



UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E CONSERVAÇÃO DA
BIODIVERSIDADE

MAPEANDO O (DES)CONHECIMENTO DA BIODIVERSIDADE DE
ORTHOPTERA (ARTHROPODA, INSECTA)

RODRIGO ANTÔNIO CASTRO SOUZA

CUIABÁ – MT
JANEIRO – 2025

RODRIGO ANTÔNIO CASTRO SOUZA

**MAPEANDO O (DES)CONHECIMENTO DA BIODIVERSIDADE DE
ORTHOPTERA (ARTHROPODA, INSECTA)**

Tese apresentada ao Curso de Pós-Graduação do Instituto de Biociências-UFMT, para obtenção do título de Doutor em Ecologia e Conservação da Biodiversidade, área de Concentração: Ecologia (organização da diversidade biológica).

Orientador: Prof.^o Dr.^o Thadeu Sobral-Souza

Coorientador: Prof.^o Dr.^o Neucir Szinwelski

Coorientadora: Prof.^a Dr.^a Geiziane Tessarolo

**CUIABÁ – MT
JANEIRO – 2025**

FICHA CATALOGRÁFICA

Dados Internacionais de Catalogação na Fonte.

S729m Souza, Rodrigo Antônio Castro.
Mapeando o (Des)Conhecimento da Biodiversidade de Orthoptera (Arthropoda, Insecta)
[recurso eletrônico] / Rodrigo Antônio Castro Souza. -- Dados eletrônicos (1 arquivo : 161 f.,
il. color., pdf). -- 2025.

Orientador: Thadeu Sobral-Souza.
Coorientador: Neucir Szinwelski.
Coorientadora: Geiziane Tessarolo.
Tese (doutorado) - Universidade Federal de Mato Grosso, Instituto de Biociências,
Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Cuiabá 2025.
Modo de acesso: World Wide Web: <https://ri.ufmt.br>.
Inclui bibliografia.

1. Taxonomia. 2. Macroecologia. 3. Lacunas de Conhecimento da Biodiversidade. 4.
Qualidade dos Dados. 5. Ortópterologia. I. Sobral-Souza, Thadeu, *orientador*. II. Szinwelski,
Neucir, *coorientador*. III. Tessarolo, Geiziane, *coorientador*. IV. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE MATO GROSSO
PRÓ-REITORIA DE ENSINO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E CONSERVAÇÃO DA BIODIVERSIDADE

FOLHA DE APROVAÇÃO

TÍTULO: "MAPEANDO O (DES)CONHECIMENTO DA BIODIVERSIDADE DE ORTHOPTERA (ARTHROPODA, INSECTA)"

AUTOR: DOUTORANDO RODRIGO ANTÔNIO CASTRO SOUZA

Tese defendida e aprovada em **15 de Janeiro de 2025**.

COMPOSIÇÃO DA BANCA EXAMINADORA

1. Doutor Thadeu Sobral de Souza [Presidente Banca/Orientador]

INSTITUIÇÃO: Universidade Federal de Mato Grosso

2. Doutora Carla da Costa Siqueira (Examinadora Externa)

INSTITUIÇÃO: Universidade do Estado do Rio de Janeiro

3. Doutor Edison Zefa (Examinador Externo)

INSTITUIÇÃO: Universidade Federal de Pelotas

4. Doutor Sidney Feitosa Gouveia (Examinador Externo)

INSTITUIÇÃO: Universidade Federal de Sergipe

5. Doutor Victor Mateus Prasniewski (Examinador Externo)

INSTITUIÇÃO: Universidade Federal da Fronteira Sul

6. Doutor Domingos de Jesus Rodrigues (Examinador Suplente)

INSTITUIÇÃO: Universidade Federal de Mato Grosso

7. Doutor Jessie Pereira dos Santos (Examinador Suplente)

INSTITUIÇÃO: Universidade de Campinas

Cuiabá - MT, 15 de Janeiro de 2025.



Documento assinado eletronicamente por **THADEU SOBRAL DE SOUZA, Coordenador(a) de Programas de Pós-Graduação em Ecologia e Conservação da Biodiversidade - IB/UFMT**, em 15/01/2025, às 17:20, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Víctor Mateus Prasniewski, Usuário Externo**, em 15/01/2025, às 17:32, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Carla da Costa Siqueira, Usuário Externo**, em 15/01/2025, às 17:42, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Sidney Feitosa Gouveia, Usuário Externo**, em 15/01/2025, às 17:47, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **EDISON ZEFA, Usuário Externo**, em 15/01/2025, às 18:14, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site

[http://sei.ufmt.br/sei/controlador_externo.php?](http://sei.ufmt.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)

[acao=documento_conferir&id_orgao_acesso_externo=0](http://sei.ufmt.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **7481530** e o código CRC **20B99C9C**.

AGRADECIMENTOS

Inicialmente, gostaria de agradecer aos colaboradores indiretos desta tese:

À minha mãe e ao meu pai† por todo o apoio e compreensão, que, sem dúvida, foram fundamentais. Em especial ao Sr. Sebastião de Jesus de Souza†, que, apesar de ter tido pouquíssima escolaridade, me ensinou valores que não se encontram nos livros.

Aos meus Tios, Tias, Primos e Irmão, pelos cafezinhos de domingo, que foram essenciais para recarregar as energias ao longo dessa jornada.

À minha companheira, Fernanda, por ser um pilar na minha vida desde o início do doutorado, sempre me apoiando, me aturando nos momentos difíceis e me ensinando a enxergar a vida com mais amor e carinho.

Ao meu enteado, Theo, por toda bagunça e pelos momentos felizes que compartilhamos no dia a dia.

Aos meus Sogros, pelo acolhimento, conversas filosóficas, e momentos felizes compartilhados.

Aos amigos, que encontrei passando pelo MacrEco, Afonso, Andréia Ceci, Geovana, Lana, Nicolas, Rafaela, Thairik, Victor, Vinicius, obrigado pelas preciosas discussões dentro e fora do mundo da ciência, em especial aos insights filosóficos e cafezinhos.

Aos amigos, que encontrei passando pelo Museu Nacional de Ciencias Naturales (MNCN), Alejandra, André, Cristine, Genoveva, Karoline, Monica, Raquel, Tainá, obrigado pelo acolhimento, cafezinhos e discussões.

Ao amigo Matheus, pesquisador incrível que encontrei passando na UFG, obrigado pelas conversas e conselhos.

A Deus, apesar de não ser uma pessoa de muita fé.

E aos colaboradores diretos desta tese:

Ao meu orientador, Prof.º Dr.º Thadeu Sobral de Souza, por me acolher desde o mestrado, por se oferecer para me orientar no doutorado e por ter me dado todo o suporte, independentemente de onde eu estivesse. Obrigado por ser uma grande inspiração para mim ao longo desses anos, por me conectar com a Macroecologia e inúmeros pesquisadores de referência.

Ao meu coorientador, Prof.º Dr.º Neucir Szinwelski, por seu acolhimento desde o início do doutorado, por estar sempre disponível para ajudar e por me incentivar a levar os Orthoptera a um lugar de destaque na Ecologia.

À minha coorientadora, Prof.^a Dr.^a Geiziane Tessarolo, por me orientar durante as primeiras análises, me receber de braços abertos na Universidade Federal de Goiás (UFG) e me ensinar a criar meu primeiro loop em R. Obrigado pela paciência e pela valiosa ajuda.

Ao Prof. José Alexandre Felizola Diniz-Filho, por me receber de braços abertos na UFG, por me incentivar a trazer os Orthoptera para o campo da Macroecologia e por me convidar a participar dos workshops sobre lacunas de conhecimento da biodiversidade, que foram inestimáveis para a construção desta tese. É uma honra ter conhecido uma das maiores referências vivas da Macroecologia.

Ao Prof. Dr. Joaquín Hortal, uma das minhas maiores referências como pesquisador, por ter me acolhido no Museu Nacional de Ciencias Naturales (MNCN), e tornado o sonho de intercambio uma realidade. Obrigado por todos os cafezinhos, e por ter me ajudado a criar um elo entre a taxonomia e macro-ecologia.

À Prof.^a Dr.^a Juliana Stropp Carneiro, pelas sugestões sempre super construtivas, obrigado por todos os ensinamentos

Ao Prof.^o Dr.^o Richad Ladle, pelas maravilhosas ideias sempre fora da caixinha, que me inspiram a pensar mais longe.

Ao Prof.^o Dr.^o Luiz Iniesta, pela imensa ajuda e discussões profundas sobre nomenclatura zoológica, taxonomia e sistemática, que foram fundamentais para estabelecer um elo entre taxonomia e macro-ecologia.

Ao Prof.^o Dr.^o Dieson Moi, pela introdução ao mundo dos modelos de equações estruturais, obrigado por todas as discussões empolgantes, pelo auxílio no R e pelos valiosos conselhos.

Aos mebros que participaram das minhas qualificações, Prof.^o Dr.^o Carlos Sperber, Prof.^o Dr.^o Edison Zefa, Prof.^o Dr.^o Fernando Vaz-de-Melo, Prof.^o Dr.^o Jessie Santos, Prof.^a Dr.^a Leila Meyer, Prof.^o Dr.^o Marcos Fianco, Prof.^a Dr.^a Rosane Collevatti, pelas preciosas sugestões, conversas e disponibilidade para ajudar.

À Universidade Federal de Mato Grosso (UFMT) e ao Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade (PPG-ECB) pelo inestimável suporte.

Sou eternamente grato a **CAPES, ICMBio/CECAV**, pelas bolsas de doutorado e sanduíche concedidas.

“To know that we know what we know, and that we do not know what we do not know, this is true knowledge.” Henry David Thoreau, 1854

RESUMO

Na era da bioinformática, tornou-se possível identificar lacunas taxonômicas-espaço-temporais no conhecimento da biodiversidade por meio de dados digitais acessíveis. Nesta tese, utilizamos esses dados para avaliar globalmente diferentes incompletudes e tendências no conhecimento mapeável sobre os Orthoptera (Arthropoda, Insecta), a sexta ordem mais rica em espécies de insetos, popularmente conhecidos como gafanhotos, grilos, e esperanças. No Capítulo I, mapeamos o estado do conhecimento taxonômico de Orthoptera, utilizando o ano de descrição das espécies como indicador de taxonomia alfa e a quantidade de sinônimas associadas a cada espécie como indicador de taxonomia beta. Esses indicadores foram combinados espacialmente, o que permitiu identificar que a taxonomia alfa predomina nas regiões tropicais e no Hemisfério Sul, enquanto a taxonomia beta é mais predominante nas regiões temperadas do Hemisfério Norte. No Capítulo II, mapeamos as lacunas no conhecimento sobre Orthoptera, utilizando medidas de completude taxonômica, amostral e temporal, bem como a equidade amostral e temporal dos inventários, integrando-as em um mapa de ignorância. Nossos resultados revelaram deficiências significativas no conhecimento biogeográfico de Orthoptera, marcadas por dados não refinados taxonomicamente, lacunas nos inventários, baixa cobertura temporal e desequilíbrio no número de registros, especialmente na região tropical. No Capítulo III, integramos drivers socioecológicos, como densidade populacional humana, densidade de áreas protegidas, desenvolvimento socioeconômico e riqueza conhecida de espécies de Orthoptera, ao longo do tempo a fim de explicar ignorância do conhecimento de Orthoptera em diferentes subdivisões biogeográficas (Global, Neártico, Neotropical, Paleártico, Afrotropical, Indo-Malaio e Australásia) por meio de modelos de equação estrutural (SEM). Nossos resultados revelaram que tanto a riqueza conhecida, quanto a ignorância de Orthoptera são diretamente influenciadas por processos antrópicos, conservacionistas, socioeconômicos e por fatores locais ao longo do tempo. Por fim, discutimos esses padrões, oferecendo insights sobre o conhecimento de Orthoptera em larga escala.

Palavras Chaves: Taxonomia; Macroecologia, Lacunas de Conhecimento da Biodiversidade, Qualidade dos Dados, Ortópterologia.

ABSTRACT

In the era of bioinformatics, it has become possible to identify taxonomic, spatial, and temporal gaps in biodiversity knowledge through digitally accessible data. In this thesis, we use these data to globally assess different types of incompleteness and trends in the mappable knowledge of Orthoptera (Arthropoda, Insecta), the sixth most species-rich order of insects, commonly known as grasshoppers, crickets, and katydids. In Chapter I, we mapped the state of taxonomic knowledge of Orthoptera, using the year of species description as an indicator of alpha taxonomy and the number of synonyms associated with each species as an indicator of beta taxonomy. These indicators were spatially combined, revealing that alpha taxonomy predominates in tropical regions and the Southern Hemisphere, while beta taxonomy is more prevalent in temperate regions of the Northern Hemisphere. In Chapter II, we mapped knowledge gaps in Orthoptera by assessing taxonomic, sampling, and temporal completeness, as well as the spatial and temporal evenness of inventories, integrating them into an ignorance map. Our results revealed significant deficiencies in the biogeographic knowledge of Orthoptera, characterized by taxonomically unrefined data, inventory gaps, low temporal coverage, and an imbalance in the number of records, especially in tropical regions. In Chapter III, we integrated socioecological drivers, such as human population density, protected area density, socioeconomic development, and known species richness of Orthoptera over time, to explain gaps in Orthoptera knowledge across different biogeographic subdivisions (Global, Nearctic, Neotropical, Palearctic, Afrotropical, Indo-Malayan, and Australasian) using structural equation models (SEM). Our results revealed that both known species richness and ignorance of Orthoptera are directly influenced by anthropogenic, conservationist, socioeconomic, and local factors over time. Finally, we discuss these patterns, providing insights into large-scale knowledge of Orthoptera.

Keywords: Taxonomy, Macroecology, Biodiversity Shortfalls, Data Quality, Orthopterology.

LISTA DE FIGURAS

INTRODUÇÃO GERAL

Figura 1. Sistematização das lacunas (1), limitações (2) e vieses (3) do conhecimento da biodiversidade, problemas diretamente associados à obtenção e coleta de dados. Em seguida, sistematizamos os déficits (4) e ruídos (5), problemas associados aos bancos de dados da biodiversidade.....47

Figura 2. Síntese sobre as Limitações, Déficits, Vieses e Ruídos do conhecimento da biodiversidade (baseado em Scott & Hallam 2003, Hey et al. 2003, Murphey et al. 2004, Hortal et al. 2008, Boakes et al. 2010, Ahrends et al. 2011, Rocchini et al. 2011, Ledle & Hortal 2013, Meyer et al. 2015, Escribano et al. 2016, Meyer et al. 2016, Tassarolo et al. 2017, Troudet et al. 2017, Ficetola et al. 2019, Siracusa et al. 2020, Hughes et al. 2021, Tassarolo et al. 2021)...48

Figura 3. Perguntas que podem ajudar a identificar incompletudes do conhecimento taxonômico, espacial e temporal associadas aos registros primários, considerando o momento de sua descoberta, formação de vieses, ruídos e déficits. Aqui, não consideramos as lacunas e limitações, tendo em vista que elas são intrínsecas ao conhecimento da biodiversidade.....49

CAPÍTULO I

Figure 1. Filtering process and identification of valid species names and synonyms in the Catalogue of Life (CoL) considering all taxonomic components (genus + specific epithet + author(s) + year). The first step consisted in removing fossils, subspecies, and synonyms of subspecies (Tax. status). In the second filter, we counted proposed names until 2023 (Tax. timeframe). Finally, the third filter consisted of removing duplicate records in valid species names and synonyms, disregarding synonyms resulting from subgenus syntax (Tax. ambiguities).....58

Figure 2. Workflow for building a taxonomic knowledge status map: (A) build a checklist of Orthoptera names (i.e., valid species names and associated synonyms) using data from the Catalogue of Life (CoL); (B) access the geographic distribution of the number of valid and total species names integrating data from the Orthoptera Species File Online (OSF) and the Global Biodiversity Information Facility (GBIF); (C) calculate two indices representing alpha and beta taxonomy: alpha temporal (AT) and beta proportion (BP); (D) build maps of alpha and beta taxonomy, rasterizing the indices of AT and BP in each 2° grid cell with data on orthoptera species; and (E) combine the AT (y axis) and BP (x axis) indices into a bivariate taxonomic knowledge status map.....62

Figure 3. Geographic distribution of the number of valid species names (A) and the total number of Orthoptera names (B). In (A) dark red colors indicate higher richness of valid species names, while light beige regions indicate lower richness. In (B) dark regions present higher total numbers of names (i.e., number of valid species names + number of associated synonymous names), while light colors indicate a lower total number of names.....64

Figure 4. Rates of taxonomic work on Orthoptera names through time. (A) accumulation of valid species names (blue), synonyms (orange), and active taxonomists (black) over time. (B)

number of currently valid species (blue), synonyms (orange), and taxonomists' (black) names per year.....66

Figure 5. Spatial distribution of Alpha Temporal (A) and Beta Proportion (B) indices for Orthoptera names. In (A) Yellow and green indicate older species names, blue represents more recent species names. In (B) yellow and green indicate a higher number of synonyms associated with valid species names, while blue corresponds to regions with lower proportions of synonyms.....67

Figure 6. Taxonomic knowledge status map resulting from the combination of alpha temporal (AT, X axis) and beta proportion (BP, Y axis). Regions in blue indicate recent species and high numbers of synonyms; regions in purple indicate older species together with many synonyms; regions in gray indicate recent species and few synonyms; and regions in yellow indicate older species and few synonyms.....68

Figure S1. Sensitivity analysis between the similarity of the names present in the synonym list and the corresponding valid species names using the distance algorithm of Damerau-Levenshtein (A), Lonest Common Substring (B), and Jaccard (C)..... 83

Figure S2. Density curves of alpha temporal (A) and beta proportion (B) indices by different thresholds.....83

Figure S3. Correlation between the number of valid species names and the total number of names.....84

Figure S4. Distribution of the Global Scientific Collections (A) and Types (i.e., holotypes, syntypes, lectotypes or neotypes) of Orthoptera (B). Data source: <https://scientific-collections.gbif.org/> and <https://orthoptera.speciesfile.org/>.....85

Capítulo II

Figure 1. Outline of all methodological and analytical steps required to construct the ignorance map for the order Orthoptera (Arthropoda, Insecta) according to digitally-accessible data. a) we compile all digitally accessible occurrence records for Orthoptera (see Table 1). b) we applied several filters to the records, which were divided into two groups according to their taxonomic resolution: 1°) species and subspecies level, and 2°) above species up to order level. c) within each 100 km-width grid cell, we calculated five different data quality indices. d) the indices were subsequently mapped. e) five of the metrics were combined by summing them in the form of an ensemble, which was then rescaled to range from 0 to 1, and finally subtracted from 1, in order to create a single final ignorance map.....114

Figure 2. Maps of the five indices (a–e) used in this study to assess the quality of digitally-accessible data for the global distribution of the order Orthoptera (Arthropoda, Insecta). a) taxonomic completeness, values close to 1 indicate that most records are identified to the species taxonomic level. b) survey completeness, values close to 1 indicate an almost complete inventory. c) temporal completeness, values close to 1 indicate maximum temporal coverage. d) survey evenness, values close to 1 indicate no local bias in the distribution of the numbers of records per species. e) temporal evenness, values close to 1 indicate no local bias in the distribution of samples per year.....115

Figure 3. Map of ignorance on the global distribution of the order Orthoptera (Arthropoda, Insecta) based on digitally-accessible data. Lower ignorance values indicate higher levels of knowledge, pinpointing areas with good data quality that only require some refinements to provide high-quality inventories.....	116
Figure 4. Density curves of the values of the data quality indices generated by different threshold configurations for taxonomic, survey and temporal completeness, and survey and temporal evenness. a) density curves for taxonomic completeness. b) density curves for survey completeness. c) density curves for temporal completeness. d) density curves for survey evenness. e) density curves for temporal evenness.....	117
Figure S1. Total number of species occurrences records after taxonomic, spatial, and temporal filtering processes (see methods).....	120
Figure S2. Total number of occurrences records above species up to order level after spatial and temporal filtering processes (see methods).....	121
Figure S3. Total number of occurrences records resulting from the sum of the total number of species occurrences records and the total number of occurrences records above the species level up to the order level per raster cell (see Figures 1 and 2 mentioned above).....	122
Figure S4. Year richness calculated from digitally accessible data for the order Orthoptera for each raster cell.....	123
Figure S5. Taxonomic Completeness (TaxC), values close to 1 indicate that most records are identified to the species taxonomic level.....	124
Figure S6. Survey Completeness (SC), values close to 1 indicate an almost complete inventory.....	125
Figure S7. Temporal Completeness (TE), values close to 1 indicate maximum temporal coverage.....	126
Figure S8. Survey Evenness (SE), values close to 1 indicate no local bias in the distribution of the numbers of records per species.....	127
Figure S9. Temporal Evenness (TE), values close to 1 indicate no local bias in the distribution of samples per year.....	128

CAPÍTULO III

Figura 1. Modelo hipotético de caminhos representando as relações direcionais analisadas e as hipóteses subjacentes, efeito positivo (+) e inverso (-) entre os diferentes drivers e a ignorância do conhecimento de Orthoptera em larga escala.....	134
Figura 2. Esboço resumindo nossas etapas metodológicas e analíticas para prever o efeito de diferentes drivers socioecológicos sobre a ignorância do conhecimento de Orthoptera ao longo do tempo. O primeiro passo consiste em mensurarmos a completude taxonômica, espacial e temporal em quatro diferentes cenários temporais (a), combinarmos essas três medidas por meio de uma soma para cada cenário temporal, reescalarmos para um intervalo entre 0 e 1 e	

subtrairmos de 1, construindo assim nossos mapas de ignorância para Orthoptera (b). O segundo passo consiste na extração de densidade de áreas protegidas, densidade populacional humana, produto interno bruto e riqueza de espécies de Orthoptera para os mesmos cenários e resolução espacial dos mapas de ignorância (c). Seguidamente, transformamos os dados para uma mesma escala numérica utilizando normalização min-max e avaliamos a correlação entre eles por meio de um teste de fator de inflação da variância (VIF) (d). Finalmente, construímos nosso modelo de equações estruturais em diferentes subdivisões biogeográficas associando os drivers preditores com a ignorância do conhecimento de Orthoptera (e).....137

Figura 3. Efeitos padronizados do Modelo de Equações Estruturais (SEM) considerando todas as interações diretas. As cores representam as diferentes subdivisões biogeográficas investigadas: Cinza – Global; Azul – Neártico; Azul claro – Neotropical; Verde – Paleártico; Amarelo – Afrotropical; Vermelho – Indo-Malaio; Vermelho escuro – Australásia. Círculos preenchidos indicam efeitos significativos ($p < 0,05$). Os resultados completos do SEM estão disponíveis na Tabela e na Figura Suplementar 1.....139

Figura S1. Modelos de Equações Estruturais (SEM) dos drivers preditores da ignorância da biodiversidade de Orthoptera. Os coeficientes com: *, ** ou *** representam caminhos significativos para o modelo(s), enquanto (NS) indicam caminho(s) não significativos ($p \geq 0,05$), (c) indicam que estimativa é fixa em todas as subdivisões biogeográficas para a relação. Os níveis de significância de cada preditor são * $p \leq 0.05$, * $p \leq 0.01$, *** $p \leq 0.001$. (M) e (Co) representam o R² Marginal e Condicional, respectivamente. Resultados completos dos modelos estão na Tabela Suplementar 1.....151

Figura S2. Efeitos padronizados do Modelo de Equações Estruturais (SEM) considerando as interações indiretas mediadas pela riqueza de espécies de Orthoptera. As cores representam as diferentes subdivisões biogeográficas investigadas: Cinza – Global; Azul – Neártico; Azul claro – Neotropical; Verde – Paleártico; Amarelo – Afrotropical; Vermelho – Indo-Malaio; Vermelho escuro – Australásia. Círculos preenchidos indicam efeitos significativos ($p < 0,05$). Os resultados completos do SEM estão disponíveis na Tabela e na Figura Suplementar 1....153

LISTA DE TABELAS

INTRODUÇÃO GERAL

Tabela 1. Síntese sobre as principais lacunas do conhecimento da biodiversidade (referências marcadas com (*) são as originais de cada lacuna)46

CAPÍTULO I

Table 1. Digital repositories accessed to build the global dataset of Orthoptera.....113

CAPÍTULO III

Tabela S1. Coeficientes de caminhos diretos, tanto padronizados quanto não padronizados, de todas as relações testadas em nosso Modelo de Equação Estrutural (SEM) para prever a ignorância da biodiversidade de Orthoptera. Os drivers incluem: cenário temporal (TEMPO), densidade da população humana (POP), densidade de áreas protegidas (PAD), produto interno bruto (GDP) e riqueza de espécies de Orthoptera (RIQ). A tabela inclui todos os caminhos significativos e não significativos (→) em diferentes subdivisões biogeográfica e as variáveis que foram permitidas para covariar (↔). * = $P < 0.05$, ** = $P < 0.01$ e *** = $P < 0.001$147

SUMÁRIO

17	HISTÓRICO DE CONSTRUÇÃO DA TESE
19	INTRODUÇÃO GERAL
50	CAPÍTULO I - Mapping the Status of Global Taxonomic Knowledge of Orthoptera (Arthropoda, Insecta)
54	Introduction
57	Methods
63	Results
69	Discussion
74	Conclusion
78	References
83	Supplementary Material
89	CAPÍTULO II - Mapping Ignorance to Uncover Shortfalls in the Knowledge on Global Orthoptera Distribution
91	Introduction
94	Results
95	Discussion
100	Methods
106	References
118	Supplementary Material
129	CAPÍTULO III - Drivers da Ignorância do Conhecimento da Biodiversidade de Orthoptera (Arthropoda, Insecta)
132	Introdução
134	Métodos
138	Resultados
140	Discussão
142	Referências
147	Material Suplementar
154	CONCLUSÃO GERAL
156	PRODUTOS ADICIONAIS

HISTÓRICO DE CONSTRUÇÃO DA TESE

Após concluir a graduação em Biologia e o mestrado em Ecologia Aplicada, pela Universidade Federal de Lavras (UFLA), onde trabalhei com taxonomia e ecologia dos “grilos da escuridão”, estava muito animado para continuar estudando esse fascinante grupo. No entanto, talvez por obra do acaso, as parcerias essenciais para dar continuidade ao projeto não se concretizaram. Assim, vi meu sonho de doutorado ser sepultado, e me senti em um verdadeiro abismo, decidindo me afastar da pesquisa.

Dois anos após o mestrado, percebi que minha vocação estava na pesquisa. Foi então que iniciei meu doutorado na Universidade Federal de Mato Grosso (UFMT), no final da pandemia de COVID-19, sendo acolhido pelo Dr. Thadeu Sobral de Souza no laboratório de Macroecologia (MacrEco). A única condição que impus foi que desejava trabalhar com todas as espécies de grilos e gafanhotos do mundo, abrangendo toda a diversidade da ordem Orthoptera.

Cheio de ideias, mergulhei em leituras e referências para estruturar um projeto sólido. Em um determinado momento, meu orientador compartilhou comigo um artigo de 1949, escrito pelo cartógrafo Samuel Whittemore Boggs, que propunha a criação de mapas de ignorância como ferramenta para avançar no conhecimento em áreas como biogeografia, geologia e climatologia. Esse insight foi decisivo e inspirou o tema central da minha tese: mapear globalmente a falta de conhecimento sobre os Ortópteros. Ao longo dessas leituras, precisei sistematizar conceitos fundamentais, como lacunas, limitações, vieses, déficits e ruídos. Esse exercício teórico resultou na introdução desta tese, que foi publicada em português, na revista *Oecologia Australis*, com a ajuda de outros colaboradores.

Para executar meu projeto de doutorado, meu orientador me conectou a duas grandes referências que se tornaram meus co-orientadores: o professor Dr. Neucir Szinwelski, um ecólogo especialista em Orthoptera, e a professora Dra. Geiziane Tassarolo, uma macroecóloga que já desenvolvia abordagens inovadoras para mapear a ignorância biogeográfica. Graças a essa rede de colaborações, minha trajetória tornou-se mais rica e possibilitou que eu visitasse minha co-orientadora na Universidade Federal de Goiás (UFG). Durante essa visita, tive a oportunidade de conhecer também um dos grandes nomes da Macroecologia, o professor Dr. José Alexandre Diniz-Filho, que, curiosamente, também havia trabalhado com grilos em seu mestrado na década de 1990. Entusiasmado com meu trabalho, ele convidou meu orientador e a mim para participar de um workshop sobre lacunas de conhecimento da biodiversidade.

Foi nesse Workshop que, ao apresentar parte do meu trabalho, conheci uma série de futuros colaboradores essenciais para a execução da minha tese, incluindo a Dra. Juliana Stropp, o Dr. Richard Ladle e o Dr. Joaquín Hortal, entre outros pesquisadores excepcionais. Durante o evento, Joaquín perguntou ao meu orientador sobre o meu interesse em realizar um intercâmbio e se disponibilizou para ser meu supervisor no exterior caso tudo desse certo. Foi exatamente nesse momento que a ideia do doutorado sanduíche nasceu. Determinado, propus ao meu orientador que elaborássemos um projeto para buscar o financiamento necessário. Talvez, por obra do acaso mais uma vez, as coisas deram certo. No último ano do meu doutorado, deixei o Brasil pela primeira vez para morar na Espanha e executar parte da minha tese no Museu Nacional de Ciências Naturais (MNCN) em Madri, sob a supervisão do Dr. Joaquín Hortal. Nesse período, buscamos conectar taxonomia e macroecologia, mapeando o estado do conhecimento taxonômico de Orthoptera, trabalho que veio a compor o primeiro capítulo da minha tese.

Além desse trabalho, dediquei-me a compilar o maior banco de dados possível sobre a distribuição global de Orthoptera. Com esses dados em mãos, foi possível mapear a ignorância, utilizando medidas de qualidade taxonômica, espacial e temporal. Este trabalho constitui o segundo capítulo da tese, e foi publicado na revista *npj Biodiversity*, do grupo Nature, e levantou reflexões importantes: por que algumas regiões são bem conhecidas, enquanto outras permanecem negligenciadas? Para responder a essa questão, desenvolvemos um modelo em larga escala que associa drivers temporais, antropogênicos, conservacionistas, socioeconômicos e de diversidade biológica aos nossos mapas de ignorância. Este esforço resultou no terceiro capítulo da tese.

De volta ao Brasil, nos meses finais do doutorado, fiz uma breve viagem ao Chile com minha família. Contudo, o que deveria ser um período de descanso transformou-se em um grande desafio: fui assaltado, tive uma arma apontada para minha cabeça e perdi todos os documentos, além do computador com centenas de códigos e materiais do doutorado. Esse episódio abalou-me profundamente. Mas com resiliência, reconstruí o que foi roubado, dia após dia, aprendendo que, nos momentos difíceis, precisamos nos tornar mais fortes.

Por fim, deixo uma mensagem motivacional para aqueles que chegaram até estas últimas linhas. Não tenham medo de sonhar, mesmo que não tenham estudado nas melhores escolas ou que venham de famílias sem tradição “acadêmica”. Meu pai era analfabeto e minha mãe, costureira. Na minha casa, não havia muitos livros ou computadores. Conheci o mundo da ciência na faculdade pública, e ela alimentou meus sonhos e minha criatividade. Acreditem no poder transformador da educação, mesmo diante dos desafios.

INTRODUÇÃO GERAL

O (Des)Conhecimento da Biodiversidade: uma Sistematização sobre Lacunas, Limitações, Vieses, Déficits e Ruídos

Artigo publicado na revista *Oecologia Australis*

Link: <http://doi.org/10.4257/oeco.2024.2803.01>

O (des)conhecimento da biodiversidade

**O (DES)CONHECIMENTO DA BIODIVERSIDADE: UMA SISTEMATIZAÇÃO
SOBRE LACUNAS, LIMITAÇÕES, VIESES, DÉFICITS E RUÍDOS**

*Rodrigo Antônio Castro-Souza^{1,4 *}, Geiziane Tessarolo⁶, Nicolas Silva Bosco^{1,4}, Andréia de
Carvalho Santos², Afonso Kempner^{1,4}, Thairik Mateus Silva Marques^{1,4}, Vinícius da Costa-
Silva^{3,5}, José Alexandre F. Diniz-Filho⁶ & Thadeu Sobral-Souza⁴*

¹Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Federal de Mato Grosso (UFMT), CEP 78060-900, Cuiabá, MT, Brasil.

²Programa de Pós-Graduação em Ecologia e Conservação, Instituto de Biociências, Universidade Federal de Mato Grosso do Sul (UFMS), CEP 79070-900, Campo Grande, MS, Brasil.

³Department of Zoology & Entomology, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa.

⁴Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso (UFMT), CEP 78060-900, Cuiabá, MT, Brasil.

⁵Laboratório de Scarabaeoidologia, Instituto de Biociências, Universidade Federal de Mato Grosso (UFMT), Cuiabá, MT, Brasil.

⁶Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás (UFG), Goiás, GO, Brasil.

E-mails: rodrigodesouzaac@gmail.com(*autor correspondente); geites@gmail.com; nicolas-bosco@hotmail.com; afonso.hk@gmail.com; andreaecologa@gmail.com; thairiksilvamarques@gmail.com; silvavinicius92@gmail.com; jafdinizfilho@gmail.com; thadeusobral@gmail.com

RESUMO: A partir dos avanços das ferramentas de bioinformática e ecoinformática tornou-se viável usar grandes quantidades de informação sobre biodiversidade para descrever padrões biológicos e tentar avaliar os processos ecológicos e evolutivos subjacentes a eles, bem como estabelecer ações mais efetivas e abrangentes para conservá-la. Contudo, as informações existentes acerca da biodiversidade ainda são incompletas, possuindo lacunas, limitações, vieses, déficits e ruídos. Neste sentido, reconhecemos a necessidade de discutir problemáticas relacionadas ao uso dos grandes bancos de dados de biodiversidade e a sua qualidade, discussões estas que precisam ser conceitualizadas de forma mais estruturada. Neste texto, iremos descrever de forma sistematizada as lacunas, limitações, e vieses do conhecimento da biodiversidade, problemas que estão diretamente associados à obtenção de informações das espécies na natureza. Em seguida, delimitaremos os tipos de déficits e ruídos, que são problemas associados aos bancos de dados da biodiversidade. Isso se deve ao fato de que as informações presentes nos registros primários e nos grandes bancos de dados frequentemente são usadas de forma inadequada. Além disso, apontamos também perspectivas sobre como abordar essas deficiências em trabalhos ecológicos/ biogeográficos, por meio de uma adequada curadoria dos bancos de dados e da construção de mapas de ignorância. Por fim, destacamos desafios futuros, como a necessidade de novas coletas de dados, a formação de novos taxonomistas e o fornecimento de subsídios financeiros para a infraestrutura e a digitalização das coleções biológicas.

Palavras-chaves: déficits de conhecimento; lacunas da biodiversidade; limitações de conhecimento; ruídos do conhecimento; vieses de conhecimento.

BIODIVERSITY (UN)KNOWLEDGE: A SYSTEMATIZATION OF GAPS, LIMITATIONS, DEFICITS, BIASES, AND NOISE

ABSTRACT: With the advancements in bioinformatics and ecoinformatics tools, it has become feasible to utilize vast amounts of biodiversity information to describe biological patterns and attempt to assess the ecological and evolutionary processes underlying them, as well as to establish more effective and comprehensive conservation actions. However, existing information regarding biodiversity remains incomplete, with gaps, limitations, biases, deficits, and noise. In this regard, we acknowledge the need to discuss issues related to the use of large biodiversity databases and their quality, discussions that need to be conceptually structured. In this text, we will systematically describe the gaps, limitations, and biases in biodiversity knowledge, issues that are directly associated with the acquisition of species information in nature. Subsequently, we will outline the types of deficits and noise, which are problems associated with biodiversity databases. This is because the information present in primary records and large databases is frequently used inadequately. Additionally, we also point out perspectives on how to address these deficiencies in ecological/biogeographical work through appropriate database curation and the construction of maps of ignorance. Finally, we highlight future challenges, such as the need for new data collections, the training of new taxonomists, and the provision of financial support for the infrastructure and digitization of biological collections.

Keywords: gaps of biodiversity; knowledge limitations; knowledge deficits; knowledge biases; knowledge noise.

INTRODUÇÃO

Estudos biogeográficos e macroecológicos, bem como suas aplicações em conservação da biodiversidade, dependem diretamente do conhecimento sobre a biodiversidade que foi construído e armazenado por diversos cientistas ao longo do tempo. A fonte básica do conhecimento da biodiversidade é definida pelos registros primários, que são informações de quais espécies estão descritas e onde elas estão (ou estavam) no tempo e no espaço (Soberón & Peterson 2004). Além das informações taxonômicas, espaciais e temporais, muitos registros podem conter também informações genéticas e de traços funcionais (veja Cornwell *et al.* 2019) ou até mesmo de bioacústica. No entanto, focaremos aqui apenas na tríade taxonomia x espaço x tempo, visto que ela é o componente fundamental em praticamente todo estudo de escopo ecológico.

A obtenção dos registros primários ocorre, *a priori*, por meio de coletas ou amostragens das espécies (p.ex., visual, acústico ou por registros indiretos) em comunidades ecológicas onde elas habitam. Dependendo do habitat e da história natural de cada grupo-alvo estudado, diferentes métodos e horários de coleta são necessários (Papavero 1994, Alves *et al.* 2021, Appel *et al.* 2021). Após a coleta, os registros, individualizados ou em lotes (amostras), devem ser obrigatoriamente etiquetados com informações sobre o(s) organismo(s) coletado(s): localização (p.ex., país, estado, cidade e coordenadas geográficas), data de coleta, metodologia utilizada, bem como o nome do(s) coletor(es) (Almeida *et al.* 2012). A identificação taxonômica prévia do material é importante, porém não obrigatória nessa etapa do processo. Em seguida, os registros precisam ser armazenados de forma adequada em recipientes com líquido fixador (por exemplo, álcool ou formol) ou em caixas secas contendo repelentes (p.ex., cânfora ou naftalina) e enviados para laboratórios ou coleções biológicas, onde serão triados, montados e identificados (Papavero 1994).

Como toda a biodiversidade na Terra ainda não é conhecida em suas múltiplas facetas do conhecimento, nos últimos anos foram identificadas e evidenciadas diversas lacunas no conhecimento da diversidade biológica. Tais lacunas conhecidas como “*shortfalls*”, em inglês, chamadas a partir de agora de lacunas de conhecimento, referem-se à falta de conhecimento ou ao conhecimento desigual da biodiversidade no tempo e espaço, no que tange as diferentes facetas da biodiversidade (Hortal *et al.* 2015, Ficetola *et al.* 2019, Lee *et al.* 2020, LopesLima 2021). Contudo, além das lacunas existem limitações que são obstáculos, restrições ou desafios que impedem o conhecimento completo da biodiversidade. Devido a existência das lacunas de

conhecimento e limitações, vão existir também os vieses de conhecimento da biodiversidade, que se referem a um conhecimento desigual da biodiversidade em relação a sua totalidade.

Com o avanço de ferramentas de bioinformática e ecoinformática pode-se facilmente construir bancos de dados robustos com milhares de registros primários presentes em artigos e coleções biológicas que estão disponibilizados em plataformas digitais (Sousa-Baena *et al.* 2014), como por exemplo GBIF (<https://www.gbif.org/>), OBIS (<https://obis.org/>) e iNaturalist (<https://www.inaturalist.org/>) que podem servir como base para preencher e/ou apontar as lacunas, as limitações e os vieses existentes no conhecimento da diversidade biológica. Mas com o advento dos bancos de dados, surgiram também novos problemas, como a falta de refinamento das informações dos registros, a partir de agora chamados de déficits do conhecimento, e os erros nas informações, a partir de agora chamados de ruídos do conhecimento.

Neste texto, iremos descrever de forma sistematizada as lacunas, limitações, e vieses do conhecimento da biodiversidade, problemas que estão diretamente associados à obtenção e coleta de informações das espécies na natureza. Seguidamente, delimitaremos os tipos de déficits e ruídos, que são problemas associados aos bancos de dados da biodiversidade (Figura 1). Isso se deve ao fato de que as informações presentes nos registros primários e nos grandes bancos de dados frequentemente são usadas de forma inadequada.

As Lacunas

Atualmente existem dez lacunas de conhecimento da biodiversidade descritas na literatura, embora futuramente ainda possam ser descritas novas (veja Hortal *et al.* 2015, Ficetola *et al.* 2019, Lee *et al.* 2020, Lopes-Lima 2021), enquanto algumas já existentes podem vir a ser sinonimizadas. Elas surgiram devido à ausência de correspondência entre o que se espera ter de informação total na natureza e o que é/foi realmente coletado e revelado pela ciência. A lacuna mais fundamental de todas é a lacuna Linneana (LL), que representa nosso desconhecimento sobre a identidade taxonômica das espécies (Cardoso *et al.* 2011, Hortal *et al.* 2015). Estima-se que cerca de 86% das espécies terrestres e 91% das espécies marinhas sejam ainda desconhecidas para a ciência (Mora *et al.* 2011). Ainda assim, estima-se que alguns táxons, como vertebrados, por exemplo, possuem quase que a totalidade das suas espécies descritas. Por outro lado, alguns grupos, como artrópodes, ainda possuem uma pequena porção das suas espécies descritas, possuindo assim alta LL (Mora *et al.* 2011, Stork *et al.* 2018). A LL engloba também inúmeros novos morfotipos que estão depositados em coleções, porém,

ainda não foram formalmente descritos (Hortal *et al.* 2015, Kemp 2015). Atualmente espécies crípticas estão sendo distinguidas em táxons distintos por meio de novas metodologias/critérios (p.ex., DNA barcoding, bioacústica e cariótipo) (veja discussão em Freeman & Pennell, 2021). A LL não existe só na natureza, mas também nos bancos de dados de registros primários, tendo em vista que nem todas as espécies atualmente descritas podem constar nessas plataformas.

Assumindo que espécies existam (veja discussão em Coyne & Orr 2004), a LL deve ser considerada a mais basal, sendo a primeira fonte de (des)conhecimento da biodiversidade. As demais lacunas são, portanto, subordinadas a essa lacuna, já que dependem do conhecimento taxonômico (descrição formal de uma espécie). Em suma, não se pode discutir padrões espaciais da biodiversidade se não sabemos com quais espécies estamos lidando. Embora a LL se demonstre importante, sabe-se hoje que o conhecimento espacial da biodiversidade, baseado na taxonomia das espécies descritas atualmente, é enviesado pela acessibilidade - locais próximos a cidades, estradas e rios são mais amostrados (Meyer *et al.* 2015, Zizka *et al.* 2020). Questões socioculturais também são direcionadores de viés de amostragem, tendo em vista que existem táxons e regiões mais “carismáticas” a serem estudados. Aves, por exemplo, é o táxon mais estudado quando comparado com outras Classes de vertebrados e invertebrados (Troudet *et al.* 2017). Por outro lado, fragmentos naturais maiores e mais conectados entre si são mais amostrados que outros dentro da Mata Atlântica (Sobral-Souza *et al.* 2021).

A lacuna Wallaceana (LW) representa o (des) conhecimento acerca da distribuição espacial das espécies (Cardoso *et al.* 2011, Hortal *et al.* 2015), que depende *a priori* do conhecimento taxonômico (LL). Para a maior parte das espécies já descritas, existem poucas ou, se não uma única localidade conhecida (localidade tipo), o que torna o conhecimento espacial da maioria das espécies tendencioso e ao mesmo tempo incerto para inúmeras regiões (Cardoso *et al.* 2011). Adicionalmente, considerando os dados de cobertura de distribuição para as principais espécies de animais terrestres e marinhos, sabemos que menos de 7% da biota terrestre e apenas 5% da biota marinha possuem informações sobre a distribuição espacial. Essas informações se concentram principalmente próximas a fatores de acessibilidade, como estradas, zonas costeiras e rotas marítimas (Hughes *et al.* 2021). Tal padrão ocorre muitas vezes devido à grande parte do conhecimento wallaceano estar pouco refinado, onde podemos ter uma quantidade substancial de registros primários com informações de localidade incertas, como por exemplo, a nível de município, estado ou província, e até mesmo país, que são

desconsiderados nos processos de filtragem que utilizam apenas dados com informações seguras de coordenadas geográficas (veja Meyer *et al.* 2016, Ronquillo *et al.* 2023).

A lacuna Prestoniana (LP) representa o (des) conhecimento sobre a dinâmica populacional das espécies ao longo do espaço e do tempo (Cardoso *et al.* 2011). Da mesma forma, não se pode inferir se não sabemos quais são as espécies (LL). Para acessar tal lacuna é necessário um método de coleta (ou observação) representativo para o grupo. Além disso as informações de abundância amostradas em um determinado local muitas vezes podem ser perdidas devido o mal aproveitamento dos dados catalogados (Cardoso *et al.* 2011) e escala de amostragem (p.ex., no caso de metapopulações e populações de distribuição irregular) (Freckleton & Watkinson 2002). Adicionalmente, vale ressaltar que a dinâmica populacional das espécies, por vezes, depende de metodologia que visa a não amostragem de espécimes, em processos de marcação e recaptura, por exemplo (Sobral-Souza *et al.* 2015). Dessa forma, esses dados muitas vezes não podem ser acessados várias vezes, em diferentes períodos de tempo por diferentes pesquisadores. Somados a isso, a dinâmica populacional é um processo local que varia ao longo das subpopulações das espécies, aumentando ainda mais a LP.

As lacunas Darwiniana (LD) e Raunkiaeriana (LR) refletem nosso desconhecimento da evolução das espécies (Diniz-Filho *et al.* 2013) e das características funcionais das espécies e suas funções ecológicas (Hortal *et al.* 2015), respectivamente. Estas lacunas também podem ser consideradas de forma subordinada à LL, e também a LW, visto que se desconhecemos as espécies em um determinado espaço geográfico, não podemos inferir relações evolutivas (posicionamento filogenético) e nem quais as características fenotípicas relacionadas as suas funcionalidades evoluíram em um dado local (embora a proposição original da LD tenha sido feita em um contexto de substituir medidas de diversidade a partir da riqueza de espécies por métricas filogenéticas, que seriam mais robustas à incertezas taxonômicas; Diniz-Filho *et al.* 2013). De modo antagônico, a falta de conhecimento a respeito da história evolutiva e características dos grupos (LD) também prejudica a compreensão da distribuição espacial atual das espécies, estando assim, diretamente relacionada com a LW (Hortal *et al.* 2015, Rudbeck *et al.* 2022, Diniz-Filho *et al.* 2023). Isso ressalta que as lacunas do conhecimento da biodiversidade são influenciadas não somente por escalas ecológicas, mas também por escalas de dimensões evolutivas (Whittaker *et al.* 2005). Contudo, mesmo existindo essas discussões, novas iniciativas biogeográficas estão surgindo como forma de interligar o conhecimento entre

LL, LD e LW, gerando conhecimento sobre padrões biogeográficos (veja Diniz-Filho *et al.* 2023).

Além das lacunas preditas, temos as lacunas Hutchinsoniana (LH) e Eltoniana (LE), que indicam nosso desconhecimento sobre a tolerância ambiental das espécies às condições abióticas e o desconhecimento das interações biológicas, respectivamente (Hortal *et al.* 2015, mas veja também Rosado *et al.* (2016)). A compreensão da LH pode auxiliar na compreensão de outras lacunas, pois algumas dinâmicas ambientais podem ser capazes de regular o tamanho das populações e interações das espécies (veja a discussão em Hortal *et al.* 2015). Em escalas amplas, a LE pode ser considerada um ruído para a construção de modelos correlativos de nicho (veja Soberón & Nakamura 2009). No entanto, o uso de traços funcionais tem demonstrado que a compreensão dos padrões locais pode ser determinante nos padrões macroecológicos e biogeográficos (Rosado *et al.* 2016). Por fim, também temos as lacunas Parkerianas (LPa), Ostromianas (LO) e Racovitšanas (LRc), formalizadas mais recentemente, que ressaltam, respectivamente, nosso (des)conhecimento sobre a história natural, a eficácia das medidas de conservação e a exploração e o mapeamento das espécies em todos os ambientes terrestres e aquáticos (Ficetola *et al.* 2019, Lee *et al.* 2020, Lopes-Lima 2021) (Tabela 1).

As lacunas no conhecimento da biodiversidade se tornaram um dos grandes desafios para as pesquisas biológicas em todo o globo, visto que impactam diretamente a interpretação e qualidade dos padrões de biodiversidade em diversas escalas ecológicas, em instâncias como a precisão, generalização e realismo dos dados (Hortal *et al.* 2015). Estes desafios podem ser ainda maiores quando considerados os efeitos hierárquicos anteriormente mencionados, uma vez que, a depender das áreas (configuração da paisagem), períodos e escalas estudadas, as lacunas do conhecimento da biodiversidade podem ter influência umas sobre as outras, correlacionando-se de forma sinérgica (Diniz-Filho *et al.* 2023).

As Limitações

As limitações são obstáculos, restrições e/ ou desafios que impedem o conhecimento completo da biodiversidade. As limitações podem ser responsáveis pela formação das lacunas de conhecimento. Aqui, vamos listar as cinco principais limitações, embora existam outras, como a falta de fotografias das espécies na natureza (Mesaglio *et al.* 2023) ou até mesmo a dificuldade de acessar a série tipo das espécies depositadas em coleções estrangeiras por pesquisadores das regiões tropicais (Nakamura *et al.* em preparação).

A primeira refere-se às limitações de exploração, que ressalta a falta de inventários da biodiversidade em alguns ambientes naturais, resultando no desconhecimento quase que total da biodiversidade em muitos lugares do planeta (p.ex., cavernas, fundo dos oceanos, manchas de florestas e desertos) (Ficetola *et al.* 2019). Certamente existem ambientes mais “atraentes” (florestas tropicais, recifes de corais) que são mais amostrados, seja pela sua beleza cênica ou pela elevada biodiversidade. Por outro lado, ambientes menos “atraentes”, como brejos, alagados e mangues, são menos amostrados sendo, portanto, menos conhecidos.

Outra limitação é a detectabilidade. Muitas espécies, e/ou assembleias, são menos detectáveis do que outras fazendo com que algumas espécies sejam subamostradas, como, por exemplo, espécies raras, com abundância reduzida e/ ou espécies migratórias (Ladle & Hortal 2013). Somados a isso, existem diferentes estratégias e finalidades na forma de inventariar espécies em ambientes naturais (Ahrends *et al.* 2011). As técnicas utilizadas para amostragens da biodiversidade podem ter atratividade específica para alguns táxons, em detrimento de outros que estão no mesmo ambiente. Isso resulta na hiper amostragem de alguns táxons como resultado da técnica de amostragem específica e não pela realidade das abundâncias relativas das espécies na comunidade local (Bach *et al.* 2023).

As limitações de similaridade ambiental também são inerentes ao conhecimento da biodiversidade. O conhecimento da distribuição das espécies diminui com o aumento da distância (espacial e/ou climática) em relação à ocorrência conhecida da espécie. Isso acontece devido à existência de uma variabilidade na composição de espécies ao longo do espaço geográfico e ambiental, como, por exemplo, em função de gradientes de habitat, clima e vegetação (Ladle & Hortal *et al.* 2013, Tassarolo *et al.* 2021). Isso deixa claro que as limitações de similaridade ambiental e a limitação de exploração são intimamente relacionadas. Embora isso seja bastante óbvio, é necessário considerar que, dependendo do tipo de extrapolação feita com os dados existentes e da falta de registros primários, o padrão real de agregação de ocorrências próximo ao local de amostragem (baixa heterogeneidade de registros e alta correlação espacial), tem sido desconsiderado na maioria dos estudos como no caso do uso de modelos de distribuição de espécies (SDMs - veja a problematização levantada em Lobo *et al.* 2018) ou construídos a partir de ocorrência enviesadas no espaço geográfico e ambiental (veja GutierrezVelez & Wiese 2020, Sobral-Souza *et al.* 2021), podendo resultar em interpretações irrealísticas de padrões e processos.

As limitações temporais dos bancos de dados podem ser explicadas pelo dinamismo natural (p.ex., distúrbios estocásticos, extinções e invasões biológicas) e antrópico (p.ex., perda de habitat, poluição e mudanças climáticas) nos sistemas naturais que altera ao longo do tempo a distribuição geográfica dos registros primários, o que acontece devido a diversos processos em diferentes escalas ecológicas (Escribano *et al.* 2016, Tassarolo *et al.* 2017). Um registro originalmente descrito para um ambiente florestal, por exemplo, pode não ser mais válido atualmente por conta da perda do habitat florestal daquela localidade após a amostragem da espécie. Outra fonte de limitação temporal é que a nomenclatura taxonômica das espécies pode sofrer mudanças devido a novos atos nomenclaturais e novas classificações taxonômicas que ocorreram após o registro. Isso gera históricos taxonômicos complexos como sinônimas e homônimas que não podem ser desconsiderados na curadoria dos bancos de dados (Tassarolo *et al.* 2017, Stropp *et al.* 2022). Além disso, com o tempo, também pode ocorrer a perda dos metadados e das informações básicas sobre as coletas de dados, além do intemperismo nos registros armazenados em coleções biológicas. Tais perdas temporais ocasionam o desaparecimento de informações vitais sobre os registros primários, como a data de coleta, coordenadas, identificação do autor e do coletor, ou até mesmo partes dos espécimes depositados (Tassarolo *et al.* 2017). Muitas vezes isso pode depender mais das condições de armazenamento dos dados do que do tempo (Greshko 2018). Todas as limitações temporais aqui listadas tornam as informações conhecidas em um banco de dados obsoletas ao longo do tempo, afetando diretamente o conhecimento da distribuição das entidades taxonômicas (Tassarolo *et al.* 2017; Stropp *et al.* 2022), bem como o estudo e acesso aos espécimes físicos estudados pelos taxonomistas (Tassarolo *et al.* 2017). Isso resulta diretamente na perda de informações preciosas sobre a biodiversidade, transformando informações conhecidas em desconhecidas.

A última limitação refere-se à limitação de disponibilidade. As coleções biológicas de diferentes partes do mundo ainda não são de domínio público e/ou não estão digitalizadas e disponíveis em repositórios digitalmente acessíveis (Krishtalka & Humphrey 2000, Nelson & Ellis 2018, Bakker *et al.* 2020). Além disso, muitos dos dados disponíveis na literatura taxonômica científica e literatura cinza (dados científicos não publicados, incluindo dissertações, teses e relatórios) também não estão digitalizados e disponíveis (Heidorn 2008, Sikes *et al.* 2016). A falta das grandes coleções biológicas nos repositórios digitais é creditada, principalmente, à falta de mão de obra especializada (e remunerada) para o preenchimento do banco de dados, além da quantidade exorbitante de material ainda a ser digitalizado. Muitas

das espécies atualmente descritas pela ciência podem também não possuir registros de sua descrição e/ou distribuição disponíveis digitalmente, uma deficiência que normalmente ocorre nos repositórios digitais da biodiversidade (Figura 2).

Embora o grau de impacto dessas limitações possa variar dependendo do tipo de estudo e enfoque da pesquisa, a maioria dos levantamentos de dados de biodiversidade vão sofrer com algumas dessas limitações. Assim, mesmo bancos de dados com milhões de registros primários, provenientes de diversos estudos, vão herdar as limitações inerentes aos dados originais, e, portanto, sofrer das mesmas limitações em maior ou menor grau, o que pode resultar na impossibilidade de responder algumas perguntas biogeográficas (Sobral-Souza *et al.* 2021, Ronquillo *et al.* 2023). Por exemplo, a falta de informações sobre a data de coleta de determinados registros pode impossibilitar qualquer interpretação associada à cobertura temporal. Da mesma forma, a inclusão apenas de registros que possuem espécimes depositados em coleções biológicas pode negligenciar o papel da ciência cidadã na construção do conhecimento ao longo do espaço e do tempo.

Os Vieses

As limitações dos dados geram as lacunas no conhecimento. Por sua vez, as lacunas podem se propagar para os bancos de dados em forma de vieses. Aqui, definimos vieses como o conhecimento desigual de parte dos dados em relação à totalidade dos dados existentes. Os vieses podem ocorrer na hiper ou sub amostragem de táxons, regiões geográficas e períodos temporais estudados. Os vieses normalmente são resultados da heterogeneidade no esforço dos inventários de coletas das espécies, compilação e estudo dos registros primários ao longo do planeta (Hughes *et al.* 2021), e/ou também em função do padrão heterogêneo de distribuição da biodiversidade. Muitos estudos têm demonstrado que fatores de acessibilidade - locais próximos de vias de acesso (estradas e rios), centros de pesquisa, unidades de conservação e fragmentos florestais grandes e conectados são mais amostrados, gerando um conhecimento enviesado para estes locais (Meyer *et al.* 2015, Zizka *et al.* 2020, Sobral-Souza *et al.* 2021). Além disso, a acessibilidade e construção do conhecimento da biodiversidade são creditados também ao colonialismo científico ocorrente em cada país, níveis de riqueza, educação e estabilidade política (veja Boakes *et al.* 2010; Meyer *et al.* 2015; Raja *et al.* 2022).

Historicamente as coleções biológicas dos países europeus e da América do Norte se originaram e cresceram conforme as nações imperiais estenderam seus domínios coloniais e adquiriram uma variedade de recursos de suas colônias, o que inclui direta e indiretamente a

coleta de exemplares de animais e plantas (Zizka *et al.* 2021). Tal fato é conhecido como colonialismo científico clássico. Na modernidade, esse colonialismo ocorre na forma de uma ciência de “para-quedas” onde pesquisadores de países desenvolvidos e estruturados aparecem em países menos desenvolvidos para coletar dados e desenvolver suas pesquisas, sem trazer retorno científico e tecnológico para a comunidade local. Além disso, o material biológico coletado, em muitos casos, é depositado fora de seu país de origem (Raja *et al.* 2022). Nesse sentido, como os períodos de colonização e expedições foram diferentes ao longo do planeta, o colonialismo científico foi/é mais forte em algumas regiões, gerando vieses espaciais, taxonômicos e temporais (Zizka *et al.* 2021). Além disso, a disponibilização dos registros primários, mesmo estando depositados em museus, pode ocorrer de forma diferente devido a questões ligadas à macroeconomia (Meyer *et al.* 2015) e sigilo nos dados.

Os vieses podem ser divididos em pelo menos três categorias. Os vieses taxonômicos indicam que alguns táxons são mais estudados do que outros em consequência de maior financiamento, à melhor facilidade e rapidez de amostragem, maior número de especialistas, maior diversidade de espécies dentro dos grupos taxonômicos (Troudet *et al.* 2017, Siracusa *et al.* 2020) e/ou maior necessidade de conhecimentos sobre certos táxons visando políticas públicas para conservação e/ou de serviços ecossistêmicos essenciais para a manutenção do bem-estar humano. Os vieses espaciais, por sua vez, são relacionados a amostragem da biodiversidade de forma desigual ao longo do espaço geográfico. Isso resulta em lugares mais amostrados do que outros, devido a fatores como acessibilidade, condições socioeconômicas, de infraestrutura e de condições paisagísticas (Hortal *et al.* 2008, Meyer *et al.* 2015, Meyer *et al.* 2016, Sobral-Souza *et al.* 2021). Os vieses temporais descrevem que existem determinados períodos temporais que são mais estudados do que outros, devido questões ligadas ao esforço de amostragem, sazonalidade climática, grupos taxonômicos estudados e ao colonialismo científico (clássico e moderno) (Boakes *et al.* 2010, Stropp *et al.* 2016, Meyer *et al.* 2016, Ronquillo *et al.* 2020) (Figura 2).

Os Déficits

Os déficits - a diferença entre o que é esperado em termos de conhecimento da biodiversidade e o que realmente está disponível nos dados - podem estar presentes nas identificações taxonômicas, nas informações de localização espacial e temporal dos registros primários, bem como em todos os outros tipos de informações além da tríade taxonomia, espaço e tempo (p.ex., informações acústicas, de traço funcional ou genéticas).

Embora existam alguns métodos para descrever o déficit taxonômico dentro dos inventários (LW) (Soberón *et al.* 2007) e acerca da quantidade de espécies/gênero/famílias a serem descobertos (LL) (Mora *et al.* 2011, Freitas *et al.* 2021), um dos mais simples é representar a proporção de identificações em níveis taxonômicos supraespecífico (gênero, família e Ordem) em relação a todos os registros existentes (Castro-Souza *et al.* submetido). Quanto maior a proporção de registros supraespecífico no banco de dados, maior é o déficit taxonômico. O ideal é que o banco de dados contenha todos os registros à nível específico. Contudo, a exclusão dos dados supraespecíficos impossibilita o cálculo do déficit taxonômico e negligencia a incerteza taxonômica dos dados e dos padrões de biodiversidade. Tal problemática nasce dentro do sistema de classificação linneano que possui níveis mais “refinados” (p.ex., categorias de espécies e subespécies), enquanto outros são mais “amplos” (p.ex., categorias de Ordens, Classes e Filos). Para buscar padrões biogeográficos e estratégias de conservação biogeográfica, muitas vezes somos indagados a resolver questões que utilizam apenas partículas de conhecimento mais “refinadas” (Hortal *et al.* 2015).

Os déficits espaciais podem ser representados pelos registros que possuem indicação equivocadas, ou faltantes, sobre o local amostrado. Consequentemente o déficit espacial não permite apontar com acurácia geográfica onde o exemplar foi obtido. Em muitos casos os registros possuem localidades descritivas em níveis de município, estado ou país, ou informações ausentes (Rocchini *et al.* 2011, Ronquillo *et al.* 2020). Alguns outros, sendo representados principalmente por registros antigos, colocam pontes, comunidades tradicionais e fazendas como locais de coleta. Contudo, muitos locais não existem mais nos dias atuais, sendo necessário o estudo de documentos históricos, como o Guia Postal-Telegráfico do país, compilados sobre localidades descritas na literatura científica (veja Stephens & Traylor 1983) ou trabalhos que abordem a vida e história dos grandes coletores (veja Papavero 1971, 1973, Lamas 1980, Horn *et al.* 1990a, b, Evenhuis 1997). Essas deficiências nos dados primários podem muitas vezes resultar na não utilização do registro nas análises biogeográficas, transformando dados já escassos em mais escassos ainda.

Por fim, os déficits temporais podem ser representados pelos dados que contém informações incompletas a respeito de quando um espécime foi registrado, não permitindo identificar a data de sua coleta (Figura 2), ou até mesmo alguns dados mais refinados, como horário da observação (p.ex., estudos de horário de atividade e forrageamento) e/ou estações em que os dados foram coletados (p.ex., seca e chuvosa). Embora aqui nós tenhamos separados os déficits das lacunas, limitações e vieses, eles estão interligados e devem ser estudados e sumarizados de forma indissociável e não isolada.

Os Ruídos

Além das problemáticas mencionadas, existem ainda ruídos que podem estar presentes nos bancos de dados que sintetizam o conhecimento da biodiversidade. Ruídos são os “erros” que afetam a qualidade taxonômica, espacial e temporal dos dados e que podem surgir em diferentes fases da curadoria, digitalização ou coleta dos dados (Meyer *et al.* 2016). O ruído taxonômico se apresenta a partir da existência de nomes das espécies incorretos devido a rearranjos taxonômicos que ocorreram no passado e não foram atualizados nos registros ou bases de dados (Hey *et al.* 2003, Meyer *et al.* 2016). Os erros de digitação e digitalização ao longo da construção do banco de dados também podem fazer com que novas entidades taxonômicas, não existentes no mundo real, sejam criadas (mesmo não reconhecidas pelo código de nomenclatura).

O ruído de identificação surge quando as identificações taxonômicas não foram feitas por especialistas do grupo e, conseqüentemente, estão equivocadas (Scott & Hallam 2003, Meyer *et al.* 2016). Esse problema é complexo, uma vez que existem poucos especialistas para a maioria dos táxons e muitas das identificações em bancos de dados podem não conter a autoria de quem realizou ou revisou as identificações (Meyer *et al.* 2016). Além disso, as identificações conduzidas por especialistas podem estar erradas, considerando que o especialista pode estar iniciando sua carreira, ou no caso de serem espécies crípticas. Nesse sentido, iniciativas como o iNaturalist representam uma possível solução, pois permitem conhecer o grau de confiança da identificação do registro através de um consenso da sua comunidade de especialistas, naturalistas e fotógrafos. Desta forma, iniciativas como esta poderiam ser adaptadas e incorporadas em outras fontes ou até mesmo nas coleções biológicas, desde que possuam imagens ou alguma evidência da espécie acessível por outros pesquisadores, como é o caso de sinais acústicos e sequências genéticas de DNA mitocondrial. No caso das coleções, mais de um especialista poderia identificar a mesma espécie, em busca de um consenso.

O ruído de coordenadas ocorre quando as coordenadas geográficas e as localidades descritivas dos registros foram medidas ou denominadas de forma incorreta, devido a problemas com o uso de aparelhos de registros, como GPS e celulares, e principalmente devido ao uso incorreto de *datum* e projeção geográfica, bem como descrições errôneas sobre o local onde as espécies foram observadas (Murphey *et al.* 2004, Meyer *et al.* 2016), ou até mesmo transcritas de forma errônea das etiquetas originais. Esse ruído pode ser mais comum em dados antigos, onde não existiam aparelhos de GPS para determinar a localização exata da coleta.

Além disso, muitos aparelhos de GPS podem possuir uma “*Selective Availability*” (SA) para degradar intencionalmente a precisão do sinal para usuários civis (veja Pace *et al.* 1995). Por fim, existe também o ruído de data, que acontece quando as datas de coleta estão incorretas, seja por terem sido marcadas e/ou transcritas de forma errônea, ou até mesmo pela imprecisão fornecida pelo coletor (Figura 2).

Além disso, os déficits e ruídos supracitados podem ser responsáveis pela formação de muitos vieses, tendo em vista que ao descartar/filtrar dados problemáticos, podemos estar favorecendo determinados táxons, localidades, períodos temporais e/ou até autores específicos (veja Ronquillo *et al.* 2023).

DISCUSSÃO

Desafiando a Ignorância do Conhecimento

Se as lacunas, limitações, vieses, déficits e ruídos fossem desconsiderados, estaríamos assumindo que toda a biodiversidade na Terra foi amostrada e compreendida e as informações contidas nos bancos de dados estão corretas. Dificilmente isso é verdade. Portanto, destacamos que a chave é estar sempre atento a pelo menos dois tipos de problemas inerentes aos bancos de dados: a curadoria dos bancos de dados e a ignorância existente no conhecimento que eles representam. Levá-los em consideração ao avaliar quaisquer padrões ecológicos e/ou realizar análises fundamentadas no conhecimento digitalmente acessível é importante (Figura 3).

A curadoria do banco de dados é um processo demorado, complexo e extremamente essencial. Saber quais dados serão filtrados (excluídos) e quais permanecerão nas análises é fundamental. Contudo, excluir dados primários pode aumentar os vieses e lacunas; mantê-los pode aumentar os ruídos e déficits. A decisão dos processos de filtragem dos registros primários é diretamente dependente dos objetivos e hipóteses do trabalho que utilizará tais dados. Logo, o mesmo banco de dados original (pré-filtragem) poderá ser usado de forma diferente, com diferentes filtragens, em diferentes trabalhos. Sugerimos a confecção de metadados da tabela final (filtrada) que contenha os processos detalhados (com as explicações) das filtragens realizadas para que se saiba as decisões tomadas ao longo da curadoria dos dados (veja por exemplo, Ronquillo *et al.* 2020). Indicamos também que cada filtro executado dê origem a nova tabela. Assim, todos os passos ficarão registrados, permitindo a utilização do mesmo banco de dados por diferentes pesquisadores.

Uma solução proposta para compreender a ignorância existente no conhecimento e integrar as incertezas das informações dos registros primários nos trabalhos é construir mapas de ignorância (Boggs 1949, Rocchini *et al.* 2011, Ladle & Hortal 2013, Tessarolo *et al.* 2021). Esses mapas possibilitam a compreensão do que realmente sabemos e o que de fato não sabemos sobre o conhecimento da biodiversidade (Boggs 1949). Como na maioria das vezes os dados de biodiversidade apresentam algum tipo de inconsistência (Meyer *et al.* 2016), mapeá-las pode ser a base de um conhecimento mais sólido nos dias de hoje. Esses mapas podem também basear decisões acerca de investimentos para inventários e quais os táxons que devem ser priorizados. Entender os vieses taxonômicos e espaciais, em conjunto, por exemplo, pode salientar quais regiões geográficas e quais táxons estão superamostrados e quais possuem regiões e táxons subamostrados. Isso permite ordenar e priorizar locais e táxons para amostragens futuras (SobralSouza *et al.* 2021).

Existe uma variedade de métodos que nos permitem compreender parte da ignorância dos dados da biodiversidade, tanto em termos taxonômicos, espaciais e/ou temporais (veja Soberón *et al.* 2007, Sousa-Baena *et al.* 2014, Stropp *et al.* 2016, Meyer *et al.* 2015, Meyer *et al.* 2016, Lobo *et al.* 2018, Cornwell *et al.* 2019, Zizka *et al.* 2020, Freitas *et al.* 2021, Tessarolo *et al.* 2021). Nesse sentido, podemos delimitar nossa problemática científica, como, por exemplo, qual lacuna, qual limitação, qual viés, qual déficit, ou qual ruído da biodiversidade ou biogeografia desejamos investigar. Compilamos todo o conhecimento existente (registros primários) e realizamos uma curadoria nestes dados (filtragem) para poder construir mapas que permitirão representar essas diferentes questões, seja de forma isolada ou combinada em um único mapa de ignorância (veja por exemplo, Boggs 1949, Meyer *et al.* 2016, Stropp *et al.* 2016, Tessarolo *et al.* 2021). Desta forma, compreender, mapear e inserir a ignorância existente nos bancos de dados que sumarizam o conhecimento da biodiversidade dentro de pesquisas ecológicas pode ser algo extremamente valioso, contribuindo para o uso crítico dos bancos de dados e a realização de pesquisas ainda mais sólidas que considerem as incompletudes associadas aos dados.

Desafios Futuros

Uma coleta de dados de biodiversidade eficiente, que permita diminuir a(s) lacuna(s), é um dos desafios que surgem. Para isso, torna-se necessário o uso de protocolos de coleta sistematizados (p.ex., abrangendo aspectos taxonômicos, espaciais e temporais) que podem variar conforme o grupo estudado. Nesse sentido, apenas o uso do conhecimento digital

acessível (Sousa-Baena *et al.* 2014) pode não ser suficiente para o(s) grupo(s) ou região considerada(s) como modelo, tendo em vista que grande parte das coleções biológicas de diferentes partes do mundo não são de domínio público e/ou não estão digitalizadas e disponíveis online (Krishtalka & Humphrey 2000; Nelson & Ellis 2018; Bakker *et al.* 2020). Adicionalmente, os dados disponíveis na literatura taxonômica científica e literatura cinza não necessariamente significam que toda distribuição relatada nessas fontes esteja digitalizada em bancos de dados online (veja Heidorn 2008; Sikes *et al.* 2016). No entanto, como a literatura cinza não passa por uma revisão por pares, é comum que as listas de espécies contenham identificações equivocadas, tenham utilizado metodologias inadequadas para coleta dos dados e/ou tenham sido conduzidas por um profissional inexperiente para identificar espécies. Nesse contexto, a inclusão de dados da literatura cinza poderia trazer mais limitações e ruídos do que soluções, embora possa existir casos que contenha informações de boa qualidade. Assim, torna-se necessária a realização de novas coletas de dados primários em campo ou visitas a coleções biológicas ou acervos de materiais ainda não digitalizados. Isso pode, consequentemente, requerer financiamento de projeto, e dependendo do país ou instituição, pode tornar-se inviável.

Também é crucial conseguir compreender de forma precisa “*que espécie é essa?*”. O taxonomista é o principal profissional responsável por responder essa questão, a partir da descrição, nomeação e elaboração de material científico necessários para identificação de cada uma das unidades taxonômicas (nossos registros primários). Como exemplo prático, a taxonomia desempenha um papel fundamental na identificação de pragas agrícolas e espécies exóticas, evitando gastos e esforços inúteis. Além disso, é crucial na medicina, onde a identificação correta de organismos, como larvas de moscas, é vital para tratamentos utilizado na chamada terapia larval (Masiero *et al.* 2015). Ainda, a correta identificação de um organismo, como por exemplo um inseto, pode auxiliar os estudos periciais de casos envolvendo morte violenta, onde a partir da biologia do inseto estudado, pode-se determinar o intervalo pós-morte (Benecke 2001, Löbl *et al.* 2023).

No entanto, mesmo com sua notória importância no campo das ciências naturais, a taxonomia tem sido historicamente negligenciada, contando com um número cada vez menor de taxonomistas no mercado de trabalho (Wheeler 2014, Löbl *et al.* 2023), baixo impacto de suas publicações (Krell 2012), baixa perspectivas de emprego (Disney 1998, Jaspars 1998, Pearson *et al.* 2011, Salvador *et al.* 2022) e estímulos econômicos pífios para formação de novos taxônomos (Wheeler 1995, Disney 1998, Bortolus 2008, Werner 2009, Boero 2010, Salvador *et al.* 2022) frente as áreas de tecnologia e inovação.

A formação de um taxonomista, assim como de outros profissionais que atuam na biodiversidade, demanda tempo e dedicação, além de ocorrer em uma velocidade muito menor do que a capacidade dos problemas a serem resolvidos (Fontaine *et al.* 2012). Não que em outras áreas seja diferente (p.ex., genética molecular, estatística e ecoinformática), contudo, no caso da taxonomia, muitas vezes esse processo não conta com uma estrutura tecnológica mais avançada que pode ajudar a poupar um pouco de tempo. Essa incompatibilidade entre a velocidade que a sociedade contemporânea quer para as respostas e o tempo que as evidências existentes na biodiversidade demoram para aparecer, tornam essa formação cada vez mais desafiadora. Em um contexto de taxonomia integrativa (Schlick-Steiner 2010), os taxonomistas “modernos” necessitam, muitas vezes, possuir conhecimentos abrangentes de uma vasta literatura taxonômica (e histórica) em diferentes idiomas, e, além disso, dominar métodos de análises de filogenética, microscopia, e, especialmente no contexto atual, compreender as aplicações de marcadores moleculares em sistemática (Wägele 2011). Mesmo dando o devido destaque ao taxonomista, o ecólogo, geneticista, e outros pesquisadores da biodiversidade, são limitados pelo tempo e precisam estabelecer estratégias eficientes para conciliar as complexas dificuldades e responder as perguntas científicas. Nesse quesito, conseguir inovações nas linhas de pesquisa em biodiversidade parece ser um caminho.

Com o advento da era da bioinformática e ecoinformática, a formação de novos taxonomistas e cientistas de base tem se tornado ainda mais comprometida. Atualmente vivemos em um contexto científico no qual se a informação existe, ela deve ser disponibilizada para a sociedade (Boero 2010). Isso se traduz no enviesamento dos recursos financeiros para iniciativas que objetivam construir, armazenar e gerenciar conjuntos de dados digitalmente acessíveis que sumarizam diferentes facetas da biodiversidade, seja em escala local, regional ou global (Bingham *et al.* 2017). Além disso, iniciativas deste tipo, como por exemplo, “*Catalogue of Life*” ([https:// www.catalogueoflife.org/](https://www.catalogueoflife.org/)) e “*Global Biodiversity Information Facility*” (<https://www.gbif.org/>) dão a impressão equivocada de que a taxonomia recebe um apoio eficiente e suficiente (Löbl *et al.* 2023). Essa “tendência” parece ter alocado parte dos recursos disponíveis para a conservação da biodiversidade para o financiamento de propostas que objetivam o surgimento de novas tecnologias, como a bioinformática (Löbl *et al.* 2023). Em contrapartida, os recursos destinados aos estudos taxonômicos, de história natural e de ciência de base, foram diminuindo (Boero 2010) mantendo as lacunas no conhecimento. Logo, a integração entre as áreas da biodiversidade é necessária como forma de (i) demonstrar quais

as lacunas de conhecimento existentes e (ii) preenchê-las. Sem a integração entre as áreas, e com o isolamento delas, a inovação nas pesquisas ficará cada vez mais distante.

Outro desafio que surge é a curadoria contínua dos bancos de dados, que nesse caso inclui tanto as coleções biológicas físicas quanto os repositórios online. Assim, torna-se de extrema urgência a obtenção de subsídios financeiros, tanto para manutenção da infraestrutura física das coleções, mas também a fim de organizar e manter todos os espécimes em bom estado, o que se traduz na contratação de técnicos capacitados tanto em taxonomia quanto em bioinformática. Além disso, também é necessária uma infraestrutura de informática para a mobilização de dados em coleções biológicas de forma aberta e colaborativa (Krishtalka & Humphrey 2000, Nelson & Ellis 2018, Johnson *et al.* 2023), seja por meio de políticas locais de cada país ou acordos globais, como a Convenção sobre Diversidade Biológica (CDB) (<<https://www.cbd.int/convention/>>). Isso é especialmente importante considerando que acidentes em coleções biológicas resultam em perdas inestimáveis (Phillips, 2010; Phillips, 2018), principalmente quando não temos informações digitalizadas.

CONCLUSÕES

Por fim, reforçamos que novas coletas de dados são importantes para obtenção de mais informações e preenchimento das lacunas. O taxonomista é importante para descrever e detalhar informações sobre as espécies (p.ex., distribuição geográfica, história natural, habitat, entre outras). Os ecólogos são importantes para entender os processos locais e regionais que mantem e estruturam a biodiversidade. Esses fatos remontam a necessidade da interação entre as áreas. A curadoria dos dados e a bioinformática também são importantes para sumarizar e disponibilizar essa informação em grandes bancos de dados que sejam úteis para projetos que objetivem as tomadas de decisão e a descrição de padrões em larga escala. Contudo, deve-se ter em mente que os bancos de dados construídos, por serem baseados nos dados primários, sempre possuem lacunas, limitações, vieses, déficits e ruídos. Estas deficiências não podem ser negligenciadas e devem ser estudadas e discutidas entre taxonomistas, ecólogos, e diferentes pesquisadores da biodiversidade para que os padrões, processos e tomadas de decisões sejam trabalhadas em conjunto e com excelência.

AGRADECIMENTOS

Somos gratos à equipe do Laboratório de Macroecologia e Conservação da Biodiversidade (MacrEco) da Universidade Federal de Mato Grosso (UFMT) pelas valiosas discussões. Agradecemos a FAPEMAT (projeto FAPEMAT-PRO.000274/2023) e a

IABS/CECAV (edital 01/2023 - TCCE Vale 01/2022) pelo financiamento. RACS, NB e AHK agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas bolsas de doutorado concedidas e ao programa de Pós-graduação em Ecologia e Conservação da Biodiversidade (PPGECB). VCS agradece ao Programa de PósDoutorado da University of Pretoria (África do Sul). RACS, JAFDF, GT e TSS agradecem ao Instituto Nacional de Ciência & Tecnologia (INCT) em Ecologia, Evolução e Conservação da Biodiversidade (EECBio), sediado na Universidade Federal de Goiás (UFG), pelas preciosas discussões que contribuíram para o desenvolvimento deste artigo.

REFERÊNCIAS

- Ahrends, A., Rahbek, C., Bulling, MT., Burgess, ND., Platts, PJ., Lovett, JC., Kindemba, VW., Owen, N., Sallu, AN., Marshall, AR., Mhoro, BE., Fanning, E., & Marchant, R. 2011. Conservation and the botanist effect. *Biological Conservation*, 144(1), 131–140. DOI: 10.1016/j.biocon.2010.08.008
- Almeida, Almeida, LM., Ribeiro-Costa, CS., & Marinoni, L. 2012. Coleta, montagem, preservação e métodos para estudo [pp. 175-190]. In: Rafael, JA., Melo, GAR., Carvalho CJB., Casari, AS, & Constantino, R (Eds). *Insetos do Brasil: diversidade e taxonomia*, Holos Editora, Ribeirão Preto (SP), 810 p.
- Alves, CBM., Pompeu, PS., Mazzoni, R., & Brito, MFG. 2021. Avanços em métodos de coleta de peixes e caracterização de habitat de riachos tropicais. *Oecologia Australis*, 25(2), 246–265. DOI: 10.4257/oeco.2021.2502.03
- Appel, G., Capaverde, UD., de Oliveira, LQ., Pereira, LGA., Tavares, VC., López-Baucells, A., Magnusson W E., Baccaro FB., & Bobrowiec, PE. 2021. Use of complementary methods to sample bats in the Amazon. *Acta Chiropterologica*, 23(2), 499–511. DOI: 10.3161/15081109ACC2021.23.2.017
- Disney, H. 1998. Rescue plan needed for taxonomy. *Nature*, 394(6689), 120–120. DOI: 10.1038/28027
- Freckleton, R.P., & Watkinson, A.R. 2002. Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. *Journal of Ecology*, 90(3), 419–434. DOI: 10.1046/j.1365-2745.2002.00692.x
- Ficetola, GF., Canedoli, C., & Stoch, F. 2019. The Racovitzan impediment and the hidden biodiversity of unexplored environments. *Conservation Biology*, 33(1), 214–216. DOI: 10.1111/cobi.13179
- Freitas, TMDS., Stropp, J., Calegari, BB., Calatayud, J., De Marco Jr, P., Montag, LFDA., & Hortal, J. 2021. Quantifying shortfalls in the knowledge on Neotropical Auchenipteridae fishes. *Fish and Fisheries*, 22(1), 87–104. DOI: 10.1111/faf.12507
- Freeman, B.G., & Pennell, M.W. 2021. The latitudinal taxonomy gradient. *Trends in Ecology & Evolution*, 36(9), 778–786. DOI: 10.1016/j.tree.2021.05.003

- Bach, A., Mateus, L.A., Peres, C. A., Haugaasen, T., Louzada, J., Hawes, J.E., Azevedo, R.A., Lucena, E., Ferreira, J.V.A., & Vaz-de-Mello, F.Z. 2023. Bait attractiveness changes community metrics in dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae). *Ecology and Evolution*, 13(4), e9975. DOI: 10.1002/ece3.9975
- Bakker, F.T., Antonelli, A., Clarke, J.A., Cook, J.A., Edwards, S.V., Ericson, P.G.P., Faurby, S., Ferrand, N., Gelang, M., Gillespie, R.G., Irestedt, M., Lundin, K., Larsson, E., Matos-Maraví, P., Müller, J., von Proschwitz, T., Roderick, G.K., Schliep, A., Wahlberg, N., Wiedenhoeft, J., & Källersjö, M. 2020. The Global Museum: natural history collections and the future of evolutionary science and public education. *PeerJ*, 8, e8225. DOI: 10.7717/peerj.8225
- Benecke, M. 2001. A brief history of forensic entomology. *Forensic Science International*, 120, 2–114.
- Bingham, H.C., Doudin, M., Weatherdon, L.V., Despot-Belmonte, K., Wetzel, F.T., Groom, Q., Lewis, E., Regan, E., Appeltans, W., Güntsch, A., Mergen, P., Agosti, D., Penev, L., Hoffmann, A., Saarenmaa, H., Geller, G., Kim, K., Kim, H., Archambeau, A.S., Häuser, C., Schmeller, D.S., Geijzendorffer, I., Camacho, Guerra, A.G.C., Robertson, T., Runnel, V., Volland, N., & Martin, C.S. 2017. The biodiversity informatics landscape: elements, connections and opportunities. *Research Ideas and Outcomes*, 3, e14059. DOI: 10.3897/rio.3.e14059
- Boakes, E.H., McGowan, P.J., Fuller, R.A., Changqing, D., Clark, N.E., O'Connor, K., & Mace, G.M. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS biology*, 8(6), e1000385. DOI: 10.1371/journal.pbio.1000385
- Boero, F. 2010. The study of species in the era of biodiversity: a tale of stupidity. *Diversity*, 2(1), 115–126. DOI: 10.3390/d2010115
- Boggs, S.W. 1949. An atlas of ignorance: a needed stimulus to honest thinking and hard work. *Proceedings of the American Philosophical Society*, 93(3), 253–258.
- Bortolus, A. 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *AMBIO: A journal of the human environment*, 37(2), 114–118. DOI: 10.1579/0044-7447(2008)37[114:ECITBS]2.0.CO;2
- Cardoso, P., Erwin, T.L., Borges, P.A., & New, T.R. 2011. The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, 144(11), 2647–2655. DOI: 10.1016/j.biocon.2011.07.024
- Cornwell, W.K., Pearse, W.D., Dalrymple, R.L., & Zanne, A.E. 2019. What we (don't) know about global plant diversity. *Ecography*, 42(11), 1819–1831. DOI: 10.1111/ecog.04481
- Coyne J.A., & Orr H.A. 2004. Species: Reality and Concepts. In: J. A. Coyne & H. A. Orr (Eds.), *Speciation*. pp. 9–54. Sunderland, MA: Sinauer Associates.
- Diniz-Filho, J.A.F., Loyola, R.D., Raia, P., Mooers, A.O., & Bini, L.M. 2013. Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*, 28(12), 689–695. DOI: https://doi.org/10.1016/j.tree.2013.09.003
- 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., Carvalho, C.J.B., Ladle, R.J., & Hortal, J. 2023. Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Frontiers of Biogeography*, 15(2), e59566. DOI:10.21425/F5FBG59566

- Disney, RHL. 1998. Insect biodiversity and the demise of alpha taxonomy. *Antenna*, 23, 84–88.
- Escribano, N., Ariño, AH., & Galicia, D. 2016. Biodiversity data obsolescence and land uses changes. *PeerJ*, 4, e2743. DOI: 10.7717/peerj.2743
- Evenhuis, NL. 1997. *Litteratura taxonomica dipterorum (1758–1930)*. Volume II. L–Z. Backhuys, Leiden, pp. 427–871.
- Fontaine, B., Perrard, A., & Bouchet, P. 2012. 21 years of shelf life between discovery and description of new species. *Current Biology*, 22(22), R943–R944. DOI: 10.1016/j.cub.2012.10.029
- Freitas, TMDS., Stropp, J., Calegari, BB., Calatayud, J., De Marco Jr, P., Montag, LFDA., & Hortal, J. 2021. Quantifying shortfalls in the knowledge on Neotropical Auchenipteridae fishes. *Fish and Fisheries*, 22(1), 87–104. DOI: 10.1111/faf.12507
- Greshko, M. 2018. Fire Devastates Brazil's Oldest Science Museum. *National Geographic* (diponível: www.nationalgeographic.com/science/article/news-museunacional-fire-riode-janeiro-natural-history).
- Gutierrez-Velez, VH., & Wiese, D. 2020. Sampling bias mitigation for species occurrence modeling using machine learning methods. *Ecological Informatics*, 58, 101091. DOI: 10.1016/j.ecoinf.2020.101091
- Heidorn, PB. 2008. Shedding light on the dark data in the long tail of science. *Library Trends*, 57(2), 280–299. DOI: 10.1353/lib.0.0036
- Hey, J., Waples, RS., Arnold, ML., Butlin, RK. & Harrison, RG. 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution*, 18(11), 597–603. DOI: 10.1016/j.tree.2003.08.014
- Horn, W., Kahle, I., Friese, G. & Gaedike, R. 1990a. *Collectiones entomologicae. Ein Kompendium über den Verbleib entomologischer Sammlungen der Welt bis 1960. Teil I: A bis K*. Berlin (Akademie der Landwirtschaftswissenschaften der Deutschen Demokratischen Republik), pp. 1–220.
- Horn, W., Kahle, I., Friese, G. & Gaedike, R. 1990b. *Collectiones entomologicae. Ein Kompendium über den Verbleib entomologischer Sammlungen der Welt bis 1960. Teil II: L bis Z*. Berlin (Akademie der Landwirtschaftswissenschaften der Deutschen Demokratischen Republik), pp. 221–573.
- Hortal, J., Jiménez-Valverde, A., Gómez, JF., Lobo, JM., & Baselga, A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, 117(6), 847–858. DOI: 10.1111/j.0030-1299.2008.16434.x
- Hortal, J., de Bello, F., Diniz-Filho, JAF., Lewinsohn, TM., Lobo, JM., & Ladle, RJ. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549. DOI: 10.1146/annurev-ecolsys-112414-054400
- Hughes, AC., Orr, MC., Ma, K., Costello, MJ., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. 2021. Sampling biases shape our view of the natural world. *Ecography*, 44, 1–14.

- Jaspars, M. 1998. Tough time for taxonomy. *Nature*, 394(6692), 413–413. DOI: 10.1038/28723
- Johnson, KR., Owens, IF., & Global Collection Group. 2023. A global approach for natural history museum collections. *Science*, 379(6638), 1192–1194. DOI: 10.1126/science.adf6434
- Krell, F-T. 2012. The Journal Impact Factor as a performance indicator. *European Science Editing*, 38, 3–6.
- Krishtalka, L., & Humphrey, PS. 2000. Can natural history museums capture the future?. *BioScience*, 50(7), 611–617. DOI:10.1641/0006-3568(2000)050[0611:CNHMCT]2.0.CO;2
- Lees, AC., Rosenberg, KV., Ruiz-Gutierrez, V., Marsden, S., Schulenberg, TS., & Rodewald, AD. 2020. A roadmap to identifying and filling shortfalls in Neotropical ornithology. *The Auk*, 137(4), ukaa048. DOI: 10.1093/auk/ukaa048
- Ladle, RJ., & Hortal, J. 2013. Mapping species distributions: living with uncertainty. *Frontiers of Biogeography*, 5(1), 8–9. DOI: 10.21425/F55112942
- Lamas, G. 1980. Introducción a la Historia de la Entomología en el Perú. II. Período de los viajeros, Colectores y Estudiosos Especializado. *Revista Peruana de Entomología*, 23(1), 25–31.
- Lobo, JM., Hortal, J., Yela, JL., Millán, A., SánchezFernández, D., García-Roselló, E., GonzálezDacosta, J., Heine, J., González-Vilas, L., & Guisande, C. 2018. KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecological Indicators*, 91, 241–248. DOI: 10.1016/j.ecolind.2018.03.077
- Löbl, I., Klausnitzer, B., Hartmann, M., & Krel, F-T. 2023. The silent extinction of the species and taxonomists - An appeal to Science policymakers and legislators. *Diversity*, 15(10): 1–17. DOI: 10.3390/d15101053
- Kemp, C. 2015. The endangered dead. *Nature* 518(7539), 292–294. DOI: 10.1038/518292a
- Lomolino MV. 2004. Conservation biogeography. In: M.V. Lomolino & L.R. (Eds.) Heaney *Frontiers of Biogeography: New Directions in the Geography of Nature*. pp. 293–96. Sunderland, MA: Sinauer.
- Lopes-Lima, M., Riccardi, N., Urbanska, M., Köhler, F., Vinarski, M., Bogan, AE., & Sousa, R. 2021. Major shortfalls impairing knowledge and conservation of freshwater molluscs. *Hydrobiologia*, 848(12–13), 2831–2867. DOI: doi.org/10.1007/s10750-021-04622-w
- Masiero, FS., Martins, DS., & Thyssen, PJ. 2015. Terapia Larval e a aplicação de larvas para cicatrização: revisão e estado da arte no Brasil e no mundo. *Revista Thema*, 12(1), 4–14.
- Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. 2015. Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, 6(1), 1–8. DOI: 10.1038/ncomms9221
- Meyer, C., Weigelt, P., & Kreft, H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*. 19(8), 992–1006. DOI: 10.1111/ele.12624
- Mora, C., Tittensor, DP., Adl, S., Simpson, AG., & Worm, B. 2011. How many species are there on Earth and in the ocean?. *PLoS Biology*, 9(8), e1001127. DOI: 10.1371/journal.pbio.1001127

- Murphey, PC., Guralnick, RP., Glaubitz, R., Neufeld, D., & Ryan, JA. 2004. Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains SpatioTemporal Database-Informatics Initiative (Mapstedi). *PhyloInformatics*, 3, 1–29. DOI: 10.5281/zenodo.59792
- Nelson, G., & Ellis, S. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*, 374, 20170391. DOI: 10.1098/rstb.2017.0391
- Pace, S., Frost, G., Lachow, I., Frelinger, D., Fossum, D., Wassem, D., & Pinto, MM. 1995. The global positioning system: assessing national policies. Santa Monica, CA: Rand: p. 368.
- Papavero, N. 1971. Essays on the history of Neotropical dipterology, with special reference to collectors (1750–1905). Vol. I. São Paulo: Museu de Zoologia, Universidade de São Paulo: p. 216.
- Papavero, N. 1973. Essays on the history of Neotropical dipterology, with special reference to collectors (1750–1905). Vol. II. São Paulo: Museu de Zoologia, Universidade de São Paulo: p. 217–426.
- Papavero, N. 1994. Fundamentos práticos de taxonomia zoológica. Sao Paulo: Editora Unesp. p. 285.
- Pearson, DL., Hamilton, AL., & Erwin, TL. 2011. Recovery plan for the endangered taxonomy profession. *BioScience*, 61(1), 58–63. DOI:10.1525/bio.2011.61.1.11
- Phillips T. 2010. Sao Paulo fire destroys one of the largest collections of dead snakes. *The Guardian*. <https://www.theguardian.com/world/2010/may/16/firedestroyssnakecollection>.
- Phillips T. 2018. Brazil National Museum: as much as 90% of collection destroyed in fire. *The Guardian*. <https://www.theguardian.com/world/2018/sep/04/brazilnationalmuseumfire-collection-destroyed-not-insured>.
- Raven, PH., & Wilson, EO. 1992. A fifty-year plan for biodiversity surveys. *Science*, 258(5085), 1099–1100. DOI: 10.1126/science.258.5085.1099
- Raja, NB., Dunne, EM., Matiwane, A., Khan, TM., Nätscher, PS., Ghilardi, AM., & Chattopadhyay, D. 2022. Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature ecology & evolution*, 6(2), 145–154. DOI: <https://doi.org/10.1038/s41559-021-01608-8>.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, JM., Jimenez-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. DOI: 10.1177/0309133311399491
- Ronquillo, C., Alves-Martins, F., Mazimpaka, V., Sobral-Souza, T., Vilela-Silva, B., Medina, N. G., & Hortal, J. 2020. Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses. *Biodiversity Data Journal*, 8, e53474. DOI: 10.3897/BDJ.8.e53474
- Ronquillo, C., Stropp, J., Medina, NG., & Hortal, J. 2023. Exploring the impact of data curation criteria on the observed geographical distribution of mosses. *Ecology and Evolution*, 13(12), e10786. DOI: 10.1002/ece3.10786

- Rosado, BH., Figueiredo, MS., de Mattos, EA., & Grelle, CE. 2016. Eltonian shortfall due to the Grinnellian view: functional ecology between the mismatch of niche concepts. *Ecography*, 39(11), 1034–1041. DOI: 10.1111/ecog.01678
- Rudbeck, AV., Sun, M., Tietje, M., Gallagher, RV., Govaerts, R., Smith, SA., Svenning, JC., & Eiserhardt, WL. 2022. The Darwinian shortfall in plants: phylogenetic knowledge is driven by range size. *Ecography*, 2022: e06142. DOI: 10.1111/ecog.06142
- Salvador, RB., Cavallari, DC., Rands, D., & Tomotani, BM. 2022. Publication practice in Taxonomy: Global inequalities and potential bias against negative results. *Plos one*, 17(6), e0269246. DOI: 10.1371/journal.pone.0269246
- Schlick-Steiner, BC., Steiner, FM., Seifert, B., Stauffer, C., Christian, E., Cozier, RH. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology* 55, 421–438. DOI: 10.1146/annurevento-112408-085432
- Scott, WA., & Hallam, CJ. 2003. Assessing species misidentification rates through quality assurance of vegetation monitoring. *Plant Ecology*, 165(1), 101–115. DOI: 10.1023/A:1021441331839
- Sikes, DS., Copas, K., Hirsch, T., Longino, JT., & Schigel, D. 2016. On natural history collections, digitized and not: a response to Ferro and Flick. *ZooKeys*, 618, 145–158. DOI: 10.3897/zookeys.618.9986
- Siracusa, PC., Gadelha, LM., & Ziviani, A. 2020. New perspectives on analysing data from biological collections based on social network analytics. *Scientific reports*, 10(1), 1–10. DOI: 10.1038/s41598-020-60134-y
- Soberón, J., & Peterson, T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1444), 689–698. DOI: 10.1098/rstb.2003.1439
- Soberón, J., Jiménez, R., Golubov, J., & Koleff, P. 2007. Assessing completeness of biodiversity databases at different spatial scales. *Ecography*, 30(1), 152–160. DOI: 10.1111/j.0906-7590.2007.04627.x
- Soberón, J., & Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, 106(supplement_2), 19644–19650.
- Sobral-Souza, T., Bastos Francini, R., Guimarães, M., & Benson, WW. 2015. Short-term dynamics reveals seasonality in a subtropical *Heliconius* butterfly. *Journal of Insects*, 761058, 1–5. DOI: 10.1155/2015/761058
- Sobral-Souza, T., Stropp, J., Santos, JP., Prasniewski, VM., Szinwelski, N., Vilela, B., ... & Hortal, J. 2021. Knowledge gaps hamper understanding the relationship between fragmentation and biodiversity loss: the case of Atlantic Forest fruit-feeding butterflies. *PeerJ*, 9, e11673. DOI: 10.7717/peerj.11673
- Sousa-Baena, MS., Garcia, LC., & Peterson, AT. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions*, 20(4), 369–381. DOI: 10.1111/ddi.12136

- Stephens, L., & Traylor, MA. 1983. Ornithological gazetteer of Peru. 2ed. Obtainable from Bird Dept., Museum of Comparative Zoology: Harvard University, vi: p. 273.
- Stork, NE. 2018. How many species of insects and other terrestrial arthropods are there on earth? *Annual Review of Entomology*, 63, 31–45. DOI: 10.1146/annurev-ento-020117-043348
- Stropp, J., Ladle, RJ., Malhado, ACM., Hortal, J., Gaffuri, J., Temperley, WH., Skøien, JO., Mayaux, P. 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. *Global Ecology and Biogeography*, 25(9), 1085–1096. DOI: 10.1111/geb.12468
- Stropp, J., Ladle, RJ., Emilio, T., Lessa, T., & Hortal, J. 2022. Taxonomic uncertainty and the challenge of estimating global species richness. *Journal of Biogeography*, 49, 1654–1656. DOI: 10.1111/jbi.14463
- Tessarolo, G., Ladle, R., Rangel, T., & Hortal, J. 2017. Temporal degradation of data limits biodiversity research. *Ecology and Evolution*, 7(17), 6863–6870. DOI: 10.1002/ece3.3259
- Tessarolo, G., Ladle, RJ., Lobo, JM., Rangel, TF., & 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. DOI: 10.1111/ecog.05793
- TrouDET, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. 2017. Taxonomic bias in biodiversity data and societal preferences. *Scientific reports*, 7(1), 1–14. DOI: 10.1038/s41598-017-09084-6
- Wägele, H., Klusmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A., & Wägele, JW. 2011. The taxonomist-an endangered race. A practical proposal for its survival. *Frontiers in zoology*, 8(1), 1–7. DOI: 10.1186/1742-9994-8-25
- Werner, YL. 2009. The aspiration to be good is bad: The “Impact Factor” hurts both science and society. *The International Journal of Science and Society*, 1, 99–105. DOI: 10.18848/1836-6236/CGP/v01i01/51264
- Wheeler, Q. 2014. Are reports of the death of taxonomy an exaggeration?. *New Phytologist*, 201(2), 370–371. DOI: 10.1111/nph.12612
- Wheeler, QD. 1995. The “Old Systematics”: Classification and phylogeny. In: *Biology, Phylogeny and Classification of Coleoptera: Papers celebrating the 80th Birthday of Roy A. Crowson*; Pakaluk, J., Slipiński, S.A. (Eds.); pp. 31–62. Muzeum i Institut Zoologii: Warszawa, Poland.
- Whittaker, RJ., Araújo, MB., Jepson, P., Ladle, RJ., Watson, JEM., & Willis, KJ. 2005. Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, 11: 3-23. DOI: 10.1111/j.1366-9516.2005.00143.x
- Zizka, A., Antonelli, A., & Silvestro, D. 2020. Sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography*, 44(1), 25–32. DOI: 10.1111/ecog.05102
- Zizka, A., Rydén, O., Edler, D., Klein, J., Perrigo, A., Silvestro, D., ... & Antonelli, A. 2021. Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. *Journal of Biogeography*, 48(11), 2715–2726. DOI: 10.1111/jbi.14256

Tabela 1: Síntese sobre as principais lacunas do conhecimento da biodiversidade (referências marcadas com (*) são as originais de cada lacuna).

LACUNA	FACETA DA BIODIVERSIDADE	BREVE DEFINIÇÃO	REFERÊNCIAS SUGERIDAS
Linneana	Espécies	Desconhecimento da identidade taxonômica das espécies;	(Raven & Wilson 1992*, Cardoso <i>et al.</i> 2011, Hortal <i>et al.</i> 2015)
Wallaceana	Distribuição geográfica	Desconhecimento da distribuição geográfica das espécies;	(Lomolino 2004*, Cardoso <i>et al.</i> 2011, Hortal <i>et al.</i> 2015)
Prestoniana	Populações	Desconhecimento da abundância e dinâmica populacional das espécies ao longo do espaço e do tempo;	(Cardoso <i>et al.</i> 2011*, Hortal <i>et al.</i> 2015)
Darwiniana	Evolução	Desconhecimento sobre a evolução das espécies;	(Diniz-Filho <i>et al.</i> 2013*, Hortal <i>et al.</i> 2015)
Raunkiaeriana	Características funcionais e funções ecológicas	Desconhecimento sobre as características funcionais das espécies e de suas funções ecológicas;	(Hortal <i>et al.</i> 2015*, mas veja Rosado <i>et al.</i> 2016)
Hutchinsoniana	Tolerâncias Abióticas	Desconhecimento sobre às tolerâncias ambientais das espécies à condições abióticas;	(Cardoso <i>et al.</i> 2011*, Hortal <i>et al.</i> 2015)
Eltoniana	Interações Ecológicas	Desconhecimento das interações biológicas das espécies e quais seus efeitos sobre a sobrevivência e o fitness;	(Hortal <i>et al.</i> 2015*, mas veja Rosado <i>et al.</i> 2016)
Parkeriana	História Natural	Desconhecimento sobre história natural das espécies;	(Lee <i>et al.</i> 2020*)
Ostromiana	Medidas de Conservação	Desconhecimento sobre a eficácia das medidas de conservação;	(Lopes-Lima 2021*)
Racovitzana	Exploração geográfica	Desconhecimento da biodiversidade devido a áreas não exploradas e não mapeadas;	(Ficetola <i>et al.</i> 2019*)



Figura 1. Sistematização das lacunas (1), limitações (2) e vieses (3) do conhecimento da biodiversidade, problemas diretamente associados à obtenção e coleta de dados. Em seguida, sistematizamos os déficits (4) e ruídos (5), problemas associados aos bancos de dados da biodiversidade.

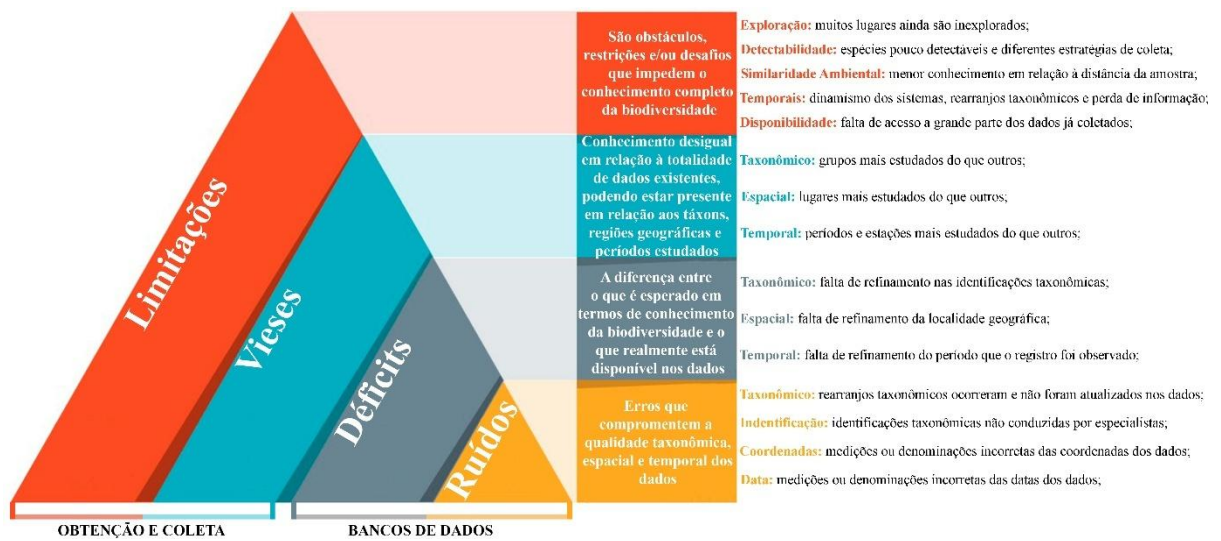


Figura 2. Síntese sobre as Limitações, Déficits, Vieses e Ruídos do conhecimento da biodiversidade (baseado em Scott & Hallam 2003, Hey et al. 2003, Murphey et al. 2004, Hortal et al. 2008, Boakes et al. 2010, Ahrends et al. 2011, Rocchini et al. 2011, Ledle & Hortal 2013, Meyer et al. 2015, Escribano et al. 2016, Meyer et al. 2016, Tessarolo et al. 2017, Troudet et al. 2017, Ficetola et al. 2019, Siracusa et al. 2020, Hughes et al. 2021, Tessarolo et al. 2021).

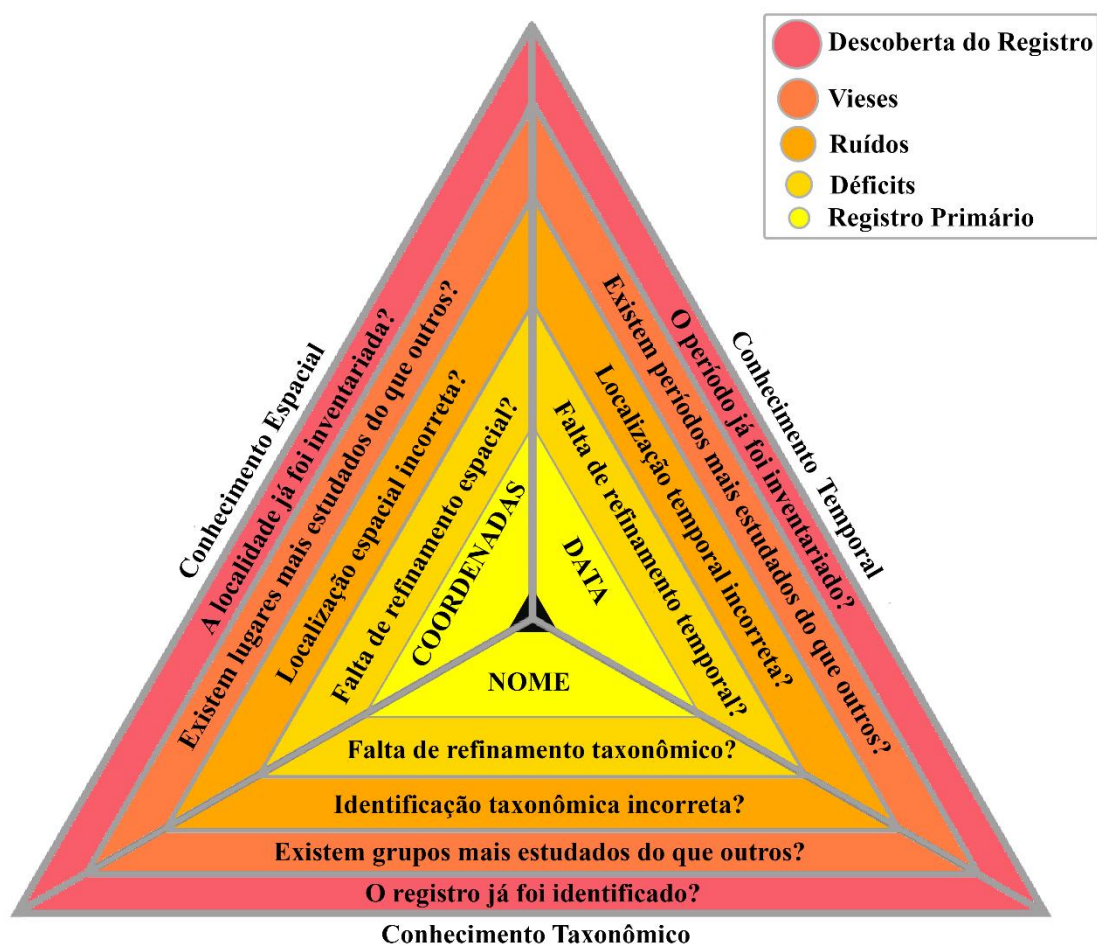


Figura 3. Perguntas que podem ajudar a identificar incompletudes do conhecimento taxonômico, espacial e temporal associadas aos registros primários, considerando o momento de sua descoberta, formação de vieses, ruídos e déficits. Aqui, não consideramos as lacunas e limitações, tendo em vista que elas são intrínsecas ao conhecimento da biodiversidade.

CAPÍTULO I

Mapping the Status of Global Taxonomic Knowledge of Orthoptera (Arthropoda, Insecta)

Manuscrito submetido e em processo de revisão na revista *Frontiers of Biogeography*

Article type: Research Article

Title: Mapping the status of global taxonomic knowledge of Orthoptera

(Arthropoda, Insecta)

Running title: Mapping the taxonomy of Orthoptera

Authors: Rodrigo Antônio Castro-Souza^{1,2,*}, Juliana Stropp³, Luiz Felipe Moretti Iniesta⁴, Richard J. Ladle⁵, Neucir Szinwelski⁶, Geiziane Tessarolo⁷, José Alexandre Diniz-Filho⁷, Thadeu Sobral-Souza², Joaquín Hortal^{8,*}

Affiliations: (1) Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Federal de Mato Grosso (UFMT), Cuiabá, Brazil.

(2) Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso, 78060-900, Cuiabá, MT, Brazil.

(3) Department of Biogeography, University of Trier, Trier, Germany.

(4) Laboratório de Coleções Zoológicas, Instituto Butantan, Av. Vital Brasil, 1500, Butantã, 05503-900 São Paulo, SP, Brazil.

(5) Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, Alagoas, Brazil.

(6) Laboratório de Orthoptera, Centro de Ciências Biológicas e da Saúde, Universidade Estadual do Oeste do Paraná, Brazil.

(7) Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil.

(8) Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain.

corresponding authors*: Rodrigo Antônio Castro-Souza,

rodrigodesouzaac@gmail.com and Joaquín Hortal, jhortal@mncn.csic.es

Abstract

The status of taxonomic knowledge varies across the globe. Quantifying and mapping the geographic patterns of taxonomic status is essential to prioritize regions that require greater attention from the taxonomic community. Here we use the date of species descriptions and the number of associated synonyms across space as proxies for alpha and beta taxonomy, and combine them to map the taxonomic knowledge of Orthoptera (Arthropoda, Insecta) species globally. We hypothesize that beta taxonomy predominates in Europe and temperate regions of the Northern Hemisphere, while alpha taxonomy prevails in tropical regions and the Southern Hemisphere. We compiled all valid orthopteran species names, and their synonyms, extracted from the Catalogue of Life (CoL), and allocated them geographically based on data from the Global Biodiversity Information Facility (GBIF) and the Orthoptera Species File (OSF), which allowed us to create measures of alpha and beta taxonomy rates and combine them spatially. Tropical regions and the Southern Hemisphere contain the majority of recently described species names with fewer associated synonyms, indicating a predominance of alpha taxonomy and highlighting the need for greater taxonomic efforts. In contrast, temperate regions, particularly in Europe and southwestern Asia, contain the majority of older names and synonyms, indicating a predominance of beta taxonomy and regions that have been more thoroughly studied. Our findings are discussed in the context of sociopolitical factors, scientific investments, and the history of taxonomy. Finally, we propose a framework that makes the links between taxonomy and macroecology accessible for biodiversity in the era of Big Data.

Highlights

- Alpha taxonomy (i.e. the description of new species) and Beta taxonomy (i.e. the revision of the taxonomic status and relationships of already described taxa) vary across space and time.
- We present a framework that connects taxonomy and macroecology, allowing to assess taxonomic trends to inform biodiversity studies in the era of Big Data.
- We mapped alpha taxonomy rates using the date of species descriptions across space, and beta taxonomy rates using the number of synonyms associated with each species.
- We combined alpha and beta taxonomy rates to understand the global taxonomic status of Orthoptera (Arthropoda, Insecta), the sixth most species-rich insect order, which includes grasshoppers, crickets, katydids and relatives.
- In tropical regions, orthopteran taxonomy is recent and has few revisions, while in temperate regions it is older and more revised.

Keywords: alpha taxonomy, beta taxonomy, biodiversity knowledge shortfalls, Orthoptera, macroecology, taxonomic uncertainty.

Introduction

Taxonomy is a theoretical and practical science that deals with biodiversity classification (Mayr 1969). It develops hypotheses about species delimitation and classification within the taxonomic hierarchy. Being hypotheses, numerous described and named species become obsolete as taxonomy progresses, as they are either duplicate names for the same biological entity or refer to two or more distinct entities (Alroy 2002). The development of taxonomy is also heterogeneous across space (Baselga et al. 2010; Stropp et al. 2022). This is due to the uneven geographic distribution of both biodiversity and taxonomists throughout history (Baselga et al. 2010; Rodrigues et al. 2010), which results in varying levels of species detectability in nature, heterogeneous rates of species discovery in the field, diverse taxonomic criteria used in the description of new species, and taxonomic uncertainties associated with those already known species (Freeman and Pennell 2021; Stropp et al. 2022; Diniz-Filho et al. 2023; Frateles et al. 2024). These aspects of taxonomic knowledge have generated biases and gaps in different geographic regions, with greater knowledge and accuracy of species in certain locations, and profound implications for studies in taxonomy, ecology, evolution, medicine, and biodiversity conservation.

The discovery and naming of new species represents the first stage of taxonomy, known as alpha taxonomy (see Mayr 1969). The large-scale spatial distribution of alpha taxonomy can be inferred by combining the years of species name descriptions with the locality where the type material was collected and is deposited. Although most species inventories still have incomplete spatial coverage, the date of species naming across space is a little-explored indicator that can help guide areas in need of new (or renewed) taxonomic efforts and inventories,

as well as provide a historical record of descriptions over time. In this context, regions hosting many recent names indicate ongoing taxonomic efforts, while areas with older names may reflect past efforts and/or a lack of new ones (see Baselga et al. 2010).

The second stage of taxonomy, known as beta taxonomy, involves understanding of relationships among already described species and higher taxa through their systematic revision (see Mayr 1969). This stage can result in: (i) taxonomic lumping – identifying species that have been redundantly described several times under different names; and (ii) taxonomic splitting –uncovering ‘new species’ contained within the populations of previously described ones (Baselga et al. 2010; Stropp et al. 2022; Lessa et al. 2024). These processes are inherent to taxonomic advance, and their distribution in space can be inferred by the number of synonyms detected for the species occurring in a given region (Baselga et al. 2010). All other things being equal, it can be hypothesized that sites with a higher number of species with synonyms will have been more studied, with comparably more effort devoted to organize, correct, and classify species through taxonomic revisions. This implies that the number of synonyms across space provides an indirect indicator of taxonomic revisions (i.e., of beta taxonomy).

Mapping alpha and beta taxonomy across space can help unveil the mechanisms that have driven or limited taxonomic development in different regions, as well as highlight areas with varying levels of taxonomic stability. This, in turn, can enable the prioritization of locations and territories that need more taxonomic attention in the form of new species descriptions, taxonomic revisions, or both. However, to date very few broad-scale investigations into these issues have been conducted, possibly due to a lack of a strong connection between taxonomic

practice and macroecological research (see Stropp et al. 2022). To create this connection, it is necessary to assess taxonomic uncertainty from the knowledge contained in catalogs, atlases, and checklists (Ronquillo et al. 2023; Lessa et al. 2024), and integrate it with the biogeographic knowledge of digitally-accessible massive biodiversity data.

Here we combine the date of description of both valid species and synonyms, and the number of synonyms associated with each valid species name, as proxies for variations in alpha and beta taxonomy across space, in order to map the global trends in the status of taxonomic knowledge. We hypothesize that beta taxonomy is predominant in Europe and temperate regions of the Northern Hemisphere, as the history of taxonomy began with European naturalists in the 18th century. The early description of a large portion of species from these regions should have allowed a subsequent focus on taxonomic revisions. Moreover, the higher technical capacity of many countries in the north temperate areas is predicted to be associated with a wider range of contemporary approaches to taxonomy, facilitating higher rates of taxonomic change. In contrast, we expect alpha taxonomy to be predominant in tropical regions, reflecting both the megadiversity yet to be documented in these regions, as well as a scarcity of taxonomists and technical capacity. These same factors should also lead to fewer taxonomic revisions.

We test the above hypotheses for the global diversity of Orthoptera (Arthropoda, Insecta), the sixth richest order of insects today, including crickets, grasshoppers, katydids, and relatives. Since the publication of “*Systema Naturae*” (Linnaeus 1758), where only 49 orthopteran species were described for science (Green 1998; Bánki et al. 2024), the taxonomy of this group has captivated a large community of orthopterists worldwide, resulting in over 29,000 valid species

currently described. In addition to valid species names for all orthopteran species, big data repositories hold description dates and synonymic lists for most of these species, enabling the large-scale mapping of the status of taxonomic knowledge.

Methods

Building a checklist of Orthoptera names

We built a checklist of valid species names (SN) and associated synonyms (SYN) using the data from the Catalogue of Life (CoL). To this end, we considered all taxonomic components of scientific name, including the binomial names of the species (genus + specific epithet), combined with authorship (author(s) + year), and applied the following filters: (i) Taxonomic status – we disregarded fossil species previously filtered in CoL, subspecies, and subspecies synonyms, as well as names lacking the year of description and/or the author, and those with potentially dubious taxonomic status (i.e., provisionally accepted and species aggregate); (ii) Taxonomic timeframe – we filtered only names proposed until the year 2023; and (iii) Taxonomic ambiguities – we removed duplicate combinations of genus + epithet from the obtained set of names. This includes disregarding synonymic names resulting from the presence of subgenus syntax (e.g., *Gryllus campestris* = *Gryllus* (*Gryllus*) *campestris*) from the synonym(s) associated with each species name (i.e. $SYN \subseteq SN$) (Fig. 1).

Based on the filtered data, we built temporal species accumulation curves, which describe the accumulation of species names over time and the quantity of new names proposed per year (see Lu and He 2017). We built these curves for valid species, currently recognized synonyms, and taxonomists' names. For synonyms, we considered both homotypic and heterotypic synonyms (i.e.

synonymic names based on, respectively, either the same type specimen or a different one). For taxonomists' names, we considered each different name associated with the species name.

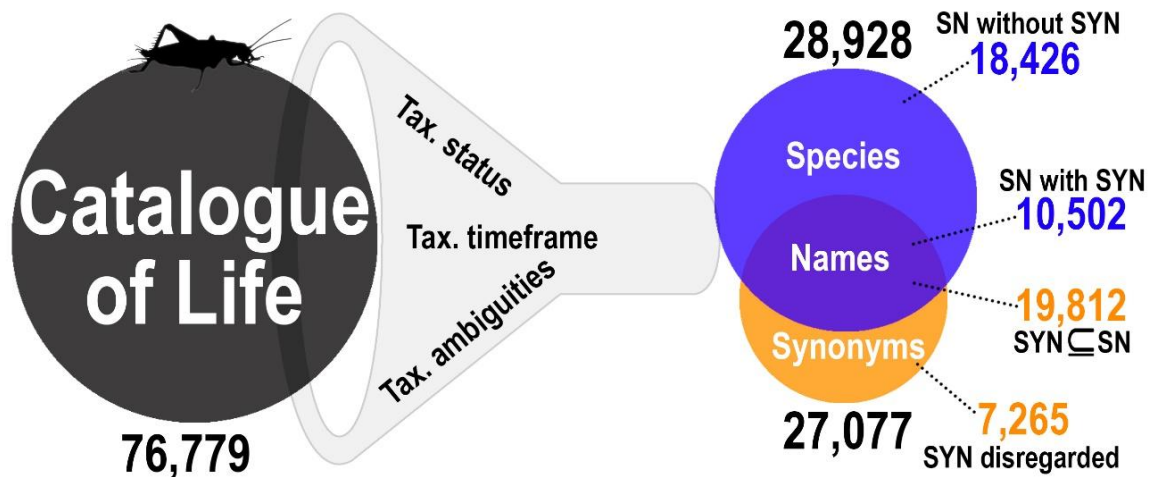


Fig 1. Filtering process and identification of valid species names and synonyms in the Catalogue of Life (CoL) considering all taxonomic components (genus + specific epithet + author(s) + year). The first step consisted in removing fossils, subspecies, and synonyms of subspecies (Tax. status). In the second filter, we counted proposed names until 2023 (Tax. timeframe). Finally, the third filter consisted of removing duplicate records in valid species names and synonyms, disregarding synonyms resulting from subgenus syntax (Tax. ambiguities).

Spatial distribution of names: valid species and total names

Using the valid species names filtered in CoL (Fig. 2A), we searched for primary records in the Orthoptera Species File (OSF; Cigliano et al. 2024) and the Global Biodiversity Information Facility (GBIF; GBIF 2024) (Fig. 2B). We filtered these databases to retain only records identified at species level, with complete information on species authority, and those with geographic coordinates with at least one decimal digit of precision. For the data sourced from GBIF, we removed records where the longitude coordinates were equal to the latitude coordinates, as well as centroid coordinates of country capitals and provinces, using the CoordinateCleaner package (Zizka et al. 2019). Subsequently, we integrated the

information of OSF and GBIF into a single dataset and removed duplicate records, considering the information of the name (binomial name + authorship) and coordinates. We included only the records that were registered in land surface, including islands. For this, we used the Natural Earth platform at a 1:10 m resolution (<https://www.naturalearthdata.com/>).

We calculated the number of valid species names within grid cells (pi) of 2° resolution (using Projected Coordinate System EPSG-4326). We also calculated the sum of synonyms associated with all valid species names in each cell, obtaining the total number of names (tn) (Fig. 2B). Note that with the term synonyms here we refer to what can be defined as “nomenclatural lumping”, which besides true synonyms may also encompass spelling errors to synonyms, ambiguous synonyms, and erroneous combinations attributed to the currently valid species name throughout taxonomic knowledge. However, it disregards synonyms created by the presence of subgeneric syntax (see above). Thus, what we here consider as lumping does not distinguish between species that have been relegated to synonyms after a taxonomic re-evaluation and nomenclatural changes, as discussed by Lessa et al. (2024) (see Supplementary Appendix 1).

Indices of alpha and beta taxonomy

We calculated two different indices as proxies for alpha and beta taxonomy: Alpha Temporal (AT) and Beta Proportion (BP) for all cells of our 2° global grid. Both indices were based on the spatial distribution of currently valid species names obtained from their records, rasterized and mapped at the same 2° grid scale (Figs 2C, D). Additionally, we also performed a sensitivity analysis of each index (see Supplementary Appendix 2 and Supplementary Fig. 1).

Alpha Temporal index (AT)

The alpha temporal index measures the temporal distribution of species naming efforts at different cells. AT assumes that cells with recent species names indicate ongoing taxonomic efforts, while cells with older species names do reflect efforts made in the past or the absence of new efforts – in the case that no recent names are present in the cell. To calculate this index, we fitted the year of description of each valid species name to an exponential decay curve, derived from a polynomial function of degree 4, considering the time span of all names. Specifically, we assigned an alpha temporal value for each valid species name AT_{sn} (Eq. 1), based on the time elapsed from the most recent authorship year to the oldest (2023 to 1758 in our data) (Fig. 2C, first graph).

$$Eq. 1 \ AT_{sn} = (t_i - t_{max}/h)^4$$

where t_i is the year of authorship of the investigated valid species name, t_{max} is the year of the most recent authorship in the checklist of the valid species names, and h is the maximum value of the difference between the years of the names. Subsequently, taking the values previously assigned to each valid species name (AT_{sn}), we calculate the average value of alpha temporal for each location (pi), based on the geographical occurrence of each species (Eq. 2).

$$Eq. 2 \ AT_{pi} = \mu AT_{sn}(pi)$$

AT values range from ~0 to 1; when AT ~0, it indicates that the majority of valid species names in a cell are recent, meaning alpha taxonomy is predominant. Conversely, values close to 1 indicate that the majority of valid species names in a cell are old, meaning alpha taxonomy is not predominant currently (Fig. 2D).

Beta Proportion index (BP)

The Beta Proportion index measures the portion of synonymized species names in a given location. We assume that the most studied sites would have more synonyms per species because of the greater effort to through taxonomic revisions, although this is not a general rule in taxonomy because of the priority rule in the Zoological Code of Nomenclature (see Braby et al. 2024; <https://www.iczn.org/>). This index can be mapped based on the known distribution of currently valid species, assuming that their associated synonyms share the same distribution ranges. Although this assumption may not hold valid in many cases, it does indicate the distribution of species that were recently revised, thereby making it appropriate as a proxy for beta taxonomic work. For each cell (π_i), we calculated Beta Proportion using Eq. 3:

$$\text{Eq. 3 BP} = 1 - [\text{sn}(\pi_i)] / [\text{tn}(\pi_i)]$$

where BP_{π_i} is the result of 1 minus the ratio of the number of valid species name occurrences (sn) to the total number of name occurrences (tn) for each location (π_i), and tn is the result of sn plus all associated synonyms for a given location (π_i) (J.Stropp, A.S.S.Pereira, T.Emilio, L.Meyer, F.Alves-Martins, R.J.Ladle & J.Hortal unpublished) (Fig. 2C).

BP values range from ~0 to 1; BP ~0 indicates a low proportion of synonyms relative to the existing species at the location, suggesting that beta taxonomy is not predominant. In contrast, values close to 1 reflect a high proportion of synonyms, indicating that beta taxonomy is predominant. It is important to note that it is not possible to determine whether beta taxonomy is recent or older (Fig. 2D).

Combining alpha and beta taxonomy to map the status of taxonomic

knowledge

To identify areas with different levels of alpha and beta taxonomy, we combined the two alpha temporal and beta proportion indices into a bivariate map. For this, we divided the AT and BP values (X and Y axes, respectively) into octiles (eight quantiles) and assigned different colors to different combinations (see Biesecker et al. 2020). Each color represents a specific combination of octile shifts (Fig. 2E). All analyses were performed in R (R Core Team 2024).

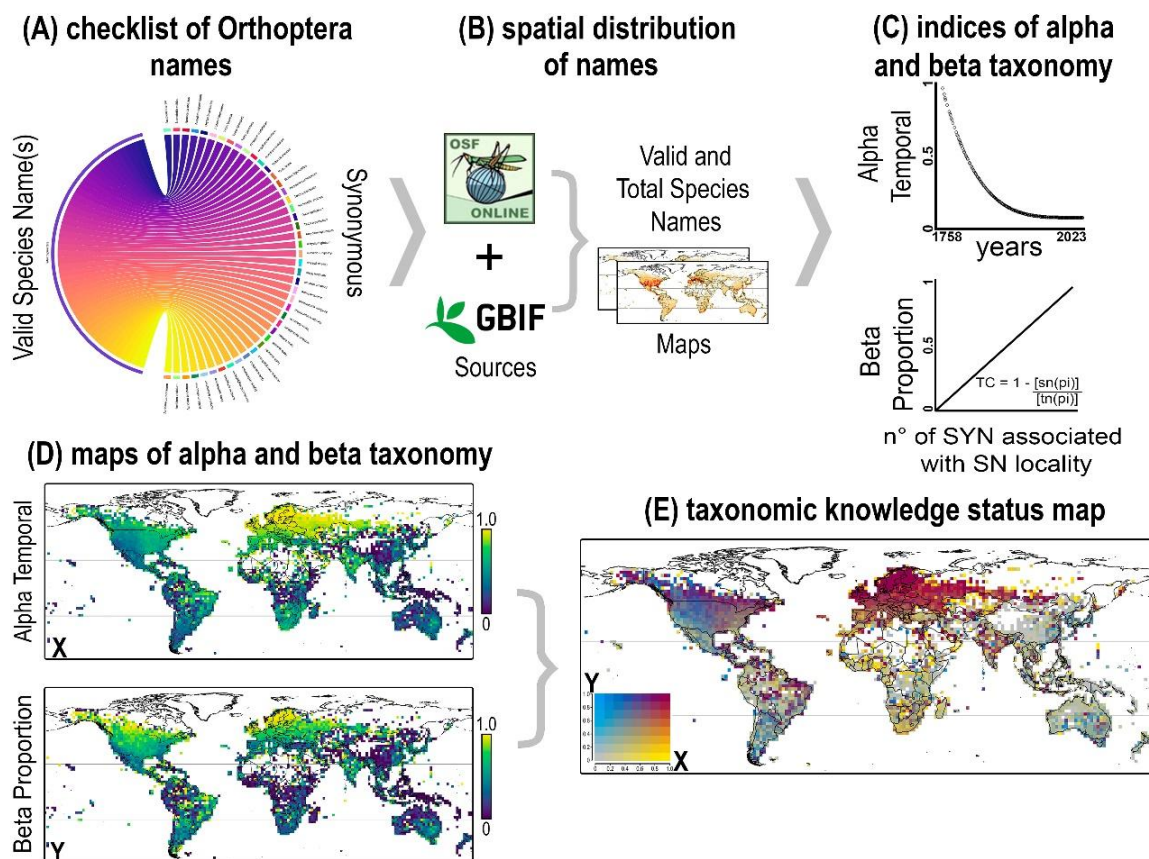


Fig 2. Workflow for building a taxonomic knowledge status map: (A) build a checklist of Orthoptera names (i.e., valid species names and associated synonyms) using data from the Catalogue of Life (CoL); (B) access the geographic distribution of the number of valid and total species names integrating data from the Orthoptera Species File Online (OSF) and the Global Biodiversity Information Facility (GBIF); (C) calculate two indices representing alpha and beta taxonomy: alpha temporal (AT) and beta proportion (BP); (D) build maps of alpha and beta taxonomy, rasterizing the indices of AT and BP in each 2° grid cell with data on orthoptera species; and (E) combine the AT (y axis) and BP (x axis) indices into a bivariate

taxonomic knowledge status map.

Results

Spatial distribution of orthopteran species names

Our dataset included 75,88% of orthopteran currently valid species names. Considering only valid species names with synonyms (i.e., $\text{SYN} \subseteq \text{SN}$), we obtained information for 8,080 out of the 10,502 valid species names in the Catalogue of Life, representing 76,96% of all currently recognized species. We obtained spatial information for 21,945 of the 28,928 orthopteran valid species names recorded in CoL. The records for valid species names of Orthoptera cover all terrestrial regions of the world, except the desert areas of northern Africa, central Asia, and the polar regions (Fig. 3A). The highest peaks in the number of valid species recorded are concentrated in tropical latitudes, notably in specific regions of Central America, western South America, Africa, and Indo-Malaysia. However, the temperate latitudes of the Northern Hemisphere also show a high density of valid species, mainly in the United States of America and Europe. Whereas in the Southern Hemisphere the places with high numbers of records for valid species are sparsely distributed along the Southern Hemisphere, with notable concentrations in Paraguay, southeastern Brazil, South Africa, and southeastern Australia (Fig. 3A).

The spatial distribution of the total number of names is quite similar to that of species names, though the total number of names is in some cases up to five times larger than the number of valid names (see the correlation in Supplementary Fig. 3). The highest peak of total names occurs in tropical latitudes, while the regions with the highest density of names are in the temperate latitudes of the Northern Hemisphere, covering the United States of America and Europe. In the Southern Hemisphere, places with high numbers of total names are also scattered (Fig. 3B).

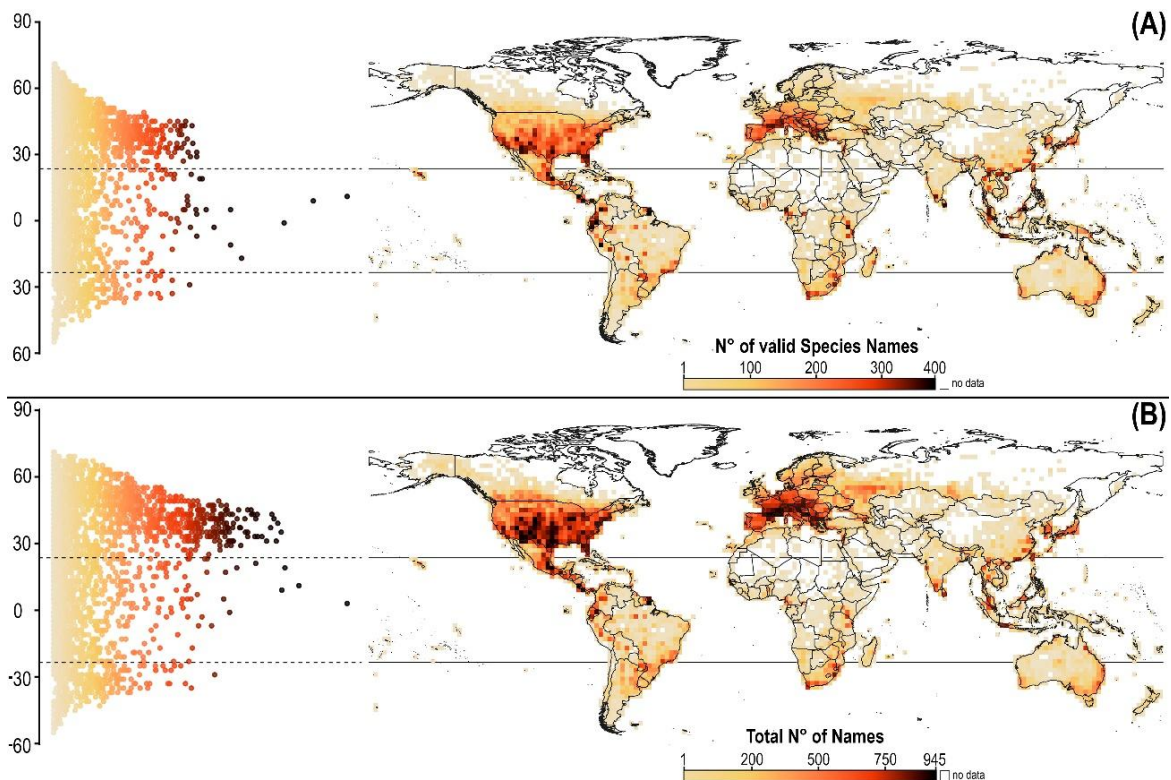


Fig 3. Geographic distribution of the number of valid species names (A) and the total number of Orthoptera names (B). In (A) dark red colors indicate higher richness of valid species names, while light beige regions indicate lower richness. In (B) dark regions present higher total numbers of names (i.e., number of valid species names + number of associated synonymous names), while light colors indicate a lower total number of names.

Historical rates of alpha and beta taxonomy

The accumulation of valid species names of Orthoptera has increased exponentially since the mid-19th century, with a brief decline just after World War II (Fig. 4A), although the number of valid species described each year shows large fluctuations over time (Fig. 4B). Most recent valid species names (i.e., $AT \leq 0.5$) are found in tropical latitudes, whereas most valid species names are older (i.e., $AT \geq 0.75$) in the temperate latitudes of the Northern Hemisphere, particularly in Europe and southwestern Asia (Fig. 5A). In the Southern Hemisphere, moderately older valid species names (i.e., $0.5 \geq AT \geq 1.0$) are found in South Africa and southeastern Australasia.

The temporal accumulation of synonyms exceeded the growth of valid species names until shortly after World War II, when it began to decline and become lower (Fig. 4A). The number of synonyms per year also shows large fluctuations, with the highest peaks occurring in the 19th century (Fig. 4B). Most regions hold low proportions of synonyms (i.e., $BP \leq 0.5$), with some exceptions in northern South America and southern India (Fig. 5B). In contrast, most regions in the temperate latitudes of the Northern Hemisphere show high proportions of synonyms (i.e., $BP \geq 0.75$), including northern Europe, southern Alaska, Canada, and Russia. In the Southern Hemisphere, the pattern ranges from moderate to strong (i.e., $0.5 \leq BP \leq 1.0$), with more pronounced levels of synonyms mainly in southern Argentina and southeastern Australia (Fig. 5B).

The accumulation of taxonomists' names increases over time, although to a much lesser extent compared to species and associated synonyms (Fig. 4A). The number of active taxonomists per year shows significant reductions during the World Wars I and II, but a strong growth afterwards, a trend that continues into the 21st century (Fig. 4B).

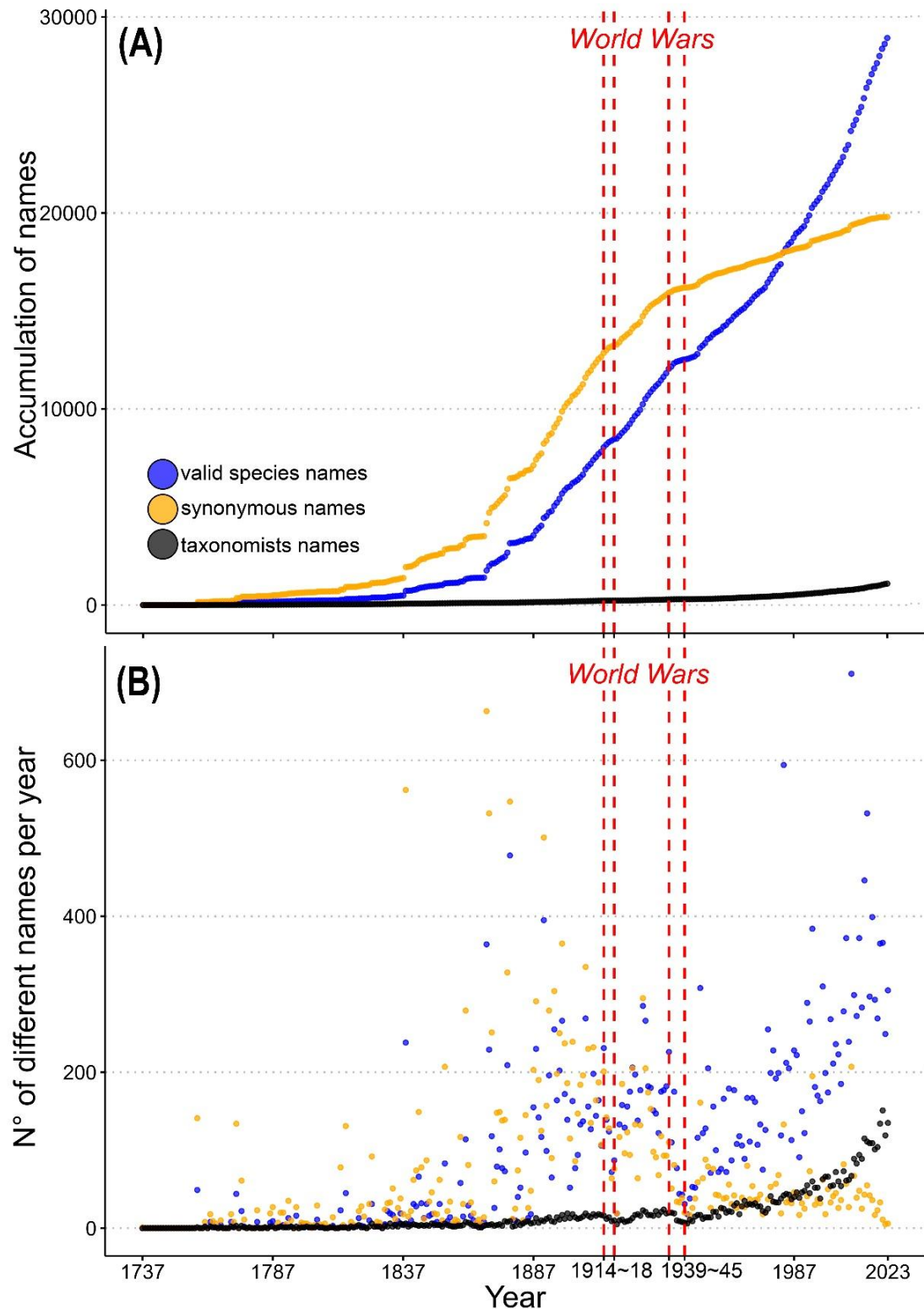


Fig 4. Rates of taxonomic work on Orthoptera names through time. (A) accumulation of valid species names (blue), synonyms (orange), and active taxonomists (black) over time. (B) number of currently valid species (blue), synonyms (orange), and taxonomists' (black) names per year.

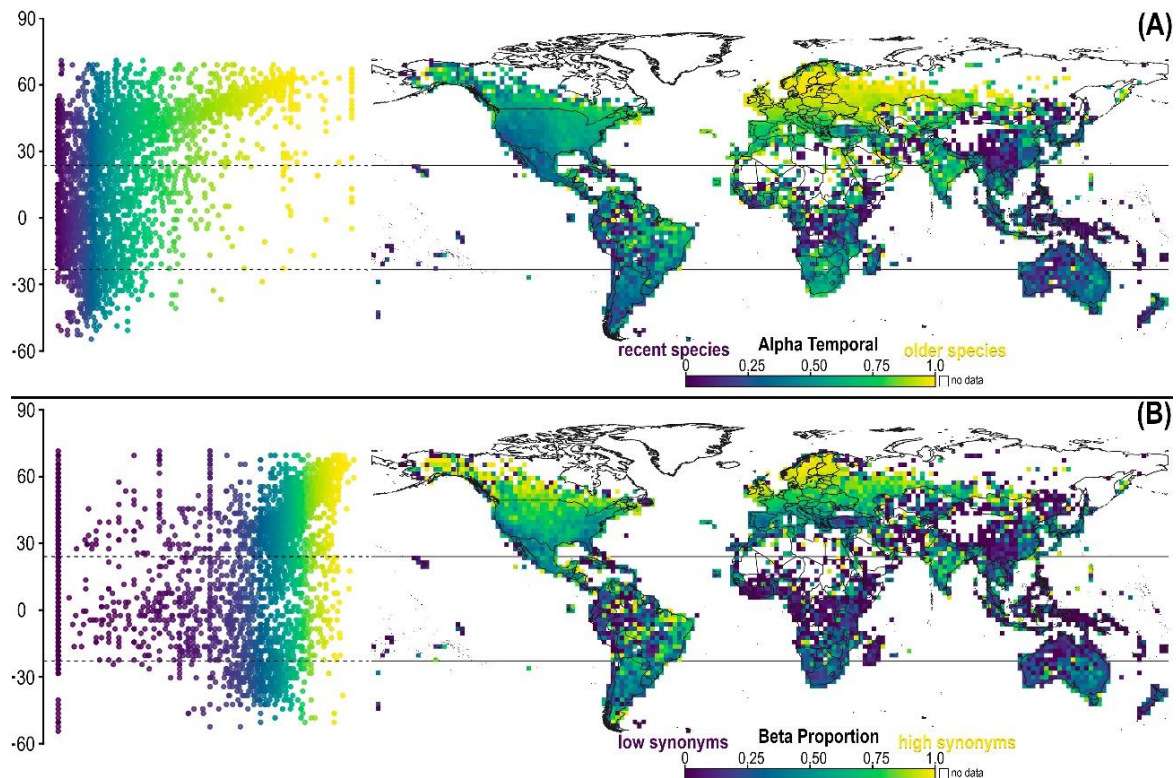


Fig 5. Spatial distribution of Alpha Temporal (A) and Beta Proportion (B) indices for Orthoptera names. In (A) yellow and green indicate older species names, blue represents more recent species names. In (B) yellow and green indicate a higher number of synonyms associated with valid species names, while blue corresponds to regions with lower proportions of synonyms.

Mapping the status of taxonomic knowledge

The bivariate map of alpha and beta taxonomy reveals at least four main types of taxonomic trends (Fig. 6). Regions in process of inventory, characterized by recent species and few synonyms (i.e., low AT and BP; gray in Fig. 6), are located mainly in the tropics latitudes, but extend also to some temperate regions of South America and Australasia, and many Asian countries, such as China, Mongolia, Afghanistan, and Iran. In contrast, the regions with well-established taxonomies based on older valid species and many synonyms (i.e., high AT and BP, purple in Fig. 6), are nearly absent in tropical latitudes (except for Brazil and India) and are more prevalent in temperate latitudes of the Northern Hemisphere, mainly in Europe and southwestern Asia.

Interestingly, many regions have received intense taxonomic work in the last decades, coupling species descriptions with taxonomic reviews, characterized by recent species and many synonyms (i.e., low AT and high BP, blue in Fig. 6). This is the case of central and western North America and, to a lesser extent, southern South America and South and East Australia and New Zealand. Besides these regions, only a few areas scattered along East Asia and Africa show a similar pattern of coupled description and revisionary work in recent times. Finally, places with old species inventories that have not been subject to revision, which have older species and few synonyms (i.e., high AT and low BP), are relatively rare except in Africa and Central Asia, plus some cells scattered worldwide (Fig. 6, yellow).

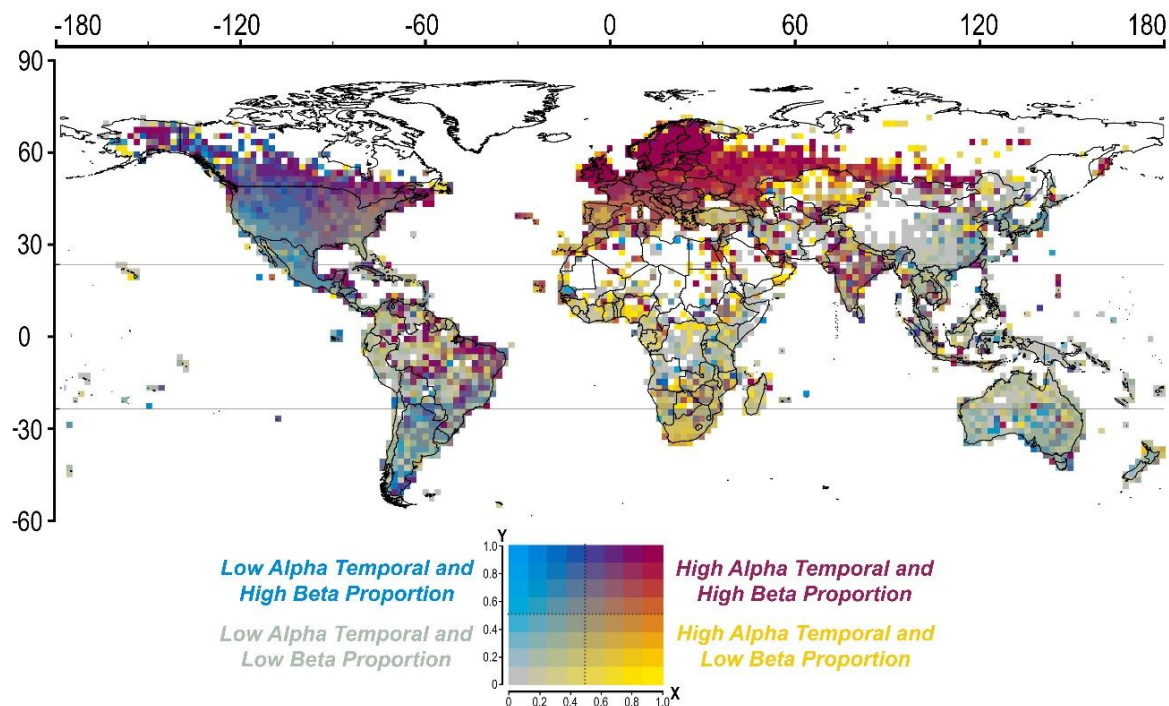


Fig 6. Taxonomic knowledge status map resulting from the combination of alpha temporal (AT, X axis) and beta proportion (BP, Y axis). Regions in blue indicate high numbers of recently described species and synonyms; regions in purple indicate older species names together with many synonyms; regions in gray indicate recent species and few synonyms; and regions in yellow indicate older species and few synonyms.

Discussion

We show that tropical regions and the Southern Hemisphere contain the majority of recently described valid species names for Orthoptera, with fewer associated synonyms. This supports our hypothesis that alpha taxonomy currently predominates in tropical regions, as well as in other areas of the Southern Hemisphere beyond the tropics. It also highlights the significant need for extensive taxonomic revisionary efforts in these areas. In contrast, temperate regions of the Northern Hemisphere, particularly Europe and Southwest Asia, have a high concentration of older valid species names and synonyms, corroborating our hypothesis that beta taxonomy is predominant in Europe and other regions of the Northern Hemisphere. However, our results also indicate a strong longitudinal gradient in the Western Hemisphere, in which east coasts are taxonomically better known, and hold more taxonomic revisions, in both North and South America, reflecting the history of colonization of these regions.

The widespread distribution of areas with recent names and few synonyms reflects an emerging orthopteran alpha taxonomy but little beta taxonomy, characterized by species discoveries which are not yet accompanied by comparing taxa across regions and refining their classifications through taxonomic revisions. This can be attributed to a historical shortage of taxonomists in the megadiverse regions of the Global South – the taxonomic impediment *sensu* Engel et al. (2021), where the rich diversity of species has so far outpaced the available taxonomic expertise. Due to this, many sites in South America, Southeast Asia, and Australasia are currently undergoing extensive activity in alpha taxonomy. In these areas taxonomic revisions are now crucial to produce more robust taxonomies. However, besides the lack of taxonomic capacity, this task is also hampered by the

location of most of the reference natural history collections in countries from the Global North, which still store much of the material collected in tropical regions by early naturalists (Hawksworth 1995; Bakker et al. 2020; Raja et al. 2021; Weber 2021) (see Supplementary Appendix 3, Supplementary Fig. 4A). This unfortunate by-product of colonialism adds to the taxonomic impediment, creating a logistic obstacle for researchers from tropical regions, who often need access to type material deposited in Old World collections to advance their taxonomic revisions (e.g., Raja et al. 2021; Nakamura et al. 2024; Supplementary Fig. 4B).

This contrasts with the situation in the historical centers of taxonomic research in Europe and Southwest Asia, which are currently more active on beta taxonomy. These areas already hold a long history of taxonomic efforts, with a large number of taxonomic changes that result in a combination of older valid names and many synonyms. It is most likely that these regions are taxonomically saturated to some extent, with a significant proportion of their actual species richness already known, described, and included in taxonomic reviews or catalogs (see Green 1998; Heller et al. 1998; <https://www.grasshoppersofeurope.com/>). Indeed, most of the Northern Hemisphere is taxonomically advanced and well-acquainted for the majority of the Orthoptera fauna (see Waloff and Popov 1990; Song 2010). Thus, there the current focus is on describing species from remote areas and inaccessible habitats, such as caves and mesovoid soil substrate, refining existing knowledge, and conducting broader analyses on systematics and evolutionary relationships, using state-of-the-art techniques such as “omics” (e.g. phylogenomics, museomics or transcriptomics; Song et al. 2020; Kociński et al. 2021; Kim et al. 2024).

Strikingly, a large number of areas hold, on average, recent names and numerous synonyms. In these regions the active research in both alpha and beta

taxonomy couples the discovery of new species with taxonomic reviews and the publication of synonym lists. In some regions, such as in West North America, these areas are adjacent to territories with older names and a large proportions of synonyms. This may indicate that, although regional checklists were previously thought to be complete (e.g. Song 2010), the fine-scale exploration of these territories may still uncover previously unknown taxonomic gaps that end up in the discovery of new species (see Castro-Souza et al. 2024). Such discoveries have likely been facilitated in the USA by the high presence of reference biological collections and the quantity of Orthoptera types deposited in their natural history collections (see Supplementary Appendix 3, Supplementary Fig. 4A, B). For certain countries, these taxonomic contributions are due to the significant legacy of local taxonomists such as María Marta Cigliano (Cigliano et al. 2000; Cigliano et al. 2024) in Argentina, David Rentz (Rentz and Su 2019) in Australia, and Carlos Salvador Carbonell (Bentos-Pereira and Lorier 2019) in Uruguay and Zhe-Min Zheng (Huang et al. 2022) in China. Their enormous labor continues to influence new Orthoptera taxonomists. In other parts of the globe, these regions possibly reflect legacies of European colonialism in megadiverse locations (Weber 2021), or higher levels of international scientific collaboration. In any case, this level of taxonomic development is less concerning than that of areas where only alpha taxonomy has been conducted. Behind the large number of synonyms and recent valid names in these areas may lie studies utilizing an integrative and more modern taxonomic approach, such as bioacoustics and/or molecular techniques (e.g., Nattier et al. 2011; Husemann et al. 2013; Braswell et al. 2016; Heller et al. 2017; Riede 2018; Uluar et al. 2021; Kim et al. 2024), which are based on novel criteria and species concepts that end up generating numerous taxonomic adjustments even in well-

known species.

Very few areas hold a combination of older names and few synonyms, which indicates only past inventory effort but little or no revisionary work. This may be due to two distinct, yet non-complementary, phenomena. In the best-case scenario, in these areas current taxonomic knowledge comes from studies that produced classifications well-structured in terms of nomenclature and taxonomic relationships in the past, which have persisted until the present day, even with evidence coming from new taxonomic approaches. Alternatively, the current regional checklists in these regions come from taxonomic work conducted in the past that has not been revisited or revised later on. Importantly, a significant part of these areas is distributed in countries where socio-political conflicts have occurred in the past or are occurring nowadays, such as Russia, Kazakhstan, Afghanistan, and various countries in Africa (Institute for Economics and Peace 2023). This not only hinders inventorying their biodiversity (Hanson et al. 2009), but also impacts taxonomy due to lack of trained taxonomists, low and discontinued investment, and the discouragement of research and collaboration from international taxonomists due to high personal security risks and logistical issues (see Hilario-Husain et al. 2024). This type of taxonomic status is the most critical, compared to the other types identified here, and the areas holding it require strong attention from the taxonomic community. In the areas lacking capacity or investment, these efforts could be facilitated by creating partnerships with local scholars and institutions (see Hanson et al. 2009), even in the absence of regional Orthoptera taxonomists. Other areas holding such a status in conflict-free countries, such as Canada, Australia, New Zealand, and some small parts of Europe (Institute for Economics and Peace 2023) correspond to spatial gaps within well-known regions, and could be easily revisited

by regional taxonomists.

It is worth noting that the global distribution of taxonomic knowledge depicted by our work is not only the result of taxonomic work, but also of sampling effort. Regions with frequent field surveys are likely to have more species recorded into their checklists, leading to more opportunities for taxonomic effort, such as the identification of new species or the revision of already described ones. Conversely, areas with historically low levels of sampling effort can appear underrepresented in terms of the occurrence of newer species simply because fewer surveys have been conducted there, not because they have fewer species. The regions with continuous taxonomic effort often have a history of consistent sampling and research investment, leading to more refined taxonomic studies (Wege et al. 2015; Wilkinson et al. 2021). Thus, while we recognize the key role of sampling effort in determining the accumulation of both valid species names and synonyms, we believe that the observed patterns provide a meaningful indication of the history and intensity of the taxonomic research across the world.

Interestingly, our global map of taxonomic status identifies some geographic patterns that can be attributed to the history of biodiversity explorations. The longitudinal gradient in well-known sites holding revisions (old AT and high BP) between the Americas and Europe indicates the influence of European colonialism and naturalist expeditions over time. These expeditions typically landed on the eastern coast of the Americas and moved inland, documenting the biodiversity of the New World. This is somehow mimicked by the longitudinal gradient shown by poorly developed old taxonomies (old AT and low BP) reflecting the efforts of naturalists from that era to document the biodiversity of the Old World, although these efforts have not continued into the present day. In contrast, the areas with

recent AT and high BP evidence the existence of a new wave of biodiversity researchers that revise the classification at the same time as they describe species, many of these researchers are academic descendants of local taxonomists. Nonetheless, although we only analyzed Orthoptera taxonomy, the situation is likely to be quite similar for other arthropod groups with historical taxonomic studies in tropical regions primarily conducted by European and North American researchers, such as spiders, mites, millipedes, centipedes, or many others (Jacinavicius et al. 2018; Iniesta et al. 2023a, b; World Spider Catalog 2024).

Conclusion

Even though taxonomy has played a pivotal role in biological science since the 18th century proposal of Linnaeus' *Systema Naturae*, assessing the current status of taxonomic knowledge for any taxon remains crucial. It helps identifying understudied and undersampled areas requiring further attention and anticipates future trends in systematics, especially in relation to related fields such as molecular phylogenetics, biogeography, and ecology. In general, spatial biases in both alpha and beta taxonomy can be influenced by various factors, including a lack of funding for monitoring and biodiversity studies, and the shortage of trained taxonomists. Importantly, taxonomic advancements have largely been driven by increased knowledge from regions with consolidated and emerging research groups, for Orthoptera and other arthropod taxa. This is evident in regions with high alpha temporal and beta proportion, which are known for having historical zoological collections, curatorial assistance, and scientific funding. In this sense, species richness revealed through taxonomic studies is not only linked to biodiversity patterns but also driven by social, cultural, and political factors, such as

governmental funding, logistical challenges, the exodus of type material from its original locality, and the shifting priorities of different historical periods.

We show that mapping the status of taxonomic knowledge allows the identification of regional trend in species description and revision. For the order Orthoptera, our model highlights that in South America, Southeast Asia, and Australasia, taxonomic efforts are needed not only for species description but also for reviewing and integrating phylogenetic and biogeographic studies. With this work, we aim to provide a foundation for exploring future challenges in Orthopterology, for both established taxonomists and new trainees, as well as for ecologists and biogeographers. But more importantly, the framework presented here brings together taxonomy and macroecology in an accessible way for biodiversity studies in the era of Big Data.

Acknowledgements

We are grateful to the team at the Laboratory of Macroecology and Biodiversity Conservation (MacrEco) of the Federal University of Mato Grosso (UFMT) and the National Institute of Science & Technology (INCT) in Ecology, Evolution, and Biodiversity Conservation (EECBio), based at the Federal University of Goiás (UFG), for the valuable discussions that contributed to the development of this study. RACS is grateful to the Coordination for the Improvement of Higher Education Personnel (CAPES), Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio) and National Center for Cave Research and Conservation (CECAV), for the scholarships. NS was supported by the Araucária Foundation (nº 09/2021). JH was supported by project NICED, grant PID2022-140985NB-C21 funded by MCIN/AEI/ 10.13039/501100011033 / FEDER, EU. TSS expresses

thanks to FAPEMAT (project FAPEMAT-PRO.000274/2023) and CAPES (grant 88887.964855/2024-00). GT was supported by the National Council for Scientific and Technological Development (MCTI/CNPQ/CAPES/FAPS Nº 16/2014 - PROGRAMA INCT). We thank the developers of Catalogue of Life (CoL), Global Biodiversity Information (GBIF), and Orthoptera Species File (OSF) that provided digitally accessible data.

Author Contributions

R.A.C.S., J.S., T.S.S. and J.H. conceived the study with input from all authors; investigation, all authors; data curation, R.A.C.S., T.S.S. and J.H.; formal analysis, R.A.C.S., T.S.S. and J.H.; visualization, R.A.C.S. and J.H.; funding acquisition, J.A.D.F., T.S.S. and J.H.; writing - original draft, R.A.C.S., T.S.S., L.I. and J.H.; writing - review & editing, J.S., L.I., R.J.L., G.T., N.S. and J.A.D.F.; All authors have read, discussed the results and approved the final manuscript.

Data Availability

The datasets used in this study are available in Catalogue of Life (CoL), Orthoptera Species File (OSF) and Global Biodiversity and Information (GBIF). The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Supplementary Material

The following materials are available online:

Figure S1. Sensitivity analysis between synonyms and valid species names.

Figure S2. Density curves of alpha temporal and beta proportion indices.

Figure S3. Correlation between the number of valid species names and the total number of names.

Figure S4. Distribution of global scientific collections and types of Orthoptera.

Appendix S1. The 'nomenclatural' lumping.

Appendix S2. Assessing the sensitivity of alpha temporal and beta proportion indices.

Appendix S3. Distribution of global scientific collections and types of orthoptera.

References

- Alroy, J. (2002) How many named species are valid?. *Proceedings of the National Academy of Sciences* 99: 3706–3711. <https://doi.org/10.1073/pnas.062691099>
- Bánki, O., Roskov, Y., Döring, M., et al. (2024) Catalogue of Life (Version 2024-07-18). Catalogue of Life, Amsterdam, Netherlands. <https://doi.org/10.48580/dgjy9>
- Bakker, F.T., Antonelli, A., Clarke, J.A., et al. (2020) The Global Museum: natural history collections and the future of evolutionary science and public education. *PeerJ*, 8, e8225. <https://doi:10.7717/peerj.8225>
- Baselga, A., Lobo, J.M., Hortal, J., Jiménez-Valverde, A. & Gómez, J.F. (2010) Assessing alpha and beta taxonomy in eupelmid wasps: determinants of the probability of describing good species and synonyms. *Journal of Zoological Systematics and Evolutionary Research*, 48, 40–49. <https://doi.org/10.1111/j.1439-0469.2009.00523.x>
- Bentos-Pereira, A. & Lorier, E. (2019) Carlos Salvador Carbonell Mas. *Boletín de la Sociedad Zoológica del Uruguay*, 28, 104–105. https://journal.szu.org.uy/index.php/Bol_SZU/article/view/107
- Biesecker, C., Zahnd, W.E., Brandt, H.M., Adams, S.A. & Eberth, J.M. (2020) A bivariate mapping tutorial for cancer control resource allocation decisions and interventions. *Preventing Chronic Disease*, 17, E01. <https://10.5888/pcd17.190254>
- Braby, M.F., Hsu, Y.F. & Lamas, G. (2024) How to describe a new species in zoology and avoid mistakes. *Zoological Journal of the Linnean Society*, 2024, zlae043. <https://doi:10.1093/zoolinnean/zlae043>.
- Braswell, W.E., Birge, L.M. & Howard, D.J. (2006) *Allonemobius shalontaki*, a new cryptic species of ground cricket (Orthoptera: Gryllidae: Nemobiinae) from the southwestern United States. *Annals of the Entomological Society of America*, 99, 449–456. [https://doi:10.1603/0013-8746\(2006\)99\[449:asancs\]2.0.co;2](https://doi:10.1603/0013-8746(2006)99[449:asancs]2.0.co;2)
- Castro-Souza, R.A., Tessarolo, G., Stropp, J., Diniz-Filho, J.A., Ladle, R.J., Szinwelski, N., Hortal, J. & Sobral-Souza, T. (2024) Mapping ignorance to uncover shortfalls in the knowledge on global Orthoptera distribution. *npj Biodiversity*, 3, 22. <https://doi.org/10.1038/s44185-024-00059-1>
- Cigliano, M.M., De Wysiecki, M.L. & Lange, C.E. (2000) Grasshopper (Orthoptera: Acridoidea) species diversity in the Pampas, Argentina. *Diversity & Distributions*, 6, 81–91. <https://doi:10.1046/j.1472-4642.2000.00077.x>
- Cigliano, M.M., Braun, H., Eades, D.C. & Otte, D. (2024) Orthoptera Species File. Accessed on 2024-07-31. Available at <http://orthoptera.speciesfile.org/>
- Cracraft, J. (1983) Species Concepts and Speciation Analysis. In *Current Ornithology* (p. 159–187). Boston, MA: Springer US.

Diniz Filho, J.A.F., Jardim, L., Guedes, J.J., et al. (2023) Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Frontiers of Biogeography*, 15, e59566. <https://doi:10.21425/F5FBG59566>

Dubois, A. (2000) Synonymies and related lists in zoology: general proposals, with examples in herpetology. *Dumerilia*, 4, 33–98.

Engel, M.S., Ceríaco, L.M., Daniel, G.M., et al. (2021) The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. *Zoological Journal of the Linnean Society*, 193, 381–387. <https://doi:10.1093/zoolinnea/zlab072>

Frateles, L.E.F., Tavares, G.R.G., Nakamura, G., da Silva Jr, N.J., Terribile, L.C. & Diniz-Filho, J.A.F. (2024) The interaction between the Linnean and Darwinian shortfalls affects our understanding of the evolutionary dynamics driving diversity patterns of New World coralsnakes. *Journal of Biogeography*, 0, 1–13. <https://doi.org/10.1111/jbi.15014>

Freeman, B.G. & Pennell, M.W. (2021) The latitudinal taxonomy gradient. *Trends in Ecology & Evolution*, 36, 778–786. <https://doi:10.1016/j.tree.2021.05.003>

Gardner, A.L. & Hayssen, V. (2004) A guide to constructing and understanding synonymies for mammalian species. *Mammalian species*, 739, 1–17. <https://doi:10.1644/739>

GBIF.org (2024) GBIF Occurrence Download. Accessed on 2024-07-31. Available at <https://doi.org/10.15468/dl.75vt4p>.

Green, S.V. (1998) The taxonomic impediment in orthopteran research and conservation. *Journal of Insect Conservation*, 2, 151–159.

Hanson, T., Brooks, T.M., Da Fonseca, G.A.B., et al. (2009) Warfare in biodiversity hotspots. *Conservation Biology: The Journal of the Society for Conservation Biology*, 23, 578–587. <https://doi:10.1111/j.1523-1739.2009.01166.x>

Hawksworth, D. L. (1995) The resource base for biodiversity assessments. In *Global Biodiversity Assessment* (V.H. Heywood, ed.), pp. 545–605. Cambridge, UK: Cambridge University Press.

Heller, K.G., Korsunovskaya, O., Ragge, D.R., Vedenina, V., Willemse, F., Zhantiev, R.D., & Frantsevich, L. (1998) Check-list of European orthoptera. *Articulata*, 7, 1–61.

Heller, K.G., Ingrisch, S., Liu, C.X., Shi, F.M., Hemp, C., Warchałowska-Śliwa, E. & Rentz, D.C.F. (2017) Complex songs and cryptic ethospecies: the case of the *Ducetia japonica* group (Orthoptera: Tettigonioidae: Phaneropteridae: Phaneropterinae). *Zoological Journal of the Linnean Society*, 181, 1–22. <https://doi:10.1093/zoolinnea/zlw019>

Hilario-Husain, B.A., Guerrero, S.J., Tanalgo, K., et al. (2023) Caught on the crossfire: Biodiversity conservation paradox of sociopolitical conflict. *npj Biodiversity*, 3, 1–9. <https://doi:10.21203/rs.3.rs-3490940/v1>

Husemann, M., Llucià-Pomares, D. & Hochkirch, A. (2013) A review of the Iberian Sphingonotini with description of two novel species (Orthoptera: Acrididae: Oedipodinae): Review of Iberian Sphingonotini. *Zoological Journal of the Linnean Society*, 168, 29–60. <https://doi:10.1111/zoj.12023>

Huang, J.R., Tang, J.X. & He, Z.Q. (2022) Trend of Chinese Orthoptera taxonomy research based on holotype specimens. *Biodiversity Science*, 30, 21314. <https://doi:10.17520/biods.2021314>.

Iniesta, L.F.M., Bouzan, R.S. & Brescovit A.D. (2023a) A reassessment of the Neotropical genus *Pseudonannolene* Silvestri, 1895: cladistic analysis, biogeography, and taxonomic review (Spirostreptida: Pseudonannolenidae). *European Journal of Taxonomy*, 867, 1–312. <https://doi:10.5852/ejt.2023.867.2109>

Iniesta, L.F.M., Bouzan, R.S., Means, J.C., Ivanov, K. & Brescovit A.D. (2023b) Where are they from and where are they going? Detecting areas of endemism, distribution patterns and conservation status of the order Spirostreptida in Brazil (Diplopoda, Juliformia). *Biodiversity and Conservation*, 32, 1591–1615. <https://doi:10.1007/s10531-023-02566-2>

Institute for Economics & Peace (2024) Global Peace Index 2023: Measuring Peace in a Complex World, Sydney, June 2023. Accessed on 2024-07. Available at: <http://visionofhumanity.org/resources>.

Jacinavicius, F.C., Bassini-Silva, R., Mendoza-Roldan, J.A., Pepato, A.R., Ochoa, R., Welbourn, C. & Barros-Battesti, D.C. (2018) A checklist of chiggers from Brazil, including new records (Acari: Trombidiformes: Trombiculidae and Leeuwenhoekiidae). *ZooKeys*, 743, 1–41. <https://doi:10.3897/zookeys.743.22675>

Kim, D.Y., Kim, S., Song, H. & Shin, S. (2024) Phylogeny and biogeography of the wingless orthopteran family Rhaphidophoridae. *Communications Biology*, 7, 401. <https://doi:10.1038/s42003-024-06068-x>

Kim, G., Kim, D.Y., Kim, T. & Shin, S. (2024) Taxonomic review of the genus *Hexacentrus* Serville, 1831 (Orthoptera: Tettigoniidae) in Korea: An integrative study to confirm cryptic species. *Journal of Asia-Pacific Entomology*, 27, 102249. <https://doi:10.1016/j.aspen.2024.102249>

Kociński, M., Grzywacz, B., Hristov, G. & Chobanov, D. (2021) A taxonomic outline of the *Poecilimon affinis* complex (Orthoptera) using the geometric morphometric approach. *PeerJ*, 9, e12668. <https://doi:10.7717/peerj.12668>

Lessa, T., Stropp, J., Hortal, J. & Ladle, R.J. (2024) How taxonomic change influences forecasts of the Linnean shortfall (and what we can do about it)? *Journal of Biogeography*, 51, 1365–1373. <https://doi:10.1111/jbi.14829>

Linnæus, C. (1758) *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Editio decima, reformata (tome 1). Laurentii Salvii. I–IV +1–824.

Lu, M. & He, F. (2017) Estimating regional species richness: The case of China's vascular plant species. *Global Ecology and Biogeography*, 26, 835–845. <https://doi.org/10.1111/geb.12589>

Mayr, E. (1969) *Principles of Systematic Zoology*. McGraw-Hill, New York, pp 428.

Nakamura, G., Stabile, B.H., Frateles, L.E., et al. (2024) The macroecology of knowledge: Spatio-temporal patterns of name-bearing types in biodiversity science. Preprint. 1–47. <https://doi.org/10.32942/X28D1M>

Nattier, R., Robillard, T., Amedegnato, C., Couloux, A., Cruaud, C. & Desutter-Grandcolas, L. (2011) Evolution of acoustic communication in the Gomphocerinae (Orthoptera: Caelifera: Acrididae): Evolution of acoustic communication in the Gomphocerinae. *Zoologica Scripta*, 40, 479–497. <https://doi.org/10.1111/j.1463-6409.2011.00485.x>

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

R Core Team (2024). R: A language and environment for statistical computing. Available at: <https://www.R-project.org/>.

Raja, N.B., Dunne, E.M., Matiwane, A., Khan, T.M., Nätscher, P.S., Ghilardi, A. M. & Chattopadhyay, D. (2022) Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature Ecology & Evolution*, 6, 145–154. <https://doi.org/10.1038/s41559-021-01608-8>

Rentz, D. & Su, Y.N. (2019) *A guide to crickets of Australia*. Csiro Publishing.

Riede, K. (2018) Acoustic profiling of Orthoptera: present state and future needs. *Journal of orthoptera research*, 27, 203–215. <https://doi.org/10.3897/jor.27.23700>

Rodrigues, A. S. L., Gray, C. L., Crowter, B. J., Ewers, R. M., Stuart, S. N., Whitten, T., & Manica, A. (2010) A global assessment of amphibian taxonomic effort and expertise. *Bioscience*, 60(10), 798–806. <https://doi.org/10.1525/bio.2010.60.10.6>

Ronquillo, C., Stropp, J., Medina, N. G., & Hortal, J. (2023) Exploring the impact of data curation criteria on the observed geographical distribution of mosses. *Ecology and Evolution*, 13(12), e10786. <https://doi.org/https://doi.org/10.1002/ece3.10786>

Saliba, E.M. (2023) *Understanding and managing Zoological Nomenclature in the era of Big Data and Open Science*. Doctoral dissertation, Sorbonne Université. 225pp. English. NNT: 2023SORUS460. tel-04402685.

Song, H. (2010) Grasshopper systematics: Past, present and future. *Journal of Orthoptera Research*, 19, 57–68. <https://doi.org/10.1665/034.019.0112>

Song, H., Béthoux, O., Shin, S., Donath, A., Letsch, H., Liu, S., ... Simon, S. (2020). Phylogenomic analysis sheds light on the evolutionary pathways towards acoustic communication in Orthoptera. *Nature Communications*, 11(1), 4939. <https://doi.org/10.1038/s41467-020-18739-4>

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>

Uluar, O., Yahyaoğlu, Ö. & Çiplak, B. (2021) Review of *Uvarovistia* (Orthoptera: Tettigoniidae; Tettigoniinae): genetic data indicate a new cryptic species, *Uvarovistia munzurensis* sp.n.. *Zootaxa*, 4952, 152–168. <https://doi.org/10.11646/zootaxa.4952.1.9>

Van der Loo, M.P.J. (2014) The stringdist package for approximate string matching. *The R Journal*, 6, 111.

Waloff, N. (1990) Sir Boris uvarov (1889-1970): The father of acridology. *Annual review of entomology*, 35(1), 1–24. <https://doi.org/10.1146/annurev.ento.35.1.1>

Weber, A. (2021) Natural history collections and empire. Em *The Routledge Handbook of Science and Empire* (p. 80–86). Routledge.

Wege, J.A., Thiele, K.R., Shepherd, K.A., Butcher, R., Macfarlane, T.D. & Coates, D.J. (2015) Strategic taxonomy in a biodiverse landscape: a novel approach to maximizing conservation outcomes for rare and poorly known flora. *Biodiversity and Conservation*, 24, 17–32. <https://doi.org/10.1007/s10531-014-0785-4>

Wilkinson, B.H., Ivany, L.C. & Drummond, C.N. (2021) Estimating vertebrate biodiversity using the tempo of taxonomy – a view from Hubbert’s peak. *Biological Journal of the Linnean Society*. *Linnean Society of London*, 134, 402–422. <https://doi.org/10.1093/biolinnean/blab080>

World Spider Catalog (2024) Version 25.5. Accessed on 2024-10. Available at: <https://wsc.nmbe.ch>

Zizka, A., Silvestro, D., Andermann, T., et al. (2019) CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10, 744–751. <https://doi.org/10.1111/2041-210x.13152>

Supplementary Material

Supplementary Figures

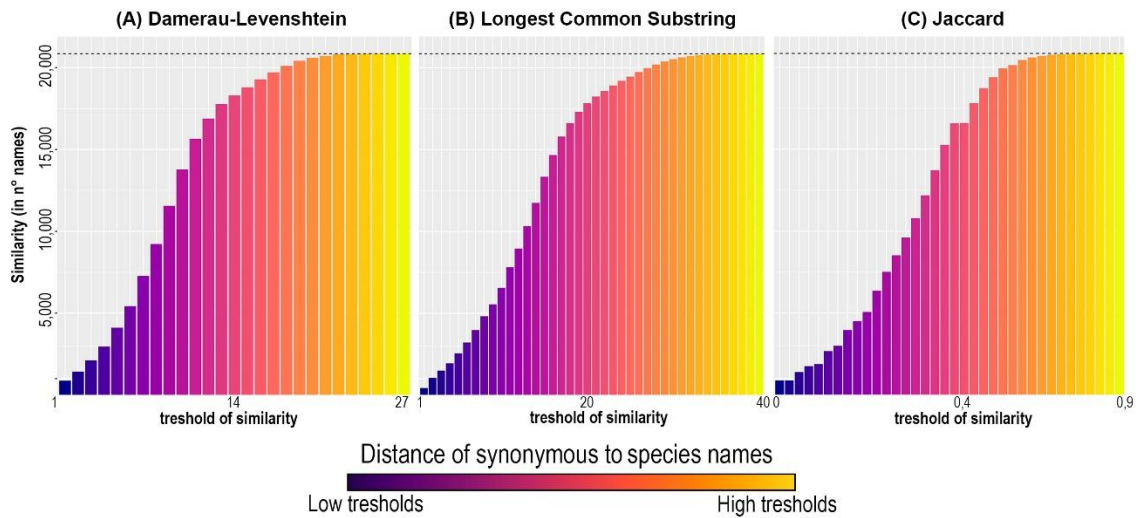


Fig S1. Sensitivity analysis between the similarity of the names present in the synonym list and the corresponding valid species names using the distance algorithm of Damerau-Levenshtein (a), Lonest Common Substring (b), and Jaccard Index (c).

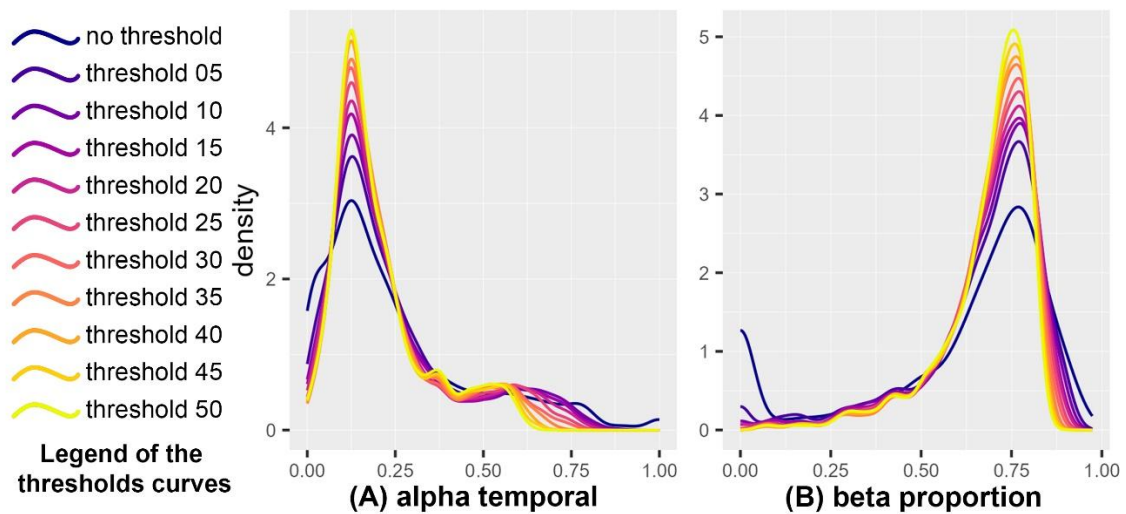


Fig S2. Density curves of alpha temporal (A) and beta proportion (B) indices by different thresholds.

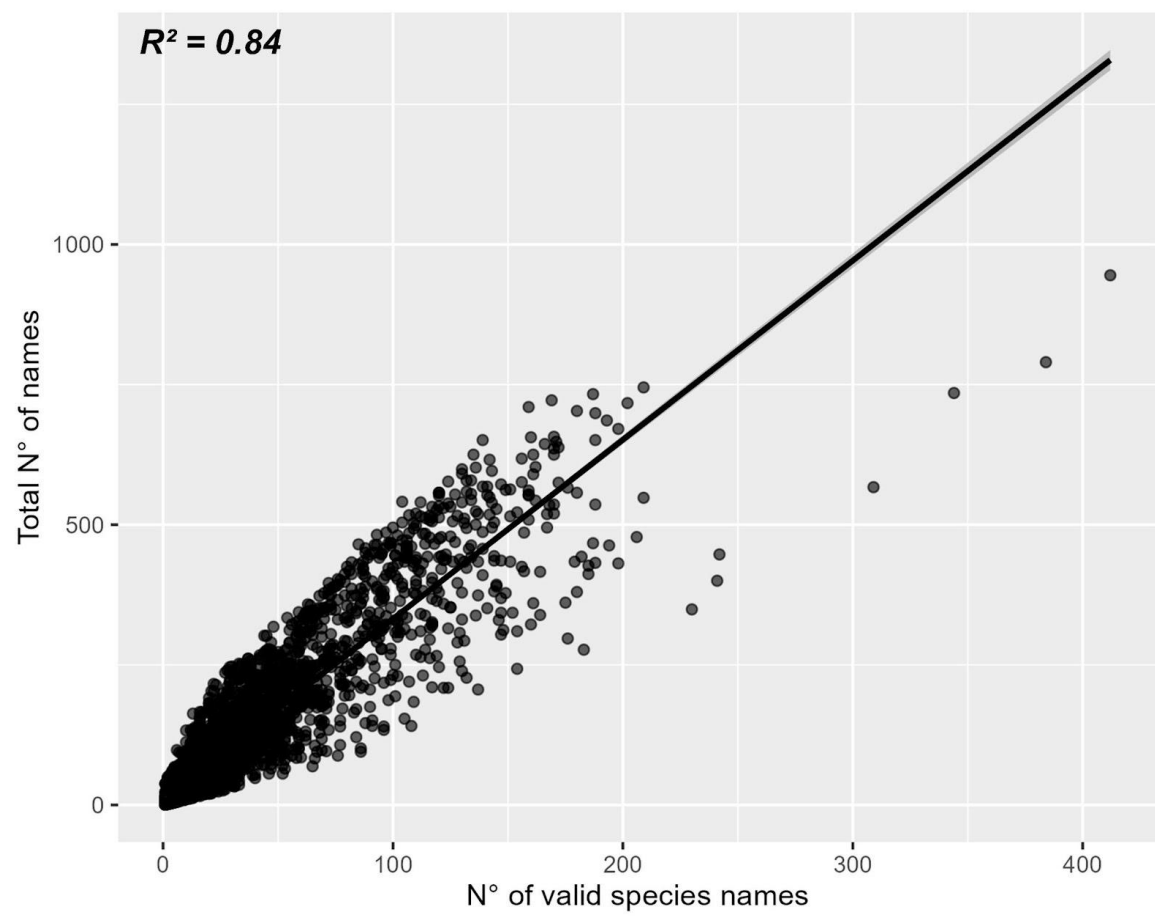


Fig S3. Correlation between the number of valid species names and the total number of names.

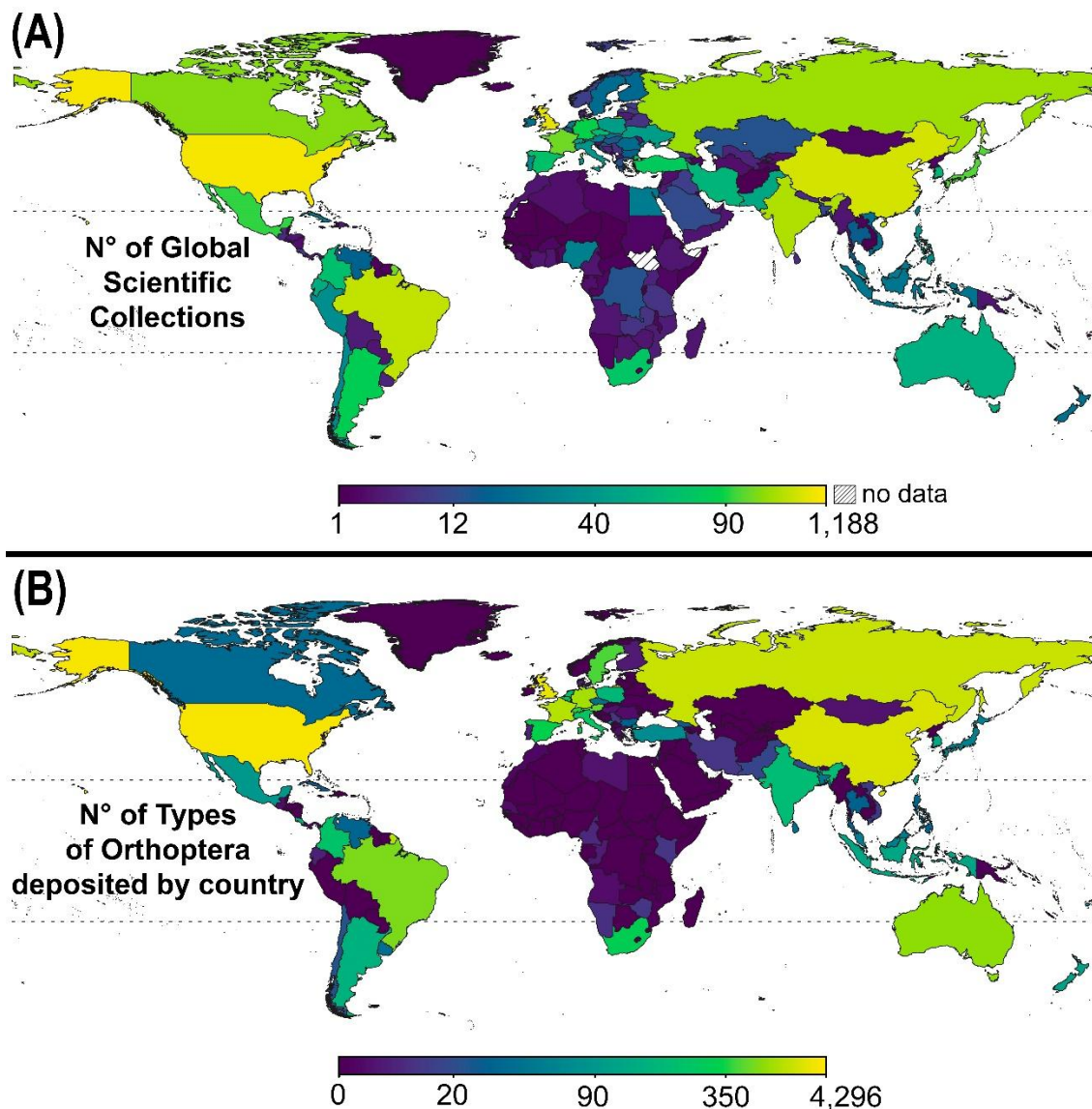


Fig S4. Distribution of global scientific collections (A) and types (i.e., holotypes, syntypes, lectotypes or neotypes) of Orthoptera (B). Data source: <https://scientific-collections.qbif.org/> and <https://orthoptera.speciesfile.org/>.

Supplementary Appendices

Appendix S1. The ‘nomenclatural’ lumping

Within the synonymic lists available for species of the order Orthoptera, both in the Catalogue of Life (CoL) and the Orthoptera Species File (OSF), we observed that, in addition to true synonyms resulting from the lumping of two or more species to a unique species (see Parsons et al. 2022; Stropp et al. 2022; Lessa et al. 2024),

there were also names that were actually ambiguous synonym (a synonym name that belongs to more than one species), spelling errors, and incorrect combinations. The concept of a synonymic list, by itself, encompasses such names in many cases in Zoology (Dubois et al. 2000; Gardner and Hayssen 2004; Saliba 2023). However, since we cannot precisely separate the other types of names from the true synonyms, here we opted to use lumping, considering the synonymic list of each currently valid species name in its entirety, or the largest amount of names it can encompass. We excluded only synonyms resulting from the presence of subgenus syntax (e.g., *Gryllus campestris* = *Gryllus* (*Gryllus*) *campestris*).

Additionally, we performed a sensitivity analysis to understand if it would be possible to extract spelling errors from the synonymic lists of species, using three different string dissimilarity methods between each synonym name and the valid name to which it belongs, with the “stringdist” package in R environment (Van der Loo 2014). We observed that, with lower thresholds, few names were filtered out, and different algorithms behaved quite similarly. This led us to opt for using the synonymic list in its entirety (i.e., without removing any names), as there was a probability that, by increasing the thresholds, the similar names identified would actually be true synonyms (Fig. S1).

Appendix S2. Assessing the sensitivity of alpha temporal and beta proportion indices.

The alpha temporal (AT) and beta proportion (BP) indices are primarily based on the geographic distribution of species names, the number of changes associated with these names, and the year of species name authorship. AT relies on the authorship year accompanying each species name across space, while BP

depends on the quantity of species names and the taxonomic changes associated with these names across space. Therefore, our metrics operate with a limited set of assumptions and do not require any predefined thresholds, allowing their application in areas with low species name richness, such as regions with extreme latitudes, high altitudes, or taxonomically understudied areas. This contrasts with other biodiversity data quality metrics that often require minimum thresholds to make predictions and extrapolations (see Soberón et al. 2007; Meyer et al. 2016; Tessarolo et al. 2021; Castro-Souza et al. 2024).

However, we conducted tests with different thresholds to assess how cutting configurations of the minimum number of valid species names (i.e., richness of names) affect the spatial coverage of index results. To calculate AT and BP, we used the following thresholds: none, 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 valid species names. With maps constructed for each threshold, we calculated density curves, allowing us to observe how each index behaved under different configurations. Based on these results (Fig. S2), we maintained the threshold (no limit) for AT D and BT models presented in this study. Thus, no grid cell with low name richness was discarded, enabling us to understand alpha and beta taxonomy across most of the globe.

References

- Soberón, J., Jiménez, R., Golubov, J. & Koleff, P. (2007) Assessing completeness of biodiversity databases at different spatial scales. *Ecography*, 30, 152–160. <https://doi:10.1111/j.2006.0906-7590.04627.x>
- Meyer, C., Weigelt, P. & Kreft, H. (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters*, 19, 992–1006. <https://doi.org/10.1111/ele.12624>
- Tessarolo, G., Ladle, R.J., 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, 1743–1755.
<https://doi.org/10.1111/ecog.05793>

Appendix S3. Distribution of global scientific collections and types of orthoptera

We examined the number of collections in each country using the Global Registry of Scientific Collections (<https://scientific-collections.gbif.org/>, accessed on 2024-10-06). Using this data, we mapped the quantity of collections per country and used this pattern as a proxy for taxonomic effort, considering that the material stored in these collections is essential for taxonomists. It is important to note that we did not exclusively filter collections containing Orthoptera; therefore, our map provides a general overview (Fig. S4A).

Subsequently, using our list of valid species names (see A checklist of Orthoptera names in methods), we searched the Orthoptera Species Files (OSF) for records of holotypes, syntypes, lectotypes, and neotypes, finding information on type material for 21,341 records, belonging to 20,106 species, and deposited in 76 countries. To associate each record with the collection where it is deposited, we matched the acronyms of non-duplicated collections from OSF with the Global Registry of Scientific Collections (<https://scientific-collections.gbif.org/>); for other collections, we conducted manual searches in OSF and the original species literature. Using this data, we mapped the quantity of Orthoptera types deposited by country and used this pattern as a proxy for taxonomic impediment (Fig. S4B).

CAPÍTULO II

Mapping Ignorance to Uncover Shortfalls in the Knowledge on Global Orthoptera Distribution

Artigo publicado na revista *npj Biodiversity*

Link: <https://doi.org/10.1038/s44185-024-00059-1>

Mapping ignorance to uncover shortfalls in the knowledge on global Orthoptera distribution

Rodrigo Antônio Castro-Souza^{1*}; Geiziane Tessarolo²; Juliana Stropp³; José Alexandre Diniz-Filho²; Richard J. Ladle⁴; Neucir Szinwelski⁵; Joaquín Hortal^{2,6} Thadeu Sobral-Souza¹

(1) Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso, 78060-900, Cuiabá, MT, Brazil

(2) Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil

(3) Department of Biogeography, University of Trier, Trier, Germany

(4) Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, Alagoas, Brazil

(5) Laboratório de Orthoptera, Centro de Ciências Biológicas e da Saúde, Universidade Estadual do Oeste do Paraná, Brazil

(6) Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

corresponding author: rodrigodesouzaac@gmail.com

Abstract

Understanding the spatiotemporal distribution of species is fundamental for ecology, evolution and conservation. However, this and other aspects of biodiversity knowledge suffer from shortcomings and biases. Quantifying and mapping biodiversity knowledge shortfalls is therefore crucial to ascertain the current quality and completeness of biodiversity data, prioritize sites for (re)sampling, or plan conservation interventions. Here, we compile a comprehensive dataset of Orthoptera occurrences, and use it to create a global ‘ignorance map’ based on taxonomic, survey and temporal completeness, and on survey and temporal evenness of the inventories. We hypothesize that knowledge of Orthopteran biodiversity is relatively poorer in tropical regions compared to temperate regions, and in the south compared to the northern hemisphere. Due to biocultural factors, we expect regions in the tropics and the Global South to have lower levels of completeness and evenness in time and space for all the studied aspects of biodiversity information. Our findings show gaps in the knowledge of orthopteran distributions, which are characterized by low survey and temporal evenness in tropical regions, but also in many temperate regions (e.g., most of the countries in temperate Asia). The combination of multiple dimensions of biodiversity knowledge (taxonomic, spatial and temporal) reveals that biogeographic interpretations based on only one component can lead to an illusion of completeness. We believe that the novel framework used in our study can guide future research towards building more accessible maps of biogeographical ignorance for entire groups.

Keywords: Biodiversity Knowledge Shortfalls; ignorance maps; Macroecology; Biogeography; Digitally Accessible Data; Data Quality; Orthoptera.

INTRODUCTION

As the world faces a global extinction crisis, good-quality data on the distribution and attributes of biodiversity is increasingly needed. In this context, quantifying how well we know biodiversity becomes imperative, for it allows scientists and policy makers to identify where biodiversity information is most accurate and complete^{1,2,3,4}, thereby providing a baseline for defining priorities for new biodiversity research and conservation^{5,6}. In the past two decades or so, an enormous volume of digitally-accessible primary biodiversity data opened new opportunities to quantify coverage and completeness of biodiversity information⁷. Freely-available biodiversity information networks compile huge amounts of primary biodiversity records from natural history collections, research articles and technical reports, among other sources, in the form of online database platforms^{8,9}. However, these digital biodiversity data are often incomplete and both spatially and temporally biased^{10,11,12}.

Addressing these limitations is of paramount importance to draw robust descriptions of biodiversity patterns from the available data, and ultimately ensure well-informed conservation decision-making. This kind of challenge can be addressed through the creation of ‘ignorance maps’ that characterize the distribution of areas where knowledge is limited¹³. The application of ignorance maps to biodiversity^{11,14} can be addressed by mapping out the various components of data limitations at broad temporal and spatial scales^{6,15,16}. Here, three key data attributes allow us to map the accuracy, coverage and completeness of biodiversity data: the species name, the location of the records, and the date they were recorded.

First, the most fundamental component of digitally-accessible biodiversity data is related to taxonomy. Records may have species names inaccurately (i.e., incorrect identifications, use of synonyms, etc.) or incompletely assigned (i.e. identified to genus or family level)^{3,12,17}. The challenge of assessing taxonomic accuracy is exacerbated by the absence or lack of standardization in the information about who identified species’ records¹², which can lead to discarding much valuable data from online and citizen science databases. Moreover, certain records refer to species that are yet to be described^{3,18}. Tropical regions bear the brunt of this challenge due to their high taxonomic uncertainty and megadiversity^{19,20}. The identification of sites requiring further taxonomic effort represents a potential strategy

to improve the coverage and completeness of taxonomic information. However, digitally-accessible biodiversity data are also biased towards certain taxa, which are either better sampled^{21,22} or more abundant in the communities²³, or have undergone more extensive digitization efforts¹². These taxonomic biases imply that data coverage assessments should be conducted specifically for each group that is sampled or studied altogether.

Second, survey coverage and completeness can range from inadequate to overly intensive, generating gaps and biases in perceived species distributions, and in the diversity of species documented for local and regional inventories across space and time^{24,25}. Here, we use species accumulation curves to assess survey coverage, assuming that inventory completeness can be measured as the rate of the adding new species to the inventory with the increment in database records in each spatial unit¹⁶. Notably, well-sampled regions often align with larger and more connected natural areas with higher accessibility^{26,27,28}. Thus, quantifying survey effort across geographic space through measures of survey completeness^{24,29} allows highlighting well- and poorly-sampled sites at large spatial scales^{8,30,31}.

Third, the temporal pattern of biodiversity surveys is a critical aspect of biodiversity knowledge^{12,32,33,34}. The quality of biodiversity knowledge decays over time due to the dynamism of natural systems, taxonomic rearrangements, and loss of information³³, a decay that is exacerbated by human-driven disturbances. Further, certain sites may exhibit skewed records for specific years¹¹. While some regions, such as Europe, North America, and Australia, boast multi-year sampling data for some taxa thanks to their long-term research tradition, others, such as Africa, hold limited temporal completeness and uneven temporal coverage^{9,12,26}. Such disparities highlight the potential importance of considering the number of sampled years (temporal richness) when assessing data quality. However, although some studies describing such quality have measured temporal coverage^{11,12}, or temporal decay of information^{16,32}, many others disregard the temporal component³⁵.

Fourth, the study of survey evenness across different regions and temporal periods can help identify sites and regions holding data appropriate for biodiversity and macroecological studies, or with enough temporal resolution so as to assess temporal shifts in their communities. Here, we use the Hill Numbers Diversity

framework³⁶ to assess both survey evenness and temporal evenness in each spatial unit, based on the relative numbers of records per species, and of years per species, respectively. The assumption behind these metrics is that sites with low evenness (i.e., an unbalanced representation of species or years) may generate biased estimates in macroecological or evolutionary models. Those areas would require further surveys or, at least, be used with caution in investigations, as they likely lack data of enough quality for assessing diversity–environment relationships, describing phylogeographic patterns, or estimating past global change impacts on biodiversity.

Orthoptera, which includes grasshoppers, crickets, katydids and relatives, have a rich taxonomic legacy, with approximately 29,500 valid species, making it the sixth most species-rich insect order³⁷. Orthopterans play several key roles in ecosystems^{38,39,40,41}, and hold high cultural significance in various countries^{42,43,44,45}, as they include both food species and agricultural pests⁴⁵. Indeed, orthopterans are relatively well-known taxonomically compared to other invertebrate groups, mainly owing to their agricultural importance⁴⁶. In addition, their taxonomy is regularly updated at the Orthoptera Species File Online^{45,46,47}. Further, information on their distribution can be obtained from multiple digitally-accessible data repositories, which make thousands of primary records available^{48,49,50} (see <https://orthoptera.speciesfile.org/>). However, research focus has been biased towards grasshopper pest species and, in some cases, pests of specific grasses or cultivars^{46,51,52}, so knowledge on the distribution of most species and higher taxa of this group may still be limited⁴², particularly in megadiverse regions.

Here we aim to construct a global ignorance map for Orthoptera by combining analyses of different biodiversity shortfalls. Based on the known geographical biases in biodiversity inventories^{3,12}, we hypothesize that the digitally accessible data for Orthoptera is less complete for tropical compared to temperate regions, and for the southern compared to the northern hemisphere. More specifically, we predict that northern temperate regions will have higher levels of taxonomic, survey and temporal completeness, while the tropics and the Global South will show higher evenness in both time and space. To test these hypotheses, we developed metrics of taxonomic, spatial, and temporal data quality. Finally, we combined all metrics into a single global

ignorance map to provide a unique biogeographic panorama of global biodiversity knowledge of orthopterans.

RESULTS

We compiled data from 23 digital repositories storing primary information on Orthoptera (Arthropoda, Insecta), totaling 4,911,359 occurrence records (Figure 1a; Table 1). The data filtering process led to the exclusion of 2,513,619 records, resulting in a final dataset of 2,397,740 species occurrences (48.82% of the initial number of records), pertaining to 16,179 valid species (Supplementary Figure 1). In addition, we were able to recover 408,377 unique records above the species level (i.e., genus, tribe, family, or order) that included information on the geographical coordinates and year of collection (Supplementary Figure 2). These records were subsequently used to calculate taxonomic completeness (TaxC). The calculation of data quality indices based on this information revealed significant data gaps and biases in many regions, which were never before measured for Orthoptera (Figure 2a-e).

Strikingly, it was not possible to calculate data quality indices for most of the extent of tropical regions, due to limited data availability in most of South America, Africa, Asia, and northern Oceania. In these regions, where data is accessible, taxonomic completeness indicates that more than 50% of the records have yet to improve taxonomic refinement. The highest TaxC values were observed in the Northern Hemisphere. In particular, central North America, most European countries, and small scattered areas in eastern Asia, South Korea and Japan exhibit significant proportions of records taxonomically refined at the species level. In the Southern Hemisphere, TaxC values are generally higher than in the tropics (i.e. $\text{TaxC} \geq 0.5$), although they still fall short of the levels observed in the Northern Hemisphere with the exception of some taxonomically consistent inventories scattered throughout South America, Africa, and Oceania (Figure 2a).

The tropical regions also show low survey completeness (SC), indicating a general need for additional surveys except perhaps in the north of Australia, which holds well-sampled inventories. In general, all the Southern Hemisphere beyond the tropics is massively under-sampled, except for some relative well-sampled sites in South Australia and New Zealand (Figure 2b). Although SC was high for most regions

of the Northern Hemisphere, parts of North America, regions bordering Europe, Asia, and Africa, as well as Japan, are still insufficiently sampled. Temporal completeness (TC) is low along the entire globe. Only a few patches scattered throughout the North America and some areas in central Europe show temporal completeness values higher than 0.25 (Figure 2c). In contrast, we did not find a clear spatial pattern of survey and temporal evenness (SE and TE, respectively) at a global scale. While some grid cells exhibit sampling biases towards few species or few time periods, others placed nearby present long-term series of surveys with similar abundances among species, independently of the region (Figure 2d, e).

The overall ignorance map shows a general lack of information across most of the tropics (Figure 3), due to the low taxonomic, sampling and temporal completeness. However, a few small isolated patches in southern Africa and northern Oceania present relatively low ignorance ($0.5 \geq IM \geq 0.25$). Ignorance values in the northern Hemisphere are low in several regions of North America and Central Europe ($IM \leq 0.25$). However, the coastal zones of the United States of America, Southern and Eastern Europe, and most regions of India and China present high ignorance values (Figure 3).

DISCUSSION

Our analyses provide evidence that digitally-accessible knowledge on the global distribution Orthopteran biodiversity is limited in tropical regions in particular, and in the Global South in general. Conversely, the temperate regions of the Global North present areas of relatively high quality and completeness of biodiversity knowledge, especially in North America and Central Europe, as well as in some extra-tropical areas of the southern hemisphere. Although these results confirm our preliminary hypothesis about the distribution of ignorance, there are also many deficient inventories in these data-rich regions, and knowledge on Orthoptera is still very limited in many temperate areas, particularly in Asia. Importantly, the comparison of our overall ignorance map with those of the indices describing different dimensions of knowledge suggests that interpretations based solely on one of these components could create an illusion of good knowledge coverage for certain regions.

The lack of taxonomic completeness in the tropics contributes to the generation of a taxonomic latitudinal bias (as proposed by Freeman and Pennell¹⁹). This bias may be partly caused by a significant 'species debt' in tropical regions due to the lack of standardized criteria for recognizing species compared to temperate regions⁵³. A significant portion of such debt is due to records that have already been cataloged but have not been refined or taxonomically revised¹⁷. This phenomenon is likely highly pronounced in small-sized and abundant animal taxa such as insects⁵⁴, and is also exacerbated by the shortage of taxonomists^{55,56}. The high species richness of tropical regions makes taxonomic refinement difficult, as the discovery, description and identification of species requires more time and a higher degree of expertise compared to temperate regions, which typically harbor fewer species and have a longer history of naturalist studies and taxonomic work^{19,53}. Nevertheless, the high taxonomic completeness observed in many temperate regions should be interpreted with caution when such values are accompanied by low survey completeness, as in, e.g. Southern Europe, the East and West Coasts of North America, or the southwest edge of Australia, because the lack of surveys may be reflecting lower levels of the taxonomic effort that should be an integral part of standardized inventories.

The low availability of digitally accessible data arises as a fundamental issue for our knowledge on the geographical data for distribution of Orthoptera, especially in tropical regions and most of Asia. The pattern of low survey completeness observed here reflects a general scarcity of species records, which may be the result of either lack of inventories or lack of data mobilization, or both, arising from limited resources or other impediments for data digitization^{9,12}. Indeed, such limited geographical coverage stands out when considering inventories with low taxonomic completeness. Both taxonomic completeness and survey completeness are affected by the challenges to Orthoptera taxonomy. Besides counting on up-to-date taxonomic keys to identify rare individuals from new samples, it is often also crucial to compare with type specimens or good descriptions⁵⁷, which may not be possible, especially for species described long time ago, or that lack information on male genitalia (a crucial characteristic for the diagnosis of many orthopterans^{58,59}). This is further complicated because many type specimens or descriptions are not accessible, there is an absence of comprehensive and up-to-date taxonomic keys⁶⁰ and low economic investment in many high-biodiversity regions^{9,61}. The exceptions are pest species and

some groups that have historically received more attention⁴⁶. This may have added an additional geographical bias to the effort devoted to orthopteran inventories, as pest species tend to predominate in grassland environments⁴⁶, which are more abundant at temperate latitudes. Thus, although the historical gaps in the study of Orthoptera in the tropics⁴² are likely caused by the generalized lack of resources for biodiversity research, they have been also exacerbated by the highest attention devoted by agricultural sciences to the orthopterans of temperate regions.

Strikingly, the high taxonomic and survey completeness in countries of the northern hemisphere is not always accompanied by high values of survey and temporal evenness, contradicting our hypothesis for these regions. This can be partly explained by the higher global availability of data on Orthoptera for certain regions, as seen in birds, mammals and amphibians⁹, in plants¹², and also in multi-taxa analyses²⁶. This tends to significantly increase the number of unique records for some local species and years already sampled, causing a significant imbalance in data evenness. Also, species records are subject to biases related to accessibility, as locations placed near to access routes (i.e. roads and rivers), research centers, conservation units, and large connected forest fragments are more frequently sampled^{9,27,28,62,63}. Further, the low taxonomic completeness that is widespread in tropical regions can distort the observed patterns of survey and temporal evenness, as many recently described species from these regions may be present in digitally-accessible data only from their type specimens, while their presence in older records is still pending from their revision using an updated taxonomy. All these factors together can result in the absence of smooth geographical patterns of data equability observed in our large-scale maps.

Our measure of temporal completeness reveals that most digitally accessible species records were collected from a limited number of years (i.e., low richness of years), for both tropical and temperate regions. Such lack of years with records hinders our ability to comprehend the decline of Orthopteran diversity over time⁶⁴, because the temporal coverage of sampling records is discontinuous and unevenly distributed along different locations. A wider temporal spread of records was only observed in some scattered areas throughout North America and Europe, similar to the historical distribution of sampling effort reported for other taxa^{9,61}, thus reflecting

that the higher financial and scientific support of the United States of America and several European countries^{65,66} extends to the field of Orthopterology. Furthermore, such capacity has been accompanied by the effects of scientific colonialism on the historical and current collection of data about Orthoptera, where taxonomic legacy has a clear impact on the global distribution of knowledge. Only recently have both biodiversity research in general, and depositions of type material in Museums in particular, started to be conducted mainly at local institutions in countries throughout the tropics⁶⁵.

The observed pattern of temporal completeness may also reflect to some extent the deficiency in data digitization, especially in the tropics and parts of temperate Asia. Here, a digitalization effort specifically tailored to retrieve data prior to the year 2000, originating from agricultural surveys, fieldwork or environmental impact reports, would complement the currently available knowledge, thus providing a more comprehensive historical coverage and temporal evenness considering the chronological gaps associated with the distribution data and the low temporal completeness of information for many locations^{25,35,67}.

Furthermore, the mapped values of different aspects of ignorance highlight that even temperate regions of the northern hemisphere, where we expected to have a good quality of biodiversity knowledge (i.e., low taxonomic, sampling, and temporal completeness, and high survey and temporal evenness), suffer from limited taxonomic refinement, inventory integrity, limited diversity of years sampled, and an imbalance in the number of records among known species and years. This suggests that low ignorance values, whether in the tropics or in temperate regions, may result from a lack of digital access to a large portion of biological collections worldwide^{65,68,69}. This represents a significant impediment to the advancement of scientific knowledge and should be rectified, at least in the collections of developed countries. Here, technological advances such as cameras, cell phones, and tablets can allow the recording of primary data by people who live and interact directly with biodiversity, such as those living in protected land, field biologists, and nature photographers⁷⁰. However, despite such potential, many orthopterans can only be accurately identified after examining the male genitalia, and this is not often possible from in situ images taken from citizen science platforms, and there is still a significant

lack of globally-available biogeographic knowledge coming from Orthoptera in databases such as EcoRegistros or iNaturalist (see Table 1).

Regardless of other alternative initiatives for gathering new data, there is clearly an urgent need to improve biodiversity data curation and to provide financial support to establish an open and collaborative data mobilization infrastructure in biological collections^{68,69,71}, through either local policies in each country, or global agreements such as the Convention on Biological Diversity (<https://www.cbd.int/convention/>). This is especially important considering that accidents in collections can result in immeasurable losses^{72,73}, that could be partially mitigated through digitized information. In addition, the already-existing platforms aimed at integrating and making literature data available need to pursue optimization initiatives. A good example is that of the Orthoptera Species File (<http://orthoptera.archive.speciesfile.org/>), which migrated thousands of data records to the Darwin Core standard (<https://orthoptera.speciesfile.org/>), the most comprehensive taxonomic standard for biodiversity data^{69,74}.

Nevertheless, the solution to the above-mentioned problems goes beyond relying solely on bioinformatics⁷⁵. Training and educating new taxonomists should also be a priority^{56,76}, along with the development and updating of user-friendly identification keys for other scientists, and the general public. Extensive engagement of taxonomists in the development of online data availability initiatives⁷, and collaboration with citizen science platforms is also necessary, both by validating and curating species identifications^{77,78}, and by applying these platforms in teaching and training activities⁷⁹. All these initiatives will ultimately engage some sectors of the general public into higher-quality citizen biodiversity science^{80,81}. Such measures will significantly contribute to the construction, security, and availability of open knowledge over time, gradually filling the gaps highlighted here, subsidizing studies in taxonomy, biogeography, and conservation³, as well as to conservation decision-making based on Orthopteran data by researchers in the future.

Lastly, it is also worth noting that assessments of ignorance for specific groups or at smaller study scales may show local differences from the general pattern observed at a global scale. Our analyses highlight the shortfalls of knowledge on the macroecological patterns that arise from the large-scale mechanisms governing

nature⁸². But these ignorance maps can be also useful to address these data gaps within specific groups and/or at smaller regions.

In conclusion, we believe that this study offers an innovative and simplified framework that can guide future research on Orthoptera, and can also be replicated for other taxa, making the construction of ignorance maps more accessible. Nevertheless, more research on the factors that cause gaps and biases in digitally accessible Orthoptera data is needed. With this work, we hope we have provided a baseline for exploring future challenges as data become more extensive and available. Without robust and reliable data, our quest to uncover the ecological and evolutionary mechanisms driving biogeographical patterns and to develop effective approaches to conserve biodiversity in a changing environment will remain fundamentally compromised.

METHODS

A global dataset of Orthoptera occurrence

We compiled data from 23 digital repositories storing primary information on the distribution of Orthoptera (Table 1). Additional information on the extraction process used for each repository can be accessed in the Supplementary Note 1. The primary records were divided into two distinct subsets: (1) occurrences encompassing species/subspecies-level data; and (2) occurrences extending beyond the species taxonomic level, up to the order level (Figure 1b).

We also built a list of currently accepted species, alongside their corresponding taxonomic arrangements (synonyms) according to the classification proposed by the Orthoptera Species File catalog - (OSF)⁴⁸, as available on the Catalogue of Life platform (<https://doi.org/10.48580/d4sw-388>) on August 2021. Using the species list, we organized the taxonomic nomenclatural information for each species and subspecies across all repositories, excluding fossil species, and synonymized subspecies with species. The final species list includes 28,309 valid extant species (Supplementary Dataset).

We applied filters to identify and remove records that lacked the necessary information for subsequent analyses: (i) *taxonomic adequacy filter*—we used the valid

species list (described above) to exclude occurrences that were not in accordance with the proposed taxonomic classification. For records without species names, we did not use taxonomic adequacy (see Data Quality analysis section); (ii) *spatial filter*—we removed occurrences without geographical coordinates, and also excluded records with only descriptive locations (e.g., municipality, state, country) or dubious coordinates; (iii) *temporal filter*—we kept only the occurrences that include the year of collection; and (iv) *land boundaries filter*—only georeferenced occurrences for the terrestrial part of the planet, including islands, were computed. To do this, we used the Natural Earth platform, at a 1:10 m resolution (<https://www.naturalearthdata.com/>). Records that did not cover these areas were excluded (Figure 1b).

Then, all occurrences were divided into two groups: species/subspecies occurrences and records with identification at a higher taxonomic level, where we applied two additional filters: (v) *ambiguities filter*—duplicate records were removed for species/subspecies and records above the species level, separately; and (vi) *fossil filter*—excluding all fossil species, which were not considered in our analyses.

Indices of data quality

We calculated five indices of data quality: taxonomic completeness (TaxC); survey completeness (SC), based on species accumulation curves^{11,16,29}; temporal completeness (TC); and survey and temporal evenness (SE and TE, respectively), both of them calculated using a Hill-Diversity method slightly adapted by us^{36,83}. All indices were computed based on the information for terrestrial grid cells at 1° resolution and then plotted into global maps (Figures 1c and 1d).

Taxonomic Completeness: For each cell (pi), we calculated taxonomic completeness using the equation where TaxC is the result of the number of primary occurrences identified at the species taxonomic level (s), divided by the total number of records (t).

$$\text{Eq. 1} \quad \text{TaxC} = [s(pi)] / [t(pi)]$$

The total number of records (t) corresponds to the sum of all species records (s) and all identifications available only above the species level (i.e., genus, tribe, family, or

order) (Supplementary Figure 3). TaxC values range from 0 to 1; when TaxC = 0 means that none of the occurrences are identified at the species level, thus indicating a large Linnean shortfall. While values close to 1 indicate that most records are identified to the species taxonomic level, thus corresponding a well-sampled cell from a taxonomic point of view. Thus, TaxC index allows us to assess, at least in part, the geographical distribution of the extent of the Linnean shortfall, based on the data quality in relation to taxonomic refinement of occurrences (Figure 1c).

Survey Completeness: We calculated survey completeness (SC) for each cell based on species accumulation curves, where the final slope angle of each curve represents the rate of accumulation of new species in the inventory²⁹. The order of entrance of records in the curves was smoothed by running 100 random permutations using the function `specaccum` from the R package `vegan`⁸⁴, and the final slope was calculated as the average difference in the number of species observed in the final and next-to last records. This final slope was subtracted from 1, thus obtaining completeness values ranging from 0 to 1. SC was performed in R environment using the function `Accum_curve` provided in Tessarolo et al.¹⁶. Values of SC close to 1 indicate an almost complete inventory (Figure 1c).

Temporal Completeness. We measured the temporal completeness (TC) at each cell (pi) based on the number of sampled years divided by the highest range of sampled years found in our database (Supplementary Figure 4), that is,

$$\text{Eq. 2} \quad \text{TC} = [S(\text{pi})] / [S(\text{pi})_{\text{max}}]$$

This index ranges from values very close to 0 to 1, with values near 1 indicating maximum temporal coverage of the period covered by the surveys (Figure 1c).

Survey Evenness. We measured survey evenness (SE) in each cell pi using the Hill Diversity model, following³⁶.

$$\text{Eq. 3} \quad \text{SE} = (\sum_{i=1}^S p_i^q)^{1/(1-q)}$$

This index is less sensitive to variations in sample size than classical evenness indices such as Shannon or Simpson^{36,83}. Here we used the parameter $q = 4$ to assign greater sensitivity to equity for the Hill diversity index. To scale the index between 0

and 1, we divided the Hill index values by the number of species within each grid cell. Values of $SE = 0$ indicate a maximum bias of occurrence abundance for some species, whereas $SE = 1$ indicates no local bias in the distribution of the numbers of records per species. Thus, SE values allow us to detect possible imbalances in the observed species abundances, which can be interpreted as a measure of bias in community data. It is important to note that we did not extrapolate our interpretations of SE to infer whether the community is more uniform or not, but rather to evaluate the potential bias of species occurrences within our inventory (Figure 1c).

Temporal Evenness: To calculate temporal evenness, we used the same Hill diversity model mentioned above, with parameter $q = 4$. However, we considered the years as "species" and the number of samples per year as "abundance". This allowed us to measure the bias in the number of samples over the years, considering the entire known temporal range of each cell. Values of TE close to 0 indicate maximum temporal bias, with occurrence samples concentrated in very few years, whereas values of $TE = 1$ indicate no local bias in the distribution of samples per year, meaning that each year holds the same number of samples (Figure 1c).

Finally, we highlight a thorough analysis of Wallacean and Linnean noise is beyond the scope of this study. Here, regardless of the source of the records and the authors who gathered and/or identified the specimens, we are considering that the taxonomic identifications, geographic localities, and the dates are correct after our filtering processes (see filters in methods). Thus, the quality and coverage of the distribution data may not be free from errors, and we recommend that they be handled and used with caution. Additionally, we recommend reading Ronquillo et al.⁶⁷ and Castro-Souza et al.⁸⁵ for a more systematic view of these issues.

Assessing the sensitivity of data quality indices to low numbers of records

Small sample sizes may result in overestimating taxonomic completeness, and can cause noise in species accumulation curves by generating artificial completeness values^{7,8,16}, and result in values of temporal coverage and survey and temporal evenness that are not representative due to the low proportions of records. Therefore, it is necessary to establish a minimum threshold of number of records per grid cell to perform all index calculations. As the thresholds used in data quality analysis can be

subjective, potentially introducing noise into the ignorance map¹⁶, we conducted tests with different thresholds to evaluate how the cutoff settings of the minimum number of occurrences affect the results of the indices. For calculating TaxC we used no threshold, 5, 10, 15, 25, 35, 45, 55, 65, and 75 total occurrences as threshold values. In SC analyses we used 10, 15, 25, 35, 45, 55, 65, and 75 species occurrences as threshold. Lower values were not tested due to the long processing time and the fact that very small values can lead to misleading estimates^{11,16}. For TC, SE, and TE analyses we also used no threshold, 5, 10, 15, 25, 35, 45, 55, 65, and 75 species occurrences. Additionally, for SE and TE, we used $q = 4$ as we were interested in placing greater weight on evenness while also considering species or year richness in our model. Finally, with the maps constructed using different thresholds for each index, we calculated density curves. This allowed us to observe how the index behaved with respect to the different threshold configurations. Based on these results (density curves shown in Figure 4, maps not shown), we selected a common threshold of 45 occurrences in for all index calculations; grid cells with less than these occurrences were thus discarded as having too poor information. All analyses mentioned above were performed in R⁸⁶.

Building a composite ignorance map

Different methods have investigated various aspects of the ignorance about the distribution of biodiversity, as conceptualized by Rocchini et al.¹⁴ and Ladle and Hortal¹⁵. These authors feed from the idea of creating of global ignorance atlas that allow identifying lesser-known areas for a specific subject of investigation (e.g., political, social, or biological) initially proposed by Boggs¹³. While Ruete⁸⁷ proposed to assess the spatial availability of primary biodiversity records through ignorance scores, Stropp et al.¹¹ focused on mapping the spatial and temporal biases of information, and Meyer et al.¹² provided a framework for analyzing gaps, uncertainties, and biases in taxonomic, spatial, and temporal information. More recently, Tessarolo et al.¹⁶ incorporated explicit measures of coverage to these dimensions, to provide measures of the data-driven uncertainty in the projections of species distribution models. Here we propose a different and easy-to-use simplified framework for the rapid quantification of the overall taxonomic, spatial and temporal

ignorance for any biological group, using the order Orthoptera as an example that may be applied to other groups.

Our composite map depicts areas of biodiversity ignorance as the result of the ensemble of the five indices described above (i.e. TaxC + SC + TC + SE + TE) (Supplementary Figures 5–9), rescaled to range from 0 to 1, and finally subtracted from 1 (Figure 1d, e). Here, values close to 1 indicate maximum ignorance (i.e. extremely poor data quality), and values close 0 indicate minimum ignorance (i.e. good data quality). By mapping these values, the ignorance map will allow us to identify issues of data quality, coverage and completeness, as it will provide a quantitative indication of the composite deficiencies in taxonomic, survey and temporal completeness, plus biases in the species abundance and years sampled (Figure 2).

DATA AVAILABILITY

The datasets used during the study are available from the corresponding author on reasonable request. The most updated versions of the original data can also be downloaded from the URLs available in Table 1.

CODE AVAILABILITY

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

CONTRIBUTIONS

R.A.C.S., G.T., N.S. and T.S.S. conceived the study with input from all authors; investigation, all authors; data curation, R.A.C.S., N.S. and T.S.S.; formal analysis, R.A.C.S., G.T. and T.S.S.; visualization, R.A.C.S. and J.H.; funding acquisition, J.A.D.F. and T.S.S.; writing - original draft, R.A.C.S., G.T., N.S. and T.S.S.; writing - review & editing, J.S., J.A.D.F., R.L., J.H.; All authors have read, discussed the results and approved the final manuscript.

ACKNOWLEDGMENTS

We are grateful to the team at the Laboratory of Macroecology and Biodiversity Conservation (MacrEco) of the Federal University of Mato Grosso (UFMT) for the valuable discussions, as well as to the National Institute of Science & Technology (INCT) in Ecology, Evolution, and Biodiversity Conservation (EECBio), based at the Federal University of Goiás (UFG), for the precious discussions that contributed to the development of this study. RACS extends its gratitude primarily to the Coordination for the Improvement of Higher Education Personnel (CAPES) and to the Brazilian Institute of Development and Sustainability (IABS), in partnership with the National Center for Cave Research and Conservation (CECAV), for the doctoral and sandwich scholarships granted, respectively. JH was supported by project NICED, grant PID2022-140985NB-C21 funded by MCIN/AEI/ 10.13039/501100011033 / FEDER, EU. TSS expresses thanks to FAPEMAT (project FAPEMAT-PRO.000274/2023). We also wish to thank the developers of the 23 platforms that provided digitally-accessible data (see Methods).

COMPETING INTERESTS

J.H. is Editor-in-Chief and J.A.D.F. is Associate Editor of *npj Biodiversity*. All other authors declare having no competing interests as defined by Nature Portfolio, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

REFERENCES

1. Singh, J. S. The biodiversity crisis: A multifaceted review. *Curr. Sci.* **82**, 638–647 (2002).
2. Dirzo, R. *et al.* Defaunation in the Anthropocene. *Science* **345**, 401–406 (2014).
3. Hortal, J. *et al.* Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 523–549 (2015).
4. Venter, O. *et al.* Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* **7**, 12558 (2016).
5. Cardoso, P., Erwin, T. L., Borges, P. A. V. & New, T. R. The seven impediments in invertebrate conservation and how to overcome them. *Biol. Conserv.* **144**, 2647–2655 (2011).

6. Hortal, J., Ladle, R. J., Stropp, J. & Tassarolo, G. Accounting for biogeographical ignorance within biodiversity modelling. *Res. Outreach* **129**, (2022).
7. Soberón, J. & Peterson, A. T. Biodiversity informatics: Managing and applying primary biodiversity data. *Philos. Trans. R. Soc. B: Biol. Sci.* **359**, 689–698 (2004).
8. 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. *Divers. Distrib.* **20**, 369–381 (2014).
9. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* **6**, 8221 (2015).
10. Hortal, J., Lobo, J. M. & Jiménez-Valverde, A. Limitations of biodiversity databases: Case study on seed-plant diversity in Tenerife, Canary Islands. *Conserv. Biol.* **21**, 853–863 (2007).
11. Stropp, J. *et al.* Mapping ignorance: 300 years of collecting flowering plants in Africa. *Glob. Ecol. Biogeogr.* **25**, 1085–1096 (2016).
12. Meyer, C., Weigelt, P. & Kreft, H. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* **19**, 992–1006 (2016).
13. Boggs, S. W. An atlas of ignorance: a needed stimulus to honest thinking and hard work. *Proc. Am. Philos. Soc.* **93**, 253–258 (1949).
14. Rocchini, D. *et al.* Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Prog. Phys. Geogr.* **35**, 211–226 (2011).
15. Ladle, R. J. & Hortal, J. Mapping species distributions: living with uncertainty. *Front. Biogeogr.* **5**, 8–9 (2013).
16. Tassarolo, G., Ladle, R. J., Lobo, J. M., Rangel, T. F. & Hortal, J. Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. *Ecography* **44**, 1743–1755 (2021).
17. Lessa, T., Stropp, J., Hortal, J. & Ladle, R. J. How taxonomic change influences forecasts of the Linnean shortfall (and what we can do about it)? *J Biogeogr* **00**, 1–9 (2024).
18. Bebbler, D. P. *et al.* Herbaria are a major frontier for species discovery. *Proc. Natl. Acad. Sci. USA* **107**, 22169–22171 (2010).
19. Freeman, B. G. & Pennell, M. W. The latitudinal taxonomy gradient. *Trends Ecol. Evol.* **36**, 778–786 (2021).

20. Stropp, J., Ladle, R. J., Emilio, T., Lessa, T. & Hortal, J. Taxonomic uncertainty and the challenge of estimating global species richness. *J. Biogeogr.* **49**, 1654–1656 (2022).
21. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132 (2017).
22. de Siracusa, P. C., Gadelha, L. M. R. & Ziviani, A. New perspectives on analysing data from biological collections based on social network analytics. *Sci. Rep.* **10**, 3358 (2020).
23. Molles, M. C. & Sher, A. Species Abundance and Diversity. in *Ecology: Concepts and Applications* 343–361 (McGraw-Hill Education, 2019).
24. Lobo, J. M. *et al.* KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecol. Indic.* **91**, 241–248 (2018).
25. Ronquillo, C. *et al.* Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses. *Biodivers. Data J.* **8**, e53474 (2020).
26. Hughes, A. C. *et al.* Sampling biases shape our view of the natural world. *Ecography* **44**, 1259–1269 (2021).
27. Sobral-Souza, T. *et al.* Knowledge gaps hamper understanding the relationship between fragmentation and biodiversity loss: The case of Atlantic Forest fruit-feeding butterflies. *PeerJ* e11673 (2021).
28. Zizka, A., Antonelli, A. & Silvestro, D. sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography* **44**, 25–32 (2021).
29. Hortal, J. & Lobo, J. M. An ED-based protocol for optimal sampling of biodiversity. *Biodivers. Conserv.* **14**, 2913–2947 (2005).
30. Moreno, C. E. & Halffter, G. Assessing the completeness of bat biodiversity inventories using species accumulation curves. *J. Appl. Ecol.* **37**, 149–158 (2000).
31. Soberón, J., Jiménez, R., Golubov, J. & Koleff, P. Assessing completeness of biodiversity databases at different spatial scales. *Ecography* **30**, 152–160 (2007).
32. Escribano, N., Ariño, A. H. & Galicia, D. Biodiversity data obsolescence and land uses changes. *PeerJ* **4**, e2743 (2016).

33. Tessarolo, G., Ladle, R., Rangel, T. & Hortal, J. Temporal degradation of data limits biodiversity research. *Ecol. Evol.* **7**, 6863–6870 (2017).
34. Stropp, J. *et al.* The ghosts of forests past and future: deforestation and botanical sampling in the Brazilian Amazon. *Ecography* **43**, 979–989 (2020).
35. Boakes, E. H. *et al.* Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol.* **8**, e1000385 (2010).
36. Roswell, M., Dushoff, J. & Winfree, R. A conceptual guide to measuring species diversity. *Oikos* **130**, 321–338 (2021).
37. Bánki, O. *et al.* Catalogue of Life. *Catalogue of Life Checklist (Version 2023-05-15)* (2023).
38. Branson, D. H., Joern, A. & Sword, G. A. Sustainable management of insect herbivores in grassland ecosystems: New perspectives in grasshopper control. *BioScience* vol. 56 743–755 (2006).
39. Lavoie, K. H., Helf, K. L. & Poulson, T. L. The biology and ecology of North American cave crickets. *J. Cave and Karst Studies* **6**, 114–134 (2007).
40. Santana, F. D., Baccaro, F. B. & Costa, F. R. C. Busy nights: High seed dispersal by crickets in a neotropical forest. *Am. Naturalist* **188**, 126–133 (2016).
41. Tan, M. K. *et al.* Overlooked flower-visiting Orthoptera in Southeast Asia. *J. Orthoptera Res.* **26**, 143–153 (2017).
42. Song, H. Grasshopper systematics: past, present and future. *J. Orthoptera Res.* **19**, 57–68 (2010).
43. Bidau, C. J. Patterns in Orthoptera biodiversity. I. Adaptations in ecological and evolutionary contexts. *J. Insect Biodivers.* **2**, 1–39 (2014).
44. Bidau, C. J. Patterns in Orthoptera biodiversity. II. The cultural dimension. *J. Insect Biodivers.* **2**, 1–15 (2014).
45. Song, H. Biodiversity of Orthoptera. in *Insect Biodiversity: Science and Society* (eds. Foottit, R. G. & Adler, P. H.) vol. 2 245–279 (John Wiley & Sons Ltd., Chichester, 2018).
46. Green, S. V. The taxonomic impediment in orthopteran research and conservation. *J. Insect Conserv.* **2**, 151–159 (1998).
47. Cigliano, M. M. & Eades, D. New technologies challenge the future of taxonomy in Orthoptera. *J. Orthoptera Res.* **19**, 15–18 (2010).

48. Cigliano, M. M., Braun, H., Eades, D. C. & Otte, D. Orthoptera Species File. (*Version 5.0/5.0*) <http://Orthoptera.SpeciesFile.org> (2022).
49. The Global Biodiversity Information Facility. *What is GBIF?* <https://www.gbif.org/what-is-gbif> (2022).
50. iNaturalist. <https://www.inaturalist.org> (2022).
51. Samways, M. J. & Lockwood, J. A. Orthoptera conservation: pests and paradoxes. *J. Insect Conserv.* **3**, 143–149 (1998).
52. Nouh, G. M. & Adly, D. Evaluation of the virulence of entomopathogenic nematodes as a biological control agents against *Gryllotalpa gryllotalpa* (Gryllotalpidae). *J. Appl. Entomol.* **145**, 1050–1056 (2021).
53. Diniz-Filho, J. A. F. *et al.* Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Front. Biogeogr.* **15**, e59566 (2023).
54. Diniz-Filho, J. A. F., de Marco, P. & Hawkins, B. A. Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect Conserv. Divers.* **3**, 172–179 (2010).
55. Gaston, K. J. & May, R. M. Taxonomy of taxonomists. *Nature* **356**, 281–282 (1992).
56. Engel, M. S. *et al.* The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. *Zool. J. Linn. Soc.* **193**, 381–387 (2021).
57. Braby, M. F., Hsu, Y.-F. & Lamas, G. How to describe a new species in zoology and avoid mistakes. *Zool. J. Linn. Soc.* 1–16 (2024) doi:10.1093/zoolinnean/zlae043.
58. Randell, R. L. On the Presence of Concealed Genitalic Structures in Female Caelifera (Insecta; Orthoptera). *Trans. Am. Entomol. Soc.* **88**, 247–260 (1962).
59. Alexander, R. D. & Otte, D. *The Evolution of Genitalia and Mating Behavior in Crickets (Gryllidae) and Other Orthoptera*. (Ann Arbor, University of Michigan, Miscellaneous Publications, No. 133, 1967).
60. Balakrishnan, R. Species concepts, Species Boundaries and Species Identification: A View from the Tropics. *Syst. Biol.* **54**, 689–693 (2005).
61. Rodrigues, A. S. L. *et al.* A global assessment of amphibian taxonomic effort and expertise. *Bioscience* **60**, 798–806 (2010).
62. Vale, M. M. & Jenkins, C. N. Across-taxa incongruence in patterns of collecting bias. *J. Biogeogr.* **39**, 1744–1748 (2012).

63. Amano, T. & Sutherland, W. J. Four barriers to the global understanding of biodiversity conservation: Wealth, language, geographical location and security. *Proc. R. Soc. B: Biol. Sci.* **280**, (2013).
64. Lewinsohn, T. M., Agostini, K., Lucci Freitas, A. V. & Melo, A. S. Insect decline in Brazil: An appraisal of current evidence. *Biol. Lett.* **18**, 20220219 (2022).
65. Bakker, F. T. *et al.* The global museum: Natural history collections and the future of evolutionary science and public education. *PeerJ* **8**, e8225 (2020).
66. Raja, N. B. *et al.* Colonial history and global economics distort our understanding of deep-time biodiversity. *Nat. Ecol. Evol.* **6**, 145–154 (2022).
67. Ronquillo, C., Stropp, J., Medina, N. G. & Hortal, J. Exploring the impact of data curation criteria on the observed geographical distribution of mosses. *Ecol. Evol.* **13**, e10786 (2023).
68. Krishtalka, L. & Humphrey, P. S. Can Natural History Museums Capture the Future? *Bioscience* **50**, 611–617 (2000).
69. Nelson, G. & Ellis, S. The history and impact of digitization and digital data mobilization on biodiversity research. *Philos. Trans. R. Soc. B: Biol. Sci. Sciences*: **374**, 20170391 (2018).
70. Jarić, I. *et al.* iEcology: Harnessing Large Online Resources to Generate Ecological Insights. *Trends Ecol. Evol.* **35** 630–639 (2020).
71. Johnson, K. R. & Owens, I. F. P. A global approach for natural history museum collections. *Science* **379**, 1192–1194 (2023).
72. Phillips, T. Sao Paulo fire destroys one of the largest collections of dead snakes. *The Guardian*, <https://www.theguardian.com/world/2010/may/16/firedestroys-snake-collection> (2010).
73. Phillips, T. Brazil National Museum: as much as 90% of collection destroyed in fire. *The Guardian*, <https://www.theguardian.com/world/2018/sep/04/brazilnational-museum-fire-collection-destroyed-not-insured> (2018).
74. Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS One* **7**, e29715 (2012).
75. Ferro, M. L. & Flick, A. J. ‘Collection Bias’ and the importance of natural history collections in species habitat modeling: A case study using *thoracophorus costalis erichson* (Coleoptera: Staphylinidae: Osoriinae), with a critique of GBIF.org. *Coleopterists Bull.* **69**, 415–425 (2015).
76. Kemp, C. The endangered dead. *Nature* **518**, 292–294 (2015).

77. Turnhout, E., Lawrence, A. & Turnhout, S. Citizen science networks in natural history and the collective validation of biodiversity data. *Conserv. Biol.* **30**, 532–539 (2016).
78. Bartoccioni, F. Big data in biogeography: From museum collection to citizen science. *Biogeographia* **32**, 1–3 (2017).
79. Chozas, S. *et al.* Rescuing Botany: using citizen-science and mobile apps in the classroom and beyond. *npj Biodiversity* **2**, (2023).
80. Fontaine, C., Fontaine, B. & Prévot, A. C. Do amateurs and citizen science fill the gaps left by scientists? *Curr. Opin. Insect. Sci.* **46** 83–87 (2021).
81. Bowler, D. E. *et al.* Decision-making of citizen scientists when recording species observations. *Sci. Rep.* **12**, 11069 (2022).
82. Brown, J. H. Macroecology: Progress and Prospect. *Oikos* **87**, 3–14 (1999).
83. Silva, F. R., Gonsalves-Souza, T., Paterno, G. B., Provete, D. B. & Vancine, M. H. *Análises Ecológicas No R*. (Nupeea, Recife, 2022).
84. Oksanen, J. *et al.* vegan: Community Ecology Package. *R package version 2.6-2* (2022).
85. Castro-Souza, R. A. *et al.* O (Des)conhecimento da Biodiversidade: uma Sistematização sobre Lacunas, Limitações, Vieses, Déficits e Ruídos. *Oecologia Australis* **28**, 159–177 (2024).
86. R Core Team. R: A language and environment for statistical computing. Preprint at <https://www.R-project.org/> (2022).
87. Ruete, A. Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. *Biodivers. Data J.* **3**, e5361 (2015).

TABLE

Table 1. Digital repositories accessed to build the global dataset of Orthoptera.

Source	Acronym	URL Link	Geographic Scope	Raw N° records	Extraction Date
Global Biodiversity Information Facility	GBIF	https://www.gbif.org/pt/	Global	3122773	03/02/2022
iNaturalist	-	https://www.inaturalist.org/	Global	857970	13/02/2022
Integrated Digitized Biocollections	IDigBio	https://www.idigbio.org/	IdgBio Sites	351592	22/04/2022
Orthoptera Species File	OSF	http://orthoptera.speciesfile.org	Global	202800	28/10/2022
National Biodiversity Network Atlas	NBN atlas	https://nbnatlas.org/	United Kingdom	190390	13/02/2022
Sistema da Informação sobre a Biodiversidade Brasileira	SiBBR	https://www.sibbr.gov.br/	Brazil	63228	13/02/2022
Atlas of Living Australia	ALA	https://www.ala.org.au/	Australia	50490	15/03/2022
Monarch	-	https://monarch.calacademy.org/	Global	15798	14/03/2022
Canadensys	-	https://www.canadensys.net/	North America	14009	15/03/2022
VertNet	-	http://vertnet.org/	Global	10999	13/02/2022
Portal da Biodiversidade - ICMbio	PortalBio	https://portaldabiodiversidade.icmbio.gov.br	Brazil	6723	26/05/2022
Ecdysis	-	https://serv.biokic.asu.edu/ecdysis/collections/index.php	Global	4185	15/03/2022
Natural History Museum Data Portal	NHM Data Portal	https://data.nhm.ac.uk/	Global	4015	04/04/2022
Smithsonian Tropical Research Institute	STRI	https://stri.si.edu/	Panama	3920	14/03/2022
speciesLink	-	https://specieslink.net/	North America	3616	13/02/2022
Ecoregistros	-	https://www.ecoregistros.org/site/index.php	Global	3203	14/03/2022
Grasshoppers, Crickets and Allied Insects (Orthoptera) of Ireland	-	https://maps.biodiversityireland.ie/Dataset/204	Ireland	3154	14/03/2022
Madrean Discovery Expedition Flora & Fauna Database	-	https://madreandiscovery.org/	Madrean Archipelago	1133	15/03/2022
Berkeley Ecoinformatics Engine	Ecoengine	https://ecoengine.berkeley.edu/	United States	1081	13/02/2022
NEON Biorepository	-	https://biorepo.neonscience.org/portal/	NEON Sites	224	14/03/2022
Ocean Biodiversity Information System	OBIS	https://obis.org/	Global	35	13/02/2022
Biodiversidad de Guatemala Data Portal	-	https://biodiversidad.gt/portal/index.php	Guatemala	15	15/03/2022

InvertEBase	-	https://invertebase.org/port al/	Global	6	15/03/2022
-------------	---	--	--------	---	------------

FIGURES

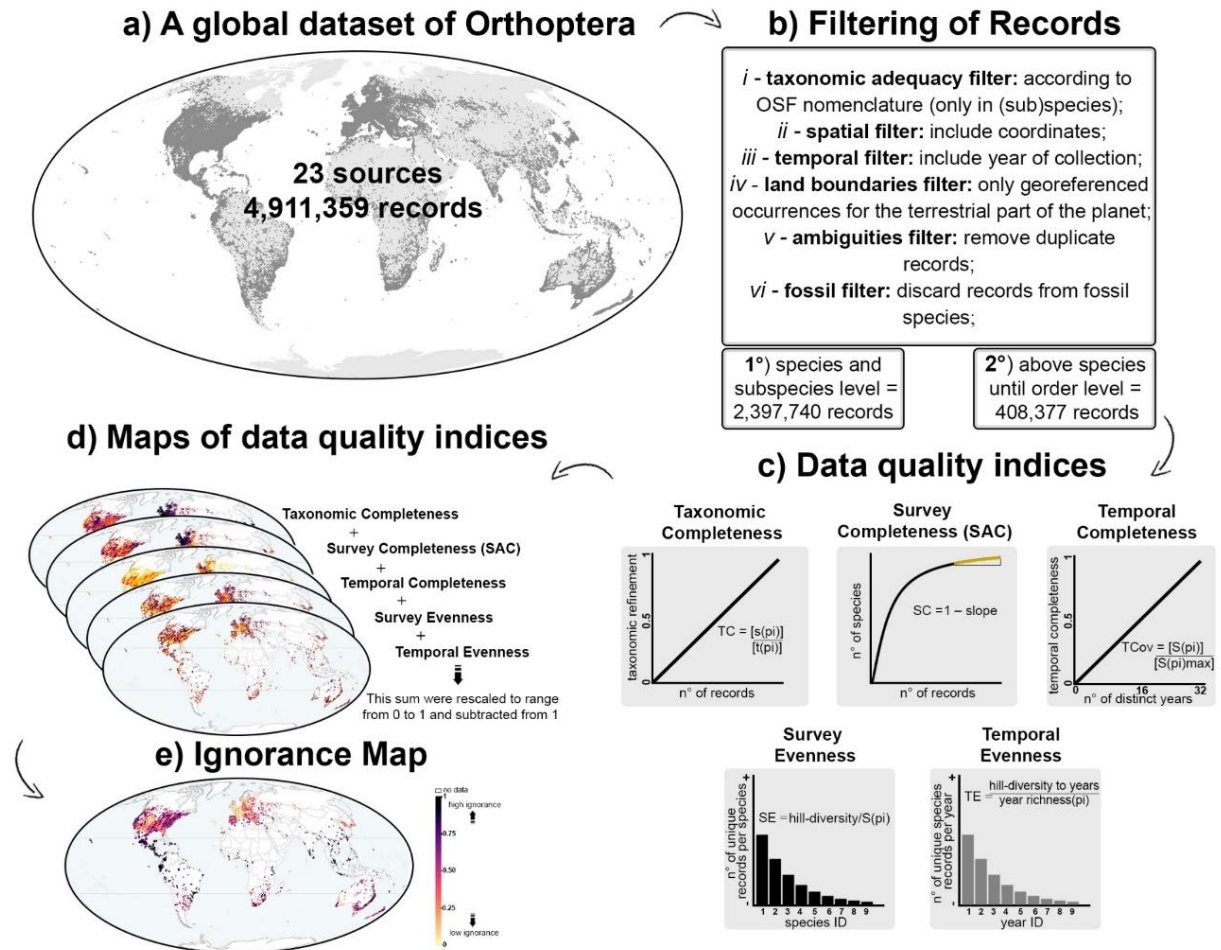


Figure 1. Outline of all methodological and analytical steps required to construct the ignorance map for the order Orthoptera (Arthropoda, Insecta) according to digitally-accessible data. a) we compile all digitally accessible occurrence records for Orthoptera (see Table 1). b) we applied several filters to the records, which were divided into two groups according to their taxonomic resolution: 1°) species and subspecies level, and 2°) above species up to order level. c) within each 100 km-width grid cell, we calculated five different data quality indices. d) the indices were subsequently mapped. e) five of the metrics were combined by summing them in the form of an ensemble, which was then rescaled to range from 0 to 1, and finally subtracted from 1, in order to create a single final ignorance map.

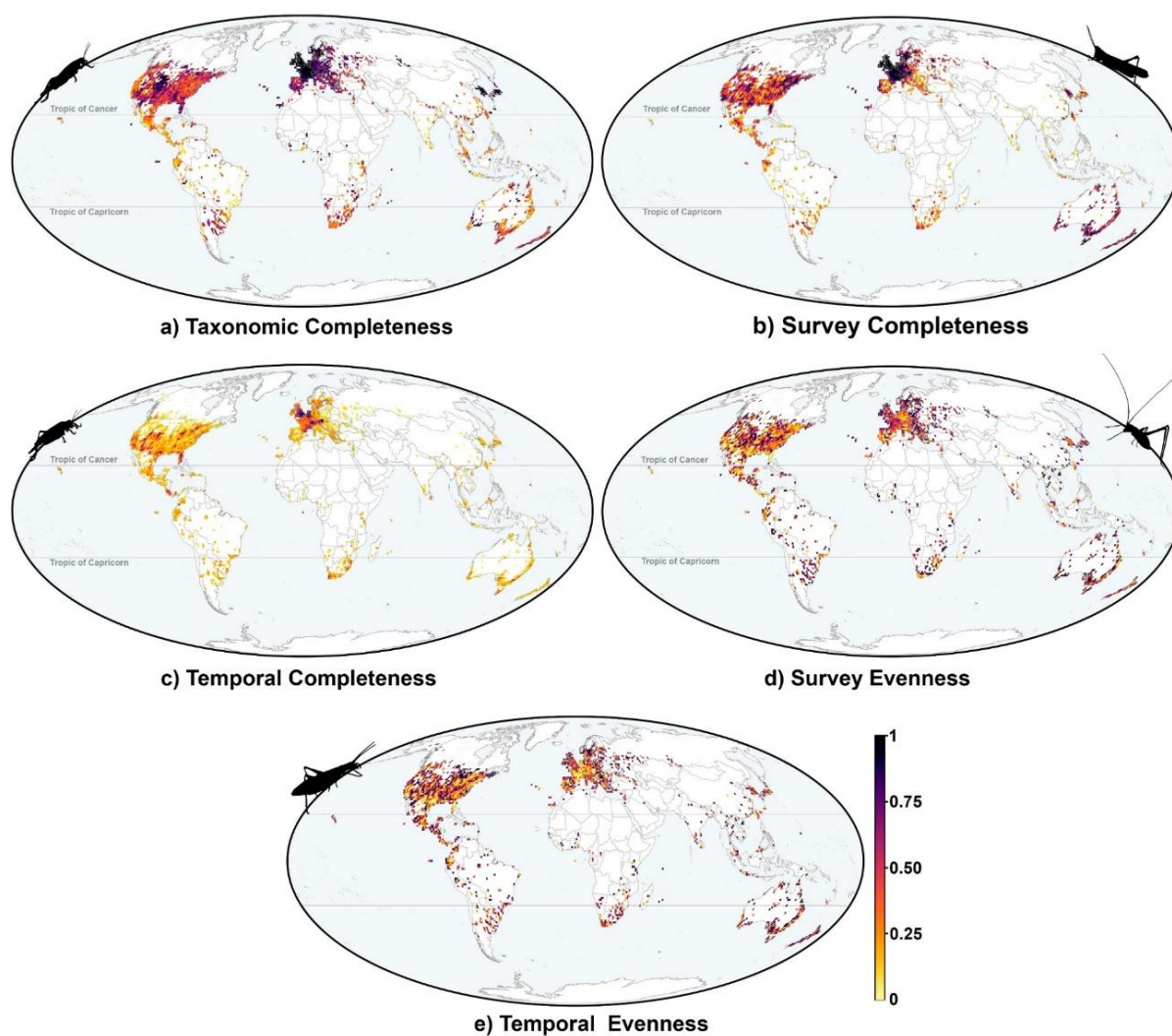


Figure 2. Maps of the five indices (a–e) used in this study to assess the quality of digitally-accessible data for the global distribution of the order Orthoptera (Arthropoda, Insecta). a) taxonomic completeness, values close to 1 indicate that most records are identified to the species taxonomic level. b) survey completeness, values close to 1 indicate an almost complete inventory. c) temporal completeness, values close to 1 indicate maximum temporal coverage. d) survey evenness, values close to 1 indicate no local bias in the distribution of the numbers of records per species. e) temporal evenness, values close to 1 indicate no local bias in the distribution of samples per year.

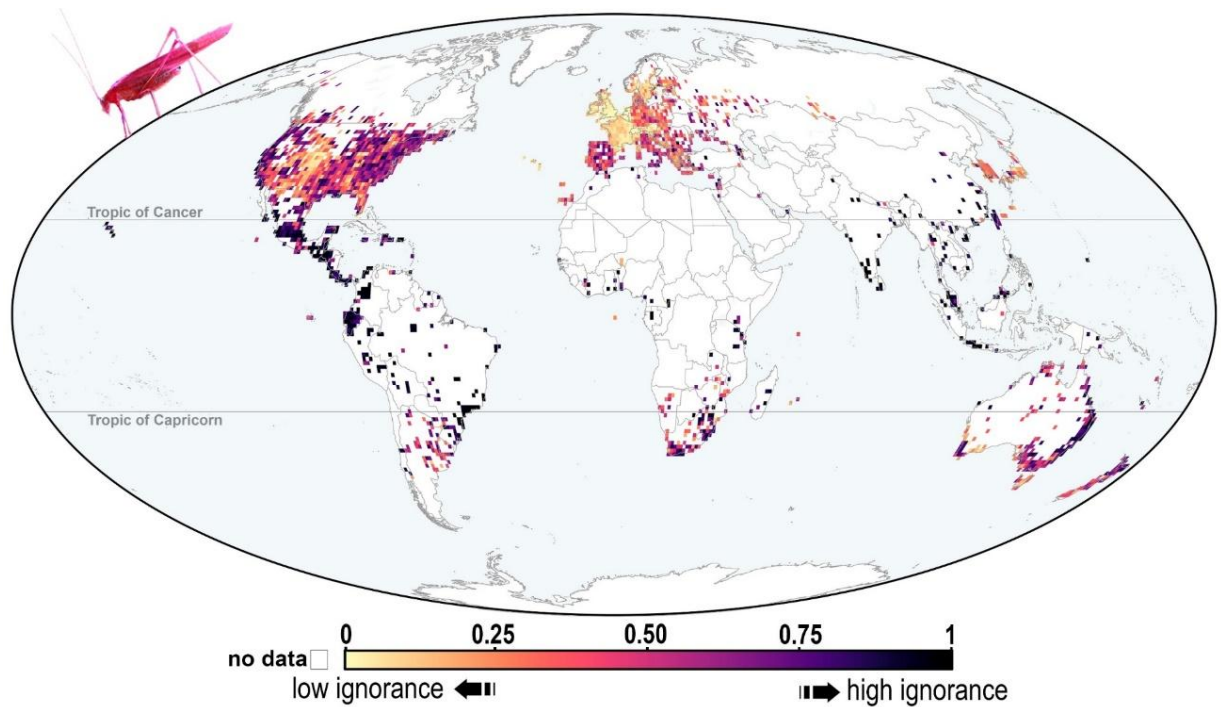


Figure 3. Map of ignorance on the global distribution of the order Orthoptera (Arthropoda, Insecta) based on digitally-accessible data. Lower ignorance values indicate higher levels of knowledge, pinpointing areas with good data quality that only require some refinements to provide high-quality inventories.

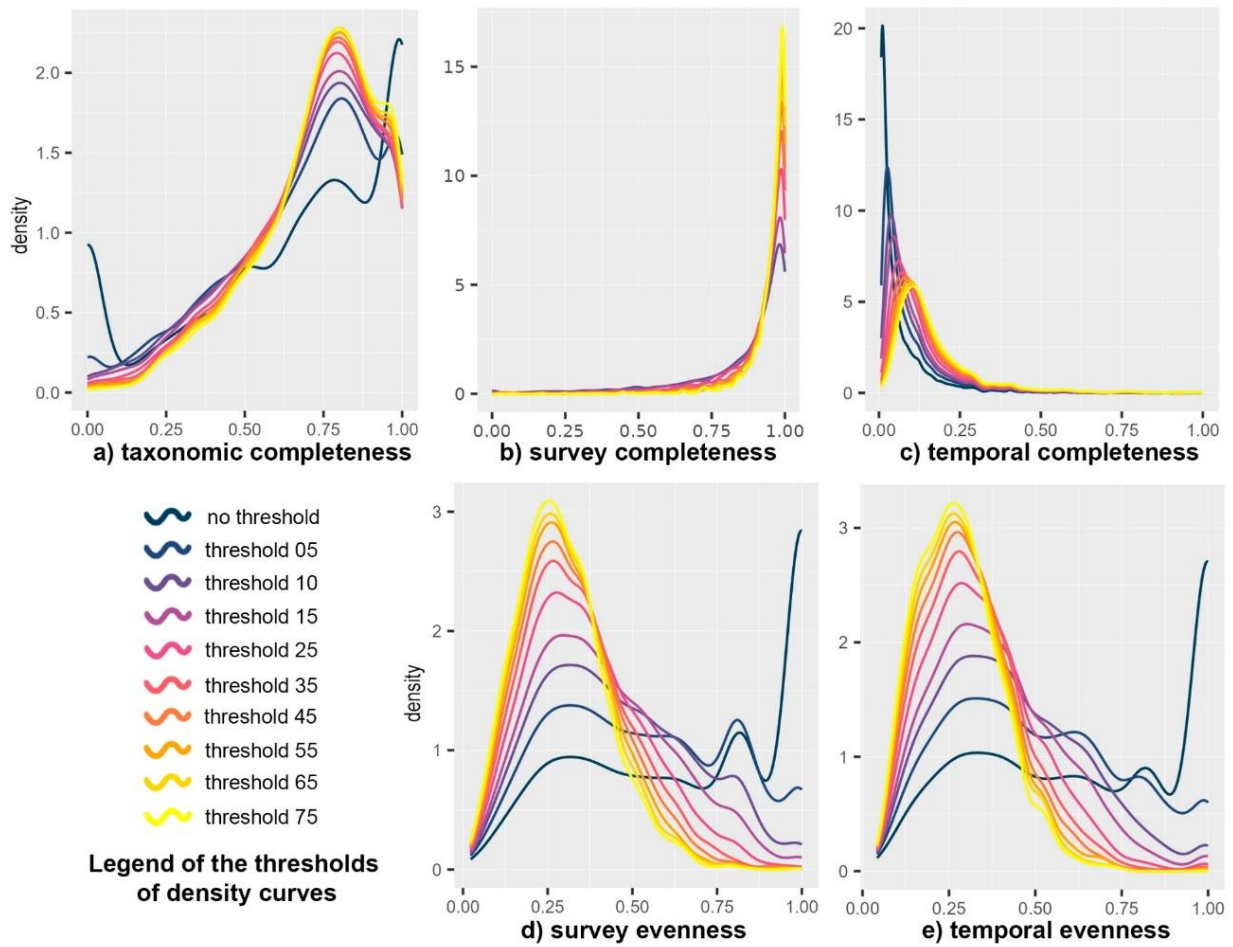


Figure 4. Density curves of the values of the data quality indices generated by different threshold configurations for taxonomic, survey and temporal completeness, and survey and temporal evenness. a) density curves for taxonomic completeness. b) density curves for survey completeness. c) density curves for temporal completeness. d) density curves for survey evenness. e) density curves for temporal evenness.

Supplementary Material

Supplementary Note 1

For most digitally accessible sources, such as: Global Biodiversity Information Facility (GBIF); Integrated Digitized Biocollections (IDigBio); National Biodiversity Network Atlas (NBN atlas); Sistema da Informação sobre a Biodiversidade Brasileira (SiBBr); Atlas of Living Australia (ALA); Monarch; Canadensys; VertNet; Portal da Biodiversidade – ICMbio (PortalBio); Ecdysis; Natural History Museum Data Portal (NHM Data Portal); Smithsonian Tropical Research Institute (STRI); speciesLink; Grasshoppers, Crickets and Allied Insects (Orthoptera) of Ireland; Madrean Discovery Expedition Flora & Fauna Database; Berkeley Ecoinformatics Engine (Ecoengine); NEON Biorepository; Ocean Biodiversity Information System (OBIS); Biodiversidad de Guatemala Data Portal; InvertEBase), searches were conducted for all available primary records for the order Orthoptera, in the format of Darwin Core spreadsheets (when available)¹, or in the format commonly used by the aforementioned databases (see Table 1 in the Manuscript).

For the citizen science platform iNaturalist, standardized searches for primary records were conducted at the "verifiable data" level for each Orthoptera family and, in some cases, for each category below the family level. A single search for the entire group was not performed due to the extraction limit on iNaturalist, which is 200,000 observations. In the case of Orthoptera, more than 900,000 records were found during the extraction period² (see Table 1 in the Manuscript). Therefore, we followed the family nomenclature proposed by the Catalogue Orthoptera Species File (OSF)³, available on the Catalogue of Life platform (<https://www.catalogueoflife.org/>), to guide each of the searches.

For the taxonomic catalog Orthoptera Species File (OSF), which uses the Species File Software (SFS) program for structuring and storing record information in a Microsoft SQL Server relational database management system⁴, we adopted a web scraping approach with the help of the RSelenium package in the R environment. This approach allowed us to collect the necessary data to create a matrix that combines taxonomic, geographic, and temporal information for cataloged and digitized orthopterans. We chose this approach because it was not feasible to extract

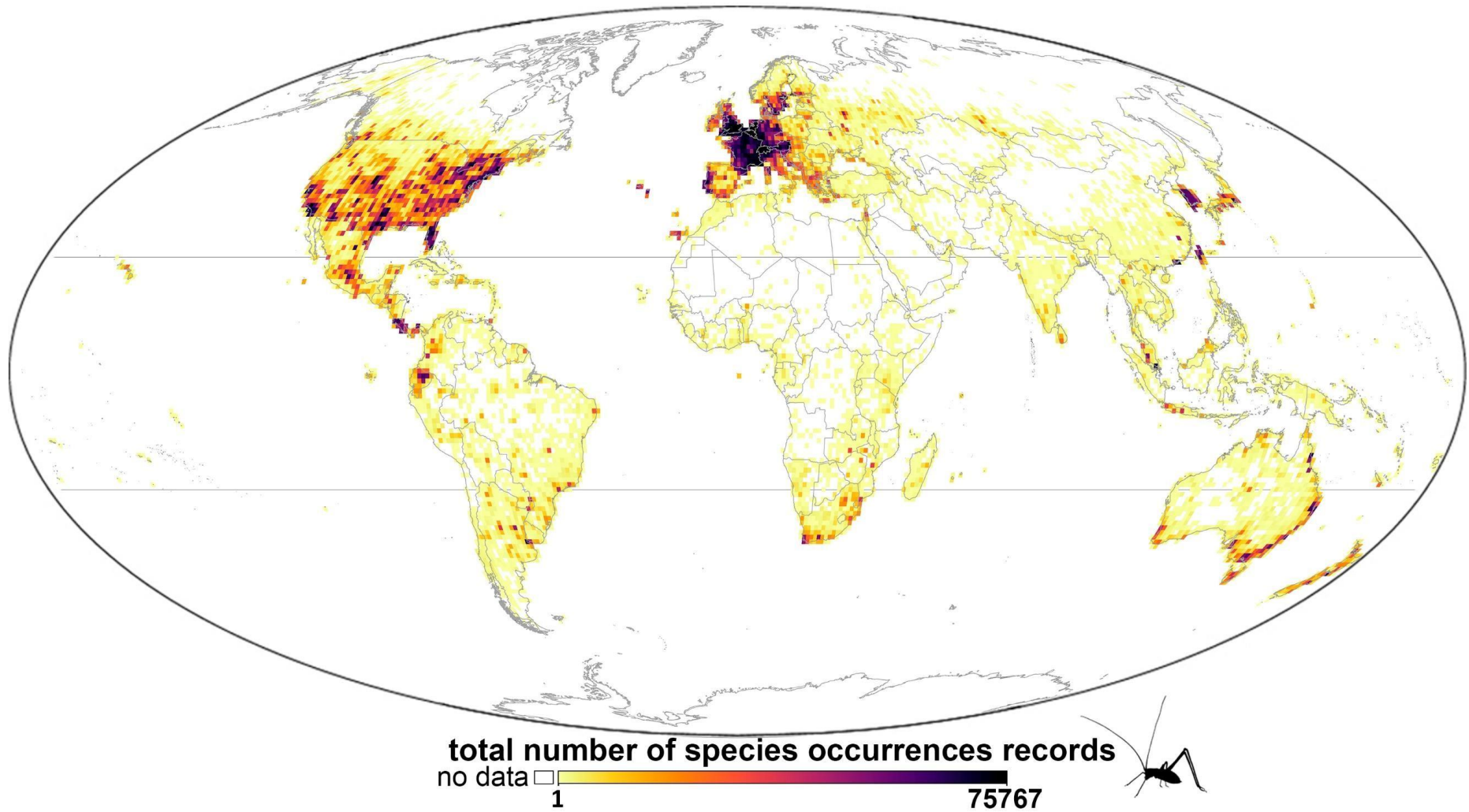
standardized lists with all biogeographical information using the tools available in OSF. However, currently, this information scanning is already almost possible, as the OSF has been migrating to a Darwin Core-based system, the TaxonWorks platform (see <https://orthoptera.speciesfile.org/>).

Additionally, for the citizen science platform EcoRegistros, we conducted manual searches for all primary Orthoptera records available up to the extraction date. This was due to the site's structural limitations, which do not allow for automated downloading of all records (see <https://www.ecoregistros.org/site/index.php>). Furthermore, the platform had a relatively small number of records, making manual extraction a viable option.

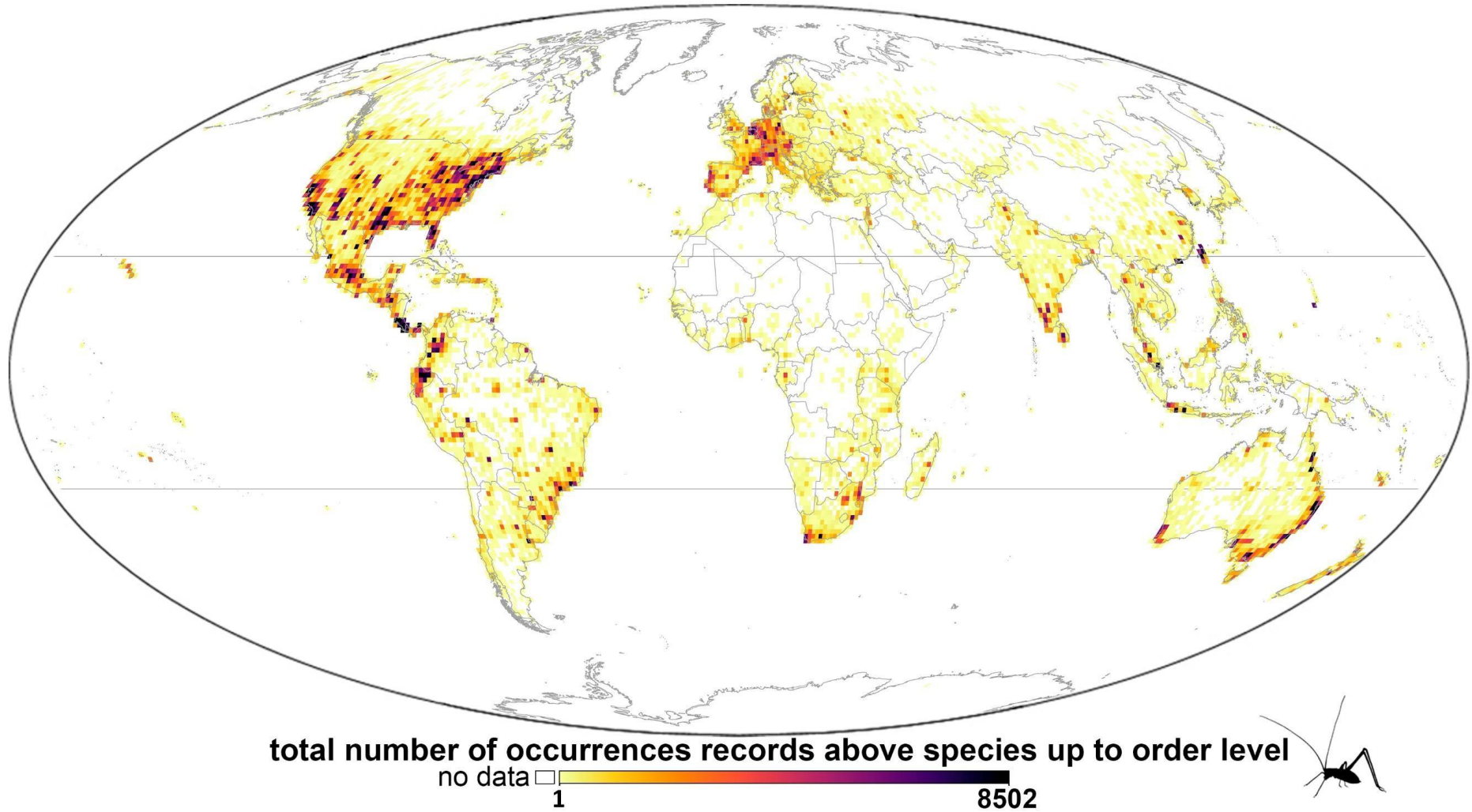
References

1. Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS One* **7**, e29715 (2012).
2. iNaturalist. <https://www.inaturalist.org> (2022).
3. Cigliano, M. M. *et al.* Catalogue of Life Checklist. <https://doi.org/10.48580/d4sw-388> (2022).
4. Song, H. Grasshopper systematics: past, present and future. *J. Orthoptera Res.* **19**, 57–68 (2010).

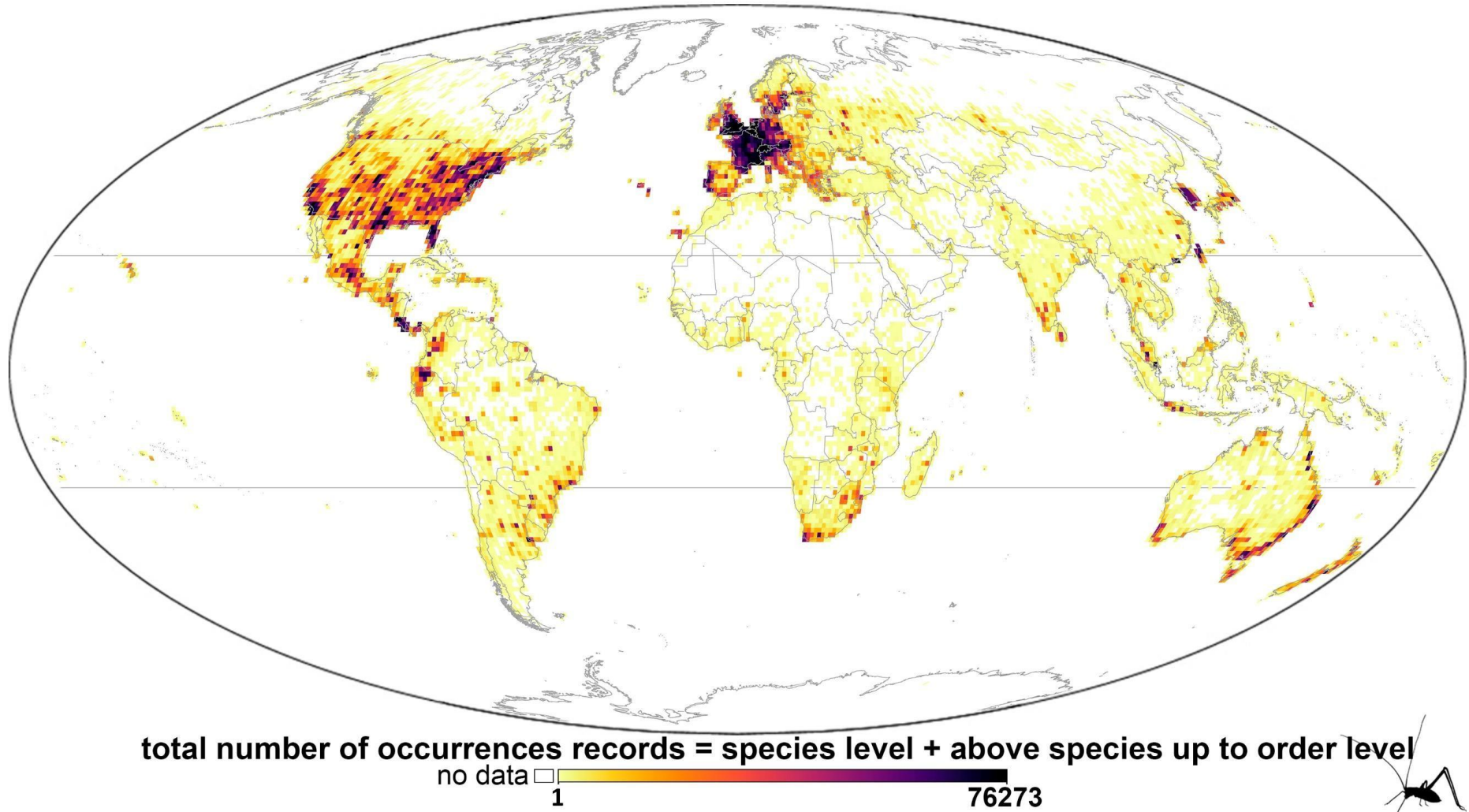
Supplementary Figures



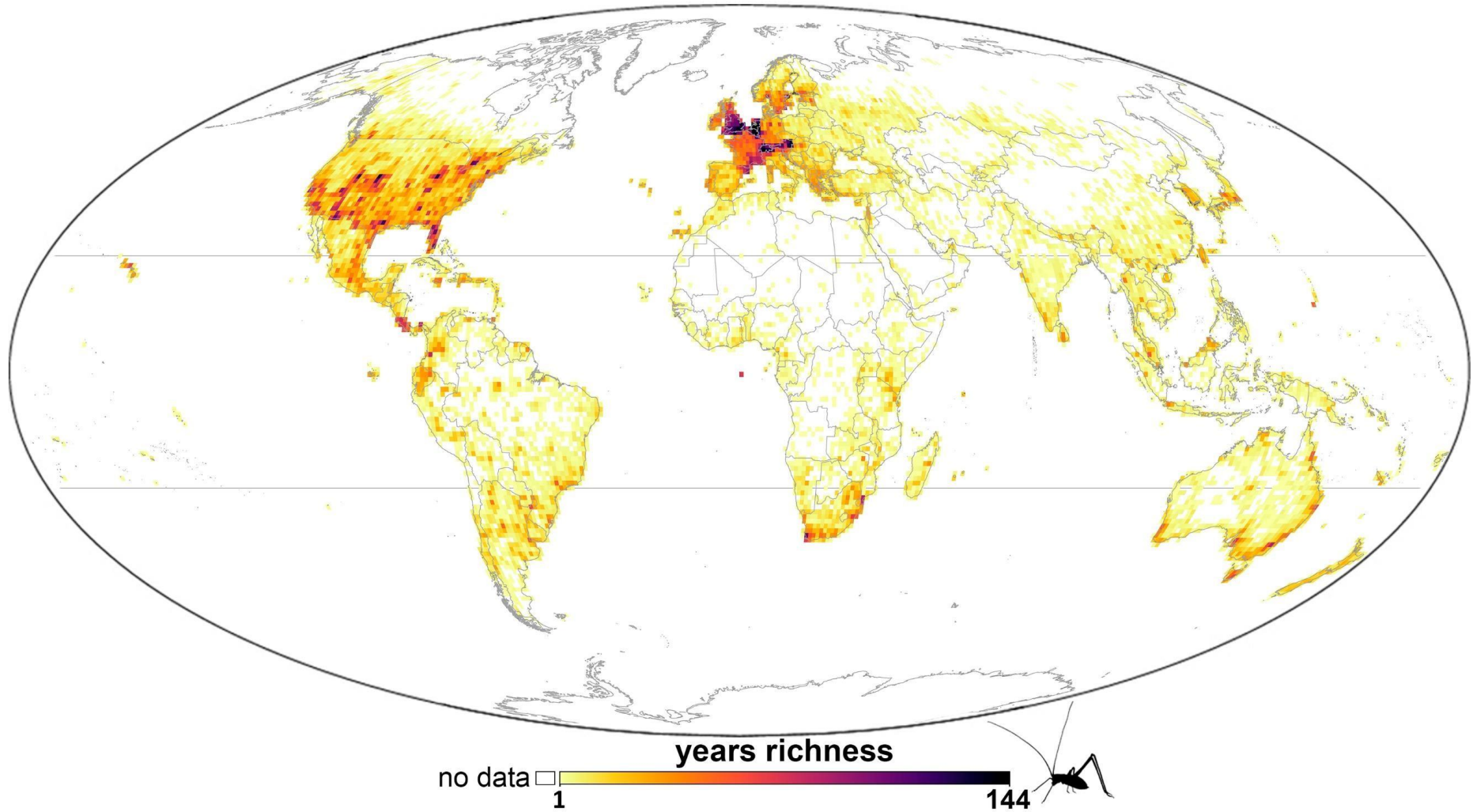
Supplementary Figure 1. Total number of species occurrences records after taxonomic, spatial, and temporal filtering processes (see methods).



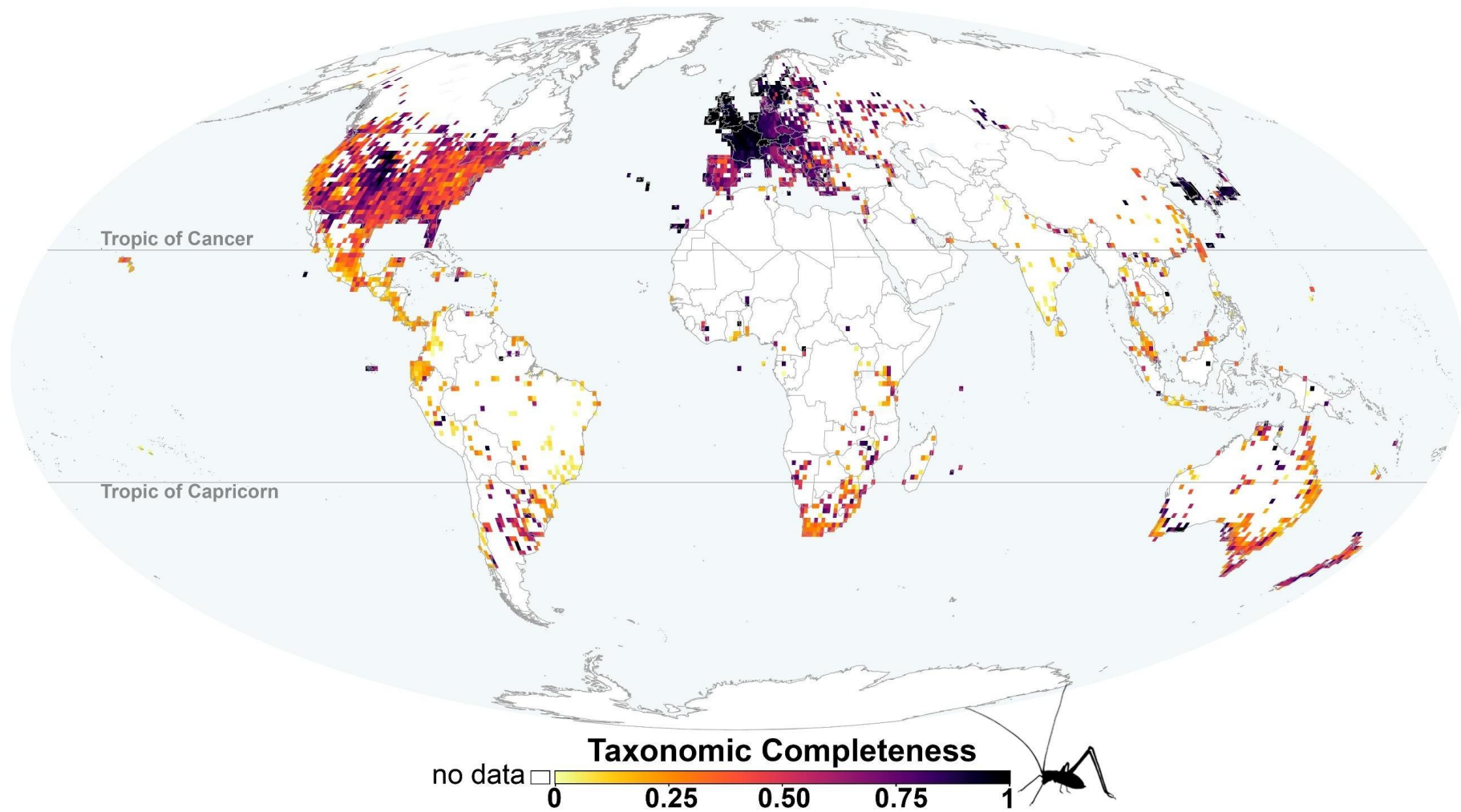
Supplementary Figure 2. Total number of occurrences records above species up to order level after spatial and temporal filtering processes (see methods).



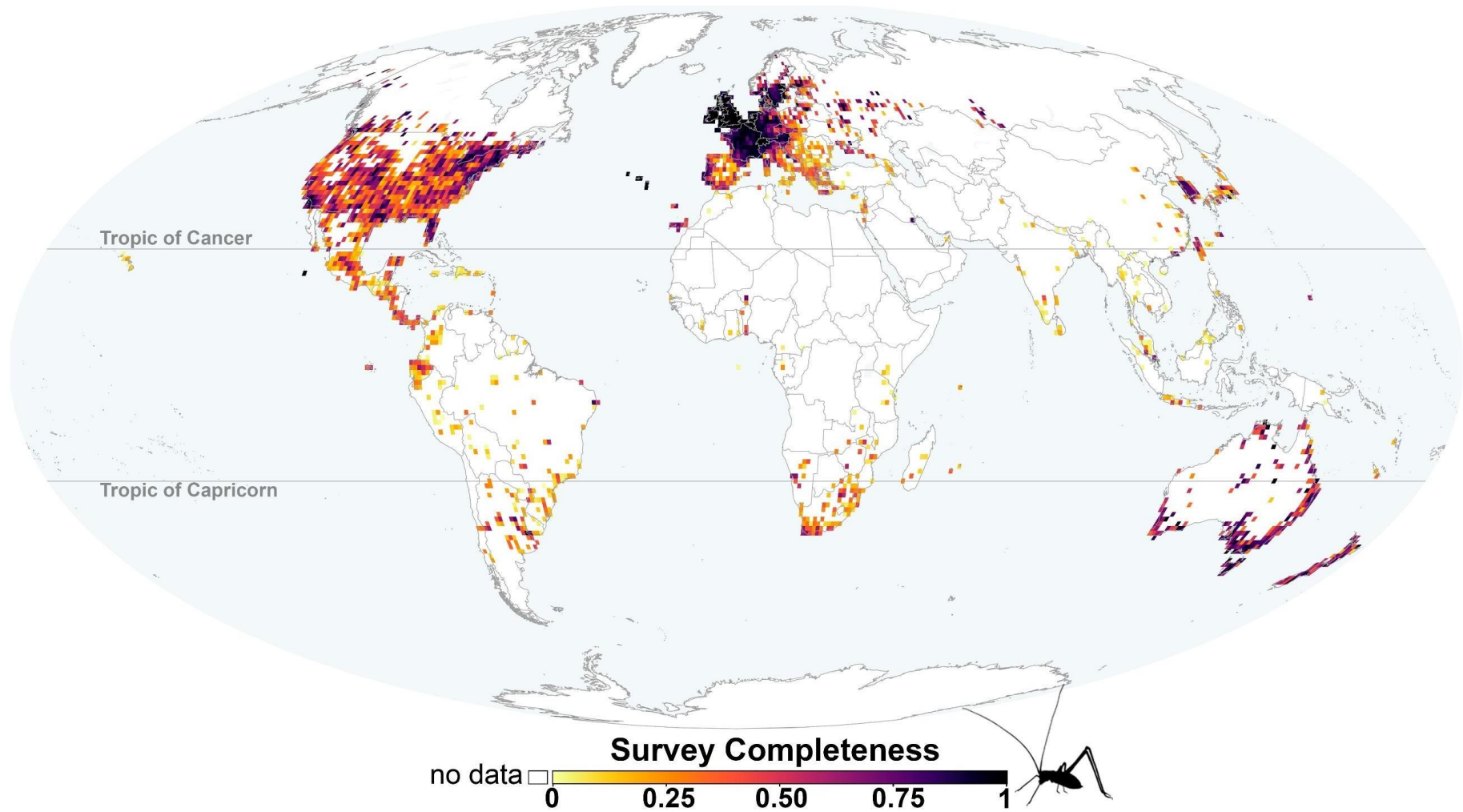
Supplementary Figure 3. Total number of occurrences records resulting from the sum of the total number of species occurrences records and the total number of occurrences records above the species level up to the order level per raster cell (see Figures 1 and 2 mentioned above).



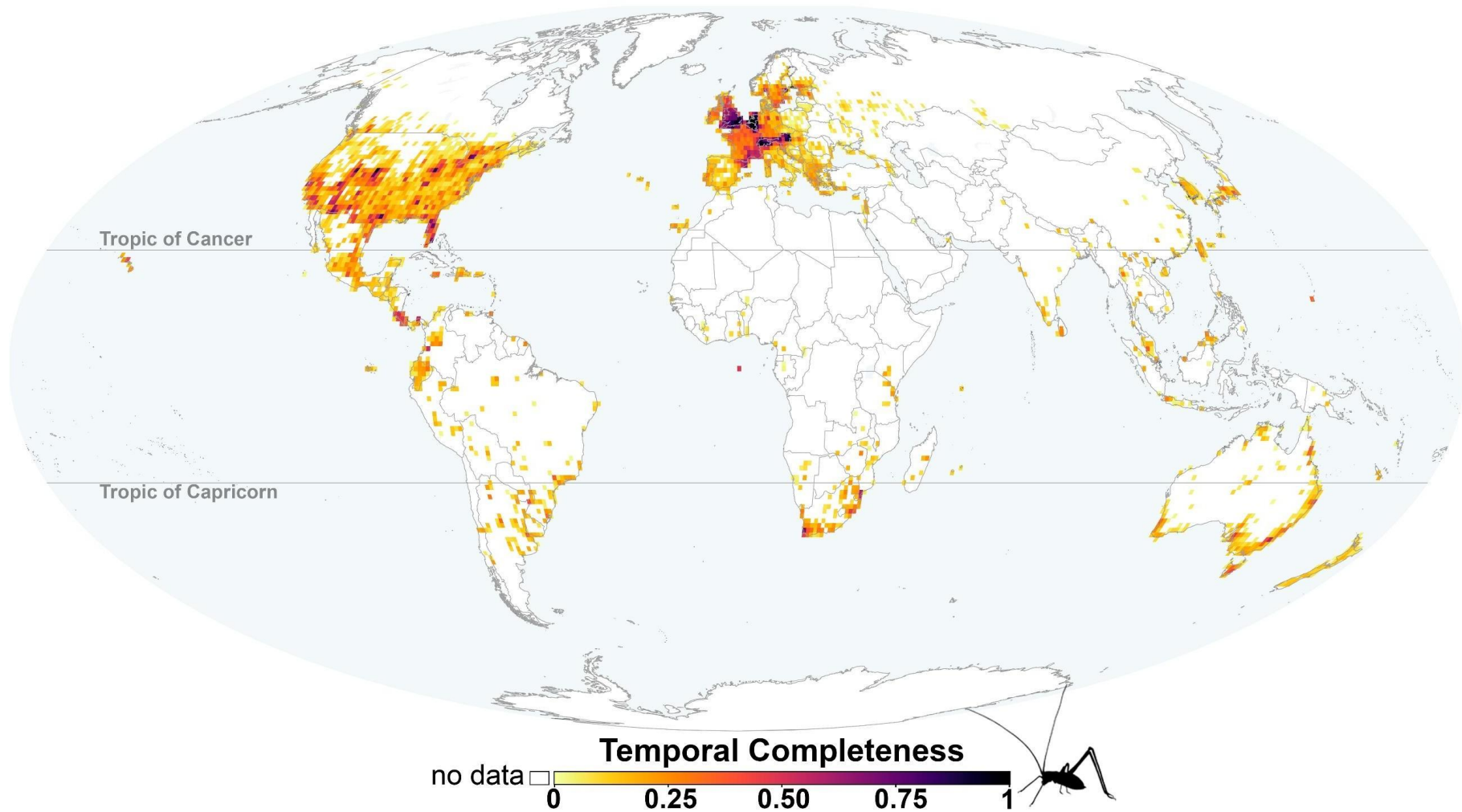
Supplementary Figure 4. Year richness calculated from digitally accessible data for the order Orthoptera for each raster cell.



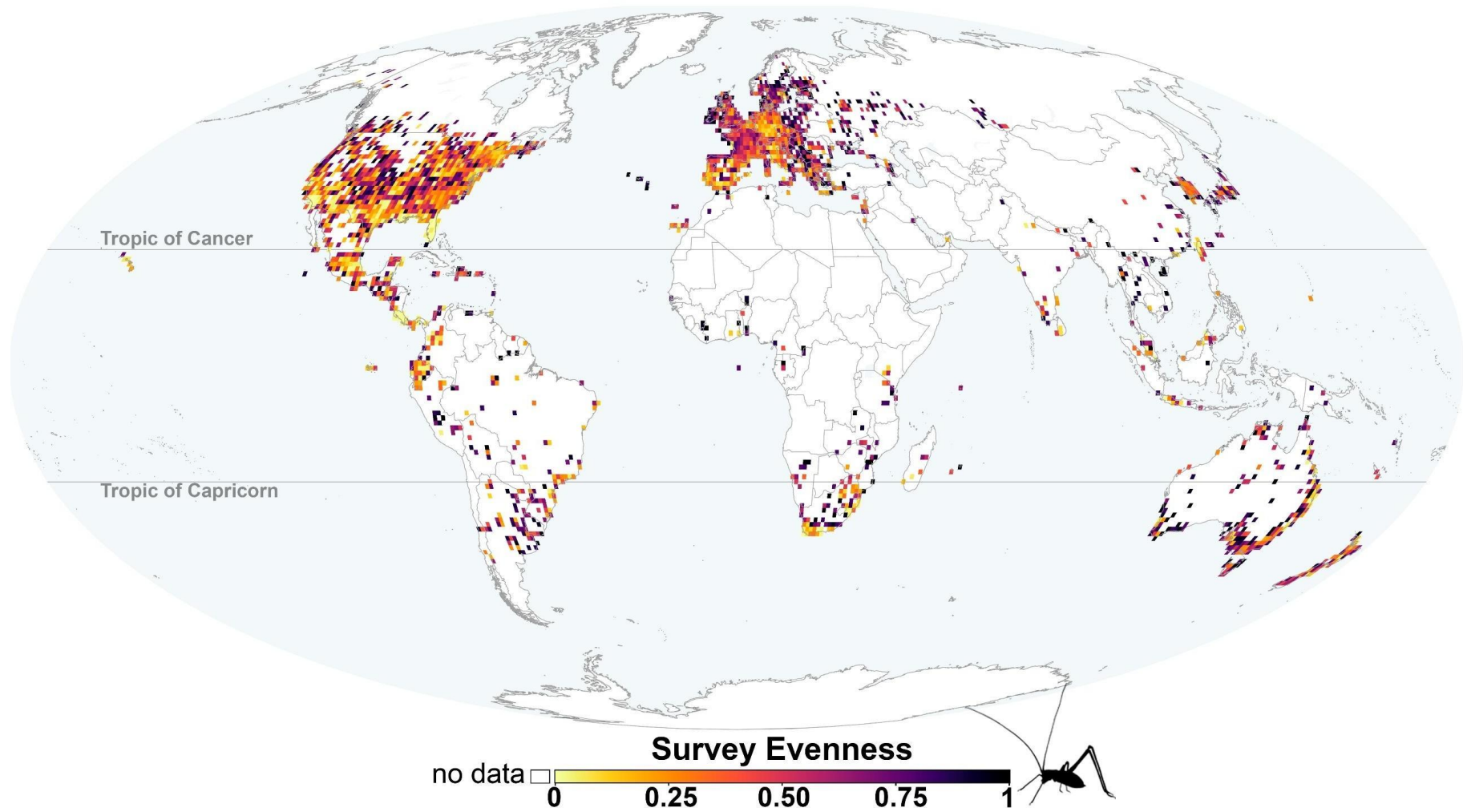
Supplementary Figure 5. Taxonomic Completeness (TaxC), values close to 1 indicate that most records are identified to the species taxonomic level.



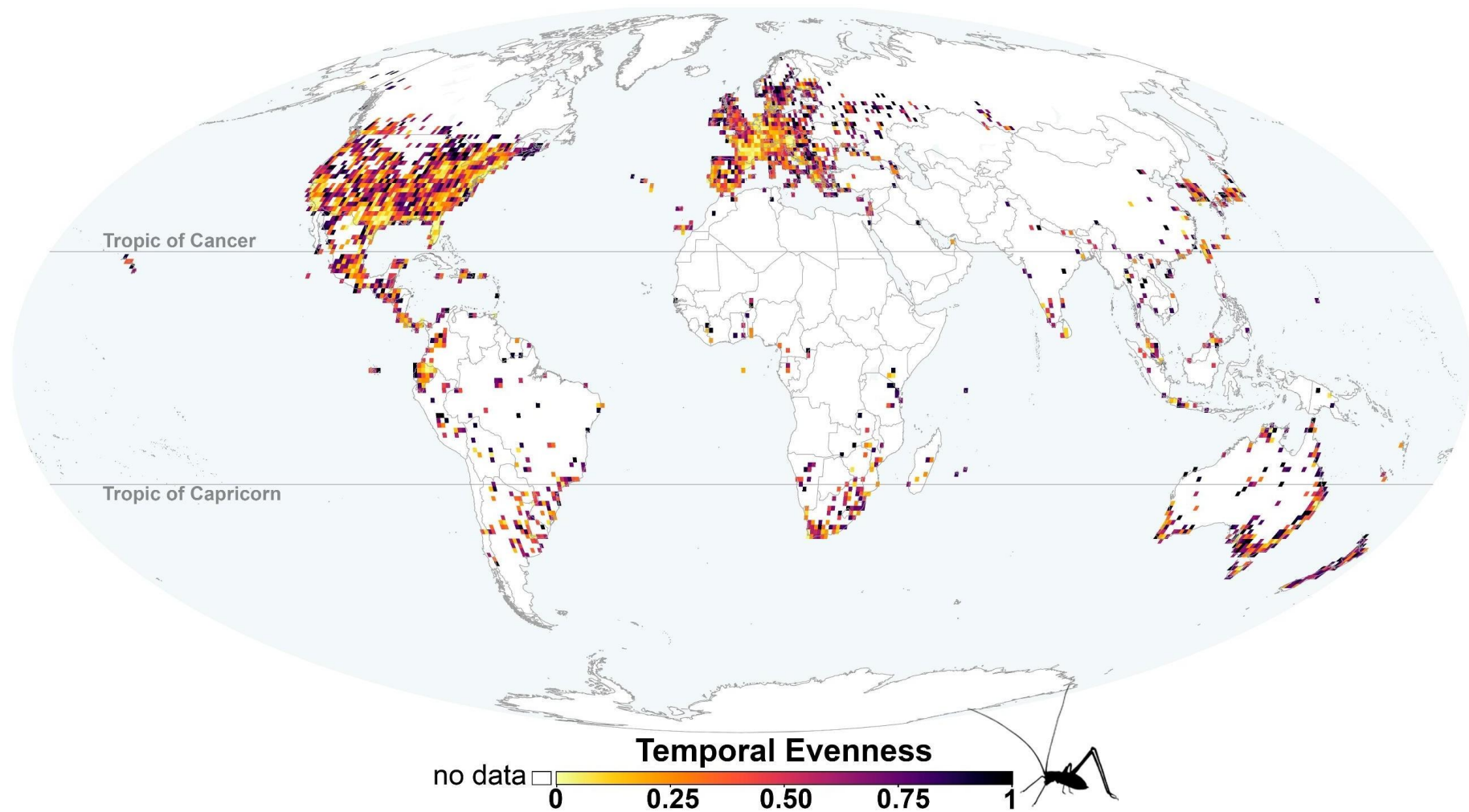
Supplementary Figure 6. Survey Completeness (SC), values close to 1 indicate an almost complete inventory.



Supplementary Figure 7. Temporal Completeness (TE), values close to 1 indicate maximum temporal coverage.



Supplementary Figure 8. Survey Evenness (SE), values close to 1 indicate no local bias in the distribution of the numbers of records per species.



Supplementary Figure 9. Temporal Evenness (TE), values close to 1 indicate no local bias in the distribution of samples per year.

CAPÍTULO III

Drivers da Ignorância do Conhecimento da Biodiversidade de Orthoptera (Arthropoda, Insecta)

Manuscrito a ser submetido à revista *Biotropica*

Drivers da Ignorância do Conhecimento da Biodiversidade de Orthoptera (Arthropoda, Insecta)

Rodrigo Antônio Castro-Souza^{1,5*}, Dieison André Moi², Geiziane Tessarolo³, Neucir Szinwelski⁴, José Alexandre Felizola Diniz-Filho³ & Thadeu Sobral-Souza⁵

(1) Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Federal de Mato Grosso (UFMT), Cuiabá, Brazil.

(2) Laboratório de Interações Multitróficas e Biodiversidade, Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual of Campinas (UNICAMP), Campinas, Brazil.

(3) Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil.

(4) Laboratório de Orthoptera, Centro de Ciências Biológicas e da Saúde, Universidade Estadual do Oeste do Paraná, Brazil.

(5) Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso, 78060-900, Cuiabá, MT, Brazil.

rodrigodesouzaac@gmail.com (*autor correspondente)

RESUMO

Compreender os drivers socioecológicos que moldam o conhecimento sobre a biodiversidade no espaço e no tempo é fundamental para a ecologia, evolução e conservação da biodiversidade. As lacunas no conhecimento sobre a distribuição de insetos como Orthoptera (Arthropoda: Insecta) refletem deficiências taxonômicas, espaciais e temporais, que juntas representam nossa ignorância em relação a este grupo. Desvendar os fatores que contribuem para aumentar ou diminuir essa ignorância representa um desafio chave na atualidade. Neste estudo, empregamos modelos de equações estruturais (SEM) para integrar drivers socioecológicos, como densidade populacional humana, densidade de áreas protegidas, desenvolvimento socioeconômico e riqueza de espécies de Orthoptera, ao longo do tempo a fim de explicar ignorância do conhecimento de Orthoptera em diferentes subdivisões biogeográficas (Global, Neártico, Neotropical, Paleártico, Afrotropical, Indo-Malaio e Australásia). Nossos resultados mostram que o tempo de amostragem aumenta a riqueza e reduz a ignorância em todas as subdivisões investigadas. A densidade populacional humana eleva a riqueza Global e Neártica, enquanto reduz a ignorância Paleártica. A densidade de áreas protegidas aumenta a riqueza Global, Neotropical, Paleártica e Indo-Malaia, ao passo que diminui a ignorância apenas na subdivisão Neotropical. Áreas mais desenvolvidas apresentam menor riqueza em todas as subdivisões. Particularmente no Global, Neártico e Paleártico, o desenvolvimento econômico está associado a uma menor ignorância, enquanto na Afrotropical, está relacionado a uma maior ignorância. Esses padrões fornecem insights valiosos para compreender e abordar as desigualdades no conhecimento sobre Orthoptera em larga escala.

Palavras Chaves: Lacunas de conhecimento da Biodiversidade; Mapas de Ignorância, Macroecologia; Conhecimento Digital Acessível; Qualidade do Dados;

INTRODUÇÃO

Em 1948, Samuel Whittemore Boggs ressaltou a necessidade da construção de mapas da ignorância para avançar no conhecimento em diferentes áreas da ciência, como biogeografia, geologia, climatologia, entre outras. Segundo ele, ao mapear o que se sabe - e o que ainda não se sabe - se torna possível compreender melhor locais com menor conhecimento e buscar soluções mais direcionadas para sanar a falta de informação (Boggs, 1949). Mais de meio século depois, na era da ecoinformática (Soberón & Peterson, 2004; Michener & Jones, 2012), quantificar a ignorância do conhecimento sobre a biodiversidade em larga escala tornou-se cada vez mais acessível devido às informações digitalmente acessíveis em bancos de dados biológicos, físicos e sociais (veja Rocchini et al., 2011; Meyer et al., 2016; Stropp et al., 2016; Tassarolo et al., 2021; Castro-Souza et al., 2024).

Ao mapear o conhecimento sobre a biodiversidade, podemos observar diferentes deficiências taxonômicas, espaciais e temporais (Ladle & Hortal, 2013; Stropp et al., 2016; Tassarolo et al., 2021; Castro-Souza et al., 2024). As deficiências taxonômicas podem estar relacionadas à falta de refinamento taxonômico nos dados coletados (Castro-Souza et al., 2024), enquanto as deficiências espaciais referem-se à completude e cobertura dos inventários que esses dados representam (Lobo et al., 2018; Tassarolo et al., 2021; Castro-Souza et al., 2024). Já as deficiências temporais dizem respeito à diversidade dos dados ao longo do tempo (Meyer et al., 2016; Castro-Souza et al., 2024). A combinação de uma ou mais dessas deficiências, mapeadas em uma perspectiva de larga escala, resulta nos chamados mapas de ignorância da biodiversidade (Ladle & Hortal, 2013; Tassarolo et al., 2021; Castro-Souza et al., 2024). Neste contexto, um novo desafio que surge na atualidade é compreender os drivers socioecológicos que podem estar associados a esses mapas de ignorância ao longo do tempo.

O aumento da densidade populacional humana em várias regiões do mundo nas últimas décadas (Schiavina et al., 2023) é um driver socioecológico diretamente relacionado ao conhecimento da biodiversidade. Esse crescimento contribuiu positivamente para uma melhor compreensão da riqueza de espécies em áreas densamente povoadas (Diniz-Filho et al., 2005; Luck et al., 2007; Colli et al., 2016; Moura et al., 2018), embora, ao mesmo tempo, muitos impactos antropogênicos possam já estar ameaçando tal biodiversidade. Além disso, a forte presença humana em muitos locais tem favorecido a redução da ignorância sobre a

biodiversidade, uma vez que facilita a realização de inventários, a acessibilidade de pesquisadores e uma melhor logística de exploração (Meyer et al., 2015; Hughes et al., 2021).

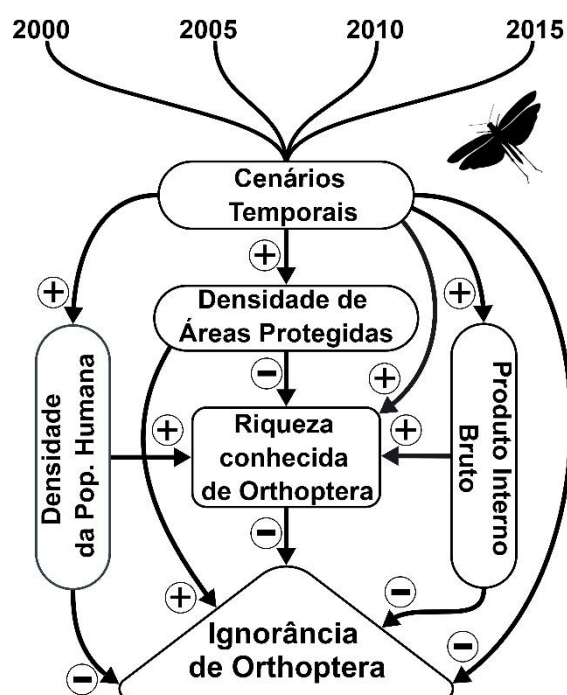
A presença de áreas protegidas também é um aspecto socioecológico diretamente ligado ao conhecimento da biodiversidade (Meyer et al., 2015; Pulido-Chadid et al., 2023; UNEP-WCMC & IUCN, 2024). A riqueza de espécies dentro das áreas protegidas pode variar em função da fragmentação histórica dos habitats, dos esforços de amostragem, dos táxons estudados e de sua conformação espacial (Collen et al., 2018; Moura et al., 2018; Sobral-Souza et al., 2021; Hughes et al., 2021). Contudo, embora muitas áreas protegidas em zonas tropicais abriguem uma rica biodiversidade, elas permanecem subexploradas cientificamente (Oliveira et al., 2017). Essa desigualdade na distribuição do conhecimento sobre a biodiversidade em áreas protegidas resulta em lacunas no entendimento da riqueza de espécies e amplia nossa ignorância sobre a biodiversidade, especialmente em regiões protegidas tropicais, devido à falta de informações nessas áreas.

O grau de desenvolvimento socioeconômico é outro driver socioecológico que influencia o conhecimento sobre a biodiversidade (Steck & Pautasso, 2008; Golding et al., 2010; Meyer et al., 2015; Moura et al., 2018). Em regiões com maior Produto Interno Bruto (PIB), a disponibilidade de recursos para pesquisa tem viabilizado um volume mais expressivo de dados primários sobre a biodiversidade (Collen et al., 2018; Meyer et al., 2015). No entanto, muitas dessas áreas são altamente antropizadas e podem já ter perdido grande parte de sua biodiversidade (Dasgupta & Levin, 2023). Assim, o maior número de registros nesses locais pode gerar um viés de confusão, levando à interpretação equivocada de que áreas com maior PIB e intensa ação antrópica abrigam mais biodiversidade e apresentam menor ignorância no conhecimento.

Diante destas problemáticas, neste estudo, empregamos modelos de equações estruturais (SEM) a fim de integrar drivers sociológicos, como densidade populacional humana, densidade de áreas protegidas, desenvolvimento socioeconômico e riqueza de espécies de Orthoptera conhecida (i.e., baseada em dados disponíveis on-line), ao longo do tempo, para explicar a falta de conhecimento (i.e., ignorância) de Orthoptera em diferentes subdivisões biogeográficas (Global, Neártico, Neotropical, Paleártico, Afrotropical, Indo-Malaio e Australásia). Hipotetizamos que a riqueza conhecida de Orthoptera seja positivamente influenciada por cenários temporais recentes, densidade populacional humana, e desenvolvimento socioeconômico, mas negativamente pela densidade de áreas protegidas. Em

relação a ignorância, hipotetizamos que a ignorância seja menor em cenários temporais mais recentes, áreas com maior densidade populacional, regiões com maior desenvolvimento socioeconômico, e em locais com maior riqueza de espécies de Orthoptera conhecidas. Por outro lado, esperamos observar maior ignorância em regiões com maior densidade de áreas protegidas (Fig. 1, modelo hipotético de caminhos).

Figura 1. Modelo hipotético de caminhos representando as relações direcionais analisadas e



as hipóteses subjacentes, efeito positivo (+) e inverso (-) entre os diferentes drivers e a ignorância do conhecimento de Orthoptera em larga escala.

MÉTODOS

Mapeando a Ignorância de Orthoptera

Combinamos três diferentes índices de qualidade dos dados conhecidos sobre Orthoptera (TaxC, SC e TC) para construir mapas globais de ignorância na resolução espacial de 1° (~100 x 100 km), em quatro cenários temporais (2000, 2005, 2010 e 2015). Utilizamos os mesmos dados e limiares (i.e., ≤ 45 registros totais para TaxC e de espécies para SC e TC) empregados por Castro-Souza et al. (2024) (Fig. 2a).

A (i) *Compleitude taxonômica* (TaxC)—indica a qualidade (refinamento) taxonômico dos dados onde valores próximos a 1 indicam dados taxonomicamente confiáveis e valores

próximos a 0 indicam dados com baixa qualidade taxonômica, demonstrando a necessidade de refinamento taxonômico. (ii) *Compleitude amostral* (SC)—indica a qualidade do inventário (esforço). Valores próximos a 1 indicam localidades bem conhecidas e com alto esforço amostral, enquanto valores próximos a 0 demonstram locais com pouco conhecimento onde se necessita de mais amostragens. (iii) *Compleitude temporal* (TC)—indica a cobertura temporal (i.e. número de anos de levantamentos). Valores próximos de 1 indicam lugares com maior cobertura de amostragem e valores próximos a 0 indicam lugares com baixa cobertura temporal (veja Castro-Souza et al., 2024).

Ao final, somamos os mapas das três métricas para cada ano e reescalamos o mapa final para variar de 0 até 1, e por fim, subtraímos de 1. Assim, os valores de ignorância próximos a 1 indicam máxima ignorância (ou seja, qualidade de dados ruim), enquanto valores próximos de 0 indicam baixa ignorância (ou seja, boa qualidade de dados) (Fig. 2b).

Drivers Preditores

Como mencionado na introdução deste estudo, processos socioecológicos podem influenciar tanto a riqueza de espécies conhecidas quanto a falta de conhecimento sobre a biodiversidade ao longo do espaço e do tempo. Com base nas premissas levantadas, analisamos como a densidade populacional humana, a densidade de áreas protegidas, o produto interno bruto e a riqueza de espécies de Orthoptera, ao longo do tempo, interagem com a ignorância de Orthoptera por meio de análises de caminho (conforme Lefcheck, 2016).

Densidade da População Humana (POP)—Extraímos os valores de *Gridded Population of the World, Version 4* (GPWv4) (<https://doi.org/10.7927/H49C6VHW>) para os anos de 2000, 2005, 2010 e 2015, disponibilizados pelo Socioeconomic Data and Applications Center (SEDAC), no formato raster com uma resolução de 0,5° (~50 x 50 km).

Densidade de Áreas Protegidas (PAD)—Filtramos as áreas protegidas globais da UNEP-WCMC & IUCN (2024) propostas até os anos de 2000, 2005, 2010 e 2015, e dividimos as áreas protegidas em células raster de 0,01° (~1 x 1 km). Seguidamente, calculamos a quantidade dessas células protegidas que ocorriam dentro de cada célula de 1° na nossa área de estudo. Por fim, dividimos a quantidade de células protegidas pela quantidade total de células de 0,01° presentes em cada célula de 1° (i.e., 10,000), obtendo valores de 0 a 1, onde 0 indica

a ausência de áreas protegidas na célula e valores entre 0 e 1 indicam desde uma baixa até uma alta densidade de áreas protegidas naquela célula.

Produto Interno Bruto (GDP)—Extraímos os valores do Produto Interno Bruto (GDP) para os anos de 2000, 2005, 2010 e 2015 disponibilizados por Kummu et al. (2018) a partir do repositório DRYAD (<https://doi.org/10.5061/dryad.dk1j0>), no formato raster com uma resolução de 0.08° (~ 8 x 8 km).

Riqueza de espécies de Orthoptera (RIQ)—Quantificamos a riqueza de espécies de Orthoptera acumulada para os anos de 2000, 2005, 2010 e 2015 utilizando os dados compilados por Castro-Souza et al. (2024), em cada célula raster de 1° (~ 100 x 100km).

Por fim, os rasters de *Densidade da População Humana* e *Produto Interno Bruto* foram reescalados para 1° (i.e., compatível com nosso mapa de ignorância) utilizando a função *aggregate* com o método *mean* do pacote “raster” (Hijmans, 2023) em ambiente R (Fig. 2c).

Análises Estatísticas

Realizamos uma normalização min-max em todas as variáveis a fim de comparar coeficientes de caminho entre variáveis de diferente magnitude. Em seguida, a partir do modelo hipotético de caminhos, verificamos a multicolinearidade, calculando o fator de inflação da variância (VIF) para cada driver (i.e., $VIF < 3$, ausência de multicolinearidade significativa) (Fig. 2, d). Assim, elaboramos um modelo de equação estrutural global a partir de um conjunto de modelos de regressão linear misto e método de ajuste Restricted maximum likelihood (REML), onde consideramos cada localidade (i.e., célula raster) como fator aleatório, e os drivers soci ecológicos de densidade populacional humana, densidade de áreas protegidas, produto interno bruto, riqueza de espécies de Orthoptera e os diferentes cenários temporais como fatores fixos. Para cada regressão utilizamos a função *lme* do pacote “nlme” (Pinheiro et al., 2017) e para construir o modelo estrutural utilizamos a função *psem* do pacote “piecewiseSEM” (Lefcheck, 2016) (Fig. 2, e).

Com o ajuste do modelo em escala global, utilizamos a abordagem “multigrupo” do SEM em partes a partir da função *multigroup*, para projetar o modelo para seis subdivisões biogeográficas do World Wildlife Fund (WWF) (Neártico, Neotropical, Paleártico Afrotropical, Indo-Malaio e Australásia; veja: <https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>) como uma

variável categórica de agrupamento (Fig. 2, e). Essa abordagem incorpora um termo de interação no modelo para avaliar se a inclusão de caminhos específicos para cada subdivisão biogeográfica melhora a qualidade do modelo. Interações significativas permitem a variação do coeficiente de caminho conforme a subdivisão biogeográfica, enquanto, no caso oposto, a estimativa é fixa (*constrain*) em todas. Avaliamos o ajuste do SEM usando a estatística C de Fisher e consideramos o modelo incompleto se o valor de p associado fosse inferior a 0,05.

A partir dos coeficientes padronizados do SEM, avaliamos os efeitos diretos entre todos os drivers, bem como os efeitos indiretos dos drivers mediados pela riqueza de espécies de Orthoptera. No caso dos efeitos indiretos, consideramos como significativos aqueles em que ambos os caminhos de conexão individuais foram significativos (i.e., $p \leq 0,05$). Representamos graficamente o SEM por meio de plots gerados com o pacote “ggplot” (Wickham, 2016) e diagramas elaborados no programa Inkscape (Inkscape, 2024). Todas as análises mencionadas acima foram conduzidas em R (R Core Team, 2024).

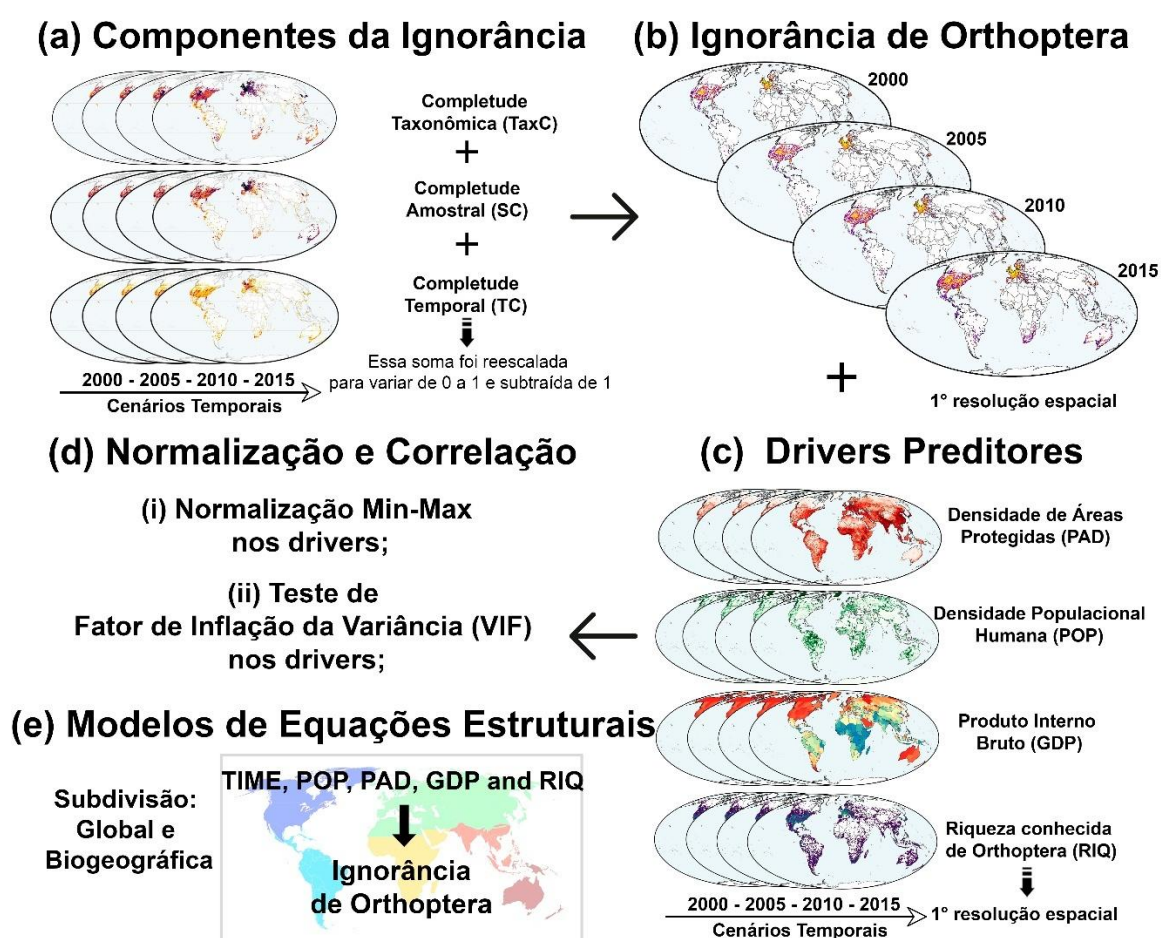


Figura 2. Esboço resumindo nossas etapas metodológicas e analíticas para prever o efeito de diferentes drivers sociológicos sobre a ignorância do conhecimento de Orthoptera ao longo

do tempo. O primeiro passo consiste em mensurarmos a completude taxonômica, espacial e temporal em quatro diferentes cenários temporais (a), combinarmos essas três medidas por meio de uma soma para cada cenário temporal, reescalarmos para um intervalo entre 0 e 1 e subtraímos de 1, construindo assim nossos mapas de ignorância para Orthoptera (b). O segundo passo consiste na extração de densidade de áreas protegidas, densidade populacional humana, produto interno bruto e riqueza de espécies de Orthoptera para os mesmos cenários e resolução espacial dos mapas de ignorância (c). Seguidamente, transformamos os dados para uma mesma escala numérica utilizando normalização min-max e avaliamos a correlação entre eles por meio de um teste de fator de inflação da variância (VIF) (d). Finalmente, construímos nosso modelo de equações estruturais em diferentes subdivisões biogeográficas associando os drivers preditores com a ignorância do conhecimento de Orthoptera (e).

RESULTADOS

O modelo construído apresentou um bom ajuste estatístico (Fisher's $C = 3.33$, $p\text{-value} = 0.18$). Os drivers utilizados explicaram 88% da ignorância sobre biodiversidade de orthoptera e 96% da riqueza de espécies conhecida de Orthoptera, considerando o R^2 condicional (ou seja, fatores fixos e aleatórios). Destacando que as características locais (escala local) desempenham um papel crucial nos padrões de ignorância da biodiversidade de Orthoptera e que os drivers socioecológicos podem atuar nessas escalas (Fig. Suplementar 1).

Os cenários temporais recentes demonstram um aumento na densidade populacional humana (POP), na densidade de áreas protegidas (PAD), no produto interno bruto (GDP) e na riqueza conhecida de espécies de Orthoptera (RIQ) em todas subdivisões biogeográficas (Figs. 3a, b, c, d). Em relação à ignorância, observou-se um efeito inverso, com a qualidade dos dados aumentando nos cenários mais recentes (Fig. 3e).

O aumento da densidade populacional (POP) aumenta a riqueza conhecida de espécies de Orthoptera (RIQ) nas subdivisões biogeográficas Global e Neotropical (Fig. 3f) e reduz a ignorância no Paleártico (Fig. 3g). Já o aumento da densidade de áreas protegidas (PAD) aumenta a riqueza conhecida de espécies de Orthoptera Globalmente e nas subdivisões Neotropical, Paleártico e Indo-Malaio (Fig. 3h). O aumento do Produto Interno Bruto (GDP) possui efeito inverso na riqueza conhecida de espécies de Orthoptera (RIQ) em todas as subdivisões biogeográficas, indicando que o conhecimento sobre a riqueza tende a ser menor nas regiões com maior GDP (Fig. 3j). Em relação à ignorância, quanto maior o GDP menor a

ignorância Global, no Neártico e no Paleártico (Fig. 3k) e maior a ignorância na subdivisão Afrotropical (Fig. 3k). Por fim, o aumento da riqueza conhecida de espécies de Orthoptera (RIQ) diminui a ignorância em todas as subdivisões biogeográficas (Fig. 3l).

Considerando os efeitos das interações indiretas dos cenários temporais mediados pela riqueza e interagindo com outros fatores, bem como os fatores socioecológicos mediados pela riqueza e afetando a ignorância, observamos relações extremamente fracas, embora algumas sejam significativas. Por esse motivo, optamos por apresentá-los apenas no Apêndice Suplementar 1 e na Figura Suplementar 2.

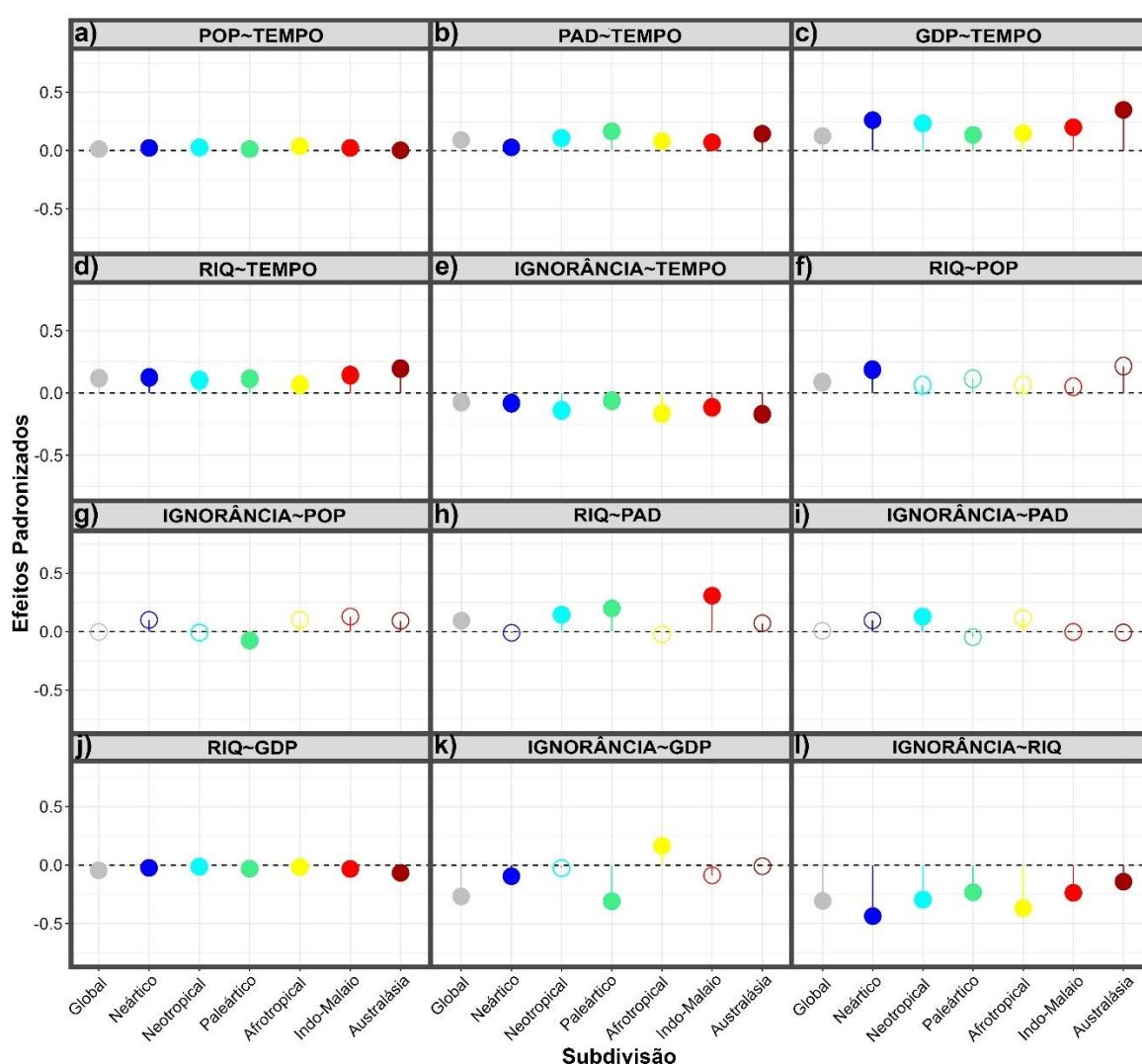


Figura 3. Efeitos padronizados do Modelo de Equações Estruturais (SEM) considerando todas as interações diretas. As cores representam as diferentes subdivisões biogeográficas investigadas: Cinza – Global; Azul – Neártico; Azul claro – Neotropical; Verde – Paleártico; Amarelo – Afrotropical; Vermelho – Indo-Malaio; Vermelho escuro – Australásia. Círculos

preenchidos indicam efeitos significativos ($p < 0,05$). Os resultados completos do SEM estão disponíveis na Tabela e na Figura Suplementar 1.

DISCUSSÃO

Nossos resultados revelaram que tanto a riqueza conhecida quanto a ignorância de Orthoptera são diretamente influenciadas por processos antrópicos, conservacionistas, socioeconômicos e fatores locais ao longo do tempo. Além disso, demonstramos que o tempo de amostragem aumenta a riqueza conhecida e reduz a ignorância em todas as subdivisões biogeográficas, corroborando nossas hipóteses temporais. Observamos que a densidade populacional humana eleva a riqueza conhecida nas subdivisões Global e Neártica, enquanto reduz a ignorância na Paleártica, corroborando parcialmente nossas hipóteses populacionais. Já a densidade de áreas protegidas aumenta a riqueza conhecida nas subdivisões Global, Neotropical, Paleártica e Indo-Malaia, o que refuta nossas hipóteses iniciais sobre os efeitos das áreas protegidas sobre a riqueza conhecida; no entanto, esse mesmo fator eleva a ignorância na subdivisão Neotropical, corroborando apenas parte de nossas previsões. Áreas economicamente mais desenvolvidas apresentam menor riqueza conhecida em todas as subdivisões, contrariando nossas hipóteses iniciais sobre o desenvolvimento econômico e riqueza conhecida. No entanto, particularmente nas subdivisões Global, Neártica e Paleártica, o desenvolvimento econômico está associado a uma menor ignorância, ao passo que na subdivisão Afrotropical, essa relação se inverte, estando associada a uma maior ignorância. Tais resultados destacam a complexidade e a heterogeneidade dos fatores que determinam as lacunas de conhecimento sobre Orthoptera ao longo do espaço e do tempo.

O efeito do tempo no aumento da riqueza de espécies conhecida e na redução da ignorância do conhecimento de Orthoptera, pode ser atribuído às iniciativas massivas de disponibilidade de registros primários de biodiversidade em diferentes regiões do mundo nas últimas décadas (Edwards et al., 2000; Soberón & Peterson, 2004; Sousa-Baena et al., 2014; Nelson & Ellis, 2018). A falta de conhecimento nos períodos mais recentes tende a ser menor, embora a digitalização de coleções biológicas e a realização de novos inventários ainda sejam necessárias em muitas partes do globo (Mandeville et al., 2021; Nuñez-Penichet et al., 2022; Castro-Souza et al., 2024). Nos cenários mais recentes também existe um avanço no uso de geotecnologias em celulares, tablets e câmeras que permitem a coleta de informações espaciais nos registros de espécies que antigamente não era possível. Além disso, os avanços na bioinformática também têm contribuído para logística mais eficiente de armazenamento e

disponibilidade de registros primários em larga escala (Soberón & Peterson, 2004). Contudo, é importante destacar que a disponibilidade de grandes volumes de dados nem sempre garante uma boa cobertura espaço-temporal do conhecimento de distribuição da biodiversidade (Meyer et al., 2016; Stropp et al., 2016; Nelson & Ellis, 2018; Castro-Souza et al., 2024) e pode não ser suficiente para melhorar a qualidade do conhecimento sobre Orthoptera dentro de um ritmo aceitável, à medida que as mudanças climáticas e os impactos humanos avançam (veja, Farooq et al., 2021).

A relação positiva entre a densidade populacional humana e a riqueza de Orthoptera indica um dos fatores de viés de conhecimento, uma vez que áreas mais povoadas são mais acessíveis, facilitando a documentação de sua biodiversidade (Luck, 2007; Oliveira et al., 2016; Hughes et al., 2021). Contudo, a ausência de relação entre a densidade populacional humana e a ignorância sobre Orthoptera (exceto no Paleártico) pode estar associada à grande falta de refinamento taxonômico, inventários e à abrangência temporal restrita do conhecimento digital acessível para Orthoptera (Castro-Souza et al., 2024), mesmo em regiões densamente povoadas. Isso pode ser explicado pelos desafios de coletar ao longo do tempo e identificar taxonomicamente organismos menores, menos chamativos e extremamente diversos, como muitas ordens de insetos, especialmente quando comparados a grupos mais chamativos, como aves e mamíferos, que possuem um conhecimento mais consolidado em áreas densamente povoadas (Luck, 2007).

Embora muitas áreas protegidas exibam extensas lacunas de distribuição da biodiversidade (Oliveira et al., 2017; Hughes et al., 2021), de acordo com nossos achados, tais ambientes tendem a apresentar maior riqueza conhecida de espécies de Orthoptera. Isso reforça a importância das áreas protegidas para a conservação da diversidade. Além disso, globalmente e no Neártico, a riqueza de Orthoptera pode estar mais ameaçada devido à interação positiva com a densidade populacional humana observada neste estudo, que consequentemente pode gerar impactos antropogênicos (Wolf et al., 2021). Por outro lado, regiões com maior densidade de áreas protegidas apresentaram maior ignorância sobre Orthoptera na subdivisão Neotropical, destacando a baixa qualidade dos dados disponíveis para regiões protegidas nas zonas (Oliveira et al., 2017).

A relação inversa entre a riqueza conhecida de Orthoptera e o desenvolvimento socioeconômico em todas as subdivisões investigadas sugere que, em regiões com elevado produto interno bruto, muitos ambientes podem estar degradados, dificultando a manutenção

de uma elevada riqueza de Orthoptera. Esse padrão contrasta com o observado anteriormente para comunidades de gafanhotos na Europa (Steck & Pautasso, 2008), mas também levanta dúvidas sobre a qualidade dos dados digitalmente acessíveis que estão disponíveis até o momento. Por outro lado, curiosamente, ao considerar a ignorância, nosso modelo indica que o desenvolvimento socioeconômico pode ser benéfico Globalmente, no Neártico e Paleártico, mas, em alguns casos, prejudicial como na subdivisão Afrotropical. Isso pode estar relacionado ao fato de que uma maior concentração de poder econômico no mundo favorece a acumulação de conhecimento sobre biodiversidade em determinadas latitudes (Collen et al., 2018; Raja et al., 2022; Nakamura et al., 2024), aqui representadas pelas regiões Neártica e Paleártica, o que também pode contribuir para a melhoria da qualidade desse conhecimento ao longo do tempo.

Por fim, ao evidenciar como esses fatores variam entre diferentes subdivisões do planeta, nosso estudo ressalta a importância de estratégias específicas para preencher as lacunas de conhecimento e promover a conservação de Orthoptera. Isso inclui o financiamento de projetos voltados à disponibilização e documentação de registros desse grupo, a importância da manutenção de áreas protegidas para resguardar sua biodiversidade e a necessidade de esforços conjuntos entre taxonomistas, ecólogos e ciência de dados. Além disso, compreender os múltiplos aspectos socioecológicos associados ao mapeamento dessas lacunas é fundamental para decifrar os padrões derivados do conhecimento digital acessível. No contexto do Antropoceno, torna-se cada vez mais evidente que os efeitos das atividades humanas estão profundamente entrelaçados com as lacunas no conhecimento sobre a biodiversidade.

AGRADECIMENTOS

Agradecemos à equipe do Laboratório de Macroecologia e Conservação da Biodiversidade (MacrEco) da Universidade Federal de Mato Grosso (UFMT) e ao Instituto Nacional de Ciência e Tecnologia (INCT) em Ecologia, Evolução e Conservação da Biodiversidade (EECBio), sediado na Universidade Federal de Goiás (UFG), pelas valiosas discussões que contribuíram para o desenvolvimento deste estudo. Também agradecemos à FAPEMAT pelo apoio financeiro (projeto FAPEMAT-PRO.000274/2023). RACS expressa sua gratidão à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Centro Nacional de Pesquisa e Conservação de Cavernas (CECAV), pelas bolsas de doutorado e sanduíche concedidas.

REFERÊNCIAS

- Boggs, S. W. (1949). An atlas of ignorance: a needed stimulus to honest thinking and hard work. *Proceedings of the American Philosophical Society*, 93(3), 253–258.
- Castro-Souza, R. A., Tessarolo, G., Stropp, J., Diniz-Filho, J. A., Ladle, R. J., Szinwelski, N., ... & Sobral-Souza, T. (2024). Mapping ignorance to uncover shortfalls in the knowledge on global Orthoptera distribution. *npj Biodiversity*, 3(1), 22.
- Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Conservation Science*, 1(2), 75–88.
- Colli, G. R., Fenker, J., Tedeschi, L. G., Barreto-Lima, A. F., Mott, T., & Ribeiro, S. L. (2016). In the depths of obscurity: Knowledge gaps and extinction risk of Brazilian worm lizards (Squamata, Amphisbaenidae). *Biological Conservation*, 204, 51–62.
- Dasgupta, P., & Levin, S. (2023). Economic factors underlying biodiversity loss. *Philosophical Transactions of the Royal Society B*, 378(1881), 20220197.
- Diniz-Filho, J. A. F., Bastos, R. P., Rangel, T. F., Bini, L. M., Carvalho, P., & Silva, R. J. (2005). Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Global Ecology and Biogeography*, 14(5), 469–477.
- Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, 289(5488), 2312–2314.
- Farooq, H., Azevedo, J. A., Soares, A., Antonelli, A., & Faurby, S. (2021). Mapping Africa's biodiversity: More of the same is just not good enough. *Systematic Biology*, 70(3), 623–633.
- Golding, J., Güsewell, S., Kreft, H., Kuzevanov, V. Y., Lehvävirta, S., Parmentier, I., & Pautasso, M. (2010). Species-richness patterns of the living collections of the world's botanic gardens: a matter of socio-economics?. *Annals of Botany*, 105(5), 689–696.
- Hijmans, R. (2023). raster: Geographic Data Analysis and Modeling. R package version 3.6-26, <<https://CRAN.R-project.org/package=raster>>.
- Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., ... & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, 44(9), 1259–1269.

- Inkscape Team (2004–2024). Inkscape: A Vector Drawing Tool. Available at <https://www.inkscape.org> (accessed 12 November 2024).
- Kummu, M., Taka, M., & Guillaume, J. H. (2018). Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. *Scientific data*, 5(1), 1–15.
- Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, 7(5), 573–579.
- Ladle, R., & Hortal, J. (2013). Mapping species distributions: living with uncertainty. *Frontiers of Biogeography*, 5(1).
- Lobo, J. M., Hortal, J., Yela, J. L., Millán, A., Sánchez-Fernández, D., García-Roselló, E., ... & Guisande, C. (2018). KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecological Indicators*, 91, 241–248.
- Luck, G. W. (2007). A review of the relationships between human population density and biodiversity. *Biological Reviews*, 82(4), 607–645.
- Mandeville, C. P., Koch, W., Nilsen, E. B., & Finstad, A. G. (2021). Open data practices among users of primary biodiversity data. *BioScience*, 71(11), 1128–1147.
- Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature communications*, 6(1), 1–8.
- Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology letters*, 19(8), 992–1006.
- Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in ecology & evolution*, 27(2), 85–93.
- Moura, M. R., Costa, H. C., Peixoto, M. A., Carvalho, A. L., Santana, D. J., & Vasconcelos, H. L. (2018). Geographical and socioeconomic determinants of species discovery trends in a biodiversity hotspot. *Biological Conservation*, 220, 237–244.
- Nakamura et al. (2024). The macroecology of knowledge: Spatio-temporal patterns of name-bearing types in biodiversity science. Preprint. 1–47.

- 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*, 374(1763), 20170391.
- Núñez-Penichet, C., Cobos, M. E., Soberón, J., Gueta, T., Barve, N., Barve, V., ... & Peterson, A. T. (2022). Selection of sampling sites for biodiversity inventory: Effects of environmental and geographical considerations. *Methods in Ecology and Evolution*, 13(7), 1595–1607.
- Oliveira, U., Paglia, A. P., Brescovit, A. D., de Carvalho, C. J., Silva, D. P., Rezende, D. T., ... & Santos, A. J. (2016). The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. *Diversity and Distributions*, 22(12), 1232–1244.
- Oliveira, U., Soares-Filho, B. S., Paglia, A. P., Brescovit, A. D., De Carvalho, C. J., Silva, D. P., ... & Santos, A. J. (2017). Biodiversity conservation gaps in the Brazilian protected areas. *Scientific reports*, 7(1), 9141.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package ‘nlme’. Linear and nonlinear mixed effects models, version, 3(1), 274.
- Pulido-Chadid, K., Virtanen, E., & Geldmann, J. (2023). How effective are protected areas for reducing threats to biodiversity? A systematic review protocol. *Environmental Evidence*, 12(1), 18.
- R Core Team (2024). R: A language and environment for statistical computing. <<https://www.R-project.org/>>.
- Raja, N. B., Dunne, E. M., Matiwane, A., Khan, T. M., Nätscher, P. S., Ghilardi, A. M., & Chattopadhyay, D. (2022). Colonial history and global economics distort our understanding of deep-time biodiversity. *Nature ecology & evolution*. 6(2), 145–154.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jimenez-Valverde, A., Ricotta, C., ... & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Progress in Physical Geography*, 35(2), 211–226.

- Schiavina M., Freire S., Carioli A. & MacManus K. (2023): GHS-POP R2023A - GHS population grid multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) PID: <http://data.europa.eu/89h/2ff68a52-5b5b-4a22-8f40-c41da8332cfe>, doi:10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE
- Soberón, J., & Peterson, T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1444), 689–698.
- Sobral-Souza, T., Stropp, J., Santos, J. P., Prasniewski, V. M., Szinwelski, N., Vilela, B., ... & Hortal, J. (2021). Knowledge gaps hamper understanding the relationship between fragmentation and biodiversity loss: the case of Atlantic Forest fruit-feeding butterflies. *PeerJ*, 9, e11673.
- Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and distributions*, 20(4), 369–381.
- Steck, C. E., & Pautasso, M. (2008). Human population, grasshopper and plant species richness in European countries. *Acta oecologica*, 34(3), 303–310.
- Stropp, J., Ladle, R. J., M. Malhado, A. C., Hortal, J., Gaffuri, J., H. Temperley, W., ... & Mayaux, P. (2016). Mapping ignorance: 300 years of collecting flowering plants in Africa. *Global Ecology and Biogeography*, 25(9), 1085–1096.
- Tessarolo, G., Ladle, R. J., 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.
- UNEP-WCMC and IUCN (2024). Protected Planet: The World Database on Protected Areas (WDPA), August of 2024, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

Wolf, C., Levi, T., Ripple, W. J., Zárrete-Charry, D. A., & Betts, M. G. (2021). A forest loss report card for the world's protected areas. *Nature Ecology & Evolution*, 5(4), 520–529.

Material Suplementar

Tabela Suplementar 1. Coeficientes de caminhos diretos, tanto padronizados quanto não padronizados, de todas as relações testadas em nosso Modelo de Equação Estrutural (SEM) para prever a ignorância da biodiversidade de Orthoptera. Os drivers incluem: cenário temporal (TEMPO), densidade da população humana (POP), densidade de áreas protegidas (PAD), produto interno bruto (GDP) e riqueza de espécies de Orthoptera (RIQ). A tabela inclui todos os caminhos significativos e não significativos ($\rightarrow$) em diferentes subdivisões biogeográfica e as variáveis que foram permitidas para covariar ($\leftrightarrow$). * = $P < 0.05$, ** = $P < 0.01$ e *** = $P < 0.001$.

Subdivisão	Preditores	Caminho	Resposta	Coeficientes Padronizados	Peso das Regressões	P-valor	Constrained
Global	TEMPO	$\rightarrow$	GDP	0.1256	0.0044	<0***	–
Global	TEMPO	$\rightarrow$	POP	0.0125	0,0001	<0***	–
Global	TEMPO	$\rightarrow$	PAD	0.0922	0.0023	<0***	–
Global	GDP	$\rightarrow$	RIQ	-0.0449	-0.0215	<0.004**	–
Global	POP	$\rightarrow$	RIQ	0.0864	0.3108	<0***	–
Global	PAD	$\rightarrow$	RIQ	0.0952	0.0641	<0***	–
Global	TEMPO	$\rightarrow$	RIQ	0.1187	0.002	<0***	–
Global	GDP	$\rightarrow$	IGNORÂNCIA	-0.2672	-0.2199	<0***	–
Global	PAD	$\rightarrow$	IGNORÂNCIA	0.0074	0.0086	0.6037	–
Global	POP	$\rightarrow$	IGNORÂNCIA	-0.003	-0.0184	0.8516	–
Global	RIQ	$\rightarrow$	IGNORÂNCIA	-0.3063	-0.5248	<0***	–
Global	TEMPO	$\rightarrow$	IGNORÂNCIA	-0.0741	-0.0021	<0***	–
Global	~~GDP	$\leftrightarrow$	~~POP	-0.041	-0.041	<0,0007***	–
Global	~~ PAD	$\leftrightarrow$	~~GDP	-0.0565	-0.0565	<0***	–
Neártico	TEMPO	$\rightarrow$	GDP	0.2624	0.0046	<0***	–
Neártico	TEMPO	$\rightarrow$	POP	0.0224	0	<0***	–
Neártico	TEMPO	$\rightarrow$	PAD	0.0291	0,0006	<0***	–

Neártico	GDP	→	RIQ	-0.0233	-0.0215	<0.004**	sim
Neártico	POP	→	RIQ	0.1856	17376	<0***	—
Neártico	PAD	→	RIQ	-0.0098	-0.0077	0.6841	—
Neártico	TEMPO	→	RIQ	0.1249	0.002	<0***	—
Neártico	GDP	→	IGNORÂNCIA	-0.0942	-0.136	<0***	—
Neártico	PAD	→	IGNORÂNCIA	0.097	0.1185	<0,0002***	—
Neártico	POP	→	IGNORÂNCIA	0.1003	14689	<0,0007***	—
Neártico	RIQ	→	IGNORÂNCIA	-0.4366	-0.6825	<0***	—
Neártico	TEMPO	→	IGNORÂNCIA	-0.0825	-0.0021	<0***	sim
Neotropical	TEMPO	→	GDP	0.2349	0.0033	<0***	—
Neotropical	TEMPO	→	POP	0.0267	0,0002	<0***	—
Neotropical	TEMPO	→	PAD	0.1098	0.0033	<0***	—
Neotropical	GDP	→	RIQ	-0.0133	-0.0215	<0.004**	sim
Neotropical	POP	→	RIQ	0.0615	0.2389	0.2438	—
Neotropical	PAD	→	RIQ	0.1452	0.1097	<0***	—
Neotropical	TEMPO	→	RIQ	0.1026	0.0023	<0***	—
Neotropical	GDP	→	IGNORÂNCIA	-0.0248	-0.0264	0.5484	—
Neotropical	PAD	→	IGNORÂNCIA	0.1294	0.0648	<0.004**	—
Neotropical	POP	→	IGNORÂNCIA	-0.0087	-0.0224	0.8867	—
Neotropical	RIQ	→	IGNORÂNCIA	-0.2955	-0.1957	<0***	—
Neotropical	TEMPO	→	IGNORÂNCIA	-0.1408	-0.0021	<0***	sim
Paleártico	TEMPO	→	GDP	0.1355	0.0034	<0***	—
Paleártico	TEMPO	→	POP	0.0122	0,0001	<0***	—
Paleártico	TEMPO	→	PAD	0.167	0.0041	<0***	—
Paleártico	GDP	→	RIQ	-0.031	-0.0215	<0.004**	sim
Paleártico	POP	→	RIQ	0.1129	0.4656	<0,0009***	—
Paleártico	PAD	→	RIQ	0.1994	0.1425	<0***	—
Paleártico	TEMPO	→	RIQ	0.114	0.002	<0***	—
Paleártico	GDP	→	IGNORÂNCIA	-0.3076	-0.4106	<0***	—
Paleártico	PAD	→	IGNORÂNCIA	-0.0469	-0.0643	0.0823	—
Paleártico	POP	→	IGNORÂNCIA	-0.0757	-0.5999	<0.02*	—

Paleártico	RIQ	→	IGNORÂNCIA	-0.2327	-0.4473	<0***	—
Paleártico	TEMPO	→	IGNORÂNCIA	-0.0622	-0.0021	<0***	sim
Afrotropical	TEMPO	→	GDP	0.15	0.0016	<0***	—
Afrotropical	TEMPO	→	POP	0.0373	0,0001	<0***	—
Afrotropical	TEMPO	→	PAD	0.0818	0.0029	<0***	—
Afrotropical	GDP	→	RIQ	-0.0168	-0.0215	<0.004**	sim
Afrotropical	POP	→	RIQ	0.0637	0.3552	0.1413	—
Afrotropical	PAD	→	RIQ	-0.0223	-0.0088	0.4423	—
Afrotropical	TEMPO	→	RIQ	0.0653	0,0009	<0***	—
Afrotropical	GDP	→	IGNORÂNCIA	0.1657	0.1968	<0.0274*	—
Afrotropical	PAD	→	IGNORÂNCIA	0.116	0.0422	0.1122	—
Afrotropical	POP	→	IGNORÂNCIA	0.102	0.5285	0.2015	—
Afrotropical	RIQ	→	IGNORÂNCIA	-0.3692	-0.343	<0***	—
Afrotropical	TEMPO	→	IGNORÂNCIA	-0.1653	-0.0021	<0***	sim
Indo- Malaio	TEMPO	→	GDP	0.2009	0.0055	<0***	—
Indo- Malaio	TEMPO	→	POP	0.0222	0,0005	<0.001***	—
Indo- Malaio	TEMPO	→	PAD	0.0721	0.0017	<0.0028**	—
Indo- Malaio	GDP	→	RIQ	-0.0322	-0.0215	<0.004**	sim
Indo- Malaio	POP	→	RIQ	0.049	0.0374	0.6914	—
Indo- Malaio	PAD	→	RIQ	0.3085	0.2453	<0.0012**	—
Indo- Malaio	TEMPO	→	RIQ	0.143	0.0026	<0***	—
Indo- Malaio	GDP	→	IGNORÂNCIA	-0.0868	-0.0576	0.2522	—
Indo- Malaio	PAD	→	IGNORÂNCIA	-0.002	-0.0016	0.9782	—
Indo- Malaio	POP	→	IGNORÂNCIA	0.128	0.097	0.2051	—
Indo- Malaio	RIQ	→	IGNORÂNCIA	-0.2368	-0.2351	<0.0011**	—
Indo- Malaio	TEMPO	→	IGNORÂNCIA	-0.1157	-0.0021	<0***	sim
Australasia	TEMPO	→	GDP	0.3505	0.0085	<0***	—
Australasia	TEMPO	→	POP	0.0028	0	<0.0264*	—
Australasia	TEMPO	→	PAD	0.1457	0.0037	<0***	—
Australasia	GDP	→	RIQ	-0.0662	-0.0215	<0.004**	sim
Australasia	POP	→	RIQ	0.2148	12973	<0,0006***	—
Australasia	PAD	→	RIQ	0.0723	0.0223	0.1085	—

Australasia	TEMPO	→	RIQ	0.1955	0.0015	<0***	—
Australasia	GDP	→	IGNORÂ NCIA	-0.0084	-0.0042	0.8832	—
Australasia	PAD	→	IGNORÂ NCIA	-0.0068	-0.0033	0.8584	—
Australasia	POP	→	IGNORÂ NCIA	0.0924	0.8686	0.1941	—
Australasia	RIQ	→	IGNORÂ NCIA	-0.1417	-0.2207	<0***	—
Australasia	TEMPO	→	IGNORÂ NCIA	-0.1714	-0.0021	<0***	sim

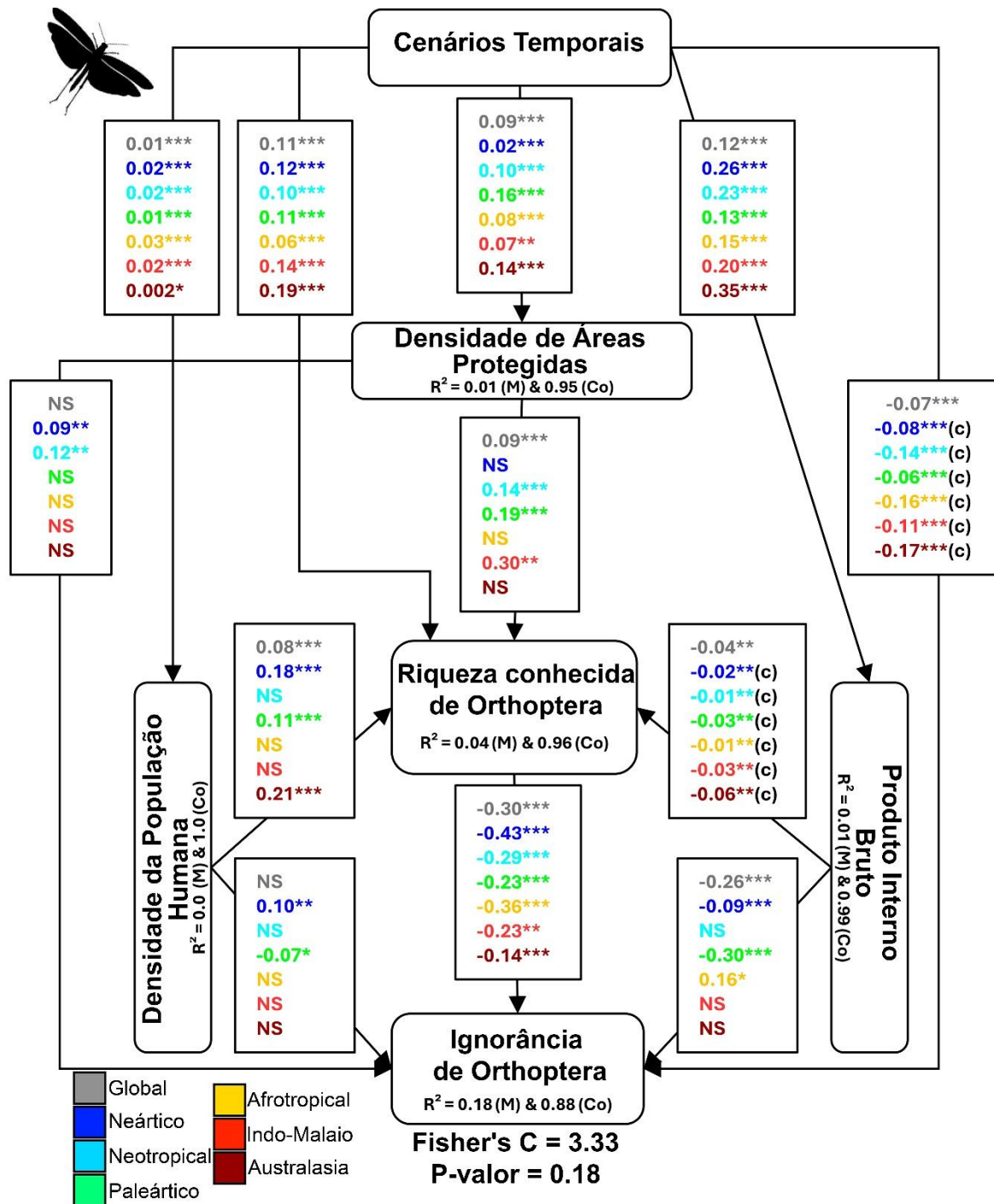


Figura S1. Modelos de Equações Estruturais (SEM) dos drivers preditores da ignorância da biodiversidade de Orthoptera. Os coeficientes com: *, ** ou *** representam caminhos significativos para o modelo(s), enquanto (NS) indicam caminho(s) não significativos ($p \geq 0,05$), (c) indicam que estimativa é fixa em todas as subdivisões biogeográficas para a relação. Os níveis de significância de cada preditor são * $p \leq 0,05$, ** $p \leq 0,01$, *** $p \leq 0,001$. (M) e (Co) representam o R² Marginal e Condicional, respectivamente. Resultados completos dos modelos estão na Tabela Suplementar 1.

Apêndice Suplementar 1. Efeitos indiretos mediados pela riqueza.

Considerando os efeitos indiretos dos cenários temporais que são mediados pela riqueza e densidade populacional humana (i.e., $\text{IGNORANCE} \leftarrow \text{RIQ} \leftarrow \text{POP} \leftarrow \text{TIME}$), observa-se uma redução da ignorância em cenários mais recentes para as subdivisões Global, Neártico, Paleártico e Australásia (Fig. S2a). De forma semelhante, em cenários temporais mais recentes, mediados pela riqueza e densidade de áreas protegidas (i.e., $\text{IGNORANCE} \leftarrow \text{RIQ} \leftarrow \text{UC} \leftarrow \text{TIME}$), essa redução ocorre nas subdivisões Global, Neotropical, Paleártico e Indo-Malaio (Fig. S2b). Por outro lado, em cenários mais recentes mediados pela riqueza e produto interno bruto (i.e., $\text{IGNORANCE} \leftarrow \text{GDP} \leftarrow \text{RIQ} \leftarrow \text{TIME}$), há um aumento da ignorância em todas as subdivisões (Fig. S2c).

Ao considerar o efeito indireto da densidade populacional humana sobre a ignorância, mediado pela riqueza de espécies (i.e., $\text{IGNORANCE} \leftarrow \text{RIQ} \leftarrow \text{POP}$), observa-se uma redução da ignorância com o aumento da densidade populacional humana nas subdivisões Global, Neártico, Paleártico e Australásia (Fig. S2d). De maneira semelhante essa redução da ignorância ocorre, para o efeito indireto do aumento da densidade de áreas protegidas mediado pela riqueza de espécies (i.e., $\text{IGNORANCE} \leftarrow \text{RIQ} \leftarrow \text{UC}$), nas subdivisões Global, Neotropical, Paleártico e Indo-Malaio (Fig. S2e). Já o efeito indireto do aumento do produto interno bruto sobre a ignorância, mediado pela riqueza (i.e., $\text{IGNORANCE} \leftarrow \text{RIQ} \leftarrow \text{GDP}$), resulta em um aumento da ignorância em todas as subdivisões (Fig. S2f).

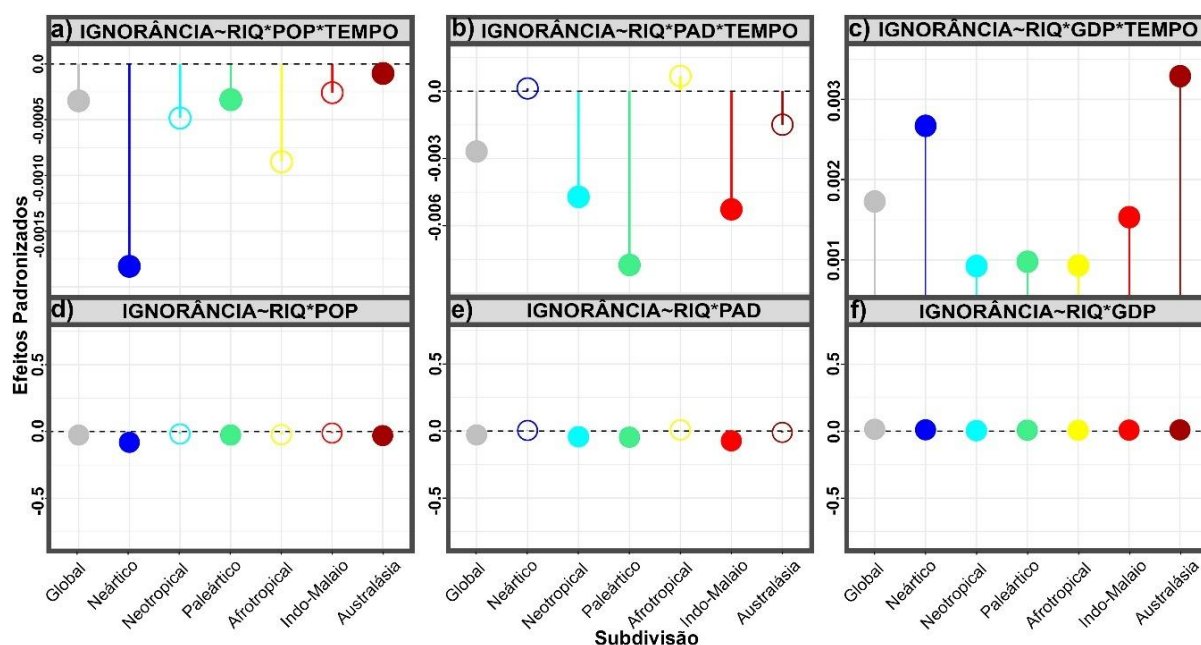


Figura S2. Efeitos padronizados do Modelo de Equações Estruturais (SEM) considerando as interações indiretas mediadas pela riqueza de espécies de Orthoptera. As cores representam as diferentes subdivisões biogeográficas investigadas: Cinza – Global; Azul – Neártico; Azul claro – Neotropical; Verde – Paleártico; Amarelo – Afrotropical; Vermelho – Indo-Malaio; Vermelho escuro – Australásia. Círculos preenchidos indicam efeitos significativos ($p < 0,05$). Os resultados completos do SEM estão disponíveis na Tabela e na Figura Suplementar 1.

CONCLUSÃO GERAL

Na introdução desta Tese, problematizamos de forma sistematizada as bases teóricas que fundamentam a ideia de desconhecimento da biodiversidade, abordando aspectos fundamentais como lacunas, limitações, vieses, déficits e ruídos. Essas questões destacam a importância da integração entre diferentes áreas, como taxonomia, ecologia e bioinformática, para alcançar um conhecimento mais sólido sobre a biodiversidade. Além disso, enfatizamos a necessidade de mapear esse desconhecimento nas pesquisas, por meio do uso de mapas de ignorância, uma abordagem que pode ser aplicada a diferentes grupos biológicos, bancos de dados e regiões geográficas.

No primeiro capítulo, utilizando o conhecimento digitalmente acessível sobre a ordem Orthoptera (Arthropoda, Insecta), evidenciamos que a taxonomia alfa (ou seja, a descrição de novas espécies) e a taxonomia beta (ou seja, a revisão do status taxonômico e das relações de táxons já descritos) se distribuem de forma heterogênea no espaço geográfico e no tempo. Nesse capítulo, detectamos que, nas regiões tropicais, a taxonomia de Orthoptera é recente e pouco revisada, enquanto nas regiões temperadas ela é mais antiga e passou por mais revisões. Esses padrões identificados nos ajudaram a destacar áreas pouco estudadas, que demandam maior atenção nos estudos taxonômicos. Além disso, discutimos como o estado da taxonomia de Orthoptera pode ser influenciado por fatores históricos, pela falta de financiamento para a documentação da biodiversidade, pela escassez de taxonomistas e por conflitos sociopolíticos.

No segundo capítulo, também utilizando o conhecimento digitalmente acessível, mapeamos a ignorância do conhecimento para ordem Orthoptera, integrando as deficiências taxonômicas, espaciais e temporais. Isso nos permitiu constatar que esse conhecimento ainda é limitado, não apenas no Sul Global, mas também em muitas regiões do Norte Global. Esses achados destacam a urgente necessidade de mobilização e disponibilização de registros primários ao redor do mundo, além da formação de novos taxonomistas, da criação de guias e chaves de identificação atualizadas, do financiamento de novas expedições, da digitalização e manutenção de instituições e coleções biológicas, e iniciativas de ciência cidadã.

No terceiro capítulo, através de modelos de equações estruturais, fomos capazes de integrar drivers socioecológicos, como densidade populacional humana, densidade de áreas protegidas, desenvolvimento socioeconômico e riqueza conhecida de espécies de Orthoptera, ao longo do tempo, para explicar a ignorância do conhecimento de Orthoptera em diferentes subdivisões biogeográficas. Isso nos permitiu constatar que tais mecanismos podem atuar de

forma diferente ao longo do globo, sendo necessário estratégias específicas para preencher as lacunas de conhecimento e promover a conservação da biodiversidade, mesmo sendo os efeitos dos drivers complexos e dependentes de escala.

Por fim, esta tese estabelece, de forma geral, uma nova abordagem integrativa entre a Taxonomia e a Macroecologia, além de introduzir, pela primeira vez, o uso aprimorado da Macroecologia no estudo da ordem Orthoptera. Apresentamos uma série de estruturas analíticas complexas que podem ser utilizadas para avaliar a qualidade e os processos associados ao conhecimento da biodiversidade. Além disso, os modelos aqui propostos podem ser replicados para outros grupos biológicos e contextos ecológicos de investigação.

PRODUTOS ADICIONAIS

Durante o período de doutorado (2021-2024), além da execução da tese, também tive a oportunidade de realizar outras contribuições científicas com diferentes colaboradores:

No Ano de 2021:



Zootaxa 4933 (1): 136–150

<https://www.mapress.com/j/zt/>

Copyright © 2021 Magnolia Press

Article

<https://doi.org/10.11646/zootaxa.4933.1.6>

<http://zoobank.org/urn:lsid:zoobank.org:pub:78BE572F-54FA-426C-A716-399EE6BE3706>

ISSN 1175-5326 (print edition)

ZOOTAXA

ISSN 1175-5334 (online edition)

The first record of *Adelosgryllus* Mesa & Zefa, 2004 (Orthoptera: Grylloidea: Phalangopsidae) from caves, with the description of a new species from Brazil

RAYANNE LAYS SANT'ANA MERLO^{1,2}, RODRIGO ANTÔNIO CASTRO-SOUZA^{1,3},
VITOR GABRIEL PEREIRA JUNTA^{1,4} & RODRIGO LOPES FERREIRA^{1*}

No Ano de 2022:



Zootaxa 5133 (1): 083–109

<https://www.mapress.com/zt/>

Copyright © 2022 Magnolia Press

Article

<https://doi.org/10.11646/zootaxa.5133.1.4>

<http://zoobank.org/urn:lsid:zoobank.org:pub:31F5A638-2922-41DC-BEA4-0988B49BE6C4>

ISSN 1175-5326 (print edition)

ZOOTAXA

ISSN 1175-5334 (online edition)

Expanding the taxonomic knowledge of *Adelosgryllus* Mesa & Zefa, 2004 (Orthoptera: Grylloidea: Phalangopsidae): description of four new species for Brazilian subterranean habitats

RAYANNE LAYS SANT'ANA MERLO¹, RODRIGO ANTÔNIO CASTRO-SOUZA², VITOR GABRIEL PEREIRA JUNTA¹ & RODRIGO LOPES FERREIRA^{1*}



Zootaxa 5094 (3): 409–434

<https://www.mapress.com/zt/>

Copyright © 2022 Magnolia Press

Article

<https://doi.org/10.11646/zootaxa.5094.3.3>

<http://zoobank.org/urn:lsid:zoobank.org:pub:50F1BC50-CEAA-491A-8E72-15F3E545BC49>

ISSN 1175-5326 (print edition)

ZOOTAXA

ISSN 1175-5334 (online edition)

A new species of *Arachnopsita* (Orthoptera: Grylloidea: Phalangopsidae) from caves in Guatemala

VITOR GABRIEL PEREIRA JUNTA¹, RODRIGO ANTÔNIO CASTRO-SOUZA² &
RODRIGO LOPES FERREIRA^{1*}

 Zootaxa 5209 (2): 211–237
<https://www.mapress.com/zt/>
 Copyright © 2022 Magnolia Press

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
 ISSN 1175-5334 (online edition)

<https://doi.org/10.11646/zootaxa.5209.2.4>
<http://zoobank.org/urn:lsid:zoobank.org:pub:BF50D677-6EBC-4447-AEFB-0A30F37EA275>

Singing crickets from Brazil (Orthoptera: Gryllidea), an illustrated checklist with access to the sounds produced

EDISON ZEFA^{1*}, LUCIANO DE PINHO MARTINS², CHRISTIAN PETER DEMARI^{1,12},
 RIULER CORRÊA ACOSTA³, ELLIOTT CENTENO⁴, RODRIGO ANTÔNIO CASTRO-SOUZA⁵,
 GABRIEL LOBREGAT DE OLIVEIRA⁶, AKIO RONALDO MIYOSHI⁷, MARCOS FIANCO⁸,
 DARLAN RUTZ REDÜ^{1,13}, VITOR FALCHI TIMM⁹,
 MARIA KÁTIA MATIOTTI DA COSTA¹⁰ & NEUCIR SZINWELSKI¹¹

 Zootaxa 5222 (1): 083–093
<https://www.mapress.com/zt/>
 Copyright © 2022 Magnolia Press

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
 ISSN 1175-5334 (online edition)

<https://doi.org/10.11646/zootaxa.5222.1.7>
<http://zoobank.org/urn:lsid:zoobank.org:pub:C1C9FE5A-865B-4B8E-A90D-55EDB321CEFD>

A new troglomorphic species of *Erebonyx* (Orthoptera: Grylloidea: Phalangopsidae) from Brazilian caves

RAYANNE LAYS SANT'ANA MERLO¹, RODRIGO ANTÔNIO CASTRO-SOUZA²,
 DIEGO DE MEDEIROS BENTO³ & RODRIGO LOPES FERREIRA^{1,4*}

No Ano de 2023:

Journal of Tropical Ecology

www.cambridge.org/tro

Persisting while changing over time: modelling the historical biogeographic of cave crickets (Orthoptera, Grylloidea) in Neotropics

Research Article

Cite this article: Castro-Souza RA, Sobral-Souza T, Rabelo LM, Zefa E, and Ferreira RL (2023). Persisting while changing over time: modelling the historical biogeographic of cave crickets (Orthoptera, Grylloidea) in Neotropics. *Journal of Tropical Ecology*. 39(e10), 1–11. <https://doi.org/10.1017/S0266467422000529>

Rodrigo Antônio Castro-Souza^{1,2} , Thadeu Sobral-Souza¹ ,
 Lucas Mendes Rabelo² , Edison Zefa³  and Rodrigo Lopes Ferreira² 

¹Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso, 78060-900, Cuiabá, MT, Brazil; ²Centro de Estudos em Biologia Subterrânea, Departamento de Ecologia e Conservação, Universidade Federal de Lavras, 37200-900, Lavras, MG, Brazil and ³Departamento de Ecologia, Zoologia e Genética, Universidade Federal de Pelotas, Instituto de Biologia, 96001-970, Capão do Leão, RS, Brazil

 Zootaxa 5263 (1): 001–039
<https://www.mapress.com/zt/>
 Copyright © 2023 Magnolia Press

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
 ISSN 1175-5334 (online edition)

<https://doi.org/10.11646/zootaxa.5263.1.1>
<http://zoobank.org/urn:lsid:zoobank.org:pub:3386FD59-2075-4F6A-8B21-B36C7F463EA9>

Three new cricket species and a new subgenus of *Endecous* Saussure, 1878 (Grylloidea: Phalangopsidae) from caves in northeastern Brazil

PEDRO HENRIQUE MENDES CARVALHO¹, VITOR GABRIEL PEREIRA JUNTA¹,
 RODRIGO ANTÔNIO CASTRO-SOUZA² & RODRIGO LOPES FERREIRA^{1*}

Two new species of *Endecous* Saussure, 1878 (Orthoptera: Grylloidea: Phalangopsidae) from caves in central-western Brazil, with the proposition of a new subgenus to accommodate *E. vitreus* Bolfarini & Campos, 2023

PEDRO HENRIQUE MENDES CARVALHO^{1,3}, RODRIGO ANTÔNIO CASTRO-SOUZA^{2,4} & RODRIGO LOPES FERREIRA^{1,*}

RESPONSE TO LAW N°. 11.861, OF AUGUST 03, 2022: CONSEQUENCES ON PANTANAL BIODIVERSITY

ABSTRACT: The Pantanal is the largest floodplain in the world and a hub of great biodiversity, but it faces constant threats from human activities. Ideally, public policy should be rooted in science, acknowledging and considering local nuances to enable and promote economic activities while safeguarding its diversity and functionality. Recently, law 11.861/2022 was proposed, which, while addressing certain environmental issues, allows for extensive livestock farming in the Pantanal without establishing adequate limits. Hence, this study aims to explore the consequences of certain management practices, as outlined in basic manuals of extensive livestock farming, on the biodiversity of key animal groups in the Pantanal. Through systematic searches on the Web of Science platform, based on global scientific literature, we assessed how the utilization of fertilizers, herbicides, pesticides, and fire impacts the biodiversity of various vertebrate and invertebrate groups that play crucial roles in the ecosystem: Birds, Fish, Mammals, Ants, and Orthoptera. Our findings indicate that all of these management methods could directly or indirectly lead to a decline in the biodiversity of these groups. Consequently, implementing the proposed law may alter the environment in ways that negatively affect several animal groups in the Pantanal Biome, resulting in environmental and economic repercussions (e.g., fishing, tourism, subsistence agriculture). Moreover, these conversions could also detrimentally impact existing traditional human populations and other economic activities that rely on biodiversity and ecosystem functions. Although some view this law as progress in meeting the demands of farmers and landowners, it may also pave the way for unregulated and disastrous practices that are already affecting the Pantanal.

Keywords: biodiversity loss; management; Web of Science.

¹ Estudante de doutorado do Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade (PPGECB) da Universidade Federal do Mato Grosso (UFMT); email: nicolas-bosco@hotmail.com; ledastefann@gmail.com; rafaelcuissi@gmail.com; rodrigodesouzaac@gmail.com; taina.souza1@gmail.com;

² Professor Adjunto na Universidade Federal de Mato Grosso (UFMT) e membro permanente do Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade (PPGECB); email: izzothiago@gmail.com

No Ano de 2024:



Zootaxa 5433 (3): 356–372

<https://www.mapress.com/zt/>

Copyright © 2024 Magnolia Press

Article

ISSN 1175-5326 (print edition)

ZOOTAXA

ISSN 1175-5334 (online edition)

<https://doi.org/10.11646/zootaxa.5433.3.4>

<http://zoobank.org/urn:lsid:zoobank.org:pub:149299C6-B272-46D1-8389-46342C2CAA6C>

Two new species of *Phaeophilacris* Walker, 1871 (Orthoptera: Grylloidea: Phalangopsidae) from caves in Kenya, with pictorial key for *Phaeophilacris* (*Taenacris*) Gorochoy, 2020

PEDRO HENRIQUE MENDES CARVALHO^{1,2}, RODRIGO ANTÔNIO CASTRO-SOUZA³, LABAN NJOROGÉ⁴, WANJA M. KINUTHIA⁴ & RODRIGO LOPES FERREIRA^{1,2*}



European Journal of Taxonomy 932: 82–111

<https://doi.org/10.5852/ejt.2024.932.2511>



ISSN 2118-9773

www.europeanjournaloftaxonomy.eu

2024 · Tavares *et al.*

This work is licensed under a Creative Commons Attribution License (CC BY 4.0).

Research article

urn:lsid:zoobank.org:pub:2597B29F-DF1C-44E0-92AC-7252E6C72E98

New species and records of *Zebragryllus* Desutter-Grandcolas & Cadena-Castañeda, 2014 (Orthoptera: Gryllidae: Gryllinae) from the Brazilian Amazon rainforest

Gustavo Costa TAVARES^{1,*}, Beatriz Harumi Kondo OYA²,
Oscar J. CADENA-CASTAÑEDA³, Marcus Paulo Alves de OLIVEIRA⁴ &
Rodrigo Antônio CASTRO-SOUZA⁵

VIESES NO CONHECIMENTO DA DISTRIBUIÇÃO GEOGRÁFICA DE CAVERNAS DO BRASIL

BIASES IN THE KNOWLEDGE OF THE GEOGRAPHICAL DISTRIBUTION OF CAVES IN BRAZIL

Rodrigo Antônio Castro-Souza

Universidade Federal de Mato Grosso Laboratório
de Macroecologia e Conservação da Biodiversidade
E-mail: rodrigodesouzaac@gmail.com

Nicolas Silva Bosco

Universidade Federal de Mato Grosso Laboratório
de Macroecologia e Conservação da Biodiversidade
E-mail: n.bosco.ambiente@gmail.com

Thadeu Sobral-Souza

Universidade Federal de Mato Grosso Laboratório
de Macroecologia e Conservação da Biodiversidade
E-mail: thadeusobral@gmail.com



Biodiversity Data Journal 12: e109785
doi: [10.3897/BDJ.12.e109785](https://doi.org/10.3897/BDJ.12.e109785)



Research Article

Amazonian amphibians: diversity, spatial distribution patterns, conservation and sampling deficits

Marcos Penhacek^{‡,§}, Thadeu Sobral Souza[‡], Jessie Pereira Santos[¶], Vinicius Guerra^{#,α}, Rodrigo Antônio Castro-Souza[‡], Domingos de Jesus Rodrigues^{‡,α}

