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

Paula Arribas[1,2,3], Carmelo Andújar[1,2,3], Antonia Salces-Castellano[1], Brent C. Emerson[1]&Alfried P. Vogler[2,3]

[1]*Island Ecology and Evolution Research Group (IPNA-CSIC), AstrofísicoFco. 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 arthropod communities are highly diverse and critical for ecosystem functioning. However, our knowledge ofspatial structure andthe underlying processes of community assembly 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 analysisof the spatial arrangementacross 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 assemblage composition differed greatly between grasslands and forests, and within each habitat showedstrong spatial structure and high endemicity. Distance-decaywashigh at all levels, even 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 similarin all regions, suggesting that uniform processes of limited dispersal determinelocal-scale community assembly. Our results from whole-community metabarcoding provide 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 currentestimations 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 these efforts have been expanded towards the biodiversity of the soil (Thakur et al., 2019). However, current knowledgeis strongly unbalanced across taxonomic groups(Cameron et al., 2018), which hampers the development of an integrative frameworkfor understanding the patterns and underlying mechanisms of soil biodiversity. In particular, there is a pronounced shortage of basic dataofspecies diversity andspatial structure for the taxonomically and functionally diverse soil arthropods.They make up a large proportion of the soilmesofauna composed of small-bodied invertebrates measuring between 0.1 – 2 mm and are found by the thousands in virtually every square meter of natural soil(Bardgett, Usher, & Hopkins, 2005; Decaëns, 2010). Conventional taxonomic approaches have been onerous, given the small body size, limited morphological variation and high local abundances of most mesofauna components. High-throughput sequencing so far has been applied mostly to microbial components of soil ecosystems (e.g. Delgado-Baquerizo et al., 2018; Ramirez et al., 2018, 2014)while 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).

Existing work on the diversity,distribution and community composition of soil arthropodshas 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 alsoreportedpurely spatial structures (independent of the measured environmental variables) orstochastic patterns (non-environmental neither spatial structures) for the soil mesofauna communities, that have been recurrentlyattributed to the contribution ofdemographic 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). However, dispersal limitation 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 candetermine 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 reflecting long-term population separation. Similar patterns of species and haplotype turnover can arise even over relatively small distancesif the scale of movement is highly constrained and if the constraints are persistent through time, as could be the case in the soil matrix. 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 communitiesmay have missedthe importance of dispersal limitation in determining the diversity patterns of soil mesofauna (but see Andújar et al., 2015; Lindo & Winchester, 2009).

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

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

approximate community profiles at the species-level, precise removal of primary DNA reads affected by sequencing errors (Andújar, Arribas, Yu, Vogler, & Emerson, 2018; Elbrecht, Vamos, Steinke, & Leese, 2018; Turon, Antich, Palacín, Præbel, & Wangensteen, 2019)and co-amplified nuclear mitochondrial copies (numts) (Andújar et al., 2020)would avert the need for clustering.Read-based dataraise the prospect of reliable haplotype information from mitochondrial COIcMBC, which represents a step change for the study of diversity patterns throughwhole-community genetic analyses at haplotype-level resolution.

Haplotype datacan be used directly for analyses of genetic diversity, or after aggregation into species-level entities for analyses of species diversity, whichpermitsthe joint analysis of turnover (beta diversity) at multiple hierarchical levels. This approach has been exploited to determine whether the composition in biological assemblages is predominantly driven by dispersal or niche-based processes (Baselga et al., 2013; Baselga, Gómez-Rodríguez, & Vogler, 2015). Local assemblages may diverge simply due to the lack of population movement which, when assessed for entire communities,results in a largely regular decay of community similarity with spatial distance for the typically neutral haplotype variation of the mitochondrial COI gene. Under ascenario where dispersal constraints determine the spatial community structure, assemblage turnover at the species level should mirror these haplotype patterns, albeit at a higher level of similarity. In contrast, niche-based processes acting on species traits produce species distributions that mainly follow environmental factors and thus differ from neutral conditions determining the haplotype distributions. This confounds the correlation (self-similarity)of distance decay at the species and haplotype levels, as each is driven by different processes. The self-similarity of distance decay of communities at genetic and species levels therefore provides a formal test to discern if a particular

spatial pattern of community assemblage is predominantly driven by stochastic dispersal or niche-based processes, as the latter will not usually produce this correlation(Baselga et al., 2013, 2015). In addition, multi-hierarchical analyses may also describe the spatial scale at which dispersal constraints act, and the variation of scale among different taxonomic groups or habitats (Gómez-Rodríguez, Miller, Castillejo, Iglesias-Piñeiro, & Baselga, 2018; Múrria et al., 2017). This framework remains to be exploited with whole-community metabarcoding.

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

MATERIALS AND METHODS

Soil sampling and mesofauna extraction

A total of 144 soil samples were collected from three regions in the southern Iberian Peninsula at Sierra de Grazalema, (GRA), Sierra de Alatoz (ALZ) and Sierra de la

AlcarriaConquense (CUE) (Fig. 1 A). In each region, 24 points were sampled, half of them in *Quercus* forest and half in wet grassland habitat, at distances of 500 m to a 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 digging the substrate of a 30 cm diameter core to 30 cm depth, comprising ca. 20 litres of soil.Samples were sifted in the field (1cm wire mesh sieve) to remove the biggest vegetation fragments and stones, and subsequentlyprocessed following the flotation–Berlese–flotation protocol (FBF) of Arribas et al. (2016) (see below for further details).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).

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

Bioinformatics read processing

Raw reads were quality checked in Fastqc(Babraham Institute, 2013). Primers were trimmed using fastx_trimmer and reads were processed in Trimmomatic(Bolger, Lohse, & Usadel, 2014) using TRAILING:20. Based on results from (Andújar, Arribas, Gray, et al., 2018) on the test of multiple tools and parameters for diverse metazoan metabarcoding samples, we further processed each library independently following several steps of the Usearch(Edgar, 2013) pipeline: reads were merged (option mergepairs – -fastq_minovlen50, -fastq_maxdiffs 15), quality-filtered (Maxee = 1), trimmed to full length amplicons of 418 bp (-sortbylength), dereplicated (-fastx_uniques) and denoised (-unoise3, -minsize 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 assignment 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 7.1.9(https://www.geneious.com/) using MAFFT and the Translation Align options, and those with insertions, deletions or stop codons disrupting the reading frame were identified and subsequently excluded.

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., 2020). The set of putative haplotypes for Acari, Collembola and Coleopterawas 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 reads while maximizing the number of real haplotypes to be further analysed (see Andújar et al., 2020 for details). Community tables of fully filtered haplotypes were then transformed into incidence (presence/absence) data, that added to the haplotype filtering before, resulting in normalised samples for further analyses.

Analysis of community composition and assembly at multiple thresholds of genetic similarity

The analyses were performed using the R-packages vegan (Oksanen et al., 2013), cluster, PMCMR, hier.part, ecodist, and betapart(Baselga & Orme, 2012). The set of filtered haplotypes was used to generate a UPGMA tree with corrected genetic distances (F84 model), and based on this tree all haplotypes were grouped into clusters of genetic similarity at different thresholds (1%, 2%, 3%, 4%, 5%, 6% and 8%). This grouping procedure based on patristic pairwise distances over a phylogenetic tree including all haplotype sequences provided multiple hierarchical levels that each can be used to estimate alpha diversity (Figure 1 provides a graphical abstract of the workflow). These

diversity measures were estimated for the richness of lineages by sample for the whole mesofauna community and also for the subsets corresponding to Acari, Collembola and Coleoptera. To test for significant differences in alpha diversity between the communities of different habitats and soil layers of each sampling point, repeated-measures ANOVAs were conducted using habitat and soil layer as grouping factors and sampling point as a within-subjects factor. For each of the three local settings, total accumulative richness (local scale richness) by habitat and soil layer and the contribution of mites, springtails and beetles was also calculated for the various levels of genetic similarity. Endemicity by sampling point was computed for each hierarchical level (once DEEP and SUP samples were combined) as the lineages present exclusively at a single sampling point in the region divided by the total number of lineages found in the region. To assess whether the endemicity by sampling point differed between the communities of forests and grasslands, Wilcoxon tests were conducted using habitat as a grouping 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).Additional NMDS ordinations and permutational ANOVAs were also conducted for each taxonomic group using the same parameters as given before.

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.

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 by removal of likely nuclear mitochondrial pseudogenes generated a conservative set of clean sequences representing the mitochondrial haplotypes.A total of 1124, 1009 and 992 haplotypes where found for the GRA, ALZ and CUE local areas respectively, and these numbers declined rapidly when haplotypes were grouped at increasing threshold values, e.g. 511, 479 and 480 lineages at 3% similarity (Table 1), but they declined only slightly further at the higher thresholds, indicating the point at which stable groups are obtained that broadly could represent the species level. The relative proportions of mites, springtails and beetles were similar across the three local settings and hierarchical levels, with Acari representing the richest group (around the 50% of clusters) followed by Collembola and Coleoptera in similar proportions (Table 1). The taxonomic composition of the samples as estimated by MEGAN is provided in Fig. S1 and included a total of 39, 44 and 40 families of Acari, 11, 9, and 10 families of

342 Collembola and 18, 23 and 20 families of Coleoptera for the GRA, ALC and CUE local
343 settings respectively.

344 The patterns of richness by sample (alpha diversity) for the different habitats,
345 soil layers and genetic thresholds were similar for the three regions, with mean values
346 between 35 - 60 haplotypes and 25 - 42 lineages at 3% per sample (Fig. 2 A, B, C and
347 Fig. S2). Superficial soils had significantly higher diversity than their corresponding
348 deep soil counterparts for the overall dataset and for both of the forest and grassland
349 habitats assessed independently (Fig. 2 A, B, C, Fig S2 and Table S3). At GRA, forest
350 habitat showed significantly higher alpha diversity per sample than grassland. However,
351 no significant differences between forest and grassland were found for ALZ and CUE
352 (Fig. 2 A, B, C and Fig S2, Table S3).

353 The local-scale cumulative richness at each region (gamma diversity) showed
354 more diverse communities for the superficial compared with deep layers, and gamma
355 diversity was generally higher for forest than grassland habitats, but the differences
356 between both habitat types were lower than observed for alpha diversity. Thus, species
357 accumulation was higher for the grassland than forest habitats, and the grassland
358 superficial layers had the highest total richness of haplotypes for ALZ and CUE, and the
359 highest for the three regions at the 3% similarity level (Fig. 2 D, E, F and Fig S3).
360 Patterns of alpha and gamma diversity for the subsets of mites, springtails and beetles
361 were similar (Fig. S3 and S4).

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

363 Compositional dissimilarity of communities within each of the three regions was high
364 and was dominated by lineage turnover β_{sim}, instead of nestednessβ_{sne} ($0.8 > \beta_{sim} > 0.95$),
365 across all hierarchical levels. NMDS showed a consistent pattern of the forest and
366 grassland habitats as the main driver of the ordination while soil layers had a secondary

role (Fig. 3, Fig. S5). 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. However, even in these two significant cases, the variance partitioning showed that uniquely explained variance in environmental distance, i.e. independently of the spatial distance, was low (5 – 9 % of explained variation at all levels) compared with the uniquely explained variance in spatial distance (23 – 31 % of explained variation).

Theendemicitywithin 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 and was 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, although 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 putative species (lineages at 3%) and nearly 3000 haplotypes of mites, springtails and beetles.Their distribution was determined across numerous sampling points, demonstrating the power of mitochondrial cMBC to overcome impediments to studying the arthropod mesofauna of the soil using conventional morphological and molecular approaches. Data analysis revealed a strong spatial community structure and high levels of endemicity at haplotype, species and supra-specific levels, even at sampling points that were mostly within a few kilometres of each other (maximum 15 km). Patterns of turnover and endemicitywere similarin all three independent study regions and in the grassland and forest biomes (that each harbour largely non-overlapping communities). Distance decay is evident at all hierarchical levels, and can be described as self-similar.

417 The correlation of community turnover at population and species levels is expected if
418 soil arthropod assemblages are predominantly driven by distance-based parameters,
419 andmovement is stronglyconstrained at the local scale and over time. In addition, the
420 overriding importance of habitat-related processes was apparent from the strong
421 differentiation of grassland and forest communities, which again was seen recurrently in
422 each of the three study regions.

423 The study extends existing comparative analyses of soil mesofaunaby improving
424 the taxonomic resolution, providing haplotype level variation, and analysing a wide
425 range of soil arthropods together. Broad surveys of invertebrate soil diversity using HTS
426 have commonly relied on markers of low species-level resolution and via eDNA
427 extracted from small soil samples(Bahram et al., 2016; Wu et al., 2011; Zinger et al.,
428 2019).Other studies have characterised specific groups of mites or springtails by
429 processing ofindividualised specimens and relying on morphological assignment to
430 generate species-level data (Caruso, Schaefer, Monson, & Keith, 2019; Ingimarsdóttir et
431 al., 2012), but see alsoYoung, Proctor, DeWaard, & Hebert(2019) on molecular species
432 assignment. HTS data now greatly increase the potential of expanding both the number
433 of species studied and the level of detail at which intra-specific variation for each is
434 captured. Our study provides haplotype level data for entire communities (one square
435 meter of leaf litter and humus and ca. 20 litres of soil per sampling point) of the three
436 most species-rich soil arthropods, which allows surveys of community composition and
437 species turnover at an unprecedented level of detail, both spatially and genetically.
438 While the correlation of COI divergence with species boundaries can vary greatly
439 between taxa, we analysed our dataset using an extensive range of hierarchical
440 thresholds (from 1 to 8% genetic similarity thresholds) as a simplified but conservative
441 approach to consider both species and supra-specific levels. We found some degree of

442 stabilization in the number of lineages defined above the 3% similarity grouping,
443 consistent with this similarity threshold as broadly representing the species level.Using
444 these data, community level responses to distance-based parameters can be assessed that
445 may be not evident in other types of studies. In addition, the combined haplotype and
446 species-level data permit the exploitation of the hierarchical framework of Baselga et al.
447 (2013, 2015) for discriminating between distance- and niche-based factors of
448 community assembly.

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

450 Existing literatureexploring the local community composition of arthropod
451 mesofaunagenerallyhas argued for selection by abiotic and/or biotic environmental
452 factors as the predominant mechanisms (seeBerg, 2012; Thakur et al., 2019 for a recent
453 review).Stochastic and purely spatial patterns have also been reported, pointing to a
454 contribution of dispersal and demographic processes at least in some local
455 settings(Caruso et al., 2012; Gao et al., 2014; Gao, Liu, Lin, & Wu, 2016; Zinger et al.,
456 2019).However, strong dispersal constraints have rarely been recognised, in part due to
457 the lower taxonomic resolution of previous community-level studies that used species
458 assignments from morphological or 18S rRNAdata (see Tang et al., 2012). Our results
459 demonstrate high community differentiation at the kilometre scalefor both genetic and
460 species levels. The key observation from the multi-hierarchical analysis is the correlated
461 distance decay at haplotype and species level. Self-similarity is expected to be eroded
462 byselection on adaptive traits at the species level, but not at the (neutral) haplotype level
463 (Baselga et al., 2015; Gómez-Rodríguez et al., 2018）. As the data largely confirm the
464 self-similarity of distance decay at haplotype and species level, this is interpreted to
465 support the predominant role of dispersal limitation driving community assembly. The
466 predominance of the dispersal constraints seems to emerge at short spatial distances

within the soil matrix, and the evident high turnover with physical distancesuggests that our sampling within each study regions (local scale) is beyond the scale of a single metacommunity. Short dispersal distancesprobably have affected a significant proportion of lineages within these communities over evolutionary timescalesin a largely stable spatial setting.The spatiotemporal continuum expected under this scenario predicts that lineages in more distant places have diverged at a more distanttime 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 distancedriving the local community assembly comes from the high microendemicityfound at all hierarchical levels, an overall picture which is not expected under a scenario with predominant environmental driversnor ecological drift without dispersal limitation.

Yet, the influence of environmental drivers cannot be discarded entirely. Whereas the recorded environmental variables did not explain the variation in community composition, a significant portion of the unexplained variance in the distance-decay curves potentiallysuggests 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, this possibility is poorly supported here, as similar spatial structures were independently found within the different habitats and regions, mirroring thedistance-decay patterns at the (neutral) haplotype level, and

hence suggesting that dispersal limitation is the main driver of the localspatial structure of the studied mesofauna communities.

The small spatial scale of turnover and endemicityis consistent with population genetic studies in soil mesofauna showing deep genetic breaks even over relatively short geographic distances (Andújar, Pérez-González, et al., 2017; Cicconardi et al., 2013; Collins, Hogg, Convey, Barnes, & McDonald, 2019). In contrast, our results are not concordant with long-distance dispersal as aprevalentprocess for soil mesofauna, as might be expected frompassive 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 recentlydeglaciatedareas (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 (Collins, Hogg, Baxter, Maggs-Kölling, & Cowan, 2019; 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 withpatterns 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 theoverall arthropod diversity(for which soil mesofauna comprises the majority of the smallest fraction), our results are not supporting the macroecologicalpredictionfora 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 ofecological and evolutionary processes driving the biodiversity ofthese 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 important role of dispersal limitation within each habitat type, the greatest assemblage differentiation was between 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, although 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 presumably are

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 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 needed to identify the processes driving the mesofaunal community turnover, but both habitat types show the signature of long-term stability 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

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 a majority of species 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 different overall settings, 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 processesreported 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 other ecoregions, the very reduced spatial scale of dispersal in

soil mesofauna could be a major contribution to the overall gamma diversity and may lead to a revised estimate of total species 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 macrobialcommunity structure into a new era of biodiversity discovery. By taking advantage of the full breadth of contemporary metabarcoding data at expanded taxonomic and geographic scales,the advances made here will 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 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 CiervaFormación Program. CA was supported by the Spanish Ministry of Economy and Competitiveness (MINECO, Spain) (CGL2015-74178-JINMINECO/FEDER, UE). 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 JesúsArribas who assisted with field sampling.

588 **REFERENCES**

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

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

596 Andújar, C., Arribas, P., & Vogler, A. P. (2017). Terra incognita of soil biodiversity:
597 unseen invasions under our feet. *Molecular Ecology*, *26*(12), 3087–3089. doi:
598 10.1111/mec.14112

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

602 Andújar, C., Creedy, T. J., Arribas, P., López, H., Salces-Castellano, A., Pérez-Delgado,
603 A. J., … Emerson, B. C. (2020). NUMT dumping : validated removal of nuclear
604 pseudogenes from mitochondrial metabarcode data. *BioRxiv*.

605 Andújar, C., Pérez-González, S., Arribas, P., Zaballos, J. P., Vogler, A. P., & Ribera, I.
606 (2017). Speciation below ground: Tempo and mode of diversification in a radiation
607 of endogean ground beetles. *Molecular Ecology*, *26*(21), 6053–6070. doi:
608 10.1111/mec.14358

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Collins, G. E., Hogg, I. D., Baxter, J. R., Maggs-Kölling, G., & Cowan, D. A. (2019). Ancient landscapes of the Namib Desert harbor high levels of genetic variability and deeply divergent lineages for Collembola. *Ecology and Evolution*, *9*(8), 4969–4979. doi: 10.1002/ece3.5103

Collins, G. E., Hogg, I. D., Convey, P., Barnes, A. D., & McDonald, I. R. (2019). Spatial and temporal scales matter when assessing the species and genetic diversity of springtails (Collembola) in Antarctica. *Frontiers in Ecology and Evolution*, *7*(MAR), 1–18. doi: 10.3389/fevo.2019.00076

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

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

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

709 Gómez-Rodríguez, C., Miller, K. E., Castillejo, J., Iglesias-Rieiro, J., & Baselga, A.
710 (2018). Understanding dispersal limitation through the assessment of diversity
711 patterns across phylogenetic scales below the species level. *Global Ecology and*
712 *Biogeography*, (October), 1–12. doi: 10.1111/geb.12857

713 Hortal, J., Roura-Pascual, N., Sanders, N. J., & Rahbek, C. (2010). Understanding
714 (insect) species distributions across spatial scales. *Ecography*, *33*(1), 51–53. doi:
715 10.1111/j.1600-0587.2009.06428.x

716 Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of
717 metagenomic data. *Genome Research*, *17*(3), 377–386. doi: 10.1101/gr.5969107

718 Ingimarsdóttir, M., Caruso, T., Ripa, J., Magnúsdóttir, O. B., Migliorini, M., &
719 Hedlund, K. (2012). Primary assembly of soil communities: disentangling the
720 effect of dispersal and local environment. *Oecologia*, *170*(3), 745–754. doi:
721 10.1007/s00442-012-2334-8

722 Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., … Yu, D. W.
723 (2013). Reliable, verifiable and efficient monitoring of biodiversity via
724 metabarcoding. *Ecology Letters*, *16*(10), 1245–1257. doi: 10.1111/ele.12162

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

728 Magilton, M., Maraun, M., Emmerson, M., & Caruso, T. (2019). Oribatid mites reveal
729 that competition for resources and trophic structure combine to regulate the
730 assembly of diverse soil animal communities. *Ecology and Evolution*, *9*(14), 8320–
731 8330. doi: 10.1002/ece3.5409

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

737 Nkem, J. N., Wall, D. H., Virginia, R. A., Barrett, J. E., Broos, E. J., Porazinska, D. L.,
738 & Adams, B. J. (2006). Wind dispersal of soil invertebrates in the McMurdo Dry
739 Valleys, Antarctica. *Polar Biology*, *29*(4), 346–352. doi: 10.1007/s00300-005-
740 0061-x

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

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

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

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

757 Ribera, I., & Vogler, A. P. (2000). Habitat type as a determinant of species range sizes:
758 the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of*
759 *the Linnean Society*, *71*(1), 33–52. doi: 10.1111/j.1095-8312.2000.tb01240.x

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

762 Schuppenhauer, M. M., Lehmitz, R., & Xylander, W. E. R. (2019). Slow-moving soil
763 organisms on a water highway: Aquatic dispersal and survival potential of
764 Oribatida and Collembola in running water. *Movement Ecology*, *7*(1), 1–14. doi:
765 10.1186/s40462-019-0165-5

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

768 Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, a., Barraclough, T. G., & Fontaneto,
769 D. (2012). The widely used small subunit 18S rDNA molecule greatly
770 underestimates true diversity in biodiversity surveys of the meiofauna.
771 *Proceedings of the National Academy of Sciences*, *109*(40), 16208–16212. doi:
772 10.1073/pnas.1209160109

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

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

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

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

785 Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H., & Garey, J. R. (2011). Molecular study
786 of worldwide distribution and diversity of soil animals. *Proceedings of the*
787 *National Academy of Sciences of the United States of America*, *108*(43), 17720–

17725. doi: 10.1073/pnas.1103824108

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

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

DATA ACCESSIBIITY STATEMENT

Raw metabarcode data deposited in the Dryad repository: https://doi.org/10.5061/dryad.vx0k6djph; primers and PCR conditions in Table S2; all supplementary tables and figures cited in the main text have been uploaded as online Supporting Information.

AUTHOR CONTRIBUTIONS

Statement of authorship: P.A., C.A. and A.P.V. conceived the work; P.A. and C.A. collected and analysed the data; P.A led the writing and all authors contributed to the discussion of results and the writing.

807 **Table 1** Total number of haplotypes and clusters for Acari, Collembola and Coleoptera
808 for each local setting (GRA, ALZ, CUE) at increasing levels of genetic divergence
809 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**

810

811

812 **FIGURE CAPTIONS**

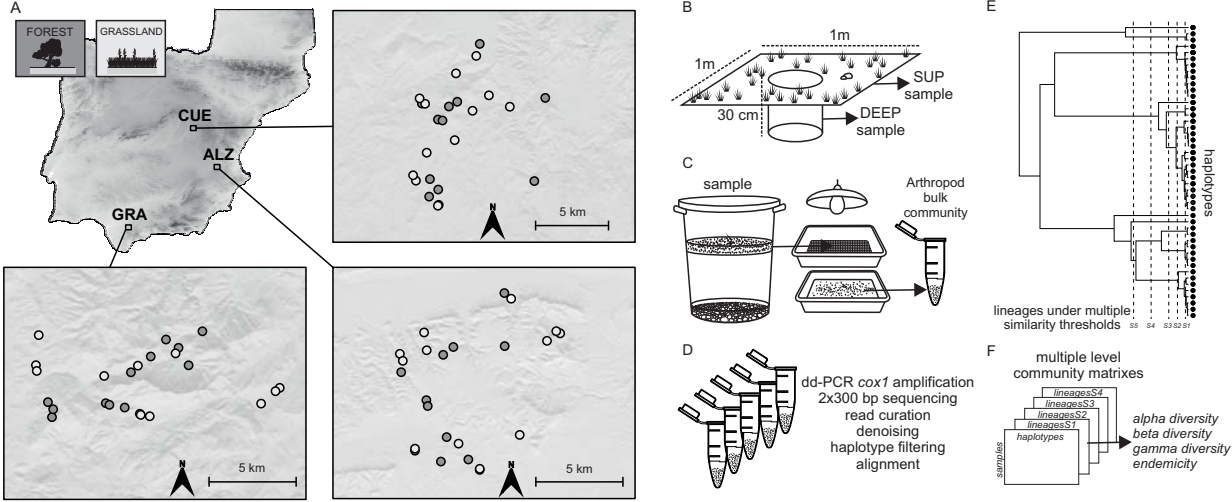
813 **Figure 1.** Sampling points in the three local settings within the Iberian Peninsula, Sierra
814 de Grazalema, (GRA), Sierra de Alatoz (ALZ) and Sierra de la AlcarriaConquense
815 (CUE). Sampling points are located within *Quercus* forest patches (dark grey) and wet
816 grassland patches (pale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

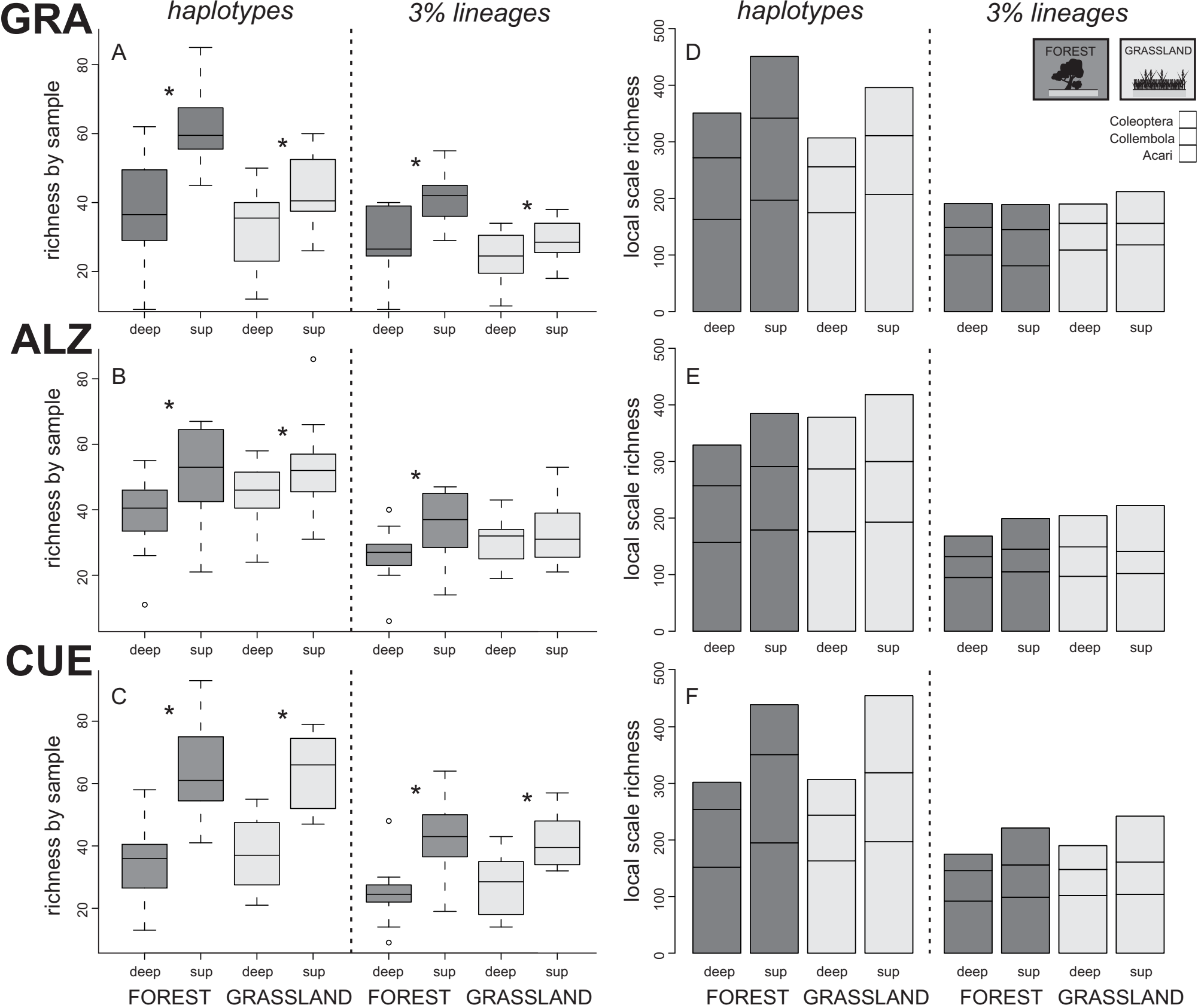
817 **Figure 2.** Richness of soil mesofauna lineages by sample (alpha diversity, A, B, C) and
818 total accumulated richness (local scale richness, D, E, F) by habitat and soil layer for the
819 three local settings (GRA, ALZ, CUE). Both measures are shown at the haplotype and
820 the 3% genetic similarity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

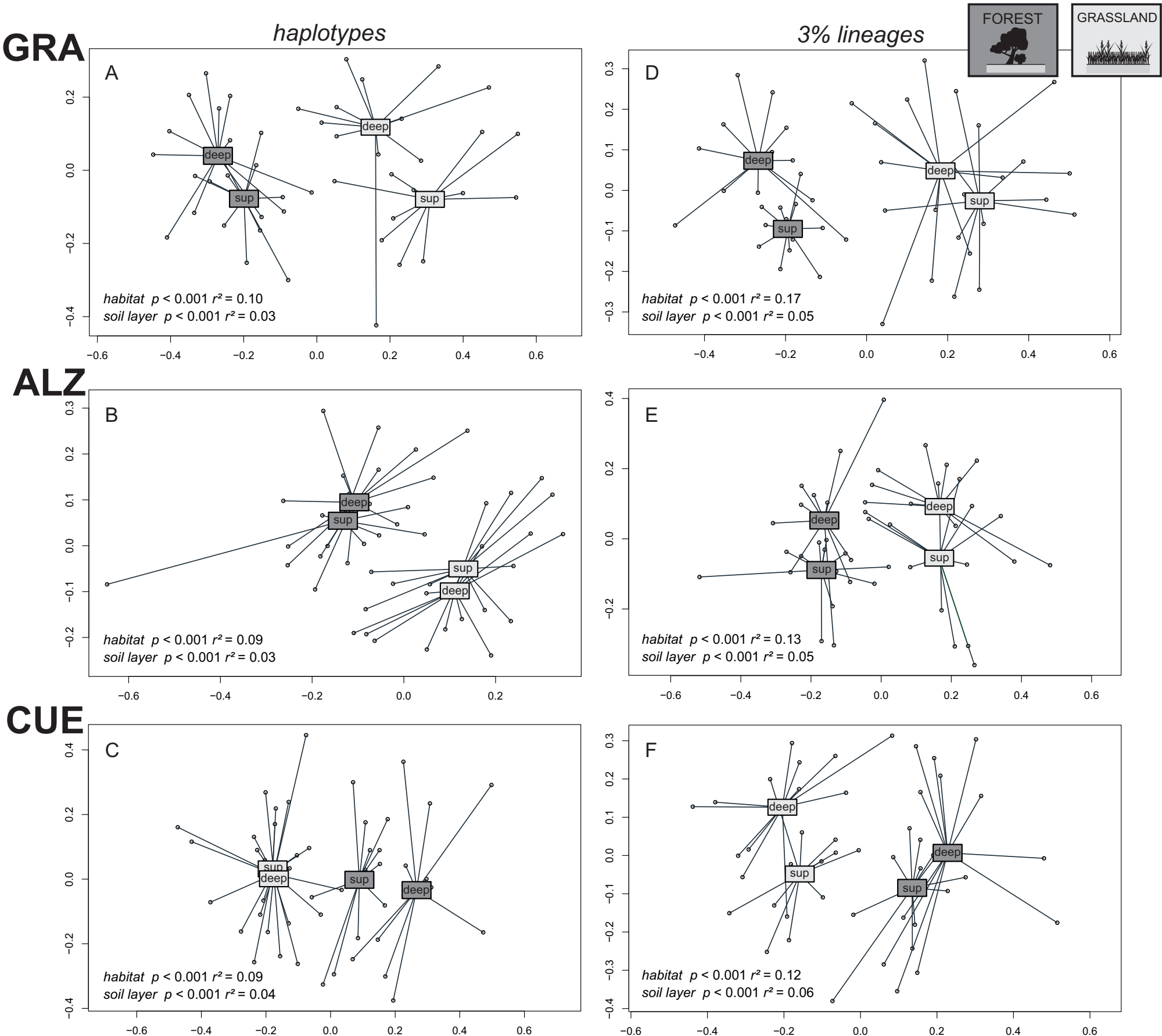
826 **Figure 3.** NMDS ordinations of the soil mesofauna samples according to the variation
827 in community composition (Simpson index, βsim) within the three local settings (GRA,
828 ALZ, CUE) and at the haplotype and the 3% genetic similarity levels. Forest habitat as
829 dark grey, grassland habitat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

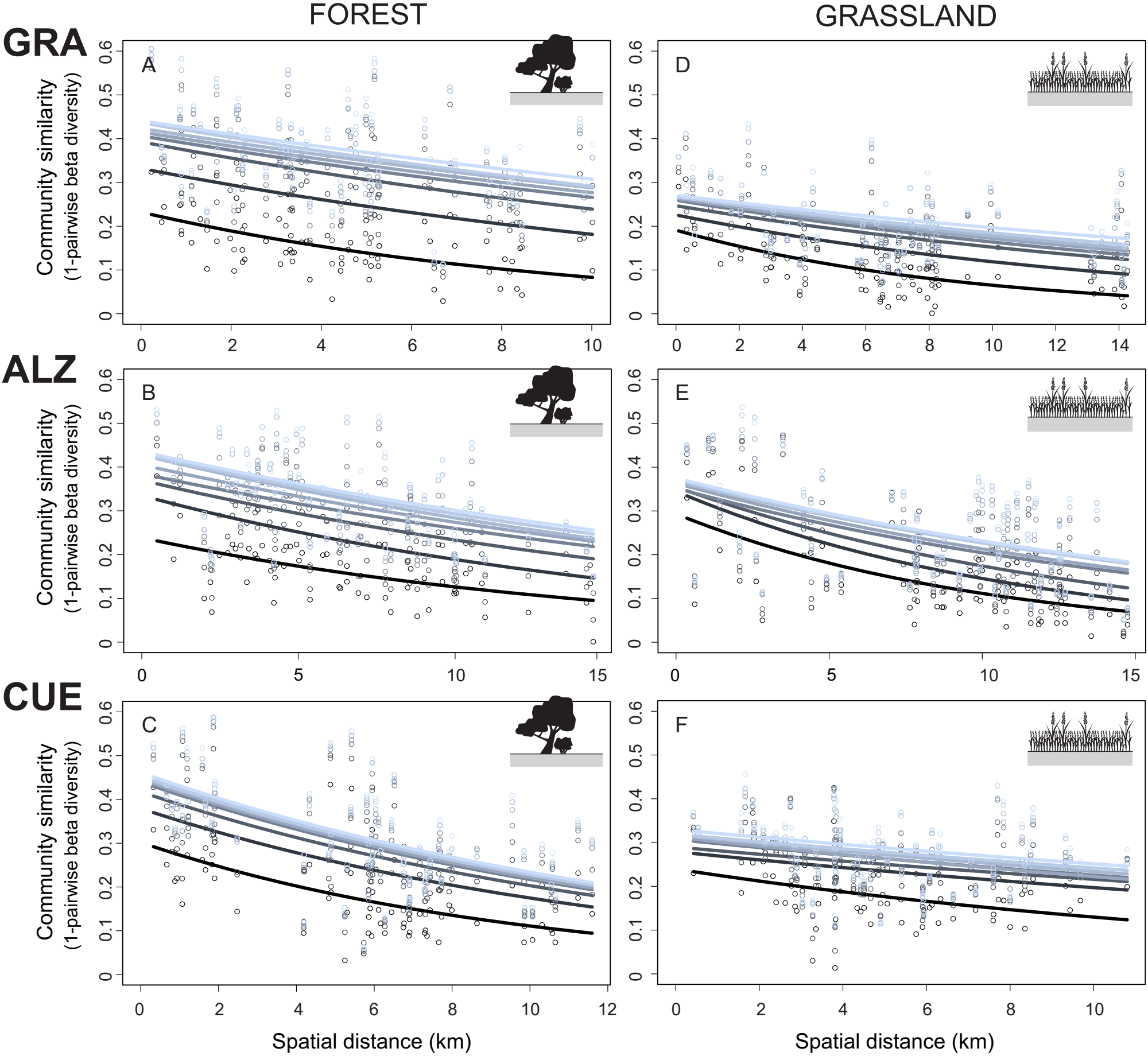
832 **Figure 4.** Distance decay of soil mesofauna community similarity at multiple levels of
833 genetic similarity (from haplotype, black to 8% genetic similarity level, pale grey)
834 within the three local settings (GRA, ALZ, CUE) and for forest and grassland habitats.

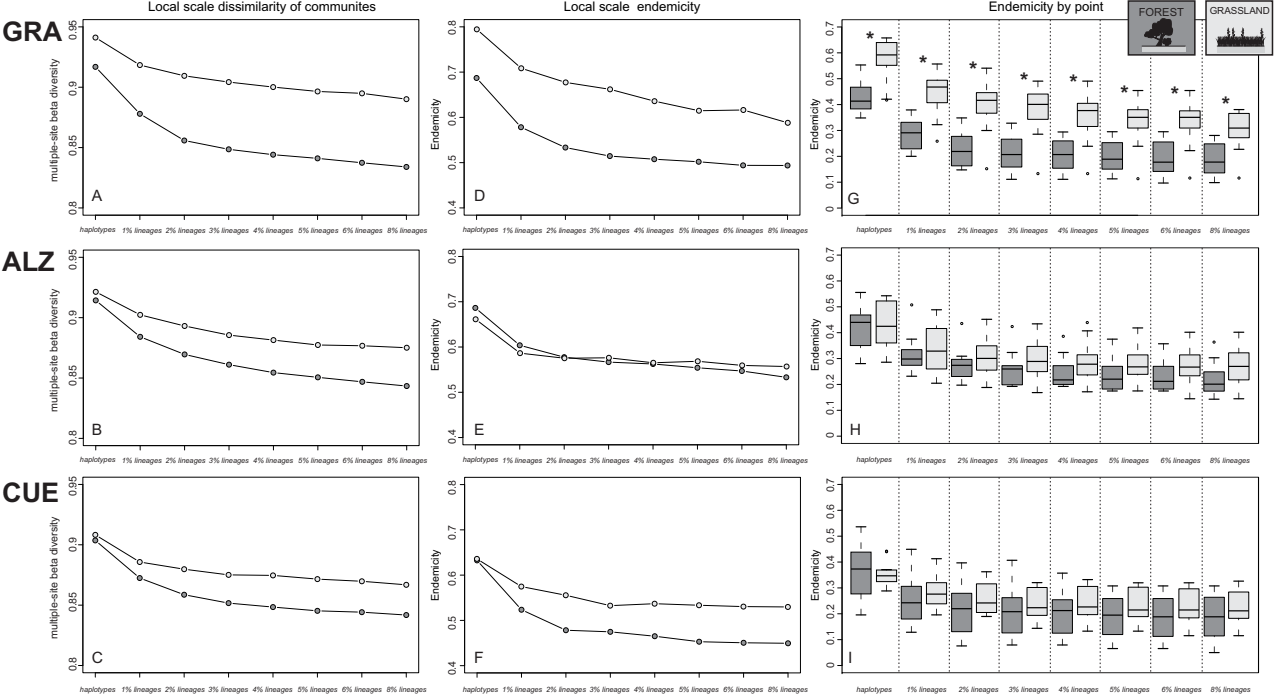
Figure 5. Dissimilarity of soil mesofauna communities (A, B, C), regional endemicity (lineages present exclusively at a single sampling point in the region divided by the total number of lineages found, D, E, F) and endemicity by sampling points (lineages present exclusively at a particular sampling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