A can of worms: estimating the global number of earthworm species

Thibaud Decaëns[1]*, George G. Brown[2], Erin K. Cameron[3], Csaba Csuzdi[4], Nico Eisenhauer[5,6], Sylvain Gérard[7], Arnaud Goulpeau[1], Mickaël Hedde[7], Samuel W. James[8], Emmanuel Lapied[9], Marie-Eugénie Maggia[10], Daniel F. Marchán[11], Jérôme Mathieu[12], Helen R. P. Phillips[13], Eric Marcon[14]

[1] CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.

[2] EMBRAPA-Floresta, Curitiba, Paraná, Brazil.

[3] Saint Mary's University, Halifax, NS, Canada.

[4] Dept. of Zoology, Eszterházy Károly Catholic University, Eger, Hungary.

[5] German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Germany.

[6] Leipzig University, Leipzig, Germany

[7] Eco&Sols, INRAE, IRD, CIRAD, l'Institut Agro Montpellier, Univ Montpellier, Montpellier, France.

[8] Maharishi International University, Department of Regenerative Agriculture, Fairfield, IA, USA.

[9] Taxonomia Biodiversity Fund, 7 rue Beccaria, 72012, Paris, France.

[10] Department of Integrative Biology, University of Guelph, Ontario, Canada.

[11] Biodiversity, Ecology and Evolution Department, Faculty of Biology, Universidad Complutense de Madrid, Madrid, Spain.

[12] Sorbonne Université, CNRS, UPEC, INRAE, IRD, Institut d'Ecologie et des Sciences de l'Environnement de Paris, Paris, France.

[13] University of Helsinki, Helsinki, Finland

[14] AgroParisTech, UMR AMAP, CIRAD, CNRS, INRAE, IRD, Univ Montpellier, Montpellier, France.

* Corresponding author. Email: thibaud.decaens@cefe.cnrs.fr

Preprint server:

Abstract

Estimating the overall number of species for a given taxon is a central issue in ecology and conservation biology. It should be particularly topical in the case of soil organisms, which represent the majority of known species, but still suffer from a considerable taxonomic knowledge deficit. We propose here an estimation of the global number of earthworm species based on the Joppa approach, which models taxonomic effort over time in order to estimate the total number of known and as yet unknown species in a given taxa. Our Bayesian estimation of this model allows us to propose a global diversity of the order of 30,000 species. However, the uncertainty around this estimate is considerable due to severe undersampling and as the model cannot unambiguously decide whether we are describing few species because of a small pool of as yet unknown species, or because of a lack of taxonomic efficiency. This means that the 5,679 species and subspecies already described only represent around 20% of the global number of earthworm species. Considering the current rate of new species description, we calculate that it would take at least 120 years to describe all the species existing on Earth. We discuss the strategies that could be developed to reduce the magnitude of this taxonomic deficit.

Keywords

Soil biodiversity; global species estimates; taxonomic deficit; French Guiana; temperate France

1. Introduction

Estimating the total number of species of a given taxon, whether on a global scale or for a given region or ecosystem, is a question of great practical and symbolic importance that has long captivated and challenged scientists [1–4]. It represents an indispensable step towards a comprehensive description of the World's biodiversity and is also necessary for the definition of efficient conservation strategies [5,6]. In this context, soil biota represents one of the most diversified and least studied compartments in terrestrial ecosystems. They are among the most diverse, comprising up to 59% of currently known living species [7,8]. They are involved in regulating a large number of ecological processes that are essential to the provision of key ecosystem services for human populations [9–11]. This has been particularly well documented for emblematic groups of soil ecosystem engineers, such as earthworms [12,13]. Paradoxically, even these well-studied groups still suffer from a glaring lack of basic studies on their biodiversity patterns [14–16], and questions as fundamental as "how many species are there?" still remain largely unanswered [17].

Earthworms are an emblematic illustration of this paradox. Some 5,700 earthworm species have been described to date [18], but the proportion of the actual diversity of these organisms that this figure represents is still poorly documented. Estimates vary, depending on the author, from a minimum of 7,000 to some 20,000 species [8,12,19,20], but most of these figures are based on expert opinion, and no mathematical formalisation has yet been used to obtain a consolidated prediction of the number of as-yet undiscovered species. Some studies have also highlighted the existence of potentially high and as yet largely undescribed regional diversity throughout the tropics [15,21], which could challenge these previously proposed estimates of global species number.

Various methods can be used to estimate the number of heretofore unknown species within a given taxon. One popular approach is to use rarefaction and extrapolation curves, combined with non-parametric estimators, to predict the taxonomic deficit and the number of species at different spatial scales [22–26]. However, to be effective, this approach requires a sufficiently

dense dataset on a global scale, which is lacking for many soil organisms such as earthworms. Another method consists of using Arrhenius' power law to estimate the number of species for large areas on the basis of a known number of species in a smaller area [27–29]. The procedure involves estimating a power parameter (z) based on the negative linear relationship between the log of compositional similarity and the log of geographical distance between small equivalent areas. It therefore requires a network of plots located at very different distances, which must be supplemented by a continuous inventory over an area corresponding to the smallest scale of the power law. Apart from the fact that this type of data is rare, log transformation is impossible for pairs of plots with a compositional similarity of 0 (i.e. no species in common), which is common in the case of earthworms, particularly in the tropics [21]. Finally, Joppa *et al.* [5] propose to use global species checklists to model the rate at which taxonomists describe new species as a function of time, and to estimate the number of species that will have been reached when this rate will drop to zero due to the exhaustion of the pool of as yet unknown species. This approach has the advantage of not being dependent on the existence of comprehensive field datasets, but requires the establishment of reliable species checklists that can be challenged by taxonomic issues.

In this work, we propose a global estimate of the number of earthworm species based on the method of Joppa *et al.* [5] and the most recent global checklist of earthworms [18,30]. In parallel, we analysed three datasets obtained from field inventories for regions equivalent in surface area in tropical and temperate regions [21,31,32]. We used these data to quantify the extent of the taxonomic deficit and estimate the actual number of species for each region. We then used these regional estimates to discuss the range of results obtained from the Joppa method.

2. Material and methods

(a) Estimating the number of species from the global checklist

For the Joppa approach, we used the checklist recently published by Misirlioğlu et al. [18] and Brown et al. [30]. This list includes 5,738 species/subspecies (5,406 species and 332 unique subspecies, i.e. not counting nomino-typical subspecies) belonging to 23 families (including one non-crassiclitellate family: Moniligastridae). We chose to consider subspecies as taxa in their own right in the analysis, because the choice of whether to consider them as infraspecific divisions or to elevate them to species rank is a matter of taxonomists' points of view. For the sake of simplicity, we referred to all the taxa in the list as “species”. After deleting certain duplicates in the table, we obtained a consolidated list of 5,752 taxa names published between 1758 and 2023, containing for each of them the authors and the year of description. For each 5-year period, we then counted the number of unique species described and the total number of authors (i.e. all the co-authors of the species, after standardising the names of authors with accents and taking account of homonyms, wherever possible).

Joppa *et al.* [5] expected the number of species (S_i) described in a given interval ($Y_i = i$, numbered sequentially from 1) to depend on the taxonomic effort, i.e. the number of taxonomists (T_i) actively describing species during that period, and the unknown number of species remaining to be described, that is the total number of species (S_T) minus those already discovered ($\sum_{j=0}^{i} S_j$ where S_0 is the number of known species before Y_1). Over time, taxonomists were expected to become more efficient, i.e. the discovery rate of unknown species per taxonomist was assumed to increase linearly. The original model was thus

$$S_i \sim T_i(\beta_1 + \beta_2 Y_i)\left(S_T - \sum_{j=0}^{i} S_j\right) ,$$

where β_1 and β_2 are unknown parameters. The model was log-transformed and estimated with a normal error of variance σ^2. The parameters to be estimated were thus β_1, β_2, S_T and σ, the total number of species S_T being our main target.

The increase in efficiency accounted for in the Joppa model is not obvious as earthworms are concerned. Methodological advancements, such as the development of integrative taxonomy, are likely offset by the increased difficulty of sampling in remote areas, namely tropical rainforests. Moreover, contemporary scientists rarely focus exclusively on taxonomy, a discipline whose publications are generally undervalued, leaving them with less time to describe new species. Consequently, we studied an alternative model with no efficiency progress (hereafter referred to as the "no-progress" model), where β_2 equals zero.

We chose Bayesian inference to obtain the distribution of the estimates of the parameters which revealed to be as important as the most probable values. We chose non-informative priors, uniformly distributed between 0 and numerically possible values β_1, β_2 and σ. According to expert opinion, the prior of S_T was forced between 6,000, i.e. the number of already discovered species, and 100,000. The parameter space was explored by a Monte-Carlo Markov Chain with 10^7 steps, 10% considered as burn-in. The analysis began with time period 1822-1827, with a previously known number of species (S_0) equal to 1 (i.e. *Lumbricus terrestris* Linneaeus, 1758) [33]. The last period is 2018-2022; species published in 2023 and 2024 were excluded from the analysis.

(b) Estimating regional diversity and the taxonomic deficit

Compared to the checklist of angiosperms used by Joppa *et al.* [5], that of earthworms is very short: data is scarce. We endeavoured to supplement the global approach with three more exhaustive regional studies.

For the analyses of regional diversity, we used the data of Goulpeau *et al.* [21,34] in French Guiana (supplemented by a small number of as yet unpublished data), and, for Southern and Northern France, those of Bouché [31] updated for taxonomy by Gérard *et al*. [32]. The choice of these three regions allows us to consider contrasting situations in terms of earthworm diversity. French Guiana is considered to be a diversity hotspot, with a high proportion of species with restricted distribution and a large taxonomic deficit [21,35–37]. By comparison,

mainland France has a lower *a priori* species diversity, but the South is nevertheless characterised by a high proportion of endemic species, some of which continue to be discovered regularly, whereas the North is characterised by great compositional uniformity [31,38–40].

Three 300 x 400 km^2 zones, each subdivided into twelve 100 x 100 km^2 cells, were delimited in each region. In each 100 x 100 km^2 cell, we counted the species present in order to obtain three occurrence tables comprising 12 rows and as many columns as species present. In French Guiana, we used the operational taxonomic units (OTUs) proposed by Goulpeau *et al.* [34] as a substitute for species in order to overcome the problem posed by the taxonomic deficit. For each region separately, we plotted incidence-based rarefaction and extrapolation curves to represent the accumulation of species numbers as a function of the number of 100 x 100 km^2 cells sampled in the study area. The 95% confidence intervals of the interpolated and extrapolated richness values were obtained by a bootstrap procedure with 200 replications. We also calculated the non-parametric Chao 2 index, which estimates the lower limit of expected asymptotic species richness [41]. The taxonomic deficit, i.e. the proportion of undiscovered species, was calculated using the formula:

$$\mathrm{TDef} = ((S_{EST} - S_{OBS}) \times 100/S_{EST} \ ,$$

with $TDef$ the taxonomic deficit, S_{OBS} the number of species currently known for a given region, and S_{EST} the richness estimated by the Chao 2 index.

For French Guiana, the taxonomic deficit was calculated by first considering S_{OBS} as the set of OTUs delimited by Goulpeau *et al.* [34], and then by restricting S_{OBS} to the number of species formally described to date (i.e. 55 species taking into account the 18 new descriptions by Decaëns *et al.* [36]. These analyses were performed using the *'iNEXT'* package for R v.4.4.1 [42,43].

3. Results

During the period from 1822 to 2022, the number of species described per 5-year period showed two peaks: the first from the end of the 19th to the middle of the 20th century, the other from the 1970s to the present day (Fig. 1A). Over the same period, the total number of authors of original species descriptions has increased exponentially (i.e. roughly linearly on a logarithmic scale; Fig. 1A). The ratio of numbers of species described per taxonomist thus shows a very marked peak in the first half of the 20th century, a period during which it generally exceeds 15 species described per taxonomist per time period, whereas everywhere else it is generally below 10 species per taxonomist (Fig. 1B).

The Joppa model yields the median posterior S_T equal to 6,814 species, with a 95% higher posterior density interval (HPDI) between 6,000 and 11,148. The no-progress model leads to quite different estimations: 26,283 species (HDPI between 6,003 and 87,425). Posterior probabilities are shown in Fig. 2.

The rarefaction curves obtained for the three study regions show a decrease in beta diversity from tropical to temperate study regions (Fig. 3). In French Guiana, the slope of the curve is steep, showing that the asymptote is far from being reached, whereas it has been or is about to be reached in the two study regions in temperate France. The diversity metrics logically follow the same trend: the cumulative number of species observed is 305 in French Guiana, 66 in Southern France, and 31 in Northern France, whereas the Chao 2 index predicts a diversity of 1,680 species in French Guiana, 99 in Southern France, and 36 in Northern France (Fig. 4A, 4B). Consequently, the taxonomic deficit is of the order of 80% in French Guiana, 33% in the south of France, and 16% in the north of France (Fig. 4C). In French Guiana, this 80% taxonomic deficit was obtained by considering as observed richness all of the OTUs mentioned in Goulpeau *et al.* [34], most of which are new to science and have not yet been named. If we consider only the 55 named species actually known from French Guiana (Fig. 4A), the estimated taxonomic deficit rises to 97% (Fig. 4C).

4. Discussion

Earthworm taxonomy experienced a golden age from the end of the 19th century to the middle of the 20th century, followed by a revival from the end of the 1970s. The first of these two periods can be credited to the great explorations that enabled zoologists of the time to explore the world and describe its biological diversity [44]. It was also made possible by the exceptional productivity of the few taxonomists of the time, who were able to describe dozens of species in a short time. By comparison, the contemporary period is characterised by a larger community of taxonomists, but with comparatively lower output. Overall, we do not observe the monotonic decline in the number of species described per taxonomist as described for plants and other animal taxa [5,45,46], which is assumed to reflect a depletion of the pool of new species to be discovered. Rather, with the exception of the notable peak in the first half of the 20th century, the rate of description of new earthworm species remained modest and constant over time, at around 5 species per taxonomist per 5-year period. Even in the early 2000s, we did not observe any clear effect from the development of molecular tools such as DNA barcoding and metabarcoding (*sensu* [47,48]). This may seem surprising, given that their usefulness in the field of earthworm taxonomy has been widely acknowledged [49–51], and their advent has been touted as the possible catalyst of a new golden age of taxonomy [52–54]. Over the last few decades, the increase in the number of new earthworm species described per time period seems to be simply correlated with the number of active taxonomists.

Our estimates of the Joppa and the no-progress model tell us two different stories. Assuming a constant progress of the efficiency of taxonomists, the global number of species would be below 7,000, i.e. around 1,000 species remaining to be described. On the other hand, assuming constant efficiency, the global number of species would rather be around 30,000, i.e. the 5,679 species of our checklist are only the beginning of the story.

In both cases, the confidence intervals around our median estimates are considerable, motivating us not to provide a single significant figure in our estimates. This can be explained

by two main causes of uncertainty. First, we lack data: more publications of new species would have allowed smaller, i.e. more, time intervals, thus more statistical power. Also, the Joppa model does not allow to completely disentangle the effects of the efficiency and the number of species: in the no-progress model, the estimates are highly correlated (Pearson correlation of the posteriors: -0,73), and the correlation is -0,55 between β_2 and S_T in the original model. This issue could not be detected by Joppa *et al.* [5] who relied on a frequentist estimation of the most likely number of species.

The analysis of regional diversity in French Guiana points in the direction of a high number of species: we estimate that the taxonomic deficit there is of the order of 97%, if only currently named species are considered in the calculation. An unknown thousand species may well all be there: a small area (actually around 1%) of Amazonia may thus be enough to find the species missing from the Joppa model. If we consider that around half of the earthworm species currently known are tropical (*i.e.* around 2,800 species), the generalisation of such a taxonomic deficit would lead us to estimate that more than 90,000 species could exist throughout the tropics. Of course, such a generalisation is highly debatable, but it does provide an argument for considering that the global number of unknown earthworm species may not be only a few thousands.

As expected, we calculate a lower taxonomic deficit in temperate regions, which have been better studied and host a more modest diversity. However, our results indicate that this deficit can still be significant—ranging from 15 to 30% in mainland France—suggesting that even in temperate regions, a substantial number of species remains undiscovered. The South of France, known for its high diversity and richness in endemic earthworms, exemplifies this, as evidenced by the recent descriptions of several new species, including some of large body size [40]. The persistence of a significant taxonomic deficit in Northern France is more unexpected. It may reflect a gap in our understanding of species distribution in the region (i.e. the Wallacean shortfall according to [55]), or the persistence of taxonomic confusion that can conceal the presence of certain species. A striking example of this is the discovery of the

existence of two cryptic species that were mistakenly considered synonymous under the name *L. terrestris* [56].

These considerations allow us to propose a conservative estimate of the global number of earthworm species of the order of 30,000 species. This is above the upper range of commonly accepted predictions, i.e. from 10,000 to 20,000 species [7,8,19,20]. Our expert opinion is that there is still a large pool of undiscovered species, but that we lack the efficiency to discover them or describe them once they have been discovered. First, a low rate of new species description per taxonomist, compared with those active at the beginning of the 20th century, is expected due the fact that taxonomy is no longer considered a 'competitive' discipline in modern science [57,58]. The pernicious effect of this situation is that contemporary authors are rarely pure taxonomists, as they are encouraged by the academic system to devote most of their research time to more 'valuable' areas of research, such as evolution or ecology. Added to this is the fact that taxonomists generally lack the time and financial resources for field sampling, particularly in the most diverse and yet least well-studied tropical regions [59]. The field scientists among us all have in mind regions that they know contain a diversity of as yet unknown species, but which they do not have the time or the resources to explore. Similarly, taxonomists have a multitude of species in their cupboards waiting to be described, but which they do not have the time to process.

Overall, our results suggest that at best we have only described one fifth of the global number of earthworm species. If we consider the current community of taxonomists (an average of 45 active taxonomists per 5-year period over the last 20 years), and the number of species they have managed to describe (an average of 4.5 species per taxonomist per time step), we calculate that it would take at least 120 years to describe all the species existing on Earth. Obviously, such an objective is certainly unattainable within a reasonable timeframe, but we can nonetheless think about the strategies that could be developed to reduce the magnitude of the taxonomic deficit. This will require maintaining a major taxonomic effort, through the training of young taxonomists, but also by raising the profile of taxonomic work in the scientific

corpus. Agnarsson & Kuntner [58] have proposed that original descriptions, and their underlying taxonomic hypotheses, should be formally cited in articles that refer to them – which is not systematically the case at present. This would allow the impact metrics of taxonomic journals to improve while more accurately reflecting their real scientific impact, thus encouraging scientists to invest more in taxonomy. We will also need to improve taxonomic efficiency, first by increasing our ability to discover as yet unknown species where they occur. This means stepping up sampling efforts, particularly in the tropics where most of this undiscovered diversity is to be found. Our results confirm the existence of a regional peak in diversity at tropical latitudes, essentially explained by a strong beta component [15,21], while emphasising that it mirrors a considerable taxonomic deficit. Finally, once sampling coverage has been improved, new taxonomic approaches based on turbotaxonomy (*sensu* [60,61]) and integrative taxonomy (*sensu* [62]) will be necessary to more rapidly describe new species thanks to automated procedures and the use of molecular tools [61,63].

In conclusion, we propose an updated estimate of the global number of earthworm species, as well as arguments for considering earthworms as a potentially hyperdiverse group of organisms, which has hitherto not been considered as such due to a strong Linnean shortfall (*sensu* [55]). This is particularly remarkable when considering that earthworms are among the most widely publicised soil macro-organisms, due to their acknowledged importance in soil functioning. It is likely that this situation could also apply to other less well-studied soil taxa, leading to a considerable reassessment of the contribution of these organisms to global biodiversity.

Acknowledgments and funding sources

The French Guiana dataset was acquired as part of different projects: DIADEMA (Dissecting Amazonian Diversity by Improving a Multiple Taxonomic Group Approach), DIAMOND (Dissecting and Monitoring Amazonian Diversity) and Wormbank (DNA barcoding earthworms in biodiversity hot spots of French Guiana) projects funded by "Investissement d'Avenir" grants managed by the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP:

ANR-10-LABX-41); "Our Planet Reviewed" French Guiana 2015 expedition [64] organised by the Muséum national d'Histoire naturelle (MNHN, Paris) and Pro-Natura international in collaboration with the Amazonian Park of French Guiana, and financed by the European Fund for Regional Development (FEDER), the Regional Council of French Guiana, the General Council of French Guiana, the Direction de l'Environnement, de l'Aménagement et du Logement and by the Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. Additional support was received from the Réserve naturelle des Nouragues (Nouragues grants 2010 and 2011), the Réserve naturelle de la Trinité and the Parc Amazonien de Guyane.

The dataset from temperate France was obtained through a project called GloWorms, mostly funded by the 'AgroEcoSystems' Scientific Department of INRAE (National Institute for Research in Agriculture, Environment and Food, Paris, France). Samplings in Normandy were funded by the INPN (National Inventory of Natural Heritage, Paris, France). Sylvain Gérard's PhD thesis was funded by a thesis grant from the Ecole Normale Supérieure (Paris, France). Daniel F. Marchán's contribution was funded by the Make Our Planet Great Again programme (Campus France, Paris), via a postdoctoral grant.

JM and GGB acknowledge the FaunaServices group jointly funded by the synthesis center CESAB of the French Foundation for Research on Biodiversity (FRB; www.fondationbiodiversite.fr)", CEBA, Sinbiose, Fapesp and CNPq (Brazil). NE gratefully acknowledges the support of iDiv, which is funded by the German Research Foundation (DFG – FZT 118, 202548816), as well as by the DFG (Ei 862/29-1, Ei 862/31-1). HRPP was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101033214 (GloSoilBio). GGB acknowledges project support by the Brazilian National Council for Scientific and Technological Development (CNPq Nos. 312824/2022-0, 441930/2020-4 and 404191/2019-3).

Funding

References

1. Dirzo R, Raven P. 2003 Global state of biodiversity and loss. *Annu Rev Environ Resour* **28**, 137–167.

2. May RM. 1992 How many species inhabit the earth? *Scientific American* **267**, 18–24.

3. Pimm SL. 2001 *The World according to Pimm: A Scientist Audits the Earth*. New York, USA: McGraw-Hill.

4. Stork N. 1993 How many species are there? *Biodiversity and Conservation* **2**, 215–232.

5. Joppa LN, Roberts DL, Pimm SL. 2011 How many species of flowering plants are there? *Proceedings of the Royal Society B* **278**, 554–559.

6. Caley MJ, Fisher R, Mengersen K. 2014 Global species richness estimates have not converged. *Trends in Ecology & Evolution* **29**, 187–188.

7. Decaëns T, Jiménez JJ, Gioia C, Measey J, Lavelle P. 2006 The values of soil animals for conservation biology. *European Journal of Soil Biology* **42**, S23–S38.

8. Anthony MA, Bender SF, van der Heijden MGA. 2023 Enumerating soil biodiversity. *Proceedings of the National Academy of Sciences of the United States of America* **120**, e2304663120.

9. Brussaard L *et al.* 1997 Biodiversity and ecosystem functioning in soil. *AMBIO* **26**, 563–570.

10. Wardle DA, Bardgett RD, Klironomos JN, Setälä H, van der Putten WH, Wall DH. 2004 Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629–1633. (doi:10.1126/science.1094875)

11. Bardgett RD, van der Putten WH. 2014 Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511. (doi:10.1038/nature13855)

12. Lavelle P, Decaens T, Aubert M, Barot S, Blouin M, Bureau F, Margerie P, Mora P, Rossi J-P. 2006 Soil invertebrates and ecosystem services. *European Journal of Soil Biology* **42**, S3–S15. (doi:10.1016/j.ejsobi.2006.10.002)

13. Lavelle P *et al.* 2016 Ecosystem engineers in a self-organized soil: A review of concepts and future research questions. *Soil Science* **18**, 91–109.

14. Cameron EK *et al.* 2018 Global gaps in soil biodiversity data. *Nature Ecology & Evolution* **2**, 1042–1043. (doi:10.1038/s41559-018-0573-8)

15. Phillips HRP *et al.* 2019 Global distribution of earthworm diversity. *Science* **366**, 480–485. (doi:10.1126/science.aax4851)

16. Phillips HRP *et al.* 2021 Global data on earthworm abundance, biomass, diversity and corresponding environmental properties. *Scientific Data* **8**, 136. (doi:10.1038/s41597-021-00912-z)

17. Decaëns T. 2010 Macroecological patterns in soil communities. *Global Ecology and Biogeography* **19**, 287–302.

18. Misirlioğlu M *et al.* 2023 Earthworms (Clitellata, Megadrili) of the world: an updated checklist of valid species and families, with notes on their distribution. *Zootaxa* **5255**, 417–438.

19. Reynolds JW. 1994 Earthworms of the world. *Global biodiversity* **4**, 11–16.

20. Blakemore RJ. 2009 Chapter 14. Cosmopolitan earthworms – a global and historical perspective. In *Annelids as model systems in the biological sciences* (ed DH Shain), pp. 257–283. New York: John Wiley & Sons, Inc.

21. Goulpeau A, Hedde M, Maggia M-E, Gérard S, Lapied E, Decaëns T. 2024 A dataset on earthworm communities in French Guiana. *Soil Organisms* **92**, 79–90.

22. Chao A. 1984 Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**, 265–270.

23. Colwell RK, Mao CX, Chang J. 2004 Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**, 2717–2727. (doi:10.1890/03-0557)

24. Chao A, Chazdon RL, Colwell RK, Shen TJ. 2005 A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* **8**, 148–159. (doi:10.1111/j.1461-0248.2004.00707.x)

25. Chao A, Jost L. 2012 Diversity measures. In *Encyclopedia of Theoretical Ecology* (eds A Hastings, L Gross), pp. 203–207. Berkeley, USA: University of California Press.

26. Cazzolla-Gatti R *et al.* 2022 The number of tree species on Earth. *Proceedings of the National Academy of Sciences of the United States of America* **119**, e2115329119. (doi:10.1073/pnas.2115329119)

27. Arrhenius O. 1921 Species and area. *Journal of Ecology* **9**, 95–99. (doi:10.2307/2255763)

28. Krishnamani R, Kumar A, Harte J. 2004 Estimating species richness at large spatial scales using data from small discrete plots. *Ecography* **27**, 637–642. (doi:10.1111/j.0906-7590.2004.03790.x)

29. Marcon E, Mirabel A, Molino J-F, Sabatier D. 2024 Estimation of the number of tree species in French Guiana by extrapolation of permanent plots richness. *Journal of Tropical Ecology* **in press**.

30. Brown GG *et al.* 2023 A checklist of megadrile earthworm (Annelida: Clitellata) species and subspecies of the world. *Zedono* **https://doi.org/10.5281/zenodo.8348860**. (doi:https://doi.org/10.5281/zenodo.8348860)

31. Bouché MB. 1972 *Lombriciens de France*. Paris: INRA.

32. Gérard S *et al.* 2024 #Vers2022: a dataset of species composition and individual mass of earthworm communities resampled 50 years after Bouché's historical work. *Ecology* **in press**.

33. Colbourne JK, Crease TJ, Weider LJ, Hebert PDN, Duferesne F, Hobaek A. 1998 Phylogenetics and evolution of a circumarctic species complex (Cladocera: Daphnia pulex). *Biological Journal of the Linnean Society* **65**, 347–365. (doi:10.1111/j.1095-8312.1998.tb01146.x)

34. Goulpeau A, Penel B, Maggia M-E, Marchán DF, Steinke D, Hedde M, Decaëns T. 2022 OTU delimitation with earthworm DNA barcodes: a comparison of methods. *Diversity* **14**, 866.

35. Decaëns T *et al.* 2016 DNA barcoding reveals diversity patterns of earthworm communities in remote tropical forests of French Guiana. *Soil Biology & Biochemistry* **92**, 171–183. (doi:10.1016/j.soilbio.2015.10.009)

36. Decaëns T, Bartz MLC, Goulpeau A, Marchan DF, Maggia M-E, Lapied E, Feijoo AM, James SW. 2024 Earthworms (Oligochaeta: Clitellata) of the Mitaraka range (French Guiana): commented checklist with description of one genus and eighteen species new to science. *Zoosystema* **46**, 195–244.

37. Maggia M-E *et al.* 2021 At each site its diversity: DNA barcoding reveals remarkable earthworm diversity in neotropical rainforests of French Guiana. *Applied Soil Ecology* **164**, 103932.

38. Mathieu J, Davies TJ. 2014 Glaciation as an historical filter of below-ground biodiversity. *Journal of Biogeography* **41**, 1204–1214. (doi:10.1111/jbi.12284)

39. Gérard S, Marchán DF, Martinez Navarro A, Hedde M, Decaëns T. 2023 Resampling Bouché's historical localities reveals three new species and helps identifying a new genus of earthworms (Oligochaeta, Hormogastridae and Lumbricidae) in Southeastern France. *Zoosystema* **45**, 749–768.

40. Marchán DF, Novo M, Domínguez J, da Silva A, Martínez A, Decaëns T. 2023 The best is yet to come: six new species within a large-bodied earthworm genus (Scherotheca, Lumbricidae) in a densely sampled region (Southwestern France). *Zoological Journal of the Linnean Society* **20**, 1–19.

41. Gotelli N, Chao A. 2013 Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In *Encyclopedia of Biodiversity* (ed SA Levin), pp. 195–211. Waltham, MA: Academic Press.

42. Hsieh TC, Ma KH, Chao A. 2019 iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.19.

43. R Core Team. 2024 R: a language and environment for statistical computing. Version 4.4.1.

44. Csuzdi C, Szlávecz K. 2016 Earthworm (Clitellata: Megadrili) taxonomy in the last 200 years: A homage to András Zicsi (1928−2015). *Opuscula Zoologica (Budapest)* **47**, 1–7.

45. Pimm SL, Jenkins CN, Joppa LN, Roberts DL, Russell GJ. 2010 How Many Endangered Species Remain to be Discovered in Brazil? *Nat. Conserv.* **08**, 71–77. (doi:10.4322/natcon.00801011)

46. Fisher MA, Vinson JE, Gittleman JL, Drake JM. 2018 The description and number of undiscovered mammal species. *Ecology and Evolution* **8**, 3628–3635.

47. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003 Biological identifications through DNA barcodes. *Proceedings of the Royal Society B-Biological Sciences* **270**, 313–321. (doi:10.1098/rspb.2002.2218)

48. Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012 Environmental DNA. *Molecular Ecology* **21**, 1789–1793. (doi:10.1111/j.1365-294X.2012.05542.x)

49. Rougerie R *et al.* 2009 DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira* **44**, 789–802. (doi:10.1590/S0100-204X2009000800002)

50. Decaëns T, Porco D, Rougerie R, Brown GG, James SW. 2013 Potential of DNA barcoding for earthworm research in taxonomy and ecology. *Applied Soil Ecology* **65**, 35–42. (doi:10.1016/j.apsoil.2013.01.001)

51. Marchán DF, Decaëns T, Domínguez J, Novo M. 2022 Perspectives in earthworm phylogeny: recent advances in Lumbricoidea and standing questions. *Diversity* **19**, 22.

52. Hebert PDN, Gregory TR. 2005 The promise of DNA barcoding for taxonomy. *Syestematic Biology* **54**, 852–859. (doi:10.1080/10635150500354886)

53. Hajibabaei M. 2012 The golden age of DNA metasystematics. *Trends in Genetics* **28**, 535–537. (doi:10.1016/j.tig.2012.08.001)

54. Pennisi E. 2019 DNA barcodes jump-start search for new species. *Science* **364**, 920–921. (doi:10.1126/science.364.6444.920)

55. Lomolino MV. 2004 Conservation biogeography. In *Frontiers of biogeography: new directions in the geography of nature* (eds MV Lomolino, LR Heaney), pp. 293–296. Sunderland, MA.

56. James SW, Porco D, Decaens T, Richard B, Rougerie R, Erseus C. 2010 DNA Barcoding Reveals Cryptic Diversity in Lumbricus terrestris L., 1758 (Clitellata): Resurrection of L. herculeus (Savigny, 1826). *Plos One* **5**, e15629. (doi:10.1371/journal.pone.0015629)

57. Wilson EO. 2004 Taxonomy as a fundamental discipline. *Philosophical Transactions of the Royal Society B - Biological Sciences* **359**, 739–739. (doi:10.1098/rstb.2003.1440)

58. Agnarsson I, Kuntner M. 2007 Taxonomy in a changing world: Seeking solutions for a science in crisis. *Systematic Biology* **56**, 531–539. (doi:10.1080/10635150701424546)

59. Guerra CA *et al.* 2020 Blind spots in global soil biodiversity and ecosystem function research. *Nature Communications* **11**, 3870. (doi:10.1038/s41467-020-17688-2)

60. Butcher BA, Smith MA, Sharkey MJ, Quicke DLJ. 2012 A turbo-taxonomic study of Thai Aleiodes (Aleiodes) and Aleiodes (Arcaleiodes) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa* , 1–232.

61. Fernandez-Triana JL. 2022 Turbo taxonomy approaches: lessons from the past and recommendations for the future based on the experience with Braconidae

(Hymenoptera) parasitoid wasps. *Zootaxa* **1087**. (doi:https://doi.org/10.3897/zookeys.1087.76720)

62. Dayrat. 2005 Towards integrative taxonomy. *Biological Journal of the Linnean Society* **85**, 407–417. (doi:https://doi.org/10.1111/j.1095-8312.2005.00503.x)

63. Padial PJ, Miralles A, de la Riva I, Vences M. 2010 The integrative future of taxonomy. *Frontiers in Zoology* **7**, 16.

64. Touroult J, Pollet M, Pascal O. 2018 Overview of Mitaraka survey: research frame, study site and field protocols. *Zoosystema* **40**, 327–365. (doi:10.5252/zoosystema2018v40a13)

Figure legends

Figure 1. Fluctuations over time, for each 5-year period from 1822 to 2017, in the numbers of described earthworm species and taxonomists (A), and in the number of species described per taxonomist (B). The loess models fitted to the data are shown in colour, with the grey area corresponding to the 95% confidence interval.

Figure 2. Results of Bayesian estimation of the global number of earthworm species according to the original Joppa model (A) and the no-progress model assuming the taxonomists' efficiency constant (B). The diagram shows the probability density of the estimated ST value. The red line represents the median and the dashed lines are the limits of the highest posterior density interval.

Figure 3. Incidence-based rarefaction curve showing cumulative numbers of earthworm species as a function of sampling effort (number of 100 x 100 km2 windows) in the three regions studied. Solid lines represent rarefaction curves (interpolated values); dashed lines represent extrapolation curves; shaded areas represent 95% confidence intervals obtained with a bootstrap of 200 replications.

Figure 4. Analysis of earthworm diversity in the three study regions: A) cumulative number of observed species , B) estimated number of species (Chao 2 index) and C) taxonomic deficit (i.e. the proportion of estimated diversity that has not been observed). In A, the grey hatched part corresponds to species from French Guiana that have already been named, while the non-hatched part of the bar corresponds to species that have been discovered but not yet described; in B, the error bars represent the 95% confidence interval obtained using a bootstrap with 200 replications; in C, the hatched part corresponds to the inflation of the taxonomic deficit in French Guiana as a result of taking into account only species that have been already named, the non-hatched part of the bar corresponding to the deficit calculated by also taking into account species that have been discovered but not yet described.

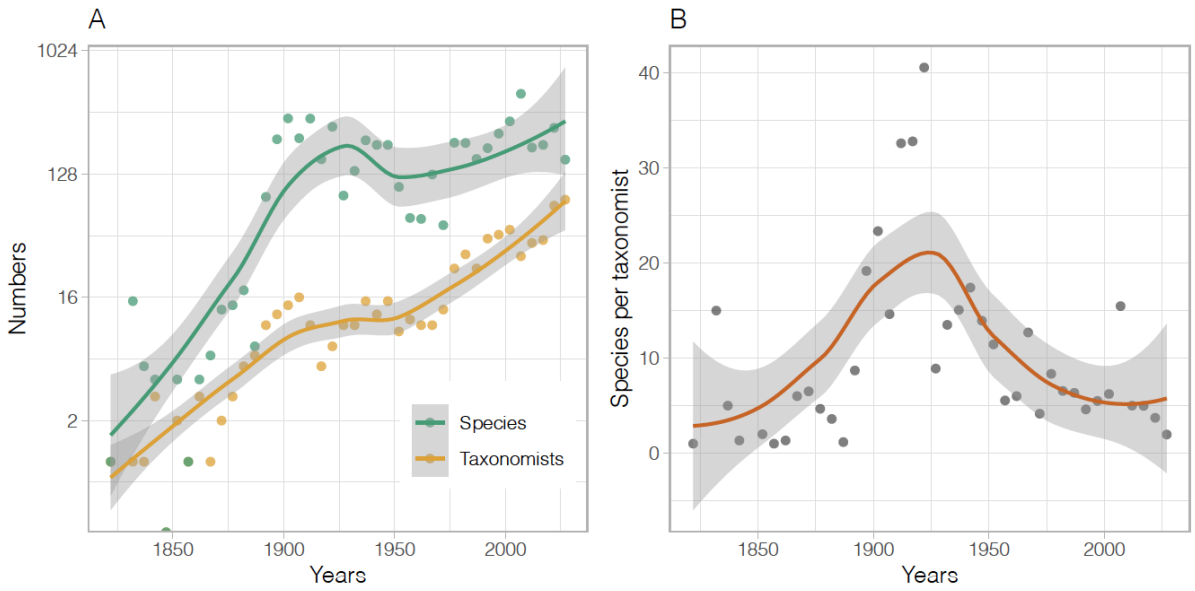
A
1024
128
16
2
Numbers
1850
1900
1950
2000
Years
Species
Taxonomists
B
40
30
20
10
0
Species per taxonomist
1850
1900
1950
2000
Years

Figure 1.

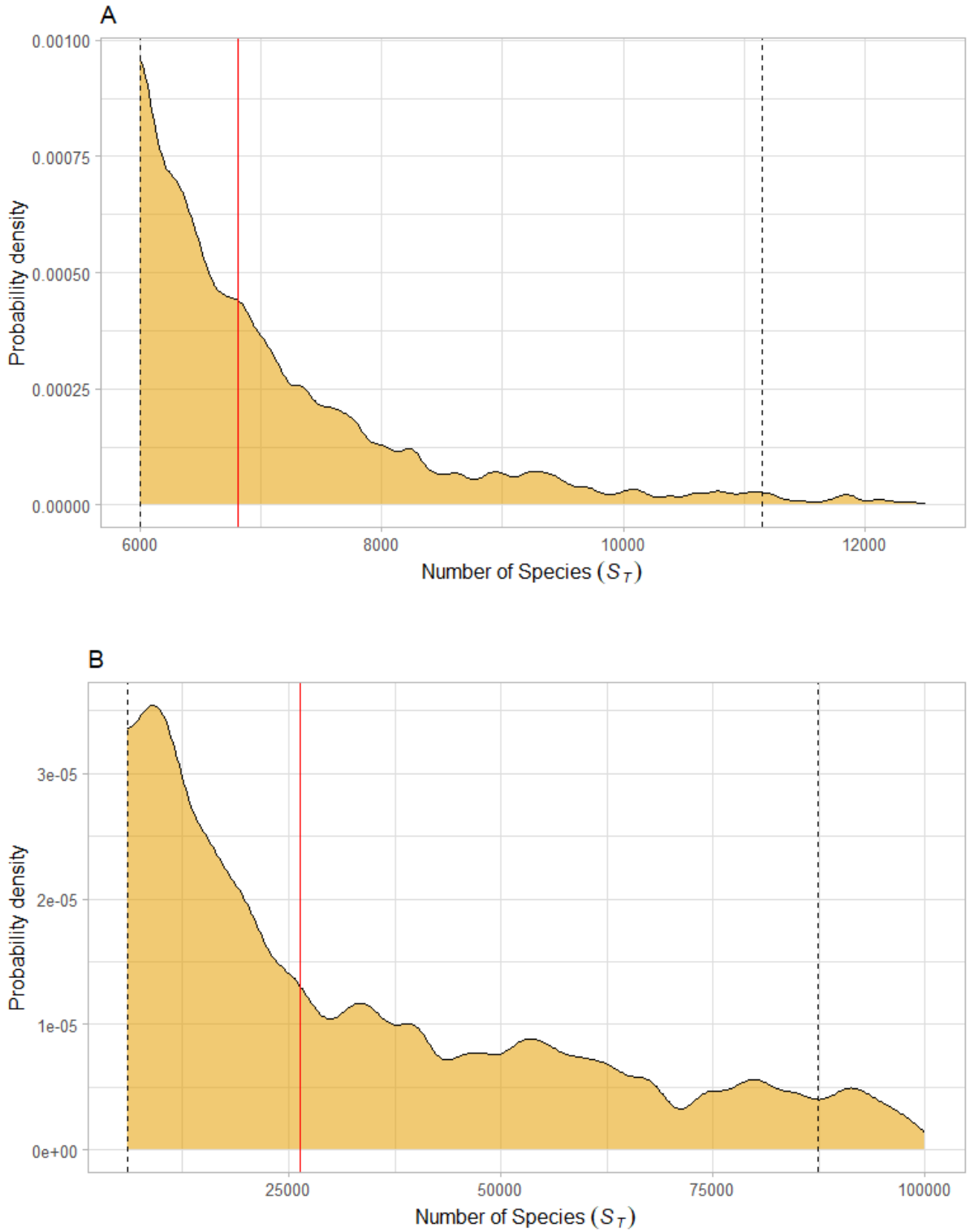
A
0.00100
0.00075
0.00050
0.00025
0.00000
Probability density
6000
8000
10000
12000
Number of Species (S_T)
B
3e-05
2e-05
1e-05
0e+00
Probability density
25000
50000
75000
100000
Number of Species (S_T)

Figure 2.

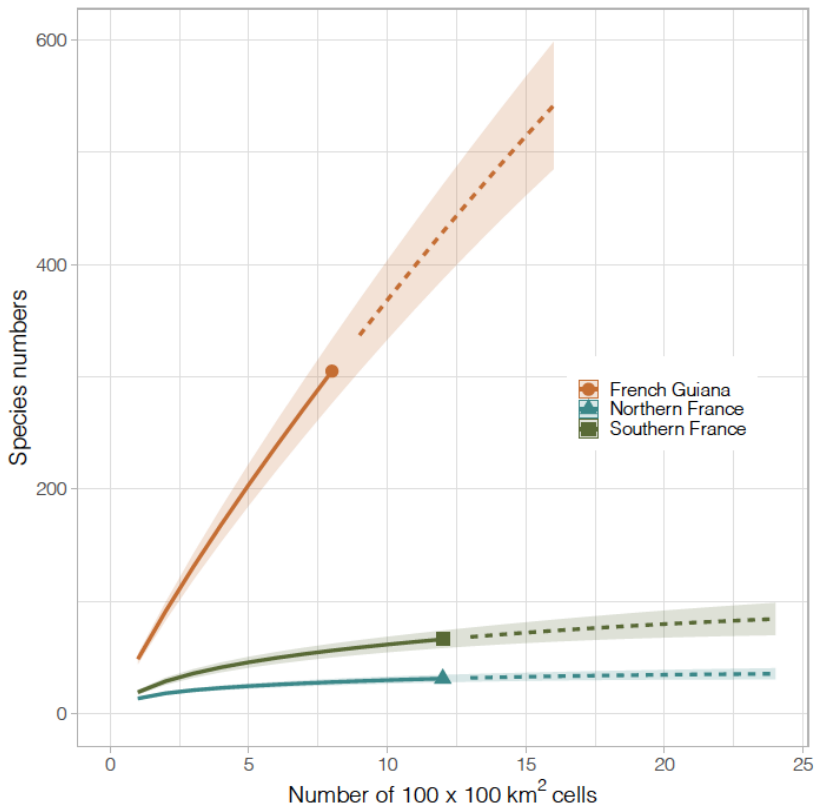
600
400
200
0
Species numbers
0
5
10
15
20
25
Number of 100 x 100 km2 cells
French Guiana
Northern France
Southern France

Figure 3.

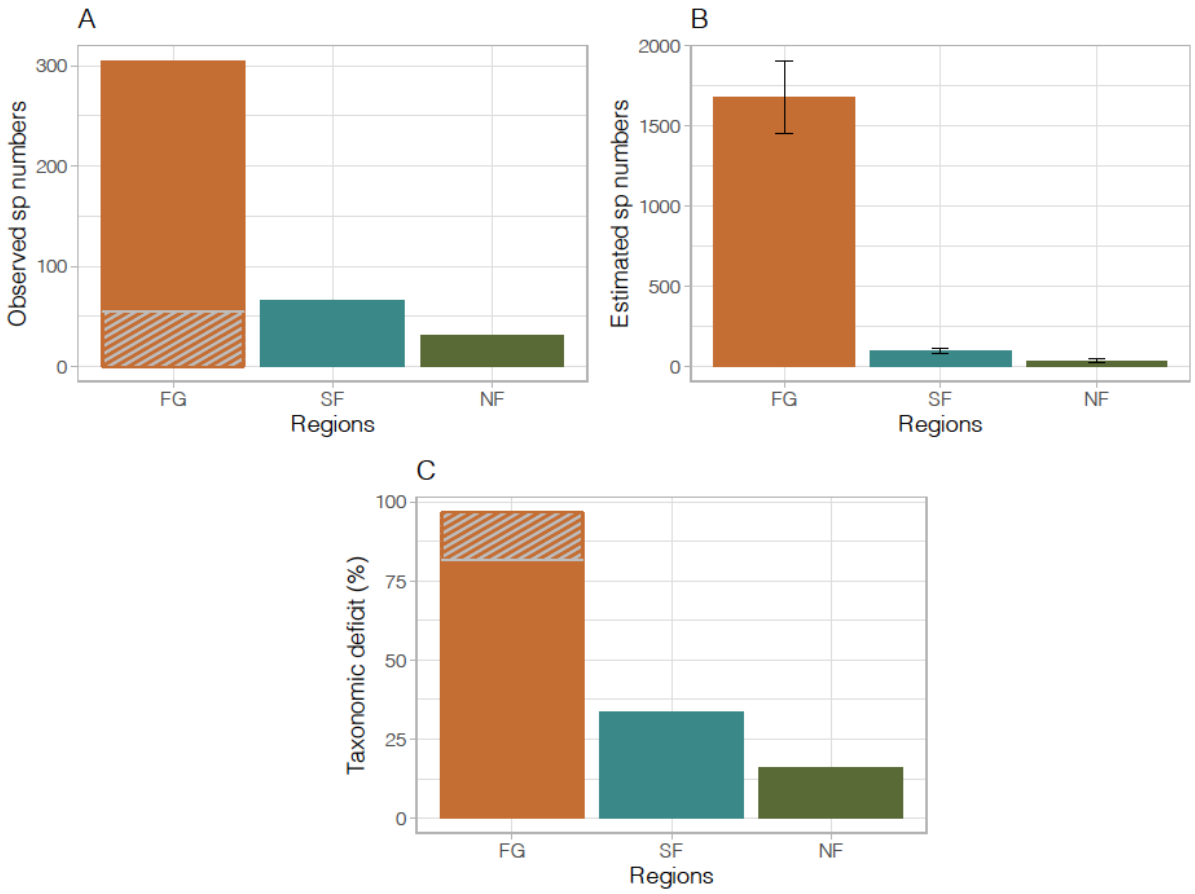
A
Observed sp numbers
300
200
100
0
FG
SF
NF
Regions
B
Estimated sp numbers
2000
1500
1000
500
0
FG
SF
NF
Regions
C
Taxonomic deficit (%)
100
75
50
25
0
FG
SF
NF
Regions

Figure 4.