The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding

Arribas P[1,2,3], Andújar C[1,2,3], Salces-Castellano A[1], Emerson BC[1] & Vogler AP[2,3]

[1]*Island Ecology and Evolution Research Group (IPNA-CSIC), Astrofísico Fco. Sánchez 3, 38206 La Laguna, Tenerife, Spain.*

[2]*Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.*

[3]*Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK.*

Corresponding author: Paula Arribas, pauarribas@ipna.csic.es.

RUNNING TITLE: Limited scale of dispersal in soil mesofauna

ABSTRACT

Soil mesofauna communities are hyperdiverse and critical for ecosystem functioning. However, our knowledge on spatial structure and underlying processes of community assembly for soil arthropods is scarce, hampered by limited empirical data on species diversity and turnover. We implement a high-throughput-sequencing approach to generate comparative data for thousands of arthropods at three hierarchical levels: genetic, species and supra-specific lineages. A joint analysis of the spatial arrangement across these levels can reveal the predominant processes driving the variation in biological assemblages at the local scale. This multi-hierarchical approach was performed using haplotype-level-COI metabarcoding of entire communities of mites, springtails and beetles from three Iberian mountain regions. Tens of thousands of specimens were

extracted from deep and superficial soil layers and produced comparative phylogeographic data for >1000 co-distributed species and nearly 3000 haplotypes. Local assemblages were highly distinctive between grasslands and forests, and within each of them showed strong spatial structures and high endemicity at the scale of a few kilometres or less. The local distance-decay patterns were self-similar for the haplotypes and higher hierarchical entities, and this fractal structure was very similar in all three regions, pointing to a significant role of dispersal limitation driving the local-scale community assembly. Our results from whole-community metabarcoding provide unprecedented insight into how dispersal limitations constrain mesofauna community structure within local spatial settings over evolutionary timescales. If generalized across wider areas, the high turnover and endemicity in the soil locally may indicate extremely high richness globally, challenging our current estimations of total arthropod-diversity on Earth.

KEYWORDS: cMBC**,** dispersal, distance-decay, endemism, haplotype, soil mesofauna, speciation scale.

INTRODUCTION

Soils are among the most biodiverse habitats on Earth, but represent probably the least well studied, and thus poorly understood, terrestrial ecosystem (Bardgett & van der Putten, 2014; Decaëns, 2010). Current understanding of terrestrial biodiversity has mainly relied on studies of aboveground organisms, but in recent years efforts have been focussed on developing an integrative understanding of soil biodiversity patterns and underlying mechanisms (Thakur et al., 2019). However, current knowledge on soil biodiversity is strongly unbalanced across taxonomic groups (Cameron et al., 2018), which hampers the development of such integrative framework of soil biodiversity. In particular, there is a pronounced shortage of basic biodiversity data, including estimations of their species diversity and how it is spatially structured, for the taxonomically and functionally diverse soil arthropod mesofauna composed of small-bodied invertebrates measuring between 0.1 – 2 mm and found by thousands in virtually every square meter of natural soil (Bardgett, Usher, & Hopkins, 2005; Decaëns, 2010). In recent years, high-throughput sequencing has been applied widely to study the microbial components of soil ecosystems and led to a solid understanding of global distribution patterns and the ecological drivers of soil microbiomes (e.g. Delgado-Baquerizo et al., 2018; Ramirez et al., 2018, 2014). In contrast, the study of soil arthropod mesofauna has seen comparatively little progress in exploiting these tools (mostly using 18S eDNA approaches Wu, Ayres, Bardgett, Wall, & Garey, 2011; Zinger et al., 2019), whereas conventional taxonomic approaches have been onerous, given the small body size, limited morphological variation and high local abundances of most mesofauna components.

Existing work on the diversity, distribution and community composition of soil arthropods has focussed on springtails and oribatid mites, and mostly has pointed to selection by abiotic and/or biotic environmental factors as major mechanisms of

community assembly at the local scale (e.g. Caruso, Trokhymets, Bargagli, & Convey, 2013; Magilton, Maraun, Emmerson, & Caruso, 2019; reviewed in Berg, 2012; Thakur et al., 2019). Different studies have also reported purely spatial structures (independent of the measured environmental variables) or stochastic patterns (non-environmental neither spatial structures) for the soil mesofauna communities, that have been recurrently attributed to the contribution of demographic processes (i.e., ecological drift without dispersal limitation) in determining the local community assembly (Bahram, Kohout, Anslan, Harend, & Abarenkov, 2016; Ingimarsdóttir et al., 2012; Widenfalk, Malmström, Berg, & Bengtsson, 2016; Zinger et al., 2019). In addition, dispersal limitations have also been suggested to contribute to some of the spatial community structures reported (Caruso, Taormina, & Migliorini, 2012; Gao, He, Zhang, Liu, & Wu, 2014), but still is rarely recognised as an important mechanism of assembly of the soil mesofauna at the local scale (Berg, 2012; Thakur et al., 2019). Beyond the aggregated distribution within the geographic ranges of the species, limitations to dispersal can determine the degree to which species pools are differentiated over spatial distance (Hortal, Roura-Pascual, Sanders, & Rahbek, 2010). These effects are frequently evident as biogeographic or phylogeographic breaks at large (regional to continent-wide) scales, but if the scale of movement is highly constrained and if the constraints are persistent through time, as could be the case in the soil matrix, purely spatial patterns of community assemblage dominated by species (and haplotype) turnover can arise even over relatively small distances. The potentially low taxonomic resolution (due to morphological species assignment or the use of 18S rRNA gene) of most of the studies on arthropod mesofauna communities may have missed the importance of dispersal limitation in determining the diversity patterns of soil mesofauna (but see Andújar et al., 2015; Lindo & Winchester, 2009).

93 The spatial scale at which dispersal constraints are effective in determining
94 species distributions and community assembly is a major "open question" in soil
95 biodiversity research (Thakur et al., 2019). For the mesofauna, small body size and high
96 local abundance may increase the probability of passive dispersal and long-distance
97 movement, and therefore dispersal constraints within the soil may be of limited
98 importance. In fact, a high prevalence of aerial, aquatic and marine rafting has been
99 demonstrated for various mesofaunal lineages (Coulson, Hodkinson, Webb, & Harrison,
100 2002; Nkem et al., 2006; Schuppenhauer, Lehmitz, & Xylander, 2019), and studies have
101 shown mesofaunal assemblages with no apparent dispersal limitation across continental-
102 scale areas (Baird, Leihy, Scheepers, & Chown, 2019), especially for the smallest-bodied
103 soil arthropods (Gan, Zak, & Hunter, 2019). On the other hand, molecular studies have
104 revealed high differentiation and ancient microendemicity even in morphologically
105 indistinguishable clades, indicating long-term constraints to dispersal (Andújar, Pérez-
106 González, et al., 2017; Cicconardi, Fanciulli, & Emerson, 2013). These empirical data
107 limited to particular mesofauna lineages and their contrasting findings highlight the
108 difficulty of establishing the role of dispersal constraints in community assembly. As
109 such, inferences regarding the distribution and diversification of edaphic species, and thus
110 generalisations regarding macroecological and macroevolutionary patterns, remain
111 challenging.

112 New approaches to the study of diverse and cryptic arthropods using whole-
113 community metabarcoding (cMBC) using the mitochondrial COI gene are now
114 revolutionizing the understanding of complex arthropod communities (Arribas et al.,
115 2016; Ji et al., 2013). The methodology involves the bulk sequencing of mixed
116 communities and subsequent clustering of DNA reads into operational taxonomic units
117 (OTUs) that broadly represent the species category. While an efficient method to

118 approximate community profiles at the species-level, precise removal of primary DNA
119 reads affected by sequencing errors (Andújar, Arribas, Yu, Vogler, & Emerson, 2018;
120 Elbrecht, Vamos, Steinke, & Leese, 2018; Turon, Antich, Palacín, Præbel, &
121 Wangensteen, 2019) and co-amplified nuclear mitochondrial copies (numts) (Andújar et
122 al. under review) obviate the need for clustering. Read-based data raise the prospect of
123 reliable haplotype information from mitochondrial COI cMBC, which represents a step
124 change for the study of diversity patterns through whole-community genetic analyses at
125 haplotype-level resolution.

126 The availability of metabarcode data at both species and haplotype levels permits
127 the joint analysis of turnover (beta diversity) at different hierarchical levels, to assess
128 whether the variation in biological assemblages is predominantly driven by dispersal or
129 niche-based processes (Baselga et al., 2013; Baselga, Gómez-Rodríguez, & Vogler,
130 2015). Local assemblages may diverge simply due to the lack of population movement
131 which, when assessed for entire communities, results in a largely regular decay of
132 community similarity with spatial distance for the typically neutral haplotype variation of
133 the mitochondrial COI gene. Under a scenario where dispersal constraints determine the
134 spatial community structure, assemblage turnover at the species level should mirror these
135 haplotype patterns, albeit at a higher level of similarity. In contrast, niche-based processes
136 acting on species traits produce species distributions that mainly follow environmental
137 factors and thus differ from neutral conditions determining the haplotype distributions.
138 This confounds the correlation (self-similarity) of distance decay at the species and
139 haplotype levels, each driven by different processes, and thus the joint analysis of
140 communities at genetic and species levels provides a formal test to discern if a particular
141 spatial pattern of community assemblage is predominantly driven by dispersal (i.e.,
142 neutral) or niche-based processes (Baselga et al., 2013, 2015). In addition, multi-

143 hierarchical analyses may also describe the spatial scale at which dispersal constraints
144 act, and the variation of scale among different taxonomic groups or habitats (Gómez-
145 Rodríguez, Miller, Castillejo, Iglesias-Piñeiro, & Baselga, 2018; Múrria et al., 2017). This
146 framework remains to be exploited with whole-community mitochondrial cMBC.

147 Here we apply the multi-hierarchical framework to study the spatial structure of
148 entire assemblages of mites (Acari), springtails (Collembola) and beetles (Coleoptera)
149 including many thousands of specimens, in a semi-natural mosaic landscape within three
150 geographically distinct mountain regions in southern and central Iberia (Fig. 1 A). Our
151 aim was to generate rigorous whole-community data at haplotype, species (OTU) and
152 supra-specific levels to evaluate the spatial turnover at the local scale (i.e. <10 km,
153 following scale definitions of Pearson & Dawson, 2003) and in two habitat types within
154 the same spatial settings. Using the three regions as natural replicates, we evaluated
155 patterns of richness, endemicity, turnover and the spatial scale of the distance decay in
156 community similarity at each hierarchical level and assessed the prevailing ecological and
157 evolutionary processes that determine the diversity and spatial distribution of soil
158 arthropod communities at the local scale.

159

160 **MATERIALS AND METHODS**

161 **Soil sampling and mesofauna extraction**

162 A total of 144 soil samples were collected from three regions in the southern Iberian
163 Peninsula at Sierra de Grazalema, (GRA), Sierra de Alatoz (ALZ) and Sierra de la
164 Alcarria Conquense (CUE) (Fig. 1 A). In each region, 24 points were sampled, half of
165 them in *Quercus* forest and half in wet grassland habitat, at distances of 500 m to a
166 maximum of 15 km (Fig. 1, Table S1). For each point, we collected i) a sample containing

the superficial soil layer (SUP), by extracting one square meter of leaf litter and humus up to 5 cm deep and ii) a sample of the corresponding deep soil layer (DEEP), by extracting the substrate of a 30 cm diameter core to 30 cm depth, comprising ca. 20 litres of soil. Within each region and habitat, sampling points were located in natural patches of similar dominant vegetation and elevation. Different variables characterising the sampling points were recorded including elevation, slope, orientation, stoniness, humus depth, qualitative porosity, roots, soil temperature and soil relative humidity (Table S1).

Superficial and deep soil samples were processed following the flotation–Berlese–flotation protocol (FBF) of Arribas et al. (2016) for the 'clean' extraction of arthropod mesofauna from a large volume of soil. Briefly, the FBF protocol is based on the flotation of soil in water, which allows the extraction of the organic (floating) matter containing the soil mesofauna from raw soil samples. Subsequently, the organic portion is placed in a modified Berlese apparatus to capture specimens alive and preserve them in absolute ethanol. The last part of the FBF protocol includes additional flotation and filtering steps of the ethanol-preserved arthropods using 1-mm and 0.45-µm wire mesh sieves to remove debris and dirt accumulated in the Berlese extract. This procedure generates two 'clean' subsamples of bulk specimens for DNA extraction, one including all adult and larval Coleoptera, and a second with the smallest mesofauna typically dominated by mites and springtails.

DNA extraction, PCR amplification and Illumina sequencing

Each bulk specimen subsample was independently homogenised and a DNA extraction was performed using the DNeasy Blood and Tissue Spin-Column Kit (Qiagen). DNA extracts were quantified using Nanodrop 8000 UV–Vis Spectrophotometer (Thermo Scientific) and the corresponding subsample pairs were combined at a ratio of 1:10 in the amount of DNA for Coleoptera to Acari plus Collembola (according to the range of

192 expected species diversity of these two fractions), in order to minimise the biomass bias
193 in the sequencing depth of the two mesofauna components. For metabarcoding, the bc3'
194 fragment corresponding to 418 bp of the 3' end of the COI barcode region was amplified.
195 Primers included a tail corresponding to the Illumina P5 and P7 sequencing adapters for
196 subsequent library preparation (see Arribas et al., 2016). For each sample, three
197 independent PCR reactions were performed and the amplicons were pooled. All
198 information regarding primers and PCR reagents and conditions is given in Table S2.
199 Amplicon pools were cleaned using Ampure XP magnetic beads, and used as template
200 for a limited-cycle secondary PCR amplification to add dual-index barcodes and the
201 Illumina sequencing adapters (Nextera XT Index Kit; Illumina, San Diego, CA, USA).
202 The resulting metabarcoding libraries were sequenced on an Illumina MiSeq sequencer
203 (2 x 300 bp paired-end reads) on ~ 1% of the flow cell each, to produce paired reads (R1
204 and R2) with a given dual tag combination for each sample. Negative controls were
205 maintained across all the different steps above and were sequenced as three independent
206 metabarcoding libraries.

207 **Bioinformatics read processing**

208 Raw reads were quality checked in Fastqc (Babraham Institute, 2013). Primers were
209 trimmed using fastx_trimmer and reads were processed in Trimmomatic (Bolger, Lohse,
210 & Usadel, 2014) using TRAILING:20. Based on results from (Andújar, Arribas, Gray, et
211 al., 2018) on the test of multiple tools and parameters for diverse metazoan metabarcoding
212 samples, we further processed each library independently following several steps of the
213 Usearch (Edgar, 2013) pipeline: reads were merged (option mergepairs – -fastq_minovlen
214 50, -fastq_maxdiffs 15), quality-filtered (Maxee = 1), trimmed to full length amplicons
215 of 418 bp (-sortbylength), dereplicated (-fastx_uniques) and denoised (-unoise3, -minsize
216 4). Denoised reads from the 48 libraries for each region, representing putative haplotypes,

were combined and dereplicated to get a collection of unique sequences for each regional dataset. The surviving reads were assigned to high-level taxonomic categories with the lowest common ancestor (LCA) algorithm implemented in MEGAN V5 (Huson, Auch, Qi, & Schuster, 2007). Each read was subjected to BLAST searches (blastn -outfmt 5 -evalue 0.001) against a reference library including the NCBI *nt* database (Accessed December 2016) plus 382 sequences corresponding to Acari and Collembola collected at Sierra de Grazalema. BLAST matches were fed into MEGAN to compute the taxonomic affinity of each read. This high level taxonomic assignation allowed extracting reads corresponding to the three target groups Acari, Collembola and Coleoptera, while excluding other taxa present in the bulk samples. Reads corresponding to the target groups were then aligned in Geneious using MAFFT and the Translation Align option, and those with insertions, deletions or stop codons disrupting the reading frame were excluded.

Surviving haplotypes from each region were further filtered to remove likely nuclear mitochondrial (numts) pseudogenes, following a protocol based on the relative abundance of co-distributed reads (Andújar et al. under review). The set of putative haplotypes for Acari, Collembola and Coleoptera was used to generate a community table with read-counts (haplotype abundance) by sample against the complete collection of reads (i.e., reads before the dereplicating and denoising steps) using Usearch (-search_exact option). Using these abundances, we firstly removed from each library those haplotypes with four or fewer reads according to the criteria used for the denoising (see above). Next, we identified haplotypes that, in all the libraries where they were present, contributed less than 1% of the total reads of the library. All reads falling in this category were then removed from the analysis, as an auxiliary criterion to define spurious copies not representing the true mitochondrial haplotypes. The 1% cut-off value removes most of the spurious numts while maximizing the number of real haplotypes to be further

242 analysed (Andújar et al. under review). Community tables of fully filtered haplotypes
243 were then transformed into incidence (presence/absence) data, that added to the haplotype
244 filtering before, resulted in normalised samples for further analyses.

245 **Analysis of community composition and assembly at multiple thresholds of genetic**
246 **similarity**

247 The set of filtered haplotypes was used to generate an UPGMA tree with corrected genetic
248 distances (F84 model), and based on this tree all haplotypes were grouped into clusters of
249 genetic similarity at different thresholds (1%, 2%, 3%, 4%, 5%, 6% and 8%). This
250 grouping procedure based on patristic pairwise distances over a phylogenetic tree
251 including all haplotype sequences provided multiple hierarchical levels that each can be
252 used to estimate alpha diversity (Figure 1 provides a graphical abstract of the workflow).
253 These diversity measures were estimated for the richness of lineages by sample for the
254 whole mesofauna community but also for the subsets corresponding to Acari, Collembola
255 and Coleoptera. To test for significant differences in alpha diversity between the
256 communities of different habitats and soil layers of each sampling point, repeated-
257 measures ANOVAs were conducted using habitat and soil layer as grouping factors and
258 sampling point as a within-subjects factor. For each of the three local settings, total
259 accumulative richness (local scale richness) by habitat and soil layer and the contribution
260 of mites, springtails and beetles was also calculated for the various levels of genetic
261 similarity. Endemicity by sampling point was computed for each hierarchical level (once
262 DEEP and SUP samples were combined) as the lineages present exclusively at a single
263 sampling point in the region divided by the total number of lineages found in the region.
264 To assess whether the endemicity by sampling point differed between the communities
265 of forests and grasslands, Wilcoxon tests were conducted using habitat as a grouping
266 factor. For each of the three local settings, the local scale endemicity, defined as the

lineages present exclusively at a particular sampling point divided by the total number of lineages in that community, was also calculated for the multiple levels of genetic similarity.

For the multi-hierarchical assessment of the variation in community composition at the local scale, the community dissimilarity matrices were generated for total beta diversity (Sorensen index, βsor) and its additive turnover (Simpson index, βsim) and nestedness (βsne) components (Baselga, 2010), for each level of genetic similarity. Community composition matrices were also used for non-parametric multidimensional scaling (NMDS) and plots were created with the *ordispider* option to visualise the compositional ordination of the communities according to the respective habitat and soil layer. To assess for significant differences, permutational ANOVAs were conducted over the community dissimilarity matrices using 999 permutations and the habitat and the soil layer as grouping factors and sampling point as a within-subjects factor. The significant relationships between the dissimilarity matrices generated for the Acari, Collembola and Coleoptera were assessed independently by permutational multiple regression on distance matrices (MRM), and additional NMDS ordinations and permutational ANOVAs as before were in this case conducted for each taxonomic group.

The analysis of the variation in community composition with spatial distance followed the 'multi-hierarchical macroecology' approach of (Baselga et al., 2013) which is based on the joint analysis of distance decay of similarity patterns across the different genetic levels. For each local setting and habitat, the relationship of community similarity between pairs of points (1 – pairwise beta diversity, see above) with their spatial distance (computed in kilometres as the Euclidean distance) was assessed independently at each level of genetic similarity (from haplotypes to 8% lineages). A negative exponential function was used to adjust a generalized linear model (GLM) with Simpson similarity as

response variable, spatial distance as predictor, log link and Gaussian error, and maintaining the spatial distances untransformed (Gómez-Rodríguez & Baselga, 2018). Finally, the existence of a fractal pattern (power law function) in the distance-decay curves across the levels of genetic similarity was assessed by a log–log Pearson correlation of genetic level and, independently: (a) number of lineages, (b) initial similarity, and (c) mean similarity. High correlation values are indicative of self-similarity in lineage branching (i.e., number of lineages) and/or spatial geometry of lineage distributional ranges (i.e., initial and mean similarity, (Baselga et al., 2015), which are predicted under a neutral process of community evolution.

In an equivalent way, these analyses were also conducted to assess the relationships of community similarity and environmental distance as computed using Gower's distance over the recorded variables characterising the sampling points (Table S1). In the cases where this relationship was significant, variance partitioning was conducted to assess the fractions of variance in community dissimilarity that are uniquely and jointly explained by spatial and environmental distance. All analyses were performed using the R-packages vegan (Oksanen et al., 2013), cluster, PMCMR, hier.part, ecodist, and betapart (Baselga & Orme, 2012).

RESULTS

Multi-hierarchical assessment of alpha and gamma diversity of soil mesofauna

Processing of 144 soil samples using the FBF protocol, followed by double dual indexing of *cox1* amplicons and Illumina MiSeq sequencing, produced 51433 to 375211 sequence reads per sample for GRA, 11307 to 159244 for ALZ, and 43562 to 128149 for CUE. Filtering of raw reads using standard protocols of read curation and denoising, followed

316 by removal of likely nuclear mitochondrial pseudogenes generated a conservative set of
317 clean sequences representing the mitochondrial haplotypes. A total of 1124, 1009 and 992
318 haplotypes where found for the GRA, ALZ and CUE local areas respectively, and these
319 numbers declined rapidly when haplotypes were grouped at increasing threshold values,
320 e.g. 511, 479 and 480 lineages at 3% similarity (Table 1), but they declined only slightly
321 further at the higher thresholds, indicating the point at which stable groups are obtained
322 that broadly represent the species level. The relative proportions of mites, springtails and
323 beetles were similar across the three local settings and hierarchical levels, with Acari
324 representing the richest group (around the 50% of clusters) followed by Collembola and
325 Coleoptera in similar proportions (Table 1).

326 The patterns of richness by sample (alpha diversity) for the different habitats, soil
327 layers and genetic thresholds were similar for the three regions, with mean values between
328 35 - 60 haplotypes and 25 - 42 lineages at 3% per sample (Fig. 2 A, B, C and Fig. S1).
329 Superficial soils had significantly higher diversity than their corresponding deep soil
330 counterparts for the overall dataset and for both of the forest and grassland habitats
331 assessed independently (Fig. 2 A, B, C, Fig S1 and Table S3). At GRA, forest habitat
332 showed significantly higher alpha diversity per sample than grassland but no significant
333 differences between forest and grassland were found for ALZ and CUE (Fig. 2 A, B, C
334 and Fig S1, Table S3).

335 The local-scale cumulative richness at each region (gamma diversity) showed
336 more diverse communities for the superficial compared with deep layers, and gamma
337 diversity was generally higher for forest than grassland habitats, but the differences
338 between both habitat types were lower than observed for alpha diversity. Thus, species
339 accumulation was higher for the grassland than forest habitats, and the grassland
340 superficial layers had the highest total richness of haplotypes for ALZ and CUE, and the

highest for the three regions at the 3% similarity level (Fig. 2 D, E, F and Fig S2). Patterns of alpha and gamma diversity for the subsets of mites, springtails and beetles were similar (Fig. S2 and S3).

Multi-hierarchical assessment of beta diversity and endemicity of soil mesofauna

Compositional dissimilarity of communities within each of the three regions was high and was dominated by lineage turnover β_{sim}, instead of nestedness β_{sne} ($0.8 > \beta_{sim} > 0.95$), across all hierarchical levels. NMDS showed a consistent pattern of the forest and grassland habitats as the main driver of the ordination while soil layers had a secondary role (Fig. 3, Fig. S4). Accordingly, for the three regions and all genetic levels, the community composition was significantly different for both habitats and soil layers but the proportion of variance explained by the forest-grassland factor was always higher (Fig. 3 and Table S4). Beta diversity matrices for mites, springtails and beetles showed high and significant correlations for each of the genetic similarity levels (Table S5), and when independently analysed, these main taxonomic groups each showed similar patterns of community composition.

Community similarity (1-pairwise beta diversity) significantly decreased with spatial distance (distance decay) at all levels of genetic similarity for both the forest and grassland habitats, and these patterns were remarkably consistent across the three local settings (Fig. 4, Table S6). The slopes of the exponential decay curves were very similar at all threshold levels, and assemblage similarity increased with each level (Fig. 4, Table S6). The levels of genetic similarity showed a high and significant log–log correlation with the number of lineages ($0.90 < r^2 > 0.96$, $p < 0.001$), initial similarity ($0.86 < r^2 > 0.96$, $p < 0.001$) and mean similarity of communities ($0.89 < r^2 > 0.95$, $p < 0.001$) for all three regions and two habitats, as expected if community variation across genetic similarity levels can be described by a fractal geometry (Baselga et al., 2013, 2015).

Comparisons of distance-decay relationships between forests and grasslands showed similar values for explained variance and for the slopes for the three regions (Fig. 4, Table S6). However, there was a consistent pattern of a lower initial community similarity in grasslands than in forests, particularly above the haplotype level (Fig. 4). Similarly, the local-scale (mean) dissimilarity of communities was always higher for grasslands than for forests and the differences between both increased across the levels of genetic similarity (Fig. 5 A, B, C). A decrease in community similarity with environmental distance was only significant in the case of the forests from ALZ and CUE, but variance partitioning showed that uniquely explained variance in environmental distance, i.e. independently of the spatial distance, was reduced (5 – 9 % of explained variation at all levels) compared with the uniquely explained variance in spatial distance (23 – 31 % of explained variation).

The endemicity within the GRA, ALZ and CUE regions ranged from 71%, 64% and 58% at the haplotype level to 55%, 53% and 46% for lineages at the 3% threshold, respectively (Table 1). Comparisons between forest and grassland habitats showed that the local scale endemicity of grassland communities was higher in the case of GRA and CUE but similar in both habitats for ALZ (Fig. 5 D, E, F). The endemicity by sampling points was consistently higher for grassland than for forest local communities particularly above the haplotype level, but the differences were significant only in the case of the GRA localities (Fig. 5 G, H, I).

DISCUSSION

In total, soil samples from three Iberian mountain regions produced over 1000 species and nearly 3000 haplotypes. Their distribution was determined across numerous sampling points, demonstrating the power of mitochondrial cMBC to overcome previous

391 impediments to studying the arthropod mesofauna of the soil using conventional
392 morphological and molecular approaches. Data analysis in a multi-hierarchical
393 framework revealed a strong spatial community structure and high levels of endemicity
394 at haplotype, species and supra-specific levels, even at sampling points that were mostly
395 within a few kilometres of each other (maximum 15 km). Patterns of turnover and
396 endemicity were similar in all three independent study regions and in the grassland and
397 forest biomes (that each harbour largely non-overlapping communities). Distance decay
398 is evident at all hierarchical levels, and can be described as self-similar. The coincidence
399 of community turnover at population level and species level is expected if soil arthropod
400 assemblages are predominantly driven by distance-based parameters, if movement is
401 strongly constrained at the local scale and over time. In addition, the overriding
402 importance of habitat-related processes was apparent from the strong differentiation of
403 grassland and forest communities, which again was seen recurrently in each of the three
404 study regions.

405 The study extends existing comparative analyses of soil mesofauna by improving
406 the taxonomic resolution, providing haplotype level variation and spanning most lineages
407 composing the soil arthropod communities. Broad surveys of invertebrate soil diversity
408 using HTS have commonly relied on markers of low species-level resolution and via
409 eDNA extracted from small soil samples (Bahram et al., 2016; Wu et al., 2011; Zinger et
410 al., 2019). Other studies have characterise specific groups of mites or springtails by
411 individualised processing of specimens and relying on morphological assignment to
412 generate species-level data (Caruso, Schaefer, Monson, & Keith, 2019; Ingimarsdóttir et
413 al., 2012), but see also Young, Proctor, DeWaard, & Hebert (2019) on molecular species
414 assignment. HTS data now greatly increase the potential of expanding both the number
415 of species studied and the level of detail at which intra-specific variation for each is

416 captured. To our knowledge, this is the first study that provides haplotype level data for
417 entire communities (one square meter of leaf litter and humus and ca. 20 litres of soil per
418 sampling point) of the three most species rich soil arthropods, which allows surveys of
419 community composition and species turnover at an unprecedented level of detail, both
420 spatially and genetically. With this data in hand, community level responses to distance-
421 based parameters can be assessed that may be not evident in other types of studies. In
422 addition, the combined haplotype and species level data permit the exploitation of the
423 hierarchical framework of Baselga et al. (2013, 2015) for discriminating between
424 distance- and niche-based factors of community assembly.

425 ***The limited spatial scale of dispersal in soil arthropods***

426 The literature exploring the community assemblage of arthropod mesofauna at the local
427 scale is generally arguing for the selection by abiotic and/or biotic environmental factors
428 as the predominant mechanisms (see Berg, 2012; Thakur et al., 2019 for a recent review)
429 but stochastic and purely spatial patterns have also been reported, pointing to a
430 contribution of dispersal and demographic processes at least in some local settings
431 (Caruso et al., 2012; Gao et al., 2014; Gao, Liu, Lin, & Wu, 2016; Zinger et al., 2019).
432 However, strong dispersal constraints have rarely been recognised as an important
433 mechanism of soil mesofauna assembly at the local scale, and this could be in part due to
434 the potentially low taxonomic resolution of most of the community-level studies in this
435 group that used morphological or 18S rRNA gene species assignment (see Tang et al.,
436 2012 on the low taxonomic resolution of this marker). Our results demonstrate high
437 community differentiation at the kilometre scale for both genetic and species levels. The
438 key observation from the multi-hierarchical analysis is the correlated distance decay at
439 haplotype and species level. Self-similarity is expected to be eroded by selection on
440 adaptive traits at the species level, but not at the (neutral) haplotype level (Baselga et al.,

2015; Gómez-Rodríguez et al., 2018). As the data largely confirm the self-similarity of distance decay at haplotype and species level, this is interpreted to support the predominant role of dispersal limitation driving community assembly. The predominance of the dispersal constraints seems to emerge at short spatial distances within the soil matrix, and the evident high turnover with physical distance suggests that our sampling within each study regions (local scale) is beyond the scale of a single metacommunity. Short dispersal distances probably have affected a significant proportion of lineages within these communities over evolutionary timescales in a largely stable spatial setting. The spatiotemporal continuum expected under this scenario predicts that lineages in more distant places have diverged at a more distant time point in evolutionary history (Baselga et al., 2013, 2015), and our findings of a largely regular distance decay at higher levels are consistent with this prediction. Additional evidence for the role of short dispersal distance driving the local community assembly comes from the high microendemicity found at all hierarchical levels, an overall picture which is not expected under a scenario with predominant environmental drivers nor ecological drift without dispersal limitation.

Yet, the influence of environmental drivers cannot be discarded entirely. Even if the recorded environmental variables did not explain the variation in community composition, a significant portion of the unexplained variance in the distance-decay curves potentially suggests the influence of non-spatial factors determining the community composition. Edaphic parameters such as soil pH or organic matter have been shown to explain a significant part of the variance observed in the distribution of the soil mesofauna communities (Caruso et al., 2012; Gao et al., 2014), and here could be driving at least part of the unexplained variation within the different habitats and regions. Edaphic environmental variables are often spatially structured and so have been also reported as potential drivers of purely spatial patterns in mesofauna communities (Caruso et al., 2019,

2012). However as exposed before, this possibility is poorly supported here, as similar spatial structures were independently found within the different habitats and regions, mirroring the distance-decay patterns at the (neutral) haplotype level, and hence suggesting that dispersal limitation is the main driver of the local spatial structure of the studied mesofauna communities.

The small spatial scale of turnover and endemicity is consistent with population genetic studies in soil mesofauna showing deep genetic breaks even over relative short geographic distances (Andújar, Pérez-González, et al., 2017; Cicconardi et al., 2013). In contrast, our results are not concordant with the extended view of long-distance dispersal as prevalent process for soil mesofauna, as might be expected from evidences of passive dispersal by air, water or in marine plankton (Decaëns, 2010; Thakur et al., 2019; Wardle, 2002). Existing reports of long-distance dispersal are mainly into virgin isolated habitats (Ingimarsdóttir et al., 2012) or recently deglaciated areas (Gan et al., 2019), or may involve the detection of mesofauna during transport (Coulson, Hodkinson, & Webb, 2003; Schuppenhauer et al., 2019). However, they do not inform about colonisation and establishment success (effective dispersal) and possibly only pertain to a few highly dispersive species. Additionally, the dispersal potential may have been overestimated due to the low resolution of morphological species identification (Cicconardi et al., 2013) leading to perceived low turnover among sites, as evident from recent large-scale barcoding studies (Young et al., 2019). Our results at the community level thus raise doubts about a generalised dispersal advantage for small-bodied arthropods and instead indicate very small dispersal distances, even over evolutionary time scales, for the majority of species that make up the complex mesofauna communities of the soil. This scale and dynamics of community assembly contrasts with patterns and processes reported for the microbial eukaryote diversity of the soil (Bahram et al., 2016) and aligns

with recent empirical evidences suggesting that at the local scale dispersal rates may be much lower for soil mesofauna than for microfauna (Zinger et al., 2019). In the context of the overall arthropod diversity (for which soil mesofauna comprises the majority of the smallest fraction), our results are not supporting the macroecological prediction for a reduce impact of dispersal limitation in the assemblage for small-bodied components compared with their bigger counterparts (de Bie et al., 2012; Ricklefs, 2004) and highlight the uniqueness of ecological and evolutionary processes driving the biodiversity of these edaphic arthropods (Andújar, Arribas, & Vogler, 2017; Andújar, Pérez-González, et al., 2017).

The role of dispersal constraints within a habitat-based framework

In spite of the evidence for an important role for dispersal limitation within each habitat type, the greatest effect was the differentiation of grassland and forest communities, which share very few species even in close (meters) spatial proximity. Previous studies also have shown great differences in soil arthropod community composition between beech forest and adjacent grassland (Caruso et al., 2012), and twice higher species richness in the forest community. The grassland-forest dichotomy in community composition is concordant with these findings, but the total diversity in either type of community was more complex: alpha diversity tended to be higher for forest habitats, but lineage accumulation across multiple sites was higher for the grasslands, resulting in higher overall landscape richness (gamma diversity). Grassland communities also had consistently lower initial and mean community similarities in the corresponding distance decay curves, together with higher levels of both point and local scale endemicity. These results are recurrent across the three sampling areas and point to slightly higher long-term dispersal constraints for the mesofauna in the grasslands studied.

Grassland species are expected to experience higher environmental variability and greater extremes, which are moderated within forested patches and thus are presumably more stable (De Frenne et al., 2019). Under the habitat stability hypothesis (Ribera & Vogler, 2000; Southwood, 1977), low species turnover is predicted in less stable habitats due to the stronger positive selection on traits promoting dispersal that are required to persist in ephemeral environments. However, our findings are not aligned with this hypothesis, suggesting similar local-scale patterns of lineage turnover within both habitats and with slightly stronger community structure for the presumably less stable grasslands. Further studies comparing the assemblages of both habitat types across gradients of stability (e.g. latitude) are fundamental to identify the similarities and disparities in the processes driving the diversity and structure of the mesofauna communities. Regardless of the potential differences between the two habitat types, the high overall turnover suggests the long-term stability of the soil environment without which the high spatial structure at multiple hierarchical levels and between grassland and forest habitats could not have arisen.

Extrapolating beyond the local scale

Disentangling the mechanisms at larger spatial scales might be challenging without the understanding of the underlying processes at relatively fine scales. The recurrence of the local patterns in each of the three study regions and across the three major taxonomic groups, corroborates the hypothesis of an underlying process of stochastic dispersal of individuals, affected by a universal type of dispersal constraint. This seems to affect to a majority of the soil arthropod fauna composing these communities, regardless of their taxonomic affinity, species traits or functional role. The soil matrix provides a common sphere in which these processes are played out, and if these soils are similar, complex communities, on average, appear to respond in a similar way. The two habitat types

clearly provide a different overall setting, obvious from the very different species present, but they also impact the respective species pool in similar ways. With the local-scale patterns and likely underlying processes reported here, questions arise about the impact for cross-regional and global spatial scales, and how these patterns and processes compare to aboveground arthropod components. If generalized across broader geographical scales and latitudes, the proposed process of a very reduced spatial scale of dispersal in soil mesofauna communities could be a major contribution to the overall gamma diversity of ecosystems and would suppose a great underestimation of global diversity on Earth. In this sense, further developments on the multi-hierarchical analysis of genetic and higher-level diversity from metabarcoding data has the potential to propel the characterisation of edaphic macrobial community structure into a new era of biodiversity discovery. By taking advantage of the full breadth of contemporary metabarcoding data at unprecedented taxonomic and geographic scales, the advances made here will be able to provide unique insights into the ecological and evolutionary processes that determine the magnitude and spatial distribution of soil arthropods.

ACKNOWLEDGMENTS

This research was funded by Newton International Program, UK to PA and the NHM Biodiversity Initiative to APV. PA was supported by two postdoctoral grants from the Royal Society (Newton International Program, UK) and the Spanish Ministry of Economy and Competitiveness (MINECO, Spain) within the Juan de la Cierva Formación Program. CA was supported by the Spanish Ministry of Economy and Competitiveness (MINECO, Spain) (CGL2015-74178-JIN). Thanks are due to the regional government of Andalucía and Castilla-la-Mancha (Spain); Alex Aitken, Andie Hall, Stephen Russell and Peter Foster (all NHM) for their technical assistance and to

565 Jesús Arribas who assisted with field sampling.

566

REFERENCES

Andújar, C., Arribas, P., Gray, C., Bruce, C., Woodward, G., Yu, D. W., & Vogler, A. P. (2018). Metabarcoding of freshwater invertebrates to detect the effects of a pesticide spill. *Molecular Ecology*, *27*(1), 146–166. doi: 10.1111/mec.14410

Andújar, C., Arribas, P., Ruzicka, F., Platt, A. C., Timmermans, M. J. T. N. M. J. T. N. J. T. N., Vogler, A. P. A. A. P., … Vogler, A. P. A. A. P. (2015). Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*, *24*(14), 3603–3617. doi: 10.1111/mec.13195

Andújar, C., Arribas, P., & Vogler, A. P. (2017). Terra incognita of soil biodiversity: unseen invasions under our feet. *Molecular Ecology*, *26*(12), 3087–3089. doi: 10.1111/mec.14112

Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the Metazoa. *Molecular Ecology*, *27*(July), 3968–3975. doi: 10.1111/mec.14844

Andújar, C., Pérez-González, S., Arribas, P., Zaballos, J. P., Vogler, A. P., & Ribera, I. (2017). Speciation below ground: Tempo and mode of diversification in a radiation of endogean ground beetles. *Molecular Ecology*, *26*(21), 6053–6070. doi: 10.1111/mec.14358

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M., Vogler, A. P., Andújar, C., … Vogler, A. P. A. P. (2016). Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution*, *7*(9), 1071–1081. doi: 10.1111/2041-210X.12557

Babraham Institute. (2013). *FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc*.

Bahram, M., Kohout, P., Anslan, S., Harend, H., & Abarenkov, K. (2016). Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME Journal*, *10*(4), 885–896. doi: 10.1038/ismej.2015.164

Baird, H. P., Leihy, R. I., Scheepers, C. J., & Chown, S. L. (2019). The ecological biogeography of indigenous and introduced Antarctic springtails. *Journal of Biogeography*, (April), 1–15. doi: 10.1111/jbi.13639

Bardgett, R. D., Usher, M. B., & Hopkins, D. W. (2005). *Biological diversity and function in soils*. Cambridge: Cambridge University Press.

Bardgett, R. D., & van der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, *515*(7528), 505–511. doi: 10.1038/nature13855

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*(1), 134–143. doi: 10.1111/j.1466-8238.2009.00490.x

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M.

605 T., & Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-
606 temporal continuum of biodiversity at species and genetic levels. *Nature*
607 *Communications*, *4*(May), 1892. doi: 10.1038/ncomms2881

608 Baselga, A., Gómez-Rodríguez, C., & Vogler, A. P. (2015). Multi-hierarchical
609 macroecology at species and genetic levels to discern neutral and non-neutral
610 processes. *Global Ecology and Biogeography*, 873–882. doi: 10.1111/geb.12322

611 Baselga, A., & Orme, C. D. L. (2012). betapart : an R package for the study of beta
612 diversity. *Methods in Ecology and Evolution*, *3*(5), 808–812. doi: 10.1111/j.2041-
613 210X.2012.00224.x

614 Berg, M. P. (2012). Patterns of biodiversity at fine and small spatial scales. In D. H.
615 Wall, R. D. Bardgett, V. M. Behan-Pelletier, & E. Al. (Eds.), *Soil Ecology and*
616 *Ecosystem Services* (pp. 136–152). Oxford: Oxford University Press.

617 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for
618 Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi:
619 10.1093/bioinformatics/btu170

620 Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Gottschall, F., …
621 Eisenhauer, N. (2018). Global gaps in soil biodiversity data. *Nature Ecology and*
622 *Evolution*, *2*(7), 1042–1043. doi: 10.1038/s41559-018-0573-8

623 Caruso, T., Schaefer, I., Monson, F., & Keith, A. M. (2019). Oribatid mites show how
624 climate and latitudinal gradients in organic matter can drive large-scale
625 biodiversity patterns of soil communities. *Journal of Biogeography*, *46*(3), 611–
626 620. doi: 10.1111/jbi.13501

627 Caruso, T., Taormina, M., & Migliorini, M. (2012). Relative role of deterministic and
628 stochastic determinants of soil animal community: a spatially explicit analysis of
629 oribatid mites. *The Journal of Animal Ecology*, *81*(1), 214–221. doi:
630 10.1111/j.1365-2656.2011.01886.x

631 Caruso, T., Trokhymets, V., Bargagli, R., & Convey, P. (2013). Biotic interactions as a
632 structuring force in soil communities: Evidence from the micro-arthropods of an
633 Antarctic moss model system. *Oecologia*, *172*(2), 495–503. doi: 10.1007/s00442-
634 012-2503-9

635 Cicconardi, F., Fanciulli, P. P. P., & Emerson, B. C. B. (2013). Collembola, the
636 biological species concept and the underestimation of global species richness.
637 *Molecular Ecology*, *22*(21), 5382–5396. doi: 10.1111/mec.12472

638 Coulson, S. J., Hodkinson, I. D., & Webb, N. R. (2003). Microscale distribution patterns
639 in high Arctic soil microarthropod communities: The influence of plant species
640 within the vegetation mosaic. *Ecography*, *26*(6), 801–809. doi: 10.1111/j.0906-
641 7590.2003.03646.x

642 Coulson, S. J., Hodkinson, I. D., Webb, N. R., & Harrison, J. A. (2002). *Survival of*
643 *terrestrial soil-dwelling arthropods on and in seawater : implications for trans-*
644 *oceanic dispersal*. 353–356.

de Bie, T., de Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., … Declerk, S. A. . (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology Letters*, *4*, 740–747. doi: 10.1111/j.1461-0248.2012.01794.x

De Frenne, P., Zellweger, F., Rodríguez-Sánchez, F., Scheffers, B. R., Hylander, K., Luoto, M., … Lenoir, J. (2019). Global buffering of temperatures under forest canopies. *Nature Ecology and Evolution*, *3*(5), 744–749. doi: 10.1038/s41559-019-0842-1

Decaëns, T. (2010). Macroecological patterns in soil communities. *Global Ecology and Biogeography*, *19*(3), 287–302. doi: 10.1111/j.1466-8238.2009.00517.x

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-gonzález, A., Eldridge, D. J., Bardgett, R. D., … Fierer, N. (2018). Bacteria Found in Soil. *Science*, *325*(February), 320–325. doi: 10.1126/science.aap9516

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998. doi: 10.1038/nmeth.2604

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644. doi: 10.7717/peerj.4644

Gan, H., Zak, D. R., & Hunter, M. D. (2019). Scale dependency of dispersal limitation, environmental filtering and biotic interactions determine the diversity and composition of oribatid mite communities. *Pedobiologia*, *74*(June 2017), 43–53. doi: 10.1016/j.pedobi.2019.03.002

Gao, M., He, P., Zhang, X., Liu, D., & Wu, D. (2014). Relative roles of spatial factors, environmental filtering and biotic interactions in fine-scale structuring of a soil mite community. *Soil Biology and Biochemistry*, *79*, 68–77. doi: 10.1016/j.soilbio.2014.09.003

Gao, M., Liu, D., Lin, L., & Wu, D. (2016). The small-scale structure of a soil mite metacommunity. *European Journal of Soil Biology*, *74*, 69–75. doi: 10.1016/j.ejsobi.2016.03.004

Gómez-Rodríguez, C., & Baselga, A. (2018). Variation among European beetle taxa in patterns of distance decay of similarity suggests a major role of dispersal processes. *Ecography*, *41*(11), 1825–1834. doi: 10.1111/ecog.03693

Gómez-Rodríguez, C., Miller, K. E., Castillejo, J., Iglesias-Piñeiro, J., & Baselga, A. (2018). Understanding dispersal limitation through the assessment of diversity patterns across phylogenetic scales below the species level. *Global Ecology and Biogeography*, (October), 1–12. doi: 10.1111/geb.12857

Hortal, J., Roura-Pascual, N., Sanders, N. J., & Rahbek, C. (2010). Understanding (insect) species distributions across spatial scales. *Ecography*, *33*(1), 51–53. doi: 10.1111/j.1600-0587.2009.06428.x

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of

685 metagenomic data. *Genome Research*, *17*(3), 377–386. doi: 10.1101/gr.5969107

686 Ingimarsdóttir, M., Caruso, T., Ripa, J., Magnúsdóttir, O. B., Migliorini, M., &
687 Hedlund, K. (2012). Primary assembly of soil communities: disentangling the
688 effect of dispersal and local environment. *Oecologia*, *170*(3), 745–754. doi:
689 10.1007/s00442-012-2334-8

690 Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., … Yu, D. W.
691 (2013). Reliable, verifiable and efficient monitoring of biodiversity via
692 metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. doi: 10.1111/ele.12162

693 Lindo, Z., & Winchester, N. N. (2009). Spatial and environmental factors contributing
694 to patterns in arboreal and terrestrial oribatid mite diversity across spatial scales.
695 *Oecologia*, *160*(4), 817–825. doi: 10.1007/s00442-009-1348-3

696 Magilton, M., Maraun, M., Emmerson, M., & Caruso, T. (2019). Oribatid mites reveal
697 that competition for resources and trophic structure combine to regulate the
698 assembly of diverse soil animal communities. *Ecology and Evolution*, *9*(14), 8320–
699 8330. doi: 10.1002/ece3.5409

700 Múrria, C., Bonada, N., Vellend, M., Zamora-Muñoz, C., Alba-Tercedor, J., Sainz-
701 cantero, C. E. C. E., … Vogler, A. P. (2017). Local environment rather than past
702 climate determines community composition of mountain stream
703 macroinvertebrates across Europe a. *Molecular Ecology*, *26*(May), 6085–6099.
704 doi: 10.1111/mec.14346

705 Nkem, J. N., Wall, D. H., Virginia, R. A., Barrett, J. E., Broos, E. J., Porazinska, D. L.,
706 & Adams, B. J. (2006). Wind dispersal of soil invertebrates in the McMurdo Dry
707 Valleys, Antarctica. *Polar Biology*, *29*(4), 346–352. doi: 10.1007/s00300-005-
708 0061-x

709 Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., …
710 Wagner, H. (2013). Vegan: Community Ecology Package. R package version 2.0-
711 10. http://cran.r-project.org/package=vegan. *R Package Ver. 2.0–8*, p. 254.

712 Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on
713 the distribution of species: are bioclimate envelope models useful? *Global Ecology
714 and Biogeography*, *12*(5), 361–371. doi: 10.1046/j.1466-822X.2003.00042.x

715 Ramirez, K. S., Knight, C. G., De Hollander, M., Brearley, F. Q., Constantinides, B.,
716 Cotton, A., … De Vries, F. T. (2018). Detecting macroecological patterns in
717 bacterial communities across independent studies of global soils. *Nature
718 Microbiology*, *3*(2), 189–196. doi: 10.1038/s41564-017-0062-x

719 Ramirez, K. S., Leff, J. W., Barberán, A., Bates, S. T., Betley, J., Thomas, W., …
720 Fierer, N. (2014). Biogeographic patterns in below-ground diversity in New York
721 City ' s Central Park are similar to those observed globally Biogeographic patterns
722 in below-ground diversity in New York City ' s Central Park are similar to those
723 observed globally. *Proceedings of the Royal Society B*, *281*, 20141988. doi:
724 http://dx.doi.org/10.1098/rspb.2014.1988

725 Ribera, I., & Vogler, A. P. (2000). Habitat type as a determinant of species range sizes:

726 the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of*
727 *the Linnean Society*, *71*(1), 33–52. doi: 10.1111/j.1095-8312.2000.tb01240.x

728 Ricklefs, R. E. (2004). A comprehensive framework for global patterns in biodiversity.
729 *Ecology Letters*, *7*(1), 1–15. doi: 10.1046/j.1461-0248.2003.00554.x

730 Schuppenhauer, M. M., Lehmitz, R., & Xylander, W. E. R. (2019). Slow-moving soil
731 organisms on a water highway: Aquatic dispersal and survival potential of
732 Oribatida and Collembola in running water. *Movement Ecology*, *7*(1), 1–14. doi:
733 10.1186/s40462-019-0165-5

734 Southwood, T. R. E. (1977). Habitat, the templet for ecological strategies? *Journal of*
735 *Animal Ecology*, *46*(2), 336–365.

736 Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, a., Barraclough, T. G., & Fontaneto,
737 D. (2012). The widely used small subunit 18S rDNA molecule greatly
738 underestimates true diversity in biodiversity surveys of the meiofauna.
739 *Proceedings of the National Academy of Sciences*, *109*(40), 16208–16212. doi:
740 10.1073/pnas.1209160109

741 Thakur, M. P., Phillips, H. R. P., Brose, U., De Vries, F. T., Lavelle, P., Loreau, M., …
742 Cameron, E. K. (2019). Towards an integrative understanding of soil biodiversity.
743 *Biological Reviews*, *31*, brv.12567. doi: 10.1111/brv.12567

744 Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2019). From
745 metabarcoding to metaphylogeography: separating the wheat from the chaff.
746 *BioRxiv*, (May), 629535. doi: 10.1101/629535

747 Wardle, D. A. (2002). *Communities and Ecosystems: Linking the aboveground and*
748 *belowground components*. Princeton: Princeton University Press.

749 Widenfalk, L. A., Malmström, A., Berg, M. P., & Bengtsson, J. (2016). Small-scale
750 Collembola community composition in a pine forest soil – Overdispersion in
751 functional traits indicates the importance of species interactions. *Soil Biology and*
752 *Biochemistry*, *103*, 52–62. doi: 10.1016/j.soilbio.2016.08.006

753 Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H., & Garey, J. R. (2011). Molecular study
754 of worldwide distribution and diversity of soil animals. *Proceedings of the*
755 *National Academy of Sciences of the United States of America*, *108*(43), 17720–
756 17725. doi: 10.1073/pnas.1103824108

757 Young, M. R., Proctor, H. C., DeWaard, J. R., & Hebert, P. D. N. (2019). DNA
758 Barcodes Expose Unexpected Diversity in Canadian Mites. *Molecular Ecology*, 0–
759 2. doi: 10.1111/mec.15292

760 Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., … Chave, J.
761 (2019). Body size determines soil community assembly in a tropical forest.
762 *Molecular Ecology*, *28*(3), 528–543. doi: 10.1111/mec.14919

763

764 **AUTHOR CONTRIBUTIONS**

765 Statement of authorship: P.A., C.A. and A.P.V. conceived the work; P.A. and C.A.

766 collected and analysed the data; P.A led the writing and all authors contributed to the

767 discussion of results and the writing.

768

769 **Table 1** Total number of haplotypes and clusters for Acari, Collembola and Coleoptera
770 for each local setting (GRA, ALZ, CUE) at increasing levels of genetic divergence
771 thresholds.

	haplotypes	1% lineages	2% lineages	3% lineages	4% lineages	5% lineages	6% lineages	8% lineages
GRA								
Total	**1124**	**693**	**559**	**511**	**487**	**470**	**458**	**436**
Acari	**540**	**354**	**286**	**260**	**243**	**233**	**226**	**219**
Collembola	**306**	**172**	**129**	**113**	**108**	**104**	**101**	**94**
Coleoptera	**278**	**167**	**144**	**138**	**136**	**133**	**131**	**123**
ALZ								
Total	**1009**	**655**	**540**	**479**	**451**	**431**	**419**	**407**
Acari	**451**	**319**	**260**	**226**	**210**	**197**	**194**	**189**
Collembola	**275**	**155**	**114**	**94**	**90**	**87**	**81**	**77**
Coleoptera	**283**	**181**	**166**	**159**	**151**	**147**	**144**	**141**
CUE								
Total	**992**	**613**	**519**	**480**	**462**	**443**	**437**	**423**
Acari	**459**	**296**	**244**	**220**	**208**	**198**	**193**	**184**
Collembola	**273**	**138**	**117**	**107**	**101**	**96**	**95**	**91**
Coleoptera	**260**	**179**	**158**	**153**	**153**	**149**	**149**	**148**

772

773

FIGURE CAPTIONS

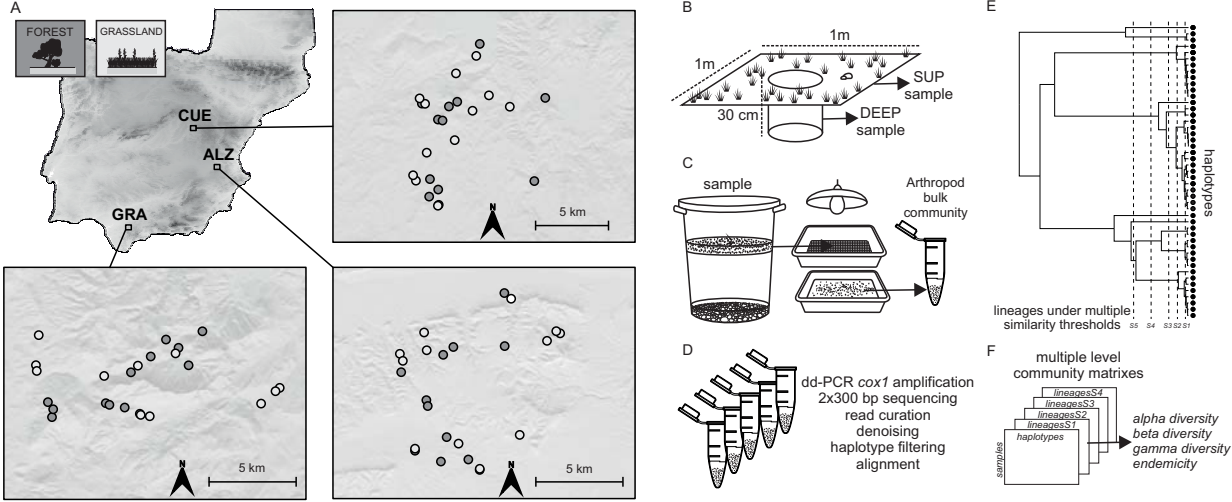
Figure 1. Sampling points in the three local settings within the Iberian Peninsula, Sierra de Grazalema, (GRA), Sierra de Alatoz (ALZ) and Sierra de la Alcarria Conquense (CUE). Sampling points are located within *Quercus* forest patches (dark grey) and wet grassland patches (pale grey).

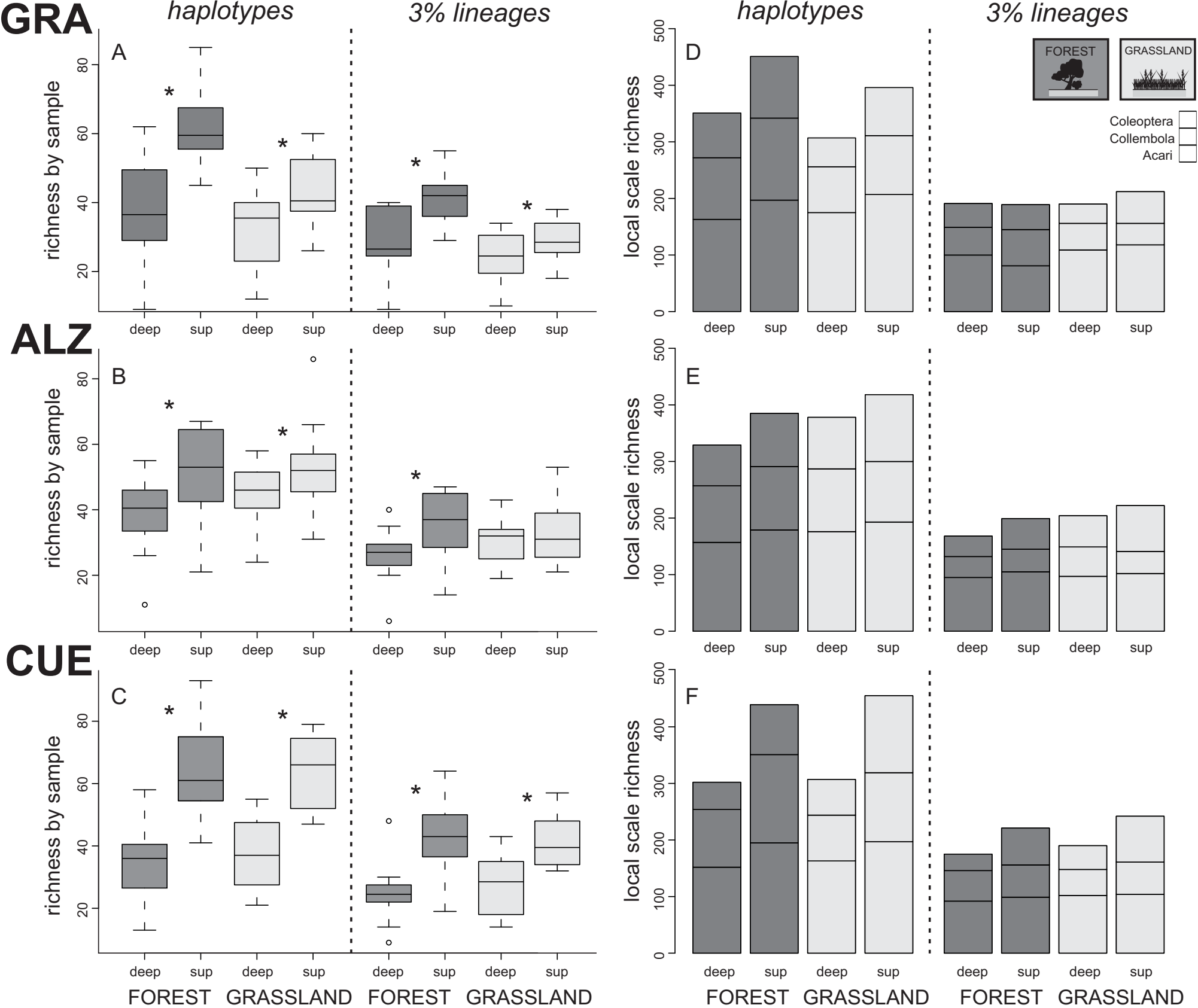
Figure 2. Richness of soil mesofauna lineages by sample (alpha diversity, A, B, C) and total accumulated richness (local scale richness, D, E, F) by habitat and soil layer for the three local settings (GRA, ALZ, CUE). Both measures are shown at the haplotype and the 3% genetic similarity levels. Forest habitat as dark grey, grassland habitat as pale gray, sup for superficial and deep for deep soil layers. Significantly different richness of lineages by sample (repeated-measures ANOVA $p < 0.05$) between deep and superficial communities of each habitat are indicated by asterisks within A, B, C panels. The contribution of Acari, Collembola and Coleoptera to the local scale richness are shown within D, E, F panels.

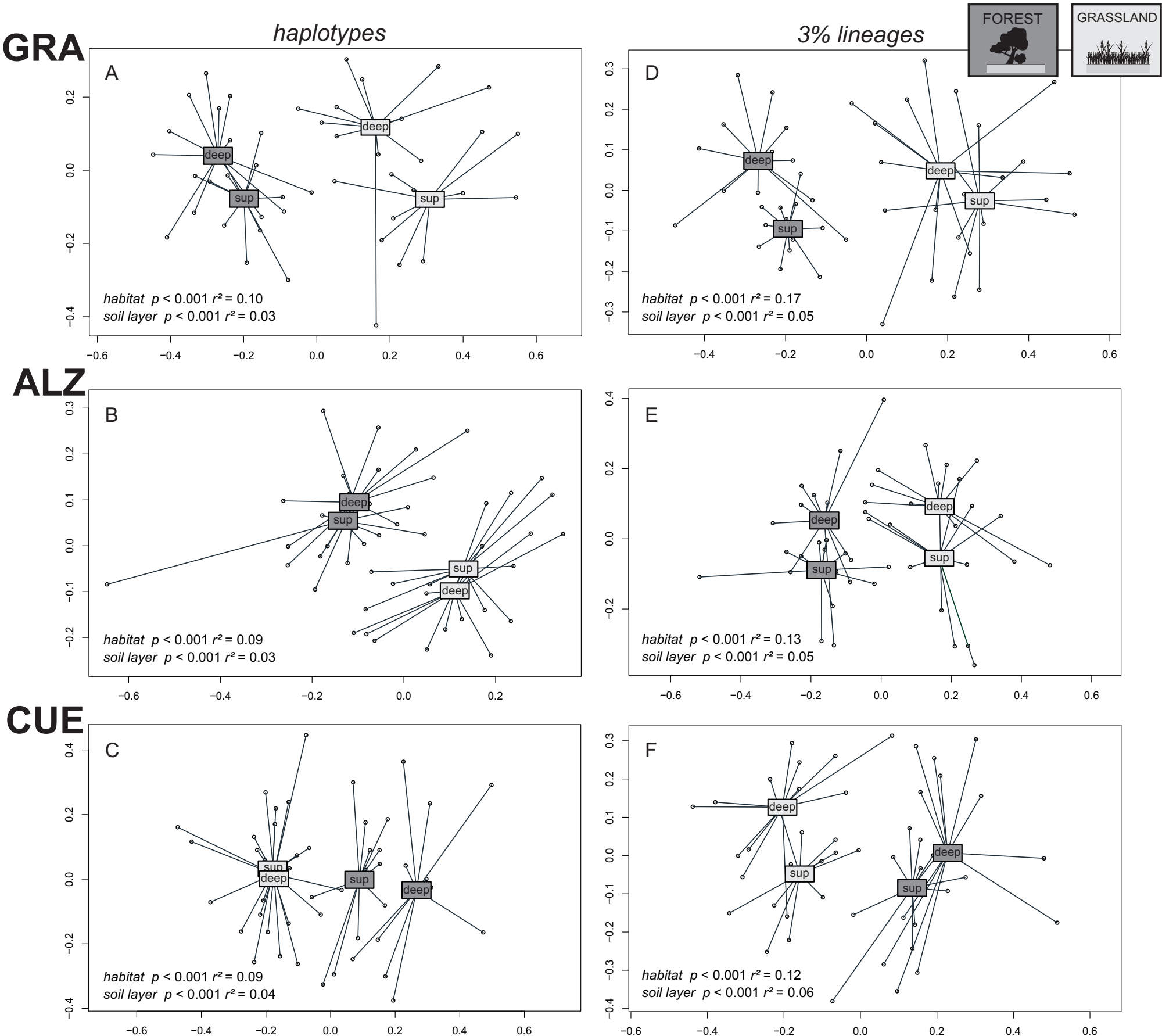
Figure 3. NMDS ordinations of the soil mesofauna samples according to the variation in community composition (Simpson index, βsim) within the three local settings (GRA, ALZ, CUE) and at the haplotype and the 3% genetic similarity levels. Forest habitat as dark grey, grassland habitat as pale grey, sup for superficial and deep for deep soil layers. Explained variation (r^2) and significance (p) of each grouping factor from the permutational ANOVAs over the community dissimilarity matrixes are shown.

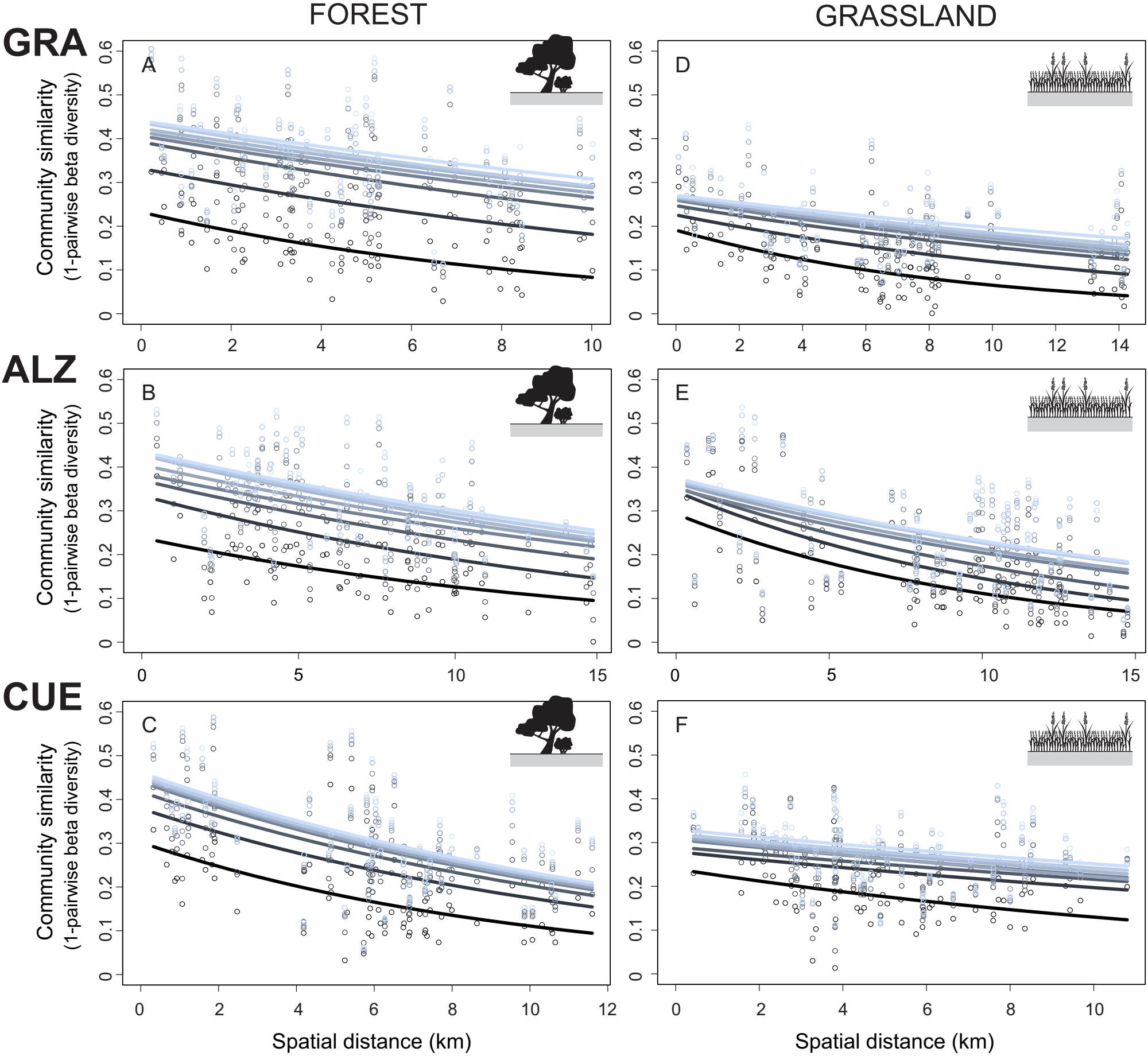
Figure 4. Distance decay of soil mesofauna community similarity at multiple levels of genetic similarity (from haplotype, black to 8% genetic similarity level, pale grey) within the three local settings (GRA, ALZ, CUE) and for forest and grassland habitats.

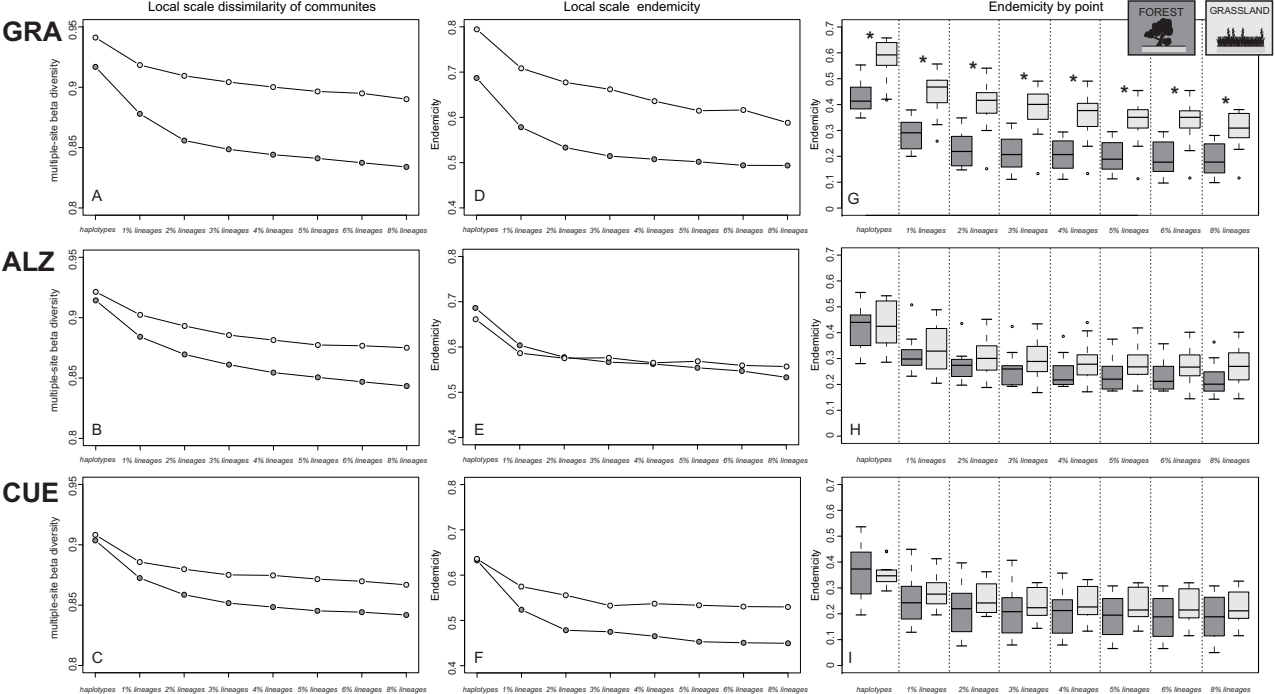
797 **Figure 5.** Dissimilarity of soil mesofauna communities (A, B, C), regional endemicity
798 (lineages present exclusively at a single sampling point in the region divided by the total
799 number of lineages found, D, E, F) and endemicity by sampling points (lineages present
800 exclusively at a particular sampling point divided by the total number of lineages in that
801 community, G, H, I) at multiple levels of genetic similarity within the three local
802 settings (GRA, ALZ, CUE) and for forest (dark grey) and grassland (pale grey) habitats.
803 Significantly different endemicity by sampling point (Wilcoxon tests $p < 0.05$) between
804 forest and grassland communities at each hierarchical level is indicated by asterisks in
805 panels G, H, I.


A
FOREST
GRASSLAND
CUE
ALZ
GRA
N
5 km
B
1m
1m
SUP
sample
30 cm
DEEP
sample
C
sample
Arthropod
bulk
community
D
dd-PCR cox1 amplification
2x300 bp sequencing
read curation
denoising
haplotype filtering
alignment
E
haplotypes
lineages under multiple
similarity thresholds
S5
S4
S3
S2
S1
F
multiple level
community matrixes
lineagesS4
lineagesS3
lineagesS2
lineagesS1
haplotypes
samples
alpha diversity
beta diversity
gamma diversity
endemicity


GRA
ALZ
CUE
haplotypes
3% lineages
haplotypes
3% lineages
A
B
C
D
E
F
richness by sample
local scale richness
deep
sup
FOREST
GRASSLAND
Coleoptera
Collembola
Acari


GRA
ALZ
CUE
haplotypes
3% lineages
FOREST
GRASSLAND
A
B
C
D
E
F
deep
sup
habitat p < 0.001 r² = 0.10
soil layer p < 0.001 r² = 0.03
habitat p < 0.001 r² = 0.09
soil layer p < 0.001 r² = 0.03
habitat p < 0.001 r² = 0.09
soil layer p < 0.001 r² = 0.04
habitat p < 0.001 r² = 0.17
soil layer p < 0.001 r² = 0.05
habitat p < 0.001 r² = 0.13
soil layer p < 0.001 r² = 0.05
habitat p < 0.001 r² = 0.12
soil layer p < 0.001 r² = 0.06


FOREST
GRASSLAND
GRA
ALZ
CUE
A
B
C
D
E
F
Community similarity
(1-pairwise beta diversity)
Spatial distance (km)
0
0.1
0.2
0.3
0.4
0.5
0.6
0
2
4
6
8
10
12
14
0
5
10
15


Local scale dissimilarity of communites
Local scale endemicity
Endemicity by point
FOREST
GRASSLAND
GRA
ALZ
CUE
multiple-site beta diversity
Endemicity
A
B
C
D
E
F
G
H
I
haplotypes
1% lineages
2% lineages
3% lineages
4% lineages
5% lineages
6% lineages
8% lineages