

Troglореgions: delimiting subterranean faunistic identities in Afrotropics and Neotropics

Denizar Alvarenga (✉ denizar.alvarenga@gmail.com)

Federal University of Lavras <https://orcid.org/0000-0002-1495-6498>

Marconi Silva

Federal University of Lavras

Rodrigo Ferreira

Universidade Federal de Lavras <https://orcid.org/0000-0003-3288-4405>

Article

Keywords: macroecological patterns, ecoregions, fauna, invertebrate families, bioregions

Posted Date: August 14th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-53597/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Studies regarding macroecological patterns involving subterranean fauna are challenging, since the environments where such species are found generally do not have the same ecological patterns observed on the surface, due to their isolation. Therefore, using data on the occurrence of invertebrate families, we tested the influence of ecological regions already established for surface environments (biogeographic domains, biomes, and ecoregions), lithology, and drainage basins as potential drivers of the similarity on the invertebrates communities. We observed that within the surface ecoregions there might still be subdivisions due to different drainage basins, a pattern that was repeated in both aquatic and terrestrial fauna. Thus, we present different bioregions in the Afrotropics and Neotropics, in which caves have distinct faunistic identities. We discuss the biogeographic relationships between the epigeal and hypogean environments that may be behind these patterns. We believe that these results can assist in conservation strategies, in which these different compositions are considered.

Introduction

The structuring of terrestrial ecosystems is determined by several biogeographic factors that define the diversity patterns of organisms¹. The processes that rule these patterns are scale-dependent and hierarchical, ranging from biotic interactions at local scales to variations in environmental conditions at large scales^{2,3}. In general, the terrestrial biota responds in a very similar way to the pressures imposed by environmental conditions⁴. Therefore, on a global scale, there is a huge variety of ecosystems shaped by historical changes on the environment, consequently presenting different biological identities⁵. The idea of representing these biological identities is old and over time it has been supported by different theories that tried to explain such patterns⁶. Previous classification systems were based on different attributes, such as temperature, precipitation, and vegetation structure^{7,8}. Currently, however, these systems primarily delimit regions with geographically distinct groups of species, communities, environmental conditions, and a unique biogeographic history⁵. For this, species richness, endemism, and shared taxonomic groups, such as genera and families are generally considered^{5,9}. For terrestrial environments, the most currently used classification system corresponds to the ecoregions proposed by Olson et al. (2001)⁵, which brings together 867 regions nested in 14 global biomes and 8 biogeographic domains. Despite certain peculiarities according to the different taxa, the "borders" of ecoregions are, until now, the most efficient way of delimiting the biodiversity patterns on the earth's surface, showing conspicuous transitions between them¹⁰.

Establishing this type of bioregional classification for subterranean environments represents an enormous challenge, since their ecological dynamics do not follow the same rules that structure surface environments, due to their relatively stable and highly selective conditions¹¹. Furthermore, there are several other obstacles related to the lack of accurate information about the taxonomy of numerous groups, their evolutionary patterns, their actual distributions, and, consequently, the community composition as a whole^{12,13}. Despite the several limitations, investigations about this subject have

shown promising results. For example, a pioneering work carried out in Germany by Stein et al. (2012)¹⁴, showed that the compositional patterns of strictly aquatic cave fauna can differ significantly from the bioregional patterns observed on the surface. Poulson & White (1969)¹⁵ theorized the general causes responsible for the differentiated diversity patterns observed for the subterranean fauna, guiding studies until today. According to these authors, in order of importance, these patterns would be guided by the time available between the dispersion, isolation and speciation; by the spatial heterogeneity and resource availability in the occupied environment and, finally, by the interactions between the organisms that compose the community. Cave communities would, therefore, represent a balance between colonization and extinction determined by variations in environmental conditions and dispersion, like an island^{15,16}.

Since then, one of the most interesting patterns observed and described for cave communities is that they comprise large ecotones between epigeal and hypogean environments^{17,18}. Thus, caves would be the most easily accessible places of a large intercommunicated empty spaces network of the most diverse dimensions, presenting themselves as peripheral zones of much larger organism distributions¹⁸. Due to a balance between environmental stability and resource availability, this transition is more easily observed in the regions close to the cave entrances, which support high biodiversity that mixes components from the epigeal and hypogean fauna¹⁷. The fact is that it is common for the taxa occurring along this interface between the surface and the underground to be phylogenetically related¹⁵. Furthermore, a substantial gradient of specializations and different niche occupations are observed in many taxa, evidenced by adaptations and behaviors according to the conditions prevailing in the occupied space¹⁹⁻²¹. The caveat is due to those relict organisms from very old distributions that had their surface-related groups extinct by some stochastic event¹⁵. Besides, in subterranean communities there is a positive relationship between local and regional taxa richness¹¹, similar to the patterns observed in surface communities²², indicating that they are unsaturated with species¹¹. It is discussed in the literature that this unsaturation indicates that subterranean communities may be influenced by the regional diversity, being somehow interconnected by species' dispersion¹¹. The definitive colonization of subterranean environments, however, occurs by those pre-adapted organisms, capable of establishing themselves under total darkness and oligotrophic conditions, among other limiting characteristics²³. With all these assumptions, the established consensus is that subterranean communities are phylogenetically related sub-samples from the regional species pool present in epigeal environments²⁴.

However, in most cases, the tests of hypotheses related to the ecological patterns of subterranean communities have involved invertebrate fauna restricted to these environments (troglobites)²⁵⁻²⁷, strictly aquatic fauna (stygobites)^{14,28,29}, or specific taxonomic groups^{30,31}. This leaves a huge gap that has not been extensively explored yet, regarding the search for more general patterns of subterranean fauna, composed mainly of organisms not restricted to subterranean environments¹⁵. From the conservation perspective, the historical focus given to obligatory species and whose distributions are very restricted may completely obscure the fact that the caves are a large reservoir of biodiversity for invertebrates in general³². Thus, mostpart of the subterranean communities, even when composed of not restricted

species, is fundamental for the maintenance of the trophic chain dynamics, a fact commonly neglected¹¹. In this context, biogeographic classifications focussed on subterranean environments that consider their whole communities can be crucial for conservation issues, so that different regions with different biological attributes can be targeted, as recommended for terrestrial external ecoregions³³.

The main objectives of our work focused, therefore, on the search for eventual compositional identities of the subterranean invertebrate communities in caves from the Afrotropical and Neotropical regions and to indicate the factors responsible for them. These regions, in particular, have a related and unique biogeographic history, since they were connected and underwent similar paleoclimatic and geological changes that shaped their current landscape configuration as well as their faunal distribution^{34,35}. To achieve these objectives, we used data on the presence and absence of invertebrate families recorded in caves in these two regions. We have explored a set of five factors, that allowed us to verify the possible influences of the surface regional pool of biological communities (Tested with the delimitation of the *biogeographical domains, biomes, and ecoregions*³³) and possible influences of factors related to the faunal dispersion (Tested with the delimitation of *drainage basins*³⁶), respecting the different cave *lithologies*³⁷. These five factors were explored both separately and nested, hierarchically encompassing the caves from more extensive to more spatially restricted scales. We considered two hypotheses *a priori*: H1) Possible large-scale compositional patterns could prove to be insignificant as each factor is considered in the analyzes using data from the occurrence of invertebrate families. The geographic distributions of the families, which are generally broad^{1,38}, could result in a low variability of the data and, consequently, in the non-significance of each factor. Therefore, an effective bioregional classification with these data would prove unfeasible; H2) Even with very simplified communities due to their highly restrictive and selective conditions¹¹, on a continental scale the compositional patterns of cave fauna could still vary markedly, reflecting the influence of different factors that characterize the surface environments features. Consequently, the establishment of a bioregional classification based on these patterns could be efficient, even when using data on the occurrence of families.

Considering the results found and the assumptions presented in the discussion, we propose a new method of bioregional classification based on the similarity patterns of cave invertebrate communities, which defines areas of faunistic identity, here called *Trogloregions*.

Results

We observed 455 invertebrates families in the 234 caves considered in our dataset. On average, the number of families per cave was 26.05 (SD=+/- 13.85). More than half of the registered families belong to only five taxa: "Acari", Coleoptera, Araneae, Diptera, and Hemiptera (Fig. 1a). Among the 455 registered families, ten occur in more than 100 caves, indicating a wide distribution (Fig. 1b). From these ten families, four belong to Araneae (Pholcidae, Theridiidae, Ctenidae, and Sicariidae). The other six are Phalangopsidae (Orthoptera), Formicidae (Hymenoptera), Tineidae (Lepidoptera), Staphylinidae (Coleoptera), Reduviidae (Hemiptera) and Psychodidae (Diptera). Pholcidae was the family with the

highest number of records, occurring in 182 of the 234 caves. Approximately half of the families were registered in only one (135), two (63), or three caves (35).

Factor testing

We observed that in the biological community as a whole, all factors significantly represented the variations in the fauna compositional similarity (biogeographic region, lithology, biome, ecoregion, drainage basin, nested combination: PERMANOVA $p < 0.001$, Table 1; nMDS Fig. 2). However, it was the nested combination of these factors that best represented the similarity patterns found, consequently providing us with the most robust spatial aggregations of caves (ANOSIM global $R = 0.732$). Thus, initially, 37 sets of caves were formed according to the different levels of the factor corresponding to the nested combination. In each formed set, the caves belong to a given drainage basin, an ecoregion, a biome, a lithology, and a biogeographic domain, respecting the spatial hierarchy of their levels.

Table 1

PERMANOVA results testing the factors separately, combined in the nested design (a hierarchical combination of factors) and the groups formed in the SIMPROF analysis, which served as the basis for the delimitation of the Trogleregions.

Factor	df	SS	MS	Pseudo-F	P(perm)	Unique perms.
Biogeographic realm	1	30414	30414	11,263	0,0001	9886
Lithology	2	17546	8773,2	3,1698	0,0001	9849
Biome	4	68108	17027	6,6224	0,0001	9812
Ecoregion	15	218420	14561	7,2394	0,0001	9659
Drainage basin	25	237760	9510,4	4,7197	0,0001	9588
Nested design	36	71180	4187,1	2,3169	0,0001	9677
Simprof groups*	16	71404	4462,7	4,0497	0,0001	9774

Considering that spatially this combination corresponds to the smallest possible aggregations of caves according to the methodology employed, this means that within the ecoregions there may still be subdivisions in the compositional patterns due to the occurrence of different drainage basins. The dependence of these subsets in relation to which ecoregions they occur was revealed when the drainage basins were tested separately, showing less explanatory power (PERMANOVA $p < 0.001$; ANOSIM $R = 0.502$). This hierarchical dependence was also evident when the ecoregions were tested separately, also showing comparatively inferior results (PERMANOVA $p < 0.001$; ANOSIM $R = 0.638$).

Given the very low correlation between the dissimilarity of cave fauna and geographical distance (DistLM $p < 0.01$, $R^2 = 0.088$), we consider the influence of spatial autocorrelation on similarity patterns observed at the family level to be negligible.

When we exclude singletons from the dataset and extend these analyses to the subsets of strictly terrestrial families and families with aquatic habits, we observe the same patterns indicated in the analyses with the complete database (supplementary material S1, supplementary tables 1, 2 and 3). The results of these different approaches highlight the robustness of our observations, reinforcing that the nested combination of factors best represented the similarity patterns of the subterranean invertebrate fauna, regardless their habits.

Troglореgion delimitation

After testing each factors' significance and explanatory power and the consequent selection of the factor corresponding to the nested combination, the next step was to make pairwise comparisons in each level of the *nested combination* factor. The nested combination of these factors, from the one with the highest number of levels to the one with the lowest number of levels (From "Drainage basin", with 26 levels, to "biogeographic domain", with two levels) resulted in the distribution of the 234 caves in 37 groups. Thus, after the invertebrate fauna was attributed to the 37 cave groups corresponding to these levels, the comparisons between these groups grouped them into 17 "supergroups" (SIMPROF $p < 0.01$), as shown in the Fig. 3 CLUSTER. In other words, we found different regions in which caves present distinct faunal compositions considering their invertebrate families. The same 17 supergroups were obtained when we re-analyzed the SIMPROF with $p < 0.05$, using the Jaccard index instead of the Sorensen-Dice index, with $p < 0.01$ and $p < 0.05$. Again, the maintenance of the same patterns in these different approaches reinforces the consistency of the results.

We adjusted the delimitation of certain supergroups resulting from the SIMPROF analysis due to the large geographical distance or spatial overlap of supergroups due only to the different lithologies (see methods section). Therefore, we have defined a total of 18 Troglореgions, which names are based on remarkable regional characteristics. They are *Kalahari, Kumasi, Acra, Lake Volta, Cape Town, Areia Branca, Upper Ribeira, Atlantic Highlands, Lower Rio Pardo, Upper Paraná, Espírito Santo, Eastern Amazon, Petén-Veracruz, Etosha, Lower São Francisco, Central Brazil, Lower Paranaíba e Apodi-Mossoró* (Fig. 5). The Troglореgions' borders correspond to the cutouts of the overlays of drainage basins with the ecoregions, in a nested way. It is important to note that naturally, ecoregions are already nested within biomes and biogeographic domains, following their borders.

Finally, the ANOSIM testing the 37 sets of caves (resulting from the nested combination of factors) using the supergroups indicated by CLUSTER and SIMPROF as the grouping factor, which served as the basis for the delimitation of the Troglореgions, obtained an $R = 0.923$ (PERMANOVA $p < 0.001$), a value much higher than that of all other factors tested in our initial explorations (Fig. 4, Table 2).

Table 2
ANOSIM results testing the factors separately, combined in the nested design (a hierarchical combination of factors) and the groups formed in the SIMPROF analysis, which served as the basis for the delimitation of the Troglореgions.

Factor	Levels	p	Global R
Biogeographic realm	2	0,001	0,704
Lithology	3	0,115	0,063
Biome	5	0,001	0,255
Ecoregion	16	0,001	0,638
Drainage basin	26	0,001	0,502
Nested design	37	0,001	0,732
Simprof groups*	17	0,001	0,923

Discussion

The results corroborated our second hypothesis. We verified that although the caves present only a filtered portion of the biodiversity observed in the surface ecosystems, along with continental scales the compositional patterns of the subterranean communities can vary dramatically. As a result, caves from different regions may have singular faunistic identities. According to the results, this occurs due to different attributes of the external environment, such as the bioregional delimitation of epigeal ecosystems with their potential colonizers (here tested with the delimitation of ecoregions, biomes and biogeographic domains) and according to divergent drainage flows (here tested with the delimitation of the drainage basins). Essentially, all the approaches showed that within the ecoregions there may still be subdivisions in the compositional patterns due to the different drainage basins, respecting the different lithologies. In addition to subterranean communities as a whole, these patterns were also evident in both strictly terrestrial communities and related or strictly aquatic organisms.

Such results, however, need to be interpreted with caution, especially regarding their biogeographical context. The first point to be noted is the large spatial discontinuities between the caves covered by this study, which is inevitable because regions that are prone to the occurrence of caves are not homogeneously spatially distributed³⁹. Besides, the subterranean biodiversity of tropical regions is still little known or underestimated when compared to certain locations in the northern hemisphere⁴⁰. The second point concerns the use of family occurrence data. Naturally, families have older origins and their occurrences reflect more generalized spatially patterns and less restricted than genera or species^{1,38}. Still, family occurrence data can show enough variability and be efficient in showing variations in ecological patterns, being a viable alternative for quick access to biodiversity⁴¹⁻⁴³. In this work, this prediction was corroborated. A positive consequence of using family data is that their occurrences may reflect different

biogeographic or paleoclimatic scenarios that have shaped their current distributions, connecting different regions that were related in the past and currently house them^{1,38}.

Although subterranean communities may be determined by regional and local processes as in any other external environment, these communities are generally very simplified due to the highly limiting conditions, and the relative environment stability, factors that end up reducing the genetic variability of the populations^{11,15}. Environmental stability is a key point, as subterranean communities can be directly affected by environmental fluctuations in both the external and subterranean environments^{30,31,44}. Apparently, environmental stability is one of the main “attractors” for populations that may come and colonize these environments, ceasing the selection of characteristics that can be advantageous under the great unpredictability of the surface environment¹⁵. Considering the cave environment, organisms can, for example, adapt their distributions according to fluctuations in temperature or relative humidity, which are much less pronounced than in the external environments^{44,45}. Regarding the environmental stability in the external habitats, climatic oscillations on the surface can extinguish epigeal populations, while hypogean populations can be spared, resulting in their isolation^{30,31}. Therefore, it is argued that the maintenance of greater subterranean biodiversity in certain regions may be the result of constant precipitation rates in the long term, favoring a greater resource availability to the hypogean environment²⁷. It has also been found that changes in temperature on the surface are one of the main factors responsible for variations in the similarity patterns of subterranean fauna, especially for non-specialized fauna, providing more similar communities in more climatically similar regions⁴⁶. Thus, it is very likely that the patterns verified with the delimitation of the Trogleregions were partly shaped by the environmental transformations and characteristics of the surface.

To support this hypothesis, there is evidence that certain groups of arthropods (Araneae, Sicariidae: *Sicarius* and *Hexophthalma*⁴⁷, and Psocoptera, Prionoglarididae: *Afrotroglia*, *Neotroglia*, and *Sensitibilla*⁴⁸) that are also found in caves, are restricted to the aridest parts of Afro and Neotropical regions. Their distributions are confined to regions of greater aridity since the vicariance event, when the distribution of their ancestors diverged^{47,48}. According to our database, the occurrence of the genera from these two families, for example, occurs largely in the caves of the “Lower San Francisco” and “Etosha” Trogleregions. It is interesting to note that the structure of the communities of these Trogleregions is highly similar, as shown by the cluster analysis (Fig. 3, groups 16 and 26, respectively). This is most likely due to the fact that these regions have similar environmental conditions, varying from arid to semi-arid⁴⁹. For the genera of Prionoglarididae, the fact that they evolved under semiarid climate and oligotrophic conditions is even related to the coevolution of some of their adaptations⁵⁰. For the Sicariidae genera, this is related to the maintenance of their ecological niche⁴⁷. Certainly, this comparative exercise can also be carried out between caves of other Trogleregions, provided that phylogenetic and biogeographic studies of the groups are taken into account, as in the example of the families previously mentioned.

In this same line of thought, the idea that in geological time, different regions of the planet were marked by changes in their environmental conditions and biogeographical histories that affected the structuring

of their ecosystems is broadly accepted⁵¹. These were factors taken into account for the delimitation of the surface ecoregions, besides, of course, of their distinct ecological communities⁵. In our results, we found that the borders of the Trogleregions also followed, in part, the borders of the ecoregions. As discussed above, we believe that this is at least partially due to the presence of potential colonizers that the different ecoregions can provide to the caves occurring within their boundaries, especially considering that the whole community was analyzed, not just the troglobitic/stygobitic organisms. Although these relationships are indirect, they indicate that the features of surface ecosystems can predict similar patterns of subterranean fauna in different regions. Knowing how and which environmental conditions were determinant for these patterns are questions that deserve further research.

Another fundamental factor observed in our analyses concerns the different drainage basins nested in the ecoregions. Hydrology is a central component for the maintenance of all the subterranean dynamics, being determinant for the resource availability, where the water flows can carry particulate or dissolved organic matter^{52,53} also determining patterns of faunistic distribution^{15,29,54}. Caves placed in larger drainage basins tend to have greater faunal diversity than caves located in smaller drainages¹⁵. Furthermore, the distribution of subterranean fauna tends to be greater in large drainages due to the greater chances of dispersion¹⁵. In our work, we found that in some situations, within the same ecoregion there may be groups of caves with dissimilar faunistic compositions because they occur in different drainage basins. Most likely, this may be the result of divergent drainage flows on large geographic scales, guiding the distribution of certain invertebrate groups to the caves downstream from the boundaries of their respective drainages. A very interesting example that strengthens this idea comes from a study carried out on the border between Italy and Slovenia, using two species of aquatic isopods of the genus *Asellus*²⁹ as a model. In that region, molecular data showed that the species distributions coincided with the delimitation between the of the Reka and Pivka river drainages²⁹. Another example comes from a study carried out in Germany, where the similarity of aquatic fauna between different regions proved to be dependent on the relief configurations, with connections across large valleys (Rhine River Valley) and between old Pleistocene basins¹⁴.

It is important to note that the examples mentioned above apply strictly to aquatic cave fauna. However, in addition to it, we also verified the dependence of drainages on the similarity patterns of strictly terrestrial fauna. Invertebrate fauna, in general, can be transported through the interstices in the rock matrix, between interconnected cave systems along drainages and from the surface to the underground, being carried involuntarily by the natural water flow and floods^{23,55}. After floods, for example, the structure of the cave community can be altered and the diversity can increase momentarily because of organisms from the epigeal fauna being carried by the water flow, coming from the external environment⁵⁵. In karst landscapes, however, on local scales the boundaries of surface drainages not always coincide with the boundaries of subterranean drainages⁵⁶. In addition to the water infiltration being quite accentuated in these landscapes due to its high porosity, subsurface flows may differ from the surface runoff patterns due to the arrangement of aquifers, which may or may not be confined within the geological barriers⁵⁷. On the other hand, on a regional scale, information about surface drainage is

still essential, as it may show zones of allogeneic recharge of subterranean aquifers, where the water flow may come from other non-karstic reliefs⁵⁸. In this sense, our analyses showed that some groups of caves in adjacent basins were grouped according to the similarity of their fauna. As discussed, a possibility for this occurrence would be the possible dispersion of fauna between them, due to subsurface drainage flows that are not aligned with surface flows. However, this is another point that needs to be more adequately explored in the future on more refined spatial scales for a more precise diagnosis.

As shown, caves from different regions can have completely different communities and this certainly needs to be taken into account when preservation strategies are devised. Considering the simplicity of the trophic dynamics, characteristic of subterranean environments¹¹, the different faunal compositions among Trogleregions imply that the trophic functions in these environments are performed by different taxa and guilds into each of them. This represents a new range of hypotheses to be explored and considered in other projects, reinforcing the potential that caves and their biological communities offer for ecological, biogeographic, and evolutionary studies²⁴. On the other hand, all this potential contrasts with the unbridled exploitation of the natural resources on the landscapes related to the caves, which locally can represent the complete destabilization of this admittedly fragile and little known ecosystems¹². This is probably the biggest practical implication of our results: the indication that the local loss of subterranean communities can be irreparable and without equivalents on a regional and even continental scale.

As an example, in Brazil, environmental defense mechanisms recommend that enterprises that cause environmental impact should present compensatory measures to mitigate them (Law Number 9985, 2000⁵⁹). However, in the legislation that regulates the exploitation of karst landscapes, there is no mention of compensatory measures that respect the faunistic identities that different regions may present (Decree No. 6640, 2008⁶⁰). Another alarming gap is the absence of any mention of measures aimed at maintaining hydrological dynamics in subterranean environments and karst landscapes, which contrasts with the fact that these landscapes are an important natural freshwater reservoir⁶¹. This gap needs to be filled as our finds show that drainage basins are one of the central pillars of large-scale subterranean fauna compositional patterns.

In the context of the subterranean environment conservation, a question inevitably arises: if caves in a certain area have been destroyed, caves from which other regions or geological units nearby can potentially present fauna with similar composition and that may be targeted for preservation actions? To solve this type of problem, methodologies similar to those we used for the delimitation of Trogleregions can represent an important guide for possible decision making, in any region of the planet that are replicated. We encourage the scientific community to replicate this methodology and improve it where possible so that the understanding of the macroecological patterns of subterranean communities can be tested, interpreted and disseminated in a more feasible way to assist conservation actions. The bioregional delimitation of surface ecosystems has already proved to be fundamental to verify which

ecoregions on the planet need greater attention for preservation actions³³. For subterranean environments, we hope to have taken some steps that can help in this task.

Methods

Study area

For the survey of biotic and abiotic data, we considered 234 caves in the Afrotropical and Neotropical regions (supplementary material S2), which are found in five countries: South Africa, Brazil, Guatemala, Ghana, and Namibia. These caves are concentrated between 17 ° N and 26 ° S latitudes (supplementary material S1, supplementary Fig. 1) and overlap with 2 biogeographic domains, 3 lithologies, 5 biomes, 16 ecoregions and 26 drainage basins (supplementary material S2).

Biotic data

For the biotic data collection, we first considered the data belonging to the repository from the Center of studies on Subterranean Biology - CEBS, (<http://www.biologiasubterranea.com.br/en/>) which comprised 191 caves distributed in Brazil, Guatemala and Namibia. Then, we considered data from scientific articles published in peer-reviewed journals, which comprised 43 other caves located in Brazil, South Africa and Ghana⁶²⁻⁶⁶. We defined three criteria for the selection of articles used to build the database, which limited considerably the inclusion of caves. Such criteria were: a) Articles in which the emphasis was clearly on surveying fauna or ecological studies involving the entire invertebrate community; b) Articles that provided georeferenced locations of the inventoried caves; c) Articles in which the methodology also indicated active collection (direct intuitive search⁶⁷), where the collection team searches the entire cave in search of organisms. Having met all the criteria, we used the data on the occurrence of invertebrate families of these caves. The use of family-level was due to two main reasons. First, due to its older origins and distributions generally broader than genera and species, enabling comparisons between areas on a wide geographic scale^{1,38}. Second, we assumed that the sampling and group identification biases have been reduced and homogenized with these measures. Therefore, further taxonomic refinements were avoided. Such refinements have a greater probability of identification errors and uncertainties regarding the distribution of species or genera. These problems have been reported as a major obstacle to studies in invertebrate ecology^{32,68}. The nomenclature of families included in the construction of the database was checked in the literature to verify their current taxonomic situation, where we were able to correct possible synonyms or reclassifications. In this case, just for the construction of the graphs in figure 1, 57 families were assigned to the taxon "Acari", which is a generic term used to designate six orders of arthropods whose phylogeny is still not well resolved⁶⁹.

Abiotic data

The abiotic data encompassed information about the surface ecological regions, lithologies, and drainage basins that overlapped the locations of the caves. The delimitations of the ecological regions were obtained from shapefiles with the biogeographical domains, biomes, and ecoregions, on the

Ecoregions2017³³ platform. The delimitations of the drainage basins were obtained from shapefiles on the HydroSHEDS³⁶ platform, where the drainage borders are provided in files with the name “bas” for each continent⁷⁰. These delimitations indicate divergent drainage flows in the landscape, according to water flow modeling from radar images from the Shuttle Radar Topography Mission (SRTM)³⁶. For contextualization, the definition of the drainage basin we used was that of “a landscape in which the surface waters converge to a single location, such as a point in a stream or river, or a single wetland, lake or other body of water”⁷¹.

The locations of the 234 caves were surveyed in the articles considered and into the CEBS database. These locations were then projected in decimal degrees and using the wgs84 datum to meet the specifications of the previously mentioned shapefiles. From the locations, the values for each shapefile were extracted with the *Point Sampling Tool* from Qgis 3.8.3 software⁷², and later used in the other steps. The caves, according to information gathered from the CEBS database and from the literature used, belong to carbonate, siliciclastic and granitoid lithologies (Dataset S1).

Statistical procedures

Before testing our hypotheses, we first verified a possible effect of spatial autocorrelation on the similarity patterns of cave fauna, where considerable positive effects are recurrent in this field of study due to the restricted patterns of subterranean fauna dispersion¹³. For this, we used Distance-based Linear Models (DistLM) to calculate a simple regression⁷³ between the similarity matrix, obtained with the Sørensen–Dice index, and the geographic distance matrix, obtained with the georeferenced locations of the caves. For these cases, these multivariate models are treated as a more robust alternative for the Mantel test⁷³, which is often used to test such influences⁷⁴. In the regression obtained with the DistLM, the criterion used for the calculation was R^2 ⁷³, with the X and Y coordinates treated as a “geographical distance” indicator and being considered the predictor variable.

Factor testing

The analysis of the factors was carried out systematically, respecting their spatial hierarchy, so that we could filter all possible significant variations within the groups of caves according to the test of each factor.

To form spatial aggregations of caves based on the factors that best represent variations in compositional similarity, we explored five different factors, both separately and combined, with their different number of levels in parentheses (infographic in figure 6): biogeographic domain (2), lithology (3), biome (5), ecoregion (16), and drainage basin (26). We have selected these five factors due to their reported influence on cave communities, where in some ways subterranean communities can be influenced by cave lithology³⁷; by the regional pool of surface species²⁴, an influence that was tested here using the boundaries of the biogeographic domains, biomes, and ecoregions; and finally, by the drainage

flows in the landscape, which influence had been only reported acting on strictly aquatic cave organisms until then^{29,54}.

The similarity between cave fauna was obtained using the Sørensen–Dice index, which is less subject to loss of sensitivity in highly heterogeneous datasets⁷⁵. In addition to being heterogeneous, we found in our data a large asymmetry in the experimental design, where there was a variation in the number of levels of a nested factor within the higher-level factor. For these reasons, we tested significant differences in the faunistic composition of the caves grouped by each factor with the Permutational Multivariate Analysis of Variance (PERMANOVA) in two different designs. PERMANOVA was chosen because it allows testing of the effect of factors by obtaining p-values in highly unbalanced experimental designs, enabling more robust interpretations for these cases^{73,76}.

In the first design, p values were obtained for each factor separately after 9999 permutations with the Type III Partial and Unrestricted Permutation of Raw Data⁷³ criteria. The second design consisted of nesting the five factors hierarchically, with the first factor fixed (biogeographic domain) and the others with their randomized levels (lithology, biome, ecoregion, and drainage basin), obtaining the p-values after 9999 permutations with the Type III Partial and Permutation of Residuals Under Reduced Model⁷³ criteria.

The nested combination of these factors, from the one with the highest number of levels to the one with the lowest number of levels (From “Drainage basin”, with 26 levels, to “biogeographic domain”, with two levels) resulted in the distribution of the 234 caves in 37 groups. Therefore, the caves of each group belong to a combination: drainage basin (Dr) “A”, ecoregion (Ec) “B”, biome (Bi) “C”, lithology (Li) “D” and biogeographic domain (Do) “E” (see the schematic model in figure 7). Despite the low influence shown by the lithology factor in the similarity patterns (PERMANOVA $p < 0.001$, ANOSIM $R = 0.063$), we opted to maintain this factor in the combination nested with the others tested factors. This standardization of cave groups regarding their lithology is important, as this is a factor that may have a more evident influence on taxonomic categories lower than the level of invertebrate families³⁷.

With the significance assessed, the global R from the Analysis of Similarity (ANOSIM) was used to verify the quality of the cave groups formed according to the factors¹⁴, where the higher the global R, the greater the difference between the levels of the factor⁷³. As with PERMANOVA, here we tested the factors separated and combined hierarchically. For ANOSIM, 999 permutations were used.

In addition to being applied to the whole biotic dataset, the aforementioned analyses were applied separately only to groups of strictly terrestrial organisms, to aquatic groups (here, we consider as aquatic groups those families with species that need or live associated with water bodies by least at some stage of its life cycle) and excluding singletons from the data. We decided to show only results of the complete dataset in the main text due to its greater robustness since we have found results that showed the same patterns in the other approaches (supplementary material S1, supplementary tables 1, 2 and 3). For these reasons, we carried out the steps described below only for the complete biotic dataset.

For this set, we graphically represented the dissimilarities between the faunistic composition of the caves with the Non-metric Multidimensional Scaling (nMDS), indicating the factors separated and combined hierarchically.

Cave groups delimitation

According to the ANOSIM results, the use of the combined factors was selected due to the higher Global R-value. The presence/absence of family data was then computed at the 37 levels resulting from the nested combination of factors, with each level representing a distinct group of caves. These cave groups served as a base spatial unit for the similarity test between the different regions covered, which was carried out through Cluster analysis (CLUSTER) along with Similarity Profile analysis (SIMPROF), both based on the Sørensen–Dice similarity index. The CLUSTER was obtained using the average linkage clustering method, condensing the samples from the closest neighbors by pairwise comparisons⁷⁷. The SIMPROF analysis added to the CLUSTER is a robust alternative, based on permutations, for the condensation of groups or samples that a priori have an unknown structure based on similarity⁷⁷. With 999 permutations and testing a significance level of 1% (results in the main text) and 5% (supplementary material S1, supplementary figure 2), respectively, SIMPROF indicated the formation of 17 “supergroups” in which caves have statistically distinct faunistic compositions, thus establishing large regional groups of invertebrate families. Based on these 17 supergroups, we carried out the final Troglореgion delimitation.

Finally, with the 37 initial groups resulting from the nested combination of factors, we calculated another PERMANOVA (9999 permutations, Type III Partial and Unrestricted Permutation of Raw Data criteria) and ANOSIM (999 permutations) using the 17 supergroups indicated with CLUSTER and SIMPROF as the clustering factor. The dissimilarity patterns between these groups were plotted with an nMDS.

To obtain a counterproof from the supergroups used to delimit the Troglореgions, we calculated another CLUSTER and SIMPROF with 1 and 5% significance levels, this time using the Jaccard index (supplementary material S1, supplementary figures 3 and 4). The Jaccard index maintains the same weight for shared and unique biological groups, resulting in comparatively lower similarity values⁷⁵. However, these new tests showed us the same supergroups, unchanged. Thus, we have decided to keep the results obtained with the Sørensen–Dice index with $p < 1\%$, which presented more robust values, in the main text.

All the above-mentioned analyzes were performed in the PRIMER 6 software with its PERMANOVA + expansion⁷³.

Map construction

Maps delimiting troglореgions were constructed respecting the hierarchical sequence of the five factors used in the other analyzes. Therefore, the initial cutouts (37, according to PERMANOVA and ANOSIM) consisted of the intersections resulting from the vector layers of the ecoregions with the overlapping

drainage basins (the ecoregions are already naturally nested in the biomes and biogeographical domains). These outlines were obtained with the *Intersection* vector tool⁷². Discontinuities in vector layers without caves were disregarded, and only features with caves were selected. For this, the *Multiparts to Single Parts* vector tool was used⁷².

From the outlines, it was possible to carry out the selection and combination of the 37 groups of caves indicated with PERMANOVA and ANOSIM in the 17 supergroups indicated with the CLUSTER and SIMPROF analyses. For the construction of the maps, we split two of the 17 supergroups for having combined caves from different ecoregions and basins which borders are separated by a great geographical distance. Therefore, we split the supergroup that combined caves from the *Cerrado* ecoregion with caves from the *Bahia coastal forests* ecoregion (formerly the “i” supergroup, now the “*Lower Rio Pardo*” and “*Upper Paraná*” Trogleregions). The other splitted supergroup had combined caves in the *Caatinga* ecoregion, in the Neotropical region, with caves in the *Angolan mopane woodlands* ecoregion, in the Afrotropical region (formerly “n” supergroup, now “*Kalahari*” and “*Lower São Francisco*” Trogleregions). Four "supergroups" in the CLUSTER/SIMPROF (A, C, F, and K) have only one cave and are considered outliers⁷⁸, they have very different compositions. This did not allow their combination with the other supergroups according to the analyses. Of these four, the supergroup “k” was combined with “l” because they overlap spatially for sharing attributes (except the lithology, supergroup k has a limestone lithology, while l has siliciclastic lithology), thus forming the “*Eastern Amazon*” Trogleregion. After these adjustments, the maps indicate 18 Trogleregions, which were named according to some striking geographic feature in their locations.

The shapefile resulting from the maps construction and Trogleregions delimitation contains all information inherent to the factors combined in its attribute table are available as supplementary files (supplementary material S3, and supplementary material S1, supplementary table 4), aiming at its application in future studies and sharing with the scientific community.

The steps described for the construction of the map were performed using Qgis 3.8.3⁷² software.

Data availability

Three supplementary files with additional information for this study are available (<https://figshare.com/s/bf641fb4cd64a99f9ae6>). Supplementary material S1 contains four supplementary figures and four supplementary tables with further data exploration. Supplementary material S2 contains a .xlsx file with all biotic and abiotic data, and the factors used. Supplementary material S3 contains the Trogleregions shapefile.

Declarations

Competing interests

The authors declare no competing interests.

Acknowledgements

Denizar Alvarenga was sponsored by VALE/SA with 24 months of grants. This paper is part of his master thesis on Applied Ecology at Federal University of Lavras. We would like to thank the researchers who contributed to the construction of the database used here, they are: Maysa Vilella, Vinícius Sperandei, Lucas Rabelo, Diego Bento, Deyvison Souza, Gabrielle Pacheco, Roberta Cerqueira, Rafael Cardoso, and Matheus Simões. Grzegorz Kopij, Rafaela Bastos and Kyle for help in obtaining data in the fieldwork carried out in Namibia and Vitor Junta for helping in the field trips in Brazil. To the authors of articles already published in journals, who made their database available in their articles, which proved to be fundamental for the execution of this project. To Paulo Pompeu and Ivan Magalhães for their contributions to the suggestion to better explore and interpret the database. We also thank the VALE/SA company for all support provided to CEBS/UFLA. RLF is grateful to CNPq for support for research N° 308334/2018-3.

Author's contribution statement

DAA and RLF conceived the Trogleregions idea and formulated the hypotheses to test the aggregation of the caves on a continental scale, based on their faunistic similarity. DAA, MSS and RLF developed the discussions pertinent to the construction of the entire article. DAA and RLF prepared the experimental design. DAA performed the analyses and designed the final version of the manuscript with the approval of all authors.

References

1. Barry Cox, C., Moore, P. D. & Ladle, R. J. Patterns of Distribution: Finding a Home. in *Biogeography, An Ecological and Evolutionary Approach* (eds. Barry Cox, C., Moore, P. D. & Ladle, R. J.) 33–88 (Wiley Blackwell, 2016).
2. Schröder, B. Challenges of species distribution modeling belowground. *J. Plant Nutr. Soil Sci.* **171**, 325–337 (2008).
3. Hortal, J., Roura-Pascual, N., Sanders, N. J. & Rahbek, C. Understanding (insect) species distributions across spatial scales. *Ecography (Cop.)*. **33**, 51–53 (2010).
4. Liu, H., Ye, Q. & Wiens, J. J. Climatic-niche evolution follows similar rules in plants and animals. *Nat. Ecol. Evol.* **4**, 753–763 (2020).
5. Olson, D. M. *et al.* Terrestrial Ecoregions of the World: A New Map of Life on Earth. *Bioscience* **51**, 933 (2001).
6. Pianka, E. R. Latitudinal Gradients in Species Diversity: A Review of Concepts. *Am. Nat.* **100**, 33–46 (1966).

7. Walter, H. & Box, E. Global classification of natural terrestrial ecosystems. *Vegetatio* **32**, 75–81 (1976).
8. Schultz, J. *The Ecozones of the World: The Ecological Divisions of the Geosphere*. (Springer, 2002).
9. Abell, R. *et al.* Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *Bioscience* **58**, 403–414 (2008).
10. Smith, J. R. *et al.* A global test of ecoregions. *Nat. Ecol. Evol.* **2**, 1889–1896 (2018).
11. Gibert, J. & Deharveng, L. Subterranean Ecosystems: A Truncated Functional Biodiversity. *Bioscience* **52**, 473–481 (2002).
12. Mammola, S. *et al.* Scientists' warning on the conservation of subterranean ecosystems. *Bioscience* **69**, 641–650 (2019).
13. Zagamajster, M., Malard, F., Eme, D. & Culver, D. C. Subterranean Biodiversity Patterns from Global to Regional Scales. in *Cave Ecology* (eds. Moldovan, O. T., Kováč, L. & Halse, S.) 195–228 (Springer Nature Switzerland AG, 2018).
14. Stein, H. *et al.* Stygoregions—a promising approach to a bioregional classification of groundwater systems. *Sci. Rep.* **2**, 1–9 (2012).
15. Poulson, T. L. & White, W. B. The Cave Environment. *Science* (80). **165**, 971–981 (1969).
16. Culver, D. C. Analysis of Simple Cave Communities I. Caves as Islands. *Evolution* (N. Y). **24**, 463–474 (1970).
17. Prous, X., Ferreira, R. L. & Martins, R. P. Ecotone delimitation: Epigeal-hypogean transition in cave ecosystems. *Austral Ecol.* **29**, 374–382 (2004).
18. Moseley, M. Are all caves ecotones? *Cave Karst Sci.* **36**, 53–58 (2010).
19. Fišer, Ž., Prevorčnik, S., Lozej, N. & Trontelj, P. No need to hide in caves: shelter-seeking behavior of surface and cave ecomorphs of *Asellus aquaticus* (Isopoda: Crustacea). *Zoology* **134**, 58–65 (2019).
20. Trontelj, P., Blejec, A. & Fišer, C. Ecomorphological convergence of cave communities. *Evolution* (N. Y). **66**, 3852–3865 (2012).
21. Mammola, S. & Isaia, M. The ecological niche of a specialized subterranean spider. *Invertebr. Biol.* **135**, 20–30 (2016).
22. Caley, M. J. & Schluter, D. The Relationship between Local and Regional Diversity. *Ecology* **78**, 70–80 (1997).
23. Trajano, E. & Cobolli, M. Evolution of Lineages. in *Encyclopedia of Caves* (eds. White, W. B. & Culver, D. C.) 295–304 (Elsevier Inc., 2012).
24. Mammola, S. Finding answers in the dark: caves as models in ecology fifty years after Poulson and White. *Ecography* (Cop.). **41**, 1–21 (2018).
25. Culver, D. C., Christman, M. C., Elliott, W. R., Hobbs, H. H. & Reddell, J. R. The North American obligate cave fauna: Regional patterns. *Biodivers. Conserv.* **12**, 441–468 (2003).
26. Christman, M. C., Culver, D. C., Madden, M. K. & White, D. Patterns of endemism of the eastern North American cave fauna. *J. Biogeogr.* **32**, 1441–1452 (2005).

27. Bregović, P. & Zagamajster, M. Understanding hotspots within a global hotspot – identifying the drivers of regional species richness patterns in terrestrial subterranean habitats. *Insect Conserv. Divers.* **9**, 268–281 (2016).
28. Zagamajster, M. *et al.* Geographic variation in range size and beta diversity of groundwater crustaceans: Insights from habitats with low thermal seasonality. *Glob. Ecol. Biogeogr.* **23**, 1135–1145 (2014).
29. Konec, M., Delić, T. & Trontelj, P. DNA barcoding sheds light on hidden subterranean boundary between Adriatic and Danubian drainage basins. *Ecohydrology* **9**, 1304–1312 (2016).
30. Leys, R., Watts, C. H. S., Cooper, S. J. B. & Humphreys, W. F. Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. *Evolution (N. Y.)* **57**, 2819–2834 (2003).
31. Bryson, R. W., Prendini, L., Savary, W. E. & Pearman, P. B. Caves as microrefugia: Pleistocene phylogeography of the troglophilic North American scorpion *Pseudouroctonus reddelli*. *BMC Evol. Biol.* **14**, 16 (2014).
32. Jaffé, R. *et al.* Reconciling mining with the conservation of cave biodiversity: A quantitative baseline to help establish conservation priorities. *PLoS One* **11**, 1–17 (2016).
33. Dinerstein, E. *et al.* An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *Bioscience* **67**, 534–545 (2017).
34. Axelrod, D. I. & Raven, P. H. Late Cretaceous and Tertiary vegetation history of Africa. in *Biogeography and Ecology of Southern Africa* (ed. Werger, M. J. A.) 77–130 (Dr W. Junk bv Publishers, 1978).
35. Lavina, E. L. & Fauth, G. Evolução Geológica da América do Sul nos Últimos 250 Milhões de Anos. in *Biogeografia da América do Sul, Padrões & Processos* (eds. Carvalho, C. J. B. & Almeida, E. A. B.) 3–13 (Roca, 2011).
36. Lehner, B., Verdin, K. & Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos, Transactions, American Geophysical Union. EOS, Trans. Am. Geophys. Union* **89**, 93–104. (2008).
37. Souza-Silva, M., Martins, R. P. & Ferreira, R. L. Cave lithology determining the structure of the invertebrate communities in the Brazilian Atlantic Rain Forest. *Biodivers. Conserv.* **20**, 1713–1729 (2011).
38. Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science (80-)*. **235**, 167–171 (1987).
39. Ford, D. & Williams, P. Introduction to Karst. in *Karst hydrogeology and geomorphology* (eds. Ford, D. & Williams, P.) 1–8 (John Wiley & Sons, 2007).
40. Deharveng, L. & Bedos, A. Biodiversity in the tropics. in *Encyclopedia of Caves* (eds. White, W. B., Culver, D. C. & Pipan, T.) 146–162 (Elsevier Inc., 2019).
41. Cardoso, P., Pekár, S., Jocqué, R. & Coddington, J. A. Global patterns of guild composition and functional diversity of spiders. *PLoS One* **6**, (2011).

42. Heino, J. & Soininen, J. Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? *Biol. Conserv.* **137**, 78–89 (2007).
43. Gaston, K. J., Williams, P. H., Eggleton, P. & Humphries, C. J. Large scale patterns of biodiversity: Spatial variation in family richness. *Proc. R. Soc. B Biol. Sci.* **260**, 149–154 (1995).
44. Ferreira, R. L., Martins, V. M., Paixão, E. A. & Souza-Silva, M. Spatial and temporal fluctuations of the abundance of neotropical cave-dwelling moth *Hypena* sp. (Noctuidae, Lepidoptera) influenced by temperature and humidity. *Subterr. Biol.* **16**, 47–60 (2015).
45. Mammola, S. & Isaia, M. Day-night and seasonal variations of a subterranean invertebrate community in the twilight zone. *Subterr. Biol.* **27**, 31–51 (2018).
46. Mammola, S. *et al.* Local-versus broad-scale environmental drivers of continental β -diversity patterns in subterranean spider communities across Europe. *Proc. R. Soc. B Biol. Sci.* **286**, (2019).
47. Magalhaes, I. L. F. *et al.* Phylogeny of Neotropical Sicarius sand spiders suggests frequent transitions from deserts to dry forests despite antique, broad-scale niche conservatism. *Mol. Phylogenet. Evol.* **140**, 106569 (2019).
48. Yoshizawa, K., Lienhard, C., Yao, I. & Ferreira, R. L. Cave insects with sex-reversed genitalia had their most recent common ancestor in West Gondwana (Psocodea: Prionoglarididae: Speleketorinae). *Entomol. Sci.* 334–338 (2019) doi:10.1111/ens.12374.
49. Beck, H. E. *et al.* Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* **5**, 1–12 (2018).
50. Yoshizawa, K., Ferreira, R. L., Yao, I., Lienhard, C. & Kamimura, Y. Independent origins of female penis and its coevolution with male vagina in cave insects (Psocodea: Prionoglarididae). *Biol. Lett.* **14**, (2018).
51. Ricklefs, R. E. A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7**, 1–15 (2004).
52. Gers, C. Diversity of energy fluxes and interactions between arthropod communities: From soil to cave. *Acta Oecologica* **19**, 205–213 (1998).
53. Souza-Silva, M., Bernardi, L. F. O., Martins, R. P. & Ferreira, R. L. Transport and Consumption of Organic Detritus in a Neotropical Limestone Cave. *Acta Carstologica* **41**, 139–150 (2012).
54. Fong, D. W. & Culver, D. C. Fine-scale biogeographic differences in the crustacean fauna of a cave system in West Virginia, USA. *Hydrobiologia* **287**, 29–37 (1994).
55. Pacioglu, O. *et al.* The multifaceted effects induced by floods on the macroinvertebrate communities inhabiting a sinking cave stream. *Int. J. Speleol.* **48**, 167–177 (2019).
56. Ford, D. & Williams, P. Analysis of Karst Drainage Systems. in *Karst Hydrogeology and Geomorphology* (eds. Ford, D. & Williams, P.) 145–203 (John Wiley & Sons, 2007).
57. Ford, D. & Williams, P. Karst Hydrogeology. in *Karst Hydrogeology and Geomorphology* (eds. Ford, D. & Williams, P.) 103–140 (John Wiley & Sons, 2007).

58. Olarinoye, T. *et al.* Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater. *Sci. Data* **7**, (2020).
59. Brazil. *Law No. 9885 of July 18, 2000.* (2000).
60. Brazil. *Law No. 6640 of November 7, 2008.* (2008).
61. Goldscheider, N. *et al.* Global distribution of carbonate rocks and karst water resources. *Hydrogeol. J.* (2020) doi:10.1007/s10040-020-02139-5.
62. Trajano, E. Cave Faunas in the Atlantic Tropical Rain Forest: Composition, Ecology, and Conservation. *Biotropica* **32**, 882–893 (2000).
63. Bichuette, M. E., Simões, L. B., Zepon, T., von Schimonsky, D. M. & Gallão, J. E. Richness and taxonomic distinctness of cave invertebrates from the northeastern state of Goiás, central Brazil: A vulnerable and singular area. *Subterr. Biol.* **29**, 1–33 (2019).
64. Sharratt, N. J., Picker, M. D. & Samways, M. J. The invertebrate fauna of the sandstone caves of the Cape Peninsula (South Africa): Patterns of endemism and conservation priorities. *Biodivers. Conserv.* **9**, 107–143 (2000).
65. Philips, T. K., DeWildt, C. S., Davis, H. & Anderson, R. S. Survey of the terrestrial arthropods found in the caves of Ghana. *J. Cave Karst Stud.* **78**, 128–137 (2016).
66. Pinto-da-Rocha, R. Invertebrados cavernícolas da porção meridional da província espeleológica do Vale do Ribeira, sul do Brasil. *Rev. Bras. Zool.* **10**, 229–255 (1993).
67. Wynne, J. J., Howarth, F. G., Sommer, S. & Dickson, B. G. Fifty years of cave arthropod sampling: Techniques and best practices. *Int. J. Speleol.* **48**, 33–48 (2019).
68. Cardoso, P., Erwin, T. L., Borges, P. A. V. & New, T. R. The seven impediments in invertebrate conservation and how to overcome them. *Biol. Conserv.* **144**, 2647–2655 (2011).
69. Krantz, G. W. & Walter, D. E. *A manual of Acarology.* (Texas Tech University Press, 2009).
70. Lehner, B., Verdin, K. & Jarvis, A. HydroSHEDS Technical Documentation. 1–27 (2006).
71. Flotemersch, J. E. *et al.* A Watershed Integrity Definition and Assessment Approach to Support Strategic Management of Watersheds. *River Res. Appl.* **32**, 1654–1671 (2016).
72. QGIS.org. QGIS Geographic Information System. (2020).
73. Anderson, M. J., Gorley, R. N. & Clarke, K. R. *Permanova + for Primer: Guide to Software and Statistical Methods.* (Primer-E Ltd, 2008).
74. Legendre, P., Fortin, M. J. & Borcard, D. Should the Mantel test be used in spatial analysis? *Methods Ecol. Evol.* **6**, 1239–1247 (2015).
75. McCune, B. & Grace, J. B. Distance Measures. in *Analysis of Ecological Communities* (eds. McCune, B. & Grace, J. B.) 45–57 (MjM, 2002).
76. Anderson, M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). *Wiley StatsRef Stat. Ref. Online* 1–15 (2017) doi:10.1002/9781118445112.stat07841.
77. Clarke, K. R. & Gorley, R. N. *Primer v6: User Manual/Tutorial.* (Primer-E Ltd, 2006).

Figures

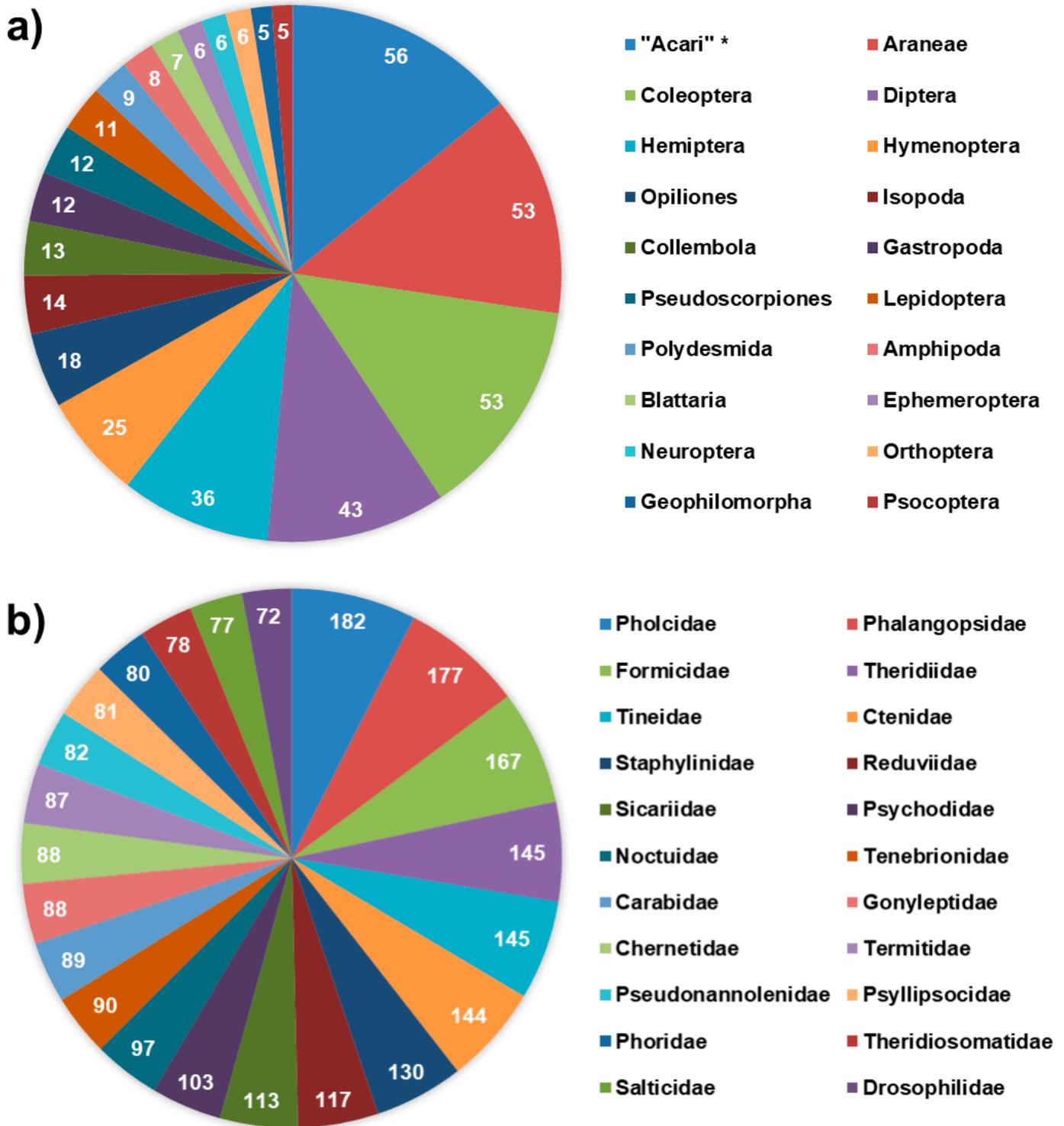


Figure 1

a) Taxa with the largest number of registered families (“Acari” is a generic name for at least six orders of arachnids). The graph shows 20 taxa with at least five families, representing 398 from the 455 families found in the study; b) Families with the highest number of occurrences in the caves included in the study. The graph shows 22 families distributed in at least 70 caves. The complete list of 455 families with their respective distributions in the 234 caves is available in supplementary material S2.

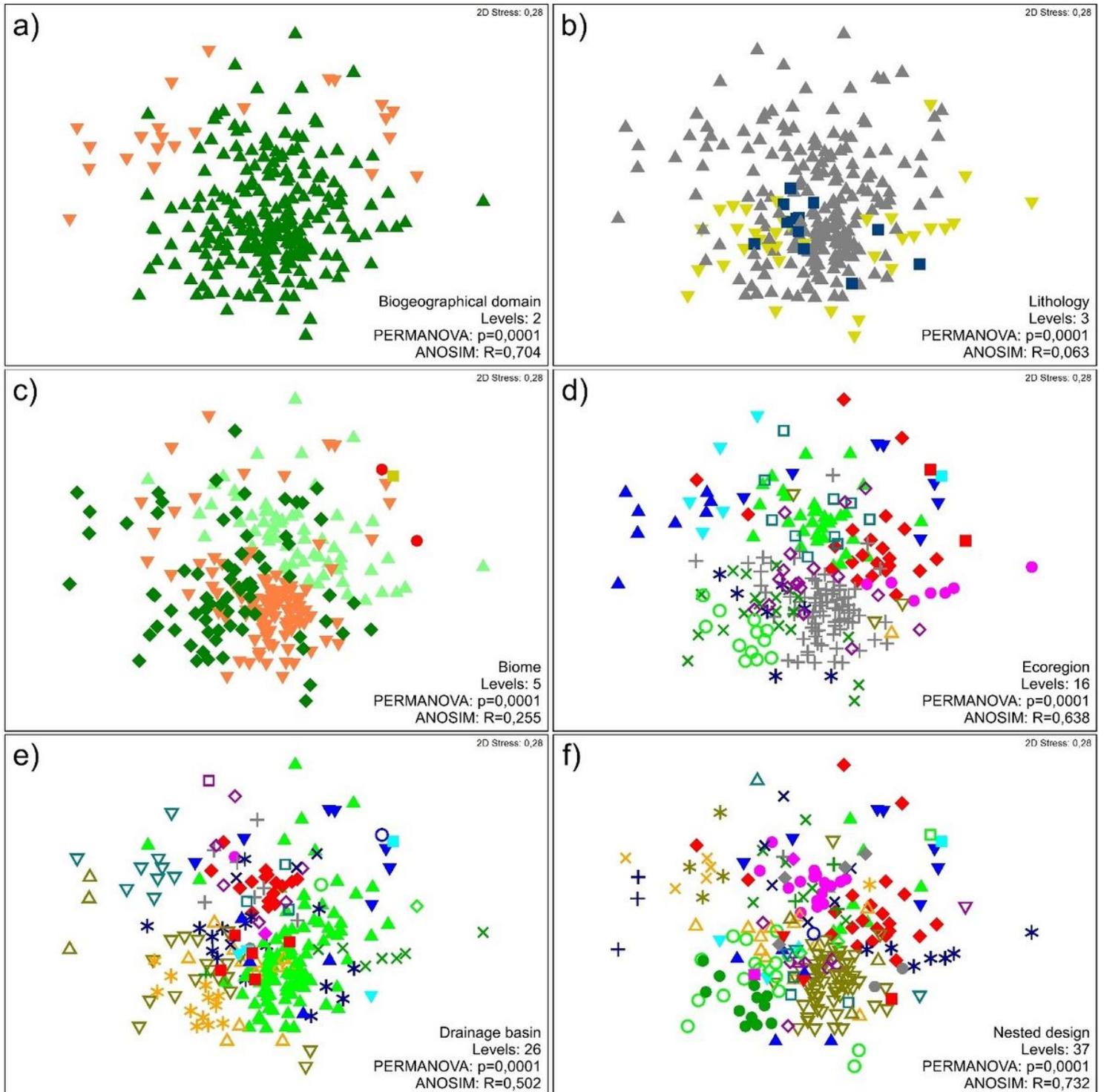


Figure 2

Non-metric Multidimensional Scaling (nMDS) indicating the similarity patterns exhibited by the invertebrate communities in the 234 caves, according to the isolated and hierarchically combined factors (Nested design). In (a), the similarity patterns according to the levels of the factor “biogeographic domain”; In (b), according to the levels of the “lithology” factor; In (c), according to the levels of the “biome” factor; In (d), according to the levels of the “ecoregion” factor; In (e), according to the levels of the “drainage basin” factor; In (f), according to the factor resulting from the nested combination of these five factors, indicating the distribution of the caves at their 37 levels. According to the PERMANOVA results, all these factors resulted in significant variations in the composition of families ($p=0.0001$). The ANOSIM showed that the factor resulting from the nested combination was the most efficient in representing these variations (Global $R=0.732$).

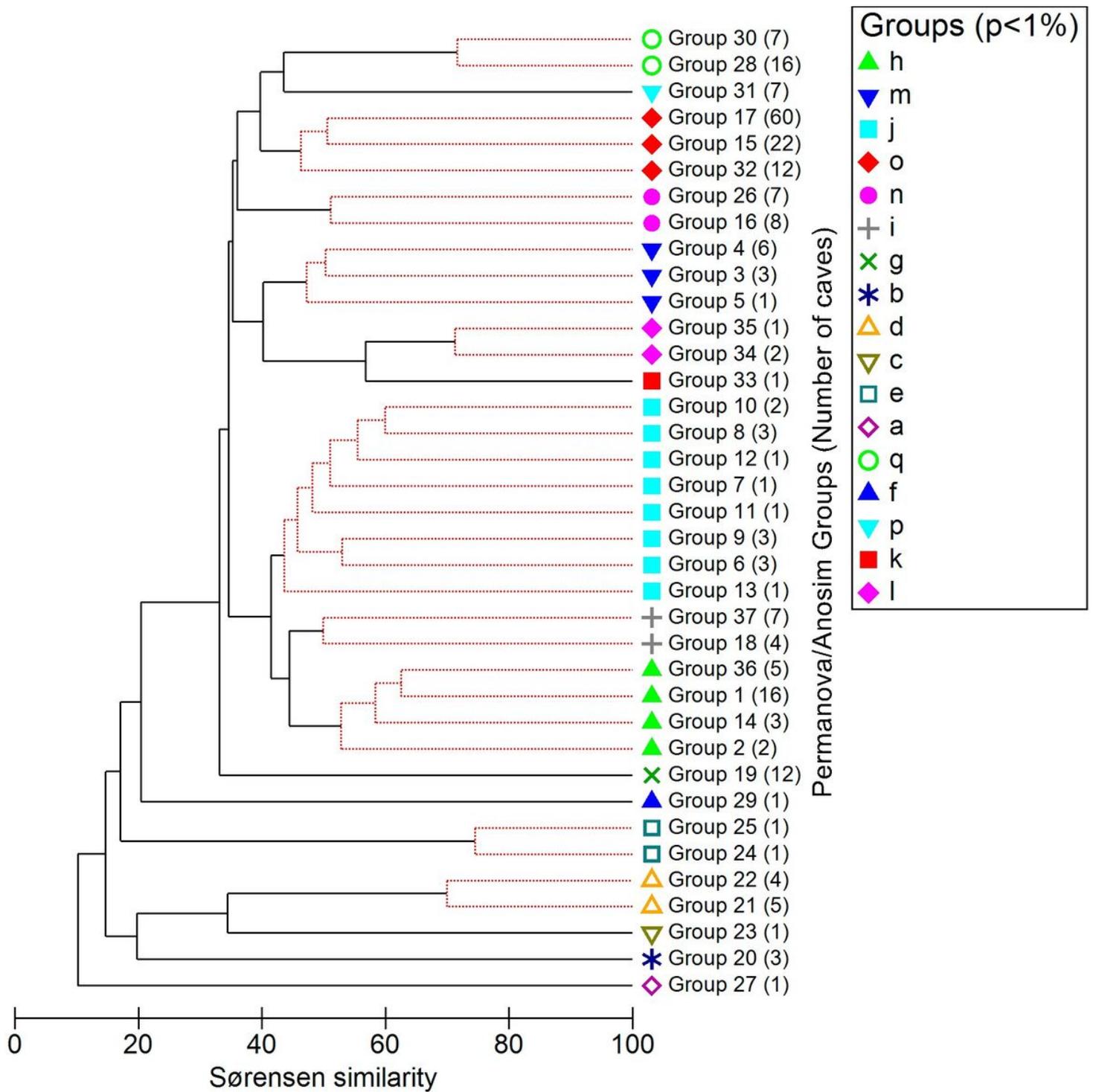


Figure 3

CLUSTER and SIMPROF analyses testing the similarity of fauna attributed to the 37 groups of caves (number of caves in parentheses) resulting from the nested combination of factors: drainage basin-ecoregion-biome-lithology-biogeographic domain. The pairwise comparisons of SIMPROF analysis indicated the formation of 17 “supergroups” (table on the right) with different faunistic compositions, evidenced by the black lines ($p < 1\%$ after 999 permutations, Sørensen–Dice index). The delimitation of Trogleregions was based on these 17 “supergroups”. Results with a $p < 5\%$ cut with the Sørensen–Dice

Index and results using the Jaccard index with a $p < 1\%$ and 5% cuts remained unchanged (See supplementary material S1, supplementary figures 2, 3 and 4).

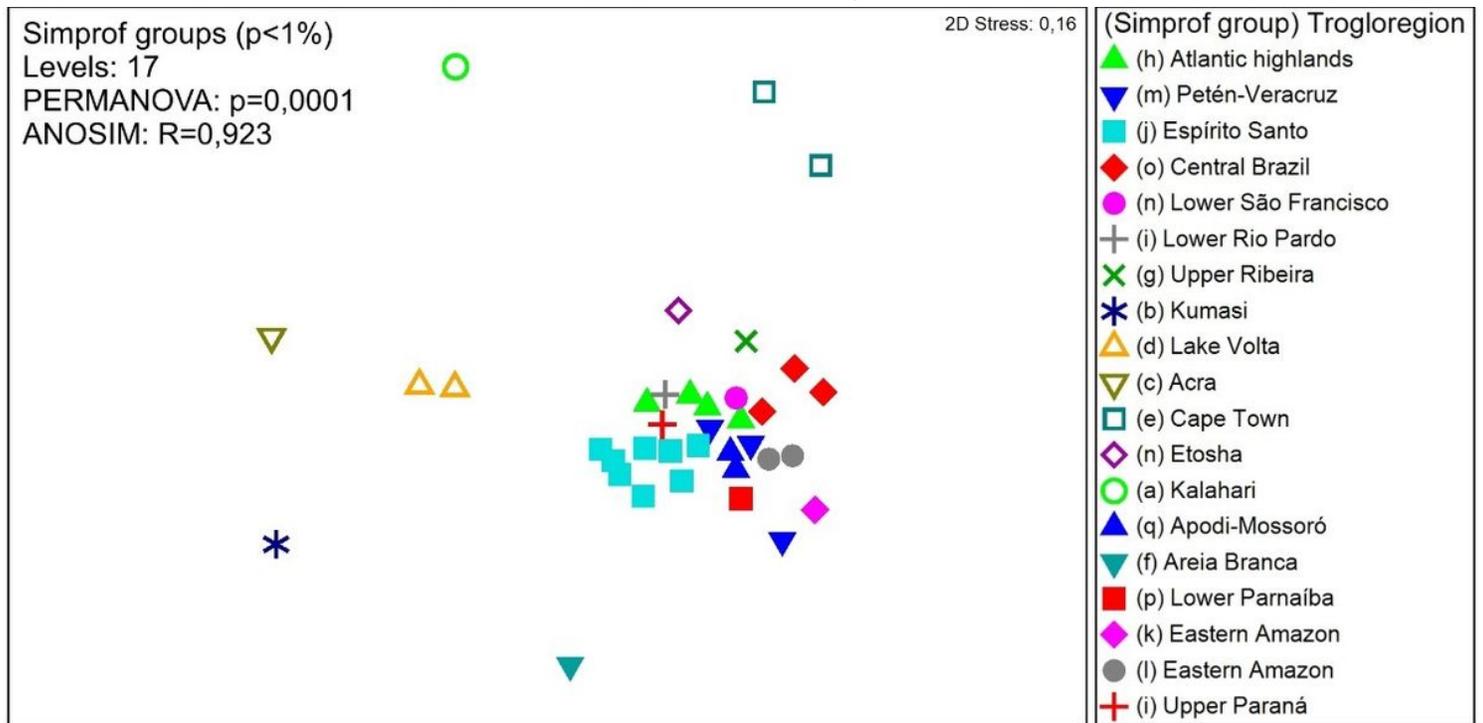


Figure 4

Non-metric Multidimensional Scaling (nMDS), indicating the similarity patterns exhibited by the invertebrate communities of the 37 cave groups resulting from the nested combination of factors (tested in PERMANOVA and ANOSIM), using the 17 “supergroups” indicated by CLUSTER and SIMPROF as clustering factor. This classification obtained the highest accuracy in showing the patterns exhibited by cave communities in the evaluated regions (PERMANOVA: $p=0.0001$; ANOSIM: Global $R=0.923$). These “supergroups” served as the basis for delimiting the Troglореgions maps (figure 5).

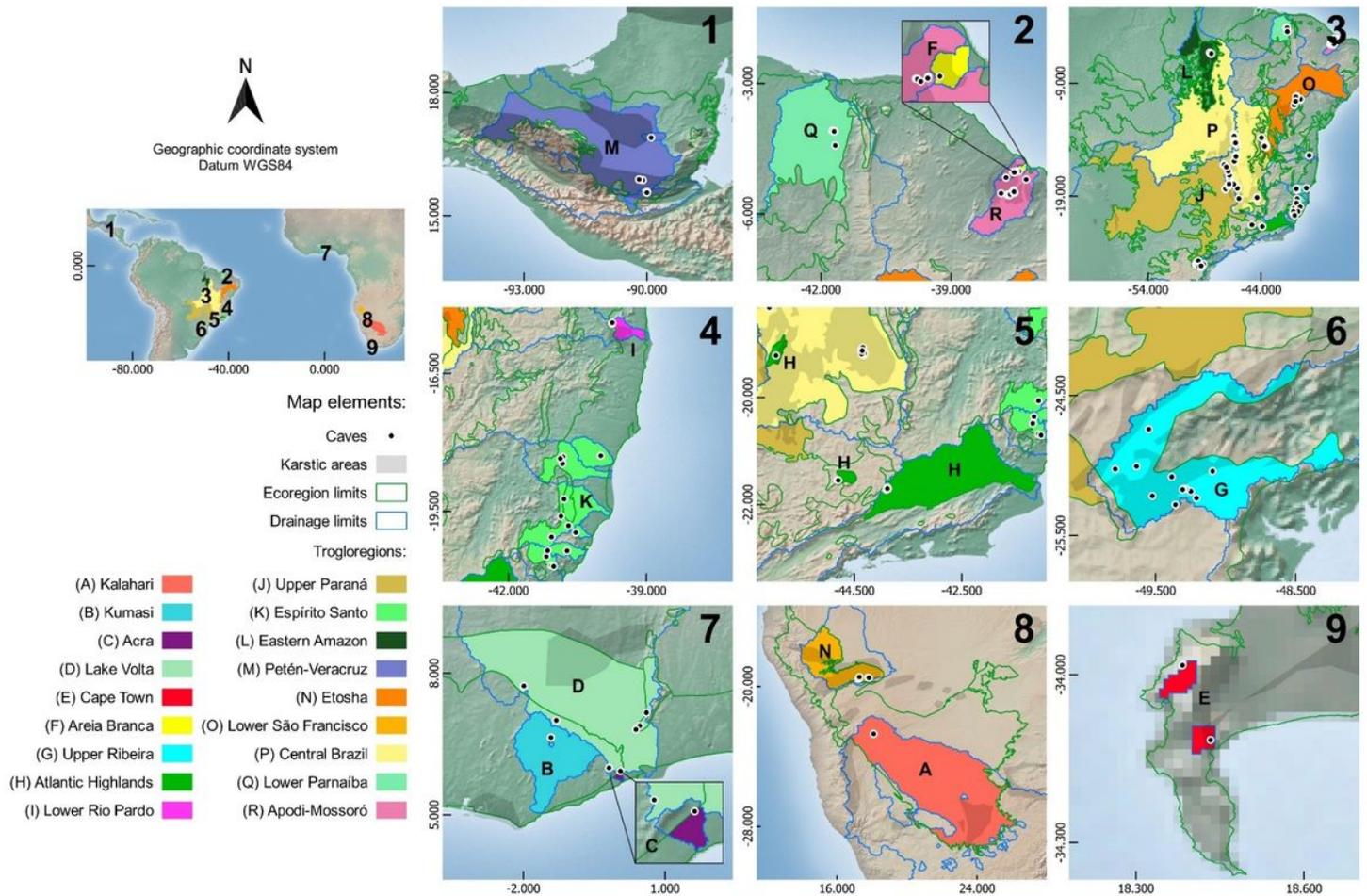


Figure 5

Delimitation of Trogleregions, merging the limits of ecoregions and drainage basins (See methods section). These delimitations correspond only to those that were used to delimit the Trogleregions and that contained caves within their limits. The limits of the ecoregions correspond to the updated version from Olson et al. (2001)5 ecoregions, made by Dinerstein et al. (2017)33. The limits of the drainage basins indicate the boundaries from which water flows are divergent on a continental scale36,70.

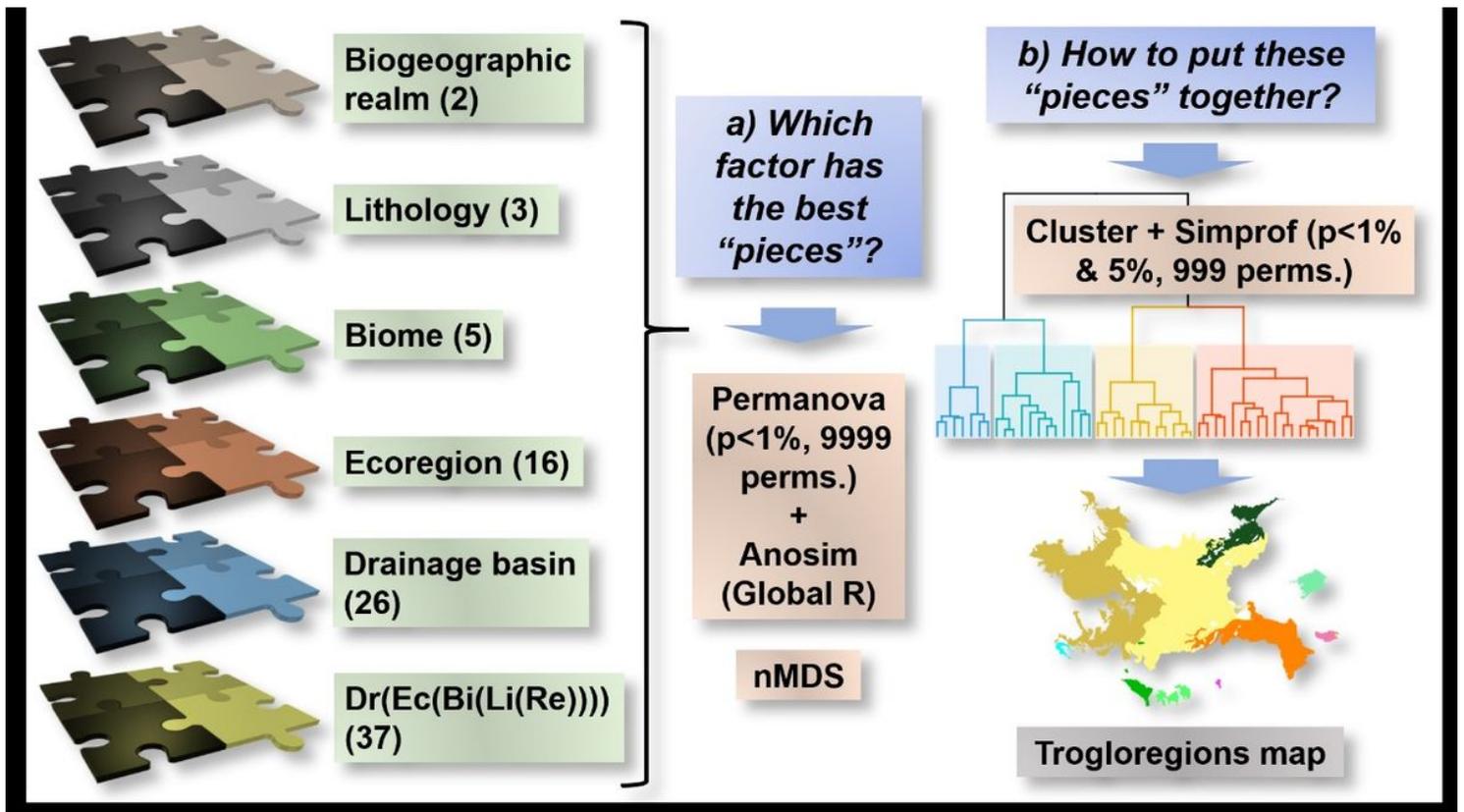


Figure 6

Infographic outlining the analyses performed according to our main questions. From the dataset of invertebrate families occurrence in 234 caves. In (a), we tested the factors separately and combined hierarchically according to their number of levels. Here, these levels are exemplified as "pieces" in a puzzle. After establishing which factor best represented the variation in the biotic data, in (b) these data were assigned to each piece, so that we could test their combination according to their compositional similarity, assembling the Troglореgion maps afterward. All analyses presented were based on data on the presence/absence of invertebrate families and using the Sørensen–Dice index. The protocols and data treatment inherent to the performance of each analysis are described in detail in the methods section.

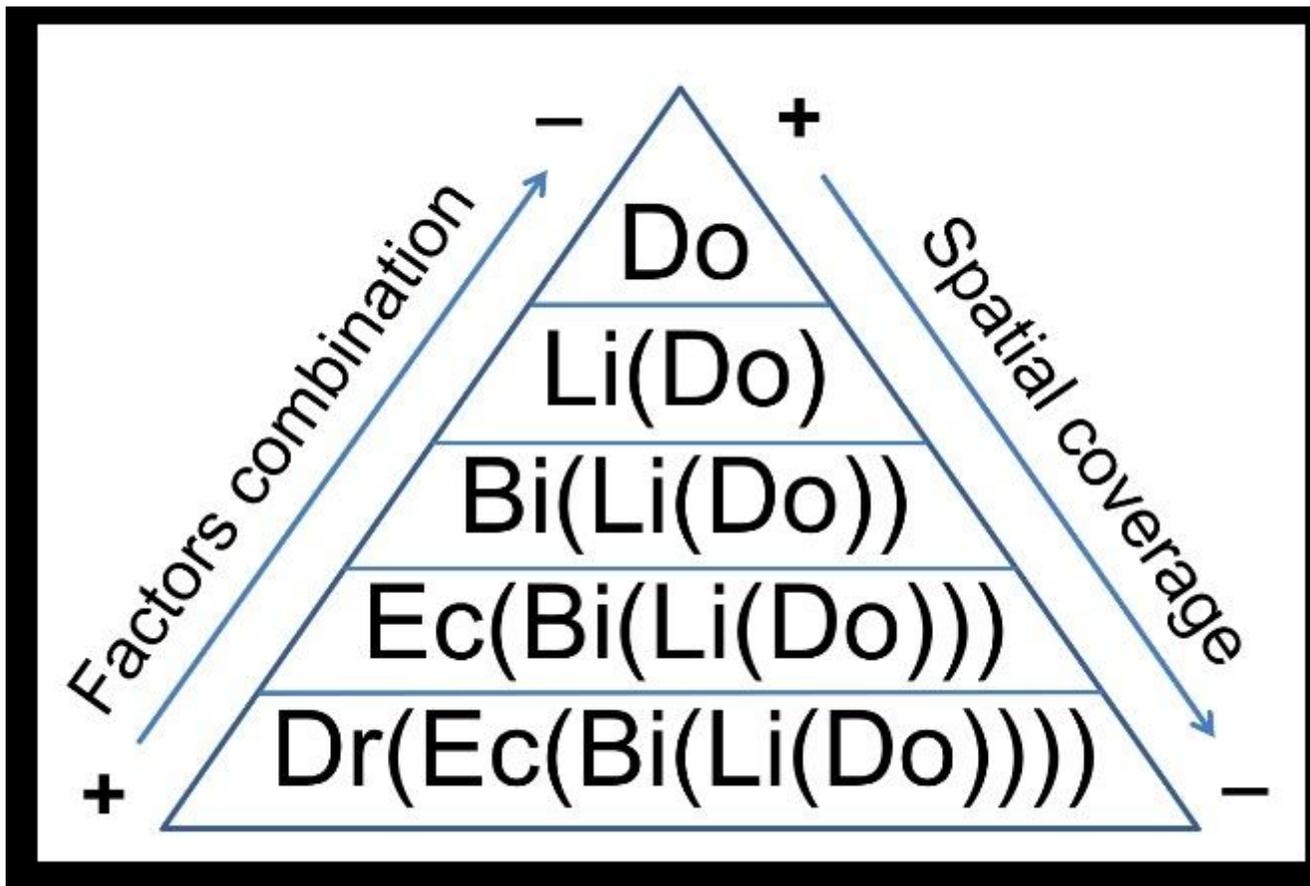


Figure 7

Schematic model representing the nested combination of factors used in the analyses: Do=Biogeographical domain; Li=Lithology; Bi=Biome; Ec=ecoregion; Dr=Drainage basin. The addition of factors in the combination implies a reduction in the spatial scope of each subset and vice versa.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [s1supplementarymaterialtrogleregionsalvarengaetal2020.docx](#)
- [s2bioticabioticfactorstrogloreregionsalvarengaetal2020.xlsx](#)
- [s3trogleregionsshapefilealvarengaetal2020.rar](#)
- [FigS1.jpg](#)
- [FigS2.jpg](#)
- [FigS3.jpg](#)
- [FigS4.jpg](#)
- [TableS1.png](#)
- [TableS2.png](#)

- [TableS3.png](#)
- [TableS4.png](#)