A dataset on earthworm communities in French Guiana

Arnaud Goulpeau, Mickael Hedde, Marie-Eugénie Maggia, Sylvain Gérard, Emmanuel Lapied, Thibaud Decaëns

SOIL ORGANISMS 96 (2) · August 2024 pp. 79–90

A dataset on earthworm communities in French Guiana

Arnaud Goulpeau[1], Mickaël Hedde[2], Marie-Eugénie Maggia[3], Sylvain Gérard[2], Emmanuel Lapied[4] & Thibaud Decaëns[1]*

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

[2] Eco&Sols, INRAE, IRD, CIRAD, Institut Agro, Montpellier, France

[3] Department of Integrative Biology, University of Guelph, N1G 2W1 Guelph, Canada

[4] Taxonomia International Foundation, 7 rue Beccaria, 72012 Paris, France

* Corresponding author, e-mail: thibaud.decaens@cefe.cnrs.fr

Received 16 April 2024 | Accepted 28 May 2024
Published online at www.soil-organisms.org 1 August 2024 | Printed version 15 August 2024
DOI 10.25674/414

Abstract

Earthworms represent a crucial taxon in soil ecosystems in terms of biomass and ecological functions. Knowledge of their diversity is growing, but the understanding of the ecological and evolutionary mechanisms underlying this diversity and its distribution patterns remains poorly understood. This is partly due to a lack of community data available on a large scale, particularly in the natural ecosystems of the most diverse tropical regions. Here we describe a large dataset containing records of 3555 georeferenced earthworm specimens, sampled in 125 one-hectare plots distributed in different habitats and localities across French Guiana. Each of these specimens was DNA barcoded targeting the cytochrome C Oxidase I subunit barcode region (COI), and these sequences were clustered into molecular operational taxonomic units (OTUs) that we used as species proxy to describe community taxonomic composition and diversity. Each community is associated in the dataset with climate and elevation data, and soil properties are also available for part of them. This dataset represents a unique opportunity for analyzing community diversity and phylogeographic patterns in neotropical rainforests.

Keywords Oligochaeta | tropical rainforest | COI barcodes | alpha diversity | beta diversity

1. Introduction

Soils are recognised as one of the most biodiverse terrestrial habitats (Decaëns 2010, Anthony et al. 2023). Within soil biota, earthworms are ecosystem engineers whose functional importance is widely acknowledged (Lavelle et al. 2016). Their bioturbation activities influence soil structure, water infiltration, organic matter cycles, and life conditions for other soil species (Eisenhauer 2010, Capowiez et al. 2024). Paradoxically, earthworms suffer from a conspicuous lack of knowledge, which affects both the level of resolution of their taxonomy and our ability to describe and understand their community patterns in natural ecosystems, particularly in tropical regions (Decaëns et al. 2006, Decaëns 2010, Cameron et al. 2018, Guerra et al. 2020). French Guiana is a perfect illustration of this situation. Indeed, recent studies using DNA barcoding have uncovered the presence of several hundred operational taxonomic units (OTUs) of specific level, most of them corresponding to species new to science (Decaëns et al. 2016, 2024), suggesting that species richness in this region is far from fully established (Maggia et al. 2021, Goulpeau et al. 2022). In addition, Maggia et al. (2021) suggested considerable turnover in the composition of species pools between different localities or landscapes. Analysing the organisation of this diversity and its underlying ecological and evolutionary mechanisms represents

ISSN 1864-6417 · **eISSN** 2509-9523

a major scientific challenge that needs to be addressed to better understand current patterns and propose an effective conservation strategy. To do this, we critically need standardised datasets covering natural ecosystems at regional geographic scales.

Here, we present and briefly describe a large dataset which comprises information on the abundance and composition of 125 earthworm communities sampled in natural ecosystems distributed in 10 localities over French Guiana. Earthworm specimens were DNA barcoded and grouped into operational taxonomic units (OTUs), which we assume to be species surrogates. Parts of this dataset have been used in previous work (Decaëns et al. 2016, Maggia et al. 2021, Goulpeau et al. 2022). Here, we publish the raw data including the complete list of specimens collected, their DNA barcodes, as well as species composition data aggregated at different spatial scales or levels of sampling hierarchy (i.e. microhabitat, community, habitat and landscape). Environmental data are also available for part of the dataset (topography, climate, soil properties and vegetation composition). This large dataset combining community data, environmental data and DNA barcodes provides a unique opportunity to scrutinize earthworm diversity patterns in neotropical rainforests from both an ecological and an evolutionary perspective.

2. Materials and methods

2.1 Study Sites

Earthworms were sampled between 2011 and 2019 in ten localities distributed over French Guiana (Fig. 1), an area of 83,846 km^2 covered by over 95% of tropical rainforests, which is located in the Amazonian basin in the eastern Guiana precambrian shield, between the Oyapock and Maroni rivers. In each locality, we identified seven to twenty-five one-hectare plots, spaced at least 500 m apart, and representing the main habitat types of the local landscape. Habitats included plateau forests (dense rainforest on deep, well-drained soils, sometimes with superficial laterite crusts), slope forests (rainforest on slopes and deep soils), lowland forests (periodically flooded rainforest on hydromorphic soils) and plant associations specific to inselbergs and other landforms such as rocky savannas (herbaceous formations dominated by terrestrial bromeliads), transitional forests and hilltop forests (low canopy rainforest on shallow soils). In most cases, habitats were replicated at least three times in each locality or landscape, giving a total of 125 plots whose distribution is described in Table 1.

2.2 Earthworm sampling

Sampling was carried out during the rainy season, which corresponds to the period of maximum earthworm activity. In each one-hectare plot, earthworms were sampled using a combination of three methods (Fig. 2; Decaëns et al., 2016): 1) three blocks of soil, 25 × 25 cm^2 area and 20 cm deep, including the litter layer, were sampled at the corners of a central triangle of 20-m sides and hand sorted on a plastic sheet; 2) for the larger soil-dwelling species, an area of 1 m^2 was excavated to a minimum depth of 40 cm, selecting where possible an area showing traces of large earthworms on the soil surface (i.e. large casts); 3) finally, qualitative sampling was carried out on the one-hectare plot, looking for earthworms in all available and accessible microhabitats (i.e. decaying trunks, litter accumulations, streamside sediments, epiphytes, etc.) for a fixed period of three hours.persons (e.g. one hour for three people). All life stages (i.e. adults, juveniles and cocoons) were collected, and specimens collected using each method and in each type of microhabitat were fixed separately in 100% ethanol. The ethanol was changed in each tube after a period of 24 h to ensure good fixation.

2.3 DNA barcoding and OTU delimitation

For each sample (i.e. individuals collected with a given method in a given microhabitat and plot), earthworms were first sorted according to external morphological features (characteristics of the clitellum when adult, size and pigmentation), and for each morpho-group obtained, up to 5 individuals per sample were DNA barcoded targeting the barcode region of the Cytochrome C Oxidase I gene (COI). The complete procedure of DNA barcoding has been described in Goulpeau et al. (2022). A proportion of 75.2% of samples were Sanger processed at the Centre for Biodiversity Genomics (Guelph, Canada) using standard protocols of the International Barcode of Life with a cocktail combining the M13 tail primer pairs LCO1490/HCO2198 (Folmer et al. 1994) and LepF1/LepR1 (Hebert et al. 2004), and failure tracking with the internal primers MLepR1/MLepF1 and LCO/HCO pairs (Hajibabaei et al. 2006). The remaining 24.8% sequences were obtained on MiSeq at the Laboratoire d'Ecologie Alpine (Grenoble, France) using the LCO1490/HCO2198 primer pair (Folmer et al. 1994) combined with a label on the 5' side to allow further sample identification. Bioinformatic analysis was done with OBITools (Boyer et al. 2016; www.grenoble.prabi.fr/trac/OBITools). All sequences are available in the BOLD dataset 'Earthworms from French Guiana - 2023 update' (DS-EWFG2023; doi.org/10.5883/DS-EWFG2023).

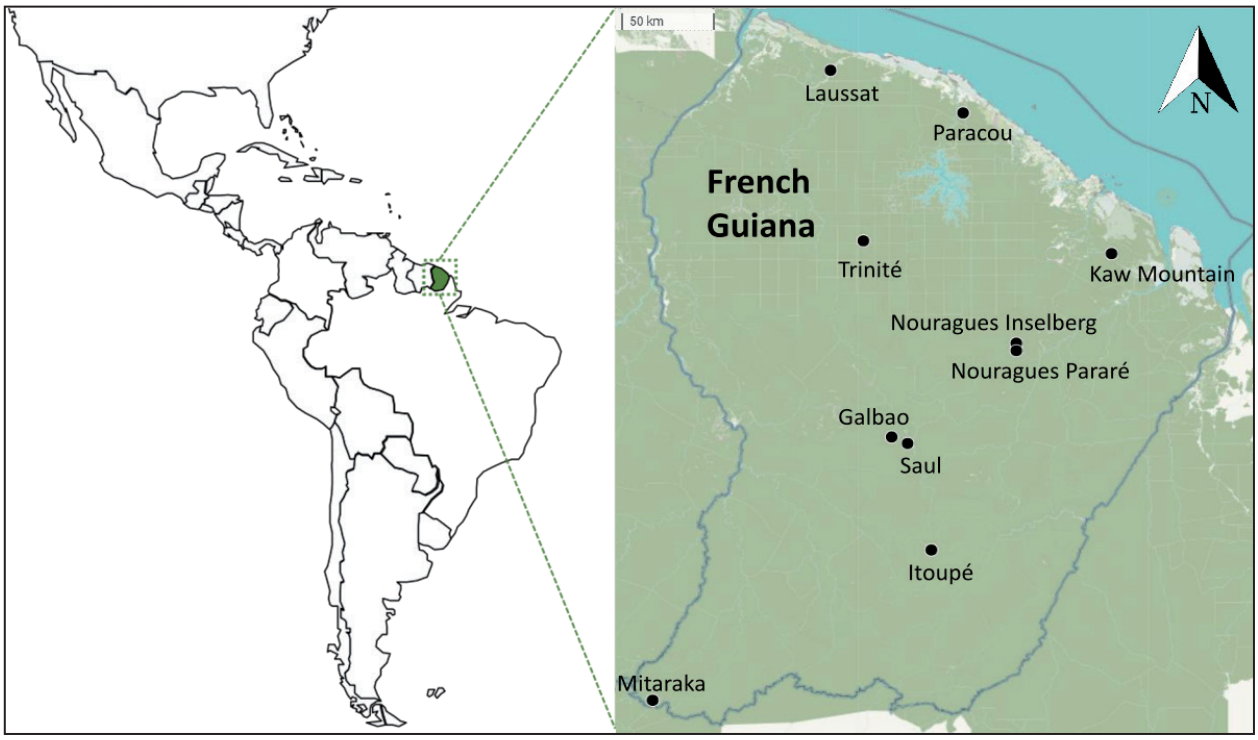

Figure 1. Location of the ten sampling localities in French Guiana.

Table 1. List of the sampling localities or landscapes, average latitudinal and longitudinal coordinates, and number of one-hectare plots sampled per habitat in each of them.

Locality / landscape	Latitude	Longitude	Lowland forests	Slope forests	Plateau forests	Inselberg habitats	Forests on white sands
Laussat	5.48117	-53.5762	2	-	2	-	3
Paracou	5.25833	-52.9346	3	4	3	-	-
Trinité	4.61871	-53.4081	3	2	3	2	-
Kaw	4.54064	-52.2152	-	2	7	-	-
Inselberg	4.08133	-52.6731	2	10	8	5	-
Pararé	4.02832	-52.6778	4	3	5	-	-
Galbao	3.60403	-53.268	2	4	2	3	-
Saül	3.5587	-53.2222	4	4	4	-	-
Itoupé	3.02696	-53.079	-	4	7	-	-
Mitaraka	2.243	-54.465	3	3	4	8	-

Sequences were further clustered into molecular operational taxonomic units (OTUs) using the Assemble Species by Automatic Partitioning (ASAP) method (Puillandre et al. 2021), which is the most suitable both for our study model and for the size of the dataset (Goulpeau et al. 2022). This was done using the ASAP web interface (https://bioinfo.mnhn.fr/abi/public/asap). It is worth mentioning that OTUs delimitations obtained for two localities (Nouragues and Mitaraka) were confronted to morphological characters, which broadly supported molecular delimitations (Decaëns et al. 2016), providing a strong argument for considering OTUs as relevant species proxies in further diversity analyses. Sequences were aligned with MUSCLE (Edgar 2004) on MEGA11 (Tamura et al. 2021). Further distance analyses were performed using a Neighbor-Joining algorithm (Saitou & Nei 1987) with the Kimura-2 parameter (K2p) model (Kimura 1980) to estimate genetic distances. The K2p model was chosen to allow a consistent comparison with most barcoding studies where this model is defined by default. The robustness of the nodes was assessed by a bootstrap re-analysis of 500 pseudo-replicates, and the Neighbor-Joining tree was drawn using the iTOL v6 online utility (Letunic & Bork 2007).

2.4 Descriptive statistics

We first described the dataset to highlight the proportion of individuals that were sampled depending on life stage, sampling method, habitat and microhabitat. The same was done for OTU composition using venn diagrams to show the proportion of shared and unique species according to different discriminating criteria. This was done using the 'euler' function in the 'eulerr' package for R (Larsson 2024, R Core Team 2023).

We then described our dataset through a hierarchy of nested spatial scales or sampling hierarchy levels (Fig. 2): 1) at the microhabitat scale, we considered the earthworms sampled in the different microhabitats in a given plot; 2) at the community scale, we grouped the specimens collected from the various microhabitats within a given plot; 3) the habitat scale refers to specimens collected in different plots of the same habitat type (i.e. plateau, slope and lowland forest, and inselberg habitats) within a given locality; 4) the landscape scale, groups together the data from all the plots sampled in the various habitats of the same locality; 5) the regional scale, corresponds to the spatial extent covered by the entire dataset.

Alpha diversity refers to species richness (or compositional diversity, sensu Whittaker 1972) at each previously defined spatial scale. Regional or gamma diversity refers to the total number of OTUs within the dataset. We used rarefaction / extrapolation curves showing how OTUs accumulate as a function of sampling effort (intrapolation) and extrapolating the shape of the curve to a hypothetical dataset where the sampling effort would have been multiplied by two (extrapolation). Sampling effort was expressed as the number of sampled landscapes at the regional scale, and as the number of sampled communities at the landscape scale. For

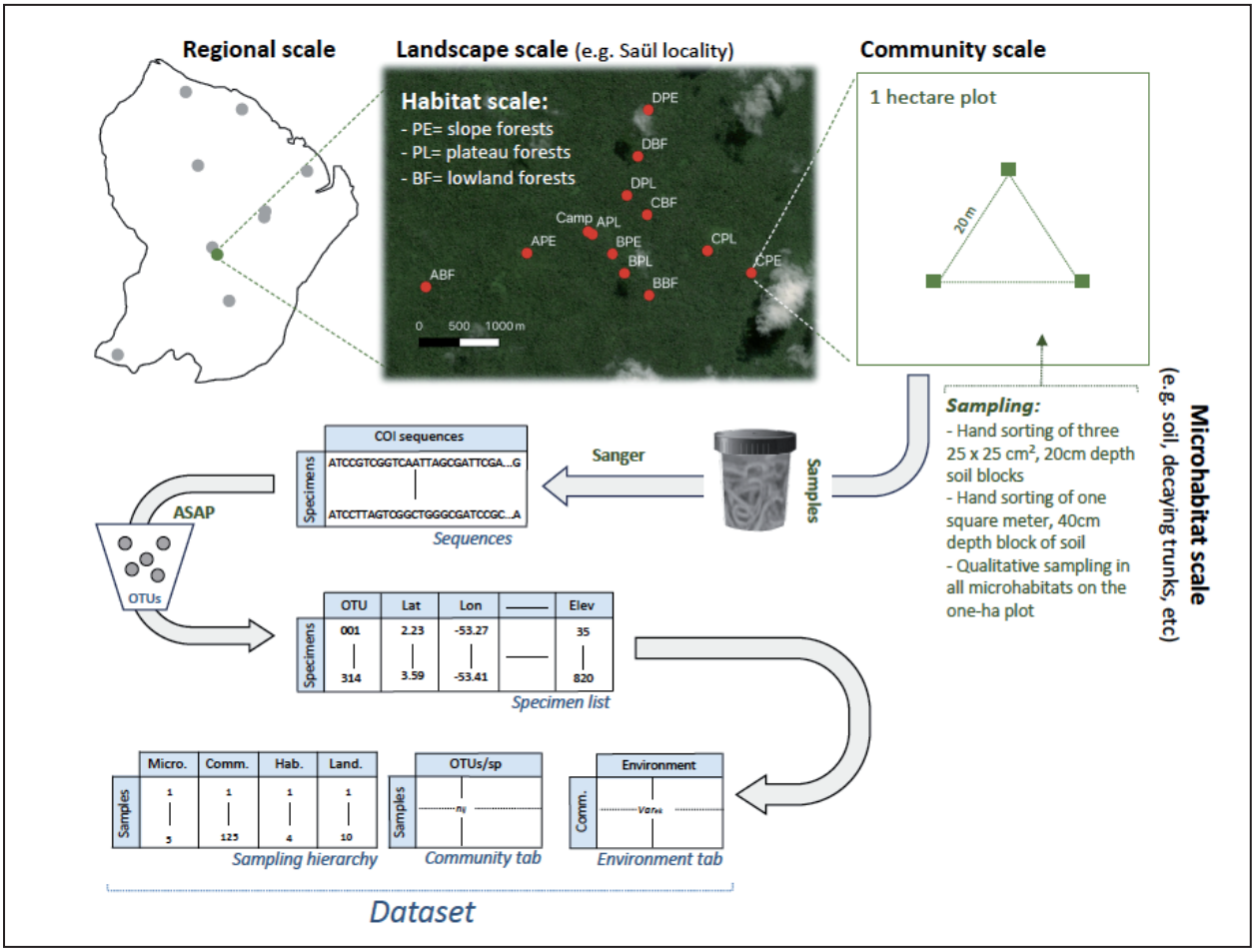

Figure 2. Framework of sample design of French Guiana earthworm with a summary of the resulting dataset: 'Sequences' tab contains the fasta sequences of DNA barcodes for all the sampled specimens (i.e. vouchers); 'Specimen list' contains OTU assignations, life stages and collecting data for all the specimens; 'Sampling hierarchy' contains the assignation of each sample to a hierarchy of spatial scales or sampling hierarchy levels, i.e. microhabitats (Micro.), communities (Comm.), habitats (Hab.) and landscapes (Land.); 'Community tab' contains the number of specimens sampled for each OTU in each sample, with n_{ij} the number of specimens of OTU *i* collected in sample *j*; 'Environment tab' contains the environmental data at the community scale, with var_{ek} the value taken by environmental variable *e* in community k. **ASAP** = Assemble Species by Automatic Partitioning.

each community, we also calculated sample coverage, which corresponds to the proportion of the community that is represented by the set of species collected in a given sample. Rarefaction curves and sample coverage calculations were done using the *'iNEXT'* package for R (Hsieh et al. 2019).

For beta diversity, we calculated the Sørensen's index of compositional dissimilarity among all possible pairs of communities: $\beta_{sor} = (b+c)/(2a + b + c)$, where a represents the number of species common to both communities, and b and c the number of species unique to the first and second community, respectively. We then partitioned beta diversity to highlight compositional dissimilarity among communities within the same habitat and landscape (beta A), among communities of different habitats in the same landscape (beta B), and among communities from different landscapes (beta C). This was done using the *'beta.pair'* function of the *'betapart'* package for R (Baselga et al. 2018). We also constructed a community network, which visualizes the compositional proximity between communities using the Sørensen similarity index, using the *'ggnet2'* function in the *'GGally'* package for R (Schloerke et al. 2024).

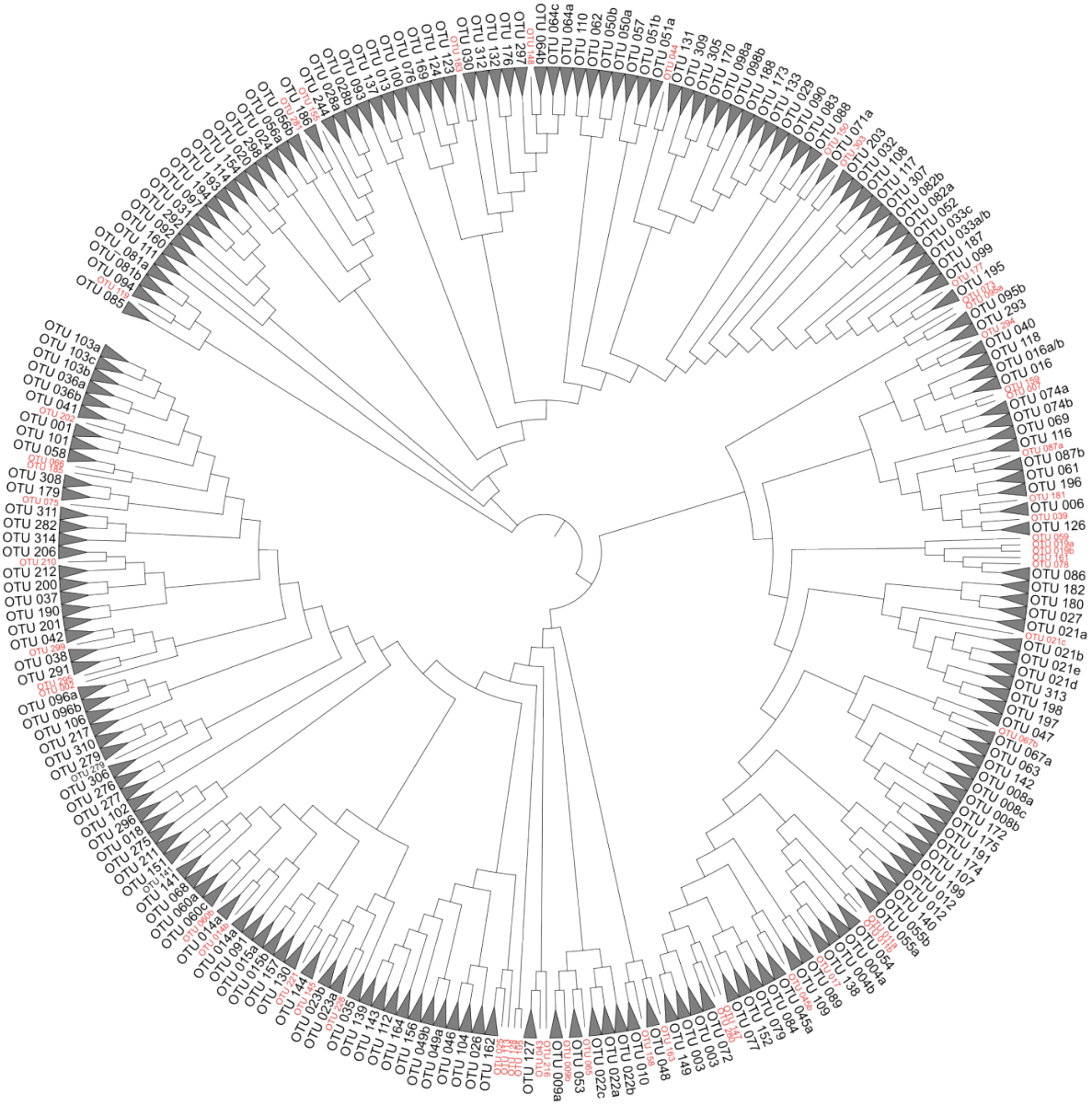

Figure 3. Neighbor-joining of the 3555 COI sequences clustered into 256 OTUs. Each OTU is represented by a triangle. Singletons, i.e. OTUs represented by a single specimen in the dataset, are represented in red.

2.5 Dataset availability statement

The complete list of specimens, associated metadata, COI sequences and GenBank accession numbers are available in the Barcode of Life Data system (BOLD) dataset 'Earthworms from French Guiana - 2023 update' (DS-EWFG2023; doi.org/10.5883/DS-EWFG2023), and the complete list of specimens, the community table and the environmental data are also deposited on the Zenodo repository (doi.org/10.5281/zenodo.10908657).

2.6 Dataset description

The dataset is organised into five tabs containing different layers of information (Fig. 2). The 'Sequences' tab contains the fasta-formated COI barcodes for the 3555 earthworm specimens collected in French Guiana. The delineation by the ASAP software grouped these sequences into 256 OTUs (Fig. 3) that were further used as species proxy in the other tabs of the dataset. 'Specimen list' tab contains 3555 lines, one for each sequence, and gives, for each of them, taxonomic assignment, life stage and collecting data. In both the 'Sequences' and 'Specimen list' tabs, samples are identified by their Sample_ID and Process-ID in BOLD. 'Community' tab contains 255 lines, each line representing a sample, i.e. the earthworm specimens collected by hand-sorting a given standardised soil block or by qualitative sampling of a given microhabitat in a given community. For each sample, the first four columns give its assignment to the modalities of each nested spatial or scale hierarchy sampling level (i.e.

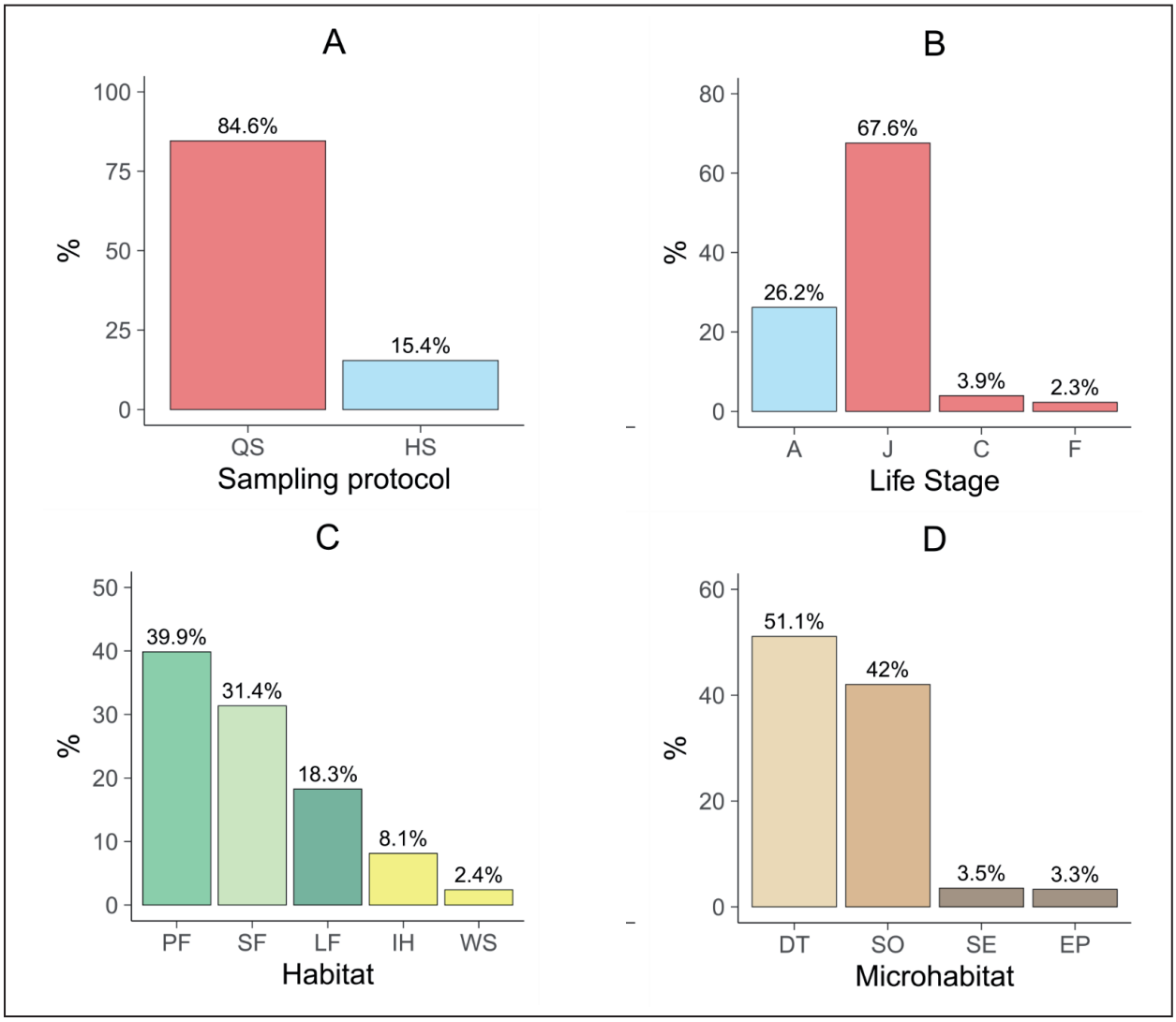

Figure 4. Histograms showing the proportion of sampled earthworms according to different criteria: A) Sampling protocol, with QS = qualitative sampling, HS = hand sorting of standardised soil blocks; (**B**) Life stage, with **A** = adults, **J** = juveniles, **C** = cocoons, **F** = unidentified fragments; (**C**) Habitat, with **PF** = plateau forests, **SF** = slope forests, **LF** = lowland forests, **IH** = inselberg habitats, **WS** = forests on white sands; (**D**) Microhabitat, with **DT** = decaying trunks, **SO** = soil, **EP** = epiphytic soils, **SE** = riverside sediments.

microhabitat, community, habitat and landscape). The following 256 columns contain the number of earthworm individuals collected for each OTU in each sample. The 'Species list' tab provides the available taxonomic information for the 256 OTUs. Finally, the 'Environment tab' gives the geolocalisation, altitude and topographical position, and mean annual precipitation for all the sampled communities, and describes, for a subset of 36 communities, the soil physico-chemical properties.

Among the 256 OTUs, 21.5 % are singletons (Fig. 3), which is in the range of what has been described in previous earthworm studies in French Guiana (Decaëns et al. 2016, Maggia et al. 2021). With over 84.6 % of the sampled specimens (Fig. 4A), the qualitative sampling approach allowed for the collection of 239 OTUs, 148 of them being found only with this method. Comparatively, hand sorting of standardised soil blocks alone only yielded 108 OTUs, among which 17 were only found with this method (Fig. 5A). This highlights the importance of combining hand sorting with qualitative sampling in studies aimed at carrying out diversity inventories, especially in forest ecosystems, as already discussed by Bartz et al. (2014) and Maggia et al. (2021). Overall specimens were in majority juveniles (67.6 %), whereas clitellate adults only represented 26.2 % (Fig. 4B). As a consequence, 117 OTUs out of 256 (i.e. 45.7 % of regional diversity) were only represented by immature specimens or fragments not identifiable from their morphology alone (Fig. 5B). This is a clear illustration of the usefulness of DNA barcoding in earthworm studies, as this approach allows assigning any specimens to its corresponding OTU or species (Richard et al. 2010, Decaëns et al. 2013). Specimens were mostly found in plateau (39.9 %), slope (31.4 %) and lowland (18.3 %) forests (Fig. 4C), and in soil (42.2 %) and decaying logs (51.1 %) (Fig. 4D). Altogether, 151 OTUs are specific to a single habitat (Fig. 5C). In line with the results of previous studies (Decaëns et al. 2016, Maggia et al. 2021), most

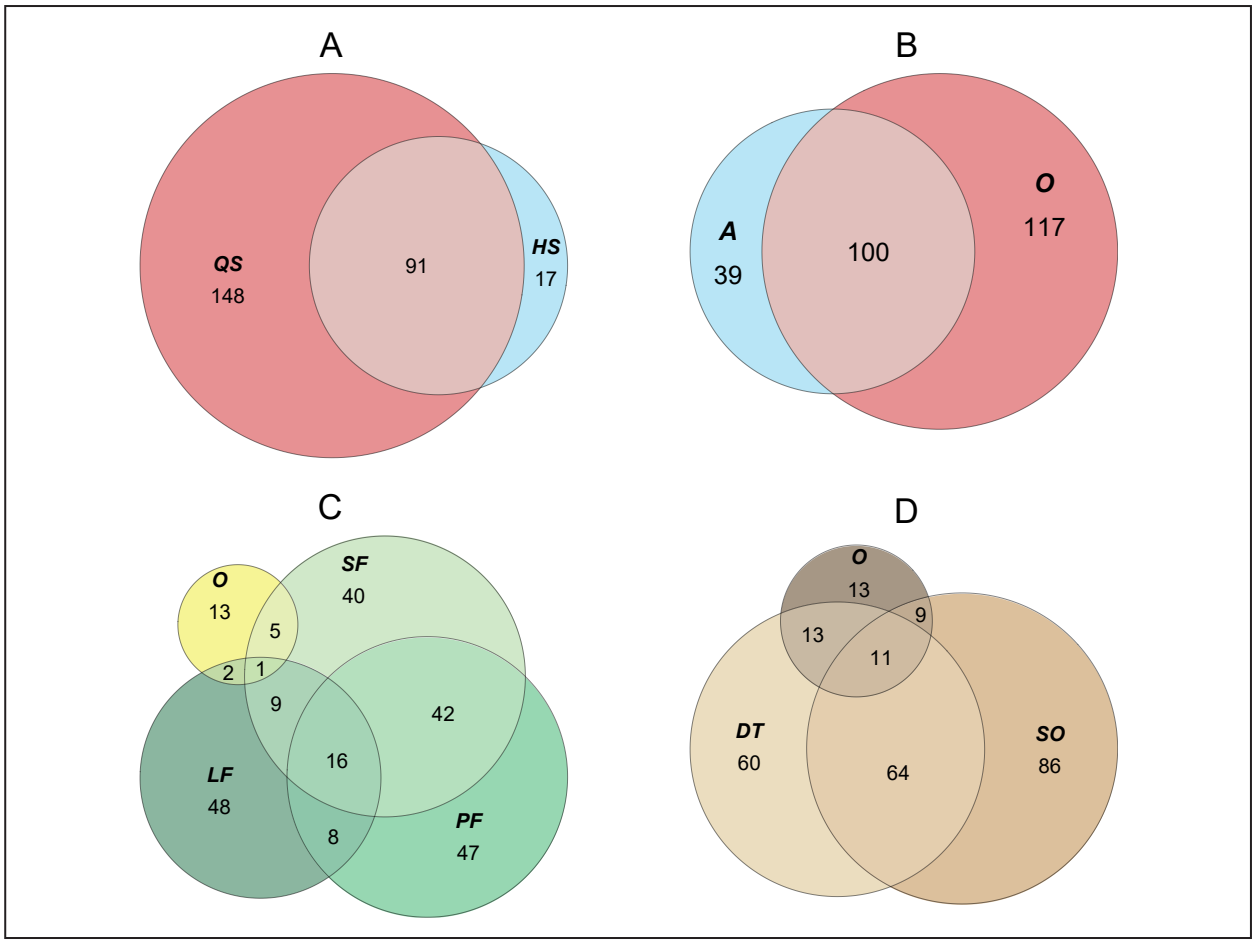

Figure 5. Venn diagrams representing the numbers of unique and shared OTUs for according to different criteria: (**A**) Sampling protocol, with **QS** = qualitative sampling, **HS** = hand sorting of standardised soil blocks; (**B**) Life stage, with **A** = adults, **O** = others (i.e. juveniles, cocoons and fragments); (**C**) Habitat, with **LF** = lowland forests, **PF** = plateau forests, **SF** = slope forests, **O** = others (i.e. inselberg habitats and forests on white sands); (**D**) Microhabitat, with **DT** = decaying trunks, **SO** = soil, **O** = others (i.e. riverside sediments and epiphytic soils). Note that in Figure 5C, the representation does not show 10 OTUs shared by the 4 habitats, 9 OTUs shared by PF, SF and O, 2 OTUs shared by LF, PF and O, and 1 OTU shared by PF and O.

of the diversity recovered in our study was in the soil with 170 OTUs, or in decaying trunks with 148 OTUs, including 75 OTUs found in both types of microhabitats (Fig. 5D).

Figure 6 illustrates the main patterns of alpha diversity across the hierarchy of spatial scales. At a regional scale, the rarefaction curve of OTU number as a function of the number of landscapes sampled (Fig. 6A) clearly shows that we are far from sampling completion at this scale. As a consequence, the 256 OTUs observed in our dataset probably represent only a small fraction of the real regional diversity. This under-sampling also applies, albeit to a lesser extent, at the landscape level (Fig. 6B). The average number of OTUs observed at the landscape scale is 33.2 (Fig. 6C), with significant differences observed from one landscape to another (Fig. 6B). The most diverse landscapes are at Pararé and Saül, both of which are characterised by a relatively low-lying topographical position compared with most of the other more hilly landscapes. This may be the result of an overall wetter environment in these landscapes, or in vegetation composition resulting in different litter quality depending on topographical position. Despite a similar topographical position, Laussat has the lowest diversity at landscape level, probably due to the predominance of forests on white sands, which likely represent adverse environmental conditions for earthworms (Fig. 6B). At finer scales,

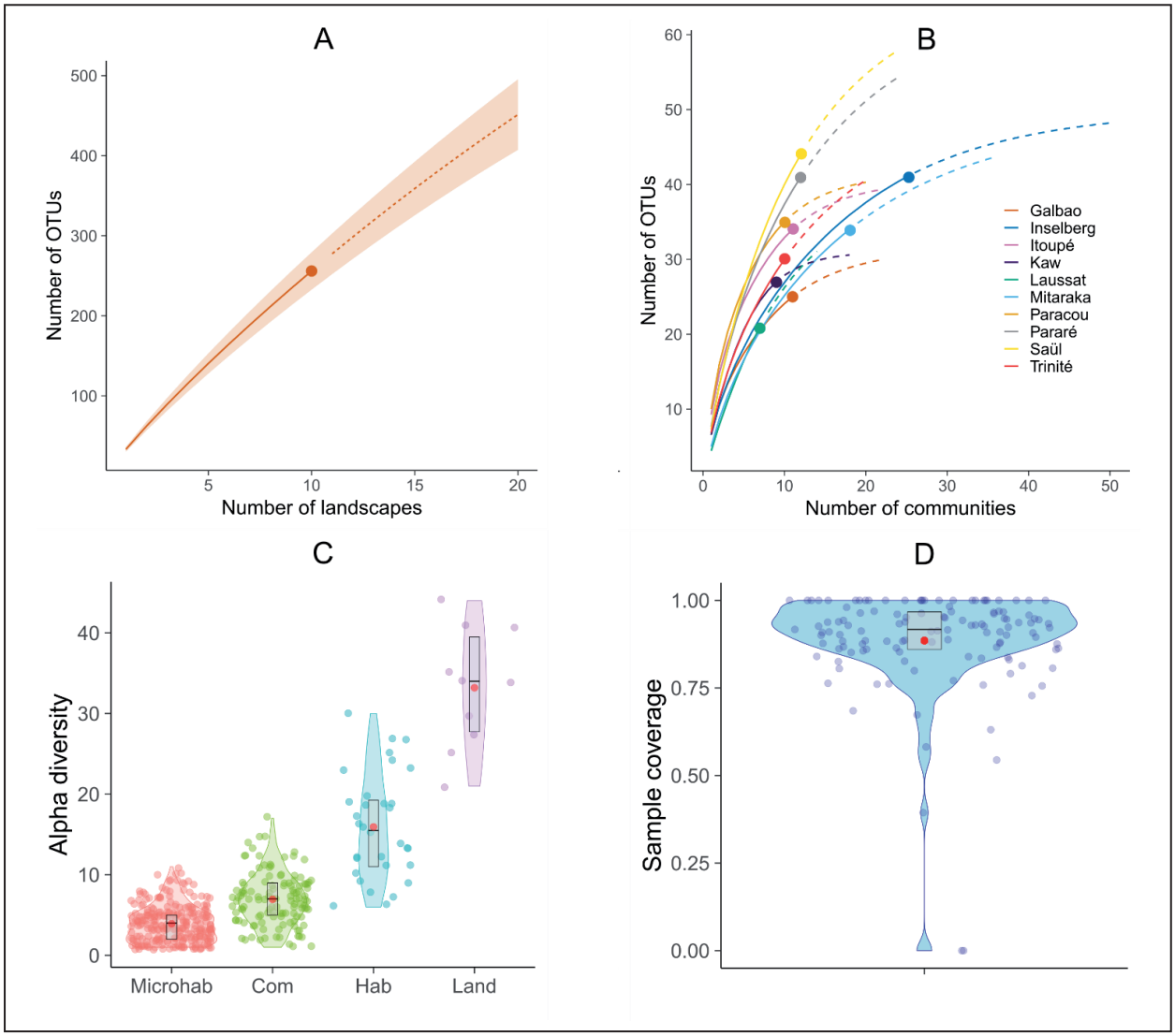

Figure 6. Patterns of earthworm alpha diversity at different spatial scales in rainforests of French Guiana: (**A**) Rarefaction curve showing how the number of OTUs observed at a regional scale increases as a function of the number of sampled landscapes; the solid line represents the rarefaction curve; the shaded area represents the 95% confidence interval based on a bootstrap with 200 replications; (**B**) Rarefaction curves showing how the number of OTUs observed at the landscape scale increases as a function of the number of sampled communities; confidence intervals were not represented to keep the figure legible; (**C**) Alpha diversity at the different spatial scales, with **Microhab** = microhabitat, **Com** = community, **Hab** = habitat and **Land** = landscape level); (**D**) Sample coverage at the community scale.

we observed on average 15.9 OTUs at the habitat scale, 7.0 OTUs at the community scale and 3.9 OTUs at the microhabitats scale (Fig. 6C). Local sampling effort is satisfactory overall, as shown by the sampling coverage which is higher than 80 % for 89.6 % of the sampled communities (Fig. 6D). This means that our sampling was globally efficient to sample the most representative species at the community level, thus validating the relevance of the dataset for more in-depth community ecology analyses.

Figure 7 presents the main patterns of composition dissimilarity (beta diversity) across the hierarchy of spatial scales. The community network shows that communities are globally dissimilar in their OTU composition, and that they are more similar when geographically closer, whether within the same landscape or between very close landscapes (e.g. between Inselberg and Pararé) (Fig. 7A). This suggests a high beta diversity among communities, as already discussed by Maggia et al. (2021). This is confirmed by the occurrence histogram which shows that the vast majority of OTUs (i.e. 94.5 %) were observed in strictly less than two localities (Fig. 7B). The analysis of Sørensen beta diversity between pairs of communities across the hierarchy of sampling levels shows that compositional dissimilarity is the lowest between communities sampled in the same habitats (beta A), then significantly increases for communities sampled in different habitats (beta B), and is the highest for communities belonging to different landscapes (beta C) (Fig. 7C). This suggests that both geographic distance

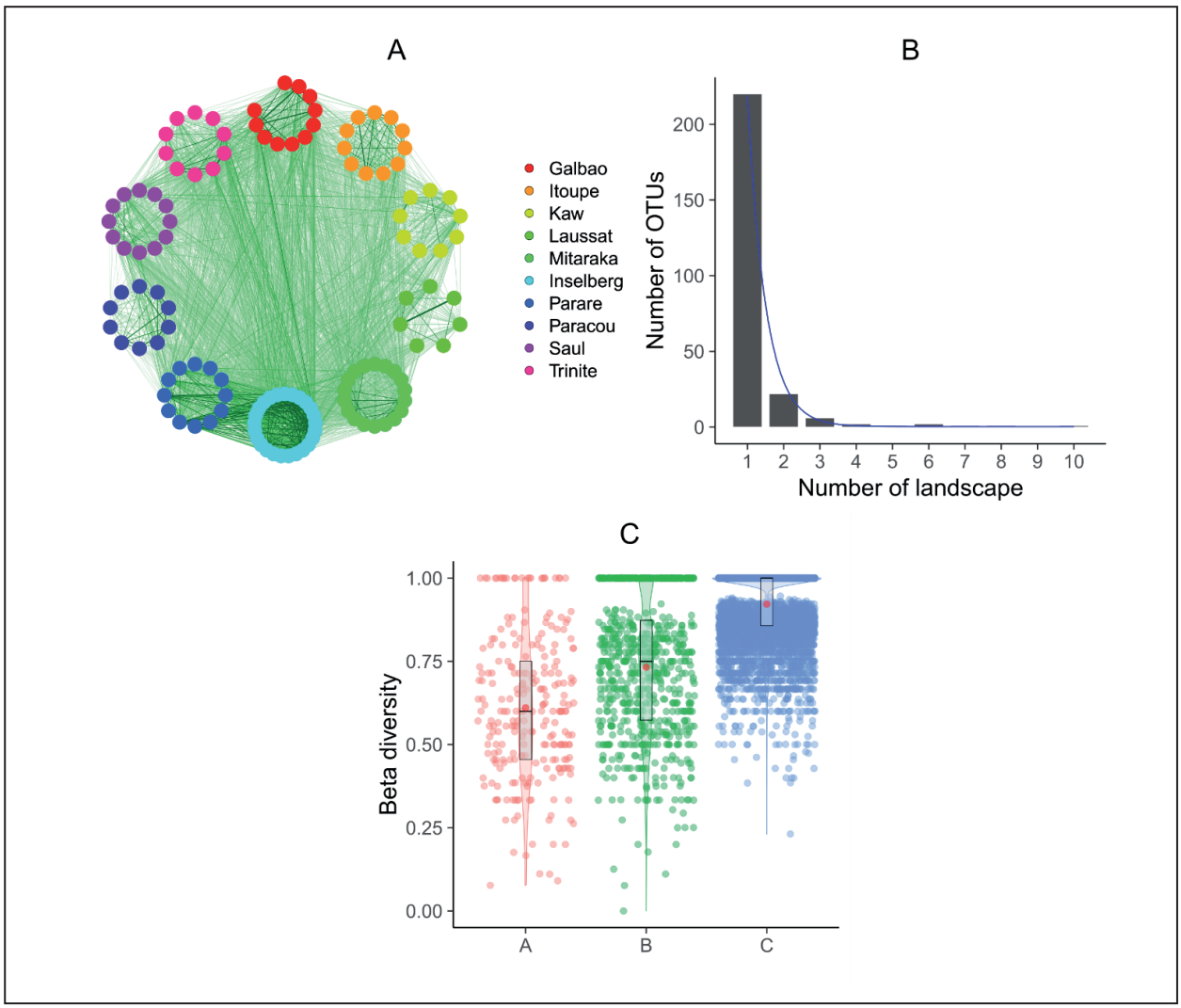

Figure 7. Patterns of earthworm beta diversity at different spatial scales in rainforests of French Guiana: (**A**) Community network; line thickness is proportional to the Sørensen similitude between pairs of communities; (**B**) Occurrence histogram showing the number of OTUs per occurrence class (i.e. number of landscapes in which each OTU was observed); blue line represents the exponential decay regression curve; (**C**) Sørensen's beta diversity between pairs of earthworm communities partitioned into different levels, with beta **A** = beta diversity between communities within the same habitat and landscape, beta **B** = beta diversity between communities of different habitats in the same landscape, beta **C** = beta diversity between communities from different landscapes.

and environmental dissimilarity could be important drivers of community species composition.

Our dataset comprises a high number of OTUs or species and a substantial number of georeferenced communities at a regional scale, rendering it a relevant dataset for an in-depth study of earthworm diversity in tropical rainforests of French Guiana. Such a dataset represents a unique opportunity for analysing the distribution of diversity across various geographic scales and to elucidate the intricate interplay among dispersal limitation, biotic interactions, and environmental filtering in the community assembly process. DNA barcodes also enable phylogeographic and population genetics approaches to shed light on historical events that may have influenced species biogeographical history and the distribution of species diversity in the study area. Furthermore, adopting an evolutionary perspective in studying diversification dynamics could provide complementary insights to phylogeographic approaches, enhancing our understanding of earthworm diversification mechanisms in Neotropical rainforests. In the near future, morpho-anatomical analysis of the various OTUs will enable each of them to be described using a range of functional traits. This information will complement that already included in the dataset describing the ecological preferences of OTUs (with respect to habitats and microhabitats), and will enable further trait-based approaches to be carried-out. The dataset published here will also be available for inclusion in more extensive data syntheses at the scale of the Neotropics or even more globally.

3. Online supplementary material

Goulpeau_et_al_dataset.xlsx

4. Acknowledgement

Part of the dataset used in this study was acquired as part of the 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). Sampling in Mitaraka was carried out as part of the 'Our Planet Reviewed' French Guiana 2015 expedition (Touroult et al. 2018) 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. At the Réserve naturelle des Nouragues, the project was supported by two Centre National de la Recherche Scientifique (CNRS) Nouragues grants in 2010 and 2011. At the Réserve naturelle de la Trinité, part of the funding was provided by the nature reserve. The authors would like to thank the Parc Amazonien de Guyane (http://www.parcamazonien-guyane.fr), the Réserve naturelle de la Trinité (http://www.reserve-trinite.fr/) and the Réserve naturelle des Nouragues (http://www.nouragues.fr/) for authorising access and collecting. Sample DNA barcoding received funding from the Canadian Centre for DNA Barcoding (CFREF-2015-00004), as well as from the CNRS through the BC-Wormbank project (APEGE 2013 call). This research is a product of the "FaunaServices" group funded by the French Foundation for Research on Biodiversity (FRB) through its synthesis center CESAB in France, and by FAPESP/CNPq in Brasil.

5. References

Anthony, M. A., S. F. Bender & M. G. A. van der Heijden (2023): Enumerating soil biodiversity. – Proceedings of the National Academy of Sciences **120**: e2304663120 [https://doi-org.inee.bib.cnrs.fr/10.1073/pnas.2304663120].

Bartz, M. L. C., G. G. Brown, M. G. Da Rosa, O. K. Filho, S. W. James, T. Decaëns & D. Baretta (2014): Earthworm richness in land-use systems in Santa Catarina, Brazil. – Applied Soil Ecology **83**: 59–70 [https://doi.org/10.1016/j.apsoil.2014.03.003].

Baselga, A., D. Orme, S. Villeger, J. De Bartoli, F. Leprieur (2018): betapart: Partitioning Beta Diversity into Turnover and Nestedness Components. R package version 1.5.1.

Boyer, F., C. Mercier, A. Bonin, Y. Le Bras, P. Taberlet & E. Coissac (2016): OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. – Molecular Ecology Resources **16**: 176–182 [https://doi.org/10.1111/1755-0998.12428].

Cameron, E. K., I. S. Martins, P. Lavelle, J. Mathieu, L. Tedersoo, F. Gottschall, C. A. Guerra, J. Hines, G. Patoine, J. Siebert & et al. (2018): Global gaps in soil biodiversity data. – Nature Ecology & Evolution **2**: 1042–1043 [https://doi.org/10.1038/s41559-018-0573-8].

Capowiez, Y., D. Marchán, T. Decaëns, M. Hedde & N. Bottinelli (2024): Let earthworms be functional – Definition of new functional groups based on their bioturbation behavior. – Soil Biology and Biochemistry **188**: 109209 [https://doi.org/10.1016/j.soilbio.2023.109209].

Decaëns, T., J. J. Jiménez, C. Gioia, J. Measey, P. Lavelle (2006): The values of soil animals for conservation biology. – European Journal of Soil Biology **42**: S23–S38 [https://doi.org/10.1016/j.ejsobi.2006.07.001].

Decaëns, T. (2010): Macroecological patterns in soil communities. – Global Ecology and Biogeography **19**: 287–302 [https://doi-org.inee.bib.cnrs.fr/10.1111/j.1466-8238.2009.00517.x].

Decaëns, T., D. Porco, R. Rougerie, G. G. Brown & S. W. James (2013): Potential of DNA barcoding for earthworm research in taxonomy and ecology. – Applied Soil Ecology **65**: 35–42 [https://doi.org/10.1016/j.apsoil.2013.01.001].

Decaëns, T., D. Porco, S. W. James, G. G. Brown, V. Chassany, F. Dubs, L. Dupont, E. Lapied, R. Rougerie, J.-P. Rossi & et al. (2016): DNA barcoding reveals diversity patterns of earthworm communities in remote tropical forests of French Guiana. – Soil Biology and Biochemistry **92**: 171–183 [https://doi.org/10.1016/j.soilbio.2015.10.009].

Decaëns, T., M. C. Bartz, A. M. Feijo, A. Goulpeau, E. Lapied, D. F. Marchán, M. E. Maggia, G. Papuga & S. W. James (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**(9): 195–244 [https://doi.org/10.5252/zoosystema2024v46a9].

Edgar, R. C. (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. – Nucleic Acids Research **32**: 1792–1797 [https://doi.org/10.1093/nar/gkh340].

Eisenhauer, N. (2010): The action of an animal ecosystem engineer: Identification of the main mechanisms of earthworm impacts on soil microarthropods. – Pedobiologia **53**: 343–352 [https://doi.org/10.1016/j.pedobi.2010.04.003].

Folmer, O., M. Black, W. Hoeh, R. Lutz & R. Vrijenhoek (1994): DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. – Molecular Marine Biology and Biotechnology **3**: 294–299.

Goulpeau, A., B. Penel, M.-E. Maggia, D. F. Marchán, D. Steinke, M. Hedde & T. Decaëns (2022): OTU delimitation with earthworm DNA barcodes: a comparison of methods. – Diversity **14**: 866 [https://doi.org/10.3390/d14100866].

Guerra, C. A., A. Heintz-Buschart, J. Sikorski, A. Chatzinotas, N. Guerrero-Ramirez, S. Cesarz, L. Beaumelle, M. C. Rillig, F. T. Maestre, M. Delgado-Baquerizo & et al. (2020): Blind spots in global soil biodiversity and ecosystem function research. – Nature Communications **11**: 3870 [https://doi.org/10.1038/s41467-020-17688-2].

Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs & P. D. N. Hebert (2006): DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences – **103**: 968–971 [https://doi.org/10.1073/pnas.0510466103].

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen & W. Hallwachs (2004): Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. – Proceedings of the National Academy of Sciences **101**: 14812–14817 [https://doi.org/10.1073/pnas.0406166101].

Hsieh, T. C., K. H. Ma & A. Chao (2019): iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.19.

Kimura, M. (1980): A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. – Journal of Molecular Evolution **16**: 111–120 [https://doi.org/10.1007/BF01731581].

Larsson, J. (2024): eulerr: area-proportional Euler and Venn diagrams with ellipses. R package version 7.0.1.

Lavelle, P., A. Spain, M. Blouin, G. G. Brown, T. Decaëns, M. Grimaldi, J. J. Jiménez, D. McKey, J. Mathieu, E. Velasquez & et al. (2016): Ecosystem engineers in a self-organized soil: A review of concepts and future research questions. – Soil Science **18**: 91–109 [https://doi.org/10.1097/SS.0000000000000155].

Letunic, I. & P. Bork (2007): Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. – Bioinformatics **23**: 127–128 [https://doi.org/10.1093/bioinformatics/btl529].

Maggia, M.-E., T. Decaëns, E. Lapied, L. Dupont, V. Roy, H. Schimann, J. Orivel, J. Murienne, C. Baraloto, K. Cottenie & 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 [https://doi.org/10.1016/j.apsoil.2021.103932].

Puillandre, N., S. Brouillet & G. Achaz (2021): ASAP: Assemble species by automatic partitioning. – Molecular Ecology Resources **21**: 609–620 [https://doi.org/10.1111/1755-0998.13281].

R Core Team (2023): R: A language and environment for statistical computing. Version 4.3.1.

Richard, B., T. Decaëns, R. Rougerie, S. W. James, D. Porco, P. D. N. Hebert (2010): Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding. – Molecular Ecology Resources **10**: 606–614 [https://doi.org/10.1111/j.1755-0998.2009.02822.x].

Saitou, N. & M. Nei (1987): The neighbor-joining method: a new method for reconstructing phylogenetic trees. – Molecular Biology and Evolution **4**: 406–425 [https://doi.org/10.1093/oxfordjournals.molbev.a040454].

Schloerke, B., D. Cook, J. Larmarange, F. Briatte, M. Marbach, E. Thoen, A. Elberg & J. Crowley (2024): GGally: Extension to 'ggplot2'. R package version 2.2.1.

Tamura, K., G. Stecher & S. Kumar, S. (2021): MEGA11: Molecular Evolutionary Genetics Analysis Version 11. – Molecular Biology and Evolution **38**: 3022–3027 [https://doi.org/10.1093/molbev/msab120].

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

Whittaker, R. H. (1972): Evolution and measurement of species diversity. – Taxon **21**: 213–251 [https://doi.org/10.2307/1218190].