Staphylinid Beetle Diversity Patterns in Afrotropical Mountains, with Particular Reference to the Genus *Megarthrus* (Coleoptera: Staphylinidae)

Håkon Knudsen

Master Thesis in Biodiversity and Systematics
60 Credits

The Natural History Museum
Department of Biosciences
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

February 2024

II

Megarthrus **sp. nov.**

Staphylinid Beetle Diversity Patterns in Afrotropical Mountains, with Particular Reference to the Genus *Megarthrus* (Coleoptera: Staphylinidae)

Håkon Knudsen

Master Thesis in Biodiversity and Systematics
60 Credits

The Natural History Museum
Department of Biosciences
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

February 2024

IV

Acknowledgements

Many thanks are in order for making this thesis not just a possibility, but also a reality. First and foremost, to my main supervisor Vladimir I. Gusarov. Thank you so much for your guidance for the past two years, and for your willingness to let me be part of this project. Throughout the project, you have thrown me head-first into new opportunities and experiences that I will always remember and appreciate.

Thank you, Torsten Hugo Struck, for accepting the role of co-supervisor for the project. Despite the entomological nature of the project, your openness to help and discuss things has been thoroughly reassuring.

Thank you, Galina Gusarova, for taking the time to make suggestions and for pointing me in the right direction regarding the statistical analysis of the thesis.

Thank you to all the members, past and present, of Comparative and Evolutionary Genomics, the research group formerly known as Frontiers in Evolutionary Zoology. Thank you for welcoming me into your group enthusiastically, and for offering assistance whenever needed.

The lab part of the thesis would not have been possible without the help of the wonderful staff at the NHM DNA-lab. In particular, I could have not have done the sequencing without you, Audun Schrøder-Nielsen. Many thanks for your help and your genuine interest in having the project succeed.

Special thanks also to my fellow students, and in particular Line Willersrud and Vineeth Krishnamoorthi, for valuable discussions and good company throughout this journey.

Finally, to the ever so patient members of my family. You have helped me keep going when things have been overwhelming and when the road ahead was unclear, more so than you know. For that, I am immensely grateful.

Håkon Knudsen

Abstract

Reports regarding the current state of biodiversity worldwide tell grim tales of anthropogenic environmental effects causing rapid decline in species abundances across the globe and extinction rates on par with those of our planet's most devastating extinction events. Despite this, only a fraction of the Earth's species have been discovered and formally described ("The Linnean Shortfall"), and knowledge is lacking regarding the distributional patterns of the species we do know about ("The Wallacean Shortfall").

Here, the Staphylinid beetle diversity patterns of the Afrotropics across the continent are explored and compared along the elevational gradients of selected mountains. Taxonomic Hill alpha and beta diversity is measured, and Redundancy Analysis (RDA) and Non-Metric Multidimensional Scaling (NMDS) are conducted on samples collected from Mt. Cameroon in West Africa and the Pare Mountains of the Eastern Arc in East Africa. The data from the younger, higher and less disturbed Mt. Cameroon suggest a monotonic decrease in alpha diversity with increasing elevation. The difference in elevation between sites is found to be the largest determinant of diversity, whereas distance between sites is less important. The Pare mountains are older, lower and highly fragmented. This is reflected in a lack of clear correlation between elevation and diversity, higher dissimilarity between samples from different slopes, as well as a much higher impact of distance between sites on beta diversity.

The phylogenetic relationships within the genus *Megarthrus* are explored by Maximum Likelihood analyses of a dataset based on sequences of 10 nuclear single copy protein-coding genes, obtained with multiplex PCR followed by sequencing on the Oxford Nanopore MinION platform. Fourty-seven species of Proteininae*,* sampled globally, are included, and the phylogeny recovers all the 23 included species of Afrotropical *Megarthrus* as one shallow clade, suggesting single dispersal of the genus into Africa, followed by profuse diversification. Within the Afrotropical clade, there are several regional subclades, restricted to the Ethiopian highlands, and more widely distributed subclades composed of very closely related species occurring across the Afrotropical region. These results imply that *Megarthrus* might be better disperses than previously assumed. The genus *Metopsia* (represented by four species) is nested within *Megarthrus*, rendering the latter paraphyletic.

Finally, one new species of *Megarthrus* from Mt. Cameroon is described, and *M. kamerunensis* Bernhauer, 1942 is redescribed.

Table of Contents

1 Introduction

1.1 Shortfalls in biodiversity knowledge

Biodiversity worldwide is currently facing its largest human mediated challenge ever. The latest global assessment from the IPBES concludes that a majority of indicators of biodiversity and ecosystems show signs of rapid decline, with about 75% of the planet's land surface significantly altered by human drivers (IPBES, 2019). The Living Planet report for 2022 reports an average decline in relative species abundances among monitored populations globally by 69% (WWF, 2022). Reported estimates of extinction rates are currently as high as they were during the five previous mass extinction events in the history of Earth (Pimm et al., 1995), with many referring to the current crisis as the sixth mass extinction (Brito, 2010; Ceballos et al., 2017).

Despite this, our knowledge of the Earth biota is critically lacking. At the core of this issue lies the fact that only a fraction of our planet's species has been discovered, formally described and properly studied (Novotny et al., 2002). Estimates of the total number of eukaryotic species on Earth often range between 5–30 million species, with the most cited estimates of around 8.7 million species (Novotny et al., 2002; Sweetlove, 2011). Of these, only around 1.5 million have been formally described, leaving an estimated 86% of land species and 91% of marine species undiscovered, an issue often referred to as the "Linnean Shortfall" (Brito, 2010; Costello et al., 2013; Lomolino, 2004). Consequently, we know equally little regarding the distribution of species globally, even when the species have already been described: this was dubbed by Lomolino (2004) as the "Wallacean Shortfall". Without knowledge and understanding of the world's species, many may go extinct before we learn of their existence, and we cannot preserve what we do not know is there. These problems are central to contemporary conservation biogeography, and biogeography as a whole (Whittaker et al., 2005).

1.2 Elevational gradients in Afrotropical mountains

Understanding why and how the number of species vary geographically has long been another major focus of biogeography (Brown & Lomolino, 1998). One striking pattern that has been extensively studied for the past decades is that of elevational gradients of biodiversity in mountains. Several characteristics make these gradients even more suitable for examining the patterns and causes of spatial variation in biodiversity than the latitudinal gradients. First, mountains function as replicates of each other. Mountains, or mountain ranges, are generally isolated, and for this reason, the generality of the explored causes of different patterns are more easily tested. Second, local scale experiments are feasible to set up, and data is easier to collect than for the global scale latitudinal gradients (Sanders & Rahbek, 2012). Finally, due to the inherent isolation of the mountains, the underlying mechanisms in different mountains do not covary along the gradients (Korner, 2007).

The biodiversity of many taxa in the mountains of the Afrotropics have yet to be thoroughly explored. The mountains in East and West Africa vary with respect to age, history, size, elevation, climate, altitude and composition. In West Africa, Mt. Cameroon contains the highest sub-Saharan peak, with an elevation of 4040 m. This volcanic mountain belongs to the Cameroon Volcanic Line (CVL), extending from the island of Annabón in the Gulf of Guinea to Mt. Oku in the western highlands of Cameroon (Déruelle et al., 1987). The massifs are of Eocene to recent origin, with Mt. Cameroon itself assessed to be of Upper Miocene age, only a few million years old (Déruelle et al., 1987; Reusch et al., 2010). Due to its location in a humid tropical area, extreme rainfall and high temperatures are commonplace all year long. The highest rainfall occurs at Debundcha in the southwest, with yearly precipitation of 10-12 m/year, deeming it the second wettest place on earth (Ako et al., 2012; Ministry of Forestry and Wildlife, 2014). However, as Mt. Cameroon is located at the coastal margins of the Gulf of Guinea, precipitation is not uniform across the massif (Wembenyui et al., 2020). In the north-east, precipitation may fall below 2000 mm/year, and precipitation also decreases with altitude (Payton, 1993). In addition, as the lithology of the mountain is simple, almost exclusively of basalt rock types, the permeable layers of basalt quickly absorb the precipitation. This makes the area a rich hydrological reservoir, and the surface dries up quicker (Ako et al., 2012). The vegetation types of Mount Cameroon range from the lowland rainforest stretching from the coast up to 800 meters, to the sub-alpine grassland extending from 3000 meters to the summit (Ministry of Forestry and Wildlife, 2014).

2

The mountains in East Africa are highly different. Generally, East Africa is characterized by a dry climate, high temperatures, and dry vegetation types. However, in mountains and mountain ranges in the region, precipitation is much higher, and fragmented forests flourish, resulting in high levels of local endemism (Wasser & Lovett, 1993). The contemporary forests are thought to be relicts of a much larger rainforest stretching across East Africa, with fragmentation starting in the Miocene some 10–15 million years ago (Demenocal, 1995). With the drying of lowland plateaus into savanna or woodland, each mountain or peak is effectively isolated from each other. East of the East African Rift System lies the Eastern Arc Mountains of Tanzania and Kenya. In contrast to the younger and taller mountains in the region, these ranges and mountains are ancient, having been mountains for over 25 million years. The peaks are lower, barely reaching 3000 meters above sea level (Bussmann, 2006). In Tanzania, east of Mt. Kilimanjaro, lies one of the blocks of the Eastern Arc, the range of Pare Mountains consisting of the North Pare Mountains, and the South Pare Mountains. These are massifs made up of multiple fragmented peaks. Unlike Mt. Cameroon, the massifs are no longer volcanically active, and they are generally dryer.

Despite the dryer climate starting in the Miocene, Trauth et al. (2005) did identify three possible periods in which the climate was warmer and wetter during the Pleistocene. These were in the periods between 2.7 to 2.5 Ma, 1.9 to 1.7 Ma and 1.1 to 0.9 Ma. During these periods, lowland forest bridges could have led to migration between the mountains of the massifs. However, high altitude species are unlikely to be able to utilize such bridges, especially those of low dispersal ability, such as flightless insects (Brühl, 1997; Elven, 2013). For such species, the Eastern Arc mountains, and similarly isolated afromontane and afroalpine mountains, are often referred to as "sky islands" alluding to their isolated nature (Elven, 2013; Grebennikov, 2021).

At the onset of studies examining the altitudinal patterns of biodiversity in mountains, results conventionally led to the conclusion that the altitudinal patterns mirror those of the latitudinal patterns, with monotonically decreasing diversity with elevation. Rahbek (1995) showed that for the most part this conclusion is wrong, with most studies when corrected yielding hump-shaped diversity gradients. More recent work point toward multiple possible patterns. In their seminal review of the topic, McCain and Grytnes (2010) categorized observed altitudinal patterns into four major categories:

1. Monotonic decrease with increased elevation.

2. Low elevation plateau followed by decrease with increased elevation.

3. Low elevation plateau followed by a mid-elevation peak.

4. Unimodal with mid-elevation peak.

The monotonically decreasing patterns evenly declines in diversity with elevation, whereas the low-plateau patterns have constant high diversity up until at least 300 meters before changing. The proportion of each pattern reported in studies differ depending on the taxa in question, with no extensive meta-analyses on plants and insects. For vertebrates, however, McCain and Grytnes (2010) report that all four patterns are common, with 45% showing mid-elevation peaks, 26% monotonically decreasing, 15% low-plateaus and 14% low-plateaus with mid-elevation peaks. The causes for these results may be of ecological basis, although some of it may be attributed to methodological issues, such as sampling deficiency, and scaling issues (McCain & Grytnes, 2010; Rahbek, 1995; Sanders & Rahbek, 2012).

1.3   Measures of species diversity

One of the key concerns in measuring and analyzing biodiversity both in general and over gradients like altitude, is choosing the proper measures and tools. The same is true for comparing diversity across assemblages in different locations, both locally and regionally. The topic of measuring and assessing biological diversity is an active one in ecological research (Magurran, 2004; Magurran & McGill, 2010), with important implications for both monitoring and conservation of diversity.

The overall diversity of a region or landscape is often referred to as that regions gamma (γ) diversity as described by Whittaker (1972). The gamma diversity is decomposed into its two major components, the alpha (α) and beta (β) diversities. The alpha component describes the level of species diversity within a focal study area, such as a community, whereas the beta diversity component describes the compositional differences between different study areas (Magurran, 2021; Whittaker, 1972).

Measuring alpha diversity relies on summary measures each emphasizing different features of the abundance distribution of species within the community (Magurran, 2021). Among the

most commonly used indices of diversity are the simple species richness, the Shannon diversity index (Shannon, 1948), the Simpson index and its variations (Simpson, 1949), and the Berger-Parker dominance (Berger & Parker, 1970). The first simply measures the number of species present in the community, whereas the latter all measure heterogeneity in the community each emphasizing dominance or evenness in the community differently. In the literature, the number of different measures and advice on which to choose and how to apply them has long been overwhelming. The topic has for this reason been one of the most controversial in ecology (Magurran & McGill, 2010).

In recent years, however, a consensus has been achieved about using the Hill number framework as a unified framework when analyzing diversity (Chao et al., 2019). The Hill number framework is a mathematical family of diversity measures, first described by Hill (1973) including the three most commonly used indices (richness, Shannon and Simpson) as special cases:

$$^qD = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)}$$

where S is the total number of species and p_i is the relative proportion of each species.

Each Hill diversity is parameterized by q, referred to as its diversity order. The diversity order determines how sensitive the number is to the relative species abundances in the community (Chao et al., 2019). When q = 0, the equation simply reduces to the species richness, as only presence/absence data are retained. As q approaches 1, the limit of the equation tends to the exponential of the classic Shannon index (Chao et al., 2014):

$$^1D = \lim_{q \to 1} {}^qD = exp\left(-\sum_{i=1}^{S} p_i \log p_i\right)$$

Similarly, for q = 2, the equation gives the inverse of the Simpson concentration:

$$^2D = 1 \Big/ \sum_{i=1}^{S} p_i^2$$

Generally, q = 1 weighs each species according to its abundance in the sample, whereas q = 2 gives more weight to the more dominant species.

The most important feature of using the Hill number diversities, giving it an advantage over other measures, is its doubling property. Most other measures of alpha diversity are indices, meaning that they themselves are not true diversities. If you combine two entirely distinct assemblages with the same relative abundances using any such index, the value of the index does not change. Using Hill numbers, the doubled intuitive diversity of such a combined community is reflected in a doubling of the Hill number value (Chao et al., 2014). As a result, Hill numbers may be expressed in units of effective number of species, meaning the number of equally abundant species needed to attain the same value of a diversity measure (Chao et al., 2014).

Diversity may be measured for every value of q. In doing so, continuous diversity profiles may be created for a given community. The curve of such a profile determines how heterogeneity affects diversity (Chao & Jost, 2015). If there is no change as q increases, there is no heterogeneity. If there is a decrease with increasing q, more leverage is given to the more dominant species. Thus, if one expects heterogeneity in a community to be significant, one expects a decrease in value as q increases.

Another issue when measuring biodiversity from samples is the issue of richness being highly sensitive to sample size. In any given community, a few species are common, whereas most species will be rare, and consequently hard to detect in samples. For this reason, observed diversity tends to be underestimates compared to true diversity (Chao et al., 2014). There are two main paths to overcoming this issue in the literature. First, an asymptotic approach seeks to estimate species richness using nonparametric estimators (Chao, 1984; Chao & Chiu, 2016; Chiu, Wang, et al., 2014). This approach is used when comparing estimates across assemblages. The second approach uses extrapolation and interpolation in order to assess samples of the same size or coverage (Chao & Jost, 2012; Colwell et al., 2012). Using coverage-based rarefaction and extrapolation seems superior to the size-based methods for several reasons. First, traditional size-based comparisons may represent one of the samples being compared well if the size is large enough to fully characterize one sample but underrepresents the other. Second, using coverage requires less sampling effort for accurate analyses. Finally, coverage-based analyses throw away less data than size-based analyses do (Chao & Jost, 2012).

1.4   Measures of compositional change

Comparing the compositional differences between study units is just as important as measuring the within-community alpha diversity. As with the measures of within-community diversity, the literature on measures of compositional change is filled with a large number of different approaches (Koleff et al., 2003). Pairwise comparisons of similarity or dissimilarity can be analyzed using simple presence/absence data, such as the Jaccard (1901) similarity measure, or with abundance data, as exemplified by the commonly used Bray-Curtis measure (Magurran, 2021). Which measures or frameworks to use depends on how they are to be applied, sampling, and partitioning schemes. Additionally, many of the simple pairwise measures have been extended to multiple samples, attempting to account for all samples in the dataset when analyzing pairs of samples (Jost et al., 2010).

Obtaining proper measures of similarity or dissimilarity also depends on the diversity decomposition, which decomposes the gamma diversity into its alpha and beta components. Utilizing Whittaker's multiplicative definition of diversity (alpha x beta = gamma), traditional definitions of beta diversity generally yielded results with a hidden dependence on alpha diversity, thus violating the multiplicative definition (Jost, 2007). In order to circumvent this problem, two main partitioning schemes have been developed based on the Hill number framework. When the new alpha and beta components are transformed into their effective number of species as described above, Whittaker's multiplicative law holds true for all indices (Jost, 2007). Routledge's (1979) alpha definition expresses alpha diversity as an assemblage-size-weighted mean of diversities from individual assemblages. In contrast, Chiu, Jost, et al.'s (2014) alpha is an unweighted approach. When comparing datasets of unweighted species relative abundances, the two definitions are equal, and either may be used. If, however, the goal of the study is to assess species raw abundance datasets, Chiu, Jost, et al.'s definition should be used (Chao et al., 2019).

Using Chiu, Jost, et al.'s scheme, beta takes values in the range [1, N], for both incidence-based (q = 0) and abundance-based (q > 0) measures and can be interpreted as the effective number of assemblages based on the data. Because of this, beta can be monotonically transformed into four classes of measures including the incidence-based Jaccard and Sørensen dissimilarity measures when q = 0, and the abundance-based Horn and Morisita-Horn measures when q = 1 and q = 2, for multiple assemblages (Chao et al., 2023). These four

classes, with mathematical definitions and interpretations, can be found in Table 1 of Chao et al. (2019). Chao et al. (2023) have extended their concept of interpolation and extrapolation of alpha diversity measures from sampling to beta diversity utilizing coverage-based methods.

1.5   Coleoptera: Staphylinidae: *Megarthrus*

In this thesis, the focus is on the mega-diverse family of rove beetles in the Afrotropics (Coleoptera: Staphylinidae). Staphylinidae are the largest beetle family, and with the recent inclusion of Silphinae as a subfamily (Cai et al., 2022), the group now entails more than 64 000 species among 35 subfamilies (Newton, 2022). The staphylinids are virtually cosmopolitan, with subfamilies present on every continent except Antarctica. The enormous taxonomic variation within the group is paralleled by an equally large variation in their ecology. Staphylinids are present in most conceivable ecological zones, have a wide variety of feeding habits, and span all elevational zones (Thayer, 2005). However, most are predators of other invertebrates. Generally, staphylinids distinguish themselves from other beetles by their short, truncate elytra, which leaves 5–6 abdominal segments exposed dorsally. Hind wings are present in most species and are folded compactly by a costal hinge proximal to the radial cell. In addition, with some exceptions, procoxae are contiguous (Newton et al., 2000).

The genus *Megarthrus* Stephens, 1829 of the subfamily Proteininae Erichson, 1839 is a genus present in all geographical zones except the South Frigid Zone. The genus occurs from sea level to high mountains in the Northern Temperate Zone, and in the montane forest belt or above in the Afrotropics, avoiding the arid lowland elevations (Thayer, 2005). Despite its widespread distribution, the diversity of *Megarthrus* remains relatively poorly described. Of the approximately 230 current proteinine species, 173 belong to *Megarthrus* (Cai et al., 2016), making it the most diverse proteinine group by far. Generally, the genus occurs in leaf litter, logs, decaying organic material and fungi. In early studies of the genus, it was believed to be a mostly temperate group, depauperate in the tropics, with Bernhauer and Paulian (1942) noting that the genus was poorly represented in Africa south of Sahara. However, with more recent revision, the genus is now represented by over 40 species in Africa south of Sahara, with much of the region still unexplored (Cuccodoro & Löbl, 1995).

Not much is known regarding the ecology of the genus. A curious behavior of water loading was observed and described by Cuccodoro (1995) where in a laboratory setting in the evening

or after being cooled to 4° C, beetles actively collect condensed water from the substrate and transfer it on to the dorsal side of the head and thorax. There it will appear as a drop that may even exceed the volume of the beetle body. Among Afrotropical species, observed species all have fully developed wings, and are found in the same variety of habitats as the group worldwide. The water loading feature has also been observed in Afrotropical species (Cuccodoro & Löbl, 1995). Morphologically, the genus is distinctly characterized by a lack of vertexal ocellus, a deep longitudinal medial impression on the pronotum, as well as the uneven and often toothed lateral edges of the pronotum.

Cladistic analyses of the genus from the 90s retrieve *Megarthrus* combined with the genus *Metopsia* Wollaston, 1854 as a single monophyletic group, with *Proteinus* Latreille, 1796 as its sister group. The three genera together make up the tribe Proteinini. Cladistics has also been performed within *Megarthrus* for certain biogeographical groupings of species (Cuccodoro, 1998). However, no molecular phylogeny has so far been constructed for the genus.

From Cameroon, two species of *Megarthrus* have been recorded to date. These are *M. wittei* Cameron, 1950, originally described from Zaire (today the Democratic Republic of the Congo); the only African species with a reported distribution spanning the continent from east to west (Cuccodoro & Löbl, 1995), as well as *M. kamerunensis* Bernhauer, 1942, described from Mt. Cameroon without illustrations, making it difficult to compare this species to the other similar species subsequently described from other parts of Africa. The only type specimen of *M. kamerunensis* was apparently lost in transit in 1969 while being returned to the Field Museum following the death of W. O. Steel (Cuccodoro & Löbl, 1995). As *M. wittei* has not been recorded on the mountain, only one species, with type material lost, has been described from Mt. Cameroon. During a field trip to Mt. Cameroon early 2023 *M. kamerunensis* was collected again, along with a new, previously undescribed species of *Megarthrus*.

1.6   Aim of the study

In this study, the goal is to investigate the Staphylinid diversity of different Afroalpine mountains, with a special focus on Mt. Cameroon and the genus *Megarthrus*. First, the Staphylinid diversity of Mt. Cameroon and the Eastern Arc Pare Mountains along the altitudinal gradient is measured by sampling data and compared under the Hill framework of α- and β-diversity. This is done in order to explore the overall patterns and differences in diversity and composition between communities in mountains of different age, size, elevation and other properties. Second, as the Staphylinid family of beetles is largely unexplored in most of the Afrotropics, thorough phylogenetic analyses using modern molecular methods are missing for most groups. Here, the phylogenetic relationships in the genus *Megarthrus* are untangled, using representative species from all around the world deposited in the Natural History Museum of Oslo, including newly collected specimens from Mt. Cameroon. Finally, two species of *Megarthrus* were collected from Mt. Cameroon. Here, *M. kamerunensis* is redescribed using new material, and the second species is described as a new species.

The thesis is written as a part of the ANTENNA-project: African Network for Training a New Generation of Entomologist in DNA-based molecular methods. The project is funded by the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (DIKU). Through the project, field trips to Mt. Cameroon were organized in July 2022 and February/March 2023, from which material was collected.

2 Material and Methods

2.1 Study area

2.1.1 Mt. Cameroon National Park

Specimens were collected from Mt. Cameroon during a field trip in early 2023 (Figure 1). Samples were collected by sifting leaf litter on three different slopes with an altitudinal interval of approximately 100 m, within a radius of 10–60 m each (Appendix A: Table 6). Fifty-five samples were collected, of which 27 were sorted and used in downstream analyses. The southern slope (above Bakingili) was studied most extensively, with samples collected in the altitudinal range of 300–2900 m. The south-east slope (above Bokwango) had samples collected between 1300–2000 m, and the northern slope (above Bova) was sampled between 1500–2600 m.
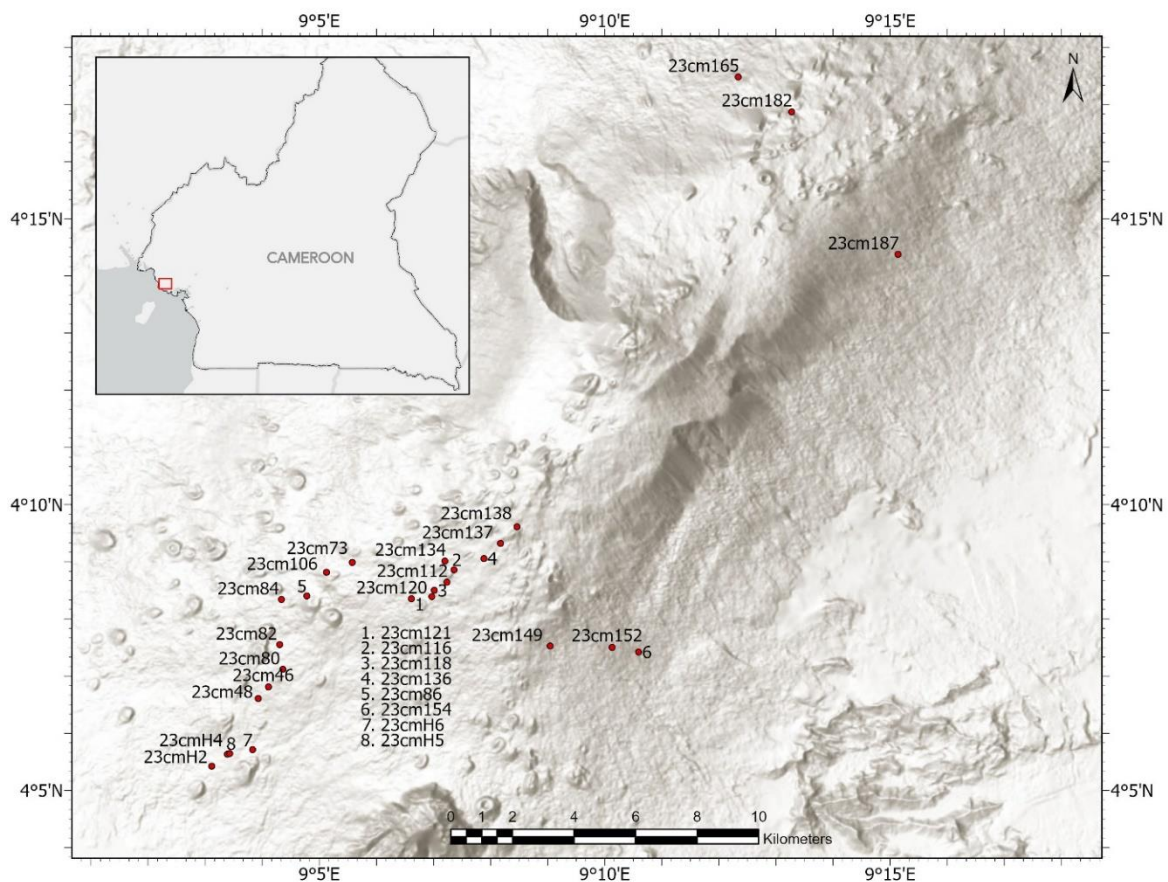
Figure 1. Sampling localities of 27 samples collected from Mt. Cameroon National Park. Samples were collected in the altitudinal range between 300–2900 m. The figure was created using ArcGIS Pro 3.1.0.

Specimens were sorted to morphospecies at the Natural History Museum in Oslo. Twenty-seven of the samples were sorted to 263 morphospecies representing 12 subfamilies. The data were recorded and processed into an abundance-based species-by-sites matrix for further analyses.

2.1.2 The Eastern Arc Pare Mountains

From East Africa, two datasets of staphylinids collected from the Pare Mountains of Tanzania in 2010 were examined and compared to the Mt. Cameroon dataset. A total of 30 samples were collected by leaf litter sifting at nine different slopes. Five slopes were sampled in the North Pare range, and four slopes were sampled in the South Pare range (Figure 2). The data were not collected based on examination of the altitudinal gradient yet cover altitudes in the range of 1300–2100 m (Appendix A: Table 7).

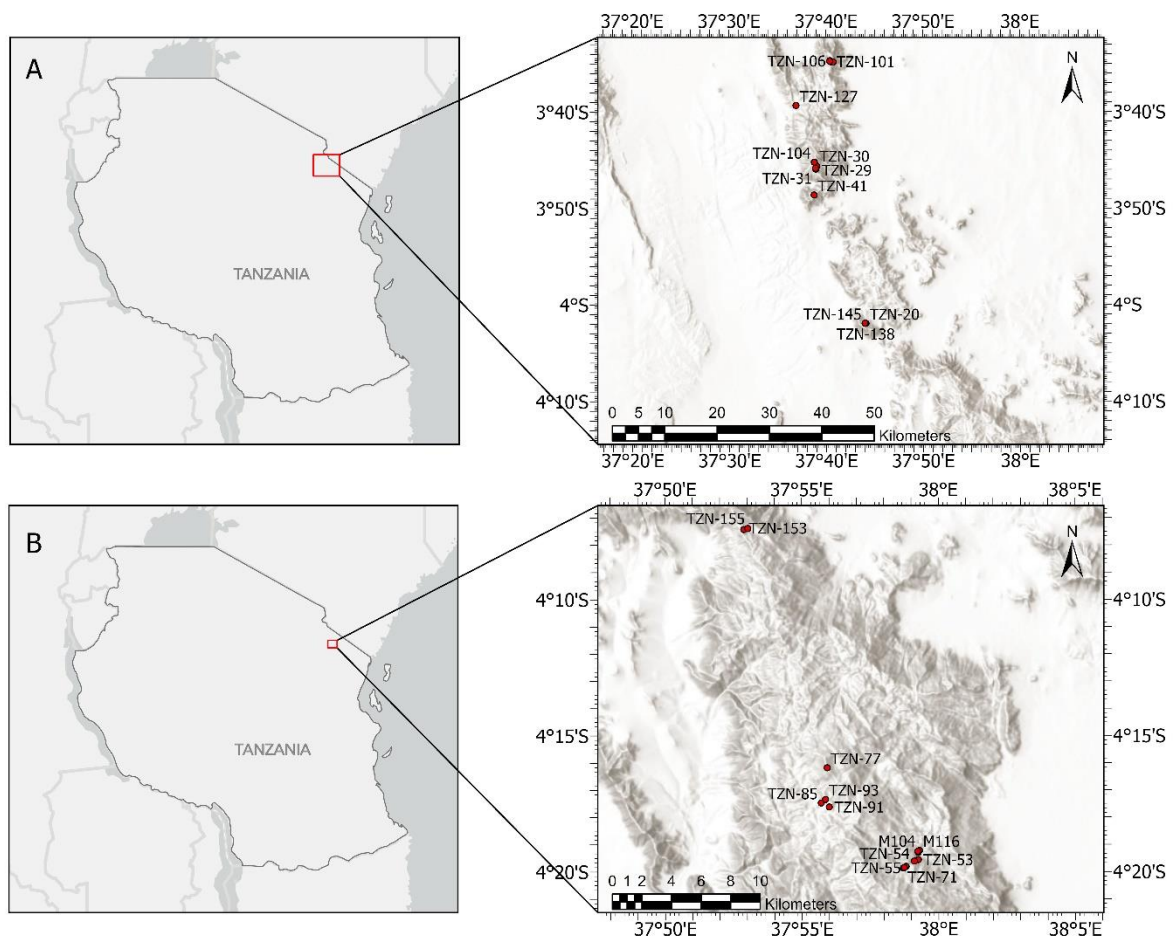**Figure 2.** Sampling localities of 30 samples collected from (**A**) North Pare and (**B**) South Pare. Samples were collected in the altitudinal range between 1300–2100 m. The figure was created using ArcGIS Pro 3.1.0. Not all sample IDs are shown.

From the Pare datasets, 184 morphospecies were sorted, representing 10 subfamilies. The data were processed into an abundance-based species-by-sites matrix for further analyses.

2.2 Diversity analysis

2.2.1 Taxonomic Hill α-diversity

Taxonomic Hill diversities for each sample were calculated in R Statistical Software v4.3.1 (R Core Team, 2023) using the *iNEXT* package (Chao et al., 2014; Hsieh et al., 2022), and the *HillR* package (Li, 2018). Diversities for diversity orders q = 0, 1, 2 were calculated from both absolute observed abundance data and asymptotic estimators for each sample. Plots of effective number of species against elevation were created using the *tidyverse* R-package (Wickham et al., 2019), and arranged using the *ggpubr* package (Kassambara, 2023).

2.2.2 Multivariate analyses

The abundance data were examined by multivariate ordination methods. Using the R package *vegan* (Oksanen et al., 2022), species abundance matrices were analyzed by non-metric multidimensional scaling (NMDS) and redundancy analysis (RDA) for examining grouping of species and sites, and the multivariate effect of elevation and diversity on the data.

RDA-plots were created using the *rda* function, with species scaled to unit variance. Here, the data were standardized by the Hellinger transformation, and elevation, species richness, Shannon diversity and Simpson diversity were included in the global models. Each global model used in the RDA-analyses was tested for significance by PERMANOVA using the vegan function *anova.cca*. Forward selection was used to assess the significance of each explanatory variable in the model, and to select which variables to include in the improved models in order to avoid overfitting and collinearity. Selection was run with a cutoff alpha level of 0.01 and with 1000 permutations. Explanatory variable collinearity was tested to confirm adequately low levels of collinearity in the improved models. Finally, the terms in the improved models were tested for significance by PERMANOVA.

Prior to NMDS-analyses, the data matrices were standardized by dividing by the margin total. Ordination plots were created using the *metaMDS* function in vegan. The commonly used Bray-Curtis distance measure was chosen for the ordination. Tests with other distance

measures were conducted, without any significant improvement in stress values. All significant explanatory variables were included in the analyses.

2.2.3 Taxonomic Hill β-diversity and compositional change

Regional effective numbers of assemblages were created, standardized by coverage and compared for each mountain using the *iNEXT.beta3D* package (Chao & Hu, 2023; Chao et al., 2023) in R version 4.3.1 (R Core Team, 2023). Beta diversities were subsequently monotonically transformed and standardized by coverage into four classes of Sørensen- and Jaccard-type dissimilarity measures for each diversity order q for comparison.

Estimated local pairwise similarities between assemblages on each mountain for relative abundance data were calculated in the online SpadeR shiny-app (Chao et al., 2015), and exported to R. Pairwise similarities were calculated using the Sørensen-type (CqN) local species overlap measures for q = 0 and q = 1. Since relative abundance data were used, the measures reduce to the traditional N-community Sørensen (1948) and Horn (1966) measures respectively. Pairwise similarities were also calculated using the Jaccard-type (UqN) regional species overlap measure for q = 0, and the Morisita-Horn and regional overlap index for q = 2 for sensitivity analysis; however, the results were similar, and these analyses have therefore been omitted.

Absolute pairwise distances between assemblages measured in kilometers were calculated in R using the *fossil* package (Vavrek, 2011). Similarly, pairwise altitudinal differences between assemblages were calculated. Scatterplots of similarity measures against absolute distances and altitudinal differences grouped by within- and between-slope assemblage-pairs were created using the *tidyverse* package and arranged using the *ggpubr* package.

All scripts created and used for the diversity analyses are presented in Appendix B.

2.3   Molecular work on the genus *Megarthrus*

2.3.1   Taxon selection and sample preparation

For initial species identification and selection, 86 specimens of the genus *Megarthrus* were selected from the NHM insect DNA grade collection for sequencing. The specimens were previously collected at different localities distributed globally, and were kept in 96% ethanol at -30 – -80° C. From each specimen, the abdomen was separated from the rest of the beetle body by Vladimir Gusarov (NHM) to be used for DNA extraction. The remaining body parts were returned to the DNA grade collection. After extraction, the abdomens were dissected to extract the genitalia, and, together, all the body parts were preserved as vouchers in genitalia microtubes with a drop of glycerol.

2.3.2   Extraction and *COI* cycle-sequencing

Total genomic DNA (gDNA) was extracted from each of the prepared specimens using the QIAGEN QIAamp DNA Micro Kit, following the manufacturer's protocol for animal tissue. The specimens were dried to evaporate all the ethanol and processed without crushing. The lysis time was extended to 24 h. 60 µl Buffer AE was used to elute the DNA. Before centrifuging, the QIAamp MinElute column loaded with Buffer AE was incubated for 10´ to increase DNA yield. The elution step was repeated to create a backup stock of DNA for storage in the NHM DNA-bank.

To select one specimen per species for phylogenetic analyses, partial sequences of the mitochondrial gene *Cytochrome Oxidase Subunit I* (*COI*) were obtained for every specimen. Polymerase Chain Reaction (PCR) was set up as 14 µl reactions with 1.2 µl BSA diluted to 1 mg/ml, 0.125 µl DreamTaq™ Hot Start DNA Polymerase (Thermofisher Scientific), 7.8 µl dsH$_2$O, 2.5 µl dNTPs and 1.2 µl of each primer per reaction. The primer pair used was LCO1490/C1-2416ra (Table 1). *COI* was amplified with a denaturation step running at 95° for 2´, followed by 40 cycles of denaturation at 95° for 30´´, annealing at 49° for 30´´, and extension at 72° for 90´´. Finally, the samples were left to extension at 72° for 10´.

Table 1. Primers used for initial *COI* amplification.

Primer name	Position[a]	Primer site	Sequence (5´-3´)	Reference
LCO1490	1514	*COI*	GGT CAA CAA ATC ATA AAG ATA TTG G	Folmer et al. (1994)
C1-2416ra	2416	*COI*	GCT AAT CAT CTA AAA ATT TTA ATT CC	Modification of C1-J-2441; Simon et al. (1994)

[a] The positions of the *COI* primers are related to the mitochondrial genome of *Drosophila yacuba* (Clary & Wolstenholme, 1985).

Presence of the targeted PCR product was verified using 1% agarose gel-electrophoresis, with the Biotium GelRed nucleic acid gel dye. Samples were compared to the Fast Ruler LR ladder (Thermofisher Scientific) for all gel-electrophoresis analyses. A total of 79 samples with verified DNA concentrations were subsequently purified in strips using Illustra™ ExoStar™ 1-Step (GE Healthcare Life Sciences). The ExoStar was diluted 10 X with MilliQ® dsH$_2$O. Each well on the strips contained an aliquot of 10 µl of PCR product, and 4 µl of diluted ExoStar. The strips were then incubated in the thermal cycler at 37° for 45´, followed by 15´ at 80°.

Purified samples were prepared for sequencing in both directions by mixing dsH$_2$O and primers (one at a time) with an aliquot of the purified DNA. A total of 158 samples (79 specimens both forward and reverse) were sequenced by StarSEQ GmbH (Mainz, Germany).

2.3.3 *COI* assembly and alignment

Raw *COI* sequences were assembled in contigs using CodonCode Aligner (v4.2.7, CodonCode Corporation). In total, 71 contigs were obtained, whereas eight sequence pairs were of too low quality and/or insufficient length to be assembled. Consensus sequences of all contigs were manually checked and edited by trimming low quality ends of the sequences, and by manual basecalling at artificially ambiguous sites. Edited contigs and unassembled sequences were exported together in fasta format for alignment in MEGA11 (Tamura et al., 2021). The MUSCLE algorithm was used to align all sequences together with publicly available *COI* sequences of *Megarthrus* retrieved from BOLD systems (Edgar, 2004;

Ratnasingham & Hebert, 2007). After alignment, primers and low quality end-reads were removed from the alignment. A neighbor-joining tree was created using all sequences to assess groupings and exclude duplicate specimens from the subsequent lab work.

2.3.4 Multiplex PCR

Twenty-four nuclear markers (Table 2) were selected for amplification using multiplex PCR and sequencing for further use in phylogenetic analyses. For this purpose, 48 new primer pairs were designed by Vladimir Gusarov based on Che et al. (2017) for running nested PCRs (Appendix D: Table 27). Each primer pair was tested on specimens of *Megarthrus* and *Bledius* through nested PCR with annealing temperature of 55° in both runs (Appendix D: Table 28). The results were visualized with 1% gel-electrophoresis using GelRed.

Subsequently, primers were run through nested PCR again for optimization of annealing temperatures. One set of primers yielding no concentration, one yielding low concentration and one yielding high concentration in the original test were tested on one specimen on a gradient of annealing temperatures ranging from 52° to 65°, with initial five cycles of lower annealing temperature (Appendix D: Tables 29–30), and visualized with 1% gel-electrophoresis using GelRed.

Of the original 86 specimens, 38 specimens from as many assumed species as possible were selected for multiplex PCR, as well as eight outgroup specimens and the two newly collected species of *Megarthrus* from Mt. Cameroon (Table 3). The outgroup specimens consisted of four species of *Metopsia*, one *Alloproteinus*, one *Austrorhysus,* one *Proteinus*, and one *Oxytelopsis* of the Oxytelinae subfamily. Many of the Afrotropical specimens are likely undescribed species, and were thus hard to identify. In Table 3, these specimens are numbered rather than named. Similarly, many species were only represented by females, limiting confident identification altogether. These specimens have not been labeled by number in Table 3. Instead, they will be referred to by their unique IDs downstream.

The 24 markers were arranged in four groups of six genes, based on band intensity in the initial test of the primers. Low intensity band genes were grouped together, and high intensity band genes were grouped together (Table 2).

Table 2. List of markers selected for amplification. The markers were grouped based on band intensity from the test PCR: 0 = no target band, 1 = weak intensity target band, 2 = strong intensity target band. Markers in bold performed well during multiplex PCR and were included in the dataset for Maximum Likelihood analyses.

Primer Group	Marker	Test PCR Performance	PCR1 Product Length (bp)	PCR2 Product Length (bp)
1	*BOP1*	0	1135	742
	Ace	0	1180	790
	Lar	**0**	**1003**	**883**
	Cap60A	0	1492	1021
	I(3)72Ab	0	1309	1138
	Hem	0	1477	1225
2	*Spt6*	1	1123	961
	Sec24	0	1081	877
	Nhel	**0**	**1189**	**1039**
	Rbcn3A	0	1522	1048
	NRX-IV	0	1624	1141
	NonC	1	1519	1388
3	*FBX011*	2	922	751
	Rols	**1**	**1060**	**826**
	Emb	**2**	**1261**	**979**
	Dhc98D	**2**	**1186**	**1012**
	CAD	**2**	**1648**	**1171**
	Brat	**2**	**1694**	**1175**
4	***Shal***	**2**	**794**	**727**
	Zip	**2**	**1297**	**841**
	RPII215	**2**	**1336**	**994**
	Trpml	2	1252	1084
	Ico	2	1348	1195
	Sxc	2	1459	1282

Table 3. List of all specimens included in the multiplex PCR. Abbreviations: HK = Håkon Knudsen, VIG = Vladimir I. Gusarov, HE = Hallvard Elven, VVG = Vasily V. Grebennikov, NY = Nikolai Yunakov, EBJ = Eilen B. Josefsen. *Oxytelopsis sp.* marked in bold was removed from downstream analyses due to poor read mapping to the reference sequences. Specimens without known collector are marked by *.

Species name	ID #	Collected by/ Identified by	Location
Megarthrus bantu	3279	HE/HK	Uganda
Megarthrus sp. 7	3823	VIG/VIG	Kenya
Megarthrus sp. cf. simiensis	4013	VIG/VIG	Ethiopia
Megarthrus sp. 4 prope rougemonti	4176	VIG/VIG	Ethiopia
Megarthrus sp. (female)	13056	*/VIG	Ethiopia
Megarthrus sp. 10 prope niloticus	13135	*/VIG	Ethiopia
Megarthrus sp. (female)	13229	VIG/VIG	Ethiopia
Megarthrus sp. (female)	13297	*/VIG	Ethiopia
Megarthrus denticollis	13566	VIG/HK	Poland
Megarthrus sp. (female)	13781	VVG/VIG	Tanzania
Megarthrus sp. 11 prope africanus	16233	*/VIG	Tanzania
Megarthrus sp. 2 prope africanus	23537	VVG/VIG	Tanzania
Megarthrus falasha	25091	*/HK	Ethiopia
Megarthrus sp. 8 prope twa	25522	*/VIG	Kenya
Megarthrus angulicollis	33467	*/HK	Canada
Megarthrus pecki	33600	*/VIG	Canada
Megarthrus longicornis	33998	*/HK	Madeira
Megarthrus americanus	36153	NY/HK	USA
Megarthrus excisus	36725	*/HK	Canada
Megarthrus sp. V1	37467	*/VIG	Vietnam
Oxytelopsis sp.	37854	*/VIG	Vietnam
Megarthrus sp. V2	37877	*/VIG	Vietnam
Austrorhysus sp.	41869	VIG/VIG	Australia

Megarthrus nitidulus	42115	*/HK	Russia
Megarthrus sp. n. prope depressus	43180	VIG/VIG	Russia
Megarthrus depressus	43677	VIG/HK	Norway
Megarthrus wollastoni	52249	VIG/HK	Canary Islands
Megarthrus bellevoyei	58473	VIG/HK	UK
Alloproteinus nigriceps	64567	VIG/VIG	Argentina
Megarthrus horticola	71893	VIG/HK	Uganda
Megarthrus sp. (female)	71895	VIG/VIG	Ethiopia
Megarthrus horticola	71896	NY/HK	Kenya
Megarthrus sp. 5 prope watutsi	71898	NY/VIG	Kenya
Megarthrus sp. 6	71899	NY/VIG	Kenya
Megarthrus sp. 3 prope africanus	71900	VVG/VIG	Tanzania
Megarthrus sp. 9 prope simiensis	71901	*/VIG	Ethiopia
Megarthrus hemipterus	71917	VIG/HK	Russia
Megarthrus sp. 2 prope africanus	71920	VVG/VIG	Tanzania
Proteinus brachypterus	71923	EBJ/HK	Norway
Metopsia cimicoides	71925	*/HK	Canary Islands
Metopsis clypeata	71929	*/HK	Portugal
Megarthrus montanus	71932	*/VIG	Russia
Metopsia ampliata	71933	VIG/HK	Madeira
Metopsia palmensis	71936	VIG/HK	Canary Islands
Megarthrus spathuliformis	71938	VIG/HK	Canary Islands
Megarthrus pictus	71940	VIG/HK	Canada
Megarthrus kamerunensis	71952	VIG/HK	Cameroon
Megarthrus sp. 1	71953	VIG/HK	Cameroon

Table 4. Optimized PCR setup for PCR1 of multiplex PCR.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	300	1
Denaturation	95	30	
Annealing	50	90	5
Extension	72	150	
Denaturation	95	30	
Annealing	55	90	25
Extension	72	150	
Final Extension	68	900	1
Infinite Hold	8	∞	1

Table 5. Optimized PCR setup for PCR2 of multiplex PCR.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	300	1
Denaturation	95	30	
Annealing	47.5	90	5
Extension	72	120	
Denaturation	95	30	
Annealing	52.5	90	25
Extension	72	120	
Final Extension	68	900	1
Infinite Hold	8	∞	1

Each reaction in the four groups was set up as 25 µl reactions using the QIAGEN Multiplex PCR Kit following the manufacturer's instructions. 12.5 µl 2 X Multiplex Master Mix, 7 µl RNase-Free Water, 2.5 µl primer mix and 3 µl DNA was added to each reaction, for PCR1. For PCR2, the product from PCR1 was diluted 20 X and 3 µl of the diluted product was used. The PCR setups were as in Tables 4–5. All PCR products were visualized with 1% gel-electrophoresis using GelRed.

PCR2 product concentrations were estimated using the Bio-Rad Image Lab 6.0 software, using the known DNA concentration of the Fast Ruler LR (Thermofisher Scientific) bands as reference. Prior to library-prep, product concentrations were normalized to ~14 ng/µl and normalized products for each sample were pooled (four products per sample) in new strips. Products with too high DNA concentrations were diluted 9 X using MilliQ® dsH$_2$O prior to pooling, and for products with too low concentrations the entire volume of PCR2 product was used. The combined products for each sample were subsequently purified using Illustra^{TM} ExoStar^{TM} 1-Step (GE Healthcare Life Sciences). For samples with total volume <25 µl, ExoStar was diluted 10 X using MilliQ® dsH$_2$O, whereas for samples with total volume ≥25 µl, 1 X ExoStar was used in order to minimize additional dilution of the products. The strips were then incubated in the thermal cycler at 37° for 45´, followed by 15´ at 80°.

The DNA concentration of normalized and purified products was measured using the Invitrogen® Qubit® 2.0 Fluorometer. Six samples were measured using 2 µl DNA (Appendix D: Table 31).

2.3.5 Oxford Nanopore MinION sequencing

The pooled samples were additionally cleaned and size-selected using AMPure XP Beads following the manufacturers protocol (Beckman Coulter Life Sciences). An AMPure XP bead ratio of 0.8 X was used in order to retain as much DNA as possible. The retained DNA concentrations were measured using the Invitrogen® Qubit® 2.0 Fluorometer on five of the same samples measured during the normalization step (Appendix D: Table 32).

Library-prep for Oxford Nanopore MinION Mk1C sequencing was carried out following the manufacturers Native Barcoding Kit 96 with Amplicons (SQK-NBD112.96) protocol (Version NBA_9137_v112_revH_01Dec2021; Oxford Nanopore Technologies, Oxford, United Kingdom), with the following modifications. During barcode ligation, the DNA

22

concentration was doubled in order to ensure a proper final concentration, based on the results from the previous DNA concentration measurements. After pooling all the samples, the pooled reaction was cleaned and size-selected again with an AMPure XP bead ratio of 0.6 X in order to remove short length artifacts from the reaction. 32 µl of eluate was removed and retained prior to adapter ligation and cleanup. 1 µl of the eluted sample was quantified to 3.98 ng/µl using the Invitrogen® Qubit® 2.0 Fluorometer. During the final AMPure XP bead cleanup step, 20 µl of eluate was retained rather than 15 µl by mistake. This error was accounted for during loading. Quantification of the final library yielded 3.26 ng/µl.

3 µl of the final library (~10ng DNA) was loaded onto a R9.4 SpotON flow cell with 1520 functioning pores, and sequenced on a MinION Mk1C (Oxford Nanopore Technologies, Oxford, United Kingdom). The sequencing was left running for ~21 h before being terminated. Due to running fast basecalling by the sequencer to save time, high quality basecalling was carried out separately using Guppy v6.5.7 (Oxford Nanopore Technologies, Oxford, United Kingdom). Reads were sorted by barcode and saved in individual barcode folders.

2.3.6 Bioinformatics

High quality basecalled reads for all barcodes were mapped to reference sequences for all markers included in the analysis using Geneious Prime 2022.1.1 (https://www.geneious.com). At the time of the analysis, no Proteininae genomes or reference sequences of the included genes were available. For this reason, the reads were first mapped to available reference sequences of the Aleocharine species *Aleochara bilineata* using the built in Geneious Prime mapper. In order to increase mapping success, a new *Megarthrus* reference sequence was constructed for each marker using consensus sequences from different barcodes made during the first mapping for each marker. These new *Megarthrus* reference sequences were then utilized for the second iteration of mapping for all the barcodes. Only reads of length >700 bp were used when mapping to a reference sequence. Each mapping profile was examined visually and if >10 complete reads mapped to a reference sequence a consensus sequence was deemed to be of acceptable quality. The markers for which most of the barcodes produced a consensus sequence (as described above) were used in downstream analyses.

Consensus sequences were extracted and aligned using the MUSCLE algorithm (Edgar, 2004). Indels that were not in multiples of three bp, and artificial Ns and gaps due to poor

mapping were manually checked and corrected or removed. After completion of the project, the sequences will be deposited in GenBank. The final alignments were subsequently concatenated in Geneious Prime.

Using Maximum Likelihood in IQ-TREE v. 1.6.12 for windows (Nguyen et al., 2015), three phylogenetic trees were constructed with the concatenated dataset applying different partitioning schemes: one based on unpartitioned data, one based on data partitioned by gene, and one based on partitioning by both gene and codon position (Appendix E: Table 34). The best partitioning schemes and model fits were automatically determined in IQ-TREE by the incorporated Model Finder Plus (MFP) option (Kalyaanamoorthy et al., 2017), and merged using the MERGE option. For each tree, branch support was calculated using Ultra-Fast Bootstrap (Hoang et al., 2018) with 1000 replicates as well as by the SH-aLRT test with 1000 replicates (Guindon et al., 2010).

The resulting single most likely unrooted ML trees were subsequently visualized in FigTree v. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and edited in Adobe Illustrator CS6 Version 16.0.0. As Silphotelini are a sister group to all the remaining Proteininae (Newton & Thayer, 1995), the trees were rooted between the Silphotelini species *Alloproteinus nigriceps* and the rest of the specimens.

2.4 Species description material and methods

All the material examined for the species descriptions is deposited at the Natural History Museum in Oslo, Norway (NHM). Specimens were cleared in KOH 10% solution. Specimen abdomens were dissected and prepared in glycerin for detailed illustrations and examinations. Habitus images were taken using a Nikon Z 7 with a Nikon 200mm f/4 ED-IF AF Micro-NIKKOR lens and a Raynox MSN-202 macro lens for magnification. Morphological structures were photographed using a ZEISS Axio Imager 2 microscope with the Axiocam 506 color camera, and the Zen 3.1 (blue edition) imaging software. Habitus images were stacked using the Zerene Stacker (Build T2023-06-11-1120). Images of morphological characters were stacked and edited using Adobe Photoshop CS6 Version 13.0, and all photos were finalized using Adobe Illustrator CS6 Version 16.0.0. Abdominal sternites and tergites are counted by morphology, and expressed by roman numerals, i.e., the last visible tergite = tergite VIII.

The species descriptions are to be considered a first draft for a subsequent formal publication. Taxonomic acts, including species descriptions, presented in a Master thesis as implemented at the University of Oslo, would not fulfill all the requirements of Article 8 of the International Code of Zoological Nomenclature (ICZN, 1999). For this reason, no formal name is given to the new species from Mt. Cameroon. Instead, it will be referred to as *Megarthrus sp. 1* throughout the rest of the thesis.

3   Results

3.1   α-diversity comparison

Hill taxonomic alpha diversities of diversity orders q = 0, 1, 2 for samples along the altitudinal gradients in the different mountains are shown in Figures 3–5. Here, samples from the unique slopes within each mountain have been pooled. The most extensively sampled mountain, Mt. Cameroon, show a clear trend of decreasing diversity with elevation for all diversity orders (Figure 3). For all but the highest elevation samples, diversity decreases with increased diversity order q, signaling a high amount of heterogeneity within communities. In the middle elevations between ~1100–1800 m, the change in diversity fluctuates more and the decrease is less pronounced.

Figure 3. Observed (**A, B, C**) and estimated (**D, E, F**) diversities for 27 samples from Mt. Cameroon against elevation. Diversities are of diversity order q = 0 (species richness), q = 1 (the exponential of Shannon diversity) and q = 2 (the inverse of Simpson diversity).

A different pattern applies to the less extensively sampled Pare Mountains, where only middle elevation altitudes have been sampled, and diversity change is less predictable given the results (Figures 4–5). Here, no general correlation between altitude and diversity have been established. Generally, as the diversity order increases, diversity decreases in these mountains as well, but only slightly, displaying a lower level of heterogeneity in the assemblages.

Figure 4. Observed (**A, B, C**) and estimated (**D, E, F**) diversities for 16 samples from North Pare against elevation. Diversities are of diversity order q = 0 (species richness), q = 1 (the exponential of Shannon diversity) and q = 2 (the inverse of Simpson diversity).
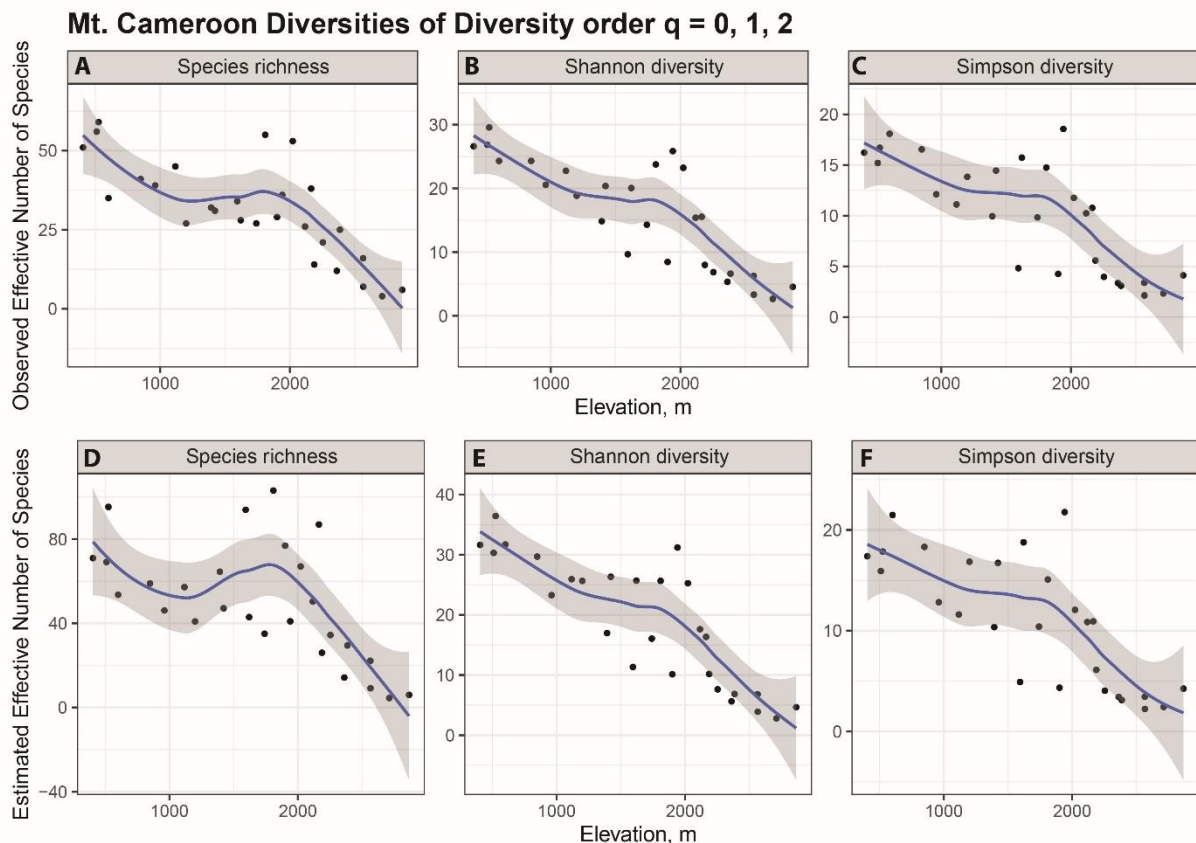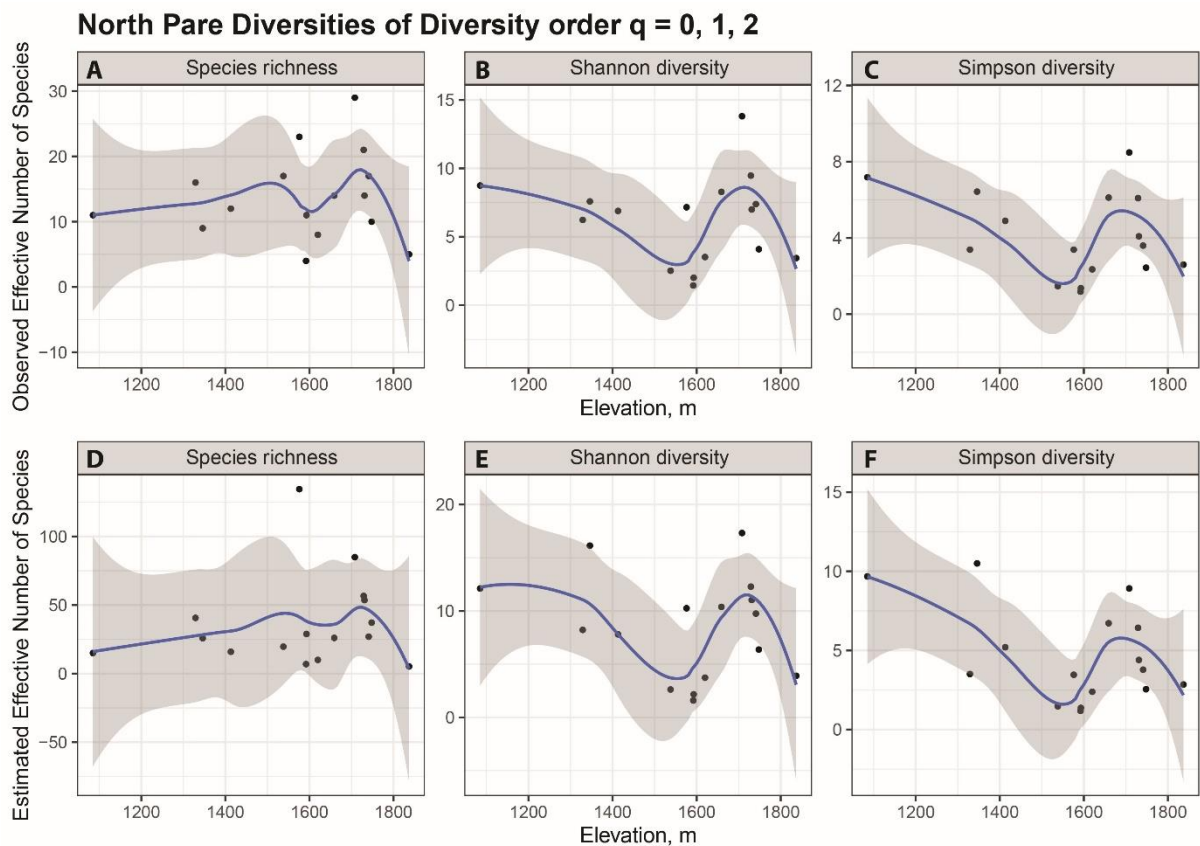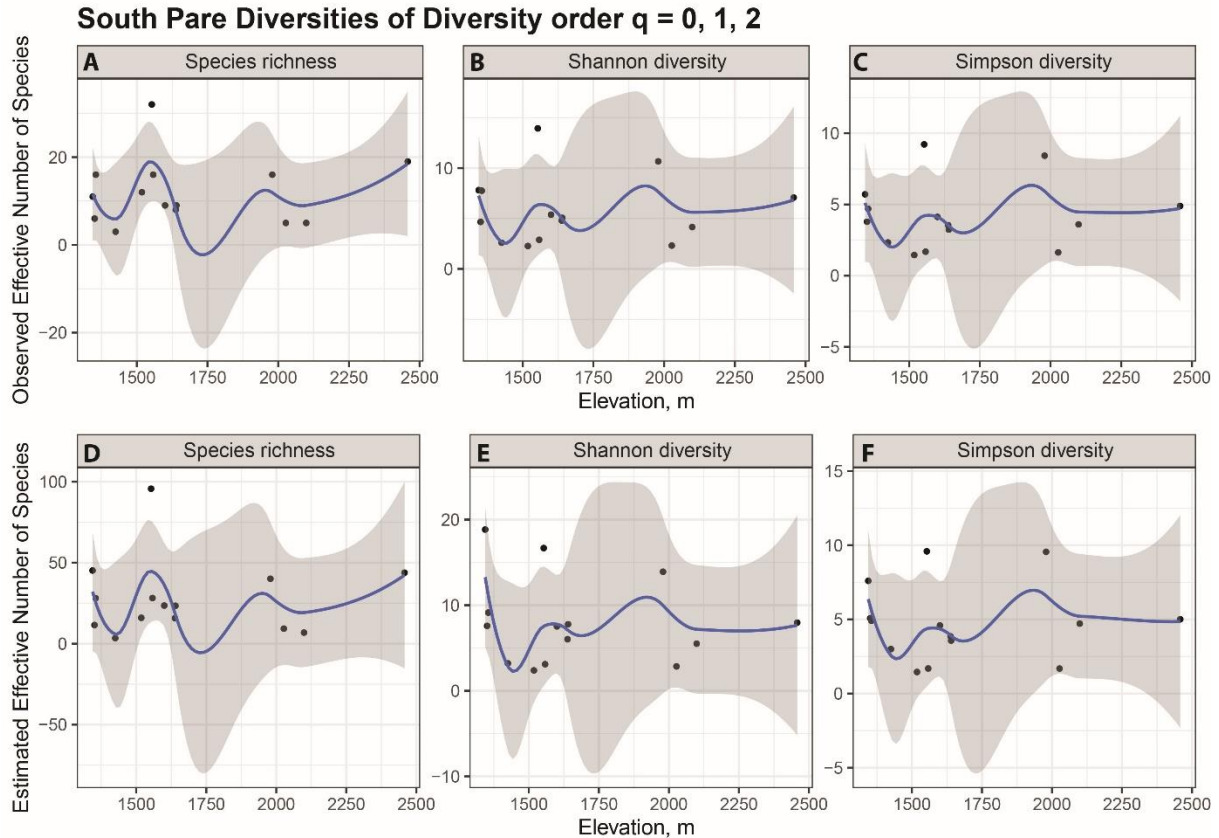
Figure 5. Observed (**A, B, C**) and estimated (**D, E, F**) diversities for 14 samples from South Pare against elevation. Diversities are of diversity order q = 0 (species richness), q = 1 (the exponential of Shannon diversity) and q = 2 (the inverse of Simpson diversity).

3.2   Redundancy Analysis & NMDS

3.2.1 Mt. Cameroon

The global model used for the Mt. Cameroon RDA was significant with pseudo-$F(4, 22)$ = 2.03, p = 0.001, and adjusted R^2 = 13.70%. Not all explanatory variables were significant, and collinearity between the diversity variables was high (sqrtVIF > 2). Forward selection retained elevation and Shannon diversity as statistically important, with a cumulative adjusted R^2 = 12.12%. The adjusted model containing only the two explanatory variables showed low levels of collinearity (sqrtVIF < 2) and was significant with pseudo-$F(2, 24)$ = 2.79, p = 0.001. Figure 6 shows the RDA triplot with the adjusted model scaled by species. The explanatory variables are negatively correlated along the first RDA axis. Distances between sites are mostly short, except for three outlier sites: 121, 106 and 73. The RDA-axes RDA1 and RDA2 explain 12.18% and 6.69% of the variation in the data respectively.

Figure 6. RDA triplot of the Mt. Cameroon dataset using the adjusted model with elevation and Shannon diversity as explanatory variables. The plot is scaled by species. Percentages on the axes are unadjusted R-squared. Sites are named by their unique ID followed by their elevation.

Figure 7. RDA of the Mt. Cameroon dataset showing sites grouped by slope. Sites are named by their unique ID followed by their elevation.

RDA of sites with grouping by slope for Mt. Cameroon is presented in Figure 7. Here, no distinct separation of sites by slope can be discerned, as overlap of sites from different slopes is high during grouping. Sites are generally spread out, and sites from different slopes occasionally are found close together.

Figure 8. NMDS of the Mt. Cameroon dataset showing the species as red plus-symbols, sites as dots, and elevation and diversity as explanatory variables. Sites are named by their unique ID followed by their elevation.

NMDS showing the groupings by Bray-Curtis distances for Mt. Cameroon is presented in Figure 8. The corresponding stress values and Shepard plot are shown in Appendix C: Figures 27–28. $R^2 = 0.98$. Once again, sites group with small distances, except three sites forming their own group. These sites are the high altitude sites 137, 136, and 138. Another high-altitude site, 182, is also positioned separately from most other samples. Additional statistical information for Mt. Cameroon is provided in Appendix C: Tables 9–14.

3.2.2 North Pare

For the North Pare data, the RDA global model was significant with pseudo-$F_{(4, 11)}$ = 2.06, p = 0.002, and adjusted R^2 = 22.00%. Like for Mt. Cameroon, not all explanatory variables in the North Pare data were significant, and collinearity between the diversity variables was high (sqrtVIF > 2). Forward selection retained elevation and Simpson diversity as statistically important, with a cumulative adjusted R^2 = 21.34%. The adjusted model showed low collinearity (sqrtVIF < 2) and was significant with pseudo-$F_{(2, 13)}$ = 3.04, p = 0.001. Figure 9 shows the RDA triplot with the adjusted model scaled by species. Here, sites form distinct groups, apparently variably spread around the explanatory variables. RDA-axes RDA1 and RDA2 explain 24.19% and 7.64% of the variation in the data respectively.
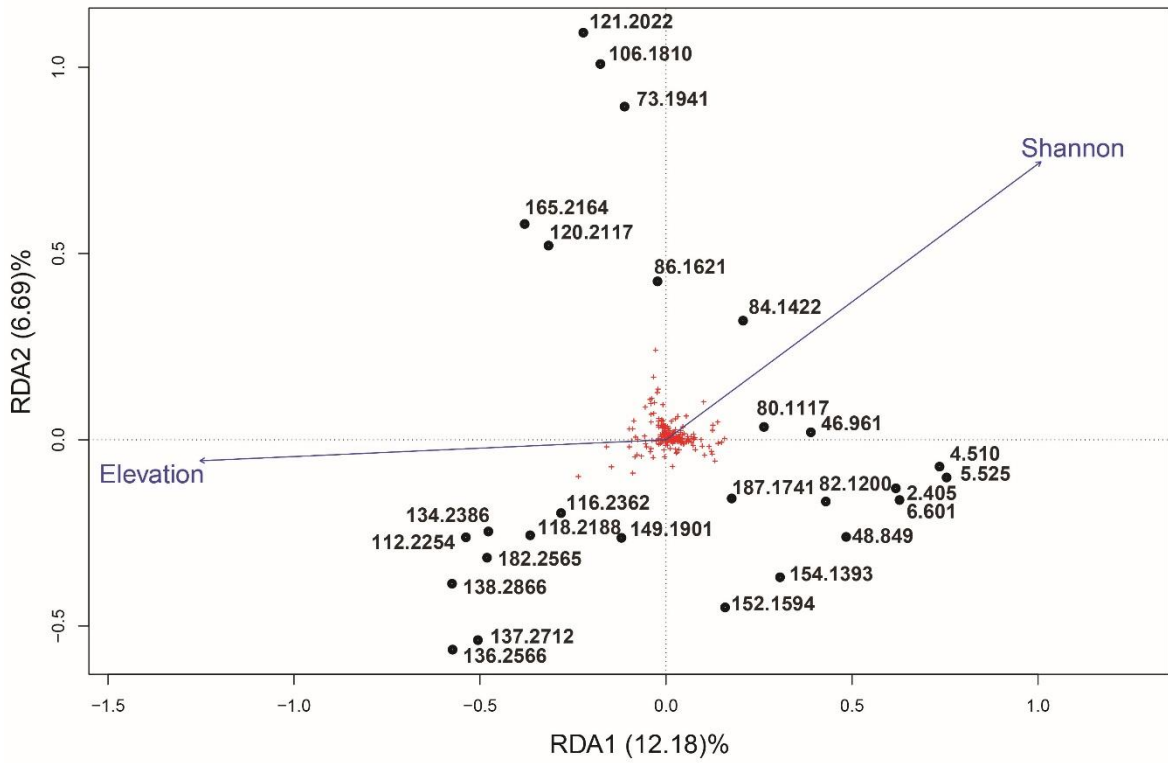

Figure 9. RDA triplot of the North Pare dataset using the adjusted model with elevation and Simpson diversity as explanatory variables. The plot is scaled by species. Percentages on the axes are unadjusted R-squared. Sites are named by their unique ID followed by their elevation.

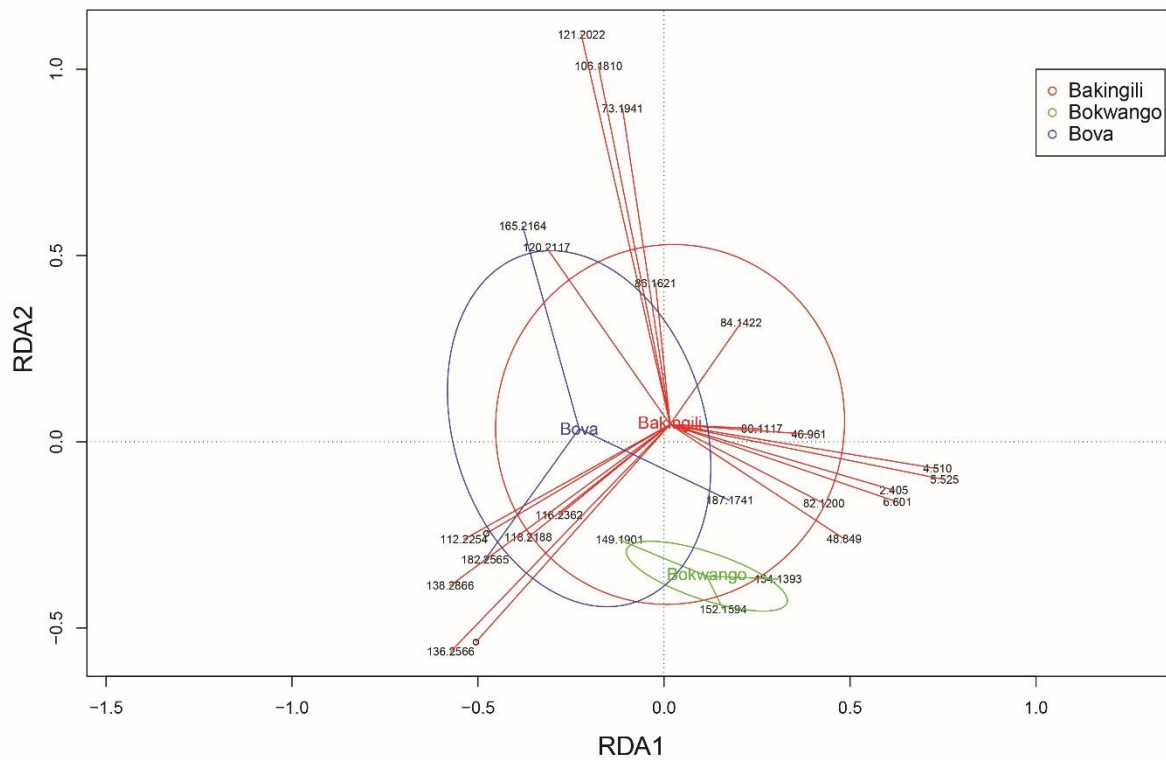**Figure 10.** RDA of the North Pare dataset showing sites grouped by slope. Sites are named by their unique ID followed by their elevation.

The North Pare RDA showing sites with grouping by slope is presented in Figure 10. Sites are clearly grouped by slope, with the Kiverenge sample close to the Kindoroko FR samples, and the Mwanga sample close to the Vumari samples.
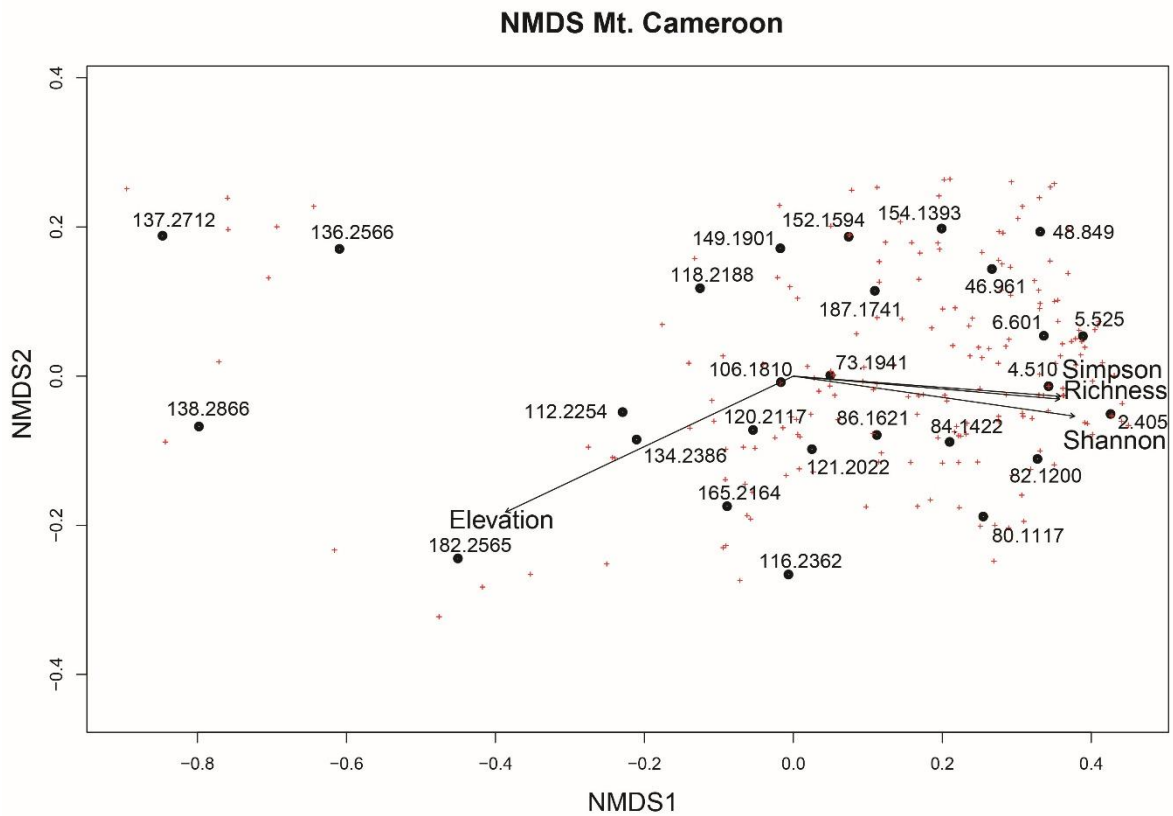
Figure 11. NMDS of the North Pare dataset showing the species as red plus-symbols, sites as dots, and elevation and Simpson diversity as explanatory variables. Sites are named by their unique ID followed by their elevation.

Figure 11 shows the NMDS with grouping by distances. The corresponding stress values and Shepard plot are shown in Appendix C: Figures 29–30. $R^2 = 0.99$. Variation around the explanatory variables is high for the data, with indistinct groupings found from the figure. The explanatory variables are negatively correlated. Additional statistical information is provided in Appendix C: Tables 15–20.

3.2.3 South Pare

Finally, in the South Pare, the RDA global model was significant with pseudo-$F_{(4, 9)} = 1.32$, $p = 0.039$, and adjusted $R2 = 8.91\%$. Again, not all explanatory variables were significant, and collinearity between the diversity variables was high (sqrtVIF > 2). Here, forward selection retained only elevation as statistically important, with a cumulative adjusted $R^2 = 10.51\%$. However, Shannon diversity was also included in the adjusted model, as the ecological importance of including a diversity measure is prioitized. New adjusted R^2 for the model was $R^2 = 10.71\%$. The adjusted model showed low collinearity (sqrtVIF < 2) and was significant with pseudo-$F_{(2, 11)} = 1.78$, $p = 0.003$. Figure 12 shows the RDA triplot with the adjusted model scaled by species. Here, explanatory variables correlate well with the RDA-axes. Sites group close to each other, except for sample 153, which acts as an outlier. RDA-axes RDA1 and RDA2 explain 17.46% and 6.99% of the variation in the data respectively.
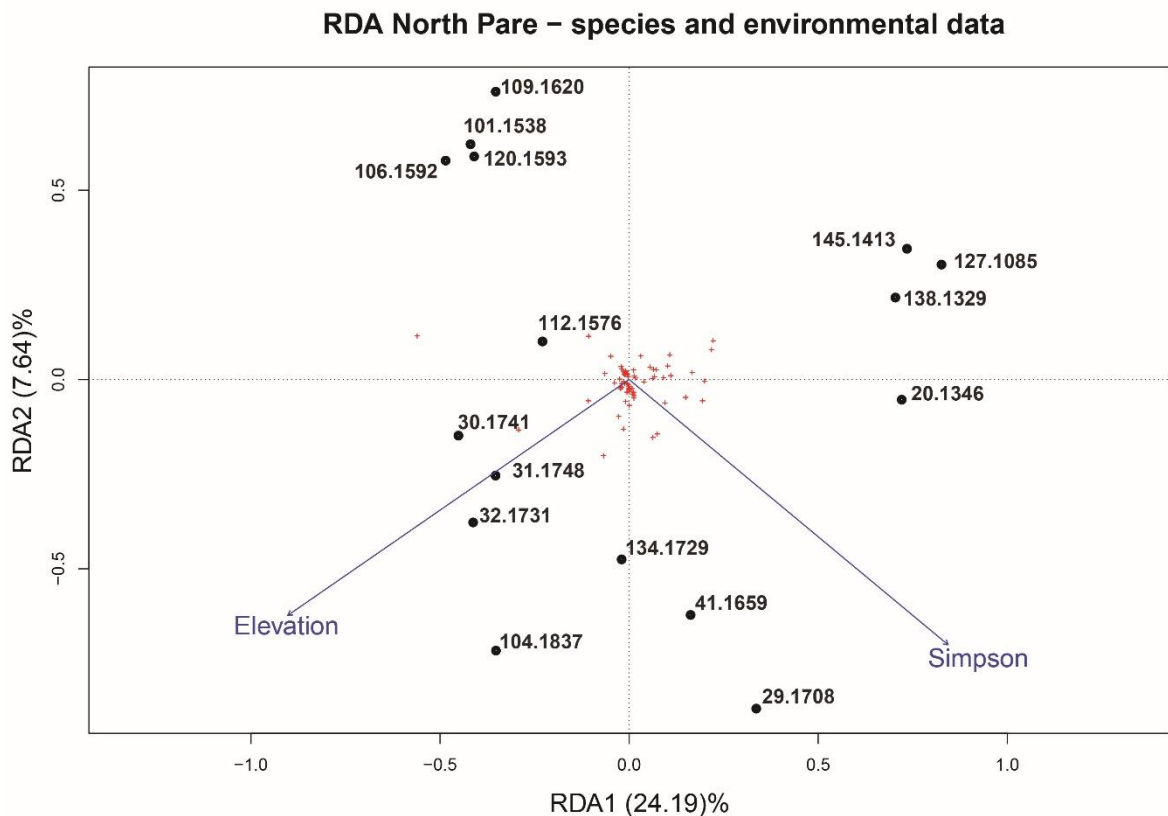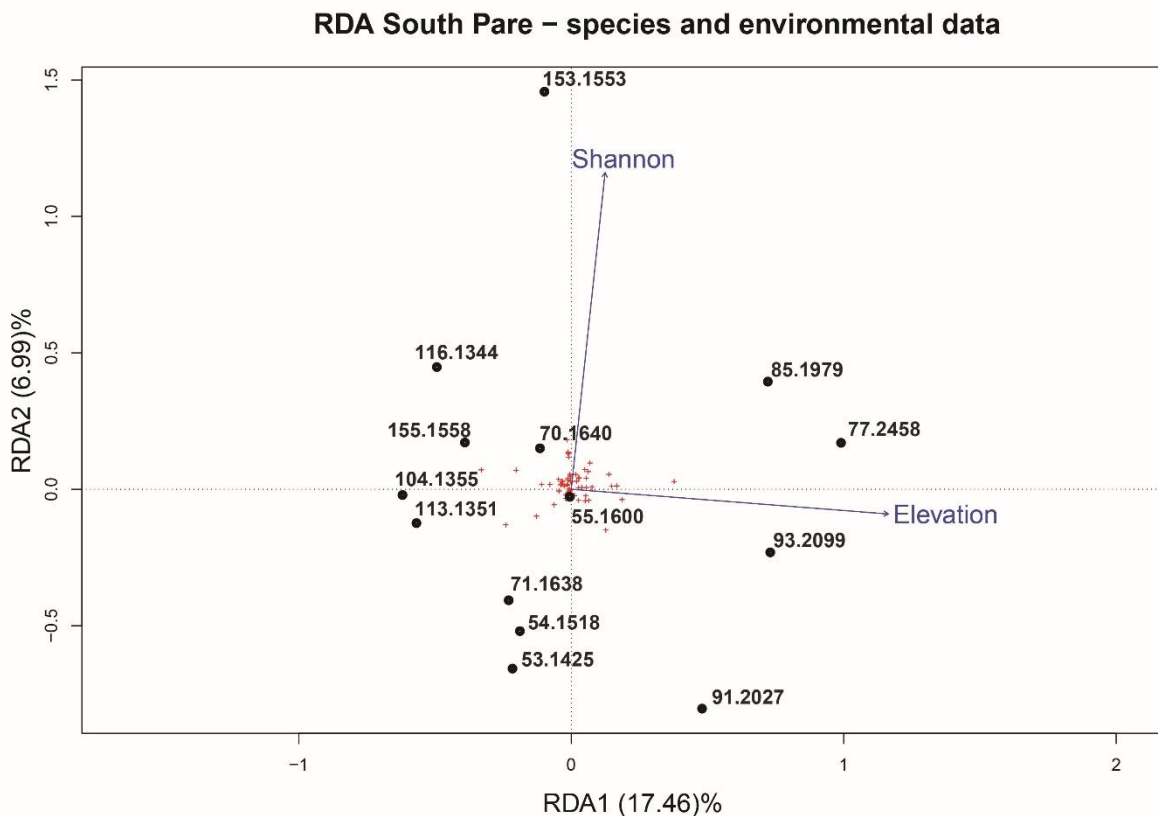

Figure 12. RDA triplot of the South Pare dataset using the adjusted model with elevation and Shannon diversity as explanatory variables. The plot is scaled by species. Percentages on the axes are unadjusted R-squared. Sites are named by their unique ID followed by their elevation.
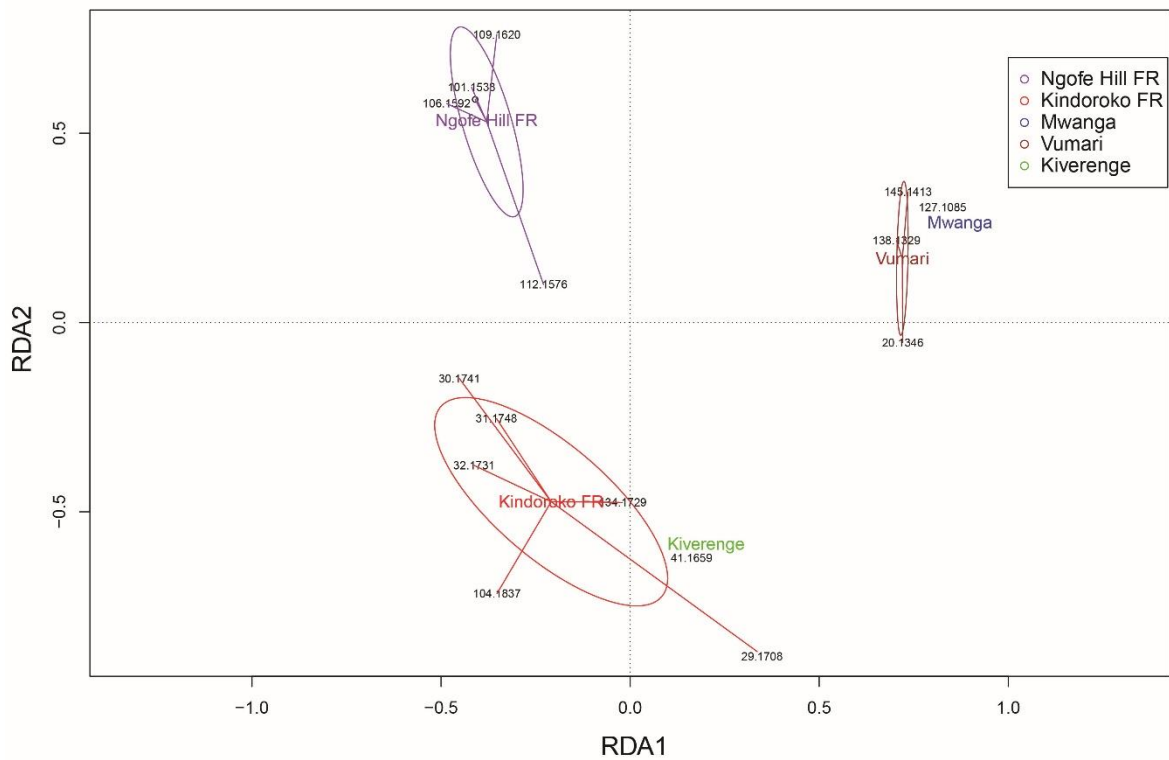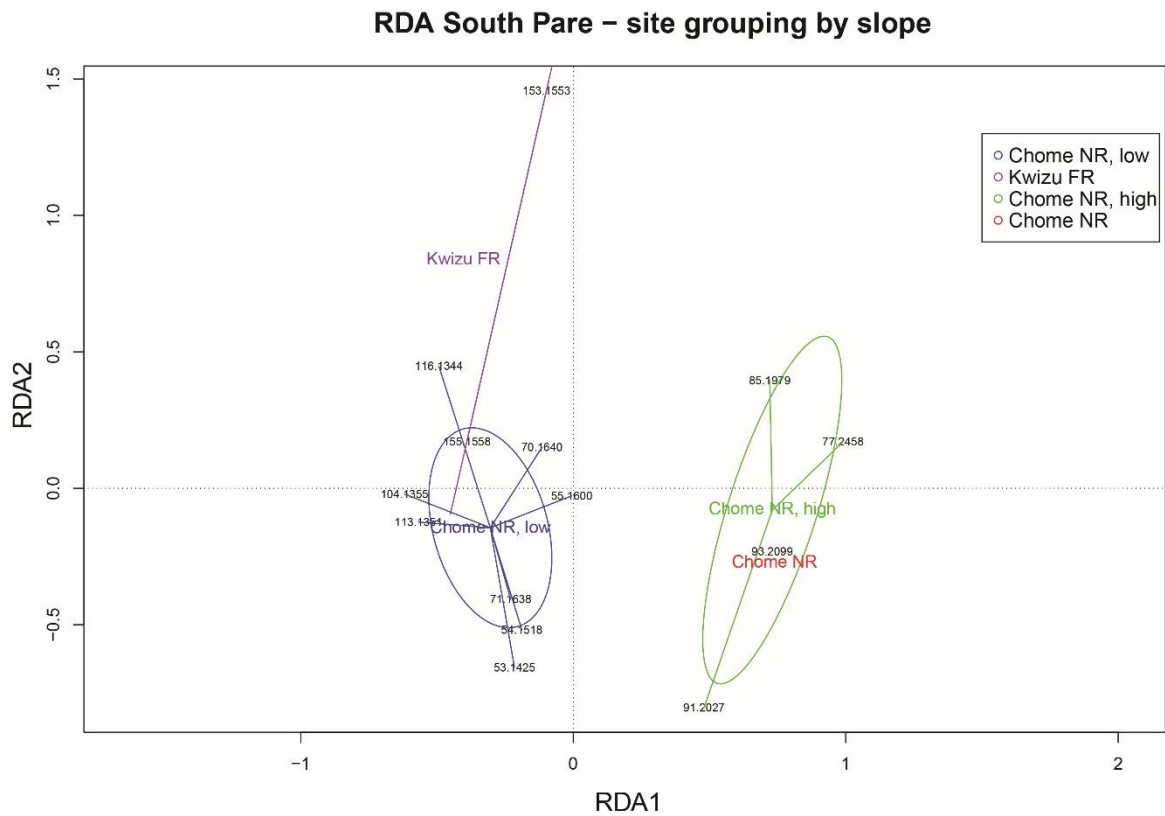
Figure 13. RDA of the South Pare dataset showing sites grouped by slope. Sites are named by their unique ID followed by their elevation.

RDA of sites with grouping by slope is presented in Figure 13. Again, sites group by slope to a certain degree. The middle and high elevations from Chome, whereas the low elevation samples from Chome group by themselves. One sample from Kwizu FR group with the low elevation sites from Chome, whereas the other is the outlier sample 153.
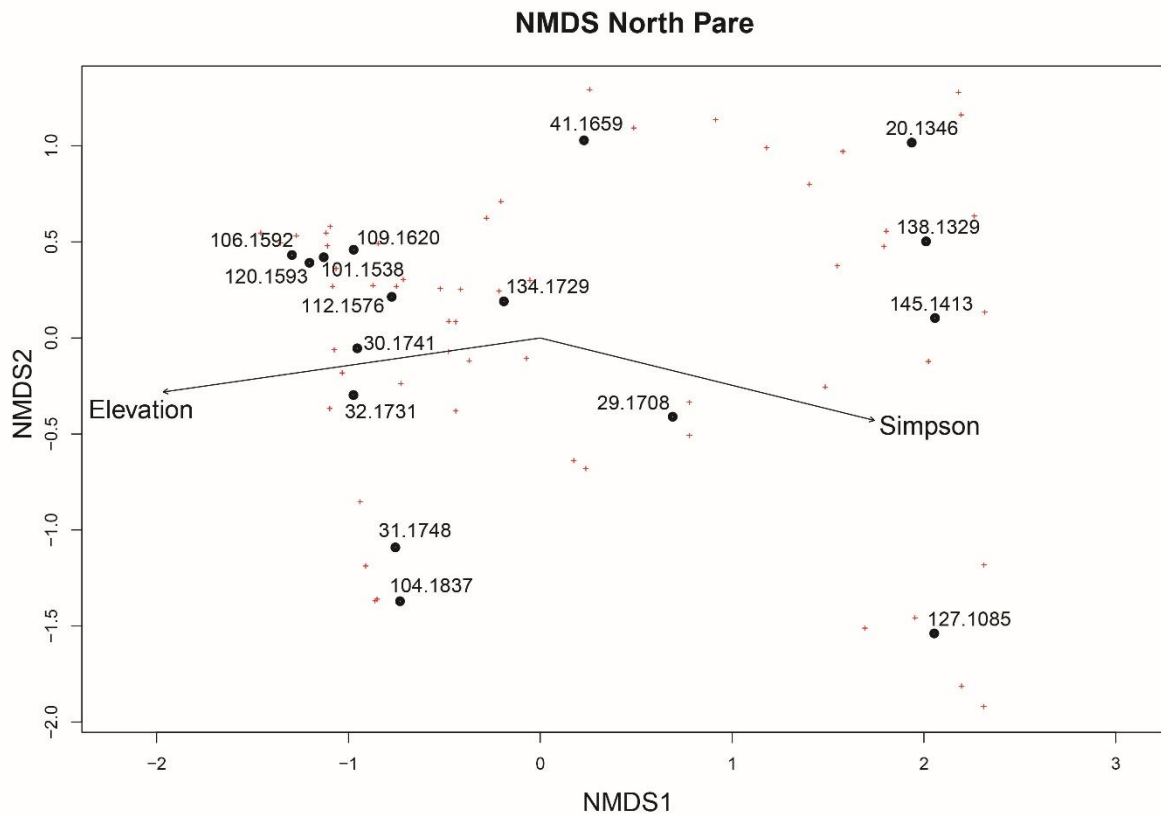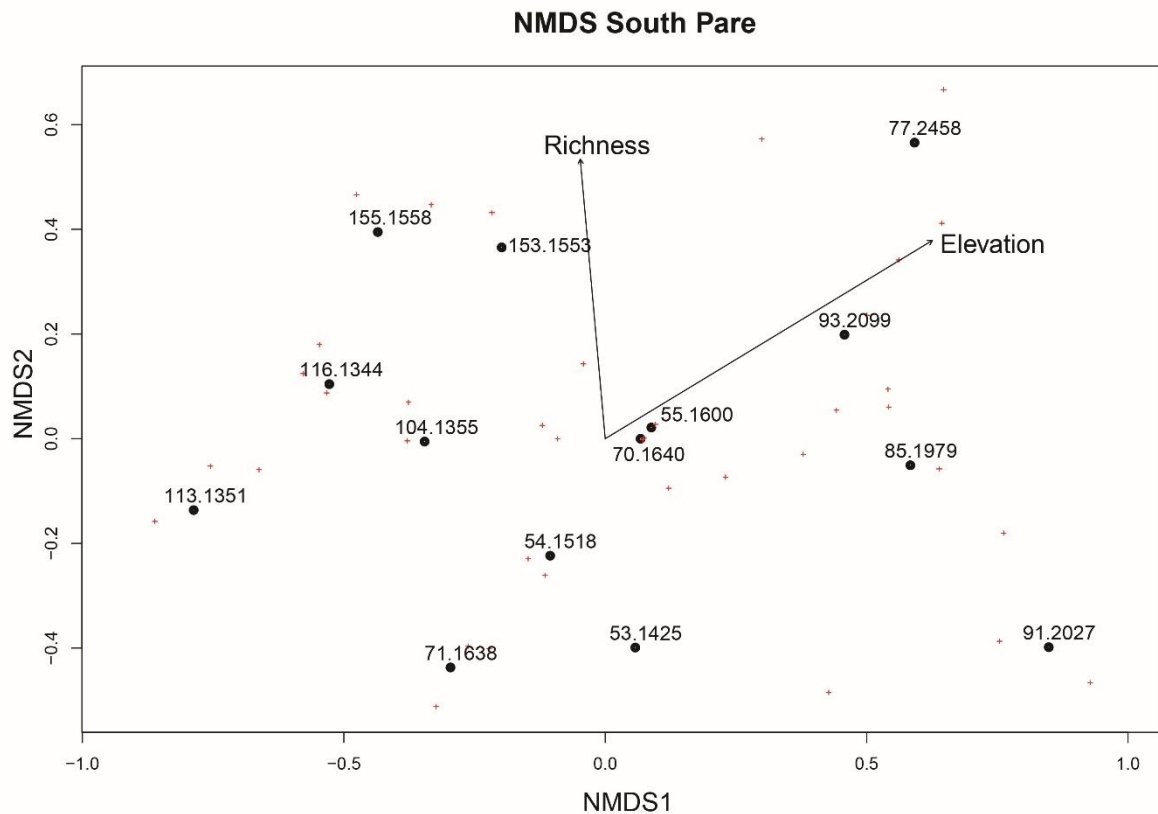
Figure 14. NMDS of the South Pare dataset showing the species as red plus-symbols, sites as dots, and elevation and species richness as explanatory variables. Sites are named by their unique ID followed by their elevation.

NMDS showing the groupings by distances for South Pare is presented in Figure 14. The corresponding stress values and Shepard plot are shown in Appendix C: Figures 31–32. $R^2 =$ 0.975. The NMDS results do not present any clear grouping of the data. The explanatory variables are not highly correlated, and both sites and species are widespread. Additional statistical information is provided in Appendix C: Tables 21–26.

3.3 β-diversity comparison

A comparison of the taxonomic gamma diversity of the three mountains and its decomposition into its alpha and beta components show different patterns of diversity for each of the components (Figure 15). Here, the data are interpolated and extrapolated by sample coverage.

Overall, gamma and alpha diversity is higher in Mt. Cameroon than in the Pare Mountains for any given sample coverage level. Generally, gamma and alpha diversity increases with higher

coverage in all the mountains for all diversity orders. The patterns of gamma and alpha diversity are similar in size and pattern for the Pare Mountains.

The beta component, the effective number of assemblages, presents a different pattern. For all diversity orders, Mt. Cameroon and South Pare share a similar amount of effective assemblages for all levels of coverage, with a trend of an increasing number of effective assemblages at with a higher emphasis on the most common species (diversity order q = 2). With respect to coverage, higher levels of coverage give a lower level of diversity for diversity orders q = 0 and q = 1, whereas the opposite is true for q = 2. Here, North Pare is the outlier, with generally a lower number of effective assemblages, and no observed change with increasing diversity order. Interestingly, for q = 0, the effective number of assemblages seem to have a unimodal pattern, with the largest number of assemblages at medial levels of coverage.
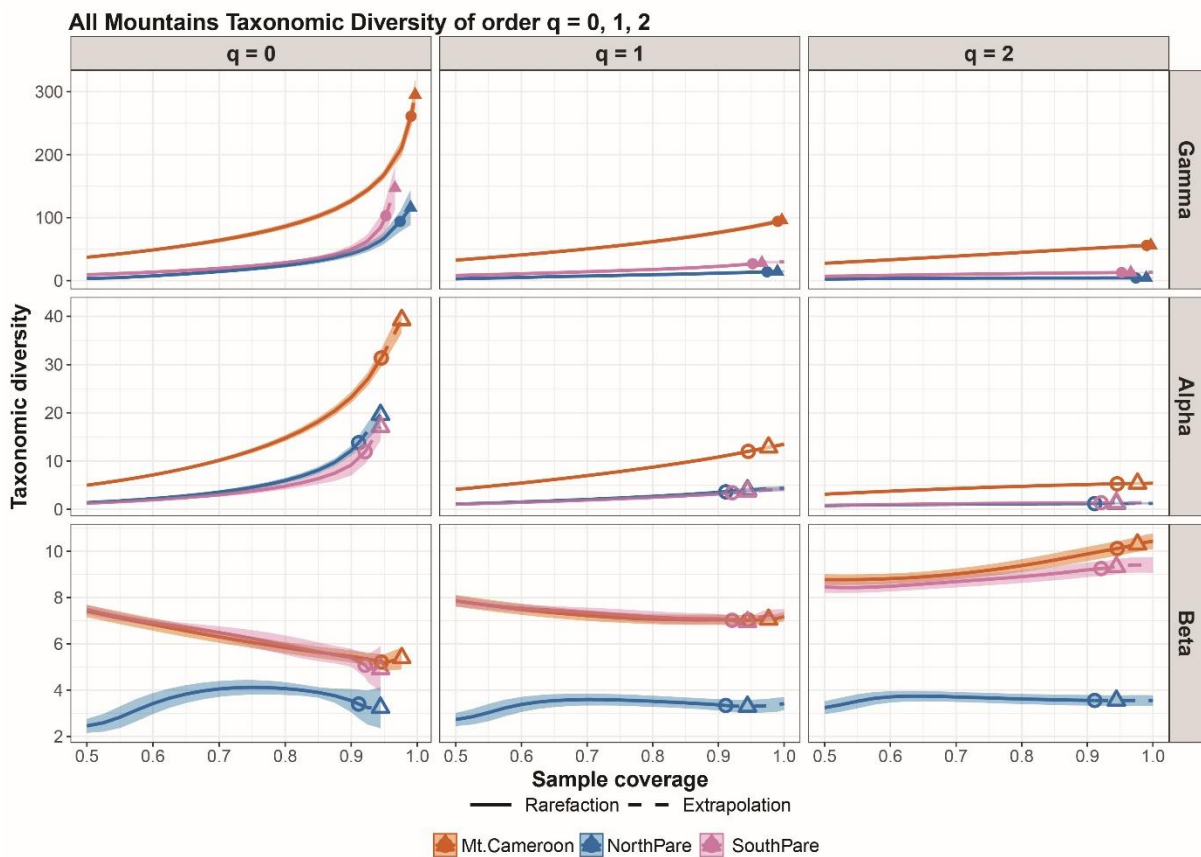
Figure 15. Gamma, alpha and beta taxonomic diversities for q = 0, q = 1 and q = 2 for all mountains, with interpolation and extrapolation by sample coverage. Circles indicate observed values (n) and triangles double sample size (2n).

The taxonomic beta diversity has also been compared as monotonically transformed measures of dissimilarity (Figure 16). These graphs present the overall pooled dissimilarities between assemblages within each mountain, for each diversity order.
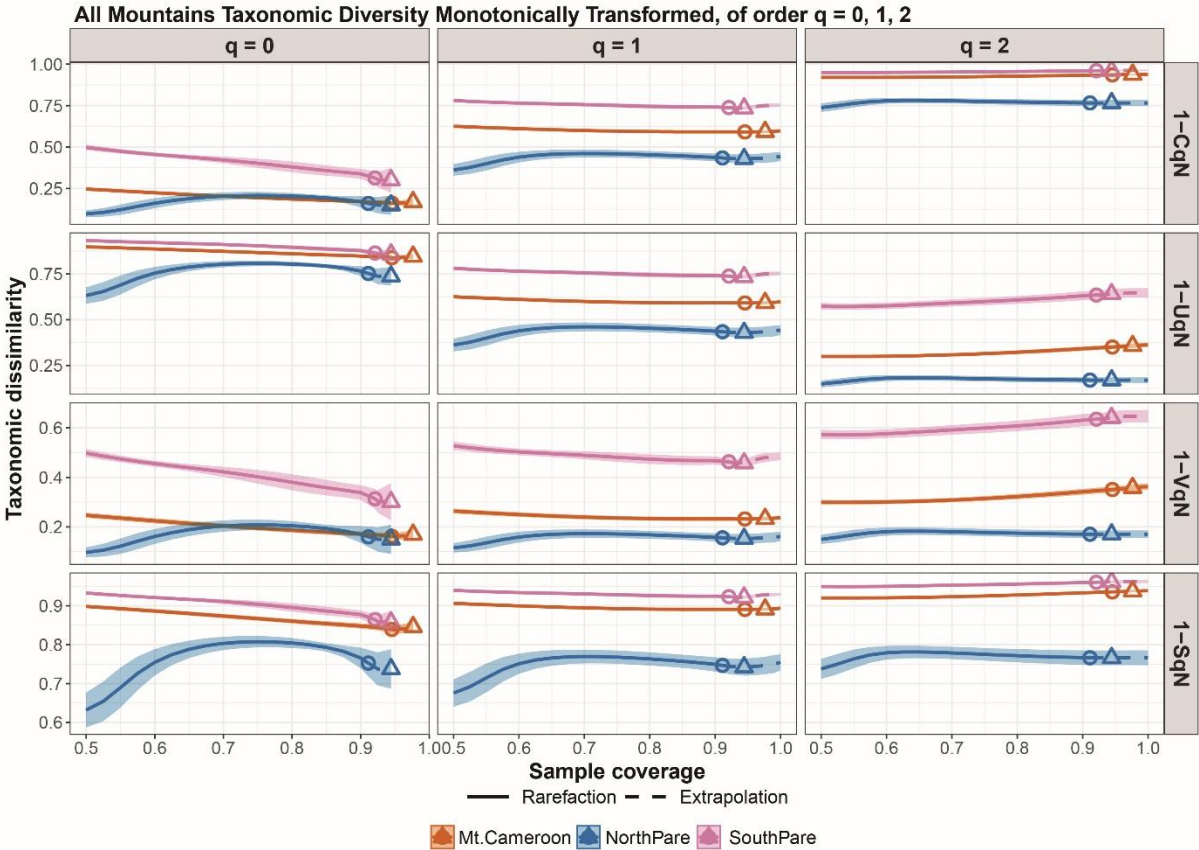
Figure 16. Taxonomic dissimilarities as monotonic transformations of beta diversities from Figure 15. 1-CqN is a Sørensen-type non-overlap measure measuring the proportion of non-overlapping species in each assemblage. 1-UqN is a Jaccard-type non-overlap measure measuring the proportion of non-overlapping species in the pooled assemblage. 1-VqN and 1-SqN are Sørensen and Jaccard-like turnover measures measuring species turnover relative to alpha and gamma respectively (Chao et al., 2019).
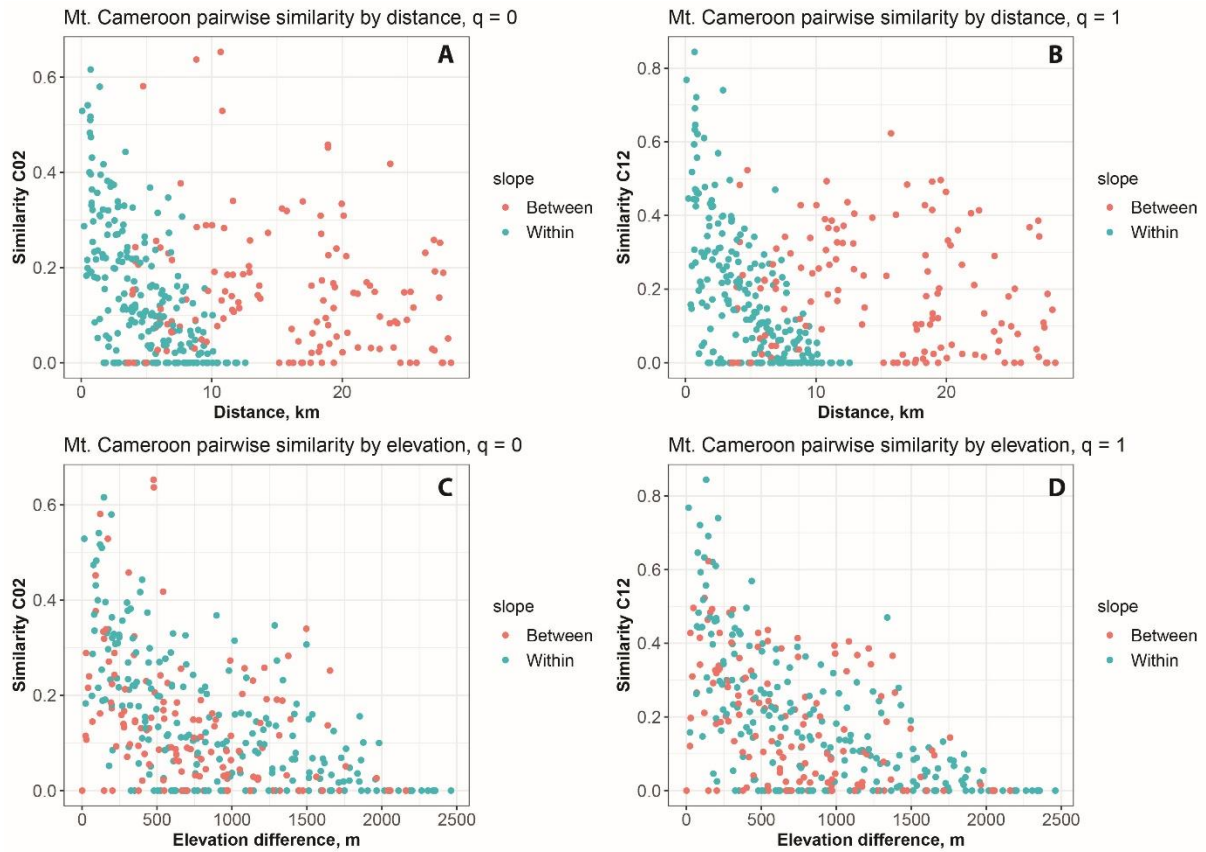
Figure 17. Sørensen-type similarity measure CqN for q = 0 and q = 1 against absolute distance (**A, B**) and elevational difference (**C, D**) for Mt. Cameroon. The pairwise similarities are colored by slope based on whether the samples are from the same slope or from different slopes.

Within Mt. Cameroon, absolute pairwise distance between assemblages does not hold a strong correlation with any decrease in similarity, neither within slopes nor between them. This holds true for different diversity orders (Figure 17A–B). Comparatively, difference in elevation between assemblages is negatively correlated with similarity. Here, no clear pattern of differences between and within slopes emerges for any diversity order (Figure 17C–D).
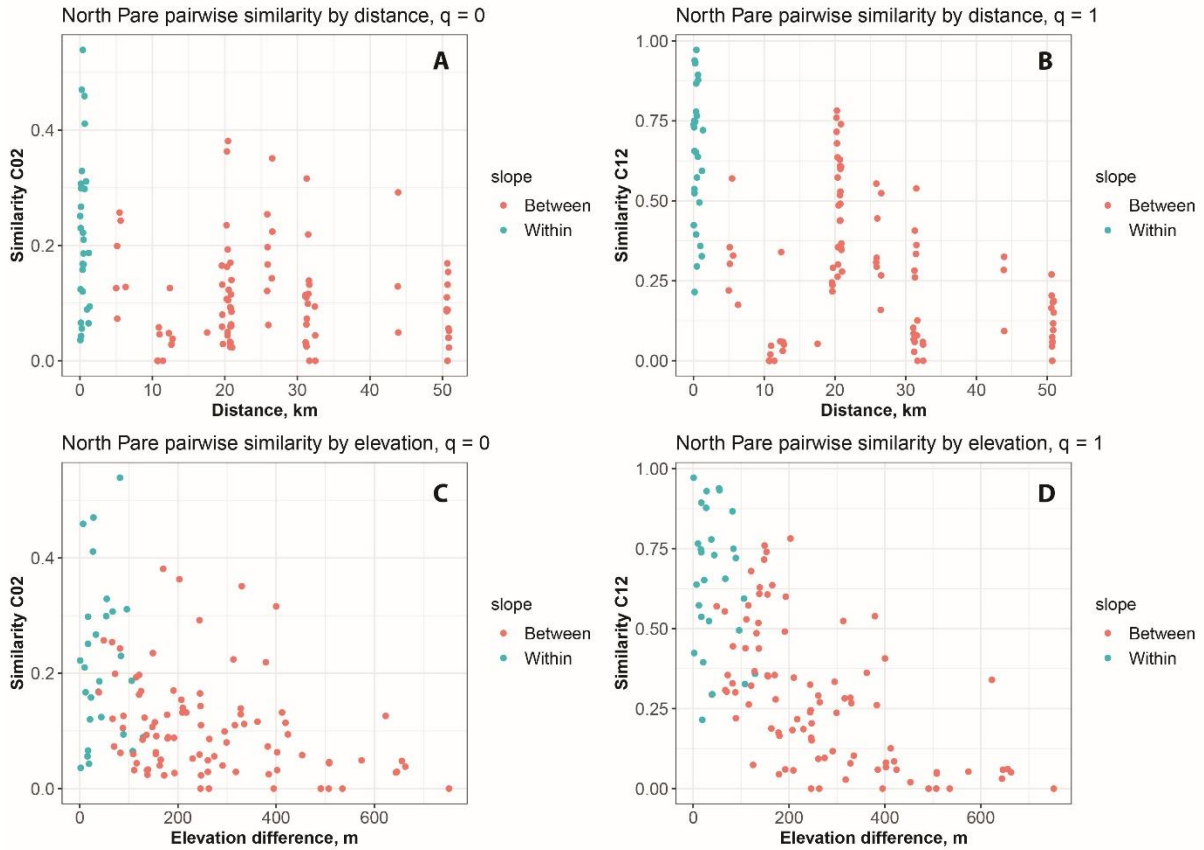
Figure 18. Sørensen-type similarity measure CqN for q = 0 and q = 1 against absolute distance (**A, B**) and elevational difference (**C, D**) for the North Pare. The pairwise similarities are colored by slope based on whether the samples are from the same slope or from different slopes.

In the Pare Mountains, assemblages closer together are generally more similar, both by absolute distance and by elevational differences (Figures 18–19). Compared to absolute distance, pairs of assemblages within the same slope are more similar than pairs between slopes. For the elevational differences, there is no clear pattern of differences based on slope despite the general decrease in similarity with increasing elevation differences.
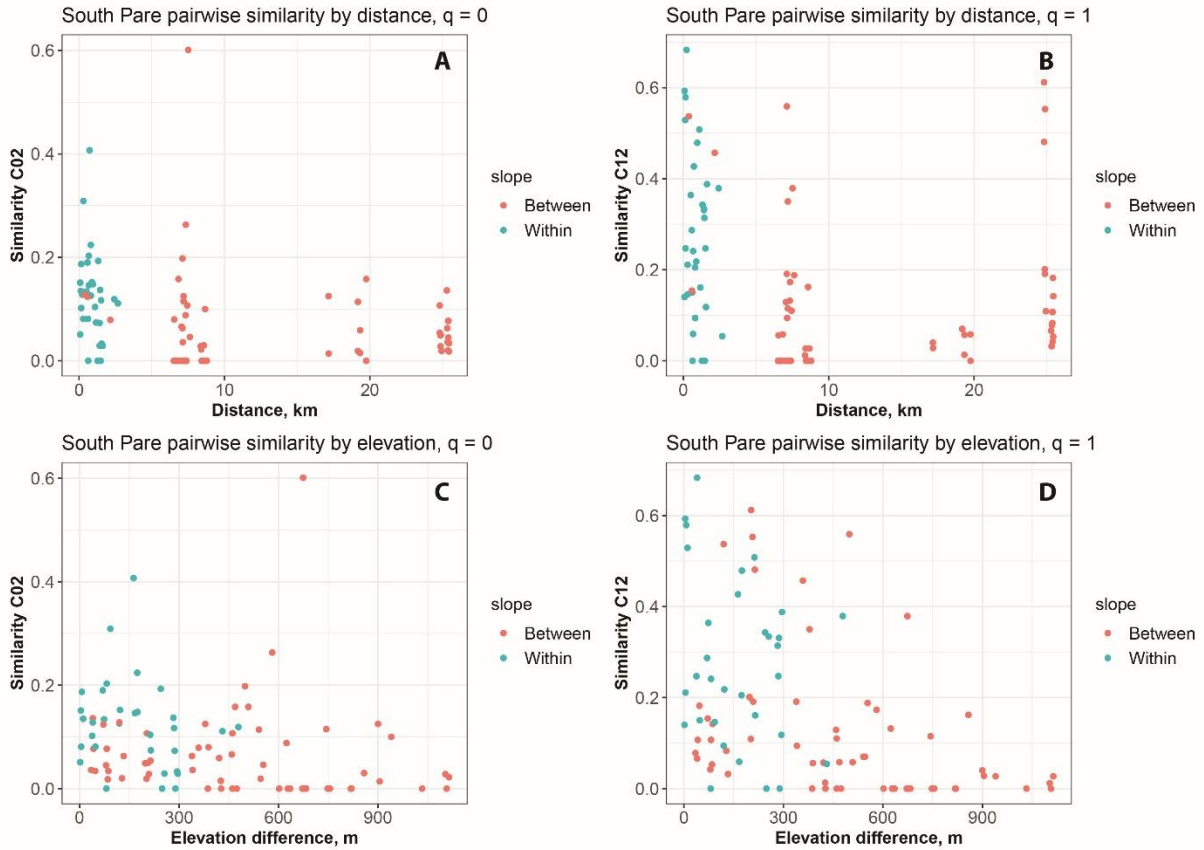
Figure 19. Sørensen-type similarity measure CqN for q = 0 and q = 1 against absolute distance (**A, B**) and elevational difference (**C, D**) for the South Pare. The pairwise similarities are colored by slope based on whether the samples are from the same slope or from different slopes.

3.4 Phylogenetic analysis

3.4.1 Multiplex PCR

From the initial test of the new primers, only 15 of the markers were successfully amplified, with a few more potential weak band results. During subsequent testing over a gradient of temperatures for annealing temperature optimization, *BOP1*, *rols* and *FBX011* all performed better at lower temperatures, with a most optimal annealing temperature of 55° C for the first PCR, and 52.5° C for the second PCR as in Tables 4–5.

Gel results from the multiplexed PCRs varied by samples and by groups with some general trends (ref Table 2). Group 2 generally performed the worst, with several samples lacking results altogether, and with low concentration for all but a few samples. Group 1 performed

better than group 2. PCRs had to be repeated for one sample from group 3, due to lab error during the first attempt. The repeated amplification was successful. Groups 3 and 4 performed much better than the other groups, with all samples yielding high DNA concentrations.

The worst performing samples in each group were generally from the same individuals across the different groups, with some exceptions.

3.4.2 Mapping to reference sequences

A total of 4.61 M reads were created and stored as fastq files after the sequencing. The number of reads created per barcode is presented in Appendix D: Table 33. The number of reads per barcode are in a range of ~20000 – 170000.

During mapping to reference sequences, only 10 of the markers mapped consistently well, with >10 reads per marker for most samples. The best performing genes are marked in bold in Table 2. Some unsuccessful markers consistently had >1000 amplified reads, yet the reads were incomplete. Most likely, in these cases, during the PCR, the highly degenerate primers annealed to unintended targets and amplified gene fragments that were shorter than intended. Other unsuccessful markers failed to amplify altogether except for occasional single reads for some barcodes.

Generally, the outgroup taxa performed slightly differently from the ingroup taxa. For some markers the primers worked better for the outgroup taxa, successfully amplifying markers that elsewise did not work. Similarly, primers interacted differently, eliminating the problem of incomplete reads for some genes. However, the same 10 markers worked consistently, and the differences between the taxa did not alter which markers were included downstream.

The outgroup Oxyteline specimen *Oxytelopsis sp.* mapped reads too poorly and was therefore removed from the dataset.

3.4.3 Maximum Likelihood

The single most likely Maximum Likelihood tree based on the dataset partitioned both by gene and by codon position is presented in Figures 20–21. Here the tree has been rooted between *Alloproteinus nigriceps* and the remaining species. The ML trees based on the dataset partitioned by gene only and the dataset with no partitioning yielded identical topologies with only slightly worse node support. These trees are shown in Appendix E: Figures 33–34. SH-aLRT and Ultrafast Bootstrap (UFB) support values were considered robust when SH-aLRT >80% and UFB >95%.

Most nodes have high or maximum support in both UFB and SH-aLRT, including the outgroup taxa. Some nodes were not supported in the topology (SH-aLRT <80% or UFB <95%). These support values are not shown in the tree.

The four species of *Metopsia* were recovered as a well supported monophyletic group (clade **F**). The tree also recovers (*Megarthrus* + *Metopsia*) as a weakly supported monophyletic group (clade **A**), with *Proteinus* as its sister group. In contrast, the clade consisting of the members of the tribe Protinini (*Megarthrus* + *Metopsia* + *Proteinus*) has robust support. *Metopsia*, however, is recovered as nested within *Megarthrus*, as sister group to clade **G** consisting of two unidentified Vietnamese species as well as *Megarthrus pictus* (clade **E**). Thus, the current *Megarthrus* is rendered paraphyletic. *Megarthrus montanus* is the sister taxon of the clade composed of all the other species of *Megarthrus* and *Metopsia,* which also has robust support. The outgroup specimens *Alloproteinus nigriceps*, *Austrorhysus sp.* and *Proteinus brachypterus* have been highlighted together as group **H** for visual simplicity.

The Afrotropical *Megarthrus* species make up a monophyletic group (clade **C**) nested within a larger clade consisting of the Afrotropical species, as well as mainly Holarctic species (clade **B**). Clade **E** is recovered as the sister group to clade **B**.

Within the Afrotropical clade (Figure 21), certain subclades are restricted geographically, like the Ethiopian subclade **L**. Other subclades contain closely related species from distantly located mountains, like Mt. Cameroon and Mt. Elgon (clade **K**) or Mt. Cameroon and Mt. Kilimanjaro (clade **N**).

44
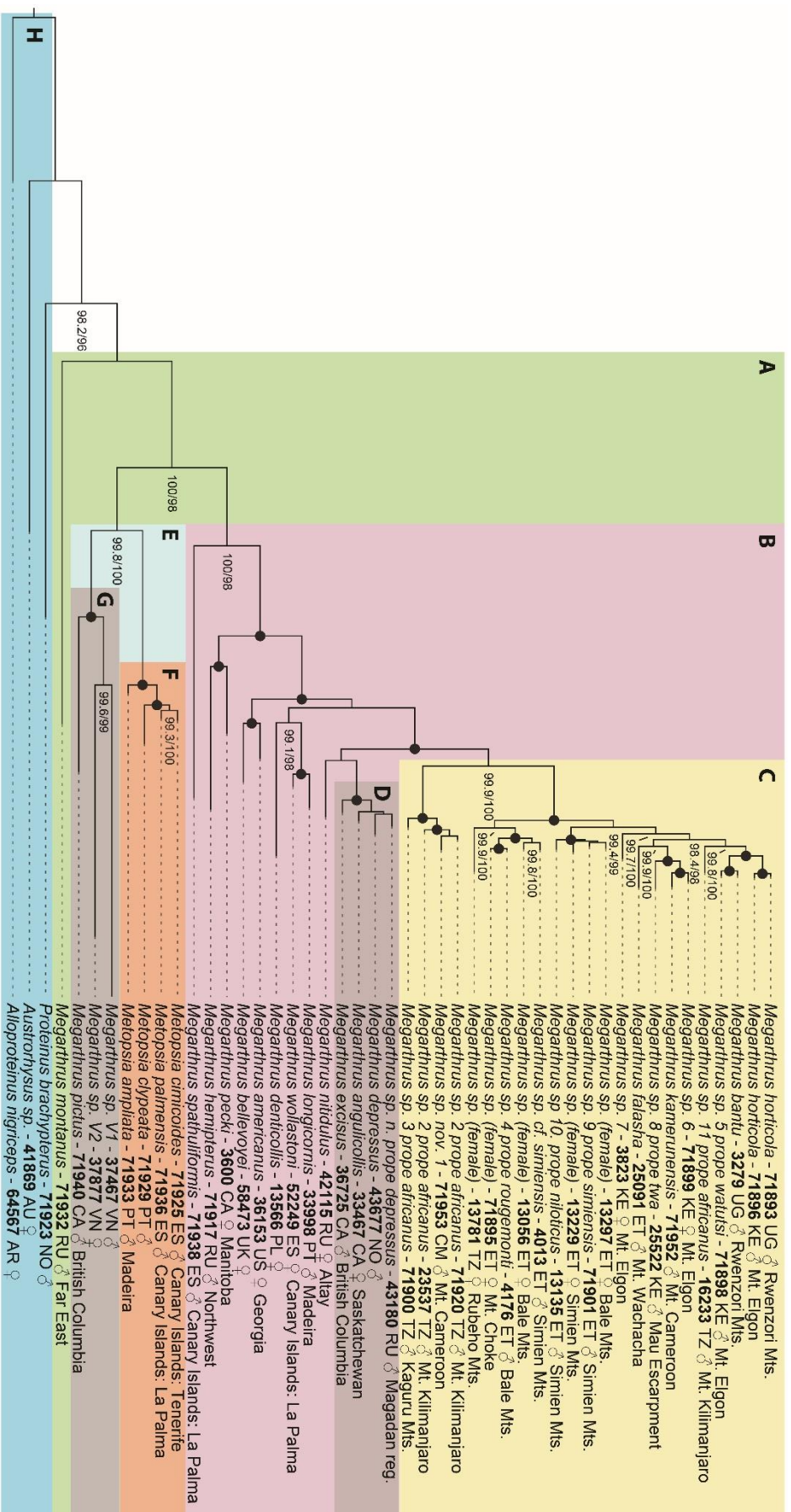
Figure 20. Single most likely Maximum Likelihood phylogeny of the 47 included Proteininae specimens, with data partitioned by gene and codon position. Only node support values with SH-aLRT > 80% and UFB > 95% are shown. Nodes with 100%/100% support are shown only as black dots. Noteworthy clades are highlighted and denoted by letter. Each specimen name is followed by its unique ID, country code, sex, and collection area.

C

Megarthrus horticola - **71893** UG ♂ Rwenzori Mts.
Megarthrus horticola - **71896** KE ♂ Mt. Elgon
Megarthrus bantu - **3279** UG ♂ Rwenzori Mts.
Megarthrus sp. 5 prope watutsi - **71898** KE ♂ Mt. Elgon
Megarthrus sp. 11 prope africanus - **16233** TZ ♂ Mt. Kilimanjaro
Megarthrus sp. 6 - **71899** KE ♀ Mt. Elgon
Megarthrus kamerunensis - **71952** CM ♂ Mt. Cameroon
Megarthrus sp. 8 prope twa - **25522** KE ♂ Mau Escarpment
Megarthrus falasha - **25091** ET ♂ Mt. Wachacha
Megarthrus sp. 7 - **3823** KE ♀ Mt. Elgon
Megarthrus sp. (female) - **13297** ET ♀ Bale Mts.
Megarthrus sp. 9 prope simiensis - **71901** ET ♂ Simien Mts.
Megarthrus sp. (female) - **13229** ET ♀ Simien Mts.
Megarthrus sp 10. prope niloticus - **13135** ET ♂ Simien Mts.
Megarthrus sp. cf. simiensis - **4013** ET ♂ Simien Mts.
Megarthrus sp. (female) - **13056** ET ♀ Bale Mts.
Megarthrus sp. 4 prope rougemonti - **4176** ET ♂ Bale Mts.
Megarthrus sp. (female) - **71895** ET ♀ Mt. Choke
Megarthrus sp. (female) - **13781** TZ ♀ Rubeho Mts.
Megarthrus sp. 2 prope africanus - **71920** TZ ♂ Mt. Kilimanjaro
Megarthrus sp. nov. 1 - **71953** CM ♂ Mt. Cameroon
Megarthrus sp. 2 prope africanus - **23537** TZ ♂ Mt. Kilimanjaro
Megarthrus sp. 3 prope africanus - **71900** TZ ♂ Kaguru Mts.

Figure 21. Subclades of the Afrotropical clade of *Megarthrus* from Figure 20. Only node support values with SH-aLRT > 80% and UFB > 95% are shown. Nodes with 100%/100% support are shown only as black dots. Noteworthy clades are highlighted and denoted by letter. Each specimen name is followed by its unique ID, country code, sex, and collection area.

3.5 Species descriptions

Megarthrus kamerunensis **Bernhauer, 1942**

Figs. 23–24 (5–21)

Type material. Neotype, male: Cameroon, Mt. Cameroon National Park, trail between Crater Lake and Mann's Spring Camps, 4°8.817'N 9°5.129'S WGS84, 1810 m, montane forest with herbaceous patches sifted, 8.iii.2023, V. I. Gusarov & L. M. Lyonga, deposited at NHM.

Additional material. Same data as neotype, 25 additional specimens deposited at NHM; Cameroon, Mt. Cameroon National Park, trail between Crater Lake & Mann's Spring Camps, 4°8.352'N 9°6.614'E WGS84, 2022 m, large, medium, small trees, shallow ravine with herbaceous vegetation, sifting, 10.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 17 specimens in NHM; Cameroon, Mt. Cameroon National Park, environment of Mann's Spring Camp, 4°8.861'N 9°7.358'E WGS84, 2362 m, montane forest, gallery forest in broad ravine, sifting, 9.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 4 specimens in NHM; Cameroon, Mt. Cameroon National Park trail between Crater Lake & Mann's Spring Camps, 4°8.389'N 9°6.971'E WGS84, 2117 m, montane forest, large, medium, small trees, sifting, 10.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 1 specimen in NHM; Cameroon, Mt. Cameroon National Park, below P&T Camp, 4°17.480'N 9°12.341'E WGS84, 2164 m, montane forest, large trees, gaps with bushes, ferns and Poaceae, sifting, 14.iii.2023, V. I. Gusarov & team, 7 specimens in NHM; Cameroon, Mt. Cameroon National Park, trail between Mann's Spring Camp and Bokwango village (Buea), 4°7.525'N 9°9.043'E WGS84, 1901 m, montane forest with small gaps, ferns, Poaceae, sifting, 13.iii.2023, V. I. Gusarov & L. M. Lyonga, 1 specimen in NHM.

Description. Habitus as in Figure 22 (1–2). Body length 2.6–2.8 mm; total length of head, pronotum and elytra combined 1.4–1.6 mm; maximum pronotal width 0.4–0.5 mm. Body light brown with yellowish lateral margins and appendages; antennae entirely dark-brown; body darker ventrally. Head narrower than pronotum at anterior margin; dorsal surface smooth and densely punctate; U-shaped frontal impression deep; supra-ocular margin sinuate in dorsal view; temples abruptly narrowed behind eyes; occipital ridge indistinct; frons slightly arcuate apically. Antennae quite long; scape broad apically; antennomere 11 almost as long as the two preceding antennomeres combined; antennomeres 5–11 densely

pubescent. Pronotum slightly narrower than elytra; not distinctly toothed at lateral margins; lightly rounded and narrowed anteriorly; more than twice as wide as long; smooth and densely punctate; slightly curved toward lateral edges, with flat mesal portion; deep and broad complete median longitudinal impression. Elytra almost twice as long as pronotum; widening posteriorly; surface smooth and densely punctate. Frons with setae oriented backwards. Elytra and pronotum with setae recumbent backwards. Median prosternal ridge absent. Mesotibiae slightly curved; all tarsi heavily pubescent.

Male: Mesotibiae with peg-like setae arranged in two rows and a cluster apically; metatibiae with peg-like setae arranged in a single row. Meta- and mesotrochanters both with peg-like setae. Protarsal segment 1 bearing adhesive setae. Abdominal sternite VIII deeply rounded as in Figure 23 (8); tergite VIII as in Figure 23 (9); hemitergites as in Figure 23 (7). Adeagus as in Figure 23 (5–6).

Female: Abdominal sternite VIII as in Figure 23 (11); tergite VIII as in Figure 23 (12). Genitalia as in Figure 23 (10).

Diagnosis. *Megarthrus kamerunensis* is distinguished from the similar *M. wittei* Cameron 1950 mainly by its larger size and lighter coloration. Males share bearing adhesive setae on the first protarsomere, the lack of peg-like setae on the protibiae, and the bearing of peg-like setae on both meso- and metatibiae and on meso- and metatrochanters. However, *M. kamerunensis* bear 6–7 peg-like setae on the metatrochanter, more than *M. wittei. Megarthrus kamerunensis* also bear peg-like setae in two rows on mesotibiae with cluster at apex and peg-like setae in one row on metatibiae, whereas *M. wittei* bears them in one row with cluster apically on mesotibiae, and in a field on metatibiae. Similarly, *M. kamerunensis* differs from the new *M. sp. 1* in size and coloration, as well as its broad scape and non-toothed lateral margins of the pronotum. Finally, *M. kamerunensis* differ from all the externally similar species in the shape and characteristics of the male tergite and sternite VIII as well as the adeagus (Figure 23 (5–9)).

Figure 22. *Megarthrus*, habitus, male; *Megarthrus kamerunensis*: dorsal (1), and ventral (2); *Megarthrus sp. 1* **sp. nov.**: dorsal (3), and ventral (4). Scale bar = 1 mm.
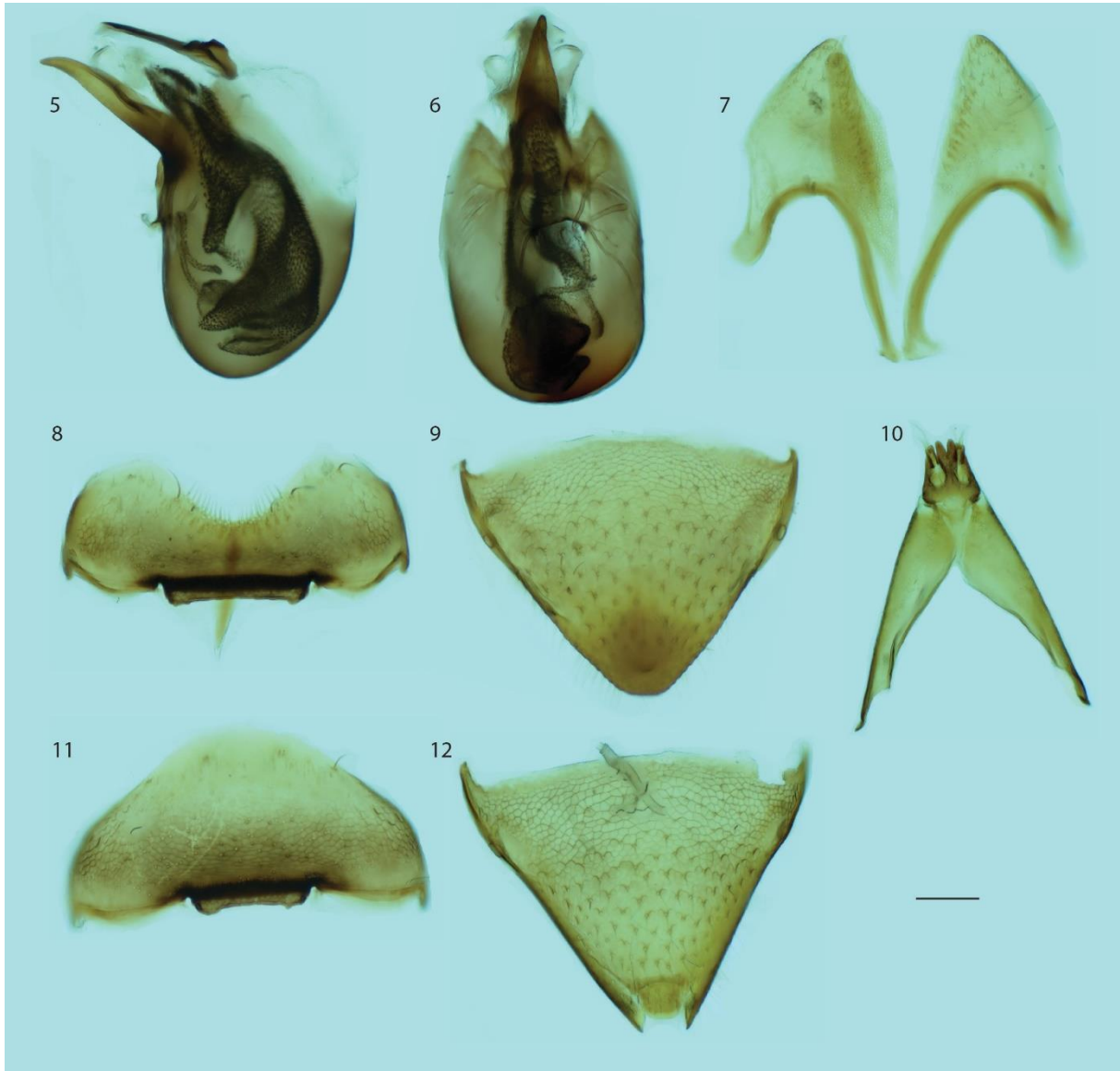
Figure 23. *Megarthrus kamerunensis* (male): adeagus in lateral (5) and ventral (6) view; hemitergite IX (7); abdominal sternite VIII in ventral view (8); abdominal tergite VIII in dorsal view (9). *Megarthrus kamerunensis* (female): genital segments (10); abdominal sternite VIII in ventral view (11); abdominal tergite VIII in dorsal view (12). Scale bar = 0.1 mm.
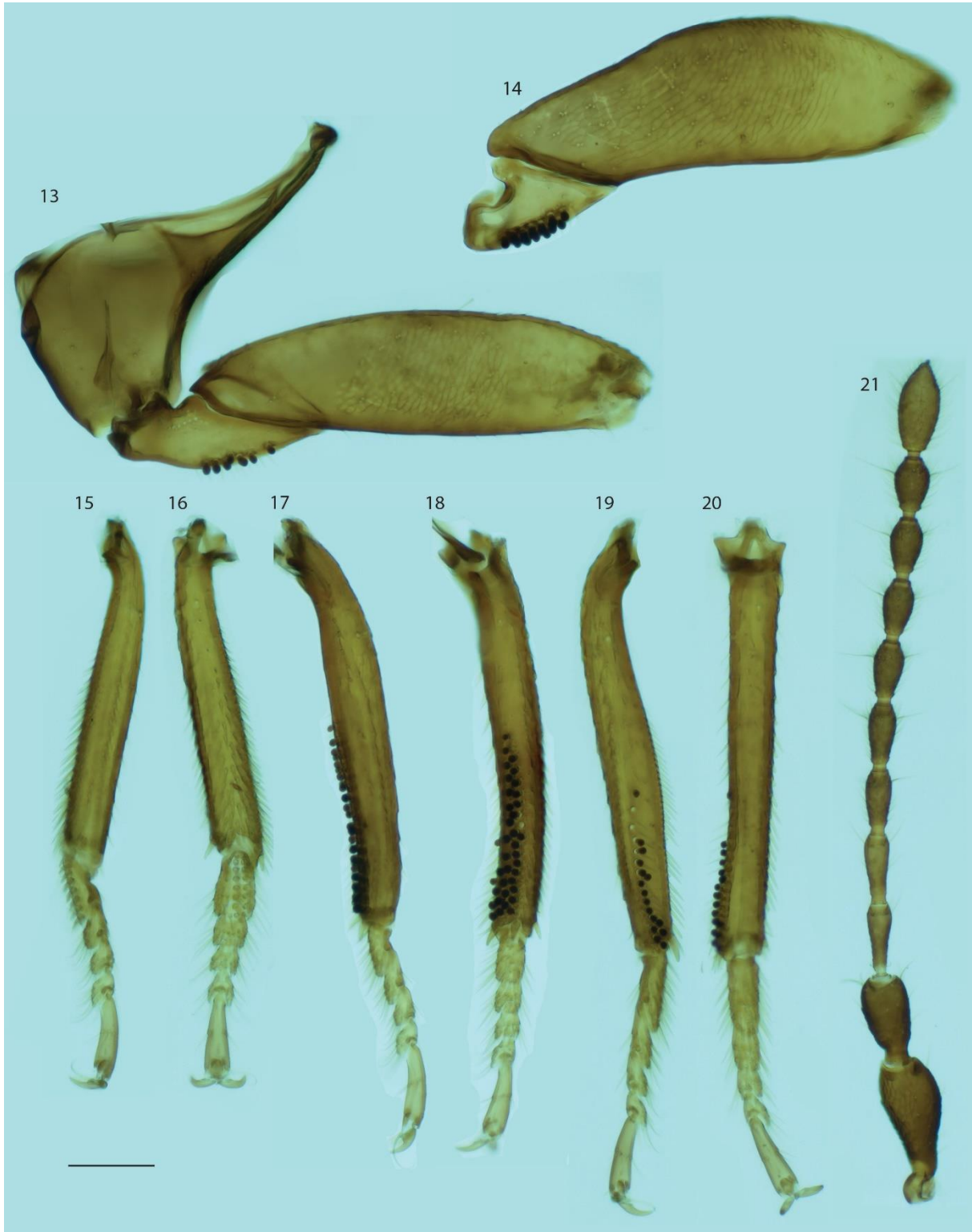
Figure 24. *Megarthrus kamerunensis* (male): metacoxae, metatrochanter and metafemura (13); mesotrochanter and mesofemura (14); protibiae and protarsi in lateral (15) and ventral (16) view; mesotibiae and mesotarsi in lateral (17) and ventral (18) view; metatibiae and metatarsi in lateral (19) and ventral (20) view; antennae (21). Scale bar = 0.1 mm.

Distribution. *Megarthrus kamerunensis* is endemic to Mt. Cameroon, with collections made in the south-east, south-west and northern part of the mountain. All collections were made by sifting leaf-litter in the elevational range of 1800–2350 m.

Comments. As documented by Cuccodoro and Löbl (1995), the only known syntype of *Megarthrus kamerunensis* was lost in transit while being returned to the Field Museum following the death of W. O. Steel in 1969. With the discovery of more than one species of *Megarthrus* on Mt. Cameroon, designating a neotype for *M. kamerunensis* becomes necessary with the express purpose of clarifying the taxonomic status of the species. Without fixing the identity of *M. kamerunensis*, the new species cannot be described. As described above, the neotype is consistent with the original description of *M. kamerunensis* (Bernhauer & Paulian, 1942). In the original description, the authors refer to a curved inscision of sternite VI of the male specimen. In reality, the authors referred to the sixth visible sternite, which is morphologically the eight. This becomes evident when the other descriptions (e.g. that of *Paederus*) in the same paper are examined. In the redescription, segments are counted by morphology, and the character is referred to as occurring on sternite VIII. In the original description, the type locality is described as "the south-east slope" at elevations 1800 – 2000 m (Bernhauer & Paulian, 1942). This is the same slope of Mt. Cameroon where the neotype was collected. The neotype is deposited in the collection of the Natural History Museum of Oslo. Based on the above, all the requirements of Article 75 of the Code (ICZN, 1999) have been met.

Megarthrus sp. 1, **sp. nov.**

Figs. 25–26 (22–39)

Type material. Holotype, male: Cameroon, Mt. Cameroon National Park, environs of Mann's Spring Camp, 4°9.013'N 9°7.203'E WGS84, 2386 m, montane forest in broad ravine, large, medium and small trees, Poaceae in understory, sifting, 11.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, deposited at NHM.

Paratypes. Same data as holotype, 26 additional specimens deposited at NHM; Cameroon, Mt. Cameroon National Park trail between Crater Lake & Mann's Spring Camps, 4°8.389'N 9°6.971'E WGS84, 2117 m, montane forest, large, medium, small trees, sifting,

10.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 4 specimens in NHM; Cameroon, Mt. Cameroon National Park, trail between Crater Lake & Mann's Spring Camps, 4°8.352'N 9°6.614'E WGS84, 2022 m, large, medium, small trees, shallow ravine with herbaceous vegetation, sifting, 10.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 11 specimens in NHM; Cameroon, Mt. Cameroon National Park, trail between Plantecam and Crater Lake Camps, 4°8.340'N 9°4.340'E WGS84, 1422 m, montane forest, herbaceous patches, sifting, 6.iii.2023, V. I. Gusarov & L. M. Lyonga, 1 specimen in NHM; Cameroon, Mt. Cameroon National Park, trail between Plantecam and Crate Lake Camps, 4°8.401'N 9°4.780'E WGS84, 1621 m, montane forest, herbaceous patches, sifting, 6.iii.2023, V. I. Gusarov & L. M. Lyonga, 3 specimens in NHM; Cameroon, Mt. Cameroon National Park, trail between Mann's Spring Camp and Bokwango village (Buea), 4°7.525'N 9°9.043'E WGS84, 1901 m, montane forest with small gaps, ferns, Poaceae, sifting, 13.iii.2023, V. I. Gusarov & L. M. Lyonga, 1 specimen in NHM; Cameroon, Mt. Cameroon National Park, trail between Crater Lake and Mann's Spring Camps, 4°8.817'N 9°5.129'E WGS84, 1810 m, montane forest, herbaceous patches, sifting, 8.iii.2023, V. I. Gusarov & L. M. Lyonga, 5 specimens in NHM; Cameroon, Mt. Cameroon National Park, environment of Mann's Spring Camp, 4°8.861'N 9°7.358'E WGS84, 2362 m, montane forest, gallery forest in broad ravine, sifting, 9.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 10 specimens in NHM; Cameroon, Mt. Cameroon National Park, environment of Mann's Spring Camp, 4°8.641'N 9°7.238'E WGS84, 2254 m, montane forest, sifting, 9.iii.2023, V. I. Gusarov & G. M. Ojong-Nkongho, 2 specimens in NHM; Cameroon, Mt. Cameroon National Park, below P&T Camp, 4°17.480'N 9°12.341'E WGS84, 2164 m, montane forest, large trees, gaps with bushes, ferns and Poaceae, sifting, 14.iii.2023, V. I. Gusarov & L. M. Lyonga, 81 specimens in NHM.

Description. Habitus as in Figure 22 (3–4). Body length 3.2–3.3 mm; total length of head, pronotum and elytra combined 2.0–2.2 mm; maximum pronotal width 1.1–1.2 mm. Body dark brown to black; appendages, abdominal segment VIII and mouthparts lighter; antennomere 11 distinctly lighter than the rest of the antennae. Head with coarse microsculpture; densely punctate; supra-ocular margin sinuate in dorsal view; temples abruptly narrowed behind eyes; occipital ridge not distinct. All antennomeres longer than broad; antennomere 11 slightly shorter than two preceding combined; antennomeres 5–11 densely pubescent. Pronotum slightly narrower than elytra; sharply toothed on lateral margins; microsculpture smooth and densely punctate; anterior angles slightly protruding; longitudinal median impression broad, deep and complete; pronotum twice as broad as long.
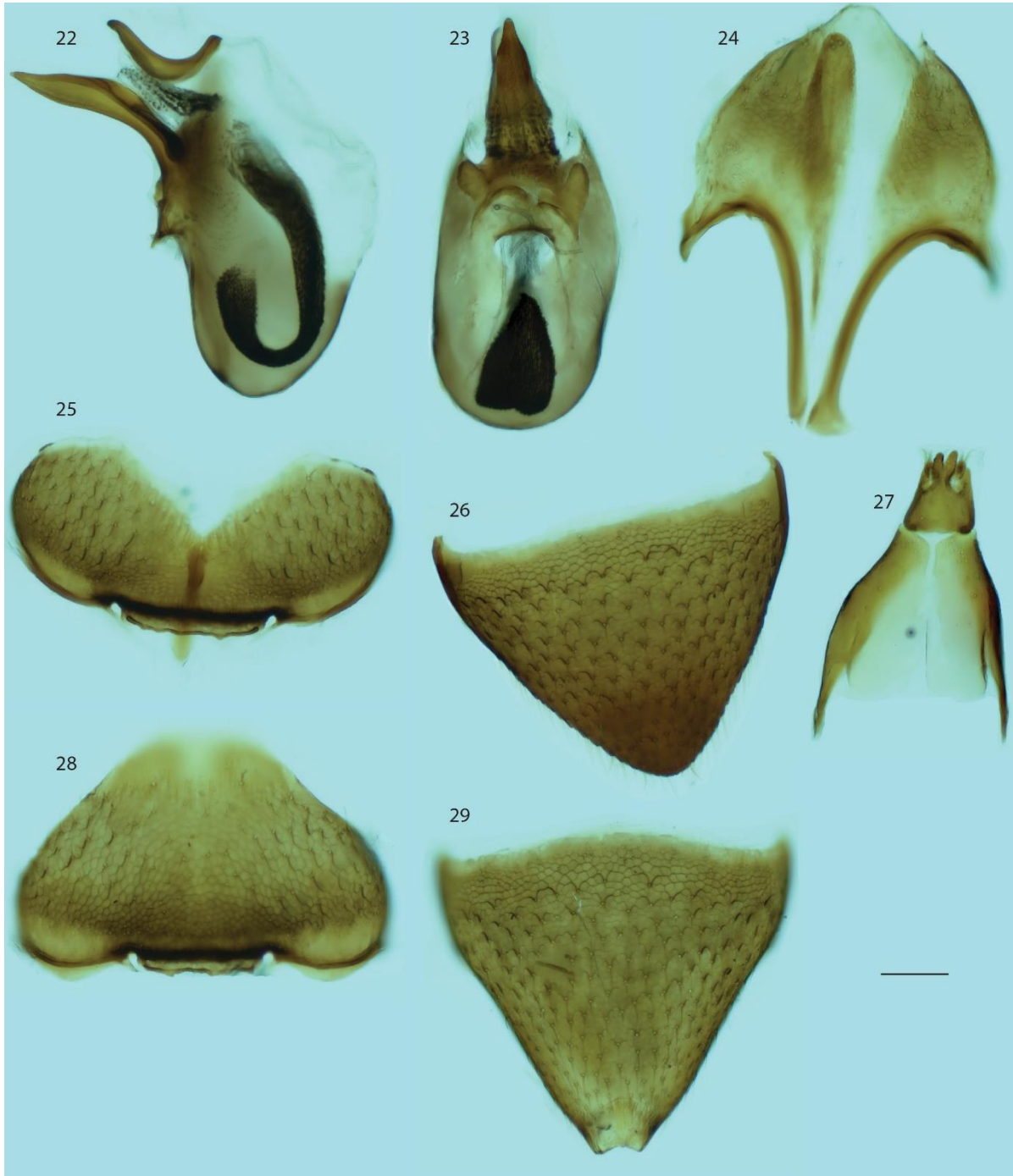
Figure 25. *Megarthrus sp. 1*, **sp. n.** (male): adeagus in lateral (22) and ventral (23) view; hemitergite IX (24); abdominal sternite VIII in ventral view (25); abdominal tergite VIII in dorsal view (26). *Megarthrus sp. 1*, **sp. n.** (female): genital segments (27); abdominal sternite VIII in ventral view (28); abdominal tergite VIII in dorsal view (29). Scale bar = 0.1 mm

Figure 26. *Megarthrus sp. 1*, **sp. n.** (male): mesotrochanter (30) and mesofemura (31); metatrochanter and metafemura (32); protibiae and protarsi in lateral (33) and ventral (34) view; mesotibiae and mesotarsi in lateral (35) and ventral (36) view; metatibiae and metatarsi in lateral (37) and ventral (38) view; antennae (39). Scale bar = 0.1 mm.

Elytra about twice as long as pronotum; widening posteriorly; suface smooth and densely punctate. Frons with setae oriented backwards. Elytra and pronotum with setae recumbent backwards. Median prosternal ridge absent. Tibiae each with two rows of setae; all tarsi highly pubescent.

Male: All tibiae slightly curved; protarsomere 1 without adhesive setae; mesotrochanter with one row of peg-like setae; metatrochanter with no peg-like setae; mesotibiae with one row of peg-like setae; metatibiae with two peg-like setae at apex. Abdominal tergite VIII as in Figure 25 (26); sternite VIII as in Figure 25 (25); hemitergites as in Figure 25 (24). Adeagus as in Figure 25 (22–23).

Female: Abdominal tergite VII long and strongly curved; abdominal tergite VIII as in Figure 25 (29), sternite VIII as in Figure 25 (28). Genitalia as in Figure 25 (27).

Diagnosis. *Megarthrus. sp. 1* can be distinguished from other Afrotropical species by its unique combination of male characters. *Megarthrus. sp. 1* does not have adhesive setae on the first protarsinere, has one row of peg-like setae on mesotrochanter, one row on mesotibiae, no peg-like setae on metatrochanter and only a few peg-like setae near the apex of the metatibiae. The species is also different from the other species of the region by its larger size and dark brown coloration, and from *Megarthrus kamerunensis* by a narrower scape and its lighter antennomere 11. *Megarthrus. sp. 1* also has a distinctive shape of the adeagus, with a curved dorsoapical sclerite of the median lobe (Figure 25 (22)).

Distribution. *Megarthrus sp. 1* is endemic to Mt. Cameroon, with collections made in the south-east, south-west and northern part of the mountain. All collections were made by sifting leaf-litter in the elevational range of 1400–2400m.

4   Discussion

4.1   α-diversity results and sampling issues

The taxonomic α-diversity pattern along the altitudinal gradient in Mt. Cameroon (Figure 3) roughly fits the monotonically decreasing pattern as described by McCain and Grytnes (2010). The middle elevation samples have a worse fit than the rest of the samples from the gradient, yet the pattern is quite robust, as most of the elevational range of the mountain has been sampled. The same cannot be said for the Pare Mountains (Figures 4–5). Here, no distinct pattern can be discerned from the data.

There are quite a few errors and inconsistencies in the workflow that may be responsible for the lack of pattern in the Pare Mountains, and which may also cause uncertainty in the otherwise clear pattern of Mt. Cameroon. First, the number of samples is quite low in the Pare Mountains. For each slope, only 1–6 samples were collected, predominantly at similar elevations. In addition, the limited size of the remaining forest fragments and extensive land use at lower elevations prevented sampling of anything but the middle elevation samples, rendering the data incomplete at best. From Mt. Cameroon, only the lower elevations below 300 m are lacking, yet these would have also been useful to include, as the lower elevation samples determine the early shape of the patterns in McCain and Grytnes' (2010) classification.

In all cases, sampling heterogeneity may also affect the results. Standardized collection using the same effort at each locality, sampling for the same amount of time, and in similar vegetation is important in order to collect unbiased data (McCain & Grytnes, 2010). Similarly, repeated samples at the same localities increases the quality of the data for each assemblage. In the datasets, sampling effort and completeness may have been suboptimal. Using Chao et al.'s (2014) iNEXT standardization, it is possible to standardize each sample to either coverage or completeness, however this most effectively requires multiple repeated samples and fewer localities for optimal results. However, within each mountain, sampling effort should be reasonably homogenous.

Finally, most of the material used in the analyses was not identified to species level. For the most part, each individual was identified to the level of subfamily, and subsequently sorted to

morphospecies. Given the limited time for the project, only half of the original samples collected from Mt. Cameroon were sorted, and errors in designating specimens to morphospecies likely occurred. Thus, the actual number of morphospecies, as well as the abundance of each morphospecies, may be inaccurate.

4.2 Groupings and abiotic factors in multivariate analysis

In the Mt. Cameroon redundancy analysis, elevation and Shannon diversity had the strongest effect on the data, with elevation correlating well with the first axis (Figure 6). Diversity did not correlate equally well with any of the axes, but is negatively correlated with elevation, as expected. The ordination of sites along the axes reflects these correlations, except for the three sites 121, 106 and 73, which act as outliers. These sites entail some of the middle elevation sites between 1800–2000 m. The sites themselves were not unique in terms of vegetation or sampling effort. As indicated by Figure 3, site 121 and 106 are the most species rich sites in the dataset excluding the lowest elevation sites. They are also the samples including the most abundant species. Looking at Shannon diversity, as in the RDA, site 73 also is more diverse than expected compared to samples at the same elevation. The two samples from site 121 and 106 are particularly similar, sharing one highly abundant species that are only found in the two samples, and one highly abundant species which are shared with only a few other samples, including the sample from site 73. These species could explain why the samples are ordinated by themselves with such high diversity, and why this is not also reflected in the NMDS. However, the result could also be an artifact of heterogeneity in sampling effort or the uncertainty in sorting of the material. Alternatively, these sites could truly be more diverse, following the patterns of a mid-elevation peak in diversity (McCain & Grytnes, 2010), despite the general results pointing toward a monotonic decrease in diversity. Sorting and processing the rest of the collected samples would help increase the resolution of these results.

Looking at sites by slope (Figure 7) does not imply any distinct grouping of sites by their respective slopes, suggesting that sites are similar regardless of absolute distance between them.

The results from the NMDS are similar to the RDA, with elevation and diversity negatively correlated, yet here with the highest elevation samples grouping by themselves (Figure 8).

These are also the sites with the poorest diversity, and interestingly, the samples 136, 137, 138 and 182 are the only samples collected above the tree line. These samples were collected from small patches of trees in the surrounding grassland and in small ravines at higher elevations, effectively isolating the samples from the montane forest belt at lower elevations. Further, site 182 is even more isolated, as it is the only of the four samples from the Bova slope rather than the Bakingili slope. Thus, absolute distance between the samples may explain some of the dissimilarity between the 182 site and the other high-altitude sites, despite the overall small effect of absolute distance in the data in general.

Similarly, in the North Pare, elevation and Simpson diversity were the only significant factors in the redundancy analyses (Figure 9). Here, none of the factors correlated well with the RDA axes, and the factors are not correlated with one another. The grouping of sites was more distinct, with the diversity in the different sites clearly grouped by slope (Figure 10). This grouping may be due to a larger effect of absolute distance and fragmentation between sites.

Higher diversity samples from one area are complemented by lower diversity samples from another at the same elevation. This fits well with the uneven distribution of sites as seen in Figure 4. The close grouping of sites from Kiverenge and Kindoroko FR reflect the small distance between the samples and the continuous forest covering both slopes. The small distance could explain the similar diversity pattern in the sites. The grouping of the Mwanga sample close to the Vumari samples can be explained by the the fact that these samples are from the lowest elevations. The Mwanga and Vumari sites are located far apart from each other, divided by lowland savanna. The Vumari samples are geographically located closer to the South Pare Mountains yet are included in the North Pare data as the samples are biogeographically more similar to this dataset. At these elevations, trees were smaller, the vegetation drier, and the amount of leaf litter much scarcer. Still, in the data, the four samples are more diverse than many of the other samples. Figure 9 shows certain species (red crosses) that do not group with the other species, and are closer to the above-mentioned sites along RDA axis 1. These include one species of *Cordalia*, which is present in most samples, but are much more abundant in the Vumari samples. Similarly, one species of *Diatrechus* is present only in the Mwanga sample and in one of the Vumari samples, with high abundance in both. These are among the species causing the ordination of the lower elevation samples. Still, additional samples from the Mwanga slope would be helpful to further explain the grouping. The NMDS shows a clearer pattern of elevation and diversity being negatively correlated, yet

the sites are less distinctly grouped (Figure 11). It seems the unevenness in sampling design and the distances between slopes erase any sign of a clear pattern in the data.

Finally, in the South Pare, the redundancy analysis found elevation as the only significant factor. The forward selection process does not account for ecological importance, however, and for this reason Shannon diversity was also included as an explanatory variable. Here the factors correlate better with both RDA1 and RDA2 (Figure 12). Seemingly, no correlation can be found between the two factors, as sites of all elevations may be either species rich of species poor. Again, the sites are grouped by slope, with one outlier being sample 153 (Figure 13). This sample is particularly species rich and include highly abundant species. The sampling site is described as disturbed forest with gaps (Appendix A: Table 8). Potentially, the fragmented nature of this site could explain the higher diversity found here. Elevation does not seem to explain the higher diversity in the site, as sites of similar elevation generally are much less diverse. The overall pattern is similar in the NMDS, with no clear grouping of sites or species of note (Figure 14).

The results from the ordinations are, however, on the weaker side, as the unconstrained variance is much higher than the constrained variance for all mountains. Clearly, additional explanatory variables would need to be recorded and analyzed in order to exhaustively explain the variation. These variables could include, e.g., temperature, moisture levels and vegetation. Similarly, models could be employed and tested in order to explain the pattern along the altitudinal gradient, such as the SAR or MDE (McCain & Grytnes, 2010). These would require data like the area of different bands at different elevations in each mountain, or the range of each species, data that were not available given the scope of this thesis.

4.3   Age and fragmentation

The results from the RDA and NMDS analyses can be partially explained by the size, age and fragmentation of the different mountains. In Mt. Cameroon, the negative correlation between elevation and diversity is strong and the pattern is the same no matter the slope. Distance between the sites does not seem to matter the way elevation does. This makes sense, as the mountain massif is young, unfragmented and to a certain degree protected (Déruelle et al., 1987; Ministry of Forestry and Wildlife, 2014; Reusch et al., 2010). The habitats and vegetation are continuous on the whole massif, and thus changing mostly based on abiotic

factors along the altitudinal gradient. The most abrupt change occurs as the forest becomes grassland at around 2300 m, which is well reflected by the NMDS groupings in Figure 8.

The same is not the case in the Pare Mountains. Here, the mountains are older, the elevational range is much narrower, the forest is patchier due to the heavy land use, and the range is less well protected (Bussmann, 2006; Demenocal, 1995). This is reflected in the much less clear correlation between elevation and diversity seen here. Similarly, the sites group well by the different slopes, reflecting the endemism of different species to the different slopes, which may be attributed to fragmentation and larger distances between slopes. Due to a combination of the lower elevations and the higher fragmentation, elevation might not be a determining factor in levels of diversity the same way it is in Mt. Cameroon.

4.4   β-diversity and the effects of altitude and distance

The overall alpha diversity of the three mountains, based on Whittaker's original multiplicative decomposition scheme and using Hill numbers (Chao et al., 2023), finds Mt. Cameroon as the most diverse of the three for any given diversity order (Figure 15). Once again, this fits with the expectations given the larger area and more protected and continuous nature of the mountain.

Beta diversity, on the other hand, is similar in Mt. Cameroon and South Pare, and the two differ extensively from the values found from North Pare. The low values from North Pare may be a result of the small size of the forest patches still left in the range, and the close proximity of the samples collected from each slope.

The analyses of pairwise similarity against absolute distance and elevational difference corroborates the findings from the previous analyses. In the continuous, protected and larger area of Mt. Cameroon, absolute distance between assemblages has little effect on the similarity between assemblages (Figure 17). Again, the difference in altitude between the sites seem to be the main driver for the change in similarity between sites. The pairs of sites with elevational differences of >2000 m have complete turnover with no species in common, and the different slopes have no effect on these results.

In contrast, the fragmented nature of the Pare Mountains makes the effect of absolute distance stronger in these areas, despite a few outliers (Figures 18-19). However, altitude still holds the strongest effect on similarity, and once again, the slope identity does not matter much.

Again, the small number of samples from the Pare Mountains may be of issue in regard to these results. These samples were not collected for the purpose of this analysis and are therefore not the best suited for it. More extensive sampling would certainly improve the quality of the analyses.

4.5 Functional and phylogenetic diversity

In the statistical analyses used in this thesis, the focus has been exclusively on the taxonomic diversity of the communities in question. While this has been the traditional way of looking at diversity distribution, focusing only on taxonomic properties of organisms may limit the quality and robustness of the results. The analyses could have also benefited from utilizing newer methods and advancements focusing on other properties of the organisms. In the current framework of diversity analysis, phylogenetic diversity and functional diversity together with taxonomic diversity represent the dimensions of diversity (Chao et al., 2023). These measures can all be analyzed under the same framework of alpha, beta and gamma diversity. However, given the scope of the current analysis, and the still ongoing development and implementation of a unified computational approach to utilizing these measures, only the taxonomic diversity has been considered here.

4.6 Primer groups and reference sequences

Only 15 of the 24 markers were successfully amplified during the initial testing of the newly designed primers, and only 10 markers were consistently recovered during mapping of sequenced reads to references. Multiple issues were detected as to why this is the case.

First, the new primer pairs themselves were untested and highly degenerate with hopes of fitting a general range of organisms. The primers were redesigned based on ones used in a previous study by Che et al. (2017), and were originally designed based on available genomic Staphylinidae data, none of which were from Proteininae. Therefore, it is likely that some of the primer pairs were not specific enough, and simply did not allow for amplification.

Second, the degeneracy of the primers allowed unfortunate interactions between sequence fragments from one marker and primers made for other markers, prematurely cutting certain sequences short. These interactions depend on the grouping of primer pairs during multiplexing, and on the taxa in question, and are therefore hard to prevent. In the case of *Megarthrus*, *Dhc98D* primers interacted with *FBX011* reads, and both *RPII215* and *Trplm* primers interacted with the *ico* reads, rendering both the genes unusable for downstream analyses. For future consideration, the mentioned primers should not be grouped together when amplifying *Megarthrus* markers.

Finally, no genomes of *Megarthrus* were fully sequenced and available for use as reference during mapping of the sequenced reads. For this reason, the Aleocharinae species *Aleochara bilineata* was used as reference. Using the Aleocharine reference sequences, mapping success was low, at around 40–50%. Creating new reference sequences based on consensus sequences of the successfully mapped *Megarthrus* markers did improve mapping to ~80–95%, yet this method is unfortunate as a lot of trial and error may give room to more inconsistencies and errors. In addition, some markers had many mapped reads, yet not many reads of complete length. For future work on *Megarthrus*, the markers in questions should be Sanger sequenced for reference prior to high-throughput sequencing.

4.7   Phylogenetic implications

Like the early cladistic analyses by Cuccodoro (1998), the Maximum Likelihood analysis (Figure 20) does recover a monophyletic group consisting of *Megarthrus* and *Metopsi*a, but it does not recover a monophyletic *Megarthrus*. Instead, *Metopsia* (Clade **F**) is recovered as a nested monophyletic group within *Megarthrus* (Clade **A**), rendering the current *Megarthrus* paraphyletic. This grouping leaves two options regarding the future of the classification of the genera. Either *Metopsia* should no longer be considered a separate genus, or *M. montanus* and clade **G**, consisting of *M. pictus* and the two unidentified Vietnamese species, should not be treated as members of *Megarthrus* and instead should be placed in two genera of their own.

Morphologically, based on secondary sexual characters, the species of clade **G** and *M. montanus* are similar to the rest of the *Megarthrus* specimens, whereas the *Metopsia* species are significantly different. The *Metopsia* specimens are distinguished from the *Megarthrus* specimens by the presence of a vertexal ocellus, lateral pronotal edges that are not denticulate,

and a less distinct medial pronotal impression. Similarly, the shape of the adeagus is distinctly different from that of the *Megarthrus* (Cuccodoro & Löbl, 1995; Zerche, 1998). This explains, based on morphology alone, why *Metopsia* has not be considered a clade within *Megarthrus*, but as a genus of its own. However, these character traits could all be autapomorphies, which do not act as evidence against a nested *Metopsia* within *Megarthrus*. In addition, the morphological similarity of *M. pictus* of clade **G** and *M. montanus* to the rest of the *Megarthrus* is striking and does not support separate genera for the two groups (Cuccodoro, 2011; Cuccodoro & Löbl, 1996). This, along with the robust support for the *Metopsia* clade within *Megarthrus* strongly supports the options of *Metopsia* as a subgroup of *Megarthrus*.

The current analysis includes only a portion of the known species of *Megarthrus* and *Metopsia* and is based on only 10 genetic markers (Table 2). Similarly, only Maximum Likelihood analyses have been conducted. For this reason, further analyses including additional species are recommended prior to a conclusive revision of the current classification of the group.

Clade **G** may seem unlikely despite its high support, as it consists of one Nearctic species and two Oriental species from Vietnam. However, the clade does make sense as Cuccodoro and Löbl (1996) do note that *Megarthrus pictus* is similar to East Asian species as it is the only Nearctic species possessing angulate temples. It also possesses a bicolorous elytral pattern as the only North American species. This is common in certain East Asian species, like the Japanese *M. corticalis* and *M. scriptus* (Cuccodoro, 1996).

In clade **D**, the morphology of specimen 43180 from the Magadan region in Russia did not match that of any other described species, despite its similarity to the common *M. depressus*. Thus, this species is likely a new underscribed species in the group of otherwise well known Holarctic species (Cuccodoro & Löbl, 1996, 1997).

The species of *Megarthrus* from Madeira and the Canary Islands are not recovered as monophyletic. These results are unsurprising as the Macaronesian islands traditionally have been considered sinks for the mainland European terrestrial biota during glacial periods due to their buffered climate and large size (Florencio et al., 2021). Recent work has shown that these oceanic island systems also contribute more as source for the continental mainland than

previously expected (Jønsson & Holt, 2015), explaining how dispersal to the islands, diversification, and dispersal from the islands may have been possible.

Within the Holarctic and African clade of *Megarthrus* (Clade **B**), the African species are recovered as monophyletic (Clade **C**; Figures 20–21). This is in line with Cuccodoro and Löbl's (1995) morphological classification of the Afrotropical species. They described a homogenous Afrotropical group, listing several synapomorphies, which can be found in their original revision.

The nestedness of the monophyletic Afrotropical clade (Clade **C**) suggests a single recent colonization event from the Palearctic into the Afrotropics without dispersal back to any other region. The short branch lengths among the members of the clade, along with the levels of local endemism within the fragmented montane areas, also suggests a rapid and recent diversification within the genus.

Generally, the Afrotropical *Megarthrus* have been considered poor contemporary dispersers, despite their fully developed wings. The patterns of local endemism, and the montane and temperate habitat of the group with collections predominantly in leaf litter, has been thought to render dispersal across the dryer lowland savanna unlikely (Cuccodoro & Löbl, 1995; Thayer, 2005). Following this, species from the same "sky islands" or within closely related mountains of similar age and size were expected to form monophyletic clades. Previously only one species, *M. wittei*, has been recorded with a transcontinental distribution (Cuccodoro & Löbl, 1995).

Certain subclades within the African clade are restricted to certain regions, following the above assumptions, like the Ethiopian subclade **L**. However, many of the subclades give high support to groupings of species from distantly located mountains. The prime example of this is the placement of the two West-African species from Mt. Cameroon. These do not form a distinguished monophyletic clade; *M. kamerunensis* groups with a species from the Kenyan side of Mt. Elgon (clade **K**), whereas *M. sp. 1* is more closely related to a Tanzanian species from Mt. Kilimanjaro (clade **N**). These placements erase the idea of a distinct transcontinental geographical separation of species.

Similarly, the other predominantly Ethiopian subclade (clade **M**) is not as geographically restricted. Here, species from different Ethiopian mountains form a monophyletic group

together with another species from the distantly located Rubeho Mountains in Tanzania. These clades also exemplify how species are not even monophyletic within each mountain. *Megarthrus. sp. cf. simiensis* from the Simien Mountains is more closely related to species from other Ethiopian mountains than to the rest of the species from the Simien Mountains (clade **M**).

The Afrotropical subclades and the recent diversification of the genus indicate that *Megarthrus* might be better dispersers than previously assumed, especially given the grouping of the West-African species. This could either be due to a better ability to traverse across the drier lowland savanna, or a result of better conditions for dispersal during periods of a wetter climate. As virtually all *Megarthrus* have fully developed wings, periods of warmer and wetter climate, like the ones mentioned by Trauth et al. (2005), could have been enough to create suitable lowland forest bridges for migration between the mountains.

Several subclades in the Afrotropical clade consist of conspecifics with almost no genetic divergence in the phylogeny. These are the specimens of *M. sp. 2 prope africanus* from Mt. Kilimanjaro (in clade **N**), and the two specimens of *M. horticola* from the Kenyan side of Mt. Elgon and the Ugandan side of the Rwenzori Mountains (clade **I**). These conspecifics were included as their *COI* sequences were not identical. The pair of conspecific *M. horticola* specimens are genetically very similar despite being collected from different localities, once again emphasizing their capability to disperse between mountains. However, there are also clades with almost no branch lengths where the species are not conspecific, like the *M. bantu* and *M. sp. 5 prope watutsi* clade (**J**). These specimens were collected from the same localities as *M. horticola*, yet they are clearly differentiated morphologically. These could be examples of rapid local allopatric speciation on different mountains.

Finally, *M. montanus*, known from the Russian Far East, China and Japan, is positioned as the sister taxon to the rest of the *Megarthrus* and *Metopsia* group. This is in accordance with Cuccodoro (1998), which also found *M. montanus* as the sister taxon to the other *Megarthus*. Interestingly, *M. montanus* is one of the few species of *Megarthrus* occasionally possessing reduced sized wings (Cuccodoro & Löbl, 1997).

4.8 *Megarthrus kamerunensis* and *Megarthrus sp. 1*

Every character from the original description by Bernhauer and Paulian (1942) agrees with the neotype of *Megarthrus kamerunensis*. These and additional characters were included in the species redescription. Notably, Bernhauer and Paulian counted abdominal segments from the first visible segments rather than the first morphological segment, and thus described sternite VIII as sternite VI. This becomes clear from their descriptions of some other species with emarginated sternite VIII in the same paper. Here, this description has been changed to fit the counting from morphological segments.

The redescribed *M. kamerunensis* and the new *M. sp. 1,* **sp. nov**. share similar character states with the other Afrotropical species, including the backward orientation and position of setae on the mesal portion of the frons, the supra-ocular sinuate margin, and the general characters of the elytra and pronotum (Cuccodoro & Löbl, 1995). Based on the fast evolving male sexual characters, the two species show morphological similarities to each other and the other two species, *M. wittei* and *M. congoensis*, known from less distant localities. Given the placement of *M. kamerunensis* and *M. sp. 1* in the ML phylogenetic tree, it would bee interesting in future molecular work to also include *M. wittei* and *M. congoensis* in the phylogeny to see how these species would be recovered in the tree.

4.9 State of the genera of Proteininae

Molecular work is scarce on the subfamily of Proteininae. Ideally, more species should have been included in this thesis in order to better resolve the relationships within the group. The same is true for the number of successfully amplified and aligned genetic markers. This should be doable by optimizing the primers and groupings used here. As mentioned, no fully sequenced genomes of Proteininae are currently available at the time of writing. The lack of already sequenced reference sequences from the group did highly affect mapping for the analyses, as the preliminary used sequences of *Aleochara bilineata* were too dissimilar genetically to effectively map the Proteinine reads. Sequencing Proteinine markers by traditional methods for better reference sequences should be prioritized prior to similar analyses on the group in future work. Further molecular analyses should also be conducted for the smaller genera of the subfamily in order to resolve the higher-level relationships of the group.

The genus *Megarthrus* has yet to be extensively explored. The new species *M. sp. 1.* and the redescribed *M. kamerunensis* were not rare or difficult to find during our trip to Mt. Cameroon. This emphasizes the point made by Cuccodoro and Löbl (1995) that the diversity of *Megarthrus* is greater than previously estimated. The limitations in exploring the genus lies not in its depauperate state, but rather in the lack of funding, time and efficient methods for dealing with the shortfalls in biodiversity knowledge (Löbl et al., 2023).

Similarly, many of the species examined in the analyses are likely undescribed, as made evident by the phylogeny (Figure 20). During identification, these unidentified specimens did not match previous descriptions of species from the region completely. The number of undescribed species and the lack of knowledge about them emphasizes the poor knowledge of Afrotropical arthropod biota and how much of it is left undiscovered. Already, the human mediated habitat fragmentation, especially in and around the mountains along the East-African rift systems, have likely led to extensive loss of data regarding the history of the biota in these areas. Other forms of anthropogenic environmental impact may further accelerate the loss of habitats and the extinction of species before we ever get a chance to discover and describe them.

5   Conclusion

In the context of the Hill number framework for analyzing and assessing biodiversity, newly collected staphylinid beetle material from Mt. Cameroon, as well as previously collected material from the Eastern Arc Pare Mountains, Tanzania, has been used to analyze and compare rove beetle diversity along the elevational gradient in selected mountains of East and West Africa. The analysis found a monotonically decreasing pattern of diversity with increasing elevation in Mt. Cameroon, whereas the quality of the Pare Mountain dataset made inferring such a pattern from these mountains impractical.

Redundancy Analyses (RDA) and Non-Metric Multidimensional Scaling (NMDS) ordinations reveal grouping of samples by the slope they were collected from as important for explaining the variation in diversity in the Pare Mountains, whereas this is unimportant in Mt. Cameroon. The explanatory variables of elevation and diversity were included in the constrained ordinations, revealing a strong correlation between the two in Mt. Cameroon, whereas the variables were only weakly or not at all correlated in the Pare Mountains. However, the constrained axes explained only a small proportion of the variation in the data. Including additional explanatory variables would improve the strength of the analyses.

Similarly, plots of pairwise similarity measures against the difference in elevation between sites and the distance between sites reveal elevation as the most important determinator of the beta diversity patterns in the mountains, and distance between sites as important only in the Pare Mountains. These results reflect the difference in properties between the East- and West-African mountains. On the young, protected and unfragmented Mt. Cameroon, the forest belt is continuous and distances between sites does not isolate assemblages as much as it does in the older, lower, and more fragmented Pare Mountains. Here the distances between the fragmented forest patches have a higher influence on the dissimilarity of assemblages.

From a dataset consisting of 10 nuclear single copy protein-coding genes, the phylogenetic relationships of 47 Proteininae species have here been inferred. The phylogeny recovered *Megarthrus* as paraphyletic, as the included species of the genus *Metopsia* were recovered as a subclade nested within the current *Megarthrus*. The 23 included Afrotropical specimens of *Megarthrus* form a well supported clade. Within that clade, some subclades are geographically restricted, whereas others consist of species from distantly located mountains.

The latter subclades suggest that *Megarthrus* might be better dispersers than previously expected. However, only a portion of the known species of *Megarthrus* have been included in the analyses, and further analyses including additional specimens is recommended prior to revision of the classification of the group.

Finally, the new species, *Megarthrus sp. 1,* **sp. nov**. has been described, and *M. kamerunensis* has been redescribed.

References

Ako, A., Shimada, J., Hosono, T., Kagabu, M., Akoachere, R., Nkeng, G., Takem, G., & Fouepe, A. (2012). Spring water quality and usability in the Mount Cameroon area revealed by hydrogeochemistry. *Environmental geochemistry and health*, *34*(5), 615–639. https://doi.org/10.1007/s10653-012-9453-3

Berger, W. H., & Parker, F. L. (1970). Diversity of Planktonic Foraminifera in Deep-Sea Sediments. *Science*, *168*(3937), 1345–1347. https://doi.org/10.1126/science.168.3937.1345

Bernhauer, M., & Paulian, R. (1942). Coléoptères Staphylinides du Cameroun (Mission P. Lepesme, R. Paulian et A. Villier, 1939). *Revue Zool. Bot. Afr.*, *35*, 344–375.

Brito, D. (2010). Overcoming the Linnean shortfall: Data deficiency and biological survey priorities. *Basic and Applied Ecology*, *11*(8), 709–713. https://doi.org/10.1016/j.baae.2010.09.007

Brown, J. H., & Lomolino, M. V. (1998). *Biogeography*. Sinuer Associates Publishers.

Brühl, C. (1997). Flightless insects: a test case for historical relationships of African mountains. *Journal of Biogeography*, *24*(2), 233–250. https://doi.org/10.1046/j.1365-2699.1997.00073.x

Bussmann, R. (2006). Vegetation zonation and nomenclature of African Mountains - An overview. *Lyonia*, *11*(1), 41–66. https://www.lyonia.org/viewarticle-476?articleID=476

Cai, C., Newton, A. F., Thayer, M. K., Leschen, R. A. B., & Huang, D. (2016). Specialized proteinine rove beetles shed light on insect–fungal associations in the Cretaceous. *Proc. R. Soc. B.*, *283*(1845), 20161439. https://doi.org/10.1098/rspb.2016.1439

Cai, C., Tihelka, E., Giacomelli, M., Lawrence, J. F., Ślipiński, A., Kundrata, R., Yamamoto, S., Thayer, M. K., Newton, A. F., Leschen, R. A. B., Gimmel, M. L., Lü, L., Engel, M. S., Bouchard, P., Huang, D., Pisani, D., & Donoghue, P. C. J. (2022). Integrated phylogenomics and fossil data illuminate the evolution of beetles. *R. Soc. Open Sci.*, *9*(3), 211771. https://doi.org/10.1098/rsos.211771

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc Natl Acad Sci U S A*, *114*(30), E6089–E6096. https://doi.org/10.1073/pnas.1704949114

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, *11*(4), 265–270. http://www.jstor.org/stable/4615964

Chao, A., Chiu, C.-H., Wu, S.-H., Huang, C.-L., & Lin, Y.-C. (2019). Comparing two classes of alpha diversities and their corresponding beta and (dis)similarity measures, with an application to the Formosan sika deer Cervus nippon taiouanus reintroduction

programme. *Methods in Ecology and Evolution*, *10*(8), 1286–1297. https://doi.org/10.1111/2041-210X.13233

Chao, A., & Chiu, C. H. (2016). Nonparametric Estimation and Comparison of Species Richness. In John Wiley & Sons (Ed.), *eLS* (pp. 1–11). https://doi.org/10.1002/9780470015902.a0026329

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*(1), 45–67. https://doi.org/10.1890/13-0133.1

Chao, A., & Hu, K.-H. (2023). iNEXT.beta3D: interpolation and extrapolation with beta diversity for three dimensions of biodiversity. R package version 1.0.1. http://chao.stat.nthu.edu.tw/wordpress/software_download/inext-beta3d/

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, *93*(12), 2533–2547. https://doi.org/10.1890/11-1952.1

Chao, A., & Jost, L. (2015). Estimating diversity and entropy profiles via discovery rates of new species. *Methods in Ecology and Evolution*, *6*(8), 873–882. https://doi.org/10.1111/2041-210X.12349

Chao, A., Ma, K. H., Hsieh, T. C., & Chiu, C.-H. (2015). Online Program SpadeR (Species-richness Prediction And Diversity Estimation in R). Program and User's Guide published at http://chao.stat.nthu.edu.tw/wordpress/software_download/.

Chao, A., Thorn, S., Chiu, C.-H., Moyes, F., Hu, K.-H., Chazdon, R. L., Wu, J., Magnago, L. F. S., Dornelas, M., Zelený, D., Colwell, R. K., & Magurran, A. E. (2023). Rarefaction and extrapolation with beta diversity under a framework of Hill numbers: The iNEXT.beta3D standardization. *Ecological Monographs*, *93*(4), Article e1588. https://doi.org/10.1002/ecm.1588

Che, L. H., Zhang, S. Q., Li, Y., Liang, D., Pang, H., Ślipiński, A., & Zhang, P. (2017). Genome-wide survey of nuclear protein-coding markers for beetle phylogenetics and their application in resolving both deep and shallow-level divergences. *Mol Ecol Resour*, *17*(6), 1342–1358. https://doi.org/10.1111/1755-0998.12664

Chiu, C.-H., Jost, L., & Chao, A. (2014). Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs*, *84*(1), 21–44. https://doi.org/10.1890/12-0960.1

Chiu, C.-H., Wang, Y.-T., Walther, B. A., & Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified good–turing frequency formula. *Biometrics*, *70*(3), 671–682. https://doi.org/10.1111/biom.12200

Clary, D. O., & Wolstenholme, D. R. (1985). The mitochondrial DNA molecular of Drosophila yakuba: nucleotide sequence, gene organization, and genetic code. *J Mol Evol*, *22*(3), 252–271. https://doi.org/10.1007/bf02099755

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S.-Y., Mao, C. X., Chazdon, R. L., & Longino, J. T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology*, *5*(1), 3–21. https://doi.org/10.1093/jpe/rtr044

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can We Name Earth's Species Before They Go Extinct? *Science*, *339*(6118), 413–416. https://doi.org/10.1126/science.1230318

Cuccodoro, G. (1995). Two new species of Megarthrus (Coleoptera, Staphylinidae, Proteininae) and a note on 'water loading' behaviour. *Journal of Zoology*, *236*(2), 253–264. https://doi.org/10.1111/j.1469-7998.1995.tb04492.x

Cuccodoro, G. (1996). Revision of the Japanese rove-beetles of the genus *Megarthrus* (Coleoptera, Staphylinidae, Proteininae). *Revue Suisse de Zoologie*, *103*(2), 475–524.

Cuccodoro, G. (1998). Revision and phylogeny of Megarthrus Curtis 1829 from New Guinea, New Caledonia and Fiji (Coleoptera Staphylinidae Proteininae). *Tropical Zoology*, *11*(1), 103–137. https://doi.org/10.1080/03946975.1998.10539356

Cuccodoro, G. (2011). *Megarthrus* of Taiwan, with notes on phylogenetic relationships within the genus (Coleoptera: Staphylinidae: Proteininae). *Studies and reports*, *taxonomical Series 7*(1–2), 25–92.

Cuccodoro, G., & Löbl, I. (1995). Revision of the Afrotropical rove-beetles of the genus *Megarthrus* (Coleoptera, Staphylinidae, Proteininae). *Revue Suisse de Zoologie*, *102*(3), 655–761.

Cuccodoro, G., & Löbl, I. (1996). Revision of the rove-beetles of the genus *Megarthrus* of America north of Mexico (Coleoptera, Staphylinidae, Proteininae). *Mitteilungen der Münchner Entomologischen Gesellschaft*, *86*, 145–188.

Cuccodoro, G., & Löbl, I. (1997). Revision of the Palearctic rove beetles of the genus *Megarthrus* Curtis (Coleoptera: Staphylinidae: Proteininae). *Journal of Natural History*, *31*, 1347–1415.

Demenocal, P. (1995). Plio-Pleistocene African Climate. *Science*, *270*, 53–59. https://doi.org/10.1126/science.270.5233.53

Déruelle, B., N'ni, J., & Kambou, R. (1987). Mount Cameroon: an active volcano of the Cameroon Line. *Journal of African Earth Sciences (1983)*, *6*(2), 197–214. https://doi.org/10.1016/0899-5362(87)90061-3

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Elven, H. (2013). *Molecular phylogeny of the rove beetle tribes Athetini, Geostibini, and Lomechusini (Coleoptera: Staphylinida: Aleocharinae)* [Doctoral dissertation, University of Oslo].

Florencio, M., Patiño, J., Nogué, S., Traveset, A., Borges, P. A. V., Schaefer, H., Amorim, I. R., Arnedo, M., Ávila, S. P., Cardoso, P., de Nascimento, L., Fernández-Palacios, J. M., Gabriel, S. I., Gil, A., Gonçalves, V., Haroun, R., Illera, J. C., López-Darias, M., Martínez, A., . . . Santos, A. M. C. (2021). Macaronesia as a Fruitful Arena for Ecology, Evolution, and Conservation Biology. *Front. Ecol. Evol.*, *9*. https://doi.org/10.3389/fevo.2021.718169

Folmer, O., Black, M., Wr, H., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial Cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, *3*(5), 294–299.

Grebennikov, V. V. (2021). Sky islands of the Cameroon Volcanic Line support the westernmost clade of five new Typoderus weevils (Coleoptera: Curculionidae: Molytinae). *Arthropod Systematics & Phylogeny*, *79*, 57–74. https://doi.org/10.3897/asp.79.e66021

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hill, M. O. (1973). Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Horn, H. S. (1966). Measurement of "overlap" in comparative ecological studies. *American Naturalist*, *100*(914), 419–424. https://doi.org/10.1086/282436

Hsieh, T. C., Ma, K. H., & Chao, A. (2022). iNEXT: INterpolation and EXtrapolation for species diversity. R package version 3.0.0. http://chao.stat.nthu.edu.tw/wordpress/software-download/

ICZN. (1999). *International Code of Zoological Nomenclature* (4 ed.). The International Trust for Zoological Nomenclature.

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (E. S. Brondizio, J. Settele, S. Díaz, & H. T. Ngo, Eds.). IPBES secretariat. https://doi.org/10.5281/zenodo.6417333

Jaccard, P. (1901). Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles*, *37*, 241–272. https://doi.org/10.5169/seals-266440

Jost, L. (2007). PARTITIONING DIVERSITY INTO INDEPENDENT ALPHA AND BETA COMPONENTS. *Ecology*, *88*(10), 2427–2439. https://doi.org/10.1890/06-1736.1

Jost, L., Chao, A., & Chazdon, R. L. (2010). Compositional similarity and β (beta) diversity. In A. E. Magurran (Ed.), *Biological Diversity: Frontiers in Measurement and Assessment* (pp. 66–84). Oxford University Press.

Jønsson, K. A., & Holt, B. G. (2015). Islands contribute disproportionately high amounts of evolutionary diversity in passerine birds. *Nat Commun*, *6*, Article 8538. https://doi.org/10.1038/ncomms9538

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kassambara, A. (2023). ggpubr: "ggplot2" Based Publication Ready Plots. R package version 0.6.0. https://CRAN.R-project.org/package=ggpubr

Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology*, *72*(3), 367–382. https://doi.org/10.1046/j.1365-2656.2003.00710.x

Korner, C. (2007). The use of 'altitude' in ecological research. *Trends Ecol Evol*, *22*(11), 569–574. https://doi.org/10.1016/j.tree.2007.09.006

Li, D. (2018). hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *Journal of Open Source Software*, *3*(31), Article 1041. https://doi.org/10.21105/joss.01041

Lomolino, M. V. (2004). Conservation biogeography. In M. V. Lomolino & L. R. Heaney (Eds.), *Frontiers of Biogeography: New Directions in the Geography of Nature* (pp. 293–296). Sinauer Associates.

Löbl, I., Klausnitzer, B., Hartmann, M., & Krell, F.-T. (2023). The Silent Extinction of Species and Taxonomists—An Appeal to Science Policymakers and Legislators. *Diversity*, *15*(10). https://doi.org/10.3390/d15101053

Magurran, A. E. (2004). *Measuring Biological Diversity*. Blackwell Science Ltd.

Magurran, A. E. (2021). Measuring biological diversity. *Current Biology*, *31*(19), R1174–R1177. https://doi.org/10.1016/j.cub.2021.07.049

Magurran, A. E., & McGill, B. J. (2010). *Biological Diversity: Frontiers in Measurement and Assessment*. OUP Oxford.

McCain, C. M., & Grytnes, J.-A. (2010). Elevational gradients in species richness. In *eLS*. https://doi.org/10.1002/9780470015902.a0022548

Ministry of Forestry and Wildlife. (2014). *The management plan of the Mount Cameroon National Park and its peripheral zone 2015-2019*. Retrieved from https://documents1.worldbank.org/curated/en/290571468007177234/pdf/E23260V30REPLA00Box391426B00PUBLIC0.pdf

Newton, A., & Thayer, M. (1995). Protopselaphinae new subfamily for Protopselaphus new genus from Malaysia, with a phylogenetic analysis and review of the Omaliine Group of Staphylinidae including Pselaphidae. In J. Pakaluk & S. A. Slipinski (Eds.), *Biology, Phylogeny, and Classification of Coleoptera: Papers Celebrating the 80th Birthday of Roy A. Crowson* (Vol. 2, pp. 219–230). Muzeum i Instytut Zoologii PAN.

Newton, A. F. (2022). StaphBase. In O. Bánki, Y. Roskov, M. Döring, G. Ower, D. R. Hernández Robles, C. A. Plata Corredor, T. Stjernegaard Jeppesen, A. Örn, L. Vandepitte, D. Hobern, P. Schalk, R. E. DeWalt, K. Ma, J. Miller, & T. Orrell (Eds.), *Catalogue of Life Checklist*. https://doi.org/10.48580/dfqf-3gk

Newton, A. F., Thayer, M. K., Ashe, J. S., & Chandler, D. S. (2000). Staphylinidae Latreille, 1802. In R. H. Arnett & M. C. Thomas (Eds.), *American Beetles, vol. 1, Archostemata, Myxophaga, Adephaga, Polyphaga: Staphyliniformia* (Vol. 1, pp. 272–418). CRC Press, LLC.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Novotny, V., Basset, Y., Miller, S. E., Weiblen, G. D., Bremer, B., Cizek, L., & Drozd, P. (2002). Low host specificity of herbivorous insects in a tropical forest. *Nature*, *416*(6883), 841–844. https://doi.org/10.1038/416841a

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., . . . Weedon, J. (2022). vegan: Community Ecology Package. R package version 2.6-4. https://CRAN.R-project.org/package=vegan

Payton, R. W. (1993). *Ecology, altitudinal zonation and conservation of tropical rainforests of Mount Cameroon*.

Pimm, S. L., Russell, G. J., Gittleman, J. L., & Brooks, T. M. (1995). The Future of Biodiversity. *Science*, *269*(5222), 347–350. https://doi.org/10.1126/science.269.5222.347

R Core Team. (2023). *R: A Language and Environment for Statistical Computing.* In R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rahbek, C. (1995). The elevational gradient of species richness: a uniform pattern? *Ecography*, *18*(2), 200–205. https://doi.org/10.1111/j.1600-0587.1995.tb00341.x

Ratnasingham, S., & Hebert, P. D. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Reusch, A. M., Nyblade, A. A., Wiens, D. A., Shore, P. J., Ateba, B., Tabod, C. T., & Nnange, J. M. (2010). Upper mantle structure beneath Cameroon from body wave

tomography and the origin of the Cameroon Volcanic Line. *Geochemistry, Geophysics, Geosystems*, *11*(10). https://doi.org/10.1029/2010gc003200

Routledge, R. D. (1979). Diversity indices: Which ones are admissible? *Journal of Theoretical Biology*, *76*(4), 503–515. https://doi.org/10.1016/0022-5193(79)90015-8

Sanders, N. J., & Rahbek, C. (2012). The patterns and causes of elevational diversity gradients. *Ecography*, *35*(1), 1–3. https://doi.org/10.1111/j.1600-0587.2011.07338.x

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene Sequences and a Compilation of Conserved Polymerase Chain Reaction Primers. *Annals of the Entomological Society of America*, *87*(6), 651–701. https://doi.org/10.1093/aesa/87.6.651

Simpson, E. H. (1949). Measurement of Diversity. *Nature*, *163*(4148), 688–688. https://doi.org/10.1038/163688a0

Sweetlove, L. (2011). Number of species on Earth tagged at 8.7 million. *Nature*. https://doi.org/10.1038/news.2011.498

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter. Kongelige Danske Videnskabernes Selskab*, *5*, 1–34.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Thayer, M. (2005). 11. Staphylinoidea. 11.7. Staphylinidae Latreille, 1802. In *Coleoptera, Beetles Volume 1: Morphology and Systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim). Handbook of Zoology, Vol. IV. Part 38.* (pp. 296–344). Walter de Gruyter.

Trauth, M., Maslin, M., Deino, A., & Strecker, M. (2005). Late Cenozoic Moisture History of East Africa. *Science*, *309*(5743), 2051–2053. https://doi.org/10.1126/science.1112964

Vavrek, M. (2011). Fossil: Palaeoecological and Palaeogeographical Analysis Tools. *Palaeontologia Electronica*, *14*(1), 16. https://palaeo-electronica.org/2011_1/238/index.html

Wasser, S. K., & Lovett, J. C. (1993). Introduction to the biogeography and ecology of the rain forests of eastern Africa. In J. C. Lovett & S. K. Wasser (Eds.), *Biogeography and Ecology of the Rain Forests of Eastern Africa* (pp. 3–8). Cambridge University Press. https://doi.org/10.1017/CBO9780511895692.001

Wembenyui, E. W., Collerson, K. D., & Zhao, J.-X. (2020). Evolution of Mount Cameroon volcanism: Geochemistry, mineral chemistry and radiogenic isotopes (Pb, Sr, Nd). *Geoscience Frontiers*, *11*(6), 2157–2168. https://doi.org/10.1016/j.gsf.2020.03.015

Whittaker, R. H. (1972). EVOLUTION AND MEASUREMENT OF SPECIES DIVERSITY. *TAXON*, *21*(2–3), 213–251. https://doi.org/10.2307/1218190

Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M., & Willis, K. J. (2005). Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, *11*(1), 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

WWF. (2022). *Living Planet Report 2022 - Building a nature-positive society* (R. E. A. Almond, M. Grooten, D. Juffe Bignoli, & T. Petersen, Eds.). WWF.

Zerche, L. (1998). Phylogenetisch-systematische Revision der westpaläarktischen Gattung *Metopsia* Wollaston, 1854. *Beiträge zur Entomologie*, *48*(1), 3–101.

Appendix A: Sample Information

Table 6. All sample information on samples collected on Mt. Cameroon early 2023. Sample IDs in bold were included in the statistical analysis.

Sample ID	Sample Description
23cmH1	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.285'N 9°3.039'E WGS84, h=346m, accuracy 9m, extent 20m, lowland forest, ravine, rocky area, dry, sifting [23cmH1] [Garmin 60CSx] H.Knudsen 24.ii.2023
23cmH2	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.424'N 9°3.115'E WGS84, h=405m, accuracy 10m, extent 30m, lowland forest with herbaceous gaps, flat, dry, sifting [23cmH2] [Garmin 60CSx] H.Knudsen 24.ii.2023
23cmH3	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.486'N 9°3.202'E WGS84, h=438m, accuracy 10m, extent 15m, lowland forest, on the slope, dry, sifting [23cmH3] [Garmin 60CSx] H.Knudsen 24.ii.2023
23cmH4	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.630'N 9°3.391'E WGS84, h=510m, accuracy 13m, extent 10m, lowland forest with gaps, flat, dry, small sample, sifting [23cmH4] [Garmin 60CSx] H.Knudsen 25.ii.2023
23cmH5	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.645'N 9°3.430'E WGS84, h=525m, accuracy 10m, extent 20m, lowland forest with gaps, flat, large trees, dry, small sample sifting [23cmH5] [Garmin 60CSx] H.Knudsen 25.ii.2023
23cmH6	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.711'N 9°3.830'E WGS84, h=601m, accuracy 90m, extent 10m, lowland forest, on slope, large trees, wet, small sample, sifting [23cmH6] [Garmin 60CSx] H.Knudsen 25.ii.2023
23cmH7	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.530'N 9°3.229'E WGS84, h=462m, accuracy 15m, extent 20m, lowland forest, on slope, wet, hillside, sifting [23cmH7] [Garmin 60CSx] H.Knudsen 27.ii.2023
23cmH8	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.555'N 9°3.260'E WGS84, h=480m, accuracy 25m, extent 15m, lowland forest with gaps, wet, flat, sifting [23cmH8] [Garmin 60CSx] H.Knudsen 27.ii.2023
23cmH9	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.973'N 9°3.785'E WGS84, h=654m, accuracy 24m, extent 15m, lowland forest, on slope, dry, sifting [23cmH9] [Garmin 60CSx] H.Knudsen 27.ii.2023
23cmH10	CAMEROON, Mt.Cameroon Nat. Park, trail btw Bamboo & Drink Harry Camps, 4°5.925'N 9°3.757'E WGS84, h=641m, accuracy 13m, extent 10m, lowland forest, valley under canopy, dry, sifting [23cmH10] [Garmin 60CSx] H.Knudsen 27.ii.2023
23cm45	CAMEROON, Mt.Cameroon Nat. Park, trail btw Drink Harry & Plantecam Camps, 4°6.925'N 9°4.161'E WGS84, h=1018m, accuracy 5m, extent 30m, montane forest with herbaceous gaps, large to small trees, dry, sifting [23Cm45] [Garmin 62CSx] V.I.Gusarov & P.A.Mballa Ndzie 1.iii.2023
23cm46	CAMEROON, Mt.Cameroon Nat. Park, trail btw Drink Harry & Plantecam Camps, 4°6.810'N 9°4.108'E WGS84, h=961m, accuracy 7m, extent 30m, montane forest with

small gaps, dry, sifting [23Cm46] [Garmin 62CSx] V.I.Gusarov & P.A.Mballa Ndzie 1.iii.2023

23cm47	CAMEROON, Mt.Cameroon Nat. Park, trail btw Drink Harry & Plantecam Camps, 4°6.700'N 9°4.012'E WGS84, h=908m, accuracy 6m, extent 30m, montane forest, large to small trees, dry, sifting [23Cm47] [Garmin 62CSx] V.I.Gusarov & P.A.Mballa Ndzie 1.iii.2023
23cm48	CAMEROON, Mt.Cameroon Nat. Park, trail btw Drink Harry & Plantecam Camps, 4°6.605'N 9°3.932'E WGS84, h=849m, accuracy 11m, extent 30m, montane forest, dry, sifting [23Cm48] [Garmin 62CSx] V.I.Gusarov & P.A.Mballa Ndzie 1.iii.2023
23cm49	CAMEROON, Mt.Cameroon Nat. Park, trail btw Drink Harry & Plantecam Camps, 4°6.493'N 9°3.827'E WGS84, h=784m, accuracy 7m, extent 30m, montane forest, wet, sifting [23Cm49] [Garmin 62CSx] V.I.Gusarov & P.A.Mballa Ndzie 1.iii.2023
23cm50	CAMEROON, Mt.Cameroon Nat. Park, trail btw Drink Harry & Plantecam Camps, 4°6.341'N 9°3.729'E WGS84, h=746m, accuracy 6m, extent 30m, montane forest, wet, sifting [23Cm50] [Garmin 62CSx] V.I.Gusarov & P.A.Mballa Ndzie 1.iii.2023
23cm72	CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring and Crater Lake Camps, 4°8.623'N 9°5.977'E WGS84, h=1958m, accuracy 4m, extent 60m, montane forest, large trees with gaps with herbaceous vegetation & bushes, sifting [23Cm72] [Garmin 62CSx] V.I.Gusarov 5.iii.2023
23cm73	CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring and Crater Lake Camps, 4°8.985'N 9°5.576'E WGS84, h=1941m, accuracy 7m, extent 30m, montane forest, large trees with gaps with herbaceous vegetation & bushes, sifting [23Cm73] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho 5.iii.2023
23cm77	CAMEROON, Mt.Cameroon Nat. Park, trail nr Plantecam Camp, 4°7.050'N 9°4.303'E WGS84, h=1091m, accuracy 6m, extent 40m, montane forest with dense understory, dry, sifting [23Cm77] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho 6.iii.2023
23cm80	CAMEROON, Mt.Cameroon Nat. Park, trail btw Plantecam & Crater Lake Camps, 4°7.119'N 9°4.362'E WGS84, h=1117m, accuracy 7m, extent 25m, montane forest with fermenting figs on ground, sifting [23Cm80] [Garmin 62CSx] V.I.Gusarov & team 6.iii.2023
23cm82	CAMEROON, Mt.Cameroon Nat. Park, trail btw Plantecam & Crater Lake Camps, 4°7.550'N 9°4.310'E WGS84, h=1200m, accuracy 6m, extent 25m, montane forest / herbaceous patches, sifting [23Cm82] [Garmin 62CSx] V.I.Gusarov & team 6.iii.2023
23cm83	CAMEROON, Mt.Cameroon Nat. Park, trail btw Plantecam & Crater Lake Camps, 4°8.086'N 9°4.163'E WGS84, h=1303m, accuracy 6m, extent 25m, montane forest / herbaceous patches, sifting [23Cm83] [Garmin 62CSx] V.I.Gusarov & team 6.iii.2023
23cm84	CAMEROON, Mt.Cameroon Nat. Park, trail btw Plantecam & Crater Lake Camps, 4°8.340'N 9°4.340'E WGS84, h=1422m, accuracy 3m, extent 25m, montane forest / herbaceous patches, sifting [23Cm84] [Garmin 62CSx] V.I.Gusarov & team 6.iii.2023
23cm85	CAMEROON, Mt.Cameroon Nat. Park, trail btw Plantecam & Crater Lake Camps, 4°8.412'N 9°4.541'E WGS84, h=1505m, accuracy 6m, extent 30m, montane forest / herbaceous patches, sifting [23Cm85] [Garmin 62CSx] V.I.Gusarov & team 6.iii.2023
23cm86	CAMEROON, Mt.Cameroon Nat. Park, trail btw Plantecam & Crater Lake Camps, 4°8.401'N 9°4.780'E WGS84, h=1621m, accuracy 10m, extent 30m, montane forest / herbaceous patches, sifting [23Cm86] [Garmin 62CSx] V.I.Gusarov & team 6.iii.2023

23cm93   CAMEROON, Mt.Cameroon Nat. Park, env. of Crater Lake Camp, 4°8.427'N 9°4.881'E WGS84, h=1639m, accuracy 13m, extent 25m, montane forest, huge trees with herbaceous gaps, sifting [23Cm93] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho 7.iii.2023

23cm105   CAMEROON, Mt.Cameroon Nat. Park, trail btw Crater Lake & Mann's Spring Camps, 4°8.610'N 9°4.971'E WGS84, h=1677m, accuracy 4m, extent 40m, montane forest, patch of large trees, figs on ground, sifting [23Cm105] [Garmin 62CSx] V.I.Gusarov & team 8.iii.2023

23cm106   CAMEROON, Mt.Cameroon Nat. Park, trail btw Crater Lake & Mann's Spring Camps, 4°8.817'N 9°5.129'E WGS84, h=1810m, accuracy 7m, extent 25m, montane forest / herbaceous patches, sifting [23Cm106] [Garmin 62CSx] V.I.Gusarov & team   8.iii.2023

23cm112   CAMEROON, Mt.Cameroon Nat. Park, env. of Mann's Spring Camp, 4°8.641'N 9°7.238'E WGS84, h=2254m, accuracy 6m, extent 20m, montane forest, sifting [23Cm112] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   9.iii.2023

23cm113   CAMEROON, Mt.Cameroon Nat. Park, env. of Mann's Spring Camp, 4°8.564'N 9°7.261'E WGS84, h=2245m, accuracy 9m, extent 30m, montane forest, large to small trees, sifting [23Cm113] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   9.iii.2023

23cm116   CAMEROON, Mt.Cameroon Nat. Park, env. of Mann's Spring Camp, 4°8.861'N 9°7.358'E WGS84, h=2362m, accuracy 6m, extent 30m, montane forest, gallery forest in broad ravine, sifting [23Cm116] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   9.iii.2023

23cm118   CAMEROON, Mt.Cameroon Nat. Park, trail btw Crater Lake & Mann's Spring Camps, 4°8.499'N 9°7.015'E WGS84, h=2188m, accuracy 7m, extent 25m, montane forest, large, medium & small trees, 50-100 m to forest edge, sifting [23Cm118] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   10.iii.2023

23cm120   CAMEROON, Mt.Cameroon Nat. Park, trail btw Crater Lake & Mann's Spring Camps, 4°8.389'N 9°6.971'E WGS84, h=2117m, accuracy 10m, extent 40m, montane forest, large, medium & small trees, sifting [23Cm120] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   10.iii.2023

23cm121   CAMEROON, Mt.Cameroon Nat. Park, trail btw Crater Lake & Mann's Spring Camps, 4°8.352'N 9°6.614'E WGS84, h=2022m, accuracy 10m, extent 30m, montane forest, large, medium & small trees, shallow ravine with herbaceous vegetation, sifting [23Cm121] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   10.iii.2023

23cm134   CAMEROON, Mt.Cameroon Nat. Park, env. of Mann's Spring Camp, 4°9.013'N 9°7.203'E WGS84, h=2386m, accuracy 7m, extent 30m, montane forest in broad ravine, large, medium & small trees, Poaceae in understory, sifting [23Cm134] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   11.iii.2023

23cm135   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Camp One, 4°8.912'N 9°7.465'E WGS84, h=2422m, accuracy 4m, extent 20m, montane forest patch in small ravine surrounded by treeless grassland, Asplenium, sifting [23Cm135] [Garmin 62CSx] V.I.Gusarov & team   12.iii.2023

23cm136   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Camp One, 4°9.057'N 9°7.885'E WGS84, h=2566m, accuracy 4m, extent 30m, scattered patches of Hypericum trees surrounded by burnt patches, sifting [23Cm136] [Garmin 62CSx] V.I.Gusarov & team   12.iii.2023

23cm137   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Camp One, 4°9.319'N 9°8.173'E WGS84, h=2712m, accuracy 4m, extent 15m, patch of Hypericum

trees surrounded by treeless grassland, Pteridium, moss, sifting [23Cm137] [Garmin 62CSx]
V.I.Gusarov & team   12.iii.2023

23cm138   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Camp One,
4°9.613'N 9°8.465'E WGS84, h=2866m, accuracy 3m, extent 15m, small wet ravine with
Hypericum trees surrounded by treeless grassland, ferns, moss, sifting [23Cm138] [Garmin
62CSx] V.I.Gusarov & team   12.iii.2023

23cm148   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.497'N 9°8.770'E WGS84, h=2003m, accuracy 7m, extent 40m, montane forest,
large to small trees, 50-100 m to forest edge, sifting [23Cm148] [Garmin 62CSx]
V.I.Gusarov & team   13.iii.2023

23cm149   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.525'N 9°9.043'E WGS84, h=1901m, accuracy 7m, extent 25m, montane forest
with small gaps, ferns, Poaceae, sifting [23Cm149] [Garmin 62CSx] V.I.Gusarov & team
13.iii.2023

23cm150   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.531'N 9°9.415'E WGS84, h=1800m, accuracy 10m, extent 40m, montane
forest, small to medium trees, old lava flow, Dracaena-like plants, sifting [23Cm150]
[Garmin 62CSx] V.I.Gusarov & team   13.iii.2023

23cm151   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.505'N 9°9.849'E WGS84, h=1714m, accuracy 7m, extent 25m, montane forest,
small to large trees, ferns, herbaceous vegetation, sifting [23Cm151] [Garmin 62CSx]
V.I.Gusarov & team   13.iii.2023

23cm152   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.504'N 9°10.129'E WGS84, h=1594m, accuracy 9m, extent 30m, montane
forest, small to large trees, herbaceous vegetation, sifting [23Cm152] [Garmin 62CSx]
V.I.Gusarov & team   13.iii.2023

23cm153   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.432'N 9°10.306'E WGS84, h=1495m, accuracy 9m, extent 30m, montane
forest, small to large trees, herbaceous vegetation, sifting [23Cm153] [Garmin 62CSx]
V.I.Gusarov & team   13.iii.2023

23cm154   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.421'N 9°10.594'E WGS84, h=1393m, accuracy 6m, extent 25m, montane
forest, small to large trees, herbaceous vegetation, sifting [23Cm154] [Garmin 62CSx]
V.I.Gusarov & team   13.iii.2023

23cm156   CAMEROON, Mt.Cameroon Nat. Park, trail btw Mann's Spring Camp & Bokwango vill.
(Buea), 4°7.441'N 9°11.005'E WGS84, h=1286m, accuracy 10m, extent 25m, montane
forest patch nr agricultural zone, sifting [23Cm156] [Garmin 62CSx] V.I.Gusarov & team
13.iii.2023

23cm159   CAMEROON, Mt.Cameroon Nat. Park, below P&T Camp, 4°17.388'N 9°12.546'E
WGS84, h=2301m, accuracy 5m, extent 30m, montane forest, large trees, gaps with bushes,
ferns and Poaceae, sifting [23Cm159] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-
Nkongho   14.iii.2023

23cm161   CAMEROON, Mt.Cameroon Nat. Park, below P&T Camp, 4°17.438'N 9°12.446'E
WGS84, h=2234m, accuracy 13m, extent 30m, montane forest, large trees, gaps with
bushes, ferns and Poaceae, sifting [23Cm161] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-
Nkongho   14.iii.2023

23cm165 CAMEROON, Mt.Cameroon Nat. Park, below P&T Camp, 4°17.480'N 9°12.341'E WGS84, h=2164m, accuracy 11m, extent 50m, montane forest, large trees, gaps with bushes, ferns and Poaceae, sifting [23Cm165] [Garmin 62CSx] V.I.Gusarov & team 14.iii.2023

23cm180 CAMEROON, Mt.Cameroon Nat. Park, below P&T Camp, 4°17.324'N 9°12.619'E WGS84, h=2359m, accuracy 4m, extent 30m, montane forest edge, Hypericum, ferns, Poaceae, Asplenium, sifting [23Cm180] [Garmin 62CSx] V.I.Gusarov & team 16.iii.2023

23cm181 CAMEROON, Mt.Cameroon Nat. Park, env. of P&T Camp, 4°16.973'N 9°12.785'E WGS84, h=2535m, accuracy 3m, extent 20m, montane forest edge, small Hypericum trees, ferns, sifting [23Cm181] [Garmin 62CSx] V.I.Gusarov & team   16.iii.2023

23cm182 CAMEROON, Mt.Cameroon Nat. Park, env. of P&T Camp, 4°16.872'N 9°13.273'E WGS84, h=2565m, accuracy 5m, extent 25m, slope with montane forest patch, small Hypericum trees, Poaceae, sifting [23Cm182] [Garmin 62CSx] V.I.Gusarov & G.M.Ojong-Nkongho   16.iii.2023

23cm187 CAMEROON, Mt.Cameroon Nat. Park, road btw P&T Camp & Bova vill. (Buea), 4°14.377'N 9°15.139'E WGS84, h=1741m, accuracy 5m, extent 30m, wet montane forest, sifting [23Cm187] [Garmin 62CSx] V.I.Gusarov & team   17.iii.2023

23cm189 CAMEROON, Mt.Cameroon Nat. Park, road btw P&T Camp & Bova vill. (Buea), 4°13.838'N 9°15.449'E WGS84, h=1472m, accuracy 6m, extent 50m, wet montane forest, sifting [23Cm189] [Garmin 62CSx] V.I.Gusarov & team   17.iii.2023

Table 7. All sample information on samples collected in the North Pare Mountains. Sample IDs in parentheses are sample IDs from the original dataset.

Sample ID	Sample Description
TZN-101	TANZANIA, Kilimanjaro Reg., 14 km NE Mwanga, North Pare Mts., Ngofe Hill Forest (Reserve), above Vuchama (Sofe) village, 37M 352972 9604096 ARC60, 3º35.015'S 37º40.626'E WGS84, accuracy 4m, extent 25m, h=1538 m, open forest, small and some medium size trees [4602] [Garmin 60CSx] V.I.Gusarov 17.xi.2010
TZN-112	TANZANIA, Kilimanjaro Reg., 14 km NE Mwanga, North Pare Mts., Ngofe Hill Forest (Reserve), above Vuchama (Sofe) village, 37M 352662 9604403 ARC60, 3º34.848'S 37º40.458'E WGS84, accuracy 3m, extent 30m, h=1576m, forest close to wetland, small & mid size trees, sifting leaf litter [4613] [Garmin 60CSx] V.I.Gusarov 18.xi.2010
TZN-106	TANZANIA, Kilimanjaro Reg., 14 km NE Mwanga, North Pare Mts., Ngofe Hill Forest (Reserve), above Vuchama (Sofe) village, 37M 352846 9604222 ARC60, 3º34.946'S 37º40.558'E WGS84, accuracy 3m, extent 25m, h=1592m [4607] [Garmin 60CSx] V.I.Gusarov 18.xi.2010
TZN-120	TANZANIA, Kilimanjaro Reg., 14 km NE Mwanga, North Pare Mts., Ngofe Hill Forest (Reserve), above Vuchama (Sofe) village, 37M 353268 9604183 ARC60, 3º34.968'S 37º40.786'E WGS84, a accuracy 7m, extent 8m, h=1562 m, on rock face with seeping water [4619] [Garmin 60CSx] V.I.Gusarov 19.xi.2010
TZN-109	TANZANIA, Kilimanjaro Reg., 14 km NE Mwanga, North Pare Mts., Ngofe Hill Forest (Reserve), above Vuchama (Sofe) village, 37M 352614 9604320 ARC60, 3º34.893'S 37º40.432'E WGS84, accuracy 3m, extent 25m, h=1620 m, forest with gaps, mid size trees, Poaceae, sifting leaf litter [4610] [Garmin60CSx] V.I.Gusarov 18.xi.2010
TZN-127	TANZANIA, Kilimanjaro Reg., 4 km E Mwanga, North Pare Mts., 37M 346175 9595841 ARC60, 3º39.489'S 37º36.948'E WGS84, accuracy 3m, extent 15m, h=1085 m, ravine forest, Acacia, Aloe, sifting leaf litter [4627] [Garmin 60CSx]V.I.Gusarov 20.xi.2010
TZN-29	TANZANIA, Kilimanjaro Reg., 14 km SSE Mwanga, North Pare Mts., Kindoroko Forest Reserve, above Kilomeni village, 37M 350146 9584153 ARC60, 3º45.835'S 37º39.083'E WGS84, accuracy 5m, extent 30m, h=1708 m, sifting forest litter, forest edge, 20-50 m from Eucalyptus row [4525] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 6.xi.2010
TZN-30	TANZANIA, Kilimanjaro Reg., 14 km SSE Mwanga, North Pare Mts., Kindoroko Forest Reserve, above Kilomeni village, 37M 350146 9584303 ARC60, 3º45.754'S 37º39.083'E WGS84, accuracy 4m, extent 30m, h=1741 m, sifting forest litter, 250 m from forest edge, medium size trees, with Dracaena, Acacia, [4526] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 6.xi.2010
TZN-31	TANZANIA, Kilimanjaro Reg., 14 km SSE Mwanga, North Pare Mts., Kindoroko Forest Reserve, above Kilomeni village, 37M 349966 9583698 ARC60, 3º46.081'S 37º38.986'E WGS84, accuracy 3m, extent 25m, h=1748 m, sifting forest litter, forest edge, 20-50 m from Eucalyptus row, medium size trees, with Acacia [4527] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 6.xi.2010
TZN-32	TANZANIA, Kilimanjaro Reg., 14 km SSE Mwanga, North Pare Mts., Kindoroko Forest Reserve, above Kilomeni village, 37M 349963 9583843 ARC60, 3º46.003'S 37º38.984'E WGS84, accuracy 5m, extent 25m, h=1731m, sifting forest litter, 250 m from forest edge, big trees [4528] [Garmin60CSx] V.I.Gusarov & O.M.Nniwako 6.xi.2010

TZN-134   TANZANIA, Kilimanjaro Reg., 14 km SSE Mwanga, North Pare Mts., Kindoroko Forest
             Reserve, above Kilomeni village, 37M 349907 9583856 ARC60, 3º45.996'S 37º38.954'E
             WGS84, accuracy 3m, extent 25m, h=1729 m, gap in the forest, Vernonia, Lobelia, sifting
             leaf litter [4635] [Garmin60CSx] V.I.Gusarov   21.xi.2010

TZN-104   TANZANIA, Kilimanjaro Reg., 14 km SSE Mwanga, North Pare Mts., Kindoroko Forest
(bag M104)   Reserve, above Kilomeni village, 37M 349686 9585016 ARC60, 3° 45,7546'S 37°
             39,0839'E, accuracy 6m, extent 25m, h=1837 m, sifting forest litter [4540a] [Garmin eTrex]
             O.M.Nniwako   7.xi.2010

TZN-41   TANZANIA, Kilimanjaro Reg., 18 km SSE Mwanga, 4.5 km SE Lembeni, North Pare Mts.,
             Kiverenge Forest, 37M 349645 9578739 ARC60, 3º48.772'S 37º38.808'E WGS84,
             accuracy 7m, extent 25m, h=1659 m, sifting forest litter; medium size and few large trees
             [4537] [Garmin 60CSx] V.I.Gusarov   7.xi.2010

TZN-20   TANZANIA, Kilimanjaro Reg., 4 km N Same, South Pare Mts., Vumari Forest Reserve,
(TZN-2)   37M 359537 9554249 ARC60, 4º02.071'S 37º44.132'E WGS84, accuracy 5m, extent 10m,
             h=1346 m, sifting leaf litter, ravine close to forest edge (100-200m), some big trees [4500]
             [Garmin 60CSx] V.I.Gusarov   3.xi.2010

TZN-138   TANZANIA, Kilimanjaro Reg., 4 km N Same, South Pare Mts., Vumari Forest Reserve,
             37M 359500 9554254 ARC60, 4º02.068'S 37º44.112'E WGS84, accuracy 12m, extent
             100m, h=1329m, open forest, small and medium size trees, sifting forest litter [4640]
             [Garmin 60CSx] V.I.Gusarov   22.xi.2010

TZN-145   TANZANIA, Kilimanjaro Reg., 4 km N Same, South Pare Mts., Vumari Forest Reserve,
             37M 359392 9554271 ARC60, 4º02.059'S 37º44.054'E WGS84, accuracy 4m, extent 15m,
             h=1413m, sifting forest litter, nr. large boulder, fruits on ground [4647] [Garmin 60CSx]
             V.I.Gusarov   22.xi.2010

Table 8. All sample information on samples collected in the South Pare Mountains. Sample IDs in parentheses are sample IDs from the original dataset.

Sample ID	Sample Description
M116 (bag M116 9.xi)	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 387731 9522436 ARC60, 4° 19,3589'S 37° 59,3476'E, accuracy 5m, extent 25m, h=1344 m, sifting forest litter, forest close to river [4551c] [Garmin eTrex] O.M.Nniwako 9.xi.2010
M113 (bag M113 9.xi)	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 387652 9522316 ARC60, 4° 19,424'S 37° 59,3048'E, accuracy 4m, extent 25m, h=1351 m, sifting forest litter, forest close to river [4551b] [Garmin eTrex] O.M.Nniwako 9.xi.2010
M104 (bag M104 9.xi)	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 387626 9522386 ARC60, 4° 19,386'S 37° 59,2907'E, accuracy 6m, extent 25m, h=1355 m, sifting forest litter, along river [4551a] [Garmin eTrex] O.M.Nniwako 9.xi.2010
TZN-53	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 387657 9521815 ARC60, 4º19.695'S 37º59.306'E WGS84, accuracy 6m, extent 25m, h=1425 m, sifting forest litter, large and medium size trees [4551] [Garmin 60CSx] O.M.Nniwako 9.xi.2010
TZN-54	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 387385 9521712 ARC60, 4º19.750'S 37º59.159'E WGS84, accuracy 4m, extent 25m, h=1518 m, sifting forest litter, large and very small trees [4552] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 9.xi.2010
TZN-55	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 386827 9521355 ARC60, 4º19.944'S 37º58.857'E WGS84, accuracy 4m, extent 25m, h=1600 m, sifting forest litter, small ravine with tree ferns [4553] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 9.xi.2010
TZN-70	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 386644 9521243 ARC60, 4º20.005'S 37º58.758'E WGS84, accuracy 3m, extent 25m, h=1640 m, sifting forest litter, 25 m from boundary of burnt area [4568] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 11.xi.2010
TZN-71	TANZANIA, Kilimanjaro Reg., 39 km SE Same, 17 km NEE Makanya, South Pare Mts., Chome Nature Reserve, 37M 386707 9521278 ARC60, 4º19.985'S 37º58.792'E WGS84, accuracy 3m, extent 25m, h=1638 m, sifting forest litter, inside burnt area [4569] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 11.xi.2010
TZN-85	TANZANIA, Kilimanjaro Reg., 32 km SE Same, 13 km NE Makanya, South Pare Mts., Chome Nature Reserve, 37M 381085 9525651 ARC60, 4º17.608'S 37º55.756'E WGS84, accuracy 5m, extent 25m, h=1979m, sifting forest litter, 25 m to forest boundary, large, medium & small tree [4584] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 14.xi.2010
TZN-91	TANZANIA, Kilimanjaro Reg., 32 km SE Same, 13 km NE Makanya, South Pare Mts., Chome Nature Reserve, 37M 381628 9525375 ARC60, 4º17.758'S 37º56.049'E WGS84, accuracy 5m, extent 10m, h=2027m, sifting leaf litter in gap with Lobelia gibberoa, bushes and herbs [4590] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 15.xi.2010
TZN-93	TANZANIA, Kilimanjaro Reg., 32 km SE Same, 14 km NE Makanya, South Pare Mts., Chome Nature Reserve, 37M 381365 9525889 ARC60, 4º17.479'S 37º55.908'E WGS84,

accuracy 3m, extent 25m, h=2099m, sifting forest litter, large and small trees, on ridge [4592] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako   15.xi.2010

TZN-77   TANZANIA, Kilimanjaro Reg., 32 km SE Same, 15km NE Makanya, South Pare Mts., Chome Nature Reserve, Mt. Shengena, summit, 37M 381481 9528024 ARC60, 4º16.321'S 37º55.972'E WGS84, accuracy 5m, extent 25m, h=2458m, sifting forest litter, small and short trees with a lot of moss [4576] [Garmin 60CSx] V.I.Gusarov & O.M.Nniwako 13.xi.2010

TZN-153   TANZANIA, Kilimanjaro Reg., 17 km SEE Same, 6 NE Mwembe, South Pare Mts., Kwizu Forest Reserve, 37M 375836 9544144 ARC60, 4º07.568'S 37º52.932'E WGS84, accuracy 3m, extent 25m, h=1553m, disturbed forest with gaps, sifting leaf litter [4656] [Garmin 60CSx] V.I.Gusarov   23.xi.2010

TZN-155   TANZANIA, Kilimanjaro Reg., 17 km SEE Same, 6 NE Mwembe, South Pare Mts., Kwizu Forest Reserve, 37M 376106 9544224 ARC60, 4º07.525'S 37º53.079'E WGS84, accuracy 3m, extent 25m, h=1558m, medium size tree, wild banana, Dracaena, Aloe, sifting leaf litter [4658] [Garmin 60CSx] V.I.Gusarov   23.xi.2010

Appendix B: R-scripts

The following section contains the R-scripts written for the analyses performed in the thesis.
The first script contains the alpha and beta diversity analyses based on the Hill number
framework:

```r
library(ggplot2)
library(iNEXT)
library(tidyverse)
library(ggpubr)
library(iNEXT.beta3D)
library(fossil)
library(SpadeR)

# FULL CAMEROON DATA READ
data_full = read.csv("CM23_27SAMPLE_21_7_23_2.csv", sep = ";")

# FORMATTING THE DATA
names(data_full) = data_full[1,]
names(data_full)[names(data_full) == "Label"] = "Species"
names(data_full)[names(data_full) == ""] = "Subfamily"
data = data_full[-c(1,2,3),-c(1,2)]
k = 1:length(data)
data[,k] = apply(data[,k], 2, function(x) as.numeric(as.character(x)))

# SETTING UP THE DIVERSITY DATA
iNEXT_data = iNEXT(data, q = c(0,1,2), datatype = "abundance")

# SETTING UP THE ELEVATION DATA
elevation_CM = as.list(data_full[3,])
elevation_CM = elevation_CM[-c(1,2)]
elevation_CM = as.numeric(elevation_CM)
multi_elevation = rep(elevation_CM, each = 3)
iNEXT_data$AsyEst$Elevation = multi_elevation

# SETTING UP THE SLOPE DATA
slope = as.list(data_full[2,])
slope = slope[-c(1,2)]
slope = as.character(slope)
multi_slope = rep(slope, each=3)
iNEXT_data$AsyEst$Slope = multi_slope

# ORDERING MULTIFACETED PLOTS
iNEXT_data$AsyEst$Diversity = factor(iNEXT_data$AsyEst$Diversity, levels = c("Species richn
ess", "Shannon diversity", "Simpson diversity"))

# SETTING UP ALL PLOTS
plot5 = ggplot(iNEXT_data$AsyEst, aes(Elevation, Estimator)) + geom_point() + facet_wrap(~D
iversity, scales = "free") + xlab("Elevation, m") + ylab("Estimated Effective Number of Spe
cies") + geom_smooth(aes(Elevation, Estimator)) + theme_bw(base_size = 15) + theme(plot.tit
le = element_text(size = 20, face = "bold"))

plot6 = ggplot(iNEXT_data$AsyEst, aes(Elevation, Observed)) + geom_point() + facet_wrap(~Di
versity, scales = "free") + xlab("Elevation, m") + ylab("Observed Effective Number of Speci
es") + labs(title = "Mt. Cameroon Diversities of Diversity order q = 0, 1, 2") + geom_smoot
h(aes(Elevation, Observed)) + theme_bw(base_size = 15) + theme(plot.title = element_text(si
ze = 20, face = "bold"))
```

```
# MULTIARRANGING ALL PLOTS
cm_plot3 = ggarrange(plot6, plot5, ncol = 1, nrow= 2)
cm_plot3

# FULL NORTH PARE DATA READ
data_fullNP = read.csv("NorthPTest.csv", sep = ";")

# FORMATTING THE DATA
names(data_fullNP) = data_fullNP[1,]
names(data_fullNP)[names(data_fullNP) == "Label"] = "Species"
names(data_fullNP)[names(data_fullNP) == ""] = "Subfamily"
dataNP = data_fullNP[-c(1,2,3,5),-c(1,2)]
dataNP = dataNP[-1,]
k2 = 1:length(dataNP)
dataNP[,k2] = apply(dataNP[,k2], 2, function(x) as.numeric(as.character(x)))
dataNP = replace(dataNP, is.na(dataNP), 0)
dataNP = dataNP[rowSums(dataNP[])>0,]

# SETTING UP DIVERSITY DATA
iNEXT_dataNP = iNEXT(dataNP, q = c(0,1,2))

# SETTING UP ELEVATION DATA
elevation_NP = as.list(data_fullNP[3,])
elevation_NP = elevation_NP[-c(1,2)]
elevation_NP = as.numeric(elevation_NP)
multi_elevationNP = rep(elevation_NP, each = 3)
iNEXT_dataNP$AsyEst$Elevation = multi_elevationNP

# SETTING UP SLOPE DATA
slopeNP = as.list(data_fullNP[2,])
slopeNP = slopeNP[-c(1,2)]
slopeNP = as.character(slopeNP)
multi_slopeNP = rep(slopeNP, each=3)
iNEXT_dataNP$AsyEst$Slope = multi_slopeNP

# ORDERING MULTIFACETED PLOTS
iNEXT_dataNP$AsyEst$Diversity = factor(iNEXT_dataNP$AsyEst$Diversity, levels = c("Species r
ichness", "Shannon diversity", "Simpson diversity"))

# SETTING UP ALL PLOTS
plot9 = ggplot(iNEXT_dataNP$AsyEst, aes(Elevation, Estimator)) + geom_point()+ facet_wrap(~
Diversity, scales = "free") + xlab("Elevation, m") + ylab("Estimated Effective Number of Sp
ecies") + geom_smooth() + theme_bw(base_size = 15) + theme(plot.title = element_text(size =
20, face = "bold"))

plot10 = ggplot(iNEXT_dataNP$AsyEst, aes(Elevation, Observed)) + geom_point()+ facet_wrap(~
Diversity, scales = "free") + xlab("Elevation, m") + ylab("Observed Effective Number of Spe
cies") + labs(title = "North Pare Diversities of Diversity order q = 0, 1, 2") + geom_smoot
h() + theme_bw(base_size = 15) + theme(plot.title = element_text(size = 20, face = "bold"))

# MULTIARRANGING ALL PLOTS
np_plot2 = ggarrange(plot10, plot9, ncol = 1, nrow = 2)
np_plot2

# FULL SOUTH PARE DATA READ
data_fullSP = read.csv("SouthPTest.csv", sep = ";")

# FORMATTING THE DATA
names(data_fullSP) = data_fullSP[1,]
names(data_fullSP)[names(data_fullSP) == "Label"] = "Species"
names(data_fullSP)[names(data_fullSP) == ""] = "Subfamily"
dataSP = data_fullSP[-c(1,2,3),-c(1,2)]
dataSP = dataSP[-1,]
```

```
k3 = 1:length(dataSP)
dataSP[,k3] = apply(dataSP[,k3], 2, function(x) as.numeric(as.character(x)))
dataSP = replace(dataSP, is.na(dataSP), 0)
dataSP = dataSP[rowSums(dataSP[])>0,]

# SETTING UP DIVERSITY DATA
iNEXT_dataSP = iNEXT(dataSP, q = c(0,1,2))

# SETTING UP ELEVATION DATA
elevation_SP = as.list(data_fullSP[3,])
elevation_SP = elevation_SP[-c(1,2)]
elevation_SP = as.numeric(elevation_SP)
multi_elevationSP = rep(elevation_SP, each = 3)
iNEXT_dataSP$AsyEst$Elevation = multi_elevationSP

# SETTING UP SLOPE DATA
slopeSP = as.list(data_fullSP[2,])
slopeSP = slopeSP[-c(1,2)]
slopeSP = as.character(slopeSP)
multi_slopeSP = rep(slopeSP, each=3)
iNEXT_dataSP$AsyEst$Slope = multi_slopeSP

# ORDERING MULTIFACETED PLOTS
iNEXT_dataSP$AsyEst$Diversity = factor(iNEXT_dataSP$AsyEst$Diversity, levels = c("Species r
ichness", "Shannon diversity", "Simpson diversity"))

# SETTING UP ALL PLOTS
plot13 = ggplot(iNEXT_dataSP$AsyEst, aes(Elevation, Estimator)) + geom_point()+ facet_wrap(
~Diversity, scales = "free") + xlab("Elevation, m") + ylab("Estimated Effective Number of S
pecies") + geom_smooth() + theme_bw(base_size = 15) + theme(plot.title = element_text(size
= 20, face = "bold"))

plot14 = ggplot(iNEXT_dataSP$AsyEst, aes(Elevation, Observed)) + geom_point()+ facet_wrap(~
Diversity, scales = "free") + xlab("Elevation, m") + ylab("Observed Effective Number of Spe
cies") + labs(title = "South Pare Diversities of Diversity order q = 0, 1, 2") + geom_smoot
h() + theme_bw(base_size = 15) + theme(plot.title = element_text(size = 20, face = "bold"))

# MULTIARRANGING ALL PLOTS
sp_plot2 = ggarrange(plot14, plot13, ncol = 1, nrow = 2)
sp_plot2

# SETTING UP BETA3D FOR MT. CAMEROON
CM_list = list(Mt.Cameroon = data)
beta_data = iNEXTbeta3D(CM_list)
ggiNEXTbeta3D(beta_data) + labs(title = "Mt. Cameroon Taxonomic Diversity of Diversity orde
r q = 0, 1, 2")

# BETA DIVERSITY FOR ALL MOUNTAINS COMPARED
data_list = list(Mt.Cameroon = data, NorthPare = dataNP, SouthPare = dataSP) # change betwe
en data and data_10_19 for limited cam data
beta_data_mountains = iNEXTbeta3D(data_list)
beta_res = ggiNEXTbeta3D(beta_data_mountains) + labs(title = "All Mountains Taxonomic Diver
sity of order q = 0, 1, 2")

beta_trans_res = ggiNEXTbeta3D(beta_data_mountains, type = "D") + labs(title = "All Mountai
ns Taxonomic Diversity Monotonically Transformed, of order q = 0, 1, 2")

# CALCULATING ABSOLUTE DISTANCES (Km) BETWEEN COORDINATES FOR ALL MTS WITH FOSSIL
# For CM
coords_cm_full = read.csv("coords_CM.csv", sep = ";")
coords_cm = coords_cm_full[,-1]
distance_mat_cm = earth.dist(coords_cm, dist = TRUE)
distance_list = as.list(distance_mat_cm)
```

90

```r
distance_list = as.numeric(distance_list)
# For NP
coords_np_full = read.csv("coords_NP.csv", sep = ";")
coords_np = coords_np_full[,-1]
distance_mat_np = earth.dist(coords_np, dist = TRUE)
distance_list_np = as.list(distance_mat_np)
distance_list_np = as.numeric(distance_list_np)
# For SP
coords_sp_full = read.csv("coords_SP.csv", sep = ";")
coords_sp = coords_sp_full[,-1]
distance_mat_sp = earth.dist(coords_sp, dist = TRUE)
distance_list_sp = as.list(distance_mat_sp)
distance_list_sp = as.numeric(distance_list_sp)

# CREATING BETWEEN-WITHIN LISTS FOR ALL PAIRS IN ALL MTS
# For CM
dist_mat_addon = as.matrix(distance_mat_cm)
rownames(dist_mat_addon) = slope
colnames(dist_mat_addon) = slope
xy <- t(combn(colnames(dist_mat_addon), 2))
slope_match_cm = data.frame(xy, dist=dist_mat_addon[xy])
addon = ifelse(slope_match_cm$X1 == slope_match_cm$X2, "Within", "Between")
addon_list = as.character(addon)
# For NP
dist_mat_addon_np = as.matrix(distance_mat_np)
rownames(dist_mat_addon_np) = slopeNP
colnames(dist_mat_addon_np) = slopeNP
xy2 <- t(combn(colnames(dist_mat_addon_np), 2))
slope_match_np = data.frame(xy2, dist=dist_mat_addon_np[xy2])
addon_np = ifelse(slope_match_np$X1 == slope_match_np$X2, "Within", "Between")
addon_list_np = as.character(addon_np)
# For SP
dist_mat_addon_sp = as.matrix(distance_mat_sp)
rownames(dist_mat_addon_sp) = slopeSP
colnames(dist_mat_addon_sp) = slopeSP
xy3 <- t(combn(colnames(dist_mat_addon_sp), 2))
slope_match_sp = data.frame(xy3, dist=dist_mat_addon_sp[xy3])
addon_sp = ifelse(slope_match_sp$X1 == slope_match_sp$X2, "Within", "Between")
addon_list_sp = as.character(addon_sp)

## START CAMEROON ##

# Getting labels
data_spade = SimilarityMult(data, "abundance")
data_spade_cm_c22_estimates = data_spade$pairwise$C22[,1]
df_cm_spade_q2 = data.frame(distance_list, data_spade_cm_c22_estimates)

# Plotting distance over C02 similarity from SpadeR-shinyapp (q0, Sorensen, local)
q0_spade_cm_estimates = read.csv("SpadeR_q0_C02_Estimates.csv", header = FALSE)
q0_spade_cm_estimates = as.numeric(q0_spade_cm_estimates[,1])
df_cm_spade_q0 = data.frame(distance_list, q0_spade_cm_estimates)
df_cm_spade_q0$slope = addon_list

CM_C02_dist = ggplot(df_cm_spade_q0, aes(distance_list, q0_spade_cm_estimates, label = row.
names(df_cm_spade_q2), col = slope)) + geom_point() + xlab("Distance, km") + ylab("Similari
ty C02") + labs(title = "Mt. Cameroon pairwise similarity by distance, q = 0") + theme_bw()
+ theme(legend.text = element_text(size = 12), legend.title = element_text(size = 12), axis
.text=element_text(size=12), axis.title=element_text(size=12,face="bold"), plot.title = ele
ment_text(size=14))

# Same for C12 (q1, Horn, local)
q1_spade_cm_estimates = read.csv("SpadeR_q1_C01_Estimates.csv", header = FALSE) # Error in
file name here (should be C12)
q1_spade_cm_estimates = as.numeric(q1_spade_cm_estimates[,1])
```

```
df_cm_spade_q1 = data.frame(distance_list, q1_spade_cm_estimates)
df_cm_spade_q1$slope = addon_list

CM_C12_dist = ggplot(df_cm_spade_q1, aes(distance_list, q1_spade_cm_estimates, label = row.
names(df_cm_spade_q2), col = slope)) + geom_point() + xlab("Distance, km") + ylab("Similari
ty C12") + labs(title = "Mt. Cameroon pairwise similarity by distance, q = 1") + theme_bw()
+ theme(legend.text = element_text(size = 12), legend.title = element_text(size = 12), axis
.text=element_text(size=12), axis.title=element_text(size=12,face="bold"), plot.title = ele
ment_text(size=14))

# Setting up elevational differences between pairs of samples
elevation_diff = as.numeric(dist(elevation_CM))

# Testing this on equal weight Horn (q=1, C12)
df_cm_spade_q1$elevation_diff = elevation_diff

CM_C12_ele = ggplot(df_cm_spade_q1, aes(elevation_diff, q1_spade_cm_estimates, label = row.
names(df_cm_spade_q2), col = slope)) + geom_point() + xlab("Elevation difference, m") + yla
b("Similarity C12") + labs(title = "Mt. Cameroon pairwise similarity by elevation, q = 1")
+ theme_bw() + theme(legend.text = element_text(size = 12), legend.title = element_text(siz
e = 12), axis.text=element_text(size=12), axis.title=element_text(size=12,face="bold"), plo
t.title = element_text(size=14))

# Same for Sorensen
df_cm_spade_q0$elevation_diff = elevation_diff

CM_C02_ele = ggplot(df_cm_spade_q0, aes(elevation_diff, q0_spade_cm_estimates, label = row.
names(df_cm_spade_q2), col = slope)) + geom_point() + xlab("Elevation difference, m") + yla
b("Similarity C02") + labs(title = "Mt. Cameroon pairwise similarity by elevation, q = 0")
+ theme_bw() + theme(legend.text = element_text(size = 12), legend.title = element_text(siz
e = 12), axis.text=element_text(size=12), axis.title=element_text(size=12,face="bold"), plo
t.title = element_text(size=14))

# Multifacet plot with elevation and distance for Sorensen (C02) and Horn (C12)
CM_ele_dist = ggarrange(CM_C02_dist, CM_C12_dist, CM_C02_ele, CM_C12_ele, ncol = 2, nrow =
2)
## END CAMEROON ##

## START NP ##
q0_spade_np_estimates = read.csv("SpadeR_NP_q0_C02_Estimates.csv", header = FALSE)
q0_spade_np_estimates = as.numeric(q0_spade_np_estimates[,1])
elevation_diff_NP = as.numeric(dist(elevation_NP))
df_np_spade_q0 = data.frame(elevation_diff_NP, q0_spade_np_estimates, distance_list_np, "sl
ope" = addon_list_np)

NP_C02_ele = ggplot(df_np_spade_q0, aes(elevation_diff_NP, q0_spade_np_estimates, col = slo
pe)) + geom_point() + xlab("Elevation difference, m") + ylab("Similarity C02") + labs(title
= "North Pare pairwise similarity by elevation, q = 0") + theme_bw() + theme(legend.text =
element_text(size = 12), legend.title = element_text(size = 12), axis.text=element_text(siz
e=12), axis.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

NP_C02_dist = ggplot(df_np_spade_q0, aes(distance_list_np, q0_spade_np_estimates, col = slo
pe)) + geom_point() + xlab("Distance, km") + ylab("Similarity C02") + labs(title = "North P
are pairwise similarity by distance, q = 0") + theme_bw() + theme(legend.text = element_tex
t(size = 12), legend.title = element_text(size = 12), axis.text=element_text(size=12), axis
.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

q1_spade_np_estimates = read.csv("SpadeR_NP_q1_C12_Estimates.csv", header = FALSE)
q1_spade_np_estimates = as.numeric(q1_spade_np_estimates[,1])
df_np_spade_q1 = data.frame(elevation_diff_NP, q1_spade_np_estimates, distance_list_np, "sl
ope" = addon_list_np)
```

```
NP_C12_ele = ggplot(df_np_spade_q1, aes(elevation_diff_NP, q1_spade_np_estimates, col = slo
pe)) + geom_point() + xlab("Elevation difference, m") + ylab("Similarity C12") + labs(title
= "North Pare pairwise similarity by elevation, q = 1") + theme_bw() + theme(legend.text =
element_text(size = 12), legend.title = element_text(size = 12), axis.text=element_text(siz
e=12), axis.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

NP_C12_dist = ggplot(df_np_spade_q1, aes(distance_list_np, q1_spade_np_estimates, col = slo
pe)) + geom_point() + xlab("Distance, km") + ylab("Similarity C12") + labs(title = "North P
are pairwise similarity by distance, q = 1") + theme_bw() + theme(legend.text = element_tex
t(size = 12), legend.title = element_text(size = 12), axis.text=element_text(size=12), axis
.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

NP_ele_dist = ggarrange(NP_C02_dist, NP_C12_dist, NP_C02_ele, NP_C12_ele, ncol = 2, nrow =
2)
## END NP ##

## START SP ##
q0_spade_sp_estimates = read.csv("SpadeR_SP_q0_C02_Estimates.csv", header = FALSE)
q0_spade_sp_estimates = as.numeric(q0_spade_sp_estimates[,1])
elevation_diff_SP = as.numeric(dist(elevation_SP))
df_sp_spade_q0 = data.frame(elevation_diff_SP, q0_spade_sp_estimates, distance_list_sp, "sl
ope" = addon_list_sp)

SP_C02_ele = ggplot(df_sp_spade_q0, aes(elevation_diff_SP, q0_spade_sp_estimates, col = slo
pe)) + geom_point() + xlab("Elevation difference, m") + ylab("Similarity C02") + labs(title
= "South Pare pairwise similarity by elevation, q = 0") + theme_bw() + theme(legend.text =
element_text(size = 12), legend.title = element_text(size = 12), axis.text=element_text(siz
e=12), axis.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

SP_C02_dist = ggplot(df_sp_spade_q0, aes(distance_list_sp, q0_spade_sp_estimates, col = slo
pe)) + geom_point() + xlab("Distance, km") + ylab("Similarity C02") + labs(title = "South P
are pairwise similarity by distance, q = 0") + theme_bw() + theme(legend.text = element_tex
t(size = 12), legend.title = element_text(size = 12), axis.text=element_text(size=12), axis
.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

q1_spade_sp_estimates = read.csv("SpadeR_SP_q1_C12_Estimates.csv", header = FALSE)
q1_spade_sp_estimates = as.numeric(q1_spade_sp_estimates[,1])
df_sp_spade_q1 = data.frame(elevation_diff_SP, q1_spade_sp_estimates, distance_list_sp, "sl
ope" = addon_list_sp)

SP_C12_ele = ggplot(df_sp_spade_q1, aes(elevation_diff_SP, q1_spade_sp_estimates, col = slo
pe)) + geom_point() + xlab("Elevation difference, m") + ylab("Similarity C12") + labs(title
= "South Pare pairwise similarity by elevation, q = 1") + theme_bw() + theme(legend.text =
element_text(size = 12), legend.title = element_text(size = 12), axis.text=element_text(siz
e=12), axis.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

SP_C12_dist = ggplot(df_sp_spade_q1, aes(distance_list_sp, q1_spade_sp_estimates, col = slo
pe)) + geom_point() + xlab("Distance, km") + ylab("Similarity C12") + labs(title = "South P
are pairwise similarity by distance, q = 1") + theme_bw() + theme(legend.text = element_tex
t(size = 12), legend.title = element_text(size = 12), axis.text=element_text(size=12), axis
.title=element_text(size=12,face="bold"), plot.title = element_text(size=14))

SP_ele_dist = ggarrange(SP_C02_dist, SP_C12_dist, SP_C02_ele, SP_C12_ele, ncol = 2, nrow =
2)
## END SP ##
```

The second script contains the code written for the ordination methods performed in the thesis:

```r
library(tidyverse)
library(hillR)
library(vegan)
library(ggplot2)
library(MASS)
library(packfor)

# SETTING UP DATA
cm23_27S = read.csv("cm23_27S_12_10_23.csv", header = TRUE, row.names = 1, sep = ";")
names_cm = c("106.1810", "112.2254", "116.2362", "118.2188", "120.2117", "121.2022", "134.2
386", "136.2566", "137.2712", "138.2866", "149.1901", "152.1594", "154.1393", "165.2164", "
182.2565", "187.1741", "46.961", "48.849", "73.1941", "80.1117", "82.1200", "84.1422", "86.
1621", "2.405", "4.510", "5.525", "6.601")
cm23_27S_abun = (cm23_27S[-1,-c(1:2)])
cm23_27S_abun = cm23_27S_abun %>% mutate_all(as.numeric)
rownames(cm23_27S_abun) = names_cm


cm23_27S_ele = cm23_27S[-1,]
rownames(cm23_27S_ele) = names_cm


# CALCULATING HILL DIVERSITIES OBSERVED
hill0 = hill_taxa(comm = cm23_27S_abun, q = 0, MARGIN = 1)
hill1 = hill_taxa(comm = cm23_27S_abun, q = 1, MARGIN = 1)
hill2 = hill_taxa(comm = cm23_27S_abun, q = 2, MARGIN = 1)


hill_df0 = as.data.frame(hill0)
hill_df1 = as.data.frame(hill1)
hill_df2 = as.data.frame(hill2)
hill_df = cbind(hill_df0, hill_df1, hill_df2)
cm23_27S_hill = cbind(cm23_27S_ele, hill_df)


# TRANSFORMING DATA TOTAL AND HELLINGER
cm23_27S_t = decostand(cm23_27S_abun, method = "total")
cm23_27S_h = decostand(cm23_27S_abun, method = "hellinger")


# SETTING UP ENVIRONMENTAL EXPLANATORY VARIABLES
sp_env = cbind(cm23_27S_hill$Elevation..m, cm23_27S_hill$hill0, cm23_27S_hill$hill1, cm23_2
7S_hill$hill2) %>% as.data.frame(apply(.,2, as.numeric))
names(sp_env) = c("Elevation", "Richness", "Shannon", "Simpson")


# CREATING NMDS DATA
cm23_nmds = metaMDS(cm23_27S_t, distance = "bray")
fit_E = envfit(cm23_nmds, sp_env, perm = 999)
scores(fit_E, "vectors")
stressplot(cm23_nmds)


# PLOTTING NMDS DATA WITH EXPLANATORY VARIABLES
plot(cm23_nmds, main = "NMDS Mt. Cameroon", cex.main = 1.5, cex.lab = 1.3)
plot(fit_E, p.max = 0.05, col = "black")
ordipointlabel(cm23_nmds, display = "sites", add = TRUE)


# CREATING RDA AND LOOKING AT THE DATA
global_model_cm = (rda(cm23_27S_h ~ Elevation + Shannon + Richness + Simpson, data = sp_env
))
anova.cca(global_model_cm, step = 1000)
anova.cca(global_model_cm, step = 1000, by = "term")
global_coll_cm = sqrt(vif.cca(global_model_cm))
global_r2_cm = RsquareAdj(global_model_cm)$adj.r.squared
fwd_cm = forward.sel(cm23_27S_h, sp_env, adjR2thresh = global_r2_cm, alpha = 0.01, nperm =
```

```
999)
adj_model_cm = rda(cm23_27S_h ~ Elevation + Shannon, data = sp_env)
adj_coll_cm = sqrt(vif.cca(adj_model_cm))
adj_r2_cm = RsquareAdj(adj_model_cm)
anova.cca(adj_model_cm, step = 1000)

perc_cm = round(100*(summary(adj_model_cm)$cont$importance[2, 1:2]), 2)
sc_si = scores(adj_model_cm, display = "sites", choices = c(1,2), scaling = 2)
sc_sp = scores(adj_model_cm, display="species", choices=c(1,2), scaling=1)
sc_si_s1 = scores(adj_model_cm, display = "sites", choices = c(1,2), scaling = 1)

# PLOTTING RDA WITH EXPLANATORY VARIABLES AND ALL NAMES
plot(adj_model_cm, scaling = "species", main = "RDA Mt. Cameroon - species and environmenta
l data", xlab = paste0("RDA1 (", perc_cm[1], ")%"), ylab = paste0("RDA2 (", perc_cm[2], ")%
"), cex.lab = 1.3, cex.main = 1.5)
text(sc_si + c(0.04, 0.05), labels = rownames(sc_si), col = "grey40", font = 2, cex = 0.8)

# SETTING UP DATA
np23 = read.csv("np23_12_10_23.csv", header = TRUE, row.names = 1, sep = ";")
names_np = c("101.1538", "104.1837", "106.1592", "109.1620", "112.1576", "120.1593", "127.1
085", "134.1729", "138.1329", "145.1413", "20.1346", "29.1346", "30.1741", "31.1748", "32.1
731", "41.1659")
np23[is.na(np23)] = 0
np23_abun = (np23[,-c(1:2)])
np23_abun = np23_abun %>% mutate_all(as.numeric)
np23_abun[is.na(np23_abun)] = 0
rownames(np23_abun) = names_np

# CALCULATING HILL DIVERSITIES OBSERVED
hill0np = hill_taxa(comm = np23_abun, q = 0, MARGIN = 1)
hill1np = hill_taxa(comm = np23_abun, q = 1, MARGIN = 1)
hill2np = hill_taxa(comm = np23_abun, q = 2, MARGIN = 1)

hill_df0np = as.data.frame(hill0np)
hill_df1np = as.data.frame(hill1np)
hill_df2np = as.data.frame(hill2np)
hill_dfnp = cbind(hill_df0np, hill_df1np, hill_df2np)
np23_hill = cbind(np23, hill_dfnp)

# TRANSFORMING DATA TOTAL AND HELLINGER
np23_t = decostand(np23_abun, method = "total")
np23_h = decostand(np23_abun, method = "hellinger")

# SETTING UP ENVIRONMENTAL EXPLANATORY VARIABLES
sp_env_np = cbind(np23_hill$Elevation, np23_hill$hill0, np23_hill$hill1, np23_hill$hill2) %
>% as.data.frame(apply(.,2, as.numeric))
names(sp_env_np) = c("Elevation", "Richness", "Shannon", "Simpson")

# CREATING NMDS DATA
np23_nmds = metaMDS(np23_t, distance = "bray")
fit_E_np = envfit(np23_nmds, sp_env_np, perm = 999)
scores(fit_E_np, "vectors")
stressplot(np23_nmds)

# PLOTTING NMDS DATA WITH EXPLANATORY VARIABLES
plot(np23_nmds, main = "NMDS North Pare", cex.main = 1.5, cex.lab = 1.3)
plot(fit_E_np, p.max = 0.05, col = "black")
ordipointlabel(np23_nmds, display = "sites", add = TRUE)

# CREATING RDA AND LOOKING AT DATA
global_model_np = (rda(np23_h ~ Elevation + Shannon + Richness + Simpson, data = sp_env_np)
)
anova.cca(global_model_np, step = 1000)
anova.cca(global_model_np, step = 1000, by = "term")
```

```r
global_coll_np = sqrt(vif.cca(global_model_np))
global_r2_np = RsquareAdj(global_model_np)$adj.r.squared
fwd_np = forward.sel(np23_h, sp_env_np, adjR2thresh = global_r2_np, alpha = 0.01, nperm = 9
99)
adj_model_np = rda(np23_h ~ Elevation + Simpson, data = sp_env_np)
adj_coll_np = sqrt(vif.cca(adj_model_np))
adj_r2_np = RsquareAdj(adj_model_np)
anova.cca(adj_model_np, step = 1000)

perc_np = round(100*(summary(adj_model_np)$cont$importance[2, 1:2]), 2)
sc_si_np = scores(adj_model_np, display = "sites", choices = c(1,2), scaling = 2)
sc_sp_np = scores(adj_model_np, display="species", choices=c(1,2), scaling=1)
sc_si_s1_np = scores(adj_model_np, display = "sites", choices = c(1,2), scaling = 1)

# PLOTTING RDA WITH EXPLANATORY VARIABLES AND ALL NAMES
plot(adj_model_np, scaling = "species", main = "RDA North Pare - species and environmental
data", xlab = paste0("RDA1 (", perc_np[1], ")%"), ylab = paste0("RDA2 (", perc_np[2], ")%")
, cex.lab = 1.3, cex.main = 1.5)
text(sc_si_np + c(0.05, 0.05), labels = rownames(sc_si_np), col = "grey40", font = 2, cex =
0.8)

# SETTING UP DATA
sp23 = read.csv("sp23_12_10_23.csv", header = TRUE, row.names = 1, sep = ";")
names_sp = c("104.1355", "113.1351", "116.1344", "153.1553", "155.1558", "53.1425", "54.151
8", "55.1600", "70.1640", "71.1638", "77.2458", "85.1979", "91.2027", "93.2099")
sp23[is.na(sp23)] = 0
sp23_abun = (sp23[,-c(1:2)])
sp23_abun = sp23_abun %>% mutate_all(as.numeric)
sp23_abun[is.na(sp23_abun)] = 0
rownames(sp23_abun) = names_sp

# CALCULATING HILL DIVERSITIES OBSERVED
hill0sp = hill_taxa(comm = sp23_abun, q = 0, MARGIN = 1)
hill1sp = hill_taxa(comm = sp23_abun, q = 1, MARGIN = 1)
hill2sp = hill_taxa(comm = sp23_abun, q = 2, MARGIN = 1)

hill_df0sp = as.data.frame(hill0sp)
hill_df1sp = as.data.frame(hill1sp)
hill_df2sp = as.data.frame(hill2sp)
hill_dfsp = cbind(hill_df0sp, hill_df1sp, hill_df2sp)
sp23_hill = cbind(sp23, hill_dfsp)

# TRANSFORMING DATA TOTAL AND HELLINGER
sp23_t = decostand(sp23_abun, method = "total")
sp23_h = decostand(sp23_abun, method = "hellinger")

# SETTING UP ENVIRONMENTAL EXPLANATORY VARIABLES
sp_env_sp = cbind(sp23_hill$altitude, sp23_hill$hill0, sp23_hill$hill1, sp23_hill$hill2) %>
% as.data.frame(apply(.,2, as.numeric))
names(sp_env_sp) = c("Elevation", "Richness", "Shannon", "Simpson")

# CREATING NMDS DATA
sp23_nmds = metaMDS(sp23_t, distance = "bray")
fit_E_sp = envfit(sp23_nmds, sp_env_sp, perm = 999)
scores(fit_E_sp, "vectors")
stressplot(sp23_nmds)

# PLOTTING NMDS DATA WITH EXPLANATORY VARIABLES
plot(sp23_nmds, main = "NMDS South Pare", cex.main = 1.5, cex.lab = 1.3)
plot(fit_E_sp, p.max = 0.05, col = "black")
ordipointlabel(sp23_nmds, display = "sites", add = TRUE)

# CREATING AND RUNNING RDA WITH EXPLANATORY VAIRABLES WITH ALL NAMES
global_model_sp = (rda(sp23_h ~ Elevation + Shannon + Richness + Simpson, data = sp_env_sp)
```

```
)
anova.cca(global_model_sp, step = 1000)
anova.cca(global_model_sp, step = 1000, by = "term")
global_coll_sp = sqrt(vif.cca(global_model_sp))
global_r2_sp = RsquareAdj(global_model_sp)$adj.r.squared
fwd_sp = forward.sel(sp23_h, sp_env_sp, adjR2thresh = global_r2_sp, alpha = 0.01, nperm = 9
99)
adj_model_sp = rda(sp23_h ~ Elevation + Shannon, data = sp_env_sp)
adj_coll_sp = sqrt(vif.cca(adj_model_sp))
adj_r2_sp = RsquareAdj(adj_model_sp)
anova.cca(adj_model_sp, step = 1000)


perc_sp = round(100*(summary(adj_model_sp)$cont$importance[2, 1:2]), 2)
sc_si_sp = scores(adj_model_sp, display = "sites", choices = c(1,2), scaling = 2)
sc_sp_sp = scores(adj_model_sp, display="species", choices=c(1,2), scaling=1)
sc_si_s1_sp = scores(adj_model_sp, display = "sites", choices = c(1,2), scaling = 1)


# PLOTTING RDA WITH EXPLANATORY VARIABLES AND ALL NAMES
plot(adj_model_sp, scaling = "species", main = "RDA South Pare - species and environmental
data", xlab = paste0("RDA1 (", perc_sp[1], ")%"), ylab = paste0("RDA2 (", perc_sp[2], ")%")
, cex.lab = 1.3, cex.main = 1.5)
text(sc_si_sp + c(0.05, 0.07), labels = rownames(sc_si_sp), col = "grey40", font = 2, cex =
1)
```

Appendix C: Statistics

Table 9. Permutation test for Mt. Cameroon global RDA under reduced model. Permutation = free. Number of permutations = 999. The model includes elevation, Shannon diversity, species richness and Simpson diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Model	4	0.22992	2.0322	0.001 ***
Residual	22	0.62228		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

Table 10. Permutation test for Mt. Cameroon global RDA under reduced model by terms. Terms are added sequentially (first to last). Permutation = free. Number of permutations = 999. The model includes elevation, Shannon diversity, species richness and Simpson diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Elevation	1	0.10374	3.6675	0.001 ***
Shannon	1	0.05715	2.0204	0.001 ***
Richness	1	0.03723	1.3163	0.101
Simpson	1	0.03181	1.1245	0.251
Residual	22	0.62228		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

Table 11. Table of values of the square root of each term's variance inflation factor (VIF) from the global RDA model of Mt. Cameroon. Values >2 are considered low.

Elevation	Shannon	Richness	Simpson
1.801629	8.198294	3.437198	5.562501

Table 12. Results from the forward selection of variables from the Mt. Cameroon global RDA model. Elevation and Shannon diversity were selected as significant variables.

Variables	order	R2	R2Cum	AdjR2Cum	F	Pval
Elevation	1	0.1217287	0.1217287	0.08659784	3.465008	0.001
Shannon	3	0.0670577	0.1887864	0.12118526	1.983922	0.002

Table 13. Tables of values of the square root of each term's variance inflation factor (VIF) from the adjusted model of Mt. Cameroon. Values >2 are considered low.

Elevation	Shannon
1.790584	1.790584

Table 14. Permutation test for Mt. Cameroon adjusted RDA under reduced model. Permutation = free. Number of permutations = 999. The model includes elevation and Shannon diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Model	2	0.16088	2.7927	0.001 ***
Residual	24	0.69132		
Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1				

```
Run 0 stress 0.135525
Run 1 stress 0.1355249
... New best solution
... Procrustes: rmse 0.0009590517  max resid 0.002186493
... Similar to previous best
Run 2 stress 0.1355249
... New best solution
... Procrustes: rmse 3.87671e-05  max resid 0.0001173007
... Similar to previous best
Run 3 stress 0.1355246
... New best solution
... Procrustes: rmse 0.0007444821  max resid 0.00170627
... Similar to previous best
Run 4 stress 0.1368657
Run 5 stress 0.1506426
Run 6 stress 0.1368775
Run 7 stress 0.1368776
Run 8 stress 0.1523808
Run 9 stress 0.1509573
Run 10 stress 0.1419016
Run 11 stress 0.1355245
... New best solution
... Procrustes: rmse 0.0004241716  max resid 0.0009720526
... Similar to previous best
Run 12 stress 0.1514342
Run 13 stress 0.1387373
Run 14 stress 0.141731
Run 15 stress 0.1460907
Run 16 stress 0.1526687
Run 17 stress 0.1387375
Run 18 stress 0.1374593
Run 19 stress 0.1457479
Run 20 stress 0.1481355
*** Best solution repeated 1 times
```

Figure 27. Stress values for the Mt. Cameroon NMDS. Values are consistently <0.2.

100
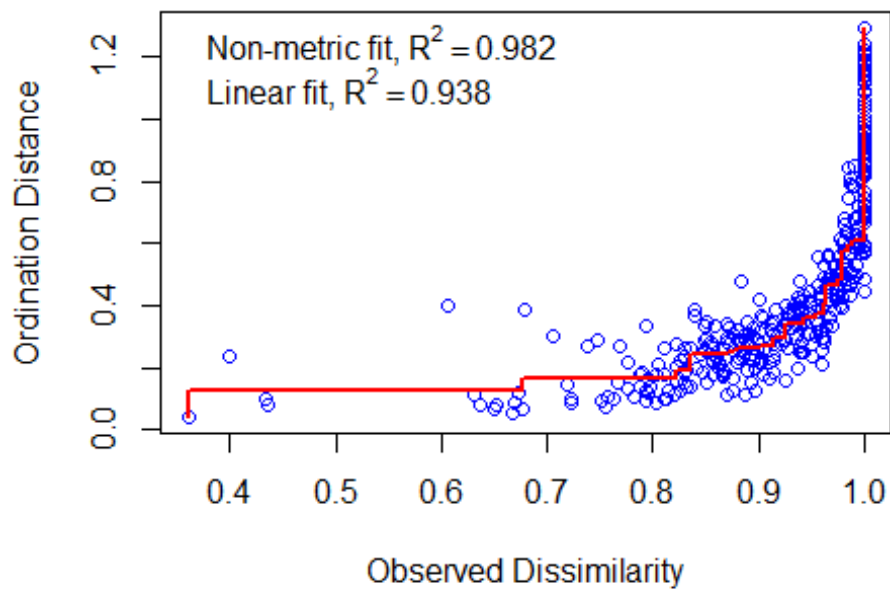
Figure 28. Shepard plot for the Mt. Cameroon NMDS. $R^2 = 0.982$.

Table 15. Permutation test for North Pare global RDA under reduced model. Permutation = free. Number of permutations = 999. The model includes elevation, Shannon diversity, species richness and Simpson diversity as explanatory variables.

	Df	**Variance**	**F**	**Pr(>F)**
Model	4	0.28441	2.0579	0.002 **
Residual	11	0.38005		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

Table 16. Permutation test for North Pare global RDA under reduced model by terms. Terms are added sequentially (first to last). Permutation = free. Number of permutations = 999. The model includes elevation, Shannon diversity, species richness and Simpson diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Elevation	1	0.12525	3.6253	0.003 ***
Shannon	1	0.07957	2.3030	0.018 *
Richness	1	0.05321	1.5401	0.118
Simpson	1	0.02637	0.7634	0.652
Residual	11	0.38005		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

Table 17. Table of values of the square root of each term's variance inflation factor (VIF) from the global RDA model of North Pare. Values >2 are considered low.

Elevation	Shannon	Richness	Simpson
1.201828	6.079894	2.290356	5.028466

Table 18. Results from the forward selection of variables from the North Pare global RDA model. Elevation and Simpson diversity were selected as significant variables.

Variables	order	R2	R2Cum	AdjR2Cum	F	Pval
Elevation	1	0.1885048	0.1885048	0.1305408	3.252104	0.002
Simpson	4	0.1298182	0.3183230	0.2134496	2.475713	0.006

Table 19. Tables of values of the square root of each term's variance inflation factor (VIF) from the adjusted model of North Pare. Values >2 are considered low.

Elevation	Simpson
1.038619	1.038619

Table 20. Permutation test for North Pare adjusted RDA under reduced model. Permutation = free. Number of permutations = 999. The model includes elevation and Simpson diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Model	2	0.21151	3.0353	0.001 ***
Residual	13	0.45295		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

```
Run 0 stress 0.1011985
Run 1 stress 0.09411699
... New best solution
... Procrustes: rmse 0.08988124  max resid 0.294703
Run 2 stress 0.1335699
Run 3 stress 0.09151782
... New best solution
... Procrustes: rmse 0.02132648  max resid 0.0584274
Run 4 stress 0.09151782
... Procrustes: rmse 1.683831e-06  max resid 2.868847e-06
... Similar to previous best
Run 5 stress 0.1418077
Run 6 stress 0.1319272
Run 7 stress 0.1341952
Run 8 stress 0.09518047
Run 9 stress 0.09151782
... New best solution
... Procrustes: rmse 3.047588e-06  max resid 7.605862e-06
... Similar to previous best
Run 10 stress 0.130013
Run 11 stress 0.09411698
Run 12 stress 0.09151782
... Procrustes: rmse 1.135749e-06  max resid 3.750389e-06
... Similar to previous best
Run 13 stress 0.09324541
Run 14 stress 0.09324541
Run 15 stress 0.1011985
Run 16 stress 0.09518047
Run 17 stress 0.09518047
Run 18 stress 0.09324541
Run 19 stress 0.09324541
Run 20 stress 0.09151782
... Procrustes: rmse 2.042251e-06  max resid 4.073832e-06
... Similar to previous best
*** Best solution repeated 3 times
```

Figure 29. Stress values for the North Pare NMDS. Values are consistently <0.2.
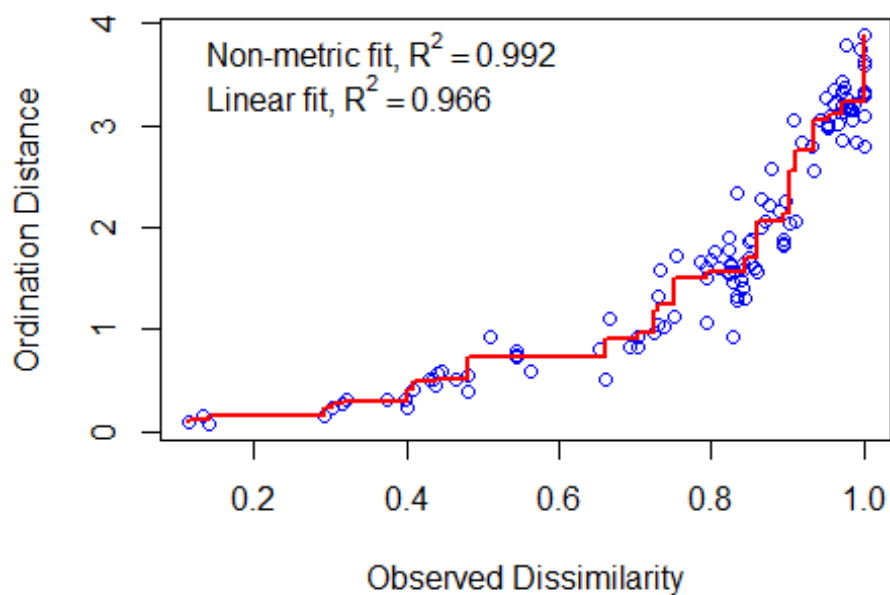
Figure 30. Shepard plot for the North Pare NMDS. $R^2 = 0.992$.

Table 17. Permutation test for South Pare global RDA under reduced model. Permutation = free. Number of permutations = 999. The model includes elevation, Shannon diversity, species richness and Simpson diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Model	4	0.30508	1.318	0.042 *
Residual	9	0.52081		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

Table 22. Permutation test for North Pare global RDA under reduced model by terms. Terms are added sequentially (first to last). Permutation = free. Number of permutations = 999. The model includes elevation, Shannon diversity, species richness and Simpson diversity as explanatory variables.

	Df	**Variance**	**F**	**Pr(>F)**
Elevation	1	0.14369	2.4830	0.001 ***
Shannon	1	0.05824	1.0065	0.462
Richness	1	0.06016	1.0397	0.405
Simpson	1	0.04299	0.7430	0.834
Residual	9	0.52081		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

Table 23. Table of values of the square root of each term's variance inflation factor (VIF) from the global RDA model of South Pare. Values >2 are considered low.

Elevation	**Shannon**	**Richness**	**Simpson**
1.096790	8.964137	2.471911	7.522421

Table 24. Results from the forward selection of variables from the South Pare global RDA model. Only elevation was selected as a significant variable.

Variables	**order**	**R2**	**R2Cum**	**AdjR2Cum**	**F**	**Pval**
Elevation	1	0.1739779	0.1739779	0.1051428	2.527457	0.001

Table 25. Tables of values of the square root of each term's variance inflation factor (VIF) from the adjusted model of South Pare. By ecological reasoning, Shannon diversity was included in the adjusted model. Values >2 are considered low.

Elevation	**Simpson**
1.0000421	1.000421

Table 26. Permutation test for South Pare adjusted RDA under reduced model. Permutation = free. Number of permutations = 999. The model includes elevation and Shannon diversity as explanatory variables.

	Df	Variance	F	Pr(>F)
Model	2	0.20193	1.7799	0.003 **
Residual	11	0.62397		

Significance codes: "***" = 0.001, "**" = 0.01, "*" = 0.05, "." = 0.1

```
Run 0 stress 0.1679106
Run 1 stress 0.1723018
Run 2 stress 0.15947
... New best solution
... Procrustes: rmse 0.171777  max resid 0.4904224
Run 3 stress 0.1594701
... Procrustes: rmse 8.664013e-05  max resid 0.0001976656
... Similar to previous best
Run 4 stress 0.1688311
Run 5 stress 0.1669038
Run 6 stress 0.1688309
Run 7 stress 0.2071221
Run 8 stress 0.1594701
... Procrustes: rmse 0.00016923  max resid 0.000401966
... Similar to previous best
Run 9 stress 0.1599674
... Procrustes: rmse 0.1090937  max resid 0.2734682
Run 10 stress 0.1640185
Run 11 stress 0.15947
... Procrustes: rmse 5.080821e-05  max resid 7.891575e-05
... Similar to previous best
Run 12 stress 0.1599885
Run 13 stress 0.1669037
Run 14 stress 0.1679107
Run 15 stress 0.1788075
Run 16 stress 0.2224292
Run 17 stress 0.1594702
... Procrustes: rmse 0.0002181951  max resid 0.000485127
... Similar to previous best
Run 18 stress 0.1688648
Run 19 stress 0.15947
... Procrustes: rmse 5.12478e-05  max resid 0.0001113255
... Similar to previous best
Run 20 stress 0.1628269
*** Best solution repeated 5 times
```

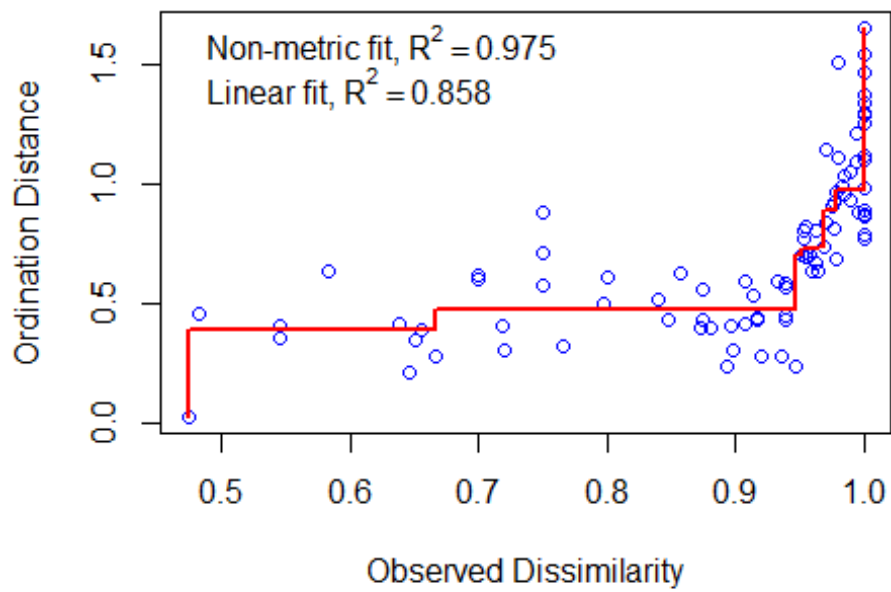**Figure 31.** Stress values for the South Pare NMDS. Values are mostly <0.2.

Figure 32. Shepard plot for the South Pare NMDS. $R^2 = 0.975$.

Appendix D: Lab Supplementary

Table 27. List of 54 newly designed primers used in the thesis. Each primer name represents the marker amplified, the direction of transcription and whether it is to be used during PCR1 or PCR2.

Primer Name	Sequence (5´-3´)
CAD-F1-La	TGMRRGGNTGGAARGARGTNGARTAYGARGTNGT
CAD-R1-La	CDGARTGNACNCCNGCRTTYTCNACRTG
CAD-F2-La	GTNTGYAAYATGGARAAYGTNGAYCCNYTNGG
CAD-R2-La	NGCNATRTTRTTNGGNARYTGNCCNCCCAT
Hem-F1-L	TTYCAYATGTGYYTNGARTTYCCNGCNCARAA
Hem-R1-L	CTNGANGANGCNARNGCNARRAAYTCYTTCAT
Hem-F2-1	CATGAACTGTGYCCNGARGARMGNCAYCAYAT
Hem-R2-1	GHYKNGGDATNGANACNGCNACRAANACCAT
sxc-F1-L	CVAAYAATGCNGTNGTNCAYGGNAAYCTSGC
sxc-R1-L	ACVRRMGGNGANARRTCNGTNGCRTTNATNATNGC
sxc-F2-1	GCARGGNCTCATYGAYYTGGCNATYGAYAC
sxc-R2-1	ARTANGTRTAHGGCATRTANGCNARYTTYTCRCT
Trpml-F1-L	CTGCARTTYTTYTTYATGAAYCCNATYGARAARTGG
Trpml-R1-L	GARTANARRTANATNYKRCKRAACCACCANARCAT
Trpml-F2-1	CAGCMGNTAYAAYCAYGTNAAYTAYACNTGGGAYAA
Trpml-R2-1	CATDATNGARAANGTNGCRAACATRTCRTCNCC
l(3)72Ab-F1-L	GGYGARGGYATYYTNATHACNAAYCAYAGYGA
l(3)72Ab-R1-L	ARRTGYCKRAANGANACNGGNARYTGNGTYTC
l(3)72Ab-F2-1	TACCTNTCNYTRYTVAAYCARCARYTNCCNATHGA
l(3)72Ab-R2-La	GTYTCNGCNCCDATCCANCKRTCNGANAC
Dhc98D-F1-L	TCCNGCNGGNACNGGNAARACNGARAC
Dhc98D-R1-L	GCNCKNGANACNGTNGCNGGNGANGC

108

Dhc98D-F2-1	GYRTNGTNACNAAYTGYGGNGARGGNATGGA
Dhc98D-R2-1	GTNARNAGYTTRTTRTCRTCCATNACNGARTTCAT
brat-F1-L	GCBATNCARGARGCNAARDCNAARGCNACNGA
brat-R1-L	TTRTCRTTVACNACNACNCCRTTNGGRAAYTC
brat-F2-1	GCTGCAGGTGCARTAYCABAARGCNCARAMBGA
brat-R2-1	TGRTCRAADATDATNACNCKCATNACYTTRCAYTC
Nrx-IV-F1-L	TYMRRATHGAYGTNGAYGGNAGYGGNCC
Nrx-IV-R1-L	TGGAABARYAAYTTNARNGGRTADATRTCRTCRAAYTC
Nrx-IV-F2-1	YNGNGAYMCNACNAARTGGTGYAAYTGYGA
Nrx-IV-R2-1	ATRTACATRTAYTGKATRTTRTTRAAYTGNGCRTCNGC
Rbcn-3A-F1-L	GCNATNTAYGARTGYGARTCNACNGGNGG
Rbcn-3A-R1-L	AANCKYTCNGCNARRTCNGTRTTRCANGTNGANAC
Rbcn-3A-F2-1	GTNCARYTNGAYTGGGTNTCNAAYGARGAYGG
Rbcn-3A-R2-1	GTRTCNGCYARNGCNARNARRTGCATYTGRTC
nonC-F1-L	CANCCDTGGAARRTNATHATHCCNCARYTRTTYTC
nonC-R1-L	GGNGTHARNCKGAANGGNACYTTYTCNGGNAC
nonC-F2-L	CAYYTVATYACNTTYCCNGCNGTNGTNGGNGC
nonC-R2-L	CTTNCCYTTYTCRAARCANACRTTRTARTCDATRTG
Cap60A-F1-L	GARTCNGCBATYGAAGCNYTNAARGARTAYGA
Cap60A-R1-L	TCAGCNGTNGMYTTRTTRTCNCCRGTRATNACRAT
Cap60A-F2-1	CNCCNYTNCARCARAARYTNGAYGARTTYGG
Cap60A-R2-1	AARACYTCYTTRCGNGGNGGRTCCARCAT
ico-F1-L	ATNGMNGMNATGCAYGARGTNAARGGNAARGAYAA
ico-R1-L	GNCCYTGNCCNCCNGAYTGYTTYTTRTG
ico-F2-1	TNGGWGCNRYNATGGAYTCNATGGARYTNGA
ico-R2-1	GCRTARATYTCNARRTGNARYTCNCCCAT

Rpll215-F1-L	GTNATYGCNTGYGTNGGNCARCARAAYGT
Rpll215-R1-L	TCRTCRTTRAARATRCARTTNARRTCRTCNCCRAA
Rpll215-F2-1	CCNCAYTTYATHAARGAYGAYTAYGGNCCNGA
Rpll215-R2-1	GRCANARNACRTTYTTNGCYTTYTCNGCRTC
zip-F1-L	GTNGARGCNATHTCNAARGCNTGYTAYGA
zip-R1-L	ARHARNGGYTTNACYTTNGTRTANARNCKCCACCA
zip-F2-1	TTYATNGGNATHYTNGAYATHGCNGGNTTYGARAT
zip-R2-1	GNGCNARNACNCCNGCNCKRAARAADATYTT
Emb-F1-L	AYNATGTGYBCNGARTTYTCNGCNATHTTYCA
Emb-R1-L	GADATCATRTANCCNACNGCYTCRTARAANGTRTG
Emb-F2-1	GNTTYYTNAAYTGGATHCCNYTNGGNTAYATHTTYGA
Emb-R2-1	TTNARNGCRATYTTDATRAANGTRTCRCANGCCAT
Nhel-F1-L	TNYTVCCNCCNATHATHTTYGARTCNGGNTAYAA
Nhel-R1-L	ACCTGYTGNGTRAAYCKNCKNGTRAARAANGGDAT
Nhel-F2-1	AAYYTRCAYAARGGNAAYTTYTTYCADAAYATHGGNTCNAT
Nhel-R2-1	TGYTCNGARTCNADNGYYTGNCCCCA
Spt6-F1-L	TGYCAYATGGGNCCNAARGTNTTYATYAAYTGYTC
Spt6-R1-L	GCRATDATYTCRTCNARRTCYTCRAAYTCYTC
Spt6-F2-1	GTNGARATHYTNGAYGGNTCNCGNGTNCAYCC
Spt6-R2-1	CNNYTTDCCYTCYTCNCKNAYRTCDATRTG
Sec24-F1-L	CGNTGYMGNGCNTGYMGNACNTAYATHAAYCC
Sec24-R1-L	CTNGANGTRTANARNARNGCNGCYTGRAARCA
Sec24-F2-1	TGYTAYCGNDTNAAYGARYTNCCNGARGARTTYCA
Sec24-R2-1	GTNGANCKNACRAARAARTTNCCRTGRAANGTRTG
Ace-F1-L	TAYTCNGGBACNARYACNYTNGANGTNTAYGA
Ace-R1-L	TGNGGNAGRTAYTTYTTCCARAANGCRCA

110

Ace-F2-1	GGNAAYCNGGNMTNTTYGAYCARATGATGGC
Ace-R2-1	AYGTARTTRATYTCRTCNGCRTGCATNACNCC
BOP1-F1-L	CCNGGNCAYGCNGARTCNTAYAAYCCNCC
BOP1-R1-L	CYKRTGRAANGCNAYNCCNCKNACNGC
BOP1-F2-1	TRCCVAGYCCNMRNGAYYTNCARCCNTTYCC
BOP1-R2-1	ACGGYTTNGTDSWNARRTCNARRTCRAACCA
rols-F1-L	CANGTNGTNGCNTAYCAYTTYTGYCARGCNGAYAA
rols-R1-L	TTNCKRTANACRTGNGCYTTNARNATRTGRTGNCC
rols-F2-1	AAYAAYACNTGYYTNGTNCCNGAYTTYRTNCAYTC
rols-R2-1	ACTCYCTRAANGANGGRTGRAARAACATRTANGTRTTRTC
Lar-F1-L	ACNAARATYACNGTNACNACNCARATGGCNGC
Lar-R1-L	AANGTYTCYTGHARNGGNCCYTGNGTNGC
Lar-F2-1	GCNWSNGARGARTAYGGNCCNATHAGYCAYTA
Lar-R2-1	ACRTAHGCRTTRTGYTTBCKRTANCCRTCRCA
FBX011-F1-L	ATHCGNCGNAAYGARATHTAYAAYGGNCAYCA
FBX011-R1-L	TTNACRCARATNGCRTTNCKRTCNGTNGTRTTRCA
FBX011-F2-1	CCNATHGTBMGNCAYAAYAARATHCAYCAYGGNCA
FBX011-R2-1	TGRCABCKRTARAARTCRTGCATNGGRAANGA
Shal-F1-L	AYGARGARDBNAARGARTAYTTYTTYGAYCGNGA
Shal-R1-L	GTRTACCARAANGCNGCNGGDATNGANGTRAA
Shal-F2-1	CCNGAYATHTTYMGNCAYATHYTNAAYTAYTAYCGYAC
Shal-R2-1	ACRTTYTTYTCNGCRTARAACATNACNGTNGC

Table 28. PCR setup for preliminary testing of the newly designed primer pairs. Numbers in parenthesis indicate values used in PCR2 rather than PCR1.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	300	1
Denaturation	95	30	
Annealing	55	90	35 (30)
Extension	72	150 (120)	
Final Extension	68	900	1
Infinite Hold	8	∞	1

Table 29. PCR setup for PCR1 during optimization of annealing temperature prior to multiplex PCR.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	300	1
Denaturation	95	30	
Annealing	47.5-60	90	5
Extension	72	150	
Denaturation	95	30	
Annealing	52.5-65	90	30
Extension	72	150	
Final Extension	68	900	1
Infinite Hold	8	∞	1

Table 30. PCR setup for PCR2 during optimization of annealing temperature prior to multiplex PCR.

Step	Temperature (°C)	Time (s)	Cycles
Initial Denaturation	95	300	1
Denaturation	95	30	
Annealing	47.5-60	90	5
Extension	72	120	
Denaturation	95	30	
Annealing	52.5-65	90	25
Extension	72	120	
Final Extension	68	900	1
Infinite Hold	8	∞	1

Table 31. Qubit concentration measures for 6 samples after normalizing and pooling of multiplex PCR products.

Sample	Concentration (ng/µl)
3279	2.57
13566	2.73
33998	2.07
43180	2.71
71898	1.61
71929	1.86

Table 32. Qubit concentration measures for 5 samples after AMPure XP-bead cleaning during library-prep.

Sample	Concentration (ng/µl)
3279	4.66
13566	4.68
33998	2.24
71898	1.94
43180	3.18

Table 33. Table of reads per barcode from resulting from the MinION sequencing.

Barcode Number	Specimen ID	Number of Reads Generated
1	3279	107323
2	3823	81602
3	4013	44000
4	4176	44735
5	13056	130391
6	13135	79549
7	13229	170671
8	13297	35213
9	13566	112314
10	13781	39973
11	16233	40593
12	23537	27105
13	25091	87344
14	25522	162002
15	33467	42918
16	33600	39393
17	33998	36897
18	36153	32248
19	36725	28825
20	37467	48732
21	37854	49326
22	37877	51774
23	41869	58764
24	42115	27989

25	43180	47558
26	43677	32993
27	52249	26965
28	58473	31685
29	64567	66291
30	71893	42569
31	71895	41752
32	71896	24935
33	71898	29753
34	71899	40526
35	71900	13258
36	71901	27731
37	71917	32088
38	71920	55180
39	71923	26185
40	71925	34835
41	71929	29704
42	71932	39137
43	71933	20836
44	71936	17222
45	71938	23785
46	71940	54354
47	71952	36517
48	71953	63979

Appendix E: Bioinformatics

Table 34. List of partitioning schemes used for each Maximum Likelihood analysis.

| Partitioning Configuration	Partitioning Scheme
No partitioning (best model by BIC)	TIM2e+I+G4
Partitioning by gene	TN+F+I+G4: brat, Emb, Lar
	TIM2+F+I+G4: CAD, Dhc98D, RPII215
	TIM3e+I+G4: Nhel, zip
	TIM2e+I+G4: rols
	TPM2+F+I+G4: Shal
Partitioning by gene and codon position	TVM+F+G4: brat3, Emb3, Shal3
	GTR+F+I+G4: brat1, CAD1, Dhc98D1, Emb1, Lar2, Nhel1, RPII2151, Shal1, zip1
	TVM+F+R2: brat2, CAD2, Dhc98D2, Emb2, Lar1, rols2, RPII2152, Shal2, zip2
	HKY+F+I+G4: CAD3, Dhc98D3, RPII2153
	TIMe+R3: Lar3, rols3, zip3
	GTR+F+G4: Nhel3
	F81+F: Nhel2
	TIM2e+G4: rols1

118
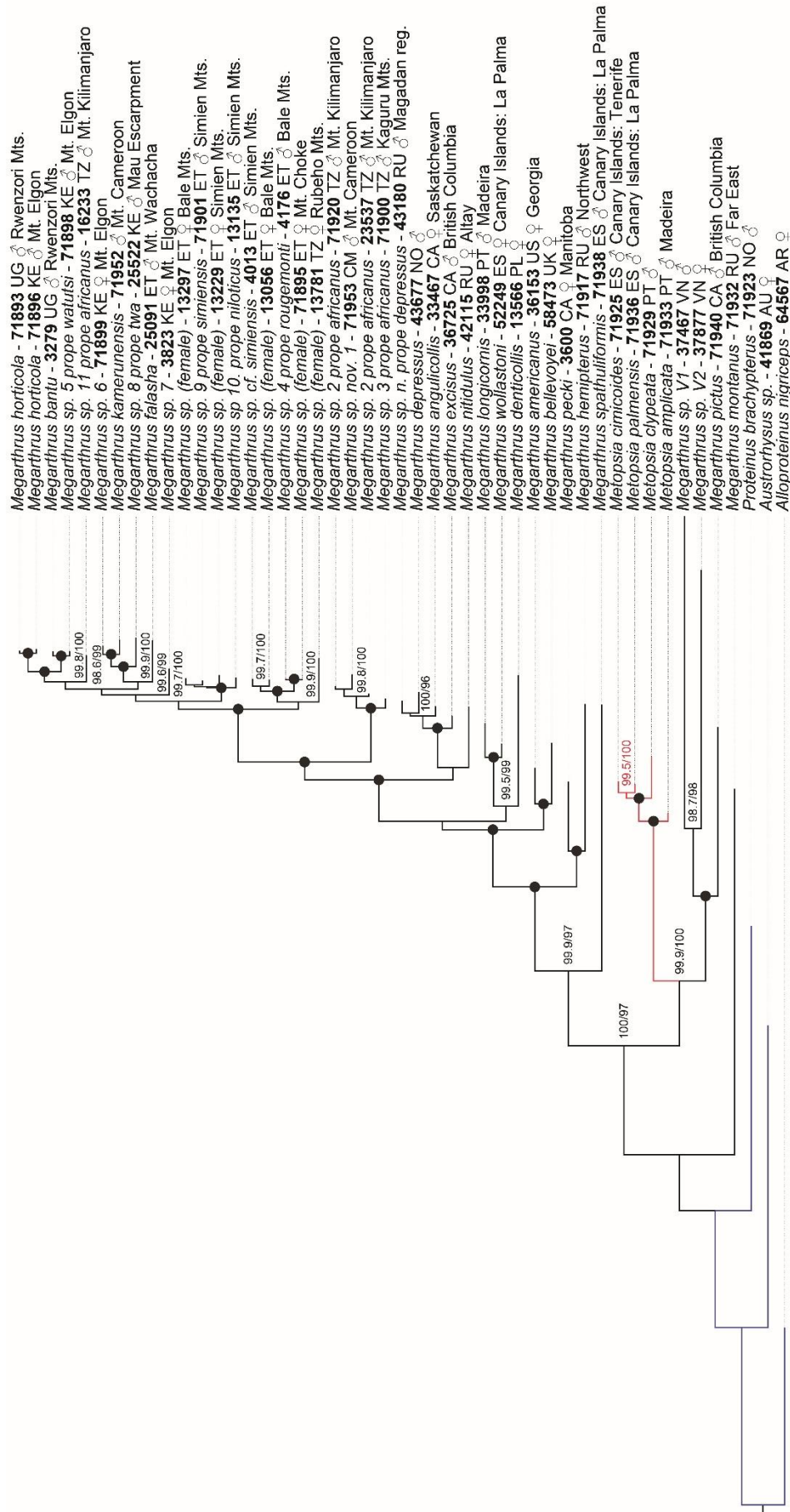
Figure 33. Single most likely Maximum Likelihood phylogeny of the 47 included Proteininae specimens, with no partitioning. Only node support values with SH-aLRT > 80% and UFB > 95% are shown. Nodes with 100%/100% support are shown only as black dots. Each specimen name is followed by its unique ID, country code, sex, and collection area.
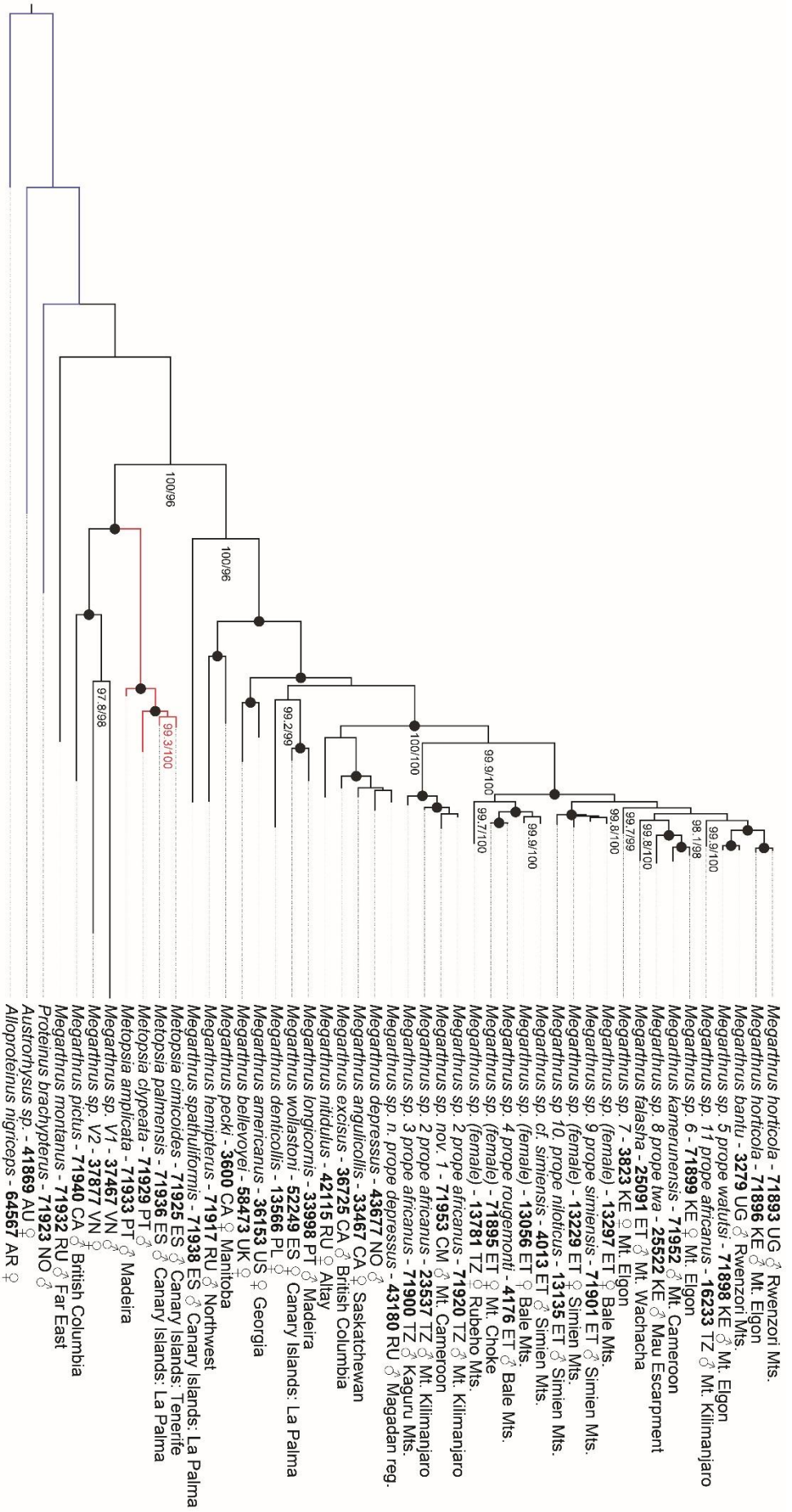
Figure 34. Single most likely Maximum Likelihood phylogeny of the 47 included Proteininae specimens, with partitioning by gene. Only node support values with SH-aLRT > 80% and UFB > 95% are shown. Nodes with 100%/100% support are shown only as black dots. Each specimen name is followed by its unique ID, country code, sex, and collection area.

119