At each site its diversity: DNA barcoding reveals remarkable earthworm diversity in neotropical rainforests of French Guiana

Authors:

Marie-Eugénie Maggia[a], Thibaud Decaëns[b], Emmanuel Lapied[c], Lise Dupont[d], Virginie Roy[d] Heidy Schimann[e], Jérôme Orivel[e], Jérôme Murienne[f], Christopher Baraloto[g], Karl Cottenie[a] and Dirk Steinke[a,h]

a Department of Integrative Biology, University of Guelph, N1G 2W1, Guelph, Canada

b CEFE, Université de Montpellier, CNRS, EPHE, IRD, Université Paul Valéry Montpellier 3, Montpellier, France

c Taxonomia International Foundation, 7 rue Beccaria, 72012, Paris, France

d Université Paris Est Créteil (UPEC), Sorbonne université, CNRS, INRAE, IRD, Institut d'Ecologie et des Sciences de l'Environnement de Paris, 94010 Créteil Cedex, France

e UMR Ecologie des Forêts de Guyane, Campus agronomique, BP 316, 97379, Kourou Cedex, France.

f Laboratoire Evolution et Diversité Biologique (EDB, UMR5174) - Université Toulouse 3 Paul Sabatier, CNRS, IRD - 118 route de Narbonne, 31062, Toulouse cedex, France.

g International Center for Tropical Botany, Department of Biological Sciences, Florida International University, Miami, FL, 33199, USA

h Centre for Biodiversity Genomics, University of Guelph, N1G 2W1, Guelph, Canada

Abstract

Despite their recognized essential role in soil, earthworms in tropical environments are still understudied. The aim of this study was to re-evaluate the diversity at the regional scale, as well as to investigate the environmental and spatial drivers of earthworm communities. We sampled earthworm communities across a range of habitats at six locations in French Guiana using three different sampling methods. We generated 1675 DNA barcodes and combined them with data from a previous study. Together, all sequences clustered into 119 MOTUs which were used as proxy to assess species richness. Only two MOTUs were common between the six locations and 20.2 % were singletons, showing very high regional species richness and a high number of rare species. A canonical redundancy analysis was used to identify key drivers of the earthworm community composition. The RDA results and beta-diversity calculations both show strong species turnover and a strong spatial effect, resulting from dispersal limitations that are responsible for the current community composition. Sampling in different microhabitats allowed the discovery of 23 MOTUs that are exclusively found in decaying trunks and epiphytes, highlighting hidden diversity of earthworms outside of soil.

Key words: DNA barcoding, Tropical rainforest, Earthworm, Community ecology, β-diversity, regional diversity

1. Introduction

Despite the fact that they host a large and complex array of species, soils still remain the most understudied habitat of terrestrial ecosystems (Bardgett and Putten, 2014; Decaëns, 2010; Wolters, 2001). Based on recent species richness estimates and because of a huge taxonomic deficit, soils

47 are considered the third biotic frontier after tropical forest canopies and oceanic abysses (André et
48 al., 1994; Giller et al., 1997; Wolters, 2001). Soil invertebrates in particular are key actors in most
49 terrestrial ecosystems, including agroecosystems (Decaëns, 2010), as their activities are essential
50 in sustaining key ecological processes (Lavelle et al., 2006). However, they remain insufficiently
51 studied in comparison with other terrestrial aboveground organisms (Wolters, 2001). As a result,
52 soil biodiversity patterns and their drivers remain largely unknown, especially at the global scale
53 (Decaëns, 2010; Phillips et al., 2019). Several studies have already pointed out that soil organisms
54 show different ecological patterns than those observed through the study of aboveground
55 organisms (Cameron et al., 2019; Fierer et al., 2009; Phillips et al., 2019).

56 Earthworms are considered major faunal actors because of their importance in the maintenance
57 of soil functions and the provisioning of soil ecosystem services (Lavelle et al., 2016). They have
58 been characterized as soil engineers, due to their capacity at altering the soil structure, with
59 important effects on its physical, chemical, and biological functioning (Jones et al., 1994; Lavelle
60 et al., 2016, 2006). They are globally distributed, present in both temperate and tropical soils, with
61 the exception of the driest and coldest regions of the planet, and can make up 60 % - 80 % of
62 overall soil biomass (Amat et al., 2008). However, there are only a few studies so far focussing on
63 earthworm community structure at a regional scale, and there is still a considerable lack of
64 knowledge on their ecology and distribution particularly for tropical ecosystems (Feijoo, 2001;
65 Fragoso, 1985; Jiménez, 1999; Lavelle, 1978).

66 One considerable roadblock to a better understanding of community ecology is the existence
67 of a taxonomic impediment on soil biodiversity in general and earthworms in particular (André et
68 al., 2001; Decaëns, 2010). Recent developments of molecular tools such as DNA barcoding have
69 the potential to overcome the barriers of traditional taxonomy and thus facilitate the acquisition of

new data that in turn can be used to describe the spatial distribution of species and communities in a rapid and comprehensive fashion. DNA barcoding uses the mitochondrial gene cytochrome c oxidase I (COI) as standard genetic marker for identification of animal species (Hebert et al., 2003). It can also be used as a mean to delineate Molecular Operational Taxonomic Units (MOTUs) in the absence of prior morphological identification. MOTUs are increasingly used to estimate taxonomic richness and to describe the spatial distribution of communities (Blaxter et al., 2005; Porco et al., 2013; Smith et al., 2005; Young et al., 2012). For instance, they were used to study diversity patterns of earthworm communities in the tropical rainforests of French Guiana (Decaëns et al., 2016). Authors were able to delimit 48 MOTUs that almost perfectly match with adult morphology, suggesting that MOTUs based on COI barcodes could in fact represent true biological species. The use of barcoding allows to take into account morphologically unidentifiable specimens such as juvenile earthworms or cocoons, as well as cryptic species, unlike the traditional taxonomy identification method. Unfortunately, the use of barcoding is still limited in the study of tropical earthworm communities, biasing the current datasets in tropical regions. With that, the classic Tropical Soil Biology and Fertility (TSBF) quantitative sampling method (Anderson and Ingram, 1989) that has been widely used in tropical studies to characterize earthworm biodiversity, but also other key members of soil biota, does not seem adapted to the context of tropical rainforests. For instance, Bartz et al. (2014) showed that more species were collected using the qualitative method and Decaëns et al. (2016) demonstrated that many species can be found in microhabitats others than the soil *sensu stricto*. Inadequate sampling methods may therefore generate a strong undersampling of earthworm species diversity in tropical ecosystems, and represent major barriers in the study and understanding of tropical earthworm communities in tropical regions, generating an underestimation of species diversity in the tropics.

For this study we generated a comprehensive data set at different spatial scales using samples collected during several expeditions in Amazonian tropical rainforests over the past few years. The sampling protocol coupled three methods comprising the traditional TSBF method associated with qualitative sampling in the soil and in other microhabitats. We wanted to analyse earthworm community patterns at regional, local, and habitat scale using newly generated DNA barcodes.

With the addition of new data for this region we expected an increase in the number of MOTUs at the regional scale (γ-diversity) and a low level of shared diversity between locations explained by a strong turnover (β-diversity); as it has been suggested that the high regional diversity of earthworms in the tropics could be the result of a high level of endemism and/or an higher beta-diversity towards the equator (Decaëns, 2010; Lavelle and Lapied, 2003). However, this pattern might not be observed at local or smaller scales (α-diversity). Indeed, some studies suggested a lack of local diversity peak in the tropics due to interspecific competition (Decaëns, 2010; Phillips et al., 2019), while others argued that a deficit of sampling, coupled with high levels of geographical turnover in community composition, could hide the existence of the latitudinal gradient (Lavelle and Lapied, 2003) or that there is no difference between tropical and temperate regions (Lavelle, 1983). In addition, we also expected that key environmental drivers such as soil properties (pH, organic carbon, soil texture etc.) or climate and habitat type (i.e. forest type) will influence earthworm diversity by shaping their community structure as it has been shown in previous studies (Mathieu and Davies, 2014; Phillips et al., 2019; Rutgers et al., 2016; Spurgeon et al., 2013).

2. Material and methods

2.1. Study sites

The sampling was performed in the Amazonian rainforests of French Guiana. This French overseas territory on the northern Atlantic coast of South America is 83,846 km^2 which more than 95 % is covered by primary rainforest. It is characterized by a relatively uniform tropical humid climate, with only two seasons: a wet season between December and June, usually interrupted in February or March by a short drier period, and a dry season between July and November.

Four different locations, Galbao, Itoupé, Mitaraka and Trinité were sampled during rainy seasons between 2015 and 2019 (Figure 1) as part of the DIADEMA (DIssecting Amazonian Diversity by Enhancing a Multiple taxonomic-groups Approach) and DIAMOND (DIssecting And MONitoring amazonian Diversity) projects of the Labex CEBA (Center for the study of Biodiversity in Amazonia) and the expedition "Our Planet Reviewed" (Touroult et al., 2018). At each location, sampling was carried out in 11 to 19 1 ha plots, spaced by at least 500 m from each other, and representing contrasting local habitat types: hilltop (low canopy tropical rainforest located on somital position on shallow soils that dry quickly), plateau (high tropical rainforest on deep and well-drained soil), slope (tropical rainforest on slope and deep soils), swamp forests (tropical rainforest on hydromorphic soils) and specific vegetation formations of the inselbergs such as transition forests and rocky savannah when present (Supplementary Table 1).

- The Mont Galbao (lon / lat: -53.2830 / 3.5886) is a mountain range spreading over 6 km inside the National Amazonian Park of French Guiana (PAG) with peaks reaching 650 to 730 m above sea level (Figure 1). A total of 11 plots were sampled in January 2019, representing hilltop, slope, plateau and swamp forests at elevations ranging from 466 to 723 m.

- The Mont Itoupé (lon / lat: -53.0834 / 3.0222) is a mountain belonging to the Tabular Mountains chain located in the centre of the PAG (Figure 1). It is the second highest peak in the territory, with an altitude of 830 m above sea level, and it is composed of a large plateau covered by cloud forest. Sampling was conducted in January 2016, in a total of 14 plots including cloud forests above 800 m, and slope and plateau forests at lower elevation ranges (i.e. 440 to 635 m).
- The Mitaraka range was sampled in 2015 around a camp (lon / lat: -54.4503 / 2.2340) set up temporarily for this occasion (Figure 1). The massif is part of the Tumuc-Humac mountain chain located at the extreme South-West of French Guiana at the border to Brazil and Surinam. The landscape is classified as “high hills and mountains” (Guitet et al., 2013) and is characterized by the presence of massive inselbergs (isolated granite rock blocks that range several hundred meters above the lowland areas) and lowland forest. Sampling was conducted in 19 plots representing the complete range of habitats targeted for this study.
- The Trinité Natural Reserve was sampled around the Aya Camp (lon / lat: -53.4132 / 4.6024), which is located in the vicinity of a large isolated inselberg with an altitude of 501 m (Figure 1). Sampling took place in 2016 in 11 plots comprising swamp, slope and plateau forests, as well as hilltop and transition forests of the inselberg.
- Two locations were sampled in the Nouragues Natural Reserve in January and June 2011 around the two permanent research stations present in the area: Pararé station (lon / lat: -52.6730 / 4.0381) and Inselberg station (lon / lat: -52.6800 / 4.0883), both 6.4 km away from each other (Decaëns et al., 2016) (Figure 1). The former is located along the Arataye River, with vegetation dominated by lowland swampy forest, while the latter is situated at

the foot of an isolated inselberg culminating at 411 m. A total of 41 plots were sampled, including inselberg habitats, plateau, slope and swamp forests (see Decaëns et al., 2016 for details).

2.2. Sampling (earthworms and soil)

In each 1 ha plot, earthworms were collected by combining three different approaches (protocol adapted from Decaëns et al., 2016): (1) Quantitative sampling (TSBF) by digging and hand-sorting three blocks of soil, each 25 x 25 x 20 cm (length x width x depth), located at the interior angles of a 20 m equilateral triangle in the center of the sampling plot ; (2) Qualitative sampling by digging and hand-sorting an area of 1 m² with a minimum depth of 20 cm, selecting an area with large earthworm casts (when available) within the sampling plot ; (3) Micro-habitat sampling by visually inspecting all available micro-habitats (such as sandy to muddy sediments of stream banks, leaf litter accumulations, decaying trunks and epiphytic soils) for three researcher-hours (e.g. one hour for three people) within the entire sampling plot. Earthworms of all life stages (adults, juveniles and cocoons) were collected and kept in 95 to 100 % ethanol. Ethanol was changed after 24 hours to ensure clean fixation.

For soil analyses, ten soil cores per plot were collected at 0–10, 10–20 and 20–30 cm depth, each 20 cm along a transect through the sampling plot at Galbao, Itoupé, Mitaraka and Trinité (Vleminckx et al., 2019). Cores were combined into a composite 500 g sample, which was dried at 25 °C and sieved at 2 mm. Physical and chemical analyses were done at CIRAD Laboratory (Montpellier, France, ("Unité de service Cirad - Analyses," n.d.)) with protocols available on their website (Pansu and Gautheyrou, 2007). Measured variables were soil texture (clay, fine silt, coarse

184 silt, fine sand and coarse sand), pH H_2O, organic carbon, total nitrogen, C/N ratio and available
185 phosphorus (Supplementary Table 1).

186 We also collected nine other climatic variables and soil properties from global databases. We
187 used seven climate layers from the CHELSA climate database (Karger et al., 2017) corresponding
188 to temperature and precipitation variables (annual mean temperature, temperature seasonality,
189 temperature annual range, annual precipitation and precipitation seasonality). In addition we
190 averaged the values of the first 30 cm for bulk density and cation exchange capacity (CEC)
191 obtained from the SoilGrids database (Hengl et al., 2017).

192

193 2.3. DNA barcoding

194 Earthworms specimens were sorted into morpho-groups (i.e. groups of individuals of similar
195 size, pigmentation and general external morphology) as a conservative approximation of the
196 taxonomic diversity in a sample. We did not attempt to group into the same morphospecies
197 immature stages (i.e. cocoons or juveniles) and adults, because the former usually lack any reliable
198 character to link them with the corresponding adults. Consequently, cocoons, juveniles and adults
199 were systematically assigned to different morphospecies even when obviously belonging to the
200 same species. Subsequently, up to five specimens per morpho-group for each sample, and all the
201 cocoons and fragments, were selected for DNA barcoding. A small piece of cutaneous tissue
202 (about 1 mm^2), or of the embryo in case of cocoons, was fixed in ethanol (100 %) and stored at -
203 20 °C before DNA extraction.

204 Lab work followed standardized protocols for DNA extraction, barcode amplification and
205 sequencing (deWaard et al., 2008). DNA was extracted using a glass-fiber column based protocol
206 (Ivanova et al., 2006). The primer cocktail C_LepFolF and C_LepFolR (Hernández Triana et al.,

2014) was used to amplify a 658 bp fragment of the COI gene. The PCR thermal regime consisted of an initial denaturation at 94°C for 1 min; five cycles at 94°C for 1 min, 45 °C for 1.5 min and 72 °C for 1.5 min; 35 cycles of 94°C for 1 min, 50 °C for 1.5 min and 72 °C for 1 min followed by a final cycle at 72 °C for 5 min. Each PCR product was cleaned up using Sephadex (Ivanova and Grainger, 2007). PCR amplicons were visualized on a 1.2% agarose gel E-Gel® (Invitrogen) and then diluted 1:10 with sterile water. Amplicons (2–5 µL) were bidirectionally sequenced using sequencing primers M13F or M13R (Messing 1983) and the BigDye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) on an ABI 3730xl capillary sequencer following manufacturer's instructions. All sequences and supporting information obtained for Galbao, Itoupé, Mitaraka and Trinité were combined with the ones obtained from samples taken at the Nouragues stations during a previous study (Decaëns et al., 2016) and deposited in the Barcode of Life Datasystems (BOLD) database (Ratnasingham and Hebert, 2007) in the dataset "Earthworms from the tropical rainforest of French Guiana" (DS-EWFG, DOI: dx.doi.org/10.5883/DS-EWFG).

2.4. Delimitation of the Molecular Operational Taxonomic Units (MOTUs)

MOTUs delimitation was done using the Automatic Barcode Gap Discovery method (ABGD, Puillandre et al., 2012). We did not use the BIN (Barcode Index Number) delimitation system from BOLD (Ratnasingham and Hebert, 2013) because as shown by Decaëns et al. (2016) who compared different methods, this one seems to not be suitable for the case of earthworms which presents a higher divergence rate compare to other groups. In a first step we used only the sequences from Nouragues and an *a priori* threshold of 12 - 14 %, to verify recovery of the 48 MOTUs found in the original study by Decaëns et al. (2016). The ABGD parameters obtained in this first run (p = 0.05, P = 0.2, relative gap width X = 0.5, distance d = K80) were subsequently

applied to the full dataset comprising sequences from all locations in order to delimited new MOTUs.

2.5. Data analyses

Data from Nouragues locations were only used to delimit the MOTUs and assess the diversity at the regional and local scales. All other following analyses were only performed for Galbao, Itoupé, Mitaraka and Trinité, as a complete dataset including soil properties was available only for these four locations and because only qualitative sampling has been performed at both Nouragues location in 2016. Also, under the category "hilltop forests" we grouped the inselberg-like habitats (i.e. hilltop and transition forests and rocky savannah) because they shared some soil characteristics and for many of this three habitats, the number of replicates was not enough to analyse them separately.

2.5.1 Alpha- to gamma-diversity estimates

To compare species diversity among different localities, habitats and microhabitats, we adjusted rarefaction and extrapolation curves for MOTU diversity using the "iNEXT" package (Hsieh et al., 2016) for R (R Core Team, 2020). At the regional scale, we used the observed and extrapolated number of MOTUs according to the number of sampling locations as a measure of sampling effort; whereas at local scale we used the number of specimens collected to account for the variability in earthworm density among habitats and microhabitats. We further computed observed richness, defined as the number of different MOTUs observed for each locality, habitat or micro-habitat, as well as the estimated species richness (Chao estimate), using the R package "vegan" (Oksanen et al., 2019).

Also, we looked at the diversity captured by each of the three sampling methods and in each kind of microhabitats, and by drawing Venn diagrams using the package "eulerr" (Larsson, 2020). This enabled us to highlight the proportion of MOTUs shared between sampling methods and microhabitats.

2.5.2. Beta-diversity analyses

We calculated the average Sorensen's index of dissimilarity using the R package "betapart" (Baselga et al., 2018) to assess the variation of MOTU composition among sites (β-diversity) and its decomposition into spatial turnover (i.e. replacement of species from location to location) and nestedness (i.e. when the composition of a site is a subset of another site hosting more species) (Baselga, 2010). This was done at three different levels to comprehensively characterise the relative contribution of distance and habitat diversity to β-diversity: (1) by adopting the same approach for each habitat separately (i.e. *between localities + within habitat* by comparing with each other the four species lists found in each locality for a given habitat); (2) the ecological β-diversity by comparing the composition of different habitat species pools within each locality (i.e. *within locality + between habitats*); (3) the local β-diversity by comparing the composition of individual communities collected in a given habitat at a given locality (i.e. *within locality + within habitat)*.

Singletons are MOTUs represented by only one individual in the dataset, and, because they are by definition present only in a single location, they are expected to inflate the indices of β-diversity when present at a high proportion. To account for this potential bias, we performed all analyses with and without singletons. As we could not find any significant differences, we decided to keep singletons present in the analyses presented herein.

276

277 2.5.3 Environmental drivers of community composition

278 Data were organized into two separate tables to perform a transformation based canonical
279 redundancy analysis (tb-RDA, Legendre and Gallagher, 2001), in order to highlight the
280 environmental parameters explaining the observed variations in earthworm community
281 composition. This analysis was done using the *rda* function of the R package "vegan" (Oksanen
282 et al., 2019) for a subset of 32 sampling plots from Galbao, Itoupé, Mitaraka and Trinité,
283 representing replicated habitats (mainly hilltop, plateau, slope and swamp forests) for which soil
284 variables were available (Supplementary Table 1). We used for this abundance data as a
285 contingency community table composed of 32 rows (i.e. the 1 ha sampling plots, see
286 Supplementary Table 1) and 81 columns (i.e. the MOTUs), and another table with the same 32
287 rows which contained the explicative variables grouped as spatial, soil texture, soil chemistry and
288 climatic variables (see Supplementary Table 2 for detail). MOTU abundance data were Hellinger-
289 transformed before computing the tb-RDA to reduce the weight of the most abundant groups in
290 the analyses. After removing the variables that were correlated based on Pearson correlation, we
291 used the function *ordiR2step* to select the explanatory variables that contribute significantly to the
292 model and permutation tests to verify the significance of the RDA model obtained. Finally, we
293 looked at the relative contribution of the different groups of explicative variables (spatial / soil
294 texture / soil chemistry) using a partial RDA ordination with the function *vpart* of the "vegan"
295 package (Borcard et al., 2018).

296

297

298

3. Results

3.1. Barcoding results and MOTUs designation

A total of 1819 earthworm specimens out of 55 sampling points from the four localities of Galbao, Itoupé, Mitaraka and Trinité were selected for DNA barcoding. We were able to obtain 1683 COI sequences (after removal of contaminated sequences), with a sequencing success of 92.52 %.

After adding the dataset from the Nouragues (Decaëns et al., 2016) and removing the sequences shorter than 300 bp, our total dataset contained 2304 sequences (409 from Galbao, 595 from Itoupé, 347 from Mitaraka, 324 from Trinité, 431 from Nouragues-Inselberg and 198 from Nouragues-Pararé) clustering into 119 MOTUs when using a 13 % threshold on ABGD (Figure 2). The dataset comprised 821 adults (35.6 %), 1 276 juveniles (55.4 %), 119 cocoons (5.2 %) and 88 fragments (3.8 %). Mean intra-MOTU divergence was 2.12 % (ranging from 0 % to 9.26 %) and mean inter-MOTU was 24.02 % (ranging from 10.56 % to 49.98 %) (Figure 2). We found 30 MOTUs (25.2 % of the total number) solely represented by juveniles, cocoons and specimen fragments. In total, 24 MOTUs (20.2 % of the total number) were singletons, 13 of which were represented only by juveniles and one by specimen fragments.

3.2. Earthworm diversity at regional scale

There were only a few shared MOTUs between the sampled locations; only two (1.7 %) MOTUs were shared between the six locations and 85 (71.4 %) were present at a single location (Figure 3). Localities that shared the most MOTUs also seemed to be closer geographically, such as both Nouragues locations (~ 5.60 km) that shared 18 MOTUs or Nouragues Inselberg and Galbao (~ 87 km) which shared 10 MOTUs (Figure 3B). However, even distant localities such as

322 Trinité and Mitaraka (~ 285 km) can share five MOTUs (Figure 3B). As a consequence, the
323 rarefaction and extrapolation curve fitted for the full dataset at regional scale shown a sharp
324 increase of MOTU counts with increasing numbers of sampling locality (Figure 4). There was no
325 evidence for any saturation of the regional species pool, even when extrapolating the number of
326 MOTUs that would result from doubling the sampling effort, and asymptotic richness estimates
327 suggested that more than 250 species could occur at this scale (Table 1). Our results therefore
328 indicated that the 119 observed MOTUs that we found during our survey might represent only less
329 than half (45.4 %) of the real number of earthworm species that may exist in the entire French
330 Guiana. However, there was a large uncertainty for this estimate (SD = 45.7).
331 The Sorensen indices of regional beta-diversity (β_{SOR}, *between locality* + *within habitat*) were
332 high and similar when comparing different forest habitats, showing a strong spatial turnover (β_{SIM})
333 at this scale (Supplementary Figure 1).
334
335 3.3. Earthworm diversity at local scale
336 Rarefaction and extrapolation curves for individual locations shown how MOTUs accumulate
337 as a function of the number of sampled individual (Figure 5). The observed richness ranged from
338 25 MOTUs in Trinité to 39 in Nouragues-Inselberg (Table 1). For almost all locations, richness
339 estimates appeared close to the observed values, except for both Nouragues locations where the
340 estimated richness was 1.3 times higher than the observed total richness. The standard error of the
341 Chao index for both Nouragues locations was also higher than for the four other locations. These
342 observations reflected the trend that the slope of the rarefaction curves of Nouragues Inselberg and
343 Pararé seemed more pronounced than that of the slopes from other localities. However, none of
344 the localities seemed to reach an asymptote.

345 The Sorensen indices of beta-diversity (β_{SOR}) calculated at local scale among the different
346 habitats inside each of the four new locations (*within habitat + within locality*) were fairly variable
347 (Supplementary Figure 2), but all showed a partitioning in favour of spatial turnover (β_{SIM}) in
348 comparison to nestedness (β_{SNE}). Overall, the ecological beta-diversity (β_{SOR} *within locality +*
349 *between habitat*) among different habitats was high and similar when comparing the different
350 localities, ranging from 0.62 from Galbao to 0.69 in Mitaraka (Supplementary Figure 3). However,
351 it was explained in greater part by nestedness in Itoupé (0.44), and by turnover in Galbao, Mitaraka
352 and Trinité (0.55, 0.51 and 0.44 respectively).
353
354 3.4. Earthworm diversity at habitat level
355 Several habitats were sampled with different levels of richness. Overall, at the regional scale,
356 plateau, slope and swamp forests seemed to harbor a similar diversity (with respectively 55, 54
357 and 55 MOTUs) higher than the diversity in hilltop forest (29 MOTUs). The extrapolation did not
358 show sign of saturation at the regional scale. However, these observed trends were not necessarily
359 conserved at the local scale, where the differences in diversity between habitats could be more
360 pronounced and for certain localities the rarefaction curves for some habitats seem to reach a
361 plateau. Indeed. a higher richness was observed for the hilltop forest in Galbao, the slope forest in
362 Itoupé, the plateau forest in Mitaraka and the swamp forest in Trinité and both Nouragues localities
363 (Supplementary Figure 4). However, for Itoupé, Trinité and Nouragues Pararé, this was
364 confounded by the number of individuals sampled. The rarefaction curves of some habitats, such
365 as the plateau forest in Galbao and Nouragues Inselberg and the slope forest in Itoupé, almost
366 reached an asymptote with a narrow standard error (Supplementary Figure 4). However, at this
367 scale, a substantial part of the MOTUs are not shared between habitats of the same location.

368

369 3.5. Earthworms diversity at microhabitat level and sampling comparison

370 On average, between 11.2 and 46.5 earthworms per meter square (SD: 28.4 – 54.6) were
371 collected per sampling plot with the TSBF method only depending on the locality, with Mitaraka
372 showing the lowest and Trinité the highest abundance (Supplementary Figure 5A). The richness
373 of MOTUs recovered with this method (TSBF) was also quite variable, as it ranged on average
374 from 0.3 in Mitaraka to 1.3 on Galbao (SD: 0.6– 1.1), with a regional mean of 0.9 (SD = 1.2) per
375 sampling plots (Supplementary Figure 5B). However, when using all sampling methods combined,
376 we found a higher average of richness (Supplementary Figure 5B) at local scale. The regional
377 mean richness was 6.1 (SD = 3.1) all sampling methods combined (Supplementary Figure 5B).
378 This trend is also observable at the habitat scale (Supplementary Figure 5 C & D).

379 Overall, the qualitative sampling approach allowed the collecting of roughly twice the number
380 of MOTUs that was recovered by the quantitative TSBF method alone (Figure 7A). All but six of
381 the 90 MOTUs found at Galbao, Itoupé, Mitaraka and Trinité were collected by qualitative
382 sampling, while the quantitative sampling only resulted in the finding of 42 MOTUs (Figure 7A).
383 Most of the diversity recovered by all sampling methods was in the soil with 69 MOTUs, but 52
384 MOTUs were however found in other types of microhabitats (Figure 7A). Also, among the 119
385 MOTUs found with the full dataset, 23 were exclusively found in other microhabitats than soil
386 (mostly decaying trunk) (Figure 7B).

387

388 3.6. Community composition

389 The climatic variables, as well as bulk density and cation exchange capacity, were not retained
390 in the RDA analysis because they resulted to be highly correlated with the spatial variables (i.e.

391 elevation and precipitation). The *ordiR2step* function selected all spatial variables (topography,
392 longitude, latitude and elevation), fine silt content, pH, total nitrogen, organic carbon and
393 phosphorus as subset of explanatory variables to explain earthworm community composition
394 (Supplementary Table 2).

395 Communities in Galbao, Itoupé, Mitaraka and Trinité were generally well separated by the first
396 four RDA axes, as well as the forest types within each location (Figures 8). With the exception of
397 swamp forests from Galbao, Mitaraka and Trinité and plateau and hilltop forests from Galbao and
398 Trinité that could be observed close together on the ordination space, forest types between
399 localities rarely seemed to share similarities in their composition and environmental properties.
400 The first two and first four canonical axes together explained 22.6 % and 36.4 % respectively of
401 the total variance of the response data, with an adjusted R^2 of 0.3697. The variable elevation played
402 an important role in the distribution of the sampling plots along the first axis (Figure 8A). Higher
403 elevation values were observed at Galbao and Itoupé, and the lowest at Trinité. The variables fine
404 silt content and hilltop topography were correlated with the second axis (Figure 8A). Soils at
405 Galbao contained high silt content. The pH was correlated with the third axis. And, the fourth axis
406 opposed the latitude and longitude to organic carbon, total nitrogen and phosphorus (Figure 8C
407 and 8E). Soils at Mitaraka showed highest content of chemical variables and Galbao the lowest.

408 The partial RDA showed that spatial variables (elevation, longitude, latitude and topography)
409 explained 19.2 % of the variance in the earthworm community composition, while soil chemical
410 variables (pH, organic carbon, total nitrogen and phosphorus) and soil texture (fine silt content)
411 accounted for 10.8 % and 6 % respectively. The remaining 63 % corresponded to the residuals,
412 i.e. the fraction of the overall variance that was not explained by our selected environmental
413 variables. This meant that the model might still displayed some dominant residual structure.

4. Discussion

4.1. Unmatched levels of earthworm diversity in French Guiana

To date, only a few publications have addressed the diversity of earthworms at regional scale in French Guiana and the neotropical region overall. In 2007 Brown and Fragoso listed 33 species on the French Guiana territory, but 12 of them have not been described. In 2012, Pavlicek & Csuzdi added one new species to the 21 previously described species from Brown and Fragoso (2007), while suggesting that this figure probably reflected nothing else than the extent of the lack of knowledge on the subject. In 2016, Decaëns et al. published a study based on an approach similar to ours, in which they described the local distribution of 48 MOTUs in the Nouragues reserve. Our results represent a significant step forward in the acquisition of knowledge on this subject, as we were able to detect 119 MOTUs in six study locations distributed all over the region. Furthermore, the rarefaction curve and richness estimates computed at the regional scale both indicate we still only recovered ~ 45 % of the true regional diversity, and that any newly added sampling location could lead to the addition of about 15 to 30 new MOTUs.

The mean observed diversity of 32.2 MOTUs (SD = 9.1), potential species, per location is one of the highest reported for earthworm communities in tropical forests. For comparison, 10 to 17 species have been reported at a similar spatial scale in the native forests of Santa Catarina and Rio Grande do Sul, Brazil (Bartz et al., 2014; Steffen et al., 2018), and in the Mexican tropical rainforest of Chiapas (Fragoso and Lavelle, 1987) using s traditional taxonomy approach. We also calculated a mean of 2.7 species (SD = 3.7) per sampling plots with the TSBF sampling method which seems to at least double what would be expected in this region according to the model used in Phillips et al. (2019) (about 1 species in French Guiana / Northern Amazonia).

437 This discrepancy could be related to the adding value of the sampling methodology we adopted
438 for our earthworm surveys. In earthworm studies, only adult specimens are typically used for
439 species richness assessment, mostly because of the difficulty to identify juveniles to species level.
440 Contrary to this classical approach, we included in our study juveniles, cocoons and fragments of
441 earthworms that represented the majority (64 %) of our dataset. Without them, we would have
442 work with about 36 % of our dataset and would have missed a quarter of the regional diversity (i.e.
443 30 MOTUs) represented by MOTUs only present in the samples as immatures or fragmentary
444 remains. This stresses the importance of an integrative approach to species richness assessment
445 that includes sampling of all life stages and the use of a molecular identification method such as
446 DNA barcoding (Decaëns et al., 2016; Richard et al., 2010).
447 An additional explanation for our high diversity compared to the model from Phillips et al.
448 (2019) is that previously published studies generally used soil hand sorting (TSBF) as the only
449 quantitative sampling method, providing only a partial picture of the composition of earthworm
450 communities. We, on the other hand, coupled three sampling methods allowing us to prospect not
451 only the soil, but also other types of microhabitats, thereby increasing the number of species that
452 we were able to detect locally. Indeed, our approach allowed us to collect twice as many species
453 as if we had only used quantitative sampling.
454 Even more interesting is the discovery of 23 MOTUs that were sampled exclusively in
455 microhabitats other than soil (decaying trunks and epiphytes). Arboriculous earthworm species
456 have been observed in tropical regions (Fragoso and Rojas-Fernández, 1996; Lavelle, 1978;
457 Lavelle and Kohlmann, 1984; Rodriguez et al., 2007), and Decaëns et al. (2016) already
458 documented that as much as 35 % of the total number of species observed in both Nouragues
459 locations may occur at least occasionally in epiphytic soils. In oligotrophic soils of the neotropics,

earthworm communities are often dominated by pigmented earthworms that prefer microhabitats where organic matter is concentrated including decaying trunks (Fragoso and Lavelle, 1992; Decaëns et al., 2016). This suggests that a significant part of the earthworm diversity in tropical regions with oligotrophic soils could live in aboveground habitats. Overall, the use of this sampling scheme appeared to be very efficient in discovering and describing earthworms richness in tropical region, as it has been previously highlighted in the context of a study of earthworm richness in agroecosystems in Southern Brazil (Bartz et al., 2014).

4.2. Few earthworm species with large geographical ranges

Only two MOTUs (#26 and #28) were present in the six sampled locations, showing a very low level of shared diversity at regional spatial scale. These two MOTUs were also the most abundant species represented by 368 (16 %) and 215 (9.3 %) individuals respectively. MOTU #26 has been identified as *Pontoscolex corethrurus* (Muller, 1856), a peregrine endogeic species that originated from the Guyana shield (a geological formation in northeastern South America that extend over Guyana, Suriname, French Guiana, southern Venezuela, as well as parts of Colombia and Brazil), and which is known to be invasive in a number of other tropical countries (Dupont et al., 2012; Marichal et al., 2010; Taheri et al., 2020). Therefore, it was not surprising that this species has a large range in its own origin area. MOTU #28 has been identified as *Nouraguesia parare* (Csuzdi and Pavlíček, 2011), a large epigeic species supposedly endemic from French Guiana and that is mostly (54 % of the times in this study) found in rotten trunks (Decaëns et al., 2016). These two species have an opposite nature, one being invasive and the other one endemic, but both in their original range show a large spatial distribution. This could certainly be explained by a high dispersal capacity of these two species, allowing them to colonise new habitats more

483 efficiently than others. This has been already described for *P. corethrurus*, which is known to
484 disperse passively with human activities, making it a formidable colonizer regardless of the
485 distance to be traveled (Dupont et al., 2012). In the case of *N. parare*, it is likely that its large size
486 and surface-dwelling behaviour also allows it to move actively over long distances.

487 Two species from the genus *Wegeneriona*, which is native of South America, were present in
488 the five locations of Galbao, Itoupé, Nouragues Inselberg, Nouragues Pararé plus Mitaraka
489 (MOTUs #12) or Trinité (MOTUs #21). *Dichogaster andina* (Cognetti de Martiis, 1904) (MOTUs
490 #35) was present in the four locations of Nouragues Inselberg and Pararé with Trinité and
491 Mitaraka. This is another invasive species, originating from Africa (Csuzdi et al., 2008). Invasive
492 species are not rare in French Guiana (Lavelle and Lapied, 2003) and even if its presence was
493 already surprising in the Nouragues locations (Decaëns et al., 2016), finding it at locations easily
494 and frequently visited by human could be explained by recent introduction events. However, its
495 presence in a region as remote as Mitaraka was unexpected. This can only be explained by older
496 traces of human activity such as trade between the Amerindian populations who inhabited this area
497 for example, or when the Wayana Indians, who currently inhabit the lower half of the Maroni
498 River and are native to the Caribbean Sea, moved upstream the river a few centuries ago, pushing
499 local tribes towards the Mitaraka Mountains (Fleury et al., 2016).
500
501 4.3. Outstanding levels of geographical turnover among earthworm communities
502 About 1/5 of all MOTUs (20.2 %) were singletons, only two MOTUs were shared between all
503 six locations and 83 MOTUs (69.8 %) were present in only one specific location. 34 MOTUs
504 (28.57 %) were shared between two to five locations, sometimes between geographically distant
505 localities and not the direct close locality. This could indicate both a high level of endemism at

506 regional scale with the presence of a significant amount of rare species, and/or a signal of
507 undersampling. Earthworms are known to exhibit higher rates of endemism compared to some
508 other invertebrates groups composing Amazonian biodiversity (Lavelle and Lapied, 2003). French
509 Guiana in particular is characterized by a high coverage of primary forest and a large water network
510 including 840 rivers stretching over a total distance of 112,000 km ("L'office de l'Eau de Guyane,"
511 n.d.). These rivers are often large enough to easily become geographic barriers to earthworm
512 dispersion, leading to the formation of isolated populations and increasing the likelihood of local
513 radiation events as it has been also shown for other taxa (Boubli et al., 2015; Bruschi et al., 2019;
514 Siqueira et al., 2013). This could therefore explain the different species pools that we observed in
515 each study location, and the importance of the spatial turnover component of regional beta
516 diversity.

517 At local scale, we also found significant levels of spatial turnover, but these were quite variable
518 among habitat. However, they all show that spatial turnover due to MOTU replacement, rather
519 than nestedness, is responsible for this local beta-diversity. Earthworm diversity is known to be at
520 least partly driven by environmental heterogeneity, as previously shown in Mexico (Fragoso and
521 Lavelle, 1987). When looking at ecological scale, our Sorensen indices between locations are very
522 similar indicating comparable variation of composition of different habitat species pools, with
523 Mitaraka showing the highest one. In contrast, the beta-diversity in Itoupé was mostly explained
524 by nestedness and not turnover. Mitaraka harbor large swamp forests that may represent adverse
525 habitats for most of the species occurring in non-flooded forests. Conversely, sampling at Itoupé
526 took place on different altitudinal levels of a single slope of the tabular mountain, with sampling
527 plots relatively close to each other, and not separated by contrasted habitats or large water barriers,
528 when compared to plots in the other locations. These characteristics of these sampling locations

529 could explain the differences observed between them and why we observed a higher nestedness in
530 the ecological beta-diversity at Itoupé.
531
532 4.4. Spatial and environmental drivers of earthworm community composition
533 The spatial effect on species composition was also confirmed by the site ordination in our
534 RDA, where each location was represented by a well-resolved cluster as a consequence of spatial
535 turnover. Habitat replicates inside locations were also quite well clustered depending on the axis,
536 showing that there were also environmental influence happening at this scale. However, even if
537 some type of forests seems to share some similarities, we did not observe a regional pattern as
538 same forest types were grouped inside a location but not between locations. The variation
539 partitioning showed that the spatial variables (longitude, latitude, elevation and topography)
540 explained 19.2 % of the variation in species composition which support the importance of spatial
541 turnover being the most important driver. Then the pH, silt content, as well as organic carbon and
542 total nitrogen also significantly contributed at explaining the species composition of earthworm
543 communities. As a result, the spatial variables must play an important role at the regional scale in
544 the variation of the earthworm community composition as shown by the RDA and beta-diversity.
545 And, the environmental variables must play an important role at a lower scale with gradients of
546 organic carbon and silt content as it has been shown before (Fragoso and Lavelle, 1992) and as
547 some of our RDA results suggest, but more sampling points are needed at local scale. It has also
548 been previously shown that temperature and precipitation determine the structure (species richness
549 and abundance) of earthworm communities (Fragoso and Lavelle, 1992; Phillips et al., 2019).
550 While this might be true at global scale, these factors are perhaps less relevant for environments
551 such as tropical rainforests where temperature remains quite constant and where annual

552 precipitation is in the range of 2,000 - 4,000 mm. Here, precipitation and temperature variables
553 were strongly correlated with the elevation and longitude / latitude, so it was hard to separate the
554 effect of each. All our results strongly agreed on the effect of spatial variables, however, increasing
555 the number of sampling plot at local scale would help to investigate the role of environmental and
556 climatic variables in structuring earthworm communities.

557 Our partial RDA reveled an important part of residual effect. We already mentioned the
558 potential effect of hydrographic network with mechanisms of local radiation in evolutionary
559 history of earthworm communities, generating high spatial turnover in communities' composition,
560 that could be part of these residual effect. Some other factors not taken into account in this study
561 could also be considered to explain the composition of earthworm communities at different scales.
562 Such as, the surface vegetation, and in particular the litter traits that are known to have a strong
563 explanatory power on the composition of the soil fauna communities in general (Korboulewsky et
564 al., 2016). But also interaction with other soil taxa such as bacteria and fungi could also play a
565 role, and these variables could reduce the residual effect.

566

567 **5. Conclusion**

568 We found a remarkable high regional species richness of earthworms in French Guiana with a
569 high proportion of rare and endemic species and a relatively low (from what would be expected in
570 tropical region compare to temperate region) local species richness. These strong levels of beta-
571 diversity seem to support the remarkable regional diversity observed in French Guiana. And, at
572 the same time, even if our estimates did not converge with those exposed by Phillips et al. (2019),
573 mainly explained by our improved sampling approach, we agreed to conclude that the explosion
574 of species diversity in equatorial ecosystems is verified at the regional scale but not at local scale.

575 Our study confirms the already mentioned usefulness of DNA barcoding to assess the diversity of
576 understudied invertebrates such as earthworms, especially in areas harbouring high diversity such
577 as the tropics. This method allows a fast investigation of the diversity at local and regional level
578 with MOTU clusters as useful species proxy, especially in the absence of further taxonomic
579 information and when investigating earlier life stages and fragmentary remains. However, as
580 mention earlier in the introduction, the use of barcoding is still limited in the study of tropical
581 earthworm communities, resulting in a lack of molecular taxonomic data of tropical earthworm
582 species. Also, the use of a single mitochondrial fragment is not sufficient and more complete data
583 at the genomic level are needed to refine the identification of molecular species. Our study also
584 highlights the importance of sampling design. The inclusion of three sampling methods including
585 the investigation of non-soil microhabitats greatly increased the assessed diversity. On the basis of
586 our results we expect that each additional sample location with different types of habitat would
587 detect 20 to 30 new MOTUs. These results are also the premises showing the richness of
588 earthworm communities in the netropical region.

589 Further analyses could be performed to investigate the historical and environmental processes
590 leading to the observed spatial patterns. Our results highlight a strong spatial turnover mechanism
591 in the Amazonian region, and the use of functional trait and/or phylogenetic approaches could
592 bring more insights on the assembling rules of earthworm communities in Amazonia. Indeed, it
593 will be interesting to see if the predominant spatial turnover responsible for the variation in the
594 taxonomic beta-diversity described here and in the previous study would also be responsible for
595 the functional and/or phylogenetic beta-diversity.

596

597

598 **Acknowledgements**

599 We are grateful to the Parc Amazonien de Guyane (http://www.parc-amazonien-guyane.fr) and
600 the Réserve Naturelle de la Trinité (http://www.reserve-trinite.fr/) for authorizations and access to
601 the sampling sites. This work was supported by 'Investissement d'Avenir' grants managed by the
602 Agence National de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-41).
603 In the Mitaraka region, material was collected as part of "Our Planet Reviewed" Guyane-2015
604 expedition, which was organized by the Muséum national d'Histoire naturelle (MNHN, Paris) and
605 Pro-Natura international in collaboration with the Parc Amazonien de Guyane, with financial
606 support of the European Regional Development Fund (ERDF), the Conseil régional de Guyane,
607 the Conseil général de Guyane, the Direction de l'Environnement, de l'Aménagement et du
608 Logement and the Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la
609 Recherche. Additional funding was provided by the Trinité Natural Reserve. At the Nouragues,
610 the project was supported by two grants CNRS Nouragues' 2010 and 2011. This study was
611 supported by funding through the Canada First Research Excellence Fund. The funders had no role
612 in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.
613 This work represents a contribution to the University of Guelph Food From Thought research
614 program.

615

616 **References**

617 Amat, J.-P., Dorize, L., Gautier, E., 2008. Éléments de géographie physique. Editions Bréal.
618 Anderson, J.M., Ingram, J.S.I., 1989. Tropical soil biology and fertility. CAB international
619 Wallingford.

André, H.M., Ducarme, X., Anderson, J., Crossley Jr, D., Koehler, H., Paoletti, M., Walter, D., Lebrun, P., 2001. Skilled eyes are needed to go on studying the richness of the soil. Nature 409, 761.

André, H.M., Noti, M.-I., Lebrun, P., 1994. The soil fauna: the other last biotic frontier. Biodivers. Conserv. 3, 45–56. https://doi.org/10.1007/BF00115332

Bardgett, R.D., Putten, W.H. van der, 2014. Belowground biodiversity and ecosystem functioning. Nature 515, 505–511. https://doi.org/10.1038/nature13855

Bartz, M.L.C., Brown, G.G., da Rosa, M.G., Filho, O.K., James, S.W., Decaëns, T., Baretta, D., 2014. Earthworm richness in land-use systems in Santa Catarina, Brazil. Appl. Soil Ecol., XVI International Colloquium on Soil Zoology & XIII International Colloquium on Apterygota, Coimbra, 2012 – Selected papers 83, 59–70. https://doi.org/10.1016/j.apsoil.2014.03.003

Baselga, A., 2010. Partitioning the turnover and nestedness components of beta diversity. Glob. Ecol. Biogeogr. 19, 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

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

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 1935–1943. https://doi.org/10.1098/rstb.2005.1725

Borcard, D., Gillet, F., Legendre, P., 2018. Numerical Ecology with R. Springer.

Boubli, J.P., Ribas, C., Lynch Alfaro, J.W., Alfaro, M.E., da Silva, M.N.F., Pinho, G.M., Farias, I.P., 2015. Spatial and temporal patterns of diversification on the Amazon: A test of the riverine hypothesis for all diurnal primates of Rio Negro and Rio Branco in Brazil. Mol.

643 Phylogenet. Evol., Phylogeny and Biogeography of Neotropical Primates 82, 400–412.
644 https://doi.org/10.1016/j.ympev.2014.09.005
645 Brown, G.G., Fragoso, C., 2007. Minhocas na América Latina: biodiversidade e ecologia.
646 Londrina: Embrapa Soja, 2007.
647 Bruschi, D.P., Peres, E.A., Lourenço, L.B., Bartoleti, L.F. de M., Sobral-Souza, T., Recco-
648 Pimentel, S.M., 2019. Signature of the Paleo-Course Changes in the São Francisco River
649 as Source of Genetic Structure in Neotropical Pithecopus nordestinus (Phyllomedusinae,
650 Anura) Treefrog. Front. Genet. 10. https://doi.org/10.3389/fgene.2019.00728
651 Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., Gottschall, F.,
652 Guerra, C.A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Ferlian, O., Kreft,
653 H., Lovejoy, T.E., Montanarella, L., Orgiazzi, A., Pereira, H.M., Phillips, H.R.P., Settele,
654 J., Wall, D.H., Eisenhauer, N., 2019. Global mismatches in aboveground and
655 belowground biodiversity. Conserv. Biol. 33, 1187–1192.
656 https://doi.org/10.1111/cobi.13311
657 Csuzdi, C., Pavlíček, T., 2011. A new earthworm genus Nouraguesia gen. nov. from French
658 Guiana with description of two new species (Oligochaeta, Glossoscolecidae). J. Nat. Hist.
659 45, 1759–1767. https://doi.org/10.1080/00222933.2011.560727
660 Csuzdi, Cs., Pavlíček, T., Nevo, E., 2008. Is Dichogaster bolaui (Michaelsen, 1891) the first
661 domicole earthworm species? Eur. J. Soil Biol. 44, 198–201.
662 https://doi.org/10.1016/j.ejsobi.2007.05.003
663 Decaëns, T., 2010. Macroecological patterns in soil communities. Glob. Ecol. Biogeogr. 19,
664 287–302. https://doi.org/10.1111/j.1466-8238.2009.00517.x

665 Decaëns, T., Porco, D., James, S.W., Brown, G.G., Chassany, V., Dubs, F., Dupont, L., Lapied,
666 E., Rougerie, R., Rossi, J.-P., Roy, V., 2016. DNA barcoding reveals diversity patterns of
667 earthworm communities in remote tropical forests of French Guiana. Soil Biol. Biochem.
668 92, 171–183. https://doi.org/10.1016/j.soilbio.2015.10.009
669 deWaard, J.R., Ivanova, N.V., Hajibabaei, M., Hebert, P.D.N., 2008. Assembling DNA
670 Barcodes, in: Martin, C.C., Martin, C.C. (Eds.), Environmental Genomics, Methods in
671 Molecular Biology. Humana Press, Totowa, NJ, pp. 275–294.
672 https://doi.org/10.1007/978-1-59745-548-0_15
673 Dupont, L., Decaëns, T., Lapied, E., Chassany, V., Marichal, R., Dubs, F., Maillot, M., Roy, V.,
674 2012. Genetic signature of accidental transfer of the peregrine earthworm Pontoscolex
675 corethrurus (Clitellata, Glossoscolecidae) in French Guiana. Eur. J. Soil Biol. 53, 70–75.
676 https://doi.org/10.1016/j.ejsobi.2012.09.001
677 Edwards, C.A., 2004. Earthworm Ecology, Second Edition. CRC Press.
678 Feijoo, A., 2001. Impacto del uso de la tierra en áreas de laderas sobre comunidades de
679 macrofauna del suelo (Caldono, Cauca, Colombia). Universidad Nacional de Colombia,
680 Palmira.
681 Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A., Cleveland, C.C., 2009. Global patterns
682 in belowground communities. Ecol. Lett. 12, 1238–1249. https://doi.org/10.1111/j.1461-
683 0248.2009.01360.x
684 Fleury, M., Alupki, T., Opoya, A., Aloïké, W., 2016. Les Wayana de Guyane française sur les
685 traces de leur histoire. Cartographie participative sur le Litani (Aletani) et mémoire orale.
686 Rev. D'ethnoécologie. https://doi.org/10.4000/ethnoecologie.2711

687 Fragoso, C., 1985. Ecología general de las lombrices terrestres (Oligochaeta: Annelida) de la
688 región Boca del Chajul. Selva Lacandona (Chiapas, México). Uni-Versidad Auton.
689 México 133.
690 Fragoso, C., Lavelle, P., 1992. Earthworm communities of tropical rain forests. Soil Biol.
691 Biochem. 24, 1397–1408. https://doi.org/10.1016/0038-0717(92)90124-G
692 Fragoso, C., Lavelle, P., 1987. The earthworm community of a Mexican tropical rain forest
693 (Chajul, Chiapas). Earthworms 281–295.
694 Fragoso, C., Rojas-Fernández, P., 1996. Earthworms inhabiting bromeliads in Mexican tropical
695 rainforests: ecological and historical determinants. J. Trop. Ecol. 12, 729–734.
696 https://doi.org/10.1017/S0266467400009925
697 Giller, K.E., Beare, M.H., Lavelle, P., Izac, A.-M.N., Swift, M.J., 1997. Agricultural
698 intensification, soil biodiversity and agroecosystem function. Appl. Soil Ecol., Soil
699 Biodiversity, Agricultural Intensification and Agroecosystem Function 6, 3–16.
700 https://doi.org/10.1016/S0929-1393(96)00149-7
701 Guitet, S., Cornu, J.-F., Brunaux, O., Betbeder, J., Carozza, J.-M., Richard-Hansen, C., 2013.
702 Landform and landscape mapping, French Guiana (South America). J. Maps 9, 325–335.
703 https://doi.org/10.1080/17445647.2013.785371
704 Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome c
705 oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B Biol.
706 Sci. 270, S96–S99. https://doi.org/10.1098/rsbl.2003.0025
707 Hendrix, P.F., Bohlen, P.J., 2002. Exotic earthworm invasions in North America: Ecological and
708 policy implications. Biosci. Oxf. 52, 801–811.

709 http://dx.doi.org.subzero.lib.uoguelph.ca/10.1641/0006-
710 3568(2002)052[0801:EEIINA]2.0.CO;2
711 Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A.,
712 Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A.,
713 Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I.,
714 Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on
715 machine learning. PLOS ONE 12, e0169748.
716 https://doi.org/10.1371/journal.pone.0169748
717 Hernández Triana, L.M., Prosser, S.W., Rodríguez-Perez, M.A., Chaverri, L.G., Hebert, P.D.N.,
718 Gregory, T.R., 2014. Recovery of DNA barcodes from blackfly museum specimens
719 (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths. Mol.
720 Ecol. Resour. 14, 508–518. https://doi.org/10.1111/1755-0998.12208
721 Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R package for rarefaction and extrapolation
722 of species diversity (Hill numbers). Methods Ecol. Evol. 7, 1451–1456.
723 https://doi.org/10.1111/2041-210X.12613
724 Ivanova, N., Grainger, C., 2007. Sequencing Protocol for DNA Barcoding.
725 Ivanova, N.V., Dewaard, J.R., Hebert, P.D.N., 2006. An inexpensive, automation-friendly
726 protocol for recovering high-quality DNA. Mol. Ecol. Notes 6, 998–1002.
727 https://doi.org/10.1111/j.1471-8286.2006.01428.x
728 Jiménez, J.J., 1999. Dinámica de las poblaciones y estructura de las comunidades de lombrices
729 de tierra de las sabanas naturales y perturbadas de los Llanos Orientales de Colombia.
730 Doctoral Thesis, Universidad Complutense de Madrid, Madrid.

731 Jones, C.G., Lawton, J.H., Shachak, M., 1994. Organisms as Ecosystem Engineers. Oikos 69,
732 373–386. https://doi.org/10.2307/3545850
733 Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann,
734 N.E., Linder, H.P., Kessler, M., 2017. Climatologies at high resolution for the earth's
735 land surface areas. Sci. Data 4, 1–20. https://doi.org/10.1038/sdata.2017.122
736 Korboulewsky, N., Perez, G., Chauvat, M., 2016. How tree diversity affects soil fauna diversity:
737 A review. Soil Biol. Biochem. 94, 94–106. https://doi.org/10.1016/j.soilbio.2015.11.024
738 Larsson, J., 2020. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses.
739 Lavelle, P., 1983. The structure of earthworm communities, in: Earthworm Ecology. Springer,
740 pp. 449–466.
741 Lavelle, P., 1978. Les vers de terre de la savane de Lamto, Côte d'Ivoire: peuplements,
742 populations et fonctions dans l'écosysteme. École normale supérieure, Laboratoire de
743 zoologie.
744 Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P.,
745 Rossi, J.-P., 2006. Soil invertebrates and ecosystem services. Eur. J. Soil Biol., ICSZ 42,
746 S3–S15. https://doi.org/10.1016/j.ejsobi.2006.10.002
747 Lavelle, P., Kohlmann, B., 1984. Etude quantitative de la macrofaune du sol dans une forêt
748 tropicale humide du Mexique (Bonampak, Chiapas). Etude Quant. Macrofaune Sol Dans
749 Une For. Trop. Humide Mex. Bonampak Chiapas 27, 377–393.
750 Lavelle, P., Lapied, E., 2003. Endangered earthworms of Amazonia: an homage to Gilberto
751 Righi: The 7th international symposium on earthworm ecology · Cardiff · Wales · 2002.
752 Pedobiologia 47, 419–427. https://doi.org/10.1078/0031-4056-00207

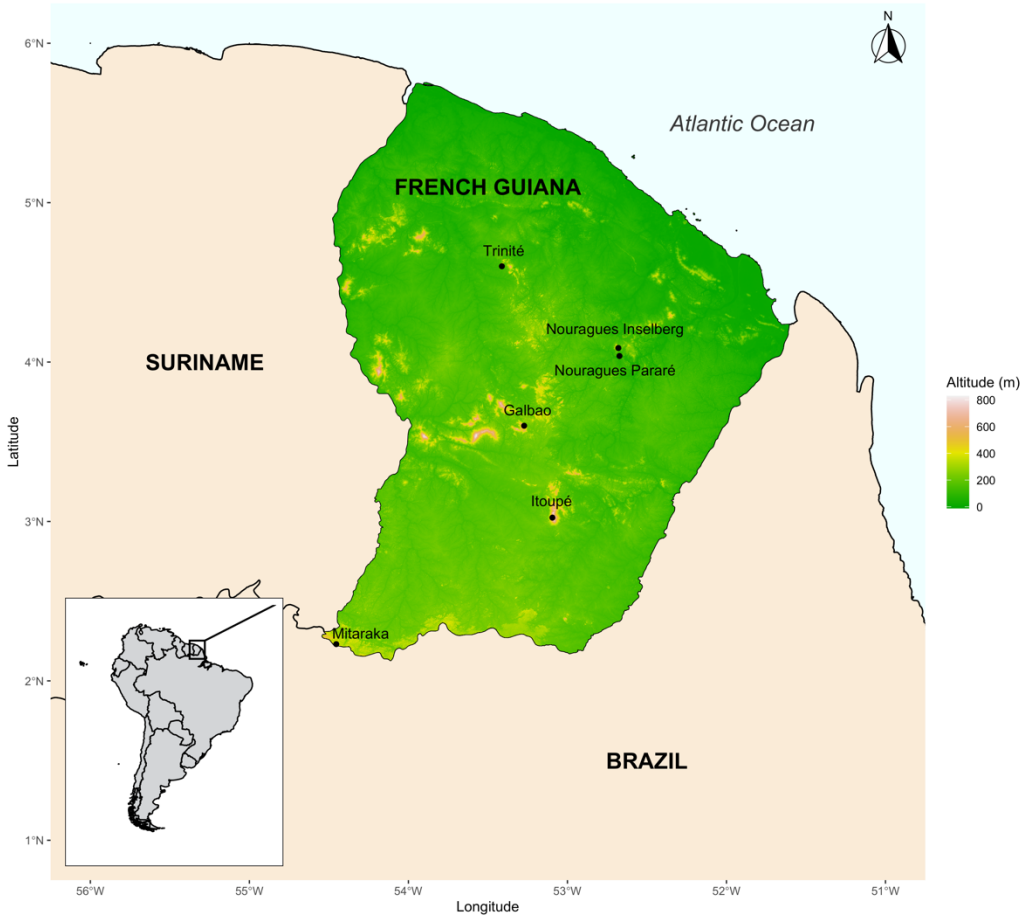
753 Lavelle, P., Spain, A., Blouin, M., Brown, G., Decaëns, T., Grimaldi, M., Jiménez, J.J., McKey,
754 D., Mathieu, J., Velasquez, E., Zangerlé, A., 2016. Ecosystem Engineers in a Self-
755 organized Soil: A Review of Concepts and Future Research Questions. Soil Sci. 181, 91–
756 109. https://doi.org/10.1097/SS.0000000000000155
757 Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of
758 species data. Oecologia 129, 271–280. https://doi.org/10.1007/s004420100716
759 L'office de l'Eau de Guyane [WWW Document], n.d. URL
760 http://eauguyane.fr/index.php?option=com_content&view=article&id=142&catid=8:actu
761 alites&Itemid=213 (accessed 5.11.20).
762 Marichal, R., Martinez, A.F., Praxedes, C., Ruiz, D., Carvajal, A.F., Oszwald, J., del Pilar
763 Hurtado, M., Brown, G.G., Grimaldi, M., Desjardins, T., Sarrazin, M., Decaëns, T.,
764 Velasquez, E., Lavelle, P., 2010. Invasion of Pontoscolex corethrurus (Glossoscolecidae,
765 Oligochaeta) in landscapes of the Amazonian deforestation arc. Appl. Soil Ecol. 46, 443–
766 449. https://doi.org/10.1016/j.apsoil.2010.09.001
767 Mathieu, J., Davies, T.J., 2014. Glaciation as an historical filter of below-ground biodiversity. J.
768 Biogeogr. 41, 1204–1214. https://doi.org/10.1111/jbi.12284
769 Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R.,
770 O'Hara, R.B., Simpson, G.L., Solymos, P., H. Stevens, M.H., Szoecs, E., Wagner, H.,
771 2019. vegan: Community Ecology Package. R package version 2.5. 4. 2019.
772 Pansu, M., Gautheyrou, J., 2007. Handbook of Soil Analysis: Mineralogical, Organic and
773 Inorganic Methods. Springer Science & Business Media.
774 Phillips, H.R.P., Guerra, C.A., Bartz, M.L.C., Briones, M.J.I., Brown, G., Crowther, T.W.,
775 Ferlian, O., Gongalsky, K.B., Hoogen, J. van den, Krebs, J., Orgiazzi, A., Routh, D.,

776 Schwarz, B., Bach, E.M., Bennett, J., Brose, U., Decaëns, T., König-Ries, B., Loreau, M.,
777 Mathieu, J., Mulder, C., Putten, W.H. van der, Ramirez, K.S., Rillig, M.C., Russell, D.,
778 Rutgers, M., Thakur, M.P., Vries, F.T. de, Wall, D.H., Wardle, D.A., Arai, M., Ayuke,
779 F.O., Baker, G.H., Beauséjour, R., Bedano, J.C., Birkhofer, K., Blanchart, E., Blossey,
780 B., Bolger, T., Bradley, R.L., Callaham, M.A., Capowiez, Y., Caulfield, M.E., Choi, A.,
781 Crotty, F.V., Dávalos, A., Cosin, D.J.D., Dominguez, A., Duhour, A.E., Eekeren, N. van,
782 Emmerling, C., Falco, L.B., Fernández, R., Fonte, S.J., Fragoso, C., Franco, A.L.C.,
783 Fugère, M., Fusilero, A.T., Gholami, S., Gundale, M.J., López, M.G., Hackenberger,
784 D.K., Hernández, L.M., Hishi, T., Holdsworth, A.R., Holmstrup, M., Hopfensperger,
785 K.N., Lwanga, E.H., Huhta, V., Hurisso, T.T., Iannone, B.V., Iordache, M., Joschko, M.,
786 Kaneko, N., Kanianska, R., Keith, A.M., Kelly, C.A., Kernecker, M.L., Klaminder, J.,
787 Koné, A.W., Kooch, Y., Kukkonen, S.T., Lalthanzara, H., Lammel, D.R., Lebedev, I.M.,
788 Li, Y., Lidon, J.B.J., Lincoln, N.K., Loss, S.R., Marichal, R., Matula, R., Moos, J.H.,
789 Moreno, G., Morón-Ríos, A., Muys, B., Neirynck, J., Norgrove, L., Novo, M., Nuutinen,
790 V., Nuzzo, V., P, M.R., Pansu, J., Paudel, S., Pérès, G., Pérez-Camacho, L., Piñeiro, R.,
791 Ponge, J.-F., Rashid, M.I., Rebollo, S., Rodeiro-Iglesias, J., Rodríguez, M.Á., Roth,
792 A.M., Rousseau, G.X., Rozen, A., Sayad, E., Schaik, L. van, Scharenbroch, B.C.,
793 Schirrmann, M., Schmidt, O., Schröder, B., Seeber, J., Shashkov, M.P., Singh, J., Smith,
794 S.M., Steinwandter, M., Talavera, J.A., Trigo, D., Tsukamoto, J., Valença, A.W. de,
795 Vanek, S.J., Virto, I., Wackett, A.A., Warren, M.W., Wehr, N.H., Whalen, J.K., Wironen,
796 M.B., Wolters, V., Zenkova, I.V., Zhang, W., Cameron, E.K., Eisenhauer, N., 2019.
797 Global distribution of earthworm diversity. Science 366, 480–485.
798 https://doi.org/10.1126/science.aax4851

799 Porco, D., Decaëns, T., Deharveng, L., James, S.W., Skarżyński, D., Erséus, C., Butt, K.R.,
800 Richard, B., Hebert, P.D.N., 2013. Biological invasions in soil: DNA barcoding as a
801 monitoring tool in a multiple taxa survey targeting European earthworms and springtails
802 in North America. Biol. Invasions 15, 899–910. https://doi.org/10.1007/s10530-012-
803 0338-2
804 R Core Team, 2020. R: A Language and Environment for Statistical Computing, R Foundation
805 for Statistical Computing. ed. Vienna, Austria.
806 Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-Based Registry for All Animal Species: The
807 Barcode Index Number (BIN) System. PLOS ONE 8, e66213.
808 https://doi.org/10.1371/journal.pone.0066213
809 Richard, B., Decaëns, T., Rougerie, R., James, S.W., Porco, D., Hebert, P.D.N., 2010. Re-
810 integrating earthworm juveniles into soil biodiversity studies: species identification
811 through DNA barcoding. Mol. Ecol. Resour. 10, 606–614. https://doi.org/10.1111/j.1755-
812 0998.2009.02822.x
813 Rodriguez, C., Borges, S., Martínez, M.A., Fragoso, C., James, S., González, G., 2007.
814 Biodiversidad y ecología de las lombrices de tierra en las islas caribeñas. Minhocas Na
815 América Lat. Biodiversidade E Ecol. EMBRAPA Soja Bras. 79–98.
816 Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A.M., Neilson, R., Boag, B.,
817 Schmidt, O., Murchie, A.K., Blackshaw, R.P., Pérès, G., Cluzeau, D., Guernion, M.,
818 Briones, M.J.I., Rodeiro, J., Piñeiro, R., Cosín, D.J.D., Sousa, J.P., Suhadolc, M., Kos, I.,
819 Krogh, P.-H., Faber, J.H., Mulder, C., Bogte, J.J., Wijnen, H.J. van, Schouten, A.J.,
820 Zwart, D. de, 2016. Mapping earthworm communities in Europe. Appl. Soil Ecol., Soil
821 biodiversity and ecosystem functions across Europe: A transect covering variations in

822 bio-geographical zones, land use and soil properties 97, 98–111.
823 https://doi.org/10.1016/j.apsoil.2015.08.015
824 Siqueira, F. de F., Sandes, S.H. de C., Drumond, M.A., Campos, S.H., Martins, R.P., Fonseca,
825 C.G. da, Carvalho, M.R.S., 2013. Genetic diversity and population genetic structure in
826 giant earthworm Rhinodrilus alatus (Annelida: Clitellata: Glossoscolecidae).
827 Pedobiologia 56, 15–21. https://doi.org/10.1016/j.pedobi.2012.08.006
828 Smith, M.A., Fisher, B.L., Hebert, P.D.N., 2005. DNA barcoding for effective biodiversity
829 assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos. Trans. R.
830 Soc. Lond. B Biol. Sci. 360, 1825–1834. https://doi.org/10.1098/rstb.2005.1714
831 Spurgeon, D.J., Keith, A.M., Schmidt, O., Lammertsma, D.R., Faber, J.H., 2013. Land-use and
832 land-management change: relationships with earthworm and fungi communities and soil
833 structural properties. BMC Ecol. 13, 46. https://doi.org/10.1186/1472-6785-13-46
834 Steffen, G.P.K., Steffen, R.B., Bartz, M.L.C., James, S.W., Jacques, R.J.S., Brown, G.G.,
835 Antoniolli, Z.I., 2018. Earthworm diversity in Rio Grande do Sul, Brazil. Zootaxa 4496,
836 562–575. https://doi.org/10.11646/zootaxa.4496.1.43
837 Taheri, S., Decaens, T., Cunha, L., Brown, G.G., Da Silva, E., Bartz, M.L.C., Baretta, D.,
838 Dupont, L., 2020. Genetic evidence of multiple introductions and mixed reproductive
839 strategy in the peregrine earthworm Pontoscolex corethrurus. Biol. Invasions 1–13.
840 Touroult, J., Pollet, M., Pascal, O., 2018. Overview of Mitaraka survey: research frame, study
841 site and field protocols. Zoosystema 40, 327–365.
842 https://doi.org/10.5252/zoosystema2018v40a13
843 Unité de service Cirad - Analyses [WWW Document], n.d. URL https://us-analyses.cirad.fr/
844 (accessed 10.8.19).

845 Wolters, V., 2001. Biodiversity of soil animals and its function. Eur. J. Soil Biol. 37, 221–227.
846 https://doi.org/10.1016/S1164-5563(01)01088-3
847 Young, M.R., Behan-Pelletier, V.M., Hebert, P.D.N., 2012. Revealing the Hyperdiverse Mite
848 Fauna of Subarctic Canada through DNA Barcoding. PLOS ONE 7, e48755.
849 https://doi.org/10.1371/journal.pone.0048755

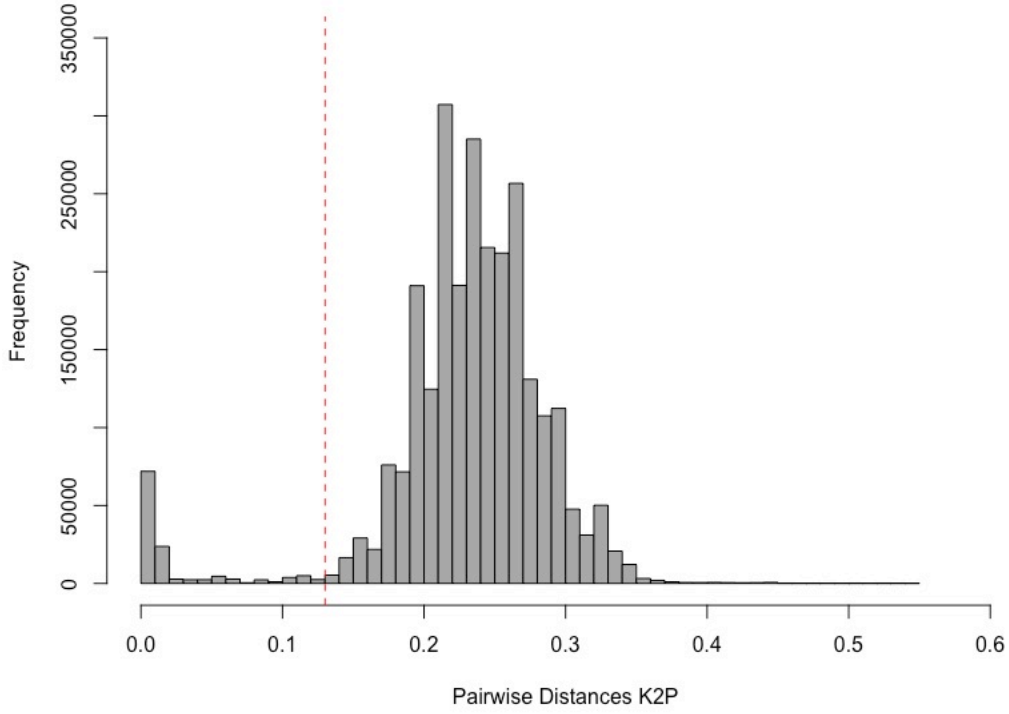


850

851 **Figure 1:** Map showing the six sampling locations in French Guiana. Data from Galbao, Itoupé,

852 Mitaraka and Trinité were generated as part of this study, data from the Nouragues were taken

853 from an earlier study (Decaëns et al., 2016).



854

855 **Figure 2:** Barplot showing the pairwise distance distribution of 2304 DNA barcode sequences.

856 The dotted red line represents the threshold value between intra-MOTU (left) and inter-MOTU

857 (right) pairs of individuals, used in ABGD (13 %).

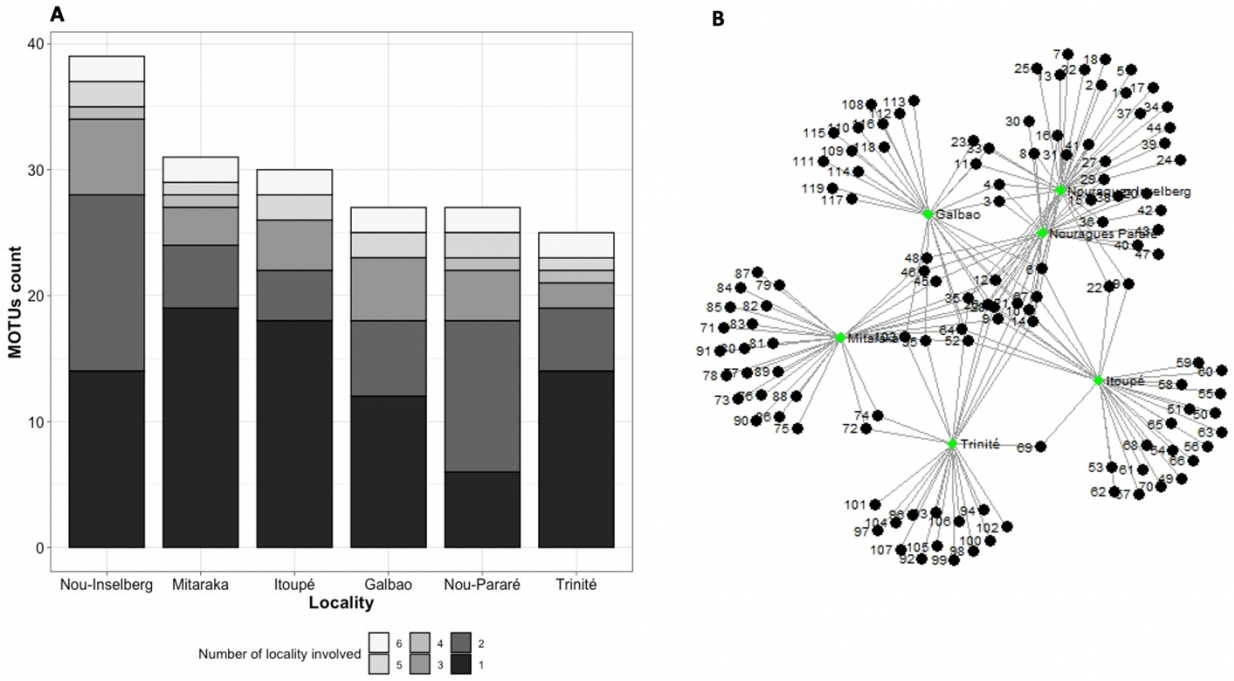


Figure 3: A) Barplot showing the total richness for each locality as well as the part that is specific to one location (1) and shared between 2 to 6 locations. B) Bipartite network of the MOTUs showing detail on how the MOTUs are shared between the four locations. Each circular node represents a MOTU and each square a locality. (Nou = Nouragues).

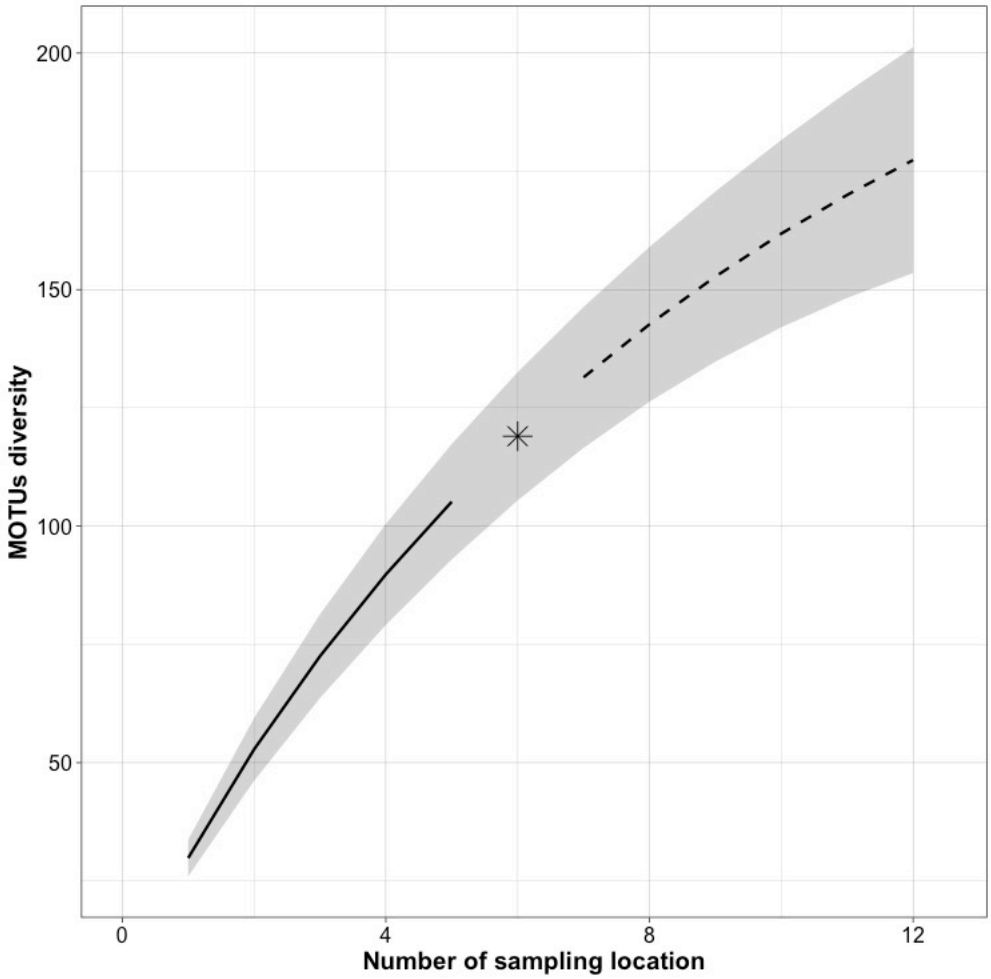
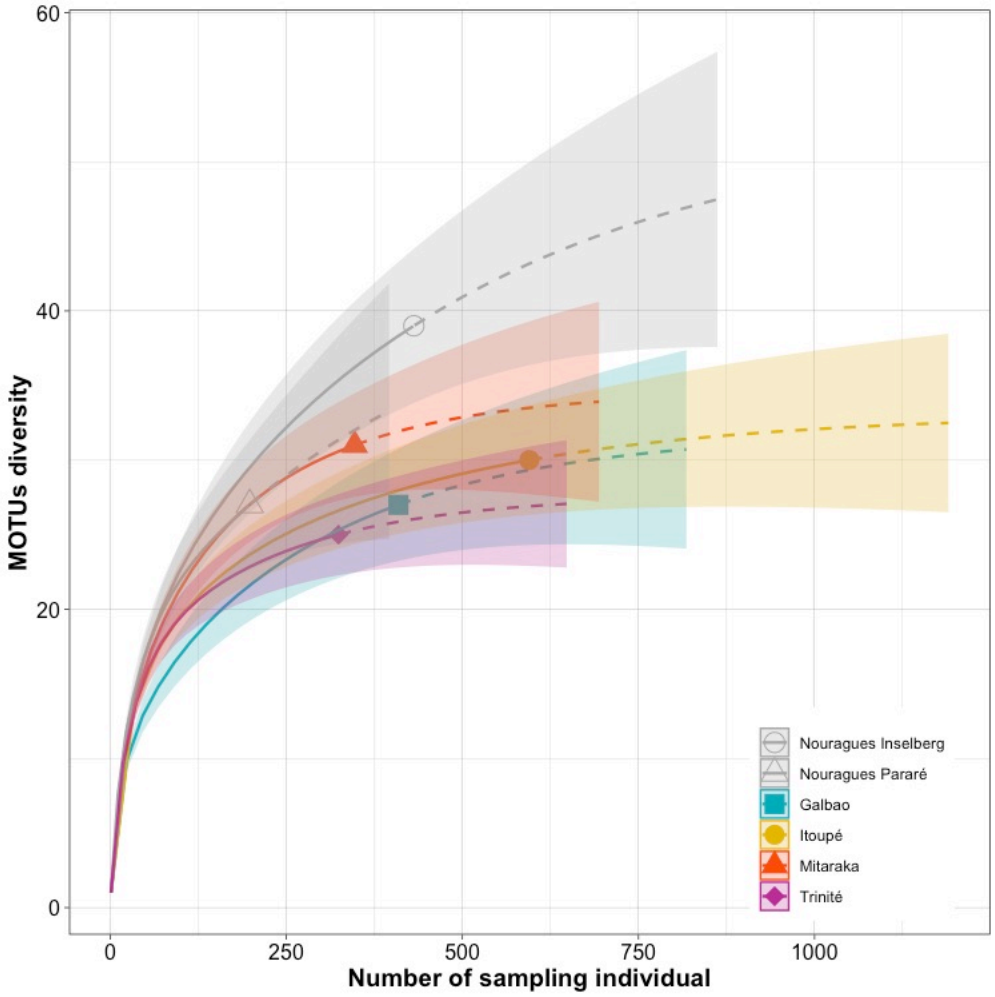
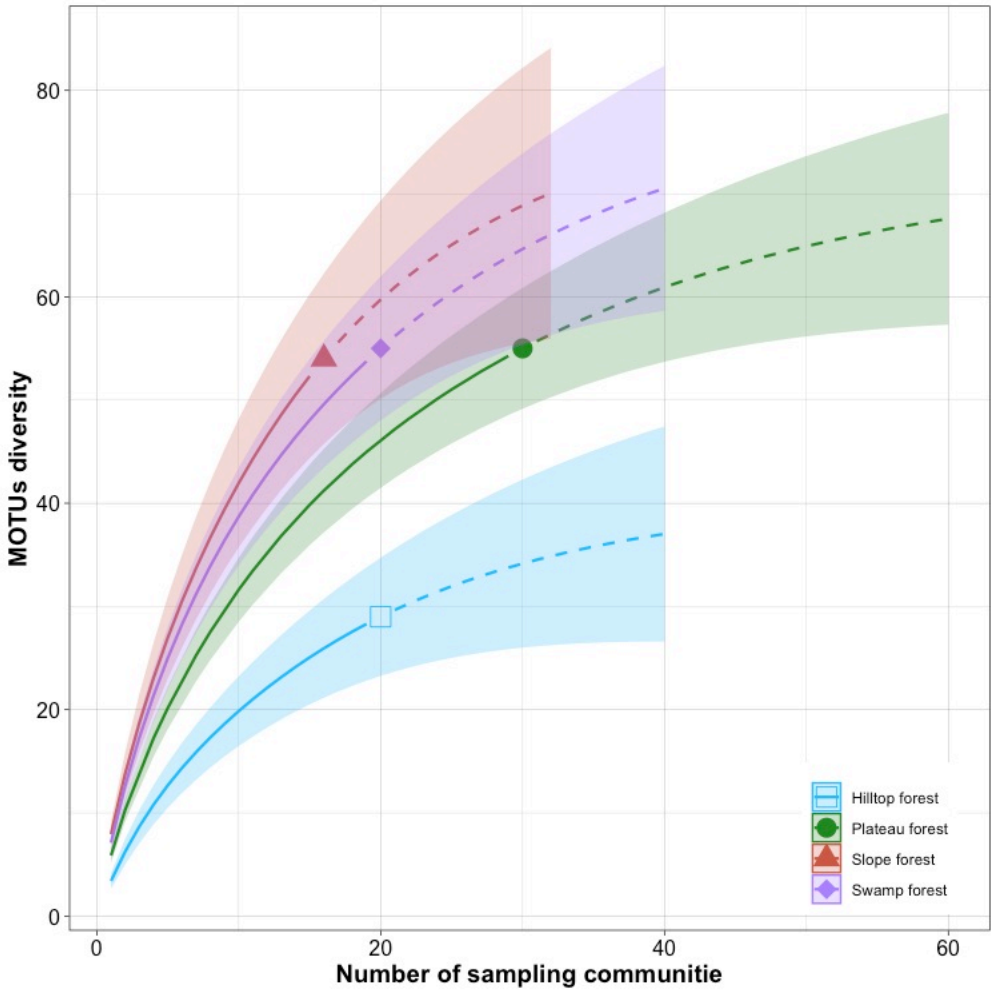


Figure 4: Rarefaction and extrapolation curves showing MOTUs accumulation according to the number of sampled locations. A total of 119 MOTUs were observed for six sampled locations (Galbao, Itoupé, Mitaraka, Trinité and Nouragues-Inselberg and Nouragues-Pararé). Solid line corresponds to rarefaction curve, dashed line to extrapolation curve; shaded area represents a 95% confidence intervals based on a bootstrap method with 200 replications.



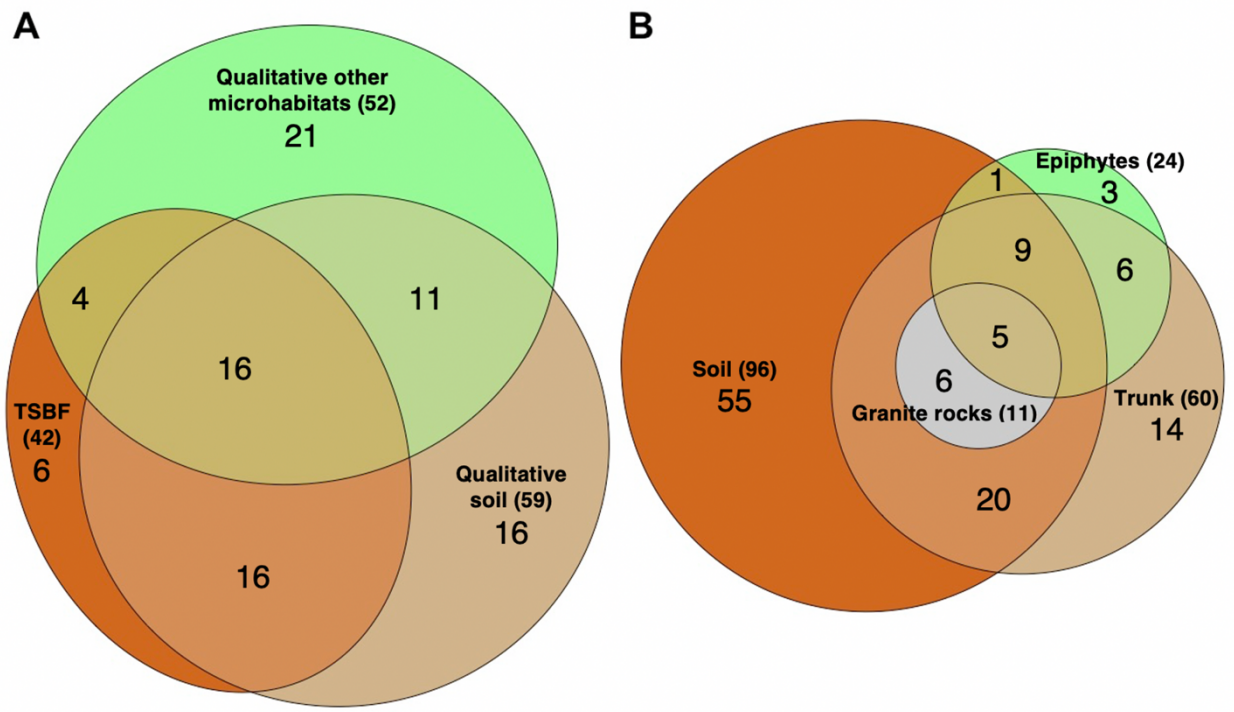
869

870 **Figure 5:** Rarefaction and extrapolation curves showing MOTU accumulation according to the
871 number of sampled individuals in the six different locations separately. Rarefaction curves are
872 represented in solid lines, extrapolation curves in dashed lines; shaded areas represent a 95%
873 confidence intervals based on a bootstrap method with 200 replications.



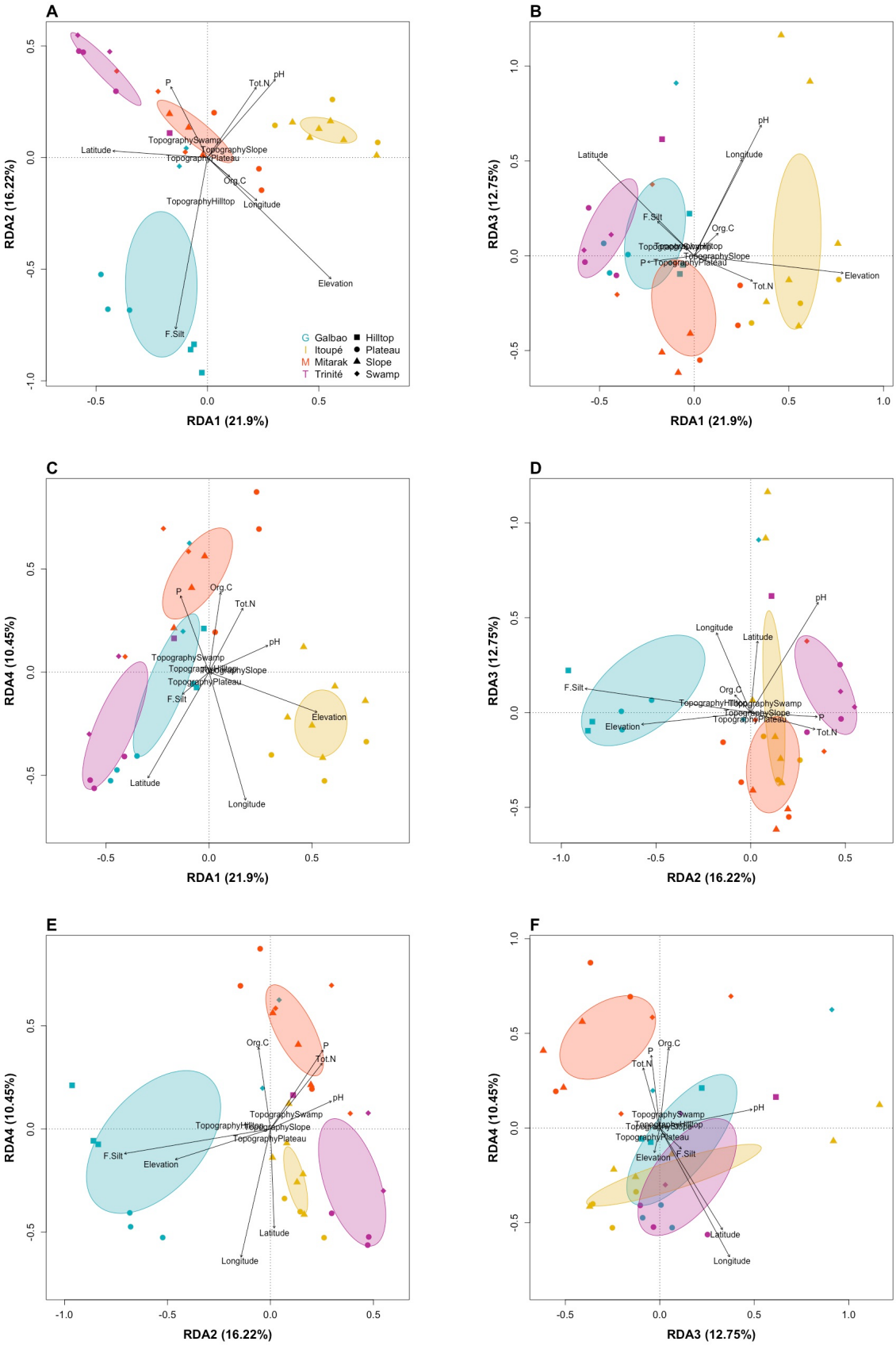
874

875 **Figure 6:** Rarefaction and extrapolation curves showing the MOTU accumulation according to
876 the number of sampled communities for each main type of habitat at the regional scale, from all
877 six available localities. Rarefaction curves are represented in solid lines, extrapolation curves in
878 dashed lines; shaded areas represent a 95% confidence intervals based on a bootstrap method
879 with 200 replications.



880

881 **Figure 7:** Venn diagram showing the number of MOTUs recovered by A) the different sampling
882 methods (TSBF = quantitative) performed in Galbao, Itoupé Mitaraka and Trinité only (total
883 MOTU = 90), as only the qualitative method has been used in the Nouragues; and B) the
884 different microhabitat types sampled with the whole range of sampling methods and all the
885 dataset (total MOTU = 119). Numbers in brackets are the total number of MOTUs (i.e. unique
886 and shared) found for a given sampling method or microhabitat.



887

888 **Figure 8:** MOTU-environment triplot of RDA after variables selection, showing the relationship
889 between MOTUs composition and environmental data (scaling 1) on the first two axes (A); the
890 relationship between the communities and the environmental data (scaling 2) on the first two
891 axes (B); the axis 1 and 4 (C); and the axis 2 and 4 (D). Arrows represent the quantitative
892 environmental variables and their length indicates the correlation between the environmental
893 variables and the ordination axes. Abbreviations: C.Sand = coarse sand, C.Silt = coarse silt, P =
894 available phosphorus and Topo = Topography.

Table 1: Number of COI sequences per locality and associated diversity indices, calculated with the package *vegan* Indices were calculated using MOTUs as species proxy. The last row "All" represents observed and estimated richness at the regional scale, for all six localities. Abbreviations: S.obs = observed richness, S.chao1 = index of estimated richness and se.chao1 = standard error for s.choa1.

Localities	**# of sequences**	**S.obs**	**S.chao1**	**se.chao1**
Galbao	409	27	30.5	3.65
Itoupé	595	30	32	2.58
Mitaraka	347	31	33.5	2.89
Trinité	324	25	26.5	2.22
Nouragues_Inselberg	425	39	50.14	8.23
Nouragues_Pararé	204	27	36.33	8.84
All	2304	119	262.35	45.70