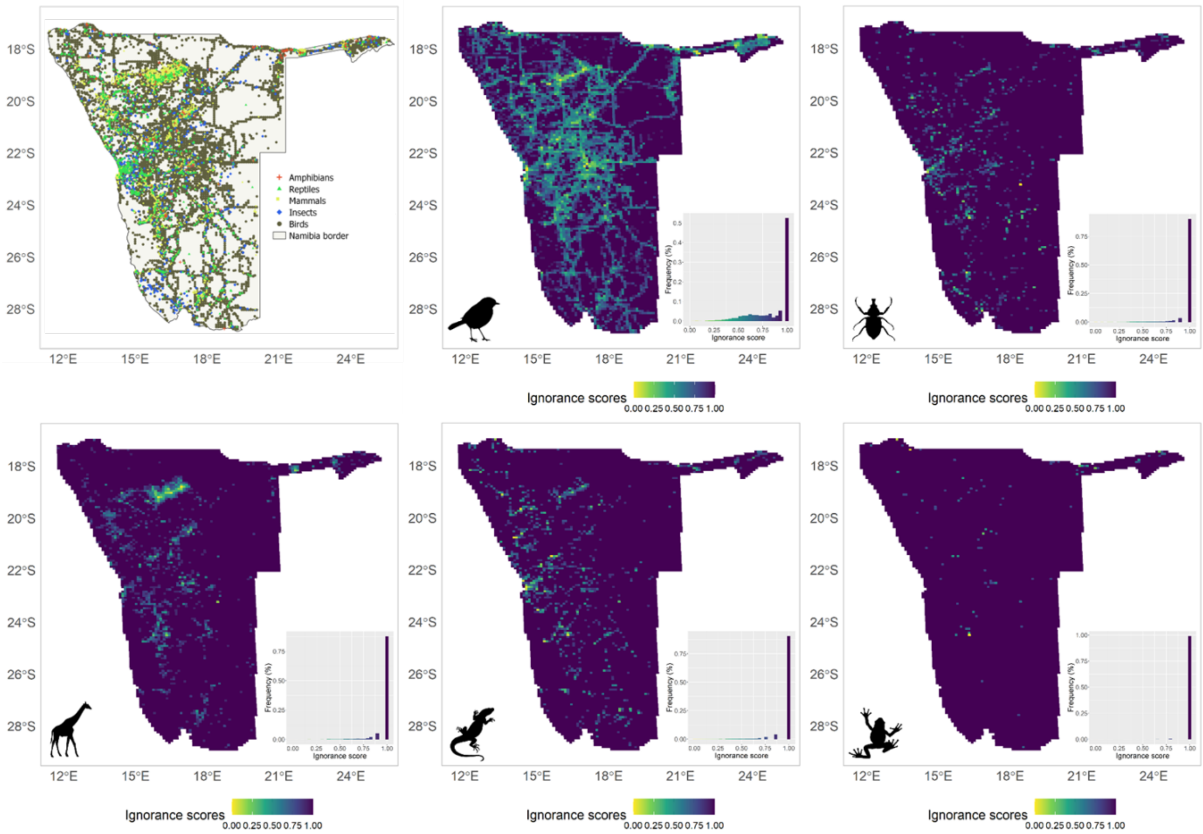


Figure 4: Spatial distribution of GBIF's records and ignorance scores for Namibia's birds, mammals, reptiles, amphibians and insects. Maps were calculated from recent dataset (2000-2021). Ignorance maps represent a gradient of ignorance scores - from cells with high ignorance scores (purple tons) to cells with low ignorance scores (yellow tons). Histograms represent the frequency of cells according to ignorance scores gradient. Silhouettes refer to taxonomic groups.

Analysing only the recent dataset (from 2000-2021), large areas of the country were underrepresented, even for birds the group with the highest number of records in GBIF (Figure 4). The ignorance maps for birds revealed that 52.4% of grid cells had no single record (ignorance score = 1), 17.3% had between 1 to 10 records (ignorance score = 0.93–0.76) and 5.7% had between 50 to 100 records (ignorance score = 0.61–0.55). Conversely, a small portion of the country was found to be overrepresented, i.e., 0.17% of 10 x 10 km2 grid cells had between 30,419 to 10,405 records (ignorance score = 0–0.1) and 1.78% of 10 x 10 km2 grid cells had between 9,185 to 1,003 records (ignorance score = 0.11–0.33) (yellowest parts of the map in Figure 4). Spatial biases in the distribution of records were more pronounced in the other reference taxonomic groups. The percentage of grid cells without any records (ignorance score = 1) was 99% to amphibians, 90.5% to reptiles, 90.2% to insects and 87.6% to mammals. In contrast, the percentage of grid cells that had ignorance score ≤ 0.5, i.e., where there is a greater recording effort, ranged between 0.1% (amphibians), 1% (insects), 1.1 % (mammals) and 1.2% (reptiles) (Figure 4).

Unsurprisingly the socio-geographical variable that best drives recording effort for all taxonomic groups analysed was road density, with an overall bias towards recording specimens in more accessible areas (higher road density, lower ignorance scores; Table 1). The distance to universities was also a significant predictor for bird and amphibian records. Ignorance scores exhibited an increase as the distance from universities increased, suggesting a tendency to record species in close proximity of these institutions. Maps of ignorance show a tendency for mammal, reptile and amphibian records to be collected in protected areas, such as national parks Etosha, Bwabwata, Waterberg Plateau, Skeleton Coast, Tsau and Namib-Naukluft. The eastern portion of Namibia (scrub savanna region), which had higher ignorance scores, is not covered by protected areas. Our statistical analyses validated these observations, demonstrating that the distance from protected areas is a significant driver of recording effort for these taxa. Ignorance scores increase with distance from PAs, indicating a decrease in recording efforts for sites located far from protected areas. Finally, the percentage of vegetation cover showed effect on recording effort only for birds, the ignorance scores for this taxon were lower in areas with less vegetation coverage, since much of Namibia's vegetation is composed of savannah, dry woodlands and desert (Table 1).

Table 1: Significant results of GAMLSS models exploring the association between ignorance scores and environmental and socio-geographical factors for the five reference taxonomic groups in Namibia ($p < 0.05$). The complete results, including non-significant associations, are available in the Supplementary Material 2.

Variable	**Reference taxonomic group**	**Coefficient estimate**	**T value**	***P***
Road density	Amphibians	-0.666	-4.767	0.000
	Birds	-1.030	-23.582	0.000
	Insects	-0.356	-5.073	0.000
	Mammals	-0.518	-8.240	0.000
	Reptiles	-0.448	-5.592	0.000
University distance	Amphibians	-0.002	-2.539	0.011
	Birds	0.001	9.156	0.000
Protected Area Distance	Amphibians	0.005	2.279	0.023
	Mammals	0.005	6.257	0.000
	Reptiles	0.004	4.408	0.000
Forest Cover	Birds	-0.041	-7.471	0.000

Discussion

Our main finding is that the volume of publicly (and thus widely) available digital information about Namibia's mainland fauna on the GBIF is still very low for most taxa and regions. Except for birds, all reference taxonomic groups evaluated here have significant temporal and spatial data shortfalls, and most of the records that have so far been added to GBIF were collected before 2000's and are therefore subject to higher rates of data degradation (Tessarolo et al., 2017). Interestingly, however, there were peaks in collection and availability of records during the period when Namibia was engaged in armed conflict for its independence (1960-1980). Although the amount of GBIF records appears to be increasing rapidly from 1990 onwards, when Namibia finally became independent, it is important to note that much of this increase is being driven by the recent influx of information on birds from the eBird citizen science platform (Bonney, 2021). eBird is also likely to be the main driver behind the decline in GBIF records during the 2020 COVID-19 pandemic when international travel was restricted and visits to national parks around the world fell precipitously (Hockings et al., 2020; Souza et al., 2021).

That Namibia has a low number and coverage of biological records is perhaps unsurprising given that it is the driest country in Sub-Saharan Africa (Simmons et al., 1998) with all of the associated challenges that this poses for biological surveying and collecting in arid, inhospitable environments with limited accessibility (Boakes et al., 2010; Lessa et al., 2019). In a recent assessment of insects' (Lepidoptera, Sphingidae) inventory completeness in Sub-Saharan Africa, a large proportion of Namibia had between 1-50 records in 200 x 200 km grid cells, and 5-30% of these were complete (Ballesteros-Mejia et al., 2013). For plants, Namibia showed a low proportion of well-sampled areas and much of the data was missing, incipient and outdated (Stropp et al., 2016). The persistence of more obsolete data compromises our understanding of the true composition of biodiversity, making conservation actions potentially inaccurate and inefficient (Escribano et al., 2016). Nevertheless, Namibia has a considerable need for publicly available high quality biodiversity information compared to other arid and dry regions – such as its neighbouring South Africa, which holds 30 times more records in GBIF, including more recent and complete data (Stropp et al., 2016) – given the enormous interest and economic importance of its wildlife industry (Schalkwyk et al., 2010).

We found that road density, a proxy of accessibility, was most strongly related to recording effort for all modelled taxonomic groups. This is a long-recognized bias for records, both historical and contemporary, and is often referred to as ‘roadside’ bias or the ‘roadside effect’ (Oliveira, et al., 2016; Petersen et al., 2021). A similar pattern was observed for the location and density of passerine birds and hawkmoths records in sub-Saharan Africa, which had a higher effort in more accessible locations: close to roads, railway lines, airports, rivers and cities (Reddy & Davalos, 2003; Ballesteros-Mejia et al., 2013). The probable mechanism behind this bias is that observations are more frequently made at short distances from roads and paths due to easier accessibility and convenience for collectors/surveyors (Kadmon et al., 2004; Petersen et al., 2021; Sastre & Lobo, 2009). This may be especially true in more inhospitable environments. Some researchers have also observed a trend of more records in densely populated areas (Luck, 2007), presumably for similar reasons. As Petersen et al., (2021) point out, the major concern over this particular bias is that areas close to roadsides may not be representative of the wider landscape, potentially leading to incorrect inferences about biodiversity patterns (but see Revermann et al. 2017).

Perhaps our most surprising result was the lack of influence of population density on ignorance scores of reference taxonomic groups as this variable was excluded by the Generalized Akaike Information Criterion (GAIC). This finding can be explained by Namibia’s very low population density. With over 2.6 million people (The World Bank, 2022), Namibia is one of the most sparsely populated countries in Africa (and the world). It has an average density of 2.5 persons per square kilometer (Wart et al., 2015), except for urban centres such as Windhoek, Rundu, Walvis Bay and Swakopmund, and certain densely populated rural areas in the central north and north-eastern areas of the country (Figure 1). Our findings indicated that 99% of Namibia territory has no amphibian records, 90.5% no reptiles, 90.2% no insects and 87.6% no mammals. In a previous study in a similarly arid region, the variables human population density and road density were spatially correlated (Oliveira et al., 2016; Correia et al., 2019). However, although the area in question – Brazilian Caatinga – has a geographic size similar to Namibia, it has almost ten times more inhabitants.

Our model also showed a relationship between recording effort and distance to universities for birds and amphibians. Again, this can be interpreted as a form of ‘convenience bias’; a high proportion of individuals contributing data to the GBIF are from the university sector (Correia et al., 2019; Liu et al., 2021) and, ceteris paribus, they will be more likely to collect records from sites close to their place of work than more distant sites. This behaviour is likely to have several underlying causes, including the practical and financial burden of mounting research expeditions to more remote areas, the increased likelihood of field stations and other research infrastructure closer to the university, and the added scientific value of working on a site that has already been partially or fully documented (dos Santos et al., 2015).

We found a clear tendency for records of all reference taxonomic groups to be collected in protected areas, such as the National Parks of Etosha, Bwabwata, Waterberg Plateau, Skeleton Coast, Tsau and Namib-Naukluft. Approximately 40% of Namibia territory has some degree of protection (Corrigan et al., 2018). In an extremely fragmented world, Namibia bucks the trend, connecting and protecting its areas in terms of ecological and economic values through ecotourism. This observed tendency is not surprising since we would expect both academics and amateur naturalists to take advantage of the superior infrastructure and accessibility available in these areas. The positive impact of protected areas on GBIF records has been noted previously (Correia et al., 2019; Oliveira et al., 2016), including in Sub-Saharan Africa (Ballesteros-Mejia et al., 2013), where several studies have shown that research sites tend to cluster close to universities in areas with some form of protection (dos Santos et al., 2015; Lessa et al., 2019). Despite the clear tendency of records to be associated with protected areas, we found very low recording effort for all reference taxonomic groups in national parks in the eastern portion of the country. This region has the lowest protected areas coverage and should be prioritized in new field works and recording efforts, or if these records already exist they should be made publicly available. Although our model did not reveal a significant association between recording effort and bird records in protected areas, these areas are still important for bird conservation according to the Atlas of Namibia (2022). Increased research effort could lead to a higher number of records if researchers make their data available through digital platforms such as GBIF.

Citizen science initiatives have played a significant recent role in mobilizing biodiversity data to the GBIF. Half of all records shared via GBIF come from datasets with significant volunteer contributions (Chandler et al., 2017). This trend is particularly notable in the case of Namibian birds and mammals, where a large amount of records came from citizen science platforms such as eBird, Southern African Bird Atlas Project 2 (SABAP2), the South African Bird Ringing Unit (SAFRING), iNaturalist and Observation.org. These platforms are almost universally used by amateur and professional wildlife watchers and photographers, resulting in a remarkable increase of the information available in GBIF (Bonney, 2021). This effect is especially notable in countries like Namibia, due to the large influx of international ecotourists. Although for other reference taxonomic groups citizen science appears to be a less common source of biodiversity data in Namibia. Specifically, amphibians and reptiles present a high proportion of records based on specimens preserved in museum collections.

Ignorance scores as a tool for visualizing biodiversity data needs

The concept of Ignorance Scores was introduced by Ruete (2015) and subsequently applied by Correia et al., (2019) for the semi-arid Caatinga biome of northeast Brazil. In contrast to alternative approaches proposed to evaluate data quality and completeness, such as Inventory Completeness (Sousa-Baena et al., 2014; Stropp et al., 2016) and MoBIs (Hortal et al., 2022; Tessarolo et al., 2021), Ignorance Scores stand apart by their unique approach. By exclusively relying on the presence of data in their metrics, without considering observed or expected species richness, ignorance scores can be applied in areas with very few data. Specifically, key distinctions between these approaches can be delineated as follows: i) Ignorance Score: bases on raw data, does not require a minimum number of records, and there are no estimations of species richness; ii) Completeness: draws upon species accumulation curves, requires a minimum number of records and estimates the number of species; and iii) MoBIs: integrates several data sources (completeness, temporal and spatial decay, and taxonomic quality) and requires a minimum number of records.

The advantage of the Ignorance Score approach is that it provides "a simple and intuitive indicator of species recording effort, allowing the assessment of taxonomic and spatial biases present in the GBIF database" (Correia et al., 2019, p. 8). Furthermore, it is ideal for regions or countries where there are large areas with few or no records where it would be impossible to compute more sophisticated measures of recording completeness based on species accumulation curves (e.g. Sousa-Baena et al., 2014). The ignorance score approach is extremely flexible and can be easily computed and mapped at different spatial scales and can be used, as in the current case study, to provide a rapid visual indicator of areas in need of further recording effort (Correia et al., 2019; Ruete, 2015).

Our study clearly shows an urgent need of collection efforts and mobilization of existing biodiversity data in the eastern and southwest portion of Namibia, especially the savanna region bordering Botswana and South Africa. Moreover, ignorance scores could provide a simple way to quantify and visualise the impact of new expeditions on biodiversity knowledge, providing a useful tool for demonstrating the value of such enterprises. Indeed, it would be extremely interesting to annually re-evaluate ignorance scores to provide a continuously updated account of progress in biological surveying and data mobilization.

As demonstrated here and elsewhere (Ruete, 2015; Correia et al., 2019), ignorance scores are also highly sensitive to spatial biases, making them useful tools to identify socio-geographical factors influencing recording effort. Nevertheless, the ignorance score algorithm also has certain limitations, the most serious of which is that it could be considered overly simplistic for many forms of analysis since they are only calculated using the number of records available in a given region over the time period of analysis. This means that valuable information on, for example, the identities or characteristics (e.g. threat status) of the species recorded, or the distribution of records within the annual cycle are not considered (Meyer et al., 2016).

A particular limitation of the Log-Normalization algorithm is that the minimum ignorance score (i.e. 0), is relative to the maximum number of records for the reference taxonomic group (Ruete, 2015), which may still be low (for example, in our database the maximum number of records in a cell for amphibians is 22). So, rather than indicating complete biodiversity knowledge, an ignorance score of 0 should be interpreted as the "best available knowledge" in any cell for the study region. In the current study we attempted to counter the inherent simplicity of the algorithm by independently considering multiple taxa and by restricting our analysis to more recent records whose collection is likely to have been driven by similar socio-geographical factors. In this context, ignorance scores provide a robust metric for measuring the importance of data mobilization efforts on biodiversity knowledge, and we would strongly recommend their use to quantify and visualize the impact of such initiatives.

Finally, it is important to highlight that the publicly available records on GBIF for Namibia only represent a fraction of the documented biodiversity in the country. Other types of institutions, such as museums, herbaria and other research centres in Namibia and elsewhere harbour biodiversity data. For example, about 1.2 million bird records were assembled during the first southern African bird atlas project (SABAP1) which was pre-2000 and are not currently included in GBIF (J. Mendelsohn, pers. comm.). In addition to the SABAP1 and SABAP2 projects (the latter already included in GBIF), which have numerous records of birds, there are many other African projects that use the efforts of citizen scientists to carry out inventories of taxonomic groups of vertebrates, invertebrates, plants and even fungi, as is the case of the Virtual Museum (Biodiversity and Development Institute, 2023).

The impediments to accessing, collecting and evaluating African biodiversity data have been previously acknowledged and reported by decision-makers represented by government, civil society and UN agencies, who have recommended a strengthening of national and international collaboration to ensure availability and usability of information and achieve conservation goals (Han et al, 2014; Stephenson et al, 2017). Notwithstanding the significance of such engagement to collect and document data, it is also imperative that these records be incorporated into online platforms and thus made available for both scientists and decision makers.

Acknowledgements

TL was funded by the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" - Brazil (CAPES) - Finance Code 001. RAC acknowledges funding from the Kone Foundation (#202101976) and the Academy of Finland (#348352). FA-M, JM-A, ACMM and RJL is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 854248.

References

Atlas of Namibia Team. (2022). Atlas of Namibia: Its land, water and life. Namibia Nature Foundation.

Ballesteros-Mejia, L., Kitching, I. J., Jetz, W., Nagel, P., & Beck, J. (2013). Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. Global Ecology and Biogeography, 22(5), 586–595. https://doi.org/10.1111/geb.12039

Barve, V., & Otegui, J. (2016). bdvis: visualizing biodiversity data in R. Bioinformatics, 32(19), 3049–3050. https://doi.org/10.1093/bioinformatics/btw333

Biodiversity and Development Institute. (2023). The Virtual Museum. https://vmus.adu.org.za

Birch, C. P. D., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modelling, 206(3–4), 347–359. https://doi.org/10.1016/j.ecolmodel.2007.03.041

Blackie, R., Baldauf, C., Gautier, D., Gumbo, D., Kassa, H., Parthasarathy, N., Paumgarten, F., Sola, P., Pulla, S., Waeber, P., & Sunderland, T. (2014). Tropical dry forests: The state of global knowledge and recommendations. Center for International Forestry Research.

Boakes, E. H., Mcgowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., Connor, K. O., & Mace, G. M. (2010). Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology, 8(6). https://doi.org/10.1371/journal.pbio.1000385

Bonney, R. (2021). Expanding the Impact of Citizen Science. BioScience, 71(5), 448–451. https://doi.org/10.1093/biosci/biab041

Carvalho, R.L., Resende, A.F., Barlow, J., et al. (2023) Pervasive gaps in Amazonian ecological research. Current Biology, S0960982223008631.

Chamberlain, S. A., & Boettiger, C. (2017). R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprint, 5, e3304v1. https://doi.org/https://doi.org/10.7287/peerj.preprints.3304v1

Chamberlain, S. A., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B. G., Marchand, P., Tran, V., Salmon, M., Li, G., & Grenié, M. (2020). taxize: Taxonomic information from around the web (R package version 0.9.98). https://github.com/ropensci/taxize

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. Biological Conservation, 213, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004

Cobos, M. E., Jiménez, L., Nuñez-Penichet, C., Romero-Alvarez, D., & Simoes, M. (2018). Sample data and training modules for cleaning biodiversity information. Biodiversity Informatics, 13, 49–50. https://doi.org/10.17161/bi.v13i0.7600

Correia, R. A., Ruete, A., Stropp, J., Malhado, A. C. M., dos Santos, J. W., Lessa, T., Alves, J. A., & Ladle, R. J. (2019). Using ignorance scores to explore biodiversity recording effort for multiple taxa in the Caatinga. Ecological Indicators, 106(June 2019), 105539. https://doi.org/10.1016/j.ecolind.2019.105539

Corrigan, C., Bingham, H., Shi, Y., Lewis, E., Chauvenet, A., & Kingston, N. (2018). Quantifying the contribution to biodiversity conservation of protected areas governed by indigenous peoples and local communities. Biological Conservation, 227, 403–412. https://doi.org/10.1016/j.biocon.2018.09.007

D'Antraccoli, M., Bedini, G., & Peruzzi, L. (2022). Maps of relative floristic ignorance and virtual floristic lists: An R package to incorporate uncertainty in mapping and analysing biodiversity data. Ecological Informatics, 67, 101512. https://doi.org/10.1016/j.ecoinf.2021.101512

Danovaro, R., Company, J. B., Corinaldesi, C., D'Onghia, G., Galil, B., Gambi, C., Gooday, A. J., Lampadariou, N., Luna, G. M., & Morigi, C. (2010). Deep-sea biodiversity in the Mediterranean Sea: the known, the unknown, and the unknowable. PloS One, 5(8), e11832.

dos Santos, J. G., Malhado, A. C. M., Ladle, R. J., Correia, R. A., & Costa, M. H. (2015). Geographic trends and information deficits in Amazonian conservation research. Biodiversity and Conservation, 24(11), 2853–2863. https://doi.org/10.1007/s10531-015-0981-x

Edwards, J. L. (2004). Research and Societal Benefits of the Global Biodiversity Information Facility. BioScience, 54(6), 485–486. https://doi.org/, https://doi.org/10.1641/0006-3568

Escribano, N., Ariño, A.H. and Galicia, D. (2016). Biodiversity data obsolescence and land uses changes. PeerJ, 4, p.e2743. doi:https://doi.org/10.7717/peerj.2743.

Gaiji S., Chavan, V., Ariño, A. H., Otegui, J., Hobern, D., Sood, R., & Robles, E. (2013). Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. Biodiversity Informatics, 8(2). https://doi.org/10.17161/bi.v8i2.4124

Gargallo, E. (2020). Community Conservation and Land Use in Namibia: Visions, Expectations and Realities. Journal of Southern African Studies, 46(1), 129–147. https://doi.org/10.1080/03057070.2020.1705617

GBIF. (2021). Namibia Occurrence Download. https://doi.org/https://doi.org/10.15468/dl.9uwfx9

GBIF. (2023). GBIF Home Page. https://www.gbif.org/grscicoll/institution/search

Giess, W. (1971). A preliminary vegetation map of South West Africa. Dinteria, 4, 5–14.

Gotelli, N. J., Booher, D. B., Urban, M. C., Ulrich, W., Suarez, A. V., Skelly, D. K., Russell, D. J., Rowe, R. J., Rothendler, M., Rios, N., Rehan, S. M., Ni, G., Moreau, C. S., Magurran, A. E., Jones, F. A. M., Graves, G. R., Fiera, C., Burkhardt, U., & Primack, R. B. (2021). Estimating species relative abundances from museum records. Methods in Ecology and Evolution, 14(2), 431–443. https://doi.org/10.1111/2041-210X.13705

Han, X., Smyth, R. L., Young, B. E., Brooks, T. M., Sánchez de Lozada, A., Bubb, P., Butchart, S. H. M., Larsen, F. W., Hamilton, H., Hansen, M. C., & Turner, W. R. (2014). A Biodiversity Indicators Dashboard: Addressing Challenges to Monitoring Progress towards the Aichi Biodiversity Targets Using Disaggregated Global Data. PLoS ONE, 9(11), e112046. https://doi.org/10.1371/journal.pone.0112046

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850–853. https://doi.org/10.1126/science.1244693

Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., Kennedy, J., Clarke, J. A., Cook, J. A., Blackburn, D. C., Edwards, S. V, & Davis, C. C. (2020). Digitization and the Future of Natural History Collections. BioScience, 70(3), 243–251. https://doi.org/10.1093/biosci/biz163

Hockings, M., Dudley, N., & Elliott, W. (2020). COVID-19 and protected and conserved areas. PARKS, 26.1, 7–24. https://doi.org/10.2305/IUCN.CH.2020.PARKS-26-1MH.en

Hopkins, M. J. G. (2019). Are we close to knowing the plant diversity of the Amazon? Anais Da Academia Brasileira de Ciências, 91.

Hortal, Joaquín, Bello, F. de, Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst., 46, 523–552. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hortal, Joaquín, Ladle, R. J., Stropp, J., & Tessarolo, G. (2022). Accounting for biogeographical ignorance within biodiversity modelling. Research Outreach, 129. https://doi.org/10.32907/RO-129-2502594871

Kadmon, R., Farber, O., & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications, 14(2), 401–413.

Klintenberg, P., & Seely, M. (2004). Land Degradation Monitoring in Namibia: A First Approximation. Environmental Monitoring and Assessment, 99(1–3), 5–21. https://doi.org/10.1007/s10661-004-3994-6

Ladle, R., & Hortal, J. (2013). Mapping species distributions: living with uncertainty. Frontiers of Biogeography, 5(1). https://doi.org/10.21425/F55112942

Lessa, T., Dos Santos, J. W., Correia, R. A., Ladle, R. J., & Malhado, A. C. M. (2019). Known unknowns: Filling the gaps in scientific knowledge production in the Caatinga. PLoS ONE, 14(7), 1–12. https://doi.org/10.1371/journal.pone.0219359

Liu, J., Zhao, Y., Si, X., Feng, G., Slik, F., & Zhang, J. (2021). University campuses as valuable resources for urban biodiversity research and conservation. Urban Forestry & Urban Greening, 64, 127255. https://doi.org/10.1016/j.ufug.2021.127255

Luck, G. W. (2007). A review of the relationships between human population density and biodiversity. Biological Reviews, 82(4), 607–645. https://doi.org/10.1111/j.1469-185X.2007.00028.x

Mair, L., & Ruete, A. (2016). Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. PLOS ONE, 11(1), e0147796. https://doi.org/10.1371/journal.pone.0147796

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. Nature Communications, 6(1), 8221. https://doi.org/10.1038/ncomms9221

Meyer, C. (2016). Limitations in global information on species occurrences. Frontiers of Biogeography, 8(2), e28195.

Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, 19(8), 992–1006. https://doi.org/10.1111/ele.12624

Millar, E. E., Hazell, E. C., & Melles, S. J. (2019). The ‘cottage effect’ in citizen science? Spatial bias in aquatic monitoring programs. International Journal of Geographical Information Science, 33(8), 1612–1632. https://doi.org/10.1080/13658816.2018.1423686

NSA (2023). Namibia Statistics Agency. https://nsa.nsa.org.na/

Nelson, G., & Ellis, S. (2019). The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1763), 20170391. https://doi.org/10.1098/rstb.2017.0391

Oliveira, U., Paglia, A. P., Brescovit, A. D., Carvalho, C. J. B. De, Silva, D. P., Rezende, D. T., Stehmann, R., Pena, P., & Remoto, C. D. S. (2016). The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Diversity and Distributions, 1–13. https://doi.org/10.1111/ddi.12489

Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. Statistical Papers, 51(1), 111–126. https://doi.org/10.1007/s00362-008-0125-4

Petersen, T. K., Speed, J. D. M., Grøtan, V., & Austrheim, G. (2021). Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. Ecological Solutions and Evidence, 2(1). https://doi.org/10.1002/2688-8319.12048

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications, 19(1), 181–197. https://doi.org/10.1890/07-2153.1

Ponder, W. F., Carter, G. A., Flemons, P., & Chapman, R. R. (2001). Evaluation of Museum Collection Data for Use in Biodiversity Assessment. Conservation Biology, 15(3), 648–657. https://doi.org/10.1046/j.1523-1739.2001.015003648.x

R Team Core. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Reddy, S., & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. Journal of Biogeography, 30(11), 1719–1727. https://doi.org/10.1046/j.1365-2699.2003.00946

Revermann, R., Wallenfang, J., Oldeland, J., & Finckh, M. (2017). Species richness and evenness respond to diverging land-use patterns - a cross-border study of dry tropical woodlands in southern Africa. African Journal of Ecology, 55(2), 152–161. https://doi.org/10.1111/aje.12333

Ribeiro, G.V.T., Teixido, A.L., Barbosa, N.P.U. & Silveira, F.A.O. (2016) Assessing bias and knowledge gaps on seed ecology research: implications for conservation agenda and policy. Ecological Applications, 26, 2033–2043.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Progress in Physical Geography, 35(2), 211–226. https://doi.org/10.1177/0309133311399491

Rocha-Ortega, M., Rodriguez, P., & Córdoba-Aguilar, A. (2021). Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. Ecological Entomology, 46(4), 718–728. https://doi.org/10.1111/een.13027

Ruete, A. (2015). Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. Biodiversity Data Journal, 3(1). https://doi.org/10.3897/BDJ.3.e5361

Sastre, P., & Lobo, J. M. (2009). Taxonomist survey biases and the unveiling of biodiversity patterns. Biological Conservation, 142(2), 462–467. https://doi.org/10.1016/j.biocon.2008.11.002

Schalkwyk, D. L. van, McMillin, K. W., Witthuhn, R. C., & Hoffman, L. C. (2010). The contribution of wildlife to sustainable natural resource utilization in Namibia: a review. Sustainability, 2(11), 3479–3499. https://doi.org/10.3390/su2113479

Simmons, R. E., Griffin, M., Griffin, R. E., Marais, E., & Kolberg, H. (1998). Endemism in Namibia: patterns, processes and predictions. Biodiversity and Conservation, 7(513–530). https://doi.org/https://doi.org/10.1023/A:1008879712736

Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014b). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions, 20(4), 369–381. https://doi.org/10.1111/ddi.12136

Souza, C. N., Rodrigues, A. C., Correia, R. A., Normande, I. C., Costa, H. C. M., Guedes-Santos, J., Malhado, A. C. M., Carvalho, A. R., & Ladle, R. J. (2021). No visit, no interest: How COVID-19 has affected public interest in world's national parks. Biological Conservation, 256, 109015. https://doi.org/10.1016/j.biocon.2021.109015

Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. Journal of Statistical Software, 23(7). https://doi.org/10.18637/jss.v023.i07

Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V., & Bastiani, F. De. (2017). Flexible Regression and Smoothing: Using GAMLSS in R (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b21973

Steege, H. t., Haripersaud, P. P., Banki, 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. https://doi.org/10.3732/ajb.1000215

Stephenson, P. J., Bowles-Newark, N., Regan, E., Stanwell-Smith, D., Diagana, M., Höft, R., Abarchi, H., Abrahamse, T., Akello, C., Allison, H., Banki, O., Batieno, B., Dieme, S., Domingos, A., Galt, R., Githaiga, C. W., Guindo, A. B., Hafashimana, D. L. N., Hirsch, T., & Hobern, D. (2017). Unblocking the flow of biodiversity data for decision-making in Africa. Biological Conservation, 213, 335–340. https://doi.org/10.1016/j.biocon.2016.09.003

Stropp, J., Ladle, R. J., M. Malhado, A. C., Hortal, J., Gaffuri, J., H. Temperley, W., Olav Skøien, J., & Mayaux, P. (2016a). Mapping ignorance: 300 years of collecting flowering plants in Africa. Global Ecology and Biogeography, 25(9), 1085–1096. https://doi.org/10.1111/geb.12468

Tessarolo, G., Ladle, R., Rangel, T., & Hortal, J. (2017). Temporal degradation of data limits biodiversity research. Ecology and Evolution, 7(17), 6863–6870.

Tessarolo, G., Ladle, R., Lobo, J. M., Rangel, T. F., & Hortal, J. (2021). Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. Ecography, 44(12), 1743–1755. https://doi.org/10.1111/ecog.05793

The World Bank. (2022). Population, total - Namibia. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=NA

Wardell-Johnson, G. (2000). Biodiversity and Conservation in Namibia into the 21st Century. In Population–Development–Environment in Namibia (pp. 17–47). International Institute for Applied Systems Analysis.

Wart, M. Van, Hondeghem, A., Schwella, E., & Suino, P. (2015). Leadership and Culture: Comparative Models of Top Civil Servant Training (M. Van Wart, A. Hondeghem, E. Schwella, & P. Suino (eds.)). Palgrave Macmillan UK. https://doi.org/10.1057/9781137454133

Yang, W., Ma, K., & Kreft, H. (2014). Environmental and socio-economic factors shaping the geography of floristic collections in China. Global Ecology and Biogeography, 23(11), 1284–1292. https://doi.org/10.1111/geb.12225

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods in Ecology and Evolution, 10(5), 744–751. https://doi.org/10.1111/2041-210X.13152

Zizka, A., Antunes Carvalho, F., Calvente, A., Rocio Baez-Lizarazo, M., Cabral, A., Coelho, J.F.R., Colli-Silva, M., Fantinati, M.R., Fernandes, M.F., Ferreira-Araújo, T., Gondim Lambert Moreira, F., Santos, N.M.C., Santos, T.A.B., Dos Santos-Costa, R.C., Serrano, F.C., Alves Da Silva, A.P., De Souza Soares, A., Cavalcante De Souza, P.G., Calisto Tomaz, E., Vale, V.F., Vieira, T.L. & Antonelli, A. (2020) No one-size-fits-all solution to clean GBIF. PeerJ, 8, e9916.

Conclusões gerais

Esta tese contribui para expandir a discussão sobre os principais motivadores de ignorância e incerteza nos dados taxonômicos e espaciais da biodiversidade. No primeiro capítulo, apresentamos uma perspectiva sobre como as mudanças taxonômicas, ocasionada pelo progresso nas práticas taxonômicas, têm levado a flutuações e imprecisões sobre o número de espécies já reconhecidas e o número espécies que realmente existente no mundo (lacuna Linneana). As revisões taxonômicas têm colaborado para os processos de divisões e agrupamentos taxonômicos, que por sua vez afetam as estimativas de riqueza de espécies. Nós recomendamos que cientistas sejam cautelosos na condução e interpretação das estimativas globais, mas também que seja imperativa a inclusão de dados sobre as mudanças taxonômicas nas estimativas da biodiversidade atual. Enfatizamos que a lacuna Linneana é um fenômeno biocultural, que será afetada por diversos aspectos como a história da área explorada e o esforço despendido na busca por táxons, ou ainda, como as espécies são definidas e delimitadas ou o nível de digitalização/mobilização de dados.

No segundo capítulo, segundo nossa métrica para avaliar da incerteza nomenclatural, observamos que mais de metade (56%) das espécies de aves globais apresentaram algum grau de incerteza relacionada ao nome científico. O cenário foi mais crítico nas classificações taxonômicas superiores, pois apenas seis das 36 Ordens de aves avaliadas não apresentaram incerteza nomenclatural. As variáveis biológicas e ecológicas que foram associadas com a incerteza nomenclatural foram: massa corporal, tamanho da área de distribuição, distinção evolutiva e status de conservação da IUCN. Sendo as espécies que apresentaram menores porcentagens de incerteza taxonômica àquelas de maior tamanho corporal, ampla distribuição geográfica, espécies não ameaçadas de extinção e espécies mais distintas evolutivamente. Discutimos que a taxonomia é uma ciência dinâmica e que as discordâncias de nomenclatura são reflexos de progresso, porém, também da falta de consenso da comunidade científica sobre o estatuto das unidades taxonômicas.

No terceiro capítulo, observamos uma desigualdade taxonômica e espacial nos registros de ocorrências da Namíbia, África. A maior parte (94%) dos registros de ocorrência dos cinco grupos taxonômicos de referência disponíveis no GBIF foi de Aves, especialmente devido ao influxo de dados em plataforma de ciência cidadã (eBird). Além disso, para maioria dos grupos taxonômicos avaliados, muitos dados podem ser considerados obsoletos, pois foram coletados antes dos anos 2000. Observamos tendências espaciais nas pontuações de ignorância que foram particularmente influenciadas pela maior densidade de rodovias, proximidade de instituições de pesquisa e de áreas protegidas. Destacamos a importância da ciência cidadã na mobilização de dados da biodiversidade, mas observamos que ainda são necessários esforços conjuntos de instituições (museus, herbários e outros centros de investigação) para acessar e disponibilizar dados da biodiversidade.

Nossos resultados, especialmente nos dos dois últimos capítulos, demonstram que muitos cientistas e coletadores estão realizando investigações taxonômicas, seja em suas zonas de conveniência e afinidade, mas principalmente pela falta de investimentos para expedições de campo, estudos taxonômicos, desenvolvimento e capacitação em novas tecnologias. Além disso, há impedimentos no acesso, recolha e usabilidade dos dados da biodiversidade, sendo extremamente recomendado que cientistas e instituições tornem os dados disponíveis para toda comunidade científica e tomadores de decisão. Finalmente, esta tese reforça a importância de investir em ciência de base para garantir que o conhecimento taxonômico e espacial das espécies seja amplamente difundido e as lacunas preenchidas.