Community metabarcoding reveals the relative role of environmental filtering and dispersal in metacommunity dynamics of soil microarthropods across a mosaic of montane forests

Victor Noguerales[1], Emmanouil Meramveliotakis[1], Adrián Castro-Insua[1], Carmelo Andujar[2], Paula Arribas[2], Thomas Creedy[3], Isaac Overcast[4], Hélène Morlon[4], Brent Emerson[2], Alfried Vogler[3], and Anna Papadopoulou[1]

[1]University of Cyprus
[2]Instituto de Productos Naturales y Agrobiología
[3]Natural History Museum
[4]Institut de Biologie de l'ENS (IBENS)

June 30, 2021

Abstract

Disentangling the relative role of environmental filtering and dispersal limitation in driving metacommunity structure across mountainous regions remains challenging, as the way we quantify spatial connectivity in topographically and environmentally heterogeneous landscapes can influence our perception of which process predominates. More empirical datasets are required to account for taxon- and context-dependency but relevant research is often compromised by coarse taxonomic resolution. We here employed haplotype-level community DNA metabarcoding, enabled by stringent filtering of Amplicon Sequence Variants (ASVs), to characterize metacommunity structure of soil microarthropod assemblages across a mosaic of five forest habitats on the Troodos mountain range in Cyprus. We found similar β diversity patterns at ASV and species (OTU, Operational Taxonomic Unit) levels, which pointed to a primary role of habitat filtering resulting in the existence of largely distinct metacommunities linked to different forest types. Within-habitat turnover was correlated to topoclimatic heterogeneity, again emphasizing the role of environmental filtering. However, when integrating landscape matrix information for the highly fragmented Golden Oak habitat, we also detected a major role of dispersal limitation imposed by patch connectivity, indicating that stochastic and niche-based processes synergistically govern community assembly. Alpha diversity patterns varied between ASV and OTU levels, with OTU richness decreasing with elevation and ASV richness following a longitudinal gradient, potentially reflecting a decline of genetic diversity eastwards due to historical pressures. Our study demonstrates the utility of haplotype-level community metabarcoding for characterising metacommunity structure of complex assemblages and improving our understanding of biodiversity dynamics across mountainous landscapes worldwide.

INTRODUCTION

Understanding the drivers of metacommunity structure across heterogeneous landscapes remains a fundamental question in ecology (Meynard et al., 2013; Viana & Chase, 2019). Under a niche selection perspective (Chase & Leibold, 2003; Hutchinson, 1959) metacommunity structure results from species sorting via environmental filtering, while under the neutral paradigm (Chase, 2005; Hubell, 2001) metacommunities are structured by stochastic dispersal and ecological drift. These processes are not mutually exclusive, as their relative importance in community assembly varies along a niche-dispersal continuum (Brown, Sokol, Skelton, & Tornwall, 2017; Cottenie, 2005; Gravel, Canham, Beaudet, & Messier, 2006). Scale dependency has been suggested to affect the perception of which process predominates, with environmental filtering prevailing at larger spatial scales and the role of stochastic processes increasing at finer scales (Viana & Chase, 2019).

1

Yet, the apparent prevalence of environmental over spatial processes at regional scales (Chase, 2014) could be biased by the tendency of most metacommunity studies to quantify dispersal as a simple function of geographic distance (Biswas & Wagner, 2012), without considering landscape features such as topography or matrix heterogeneity. Accounting for matrix resistance when quantifying connectivity (McRae, Dickson, Keitt, & Shah, 2008) may enforce the role of dispersal as the primary driver of metacommunity structure (*e.g.*, Resasco & Fletcher, 2021). This is particularly relevant when focusing on montane regions, where spatial connectivity among habitat patches can be strongly influenced by high topographic complexity and environmental heterogeneity due to steep elevational gradients (Graf, Kramer-Schadt, Fernández, & Grimm, 2007; Liu, Dudley, Xu, & Economo, 2018). At the same time, elevational gradients in topoclimatic parameters such as temperature or precipitation can impose strong environmental filtering on montane metacommunities (Hoiss, Krauss, Potts, Roberts, & Steffan-Dewenter, 2012; Leingärtner, Krauss, & Steffan-Dewenter, 2014) and are often considered to drive patterns of species richness (Peters et al., 2016; Rahbek, 1995) and community uniqueness (Wang et al., 2020) at regional scales. As montane communities are rapidly changing due to rising temperatures and anthropogenic disturbance (Rahbek et al., 2019; Steinbauer et al., 2018), it is crucial to gain a better understanding of how elevation and landscape mediate the interplay between environmental filtering and dispersal limitation as drivers of metacommunity structure across mountainous regions (*e.g.*, Gálvez-Reyes et al., 2020).

Disentangling the drivers of metacommunity structure requires comprehensive empirical datasets from different functional groups, geographic regions and landforms, as the relative importance of spatial and environmental constraints is expected to be taxon- and context-dependent (He et al., 2020; Tonkin et al., 2018). It is thus difficult to extrapolate conclusions from the limited number of well-characterised montane metacommunities that have been studied to date (*e.g.*, Benito, Fritz, Steinitz-Kannan, Vélez, & McGlue, 2018; Brodie & Newmark, 2019; Tonkin et al., 2017). More empirical studies from a diversity of montane biota across the globe are required to understand what general principles may be at play. Expanding metacommunity research to understudied geographic areas, and/or poorly known taxonomic groups, is complicated by the "taxonomic impediment" (Cicconardi, Fanciulli, & Emerson, 2013; Young, Proctor, deWaard, & Hebert, 2019), and has been often compromised by coarse taxonomic resolution (He et al., 2020; Verleyen et al., 2009). The development of metabarcoding provides new opportunities to accelerate studies on metacommunity structure across underexplored fractions of biodiversity and greatly increases their taxonomic resolution (Arribas, Andújar, Salces-Castellano, Emerson, & Vogler, 2021b; Bush et al., 2020; Martin et al., 2021; Zinger et al., 2019). Recent advances in field, laboratory and bioinformatic protocols for whole organism community DNA (wocDNA, Creedy et al., 2021) metabarcoding have led to improvements in both the efficiency for the generation of such high-resolution taxonomic inventories (Arribas, Andújar, Hopkins, Shepherd, & Vogler, 2016; Elbrecht, Vamos, Steinke, & Leese, 2018), and the reliability of α and β diversity estimates (Andújar et al., 2018; Creedy, Ng, & Vogler et al., 2019). New bioinformatic tools for the removal of noise generated by amplification and sequencing errors (*e.g.*, Callahan et al., 2016; Edgar, 2016) and the filtering of spurious sequences resulting from co-amplification of nuclear mitochondrial pseudogenes (Andújar et al., 2021) allow to move beyond classical OTU (Operational Taxonomic Unit) clustering and define haplotype-level entities, or Amplicon Sequence Variants (ASVs; Callahan, McMurdie, & Holmes, 2017). In contrast to OTUs, ASVs have intrinsic biological significance and offer the possibility for direct comparisons among studies that use the same marker (Callahan et al., 2017; Porter & Hajibabaei, 2020), while they potentially improve community diversity estimates (Joos et al., 2020). The availability of reliable whole-community ASV and OTU datasets allows comparisons of spatial structure at haplotype and species levels, which can provide insights into the prevalence of stochastic *vs.* niche-based processes in community dynamics, as similar diversity patterns at both levels are predicted under neutral scenarios (Papadopoulou et al., 2011; Baselga, Gómez-Rodríguez, & Vogler, 2015). All the latest advances in the field of wocDNA metabarcoding are expected to lead to a better understanding of species diversity and community processes, particularly in historically intractable habitats, such as the soil (Arribas et al., 2021b).

Soil biodiversity is among the most complex and poorly known terrestrial biotas on Earth (Decaëns, 2010). The high structural complexity and heterogeneity of the edaphic environment are thought to facilitate species

coexistence and drive patchy distributions of soil organisms at multiple spatial scales (Berg, 2012; Thakur et al., 2020), providing a particularly interesting template for metacommunity studies. Soil microarthropods, including the highly abundant and diverse groups of Acari, Collembola and Coleoptera, represent a major component of below-ground communities, with a broad range of functional roles in soil ecosystem services (Nielsen, 2019) and high levels of cryptic diversity (*e.g.*, Cicconardi et al., 2013; Young et al., 2019). Recent metabarcoding studies (Arribas et al., 2021b; Zinger et al., 2019) have revealed an important role of stochastic processes and dispersal limitation in community assembly of soil microarthropods at within-habitat scale, in contrast to previous research that had emphasized environmental filtering in response to soil attributes as a major driver of community composition (*e.g.*, Caruso, Schaefer, Monson, & Keith, 2019; Caruso, Taormina, & Migliorini, 2012; Gao, Liu, Lin, & Wu, 2016). Metabarcoding data also supports relatively low dispersal rates (Zinger et al., 2019) and high turnover of soil microarthropod assemblages even within short geographic distances (Arribas et al., 2021b), despite suggestions that long-distance passive dispersal might be prevalent in these small-bodied groups (Schuppenhauer, Lehmitz, & Xylander, 2019; Türke, Lange, & Eisenhauer, 2018). It remains to be assessed whether those differences among studies are due to the higher taxonomic resolution offered by metabarcoding, or due to context- or scale-dependent variation among systems (Berg, 2012; Ferrenberg, Martinez, & Faist, 2016). What both metabarcoding and morphology-based studies agree on is that habitat type can impose a strong filtering effect overriding other environmental and spatial processes, with forest *vs.* grassland habitats harbouring largely distinct metacommunities (Arribas et al., 2021b; Caruso et al., 2012; Rota et al., 2020). Yet equivalent comparisons among more structurally similar habitats (*e.g.*, different forest types) remain limited.

Here we use both OTU- and ASV-level metabarcoding of entire communities to characterize soil microarthropod assemblages (Acari, Collembola and Coleoptera) across an isolated montane forest mosaic and evaluate the relative importance of forest type, topoclimatic variation and spatial/landscape factors in driving metacommunity structure. The Troodos mountain range harbours unique and understudied Mediterranean forest habitats within the island of Cyprus, one of Europe's most vulnerable islands to climate change (Vogiatzakis, Mannion, & Sarris, 2016), and a major component of the Mediterranean biodiversity hotspot due to high levels of endemicity (Medail & Quezel, 1997). Troodos is characterised by complex topography and steep environmental gradients which, in combination with anthropogenic disturbance since ancient times (Delipetrou, Makhzoumi, Dimopoulos, & Georghiou, 2008), have created a mosaic consisting of five main forest habitat types that differ in area, altitudinal range and level of fragmentation (Figure 1). These comprise forests of: (i) the narrow endemic Cyprus Cedar - *Cedrus brevifolia* (*Cb*, hereafter), with a highly restricted distribution (˜300 ha, between 900-1,400 m) in Western Troodos; (ii) the endemic Golden Oak - *Quercus alnifolia* (*Qa*), with a broad (˜20,000 ha, 700-1,700 m) but highly fragmented distribution across Troodos; (iii) Black Pine -*Pinus nigra pallasiana* (*Pn*, ˜3,500 ha, 1,450-1,950 m) and (iv) Stinking Juniper - *Juniperus foetidissima* (*Jn*, ˜250 ha, 1,450-1,950 m), both narrowly distributed at the top of the highest peak (Chionistra, 1952 m) in Central Troodos; and finally (v) Calabrian Pine forest - *Pinus brutia* (*Pb*) the dominant habitat type, forming continuous and extensive forests across Troodos (˜90,000 ha, 400-1,400 m). We generate metabarcode data for soil microarthropods across this habitat mosaic matrix to address the following questions: (i) Is forest habitat type the primary factor shaping metacommunity structure, in a similar way to that seen between grassland and forest? (Arribas et al., 2021b; Caruso et al., 2012), (ii) What is the relative contribution of spatial *vs.* environmental processes as drivers of within-habitat metacommunity structure? (Zinger et al., 2019), (iii) Focusing on the endemic *Quercus alnifolia* habitat, which is highly fragmented across Troodos, does habitat connectivity across the heterogeneous landscape play an important role in metacommunity structure? (Resasco & Fletcher, 2021). Finally, (iv) are α and β diversity patterns obtained using ASVs and OTUs equivalent and explained by similar spatial or topoclimatic factors? Apart from elucidating soil biodiversity dynamics in those poorly studied but highly vulnerable and precious Mediterranean forests, this system can provide insights into the utility of high-resolution community metabarcoding, in combination with fine-scale topoclimatic and landscape matrix information, for disentangling the drivers of metacommunity structure across mountainous regions and complex landscapes.

MATERIALS AND METHODS

Soil sampling and sample processing

During 2019, from mid-April to mid-June, we collected soil samples from 44 sites representing the main five forest habitat types of the Troodos mountain range (described above, Figure 1; Table S1). Our sampling scheme covered the full extent of the distribution and altitudinal range that the five tree species exhibit on Troodos, spanning over 65 km along an east-west axis and 1500 m of elevation range (Figure 1; Table S1). We collected two soil samples per sampling site corresponding to the superficial (1 m^2 of leaf litter and humus, 5 cm depth) and the deep layer (30 cm diameter, 30 cm depth, comprising ~20 liters of soil) as described in Arribas et al. (2021b). The 88 soil samples were subsequently processed using a standardized flotation-Berlese-flotation protocol to extract the soil mesofauna as detailed in Arribas et al. (2016, 2021b). This protocol allows the retrieval of two subsamples of bulk arthropod specimens, divided according to their body size (typically Acari and Collembola *vs.* Coleoptera) which are suited for 'clean' extraction of whole organism community DNA. During bulk-sample processing, we additionally selected 'voucher' specimens of Acari, Collembola and Coleoptera representing broadly the morphological variation observed in these groups across samples. A total of 176 bulk subsamples and 344 'voucher' specimens were preserved at -20°C in ethanol 100% for molecular analyses.

DNA extraction, PCR amplification and sequencing

We extracted DNA from each bulk subsample using the Biosprint 96 DNA Blood Kit (Qiagen) on a Thermo KingFisher Flex automated extraction instrument. We quantified the DNA concentration using a NanoDrop spectrophotometer and combined the extracts of each pair of subsamples at a 1:10 ratio (Arribas et al., 2021b). Then, we PCR amplified the 3' end of the cytochrome c oxidase subunit I (COI) barcode region corresponding to the 418 bp bc3' fragment using the *Ill_B_F* (Shokralla et al., 2015) and *Fol-degen-rev* (Yu et al., 2012) primers. Amplifications were performed following the PCR conditions described in Arribas et al. (2016). We carried out five independent PCR replicates per sample, visualised the PCR products on an agarose gel and pooled the three amplicons per sample that presented the brightest bands, which were purified using a magnetic bead-based protocol (Agencourt AMPure XP). We included three negative controls corresponding to different wet-lab steps (lysis, DNA extraction and PCR amplification). The 88 metabarcoding samples and the three negative controls were used for a dual-indexed library preparation following the Nextera XT DNA workflow (llumina, San Diego, CA, USA) and were sequenced on a paired-end 2x300 bp lane of an Illumina MiSeq platform at the Earlham Institute (Norwich, UK).

We individually extracted DNA from each of the 'voucher' specimens as described above. We amplified the Folmers COI barcode region (658 bp, overlapping with the 418 bp metabarcoding fragment), using the primers *Fol-degen-for* and *Fol-degen-rev* (Yu et al., 2012) and following the PCR conditions described in Arribas et al. (2016). The PCR products were purified and bidirectionally Sanger sequenced (Macrogen, South Korea). The generation of a curated 'voucher' reference catalogue allowed us to improve the taxonomic assignment of the metabarcode reads and to apply a recently developed method for read filtering (metamate, Andujar et al., 2021, see below).

Illumina read processing and filtering

Detailed information on the read processing and filtering pipeline is summarized in Table S2. Briefly, we demultiplexed raw reads allowing no mismatch in the dual-index pair. Then, we used fastqc v.0.11.7 (Andrews, 2010) to quality check raw reads and cutadapt v.2.10 (Martin, 2011) to trim primers and filter out raw reads exhibiting any variation from expected primer length and composition. Subsequently, we used pear v.0.9.11 (Zhang, Kobert, Flouri, & Stamatakis, 2014) to merge forward and reverse reads. Each metabarcoding sample was then separately quality filtered, dereplicated discarding singletons, length filtered retaining only reads 416-420 bp, *de novo* chimera filtered using UCHIME3, and denoised using UNOISE3 as implemented in vsearch v.2.9.1 (Rognes, Flouri, Nichols, Quince, & Mahe, 2016). Once denoising was performed, reads from all metabarcoding samples were pooled and again dereplicated (discarding no sequences) to generate a catalogue of unique putative haplotypes (ASVs). Subsequently, we ran blast to compare all ASVs against a combined database composed of the NCBI *nt* collection (accessed November 2020) and a

4

curated reference catalogue including the 344 Sanger sequences of the 'voucher' specimens plus 561 previously available sequences corresponding to soil lineages of Acari, Collembola and Coleoptera (Arribas et al. 2016, 2021b). Based on the blast output we assigned the ASVs to high-rank taxonomic levels, by applying the weighted lowest common ancestor algorithm in megan6 (Huson et al., 2016; see also Hleap, Littlefair, Steinke, Hebert, & Cristescu, 2021). Only ASVs assigned to Acari, Collembola or Coleoptera were retained and used for downstream analyses. We further filtered the ASVs using metamate v.0.1b18 (Andujar et al., 2021), a novel approach aiming at removing putative nuclear copies of mitochondrial DNA (NUMTs; Lopez, Yuhki, Masuda, Modi, & O'Brien, 1994) and other types of low-frequency erroneous sequences from denoised metabarcoding datasets. This software allows the application of multiple read-abundance filtering strategies and posterior evaluation of their effects on the prevalence of known authentic mitochondrial haplotypes and presumed non-mitochondrial copies (e.g., those violating the reading frame or expected length, as expected for NUMTs and erroneous sequences) in the final filtered dataset (Andujar et al., 2021). We selected the most stringent filtering solution to ensure the removal of most erroneous sequences (see Supplemental Information for details on the metamate filtering). Subsequently, we used vsearch to generate a read-count community table of the metamate-filtered ASVs by matching them with a 100% identity value against the raw read dataset before dereplicating, length filtering and denoising. We further filtered these community tables by removing ASVs showing abundances of 2 or fewer reads and also those whose contribution to the total number of reads per taxonomic group and library was lower than 1%. Finally, filtered read-count community tables were converted to presence/absence tables (see Jurburg, Keil, Singh, & Chase, 2021). Negative controls were processed alongside actual samples throughout the filtering workflow.

Εστιματες οφ α ανδ β διερσιτψ ατ ΑΣ" ανδ ΟΤΥ λεελς

The resulting fully-filtered ASVs were clustered into operational taxonomic units (OTUs) based on patristic pairwise distances, using a threshold of 3% genetic divergence, which is commonly applied on arthropod COI metabarcoding datasets (Dopheide et al., 2019; Yu et al., 2012). Patristic distances were calculated based on a UPGMA tree, reconstructed using F84 model-corrected pairwise distances and a mafft FFT-NS-i alignment (Katoh & Standley, 2013). These analyses were conducted using the *adegenet* (Jombart, 2008), *phangorn* (Schliep, 2011) and *ape* (Paradis & Schliep, 2019) packages in R (R Core Team, 2020).

Using the ASV and OTU datasets, we calculated α diversity for each soil layer, sampling site and habitat type. We also calculated community dissimilarity among sampling sites and between soil layers using the Sørensen index (β_{SOR}), which we decomposed in its additive components (Simpson dissimilarity index or spatial turnover without the effect of variation in richness, β_{SIM}, and nestedness, β_{SNE}; Baselga, 2010). Then, we used these community dissimilarity matrices to calculate the local contribution to β diversity (LCBD), a comparative indicator of the uniqueness of each sample in terms of community composition (Legendre & De Cáceres, 2013). We calculated the above diversity metrics using the *betapart* (Baselga & Orme, 2012) and *adespatial* (Dray et al., 2020) R packages.

Characterisation of sampling sites using spatial, environmental and landscape variables

A number of spatial, environmental and landscape variables were calculated to characterize the sampling sites and quantify the distances among localities, taking into account the high topographic complexity and environmental/landscape heterogeneity of the study area. We calculated pairwise weighted topographic distances (SPA_{TWD}) based on a digital elevation model (DEM) at 90 m resolution using the *topoDistance* R package (Wang, 2020) as described in the Supplemental Information. We also generated a set of high-resolution environmental variables (at 90 m resolution) for Cyprus, by spatial interpolation of temperature and precipitation layers at lower resolution using the aforementioned DEM (see Supplemental Information). Specifically, we interpolated six WorldClim (annual mean temperature, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, precipitation of wettest quarter and precipitation of driest quarter; Fick & Hijmans, 2017) and three ENVIREM (climatic moisture index, Thornthwaite aridity index and topographic wetness index; Title & Bemmels, 2018) variables. These variables are known to affect the water-energy dynamics and to explain patterns of diversity in several organismal groups (Hawkins et al., 2003). We extracted values of each interpolated variable along with the elevation for all sampling sites and

5

for 500 randomly distributed points throughout Cyprus, to avoid potential biases resulting from only considering conditions at focal sites. We then applied a principal component analysis (PCA; Figure S1) to reduce the dimensionality of the dataset and eliminate covariance among variables and we retained the two first principal components (PCs). PC1 (accounting for 81.3% of variation) was positively correlated with altitude and precipitation variables and negatively with temperature variables and PC2 (8.0%) was positively correlated with the topographic wetness index (Table S3). The PC1 and PC2 scores for each sampling site were considered as topoclimatic predictors (ENV_{PC1} and ENV_{PC2}) for downstream analyses. We also calculated a Euclidean distance matrix among sampling points based on the obtained scores of the two retained PCs, which was used as a topoclimatic predictor (ENV_{PC1-2}) for matrix regression analyses.

Finally, we applied a circuit theory approach (McRae, 2006) to quantify habitat connectivity among the Qa sampling sites. We focused specifically on the Qa habitat because it is broadly distributed but highly fragmented across Troodos, in contrast to the Cb, Pn and Jn forests which are very narrowly distributed and the Pb forest which is very extensive with largely continuous distribution (Figure 1). Based on the assumption that dispersal among isolated forest fragments can be impeded by the presumed lower habitat suitability of the surrounding landscape (Brodie & Newmark, 2019), we built an isolation-by-resistance (IBR) scenario of connectivity (FRA_{IBR}) defined by the distribution of Qa forest patches according to the existing cartography (see Supplemental Information). Two alternative IBR scenarios were constructed for comparison: the TRI_{IBR} representing the topographic complexity of the study area as estimated by the terrain roughness index (Title & Bemmels, 2018) and the $NULL_{IBR}$ representing a completely "flat landscape" with a fixed resistance ($=1$) value assigned to all cells. Resistance distances among all Qa sampling points ($n = 11$) were calculated under each alternative IBR scenario (FRA_{IBR}, TRI_{IBR} and $NULL_{IBR}$) in circuitscape v.4.0.5 (McRae & Beier, 2007).

Statistical analyses

Except when specified, we performed the following statistical analyses at sampling site level, after integrating the community tables of the two soil layer samples (leaf litter and deep soil) for all three taxonomic groups (Acari, Collembola and Coleoptera) into a single matrix. Each of the analyses was performed at both ASV and OTU levels.

α diversity and community uniqueness:

We used ANOVAs to test for significant differences in α diversity (RICH) and community uniqueness (LCBD, local contribution to β diversity) between forest habitats and soil layers. We used generalized linear mixed models (GLMMs) to analyse the relationship between RICH or LCBD per site and the topoclimatic variables (ENV_{PC1} and ENV_{PC2}) as predictors, with latitude and longitude as covariates. We built GLMMs fitting forest habitat type as a random effect in order to account for non-independence among samples from the same forest habitat (see Supplemental Information).

β diversity:

The clustering of sampling sites according to their community composition was visualised using non-metric multidimensional scaling (NMDS) assuming two dimensions ($k =2$). Differences among forest habitat types were tested for significance using permutational multivariate analyses of variance (PERMANOVA). We used symmetric Procrustes analyses to statistically assess non-randomness among NMDS ordinations calculated from different community dissimilarity matrices ($β_{SOR}$ *vs.* $β_{SIM}$ and ASVs *vs.* OTUs; see Peres-Neto & Jackson, 2001). These analyses were performed using the *vegan* (Oksanen et al., 2020) and *pairwiseAdonis* (Martinez Arbizu, 2020) R packages.

The effect of forest habitat type and of the spatial and topoclimatic predictors on β diversity patterns was tested using distance-based redundancy analyses (dbRDA; Legendre & Anderson, 1999), at both across-habitats and within-habitat scales. Specifically, the response variables were the community dissimilarity matrices based on the Simpson dissimilarity index ($β_{SIM}$), and the explanatory variables were forward selected from full models containing the following sets of predictors: (i) forest habitat type (HAB); (ii) spatial

6

variables (SPA$_{\text{PCNMi}}$) derived from the transformation of the topographic weighted distance matrix using Principal Coordinates of Neighbour Matrices (PCNM, see Supplemental Information), and (iii) topoclimatic variables (ENV$_{\text{PC1}}$ and ENV$_{\text{PC2}}$). The best-fit model was selected after ensuring there were no issues of multicollinearity (Variance Inflation Factors, VIF $<$10; *e.g.,* Tonkin, Stoll, Jähnig, & Haase, 2016). Finally, the best-fit models for the across-habitat analyses were used to partition the variance explained exclusively by each variable group (forest habitat type, space and topoclimate) and their intersections using the adjusted coefficient of determination (R^2_{ADJ}) (*e.g.,* Zinger et al., 2019). These analyses were performed using the *BiodiversityR* (Kindt & Coe, 2005) and *vegan* R packages.

In order to validate the dbRDA inferences (Jupke & Schäfer, 2020), we also applied multivariate generalized linear models (mvGLMs) as implemented in the *mvabund* R package (Wang, Naumann, Wright, & Warton, 2012). Unlike dbRDA which is a distance-based approach, this model-based method fits a separate GLM per species and performs resampling-based hypothesis testing for community-level effects of predictors. Here, incidence (presence/absence) datasets were used as input and tested against the same sets of predictors as in dbRDA. Models were fitted using a binomial error distribution and a log link function. We first assessed the predictor significance using single-term models and only those predictors showing a significant effect were used to assemble a full model. The best-fit model was built following a backward stepwise selection approach (*e.g.*, How, Cowan, Teale, & Schmitt, 2019). Term significance was assessed using likelihood ratio tests, PIT-trap resampling and 999 bootstrapping iterations (Wang et al., 2012).

Finally, we assessed the effect of habitat connectivity on the community composition of the *Qa* sampling sites by applying matrix regressions with randomization (MRR; Wang, 2013). Specifically, the community dissimilarity matrices based on the Simpson dissimilarity index (β_{SIM}) were used as response variables, while the explanatory variables were selected among: (i) resistance due to habitat fragmentation (FRA$_{\text{IBR}}$), (ii) resistance due to topographic complexity (TRI$_{\text{IBR}}$), (iii) resistance due to a "flat landscape" (NULL$_{\text{IBR}}$), (iv) weighted topographic (SPA$_{\text{TWD}}$) distances and (v) topoclimatic (ENV$_{\text{PC1-2}}$) distances. We assembled a full model including all explanatory matrices and built the best-fit model following a backward stepwise selection approach, using 999 permutations for the significance tests (*e.g.*, Ortego, Gugger, & Sork, 2015). The unique contribution of each predictor to the total variance explained by the best-fit model was quantified using hierarchical variance partitioning analysis in the *hier.part* R package (Walsh & Mac Nally, 2003). Once the best-fit model was selected, we visualised the relationship between community similarity (1-β_{SIM}) and the explanatory distance/resistance matrices and fitted a distance-decay of community similarity curve (Gómez-Rodríguez & Baselga, 2018; Nekola & White, 1999). Specifically, we fitted a negative exponential function to univariate GLMs assuming a Gaussian error distribution and a log link function as implemented in the *betapart* R package.

RESULTS

Richness and uniqueness of soil microarthropod communities

After denoising and chimera filtering, the removal of putative spurious sequences by metamate, followed by additional filtering of community tables, generated a fully-filtered dataset comprising 907 ASVs which clustered into 386 OTUs (putative species) across the three taxonomic groups. Specifically, we retrieved a total of 353 ASVs and 154 OTUs of Coleoptera, which was the most diverse group at both across- and within-habitat scales. The Acari and Collembola datasets comprised 237 and 317 ASVs which clustered into 139 and 93 OTUs, respectively (Figure 2).

The average α diversity of haplotypes per sampling site differed among habitat types, with the Cyprus cedar (*Cb*) habitat showing significantly higher richness than the other four forest types (Tukey's test: *p*-value $<$0.028 in all comparisons involving *Cb*; Figure 2). At OTU level, the average α diversity per site was higher in the forest types distributed at low and mid altitudes (*Pb*, *Qa* and *Cb*) than in those habitats restricted to higher elevations (*Pn* and *Jn* Figure 2). However, the significance of this pattern was not preserved after post-hoc tests (Tukey's test: all *p*-values $>$0.123). When soil substrates were independently compared, the leaf litter layer showed significantly higher richness (α diversity) than the deep soil substrate

7

in all comparisons (Figure S2). The average local contribution to β diversity (LCBD) per sampling site differed significantly among habitats, with the Calabrian Pine (Pb) and Golden Oak (Qa) communities showing significantly higher uniqueness at both ASV and OTU level when compared to the high-altitude Pn and Jn communities (Tukey's test: p-value <0.05 in most comparisons involving Pb or Qa; Figure 2). When independently analyzed, deep soil communities showed significantly higher LCBD estimates, indicating a more unique composition than the leaf litter communities (Figure S2, inset plots), a pattern particularly evident in the high-altitude Pn and Jn communities (Figure S2).

Regression analyses (GLMMs/GLMs; see Supplemental Information) showed a significantly negative relationship between average richness (α diversity) of ASVs per site and longitude (Lon), indicating that community richness decreased towards the east of the Troodos mountain range (Table 1; Table S4). To ensure that this relationship was not biased by the Cyprus cedar (Cb) sites, which are geographically restricted to the westernmost part of the study area and have the highest ASV richness (Figures 1-2), additional analyses excluded these sites. These analyses consistently supported the significant effect of longitude on ASVs richness across forest habitats (95% CI: [-34.968] - [-4.032]). Conversely, the richness (α diversity) of OTUs per site was only explained by the topoclimatic predictor ENV_{PC2} (Table 1; Table S4). Similarly, we found that LCBD estimates at both ASV and OTU levels were significantly correlated with the topoclimatic variables, as summarized with the ENV_{PC1} and ENV_{PC2} predictors (Table 1; Table S4).

Dissimilarity in community composition among forest sites

Dissimilarity in community composition among sampling sites was high and mainly determined by spatial turnover, with a very limited contribution of nestedness (ASVs: $β_{SIM}$ = 0.974, $β_{SNE}$= 0.004; OTUs: $β_{SIM}$ = 0.961, $β_{SNE}$ = 0.006). This pattern was consistent when each taxonomic group was separately analyzed (all $β_{SIM}$ >0.952, all $β_{SNE}$ <0.015). The community dissimilarity matrices ($β_{SOR}$ or $β_{SIM}$) of the three taxonomic groups were correlated among them at both ASV and OTU level (Mantel test, all r >0.240, all p-values <0.001). As the contribution of nestedness was minimal, we only report the results of the statistical analyses obtained using the community dissimilarity matrices calculated with the Simpson dissimilarity index ($β_{SIM}$).

The NMDS analysis grouped the sampling points according to their respective forest habitat type, except those from Black Pine (Pn) and Stinking Juniper (Jn) forests which exhibited a greater overlap in the ordination (Figure 3). PERMANOVA analyses detected significant differences in community composition among forest habitat types, a factor explaining over 25-34% of the overall variation in community dissimilarity (Figure 3). These differences were significant for all habitat pairs (all p-values <0.020), as confirmed by pairwise comparisons (all p-values <0.019). The NMDS ordinations obtained from different community dissimilarity matrices ($β_{SOR}$ *vs.* $β_{SIM}$ and ASVs *vs.* OTUs) converged on highly similar solutions and were significantly concordant according to Procrustes tests (all r = 0.917, all p-values <0.001).

When dbRDA analyses were applied across all habitats, all three sets of explanatory variables (forest habitat type, spatial and topoclimatic variables) were retained as significant predictors of community dissimilarity ($β_{SIM}$) (Table 2), although the largest fraction of the variation was clearly explained by habitat type (Figure 4). Pairwise comparisons testing for the effect of forest habitat type in community dissimilarity confirmed significant differences between all habitat pairs (all p-values <0.020). When dbRDA were applied at within-habitat scale, community dissimilarity ($β_{SIM}$) of most forest types showed a significant relationship with topoclimatic predictors, except for the case of the Stinking Juniper (Jn) habitat where no predictor was significant (Table 2; Table S5). Once the Jn sampling sites were analysed together with the Pn sites according to the great overlap observed in the NMDS-based ordinations (Figure 3), a significant correlation between community dissimilarity ($β_{SIM}$) and topoclimatic predictors was confirmed (Table 2). The predominant role of topoclimatic variation in explaining $β_{SIM}$ diversity patterns within each habitat was consistent across ASVs and OTUs levels (Table 2; Table S5).

Multivariate GLMs (mvGLMs) provided highly concordant inferences with those obtained using dbRDA (Table S6). Forest habitat type was the predictor that explained the largest fraction of the variation in community composition at across-habitat scale (R^2 >0.121), with the spatial and topoclimatic predictors also

8

included in the final model, although exhibiting much less explanatory power (R^2 <0.044; Table S6). Significant differences in community composition between all habitat pairs were supported by pairwise comparisons (all p-values <0.013). Analyses at within-habitat scale showed that topoclimatic predictors significantly explained community composition within most habitat types. In those cases in which spatial predictors were also retained in the final model, their univariate contribution to the explained variance was usually lower than the one of the topoclimatic predictors (Table S6).

Matrix regressions with randomization (MRR) showed that community dissimilarity (β_{SIM}) among Golden Oak (Qa) sampling sites at both ASV and OTU levels was positively correlated with the IBR matrix based on the spatial configuration of Qa patches (FRA_{IBR}) and the topoclimatic distance matrix (ENV_{PC1-2}). Both predictors were significantly retained in the best-fit models (Table 3), and there was no correlation between them (Mantel test, r <0.103, p-value >0.188). According to the sensitivity analyses, the FRA_{IBR} scenarios that best explained the observed patterns of β_{SIM} variation were those in which the non-$Quercus$ matrix offered much higher resistance (10- or 100-fold) than the target habitat (Table S7). In concordance with the MRR analyses, distance decay models adjusted using GLMs showed a roughly linear decrease in community similarity (β_{SIM}) with increasing FRA_{IBR} or ENV_{PC1-2} distances, with both predictors being significant (Figure 5).

DISCUSSION

Whole organism community DNA (wocDNA) metabarcoding enabled characterizing the soil microarthropod assemblages of the understudied montane forests of Cyprus and provided insights into the drivers of metacommunity structure across the topographically complex region of Troodos. By integrating community metabarcoding with high-resolution topoclimatic and landscape information, we revealed that environmental filtering induced by forest habitat type and topoclimatic heterogeneity controls different facets of Troodos soil biodiversity (α, β diversity and community uniqueness), while habitat connectivity mediates dispersal dynamics across the highly fragmented Golden Oak habitat. β diversity patterns were very similar at OTU and ASV levels, but α diversity varied, with OTU richness following an altitudinal gradient and ASV richness a longitudinal one, likely indicating a decline of genetic diversity eastwards due to anthropogenic disturbance. Therefore, species richness presumably responded to niche-based processes, while haplotype richness to historical contingencies. Our results demonstrate the utility of combining OTUs with ASVs obtained by stringent filtering for characterising diversity patterns of complex assemblages across heterogeneous landscapes.

Forest type shapes regional metacommunity structure

Community composition of soil microarthropods across Troodos was largely explained by forest type, irrespective of spatial and topoclimatic variation (Table 2; Table S6; Figure 4), suggesting a primary role of habitat filtering as a driver of regional metacommunity structure. The importance of habitat type in shaping the composition of soil microarthropod assemblages has been demonstrated by previous research comparing forest $vs.$ grassland habitats (Arribas et al., 2021b; Caruso et al., 2012) or different grass and shrub species (Coulson, Hodkinson, & Webb, 2003; see also Doblas-Miranda, Sánchez-Piñero, & González-Megías, 2009), but to our knowledge, this is the first study to compare different forest types in a systematic way. Given the mosaic nature of Troodos (Figure 1), our sampling scheme revealed high turnover among nearby sampling sites of different woodland habitats (Figure 3), highlighting a major role of underlying forest-specific edaphic features for the community assembly of soil microarthropods (Eissfeller, Langenbruch, Jacob, Maraun, & Scheu, 2013). Leaf litter and vegetation type are documented to influence the soil environment through changes in microhabitat availability and physico-chemical edaphic properties (Berg & McClaugherty, 2008), and such local abiotic features might be driving a scenario of species sorting (Leibold et al., 2004) or even the existence of largely separate metacommunities inhabiting each forest type. In contrast to the general pattern, the communities of the two highland habitats of Black Pine (Pn) and Stinking Juniper (Jn) are more similar to each other (Figure 3; although still significantly different based on pairwise comparisons), suggesting that harsh climatic conditions at those high altitudes might be imposing a stronger environmental filter than forest-associated soil attributes. Alternatively, the spatial configuration of this highland woodland

9

(composed of isolated small stands of Stinking Junipers embedded in a much larger Black Pine matrix) might facilitate dispersal among habitat patches and partly counteract the effects of habitat filtering (Leibold et al., 2004; Logue, Mouquet, Peter, & Hillebrand, 2011). These two hypotheses are not mutually exclusive and may jointly contribute to the higher similarity in community composition between these highland assemblages.

Topoclimate as a driver of within-habitat structure

Analyses performed at within-habitat scale revealed the importance of both environmental and spatial variables on community composition in most cases (Table 2; Table S6), which is in line with previous research on soil microarthropod communities (Arribas et al., 2021b; Bahram et al., 2016; Ingimarsdóttir et al., 2012; Lindo & Winchester, 2009). Disentangling the relative contribution of environmental *vs.* spatial processes has proven to be challenging as environmental variation is often spatially autocorrelated, which can lead to a spurious inflation of the inferred environmental contribution (Clappe, Dray, & Peres-Neto, 2018; Vellend et al., 2014). In our study, spatial predictors generally explained less variance than topoclimatic factors in mvGLMs (Table S6), and their effect became non-significant after applying a forward selection approach in dbRDA (Table 2; Table S5). These results along with the relatively low degree of collinearity between spatial and topoclimatic axes (VIF <7; Vittinghoff, Glidden, Shiboski, & McCulloch, 2012), emphasize the role of environmental filtering as a key driver of metacommunity structure (Brown et al., 2017). Our results would complement several morphology-based studies suggesting that community composition of soil microarthropods is driven by environmental filtering, primarily in response to gradients of edaphic parameters (Caruso et al., 2019; Gao et al., 2016; Gan, Zak, & Hunter, 2019; Grear & Schmitz, 2005). However, they appear to contrast with the recent wocDNA metabarcoding study of Arribas et al. (2021b), where dispersal limitation was identified as the main driver of community assembly at within-habitat scale. This discrepancy cannot be attributed to taxonomic resolution, as both studies used very similar protocols to retrieve ASVs and OTUs, but it could be partly explained by differences in sampling scale, as the generally broader sampling extent of our study could enhance the role of environmental filtering as a consequence of encompassing higher environmental heterogeneity (Chase, 2014). Yet we also found environmental filtering to prevail in our narrowly distributed habitats (*e.g.*, *Cb*, *Pn*) with observational scales slightly smaller (<7 km) than those of Arribas et al. (2021b). Additionally, the overall stronger effect of environmental filtering in our study system may reflect context-dependency (Soininen, 2014), with environmental processes playing a more important role in systems characterized by high topoclimatic heterogeneity. While in Arribas et al. (2021b) there were only moderate altitudinal gradients (˜200-670 m elevation difference), our sampling spanned a steep elevational (1470 m elevation difference) and environmental gradient, with topoclimatic conditions varying greatly even across short distances, both within and across habitats (Figure S1). This phenomenon may be common in topographically complex regions, where dispersal limitation may actually be imposed by environmental heterogeneity rather than by geography *per se* (Liu et al., 2018), and points out the relevance of detailed topoclimatic characterization for understanding metacommunity structure within mountainous landscapes. However, it is noteworthy that the total variance explained by some models was relatively low (R^2_{ADJ}<5-10%, Table 2). This was not unexpected, as it is a common finding among metacommunity studies (Cottenie, 2005), and has been traditionally attributed to other ecological processes that are not frequently measured (Vellend, 2010). Particularly, stochastic demographic processes including ecological drift in the absence of dispersal limitation (Bahram et al., 2016; Zinger et al., 2019) or priority effects via niche preempting (Fukami, 2015) have been hypothesized as relevant forces potentially interfering with community assembly in the soil environment. Additionally, we have not considered explicitly the effect of edaphic variables (*e.g.*, organic matter, nutrient content or pH; Gao et al., 2016; Gan et al., 2019), although some of their variation is likely captured by forest habitat type and by certain topoclimatic variables, which are thought to influence specific soil attributes (horizon depth, moisture; Florinsky, 2012; Hillel, 2008).

The role of habitat connectivity in the assembly of the Golden Oak metacommunity

Habitat fragmentation has been shown to alter the relative importance of spatial *vs.* environmental processes as drivers of metacommunity structure (Jamoneau, Chabrerie, Closset-Kopp, & Decocq, 2012), but the way we measure spatial distances among habitat patches and account for the effects of the surrounding matrix

may affect our interpretation of the predominant processes (Resasco & Fletcher, 2021; Watling, Nowakowski, Donnelly, & Orrock, 2011). In the case of the highly fragmented Golden Oak (Qa) habitat, circuit theory-based connectivity modelling demonstrated that isolation estimates accounting for fragmentation and matrix resistance consistently performed better at explaining turnover than those based on topography, with the latter performing better than null models assuming a homogeneous matrix (Table 3; Table S7). This finding aligns with recent studies that have documented the ecological importance of dispersal corridors for community assembly across different taxonomic groups (Firmiano et al., 2021; Marrec et al., 2021; reviewed in Fletcher, Burrell, Reichert, Vasudev, & Austin, 2016). However, this approach relies on the assumption that all species of the metacommunity respond similarly to landscape heterogeneity. Future studies integrating metabarcoding with morphological information derived from local 'voucher' reference collections could facilitate the implementation of species-specific analyses accounting for ecological and trait variation (Brodie & Newmark, 2019; Hartfelder et al., 2020). Despite these limitations, our results provide empirical evidence of an important effect of habitat fragmentation on soil microarthropod metacommunity structure across the Golden Oak forest patches, with the role of environmental filtering remaining equally significant (Table 3). Interestingly, we also obtained equivalent distance-decay of community similarity curves based on connectivity or on topoclimatic distances (Figure 5), although the topoclimatic variables were not spatially autocorrelated (Mantel test, r <0.103, p-value >0.188) and their shared variance with connectivity-based predictors was very low (<2%; Table 3). The similar slopes of the decay curves for ASVs and OTUs based on spatial distances are compatible with a neutral dispersal-constrained model (Baselga et al., 2015), while the equivalent pattern based on topoclimatic distances could be generated under certain scenarios of high dispersal and narrow ecological niches (Baselga et al., 2013). Taken together our results suggest that dispersal limitation and niche-based processes have jointly shaped turnover patterns across the Qa habitat, although each process may affect different fractions of the metacommunity (*e.g.*, some species and/or intraspecific entities may have low dispersal propensity and wide topoclimatic niches, while others might be good dispersers with narrow niches). Future species-specific analyses may help to refine these conclusions.

The utility of ASVs for community ecology

All the above conclusions about the predominant processes shaping community composition of soil microarthropods across the Troodos forests were very similar when based on OTUs or ASVs, contributing to the broader discussion about whether OTUs should be replaced by ASVs in metabarcoding studies (Callahan et al., 2017; Porter & Hajibabaei, 2020). As recent read-filtering methods have overcome the need of clustering to account for amplification and sequencing errors, the user-defined OTUs (traditionally used as proxies of species-level entities) could be redundant. However, the exclusive use of ASVs could affect the biological conclusions drawn from biodiversity analyses, as patterns of haplotypic diversity can reflect demographic attributes of populations, and do not always coincide with diversity patterns at the species level (Martin et al., 2021). In our system, while β diversity estimates at ASV and OTU levels were correlated and likely shaped by the same ecological processes as described above, we observed distinct α diversity (richness) patterns between them, which were explained by statistically significant differences in geographic or topoclimatic predictors. OTU richness per site was primarily explained by topoclimatic conditions, with assemblages hosting fewer OTUs as elevation and precipitation increased and temperature decreased (Table 1; Figure S3), thus following the general rule of declining species richness with increasing elevation (Rahbek, 1995), commonly interpreted as an outcome of environmental filtering driven by temperature or productivity gradients (Graham et al., 2014; Peters et al. 2016). In contrast to OTUs, ASV richness varied significantly along a longitudinal axis with local communities harbouring more haplotypes westwards (Table 1; Figure S3), which might be interpreted as a signature of higher on average intraspecific genetic diversity in the Western part of the mountain range, which has historically been less affected by anthropogenic disturbance (Delipetrou et al., 2008). This finding is in accordance with population genetic studies of forest trees that observed high genetic diversity in the western populations of *Pinus brutia* (Eliades, Aravanopoulos, & Christou, 2018) and *Cedrus brevifolia* (Eliades, Gailing, Lenemann, Fady, & Finkeldey, 2011). This is potentially a consequence of the local topography facilitating the maintenance of higher effective population sizes in this region during the Pleistocene climatic oscillations, and/or of less intensive historical human impact (livestock grazing

11

and logging) than in Eastern Troodos (Eliades et al., 2018). The decline of haplotypic diversity in soil microarthropod assemblages eastwards may therefore indicate incipient biodiversity loss, as genetic variation tends to be eroded more quickly than species diversity under scenarios of global change (Balint et al., 2011). However, such differences between species and haplotype diversity were not reflected in patterns of community uniqueness, as our ASV- and OTU-based estimates of LCBD (local contribution to β diversity) were correlated and similarly explained by topoclimatic variation (Table 1; Figure S3), without any clear signature of historical contingencies, as those likely affecting longitudinal haplotypic richness patterns across Troodos. Our results therefore highlight the complementarity of OTUs and ASVs for community metabarcoding, as such side-by-side comparisons can help to detect processes that produce uncoupled patterns between the two levels of diversity (*e.g.*, Reisch & Schmid, 2019).

Our ASV-level analyses were facilitated by the application of the metamate tool that utilised local and public reference sequence databases to discard non-authentic ASVs and retain only true biological sequence variants (Andújar et al., 2021). Although applying the most stringent filtering in metamate might have caused the removal of valuable rare biological haplotypes, appreciable intraspecific genetic variation (on average 2.35 ASVs per OTU) was still retrieved and produced reasonable haplotype diversity patterns as explained above. Based on our results, we advocate stringent ASV filtering, as it can provide informative datasets without compromising the required reliability for haplotype-level metabarcoding. Even if we cannot be fully confident that all erroneous haplotypes were filtered out, as the performance of the approach depends on the completeness of the reference sequence catalogue (Andújar et al., 2021), the future incorporation of intraspecific genetic data in local reference databases will provide further confidence and the opportunities for intra-OTU analyses of ASV variation (Elbrecht et al., 2018; Zizka, Weiss, & Leese, 2020).

CONCLUSIONS

This study highlights the power of haplotype-level community metabarcoding, enabled by the application of stringent filtering strategies, for the description of spatial biodiversity patterns of complex communities in understudied regions (Cyprus) and environments (soil), overcoming previous limitations of the taxonomic impediment, low-resolution data and noise due to the presence of spurious sequences. The wide implementation of harmonised field, lab and bioinformatic protocols for community metabarcoding of unexplored assemblages will increase the comparability of datasets from across the globe (Arribas et al., 2021a), providing the basis for broad-scale analyses of metacommunity patterns that would enable drawing more general conclusions on the consistency or context-dependency of ecological processes across spatial scales and fractions of biodiversity. Additionally, the ease with which all species in local communities can be characterised at the population genetic level using metabarcoding with stringent filtering raises the prospect for modelling demographic processes for each of the component species (Overcast, Emerson, & Hickerson, 2019; Overcast et al., 2021). Such an approach has the potential to elucidate historical and contemporaneous community responses to environmental heterogeneity and dispersal limitation at a much finer resolution than the summary statistics currently applied in whole-community metabarcoding.

ACKNOWLEDGEMENTS

REFERENCES

Andrews, S. (2010). fastqc: a quality control tool for high throughput sequence data. Available at *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*

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., Creedy, T. J., Arribas, P., López, H., Salces-Castellano, A., Pérez-Delgado, A. J., ... Emerson, B. C. (2021). Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. *Molecular Ecology Resources,* in press. doi: 10.1111/1755-0998.13337

Arribas, P., Andújar, C., Bidartondo, M. I., Bohmann, K., Coissac, E., Creer, S., ... Emerson, B. (2021a). Connecting high-throughput biodiversity inventories: Opportunities for a site-based genomic framework for global integration and synthesis. *Molecular Ecology, 30*(5), 1120-1135. doi: 10.1111/mec.15797

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M., & Vogler, 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

Arribas, P., Andújar, C., Salces-Castellano, A., Emerson, B. C., & Vogler, A. P. (2021b). The limited spatial scale of dispersal in soil arthropods revealed with whole-community haplotype-level metabarcoding. *Molecular Ecology, 30*(1), 48-61. doi: 10.1111/mec.15591

Balint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change, 1*(6), 313-318. doi: 10.1038/nclimate1191

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*, 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, 24*(8), 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

Benito, X., Fritz, S. C., Steinitz-Kannan, M., Velez, M. I., & McGlue, M. M. (2018). Lake regionalization and diatom metacommunity structuring in tropical South America. *Ecology and Evolution, 8*(16), 7865-7878. doi: 10.1002/ece3.4305

Biswas, S. R., & Wagner, H. H. (2012). Landscape contrast: a solution to hidden assumptions in the metacommunity concept? *Landscape Ecology, 27*(5), 621-631. doi: 10.1007/s10980-012-9732-5

Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K., & Tedersoo, L. (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

Berg, M. P. (2012). Patterns of biodiversity at fine and small spatial scales. In Wall, D. H., Bardgett, R. D., Behan-Pelletier, V., Herrick, J. E., Jones, T. H., Ritz, K., Six, J., Strong, D. R., & van der Putte, W. H.

(Eds.), *Soil ecology and ecosystem services.* (pp. 136-152). UK: Oxford University Press.

Berg, B., & McClaugherty, C. (2008). *Plant Litter. Decomposition, humus formation, carbon sequestration* (2nd edition). Berlin: Springer-Verlag.

Brodie, J. F., & Newmark, W. D. (2019). Heterogeneous matrix habitat drives species occurrences in complex, fragmented landscapes. *The American Naturalist, 193*(5), 748-754. doi: 10.1086/702589

Brown, B. L., Sokol, E. R., Skelton, J., & Tornwall, B. (2017). Making sense of metacommunities: dispelling the mythology of a metacommunity typology. *Oecologia, 183*(3), 643-652. doi: 10.1007/s00442-016-3792-1

Bush, A., Monk, W. A., Compson, Z. G., Peters, D. L., Porter, T. M., Shokralla, S., . . . Baird, D. J. (2020). DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland wilderness. *Proceedings of the National Academy of Sciences of the United States of America, 117*(15), 8539-8545. doi: 10.1073/pnas.1918741117

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal, 11*(12), 2639-2643. doi: 10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). dada2: high-resolution sample inference from Illumina amplicon data. *Nature Methods, 13*(7), 581-583. doi: 10.1038/nmeth.3869

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. *Journal of Animal Ecology, 81*(1), 214-221. doi: 10.1111/j.1365-2656.2011.01886.x

Chase, J. M. (2005). Towards a really unified theory for metacommunities. *Functional Ecology, 19*(1), 182-186. doi: 10.1111/j.0269-8463.2005.00937.x

Chase, J. M. (2014). Spatial scale resolves the niche *versus neutral* theory debate. *Journal of Vegetation Science, 25*(2), 319-322. doi: 10.1111/jvs.12159

Chase, J. M., & Leibold, J. (2003). *Ecological niches: linking classical and contemporary approaches.* Chicago: The University of Chicago Press. Clappe, S., Dray, S., & Peres-Neto, P. R. (2018). Beyond neutrality: disentangling the effects of species sorting and spurious correlations in community analysis. *Ecology, 99*(8), 1737-1747. doi: 10.1002/ecy.2376

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

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters, 8*(11), 1175-1182. doi: 10.1111/j.1461-0248.2005.00820.x

Coulson, S. I., 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

Creedy, T. J., Andújar, C., Meramveliotakis, E., Noguerales, V., Overcast, I., Papadopoulou, A., . . . Arribas, P. (2021). Coming of age for COI metabarcoding of whole organism community DNA: towards bioinformatic harmonisation. *Authorea.* doi: 10.22541/au.162141276.61766048/v1

Creedy, T. J., Ng, W. S., & Vogler, A. P. (2019). Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. *Ecology and Evolution, 9*(6), 3105-3116. doi: 10.1002/ece3.4839

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

Delipetrou, P., Makhzoumi J., Dimopoulos, P., & Georghiou, K. (2008). Cyprus. In Vogiatzakis, I. N., Pungetti, G., & Mannion, A. M. (Eds.), *Mediterranean islands landscapes. Natural and cultural approaches.* (pp. 170-203). Berlin: Springer-Verlag.

Doblas-Miranda, E., Sánchez-Piñero, F., & González-Megías, A. (2009). Different microhabitats affect soil macroinvertebrate assemblages in a Mediterranean arid ecosystem. *Applied Soil Ecology, 41*(3), 329-335. doi: 10.1016/j.apsoil.2008.12.008

Dopheide, A., Tooman, L. K., Grosser, S., Agabiti, B., Rhode, B., Xie, D., . . . Newcomb, R. D. (2019). Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and meta-barcoding data. *Ecological Applications, 29*(4). doi: 10.1002/eap.1877

Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., . . . Wagne, H. H. (2021). *adespatial*: multivariate multiscale spatial analysis. R package version 0.3-14. Available at *https://CRAN.R-project.org/package=adespatial*

Edgar, R. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, 081257. doi: 10.1101/081257

Eissfeller, V., Langenbruch, C., Jacob, A., Maraun, M., & Scheu, S. (2013). Tree identity surpasses tree diversity in affecting the community structure of oribatid mites (Oribatida) of deciduous temperate forests. *Soil Biology and Biochemistry, 63*, 154-162. doi: 10.1016/j.soilbio.2013.03.024

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

Eliades, N.-G., Aravanopoulos, F., & Christou, A. (2018). Mediterranean islands hosting marginal and peripheral forest tree populations: the case of *Pinus brutia* Ten. in Cyprus. *Forests, 9*(9), 514. doi: 10.3390/f9090514

Eliades, N.-G. H., Gailing, O., Leinemann, L., Fady, B., & Finkeldey, R. (2011). High genetic diversity and significant population structure in *Cedrus brevifolia* Henry, a narrow endemic Mediterranean tree from Cyprus. *Plant Systematics and Evolution, 294*(3-4), 185-198. doi: 10.1007/s00606-011-0453-z

Ferrenberg, S., Martinez, A. S., & Faist, A. M. (2016). Aboveground and belowground arthropods experience different relative influences of stochastic versus deterministic community assembly processes following disturbance. *PeerJ, 4*, e2545. doi: 10.7717/peerj.2545

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology, 37*(12), 4302-4315. doi: 10.1002/joc.5086

Firmiano, K. R., Cañedo-Argüelles, M., Gutiérrez-Canovas, C., Macedo, D. R., Linares, M. S., Bonada, N., & Callisto, M. (2021). Land use and local environment affect macroinvertebrate metacommunity organization in Neotropical stream networks. *Journal of Biogeography, 48*(3), 479-491. doi: 10.1111/jbi.14020

Fletcher, R. J., Burrell, N. S., Reichert, B. E., Vasudev, D., & Austin, J. D. (2016). Divergent perspectives on landscape connectivity reveal consistent effects from genes to communities. *Current Landscape Ecology Reports, 1*, 67-79. doi: 10.1007/s40823-016-0009-6

Florinsky, I. V. (2012). Influence of topography on soil properties. In Florinsky, I.V. (Ed.), *Digital terrain analysis in soil science and geology* (2nd edition) (pp. 265-270). San Diego, CA: Academic Press-Elsevier.

Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics, 46*(1), 1-23. doi: 10.1146/annurev-ecolsys-110411-160340

Gálvez-Reyes, N., Arribas, P., Andújar, C., Emerson, B. C., Piñero, D., & Mastretta-Yanes, A. (2020). Local-scale dispersal constraints promote spatial structure and arthropod diversity within a tropical sky-island. *Authorea*. doi: 10.22541/au.160193334.45224582/v1

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

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*, 43-53. doi: 10.1016/j.pedobi.2019.03.002

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

Graf, R. F., Kramer-Schadt, S., Fernández, N., & Grimm, V. (2007). What you see is where you go? Modeling dispersal in mountainous landscapes. *Landscape Ecology, 22*(6), 853-866. doi: 10.1007/s10980-006-9073-3

Graham, C. H., Carnaval, A. C., Cadena, C. D., Zamudio, K. R., Roberts, T. E., Parra, J. L., . . . Sanders, N. J. (2014). The origin and maintenance of montane diversity: integrating evolutionary and ecological processes. *Ecography, 37*(8), 711-719. doi: 10.1111/ecog.00578

Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006). Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters, 9*(4), 399-409. doi: 10.1111/j.1461-0248.2006.00884.x

Grear, J. S., & Schmitz, O. J. (2005). Effects of grouping behavior and predators on the spatial distribution of a forest floor arthropod. *Ecology, 86*(4), 960-971. doi: 10.1890/04-1509

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J., Guegan, J. F., Kaufman, D. M., . . . Turner, J. R. G. (2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology, 84*(12), 3105-3117. doi: 10.1890/03-8006

Hartfelder, J., Reynolds, C., Stanton, R. A., Sibiya, M., Monadjem, A., McCleery, R. A., & Fletcher, R. J. J. (2020). The allometry of movement predicts the connectivity of communities. *Proceedings of the National Academy of Sciences of the United States of America, 117*(36), 22274-22280. doi: 10.1073/pnas.2001614117

He, S., Chen, K., Soininen, J., Heino, J., Ding, N., & Wang, B. (2020). Elements of metacommunity structure of diatoms and macroinvertebrates within stream networks differing in environmental heterogeneity. *Journal of Biogeography, 47*(8), 1755-1764. doi: 10.1111/jbi.13859

Hillel, D. (2008). *Soil in the environment. Crucible of terrestrial life*. Burlington, MA: Academic Press-Elsevier.

Hleap, J. S., Littlefair, J. E., Steinke, D., Hebert P. D. N., & Cristescu, M. E. (2021). Assessment of current taxonomic assignments strategies for metabarcoding eukaryotes. *Molecular Ecology Resources* , in press. doi: 10.1111/1755-0998.13407

Hoiss, B., Krauss, J., Potts, S. G., Roberts, S., & Steffan-Dewenter, I. (2012). Altitude acts as an environmental filter on phylogenetic composition, traits and diversity in bee communities. *Proceedings of the Royal Society B-Biological Sciences, 279*(1746), 4447-4456. doi: 10.1098/rspb.2012.1581

How, R. A., Cowan, M. A., Teale, R. J., & Schmitt, L. H. (2020). Environmental correlates of reptile variation on the Houtman Abrolhos archipelago, eastern Indian Ocean. *Journal of Biogeography, 47*(9), 2017-2028. doi: 10.1111/jbi.13881

Hubbell, S. P. (2001). *The unified neutral theory of species abundance and diversity*. Princeton, NJ: Princeton University Press.

Huson, D. H., Beier, S., Flade, I., Gorska, A., El-Hadidi, M., Mitra, S., . . . Tappu, R. (2016). megan Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. *Plos Computational Biology, 12*(6), e1004957. doi: 10.1371/journal.pcbi.1004957

Hutchinson, G. E. (1959). Homage to Santa-Rosalia or why are there so many kinds of animals. *The American Naturalist, 93*(870), 145-159. doi: 10.1086/282070

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

Jamoneau, A., Chabrerie, O., Closset-Kopp, D., & Decocq, G. (2012). Fragmentation alters beta-diversity patterns of habitat specialists within forest metacommunities. *Ecography, 35*(2), 124-133. doi: 10.1111/j.1600-0587.2011.06900.x

Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405. doi: 10.1093/bioinformatics/btn129

Joos, L., Beirinckx, S., Haegeman, A., Debode, J., Vandecasteele, B., Baeyen, S., . . . De Tender, C. (2020). Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genomics, 21*(1). doi: 10.1186/s12864-020-07126-4

Jupke, J. F., & Schäfer, R. B. (2020). Should ecologists prefer model- over distance-based multivariate methods? *Ecology and Evolution, 10*(5), 2417-2435. doi: 10.1002/ece3.6059

Jurburg S. D., Keil, P., Singh, B. K., & Chase, J. M. (2021). All together now: limitations and recommendations for the simultaneous analysis of all eukaryotic soil sequences. *Molecular Ecology Resources*, in press. doi: 10.1111/1755-0998.13401

Katoh, K., & Standley, D. M. (2013). mafft - Multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution, 30*(4), 772-780. doi: 10.1093/molbev/mst010

Kindt, R., & Coe, R. (2005). *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. Nairobi, Kenya: World Agroforestry Centre (ICRAF). Available at *https: // www. worldagroforestry. org/ output/ tree-diversity-analysis*

Legendre, P., & Anderson, M. J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs, 69* (1), 1-24. doi: 10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2

Legendre, P., & De Caceres, M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters, 16*(8), 951-963. doi: 10.1111/ele.12141

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., . . . Gonzalez, A. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters, 7*(7), 601-613. doi: 10.1111/j.1461-0248.2004.00608.x

Leingärtner, A., Krauss, J., & Steffan-Dewenter, I. (2014). Species richness and trait composition of butterfly assemblages change along an altitudinal gradient. *Oecologia, 175*(2), 613-623. doi: 10.1007/s00442-014-2917-7

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

Liu, C., Dudley, K. L., Xu, Z.-H., & Economo, E. P. (2018). Mountain metacommunities: climate and spatial connectivity shape ant diversity in a complex landscape. *Ecography, 41*(1), 101-112. doi: 10.1111/ecog.03067

Logue, J. B., Mouquet, N., Peter, H., & Hillebrand, H. (2011). Empirical approaches to metacommunities: a review and comparison with theory. *Trends in Ecology and Evolution, 26*(9), 482-491. doi: 10.1016/j.tree.2011.04.009

Lopez, J. V., Yuhki, N., Masuda, R., Modi, W., & O'Brien, S. J. (1994). NUMT, a recent transfer and tandem amplification of mitochondrial-DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution, 39*(2), 174-190. doi: 10.1007/BF00163806

Marrec, R., Le Roux, V., Martin, L., Lenoir, J., Brunet, J., Cousins, S. A. O., ... Decocq, G. (2021). Multiscale drivers of carabid beetle (Coleoptera: Carabidae) assemblages in small European woodlands. *Global Ecology and Biogeography, 30*(1), 165-182. doi: 10.1111/geb.13208

Martin, M. (2011). cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, 17*, 10-12.

Martin, G. K., Beisner, B. E., Chain, F. J. J., Cristescu, M. E., del Giorgio, P. A., & Derry, A. M. (2021). Freshwater zooplankton metapopulations and metacommunities respond differently to environmental and spatial variation. *Ecology, 102*(1), e03224. doi: 10.1002/ecy.3224

Martinez Arbizu, P. (2020). *pairwiseAdonis*: pairwise multilevel comparison using *adonis*. R package version 0.4. Available at:*https://github.com/pmartinezarbizu/pairwiseAdonis*

McRae, B. H. (2006). Isolation by resistance. *Evolution, 60*(8), 1551-1561. doi: 10.1111/j.0014-3820.2006.tb00500.x

McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America, 104*(50), 19885-19890. doi: 10.1073/pnas.0706568104

McRae, B. H., Dickson, B. G., Keitt, T. H., & Shah, V. B. (2008). Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology, 89*(10), 2712-2724. doi: 10.1890/07-1861.1

Medail, F., & Quezel, P. (1997). Hot-spots analysis for conservation of plant biodiversity in the Mediterranean basin. *Annals of the Missouri Botanical Garden, 84*(1), 112-127. doi: 10.2307/2399957

Meynard, C. N., Lavergne, S., Boulangeat, I., Garraud, L., Van Es, J., Mouquet, N., & Thuiller, W. (2013). Disentangling the drivers of metacommunity structure across spatial scales. *Journal of Biogeography, 40*(8), 1560-1571. doi: 10.1111/jbi.12116

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography, 26*(4), 867-878. doi: 10.1046/j.1365-2699.1999.00305.x

Nielsen, U. (2019). *Soil fauna assemblages: global to local scales (Ecology, Biodiversity and Conservation)*. Cambridge: Cambridge University Press.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2020). *vegan:* community ecology package. R package version 2.5-7. Available at*https://CRAN.R-project.org/package=vegan*

Ortego, J., Gugger, P.F., & Sork, V. L. (2015). Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography,42*(2), 328-338. doi: 10.1111/jbi.12419

Overcast I., Emerson B. C., & Hickerson M. J. (2019). An integrated model of population genetics and community ecology. *Journal of Biogeography, 46*(4), 816-829. doi: 10.1111/jbi.13541

Overcast, I., Ruffley, M., Rosindell, J., Harmon, L., Borges, P., Emerson, B. C, ... Rominger, A. J. (2021). A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. *Molecular Ecology Resources,* in press. doi: 10.1101/2020.01.30.927236

Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis, A., & Vogler, A. P. (2011). Testing the species–genetic diversity correlation in the Aegean Archipelago: toward a haplotype-based macroecology? *The American Naturalist*,*178*(2), 241-255. doi: 10.1086/660828

Paradis, E., & Schliep, K. (2019). *ape* 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526-528. doi: 10.1093/bioinformatics/bty633

Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia, 129*(2), 169-178. doi: 10.1007/s004420100720

Peters, M. K., Hemp, A., Appelhans, T., Behler, C., Classen, A., Detsch, F., . . . Steffan-Dewenter, I. (2016). Predictors of elevational biodiversity gradients change from single taxa to the multi-taxa community level. *Nature Communications, 7*, 13736. doi: 10.1038/ncomms13736

Porter, T. M., & Hajibabaei, M. (2020). Putting COI metabarcoding in context: the utility of exact sequence variants (ESVs) in biodiversity analysis. *Frontiers in Ecology and Evolution, 8*, 248. doi: 10.3389/fevo.2020.00248

R Core Team. (2020). r: a language and environment for statistical computing. r Foundation for Statistical Computing, Vienna, Austria. Available at *https://www.r-project.org/*

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

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., . . . Fjeldsa, J. (2019). Humboldt's enigma: what causes global patterns of mountain biodiversity? *Science, 365*(6458), 1108-1113. doi: 10.1126/science.aax0149

Reisch, C., & Schmid, C. (2019). Species and genetic diversity are not congruent in fragmented dry grasslands. *Ecology and Evolution*,*9*(1), 664-671. doi: 10.1002/ece3.4791

Resasco, J., & Fletcher, R. J. (2021). Accounting for connectivity alters the apparent roles of spatial and environmental processes on metacommunity assembly. *Landscape Ecology, 36*(4), 1089-1099. doi: 10.1007/s10980-021-01203-z

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahe, F. (2016). vsearch: a versatile open source tool for metagenomics. *PeerJ, 4*, e2584. doi: 10.7717/peerj.2584

Rota, N., Canedoli, C., Ferre, C., Ficetola, G. F., Guerrieri, A., & Padoa-Schioppa, E. (2020). Evaluation of soil biodiversity in alpine habitats through eDNA metabarcoding and relationships with environmental features. *Forests, 11*(7), 738. doi: 10.3390/f11070738

Schliep, K. P. (2011). *phangorn*: phylogenetic analysis in R. *Bioinformatics, 27*(4), 592-593. doi: 10.1093/bioinformatics/btq706

Schuppenhauer, M. M., Lehmitz, R., & Xylander, W. E. R. (2019). Slow-moving soil organisms on a water highway: aquatic dispersal and survival potential of Oribatida and Collembola in running water. *Movement Ecology, 7*, 20. doi: 10.1186/s40462-019-0165-5

Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., . . . Hajibabaei, M. (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports, 5,* 9687. doi: 10.1038/srep09687

Soininen, J. (2014). A quantitative analysis of species sorting across organisms and ecosystems. *Ecology, 95*(12), 3284-3292. doi: 10.1890/13-2228.1

Steinbauer, M. J., Grytnes, J. A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H., . . . Wipf, S. (2018). Accelerated increase in plant species richness on mountain summits is linked to warming. *Nature, 556*(7700),

231-234. doi: 10.1038/s41586-018-0005-6

Thakur, M. P., Phillips, H. R. P., Brose, U., De Vries, F. T., Lavelle, P., Loreau, M., . . . Cameron, E. K. (2020). Towards an integrative understanding of soil biodiversity. *Biological Reviews, 95*(2), 350-364. doi: 10.1111/brv.12567

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography, 41*(2), 291-307. doi: 10.1111/ecog.02880

Tonkin, J. D., Altermatt, F., Finn, D. S., Heino, J., Olden, J. D., Pauls, S. U., & Lytle, D. A. (2018). The role of dispersal in river network metacommunities: patterns, processes, and pathways. *Freshwater Biology, 63*(1), 141-163. doi: 10.1111/fwb.13037

Tonkin, J. D., Shah, R. D. T., Shah, D. N., Hoppeler, F., Jaehnig, S. C., & Pauls, S. U. (2017). Metacommunity structuring in Himalayan streams over large elevational gradients: the role of dispersal routes and niche characteristics. *Journal of Biogeography, 44*(1), 62-74. doi: 10.1111/jbi.12895

Tonkin, J. D., Stoll, S., Jähnig, S. C., & Haase, P. (2016). Contrasting metacommunity structure and beta diversity in an aquatic-floodplain system. *Oikos, 125*(5), 686-697. doi: 10.1111/oik.02717

Türke, M., Lange, M., & Eisenhauer, N. (2018). Gut shuttle service: endozoochory of dispersal-limited soil fauna by gastropods. *Oecologia, 186*(3), 655-664. doi: 10.1007/s00442-018-4058-x

Vellend, M. (2010). Conceptual synthesis in community ecology. *Quarterly Review of Biology, 85*(2), 183-206. doi: 10.1086/652373

Vellend, M., Srivastava, D. S., Anderson, K. M., Brown, C. D., Jankowski, J. E., Kleynhans, E. J., . . . Xue, X. (2014). Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos, 123*(12), 1420-1430. doi: 10.1111/oik.01493

Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D. A., De Wever, A., Juggins, S., . . . Sabbe, K. (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos, 118*(8), 1239-1249. doi: 10.1111/j.1600-0706.2009.17575.x

Viana, D. S., & Chase, J. M. (2019). Spatial scale modulates the inference of metacommunity assembly processes. *Ecology, 100*(2), e02576. doi: 10.1002/ecy.2576

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2012). *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models* (2nd edition). Berlin: Springer-Verlag.

Vogiatzakis, I. N., Mannion, A. M., & Sarris, D. (2016). Mediterranean island biodiversity and climate change: the last 10,000 years and the future. *Biodiversity and Conservation, 25*(13), 2597-2627. doi: 10.1007/s10531-016-1204-9

Walsh C., & Mac Nally, R. (2003). *hier.part*: hierarchical partitioning. R package version 1.0-6. Available at *https://CRAN.R-project.org/package=hier.part*

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution, 67*(12), 3403-3411. doi: 10.1111/evo.12134

Wang, I. J. (2020). Topographic path analysis for modelling dispersal and functional connectivity: calculating topographic distances using the *topoDistance* R package. *Methods in Ecology and Evolution, 11*(2), 265-272. doi: 10.1111/2041-210x.13317

Wang, J., Legendre, P., Soininen, J., Yeh, C.-F., Graham, E., Stegen, J., . . . Pan, F. (2020). Temperature drives local contributions to beta diversity in mountain streams: stochastic and deterministic processes. *Global Ecology and Biogeography, 29*(3), 420-432. doi: 10.1111/geb.13035

Wang, Y., Naumann, U., Wright, S. T., & Warton, D. I. (2012). *mvabund* - an R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution, 3*(3), 471-474. doi: 10.1111/j.2041-210X.2012.00190.x

Watling, J. I., Nowakowski, A. J., Donnelly, M. A., & Orrock, J. L. (2011). Meta-analysis reveals the importance of matrix composition for animals in fragmented habitat. *Global Ecology and Biogeography, 20*(2), 209-217. doi: 10.1111/j.1466-8238.2010.00586.x

Young, M. R., Proctor, H. C., deWaard, J. R., & Hebert, P. D. N. (2019). DNA barcodes expose unexpected diversity in Canadian mites. *Molecular Ecology, 28*(24), 5347-5359. doi: 10.1111/mec.15292

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution, 3*(4), 613-623. doi: 10.1111/j.2041-210X.2012.00198.x

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). pear: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics, 30*(5), 614-620. doi: 10.1093/bioinformatics/btt593

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

Zizka, V. M. A., Weiss, M., & Leese, F. (2020). Can metabarcoding resolve intraspecific genetic diversity changes to environmental stressors? A test case using river macrozoobenthos. *Metabarcoding & Metagenomics, 4*, 23-34. doi: 10.3897/mbmg.4.51925

DATA ACCESSIBILITY STATEMENT

Raw metabarcode data have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject xxxxxxx. Input files for all analyses are available for download from the Dryad Digital Repository (*https: // doi. org/ xx. xxxx/ dryad. xxxxxxxxxx*), including raster files of interpolated WorldClim and ENVIREM variables and distribution of *Quercus alnifolia* patches. Upon acceptance, data will be archived in the above-commented repositories and the Data Accessibility Statement completed.

AUTHOR CONTRIBUTIONS

V.N and A.P conceived and led the study. V.N and E.M performed fieldwork. V.N. performed laboratory work with help from E.M. A.C-I generated climatic and topographic information and wrote the respective methodological section. V.N. and A.P. conceived the methodological approach and V.N. performed the analyses. V.N wrote the manuscript with help from A.P. All authors contributed critically to the draft and gave final approval for publication.

ORCID

Víctor Noguerales *https: // orcid. org/ 0000-0003-3185-778X*

Emmanouil Meramveliotakis *https: // orcid. org/ 0000-0002-6399-575X*

Adrián Castro-Insua *https: // orcid. org/ 0000-0003-4184-8641*

Carmelo Andújar *https: // orcid. org/ 0000-0001-9759-7402*

Paula Arribas *https: // orcid. org/ 0000-0002-0358-8271*

Thomas Creedy *https: // orcid. org/ 0000-0002-7611-291X*

Isaac Overcast *https: // orcid. org/ 0000-0001-8614-6892*

Hélène Morlon *https://orcid.org/0000-0002-3195-7521*

Brent Emerson *https://orcid.org/0000-0003-4067-9858*

Alfried P. Vogler *https://orcid.org/0000-0002-2462-3718*

Anna Papadopoulou *https://orcid.org/0000-0002-4656-4894*

TABLE 1 Results of model selection and averaging testing for the relationship between average richness (α diversity, RICH) or local contribution to β diversity (LCBD) at ASV and OTU levels as response variables and topoclimatic variation (ENV_{PC1} and ENV_{PC2}; see Table S3) as explanatory variables. Latitude (Lat) and longitude (Lon) were included as covariates. Predictors excluding the value 0 in their 95% confidence intervals (CI) are indicated in bold and their effects were considered significant. For each final model including only predictors considered significant (in bold), marginal (R^2_m, variance explained by fixed effects) and conditional (R^2_c, variance explained by both fixed and random effects) coefficients of determination are reported.

Models	R^2_m	R^2_c	Predictors	Estimate ± adjusted standard error	Lower 95% CI	Upper 95% CI
ASVs (haplotypes)						
$RICH_{ASV}$	0.185	0.384	Lat	-6.326 ± 21.575	-48.613	35.961
			Lon	**-24.975 ± 7.970**	**-40.597**	**-9.353**
			ENV_{PC2}	-0.842 ± 1.536	-3.854	2.169
$LCBD_{ASV}$	0.698	0.807	**ENV_{PC1}**	**$-2.979 \times 10^{-4} \pm 7.938 \times 10^{-5}$**	**-4.534×10^{-4}**	**-1.423×10^{-4}**
			ENV_{PC2}	**$-6.160 \times 10^{-4} \pm 2.148 \times 10^{-4}$**	**-1.037×10^{-3}**	**-1.949×10^{-4}**
OTUs (3% lineages)						
$RICH_{OTU}$	0.278	-	Lon	-5.199 ± 5.026	-15.049	4.651
			ENV_{PC2}	**-3.115 ± 0.777**	**-4.639**	**-1.592**
$LCBD_{OTU}$	0.488	0.512	**ENV_{PC1}**	**$-5.524 \times 10^{-4} \pm 9.684 \times 10^{-5}$**	**-7.422×10^{-4}**	**-3.625×10^{-4}**

TABLE 2 Best-fit models from distance-based redundancy analyses (dbRDA) on community composition (Simpson dissimilarity index, β_{SIM}) of sampling sites at ASV and OTU levels, performed either with all sites (across habitats) or separately for each of the forest habitat types. The explanatory variables were forward selected from full models containing three sets of predictors: forest habitat type (HAB), spatial (SPA_{PCNMi}) and topoclimatic (ENV_{PCi}) variables. The adjusted coefficient of determination (R^2_{ADJ}) for each model is provided. The *Pinus nigra* (*Pn*) and *Juniperus foetidissima* (*Jn*) sampling sites were both separately and jointly analyzed according to results of NMDS-based ordinations (Figure 2). Predictor information for models with no significant variables (null) is replaced by dashes.

Dataset	*ASVs (haplotypes)* Predictors	*ASVs (haplotypes)* F	*ASVs (haplotypes)* p-value	*ASVs (ha* R^2_{ADJ}
Across habitats	HAB	3.311	<0.001	0.204
	SPA_{PCNM2}	1.824	<0.001	
	SPA_{PCNM5}	1.383	0.029	
	ENV_{PC1}	1.951	<0.001	
	ENV_{PC2}	1.565	0.006	
Pinus brutia (*Pb*)	ENV_{PC1}	4.808	0.048	0.024
Quercus alnifolia (*Qa*)	ENV_{PC1}	1.618	0.002	0.058
Cedrus brevifolia (*Cb*)	ENV_{PC1}	4.808	0.041	0.388

Pinus nigra (*Pn*)	$\mathrm{ENV_{PC1}}$	1.805	0.005	0.103
Juniperus foetidissima (*Jn*)	Null	-	-	-
Pinus nigra (*Pn*) + *Juniperus foetidissima* (*Jn*)	$\mathrm{ENV_{PC1}}$	1.459	0.047	0.030

TABLE 3 Multiple matrix regression with randomization (MRR) analyses on community composition (Simpson dissimilarity index, β_{SIM}) of *Quercus alnifolia* (*Qa*) sampling sites at ASV and OTU levels. The explanatory variables were backward selected from full models containing the following distance matrices: 'flat' scenario ($\mathrm{NULL_{IBR}}$), topography ($\mathrm{SPA_{TWD}}$), topoclimate ($\mathrm{ENV_{PC1-2}}$), topographic complexity ($\mathrm{TRI_{IBR}}$) and forest fragmentation ($\mathrm{FRA_{IBR}}$). The significance of both explanatory terms retained in the best-fit model and those rejected during the backward selection procedure are reported. The independent (R_{I}^2) and joint coefficient of determination (R_{J}^2) of each predictor retained in the best-fit final model are provided.

ASVs (haplotypes) Predictors	*ASVs (haplotypes)* Coefficient	*ASVs (haplotypes)* t	*ASVs (haplotypes)* p-value	*ASVs (haplotypes)* R_{I}^2	*ASVs (haplotypes* R_{J}^2
Explanatory terms	Explanatory terms				
$\mathrm{FRA_{IBR}}$	0.078	2.745	0.014	0.112	0.013
$\mathrm{ENV_{PC1-2}}$	0.106	3.889	0.000	0.210	0.013
Rejected terms					
$\mathrm{SPA_{TWD}}$		-0.209	0.839		
$\mathrm{TRI_{IBR}}$		1.911	0.272		
$\mathrm{NULL_{IBR}}$		-1.843	0.257		

FIGURE 1 Geographical location of sampling sites throughout the Troodos mountain range in Cyprus (top panel) and distribution of the main five forest habitat types (bottom panel). Sampling sites and forest habitat distribution are coloured as follows: *Pb*, *Pinus brutia* (light green); *Qa*, *Quercus alnifolia* (orange); *Cb*, *Cedrus brevifolia* (blue); *Pn*, *Pinus nigra* (black); *Jn*, *Juniperus foetidissima* (purple). Top panel: Background map displays elevation at 90 m resolution (SRTM Digital Elevation Data, *http://srtm.csi.cgiar.org/*). Bottom panel: Data from the Department of Forests (Ministry of Agriculture, Rural Development and Environment, Republic of Cyprus; *https://www.data.gov.cy/*). Green hatched areas represent zones where *Pinus brutia* (light green) forests are present according to our own surveys but whose extent has not been mapped in the publicly available cartography, likely because they are the result of plantations, have been affected by fires, and/or present very low tree densities.
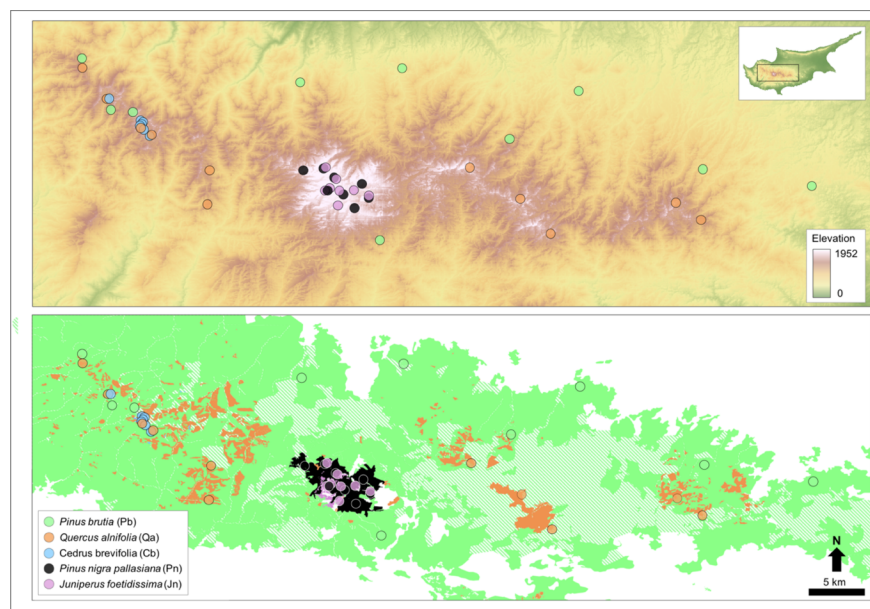
FIGURE 2 Average richness (α diversity, top panels) and community uniqueness (local contribution to β diversity - LCBD, bottom panels) across sampling sites, as estimated at ASV (left panels) and OTU (right panels) levels, per forest habitat type. Colours and habitat codes as in Figure 1. Inset graphs show the contribution of each taxonomic group (Acari, dark grey; Collembola, grey; Coleoptera, light grey) to the cumulative richness (γ diversity) per forest habitat type. Shared letters below the box-plots indicate that differences between the respective habitats are not statistically significant (p-value >0.05) after post-hoc Tukey's tests.
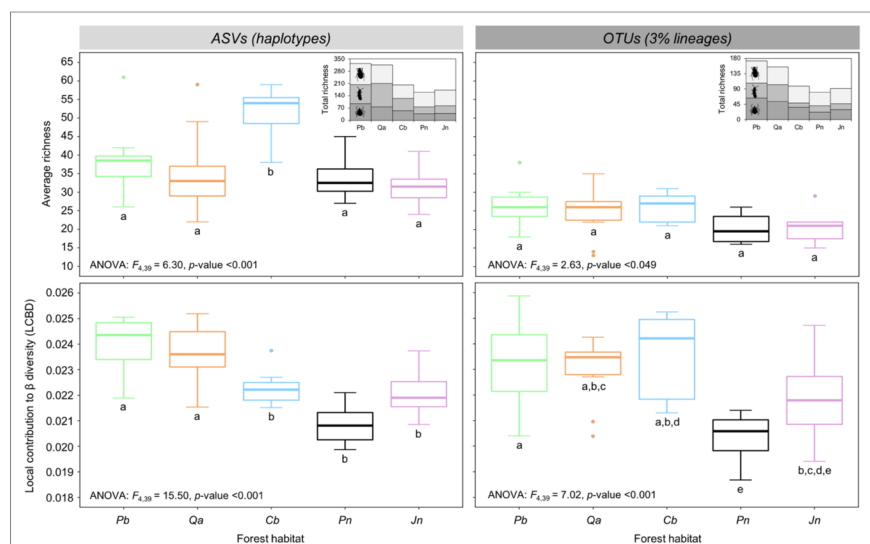

FIGURE 3 Non-metric multidimensional scaling (NMDS) ordination of sampling sites according to community dissimilarity (Simpson dissimilarity index, β_{SIM}), at ASV (left panel) and OTU (right panel) levels. Circles correspond to sampling sites, with circle size representing sample richness. The percentage of explained variation (R^2) and the significance of forest habitat type as grouping factor based on PERMANOVA are reported on the top of each plot. Colours and habitat codes as in Figure 1.
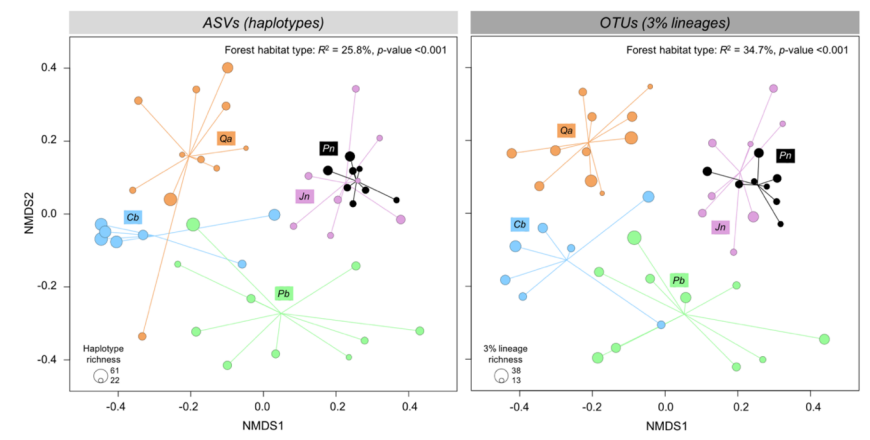
24

FIGURE 4 Venn diagram illustrating the partitioning of explained variance in community composition among the three sets of explanatory variables (forest habitat type, spatial and topoclimatic predictors) and their intersections. Variance partitioning was conducted on the best-fit dbRDA models, which only included significant variables after a forward selection procedure (Table 2). Analyses were performed on the Simpson dissimilarity index (β_{SIM}) community dissimilarity matrices at ASV and OTU levels. Percentages refer to adjusted coefficients of determination (R^2_{ADJ}). Percentages below 0.25% are not shown.
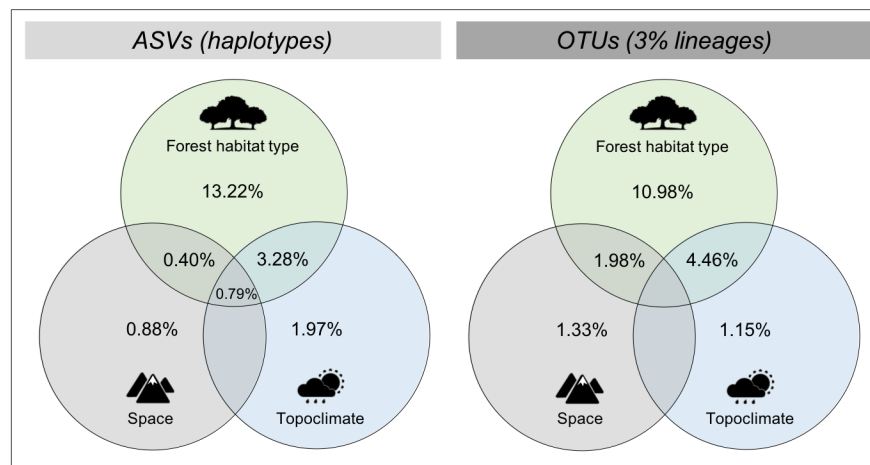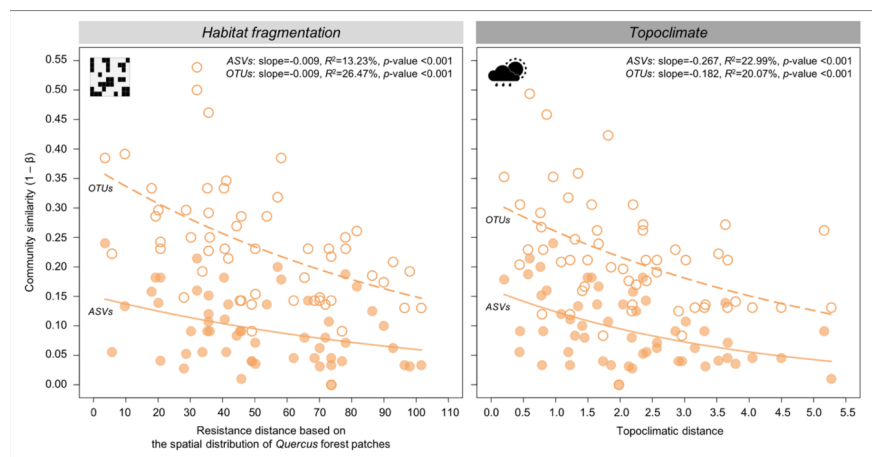

FIGURE 5 Distance decay of community similarity across the *Quercus alnifolia* (Qa) habitat. Distances among Qa sampling sites were calculated either based on the isolation-by-resistance (IBR) scenario reflecting the spatial distribution of Qa forest patches (FRA$_{IBR}$, left panel) or on the topoclimatic distances among sampling sites (ENV$_{PC1-2}$, right panel). Community similarity was estimated using the Simpson similarity index (1 - β_{SIM}) at ASV (filled circles, solid regression line) and OTU (open circles, dashed regression line) levels. For display purposes, the decay curves were fitted using generalized linear models (GLMs) with a negative exponential function. Regression slopes, coefficients of determination (pseudo-R^2) and p-values are provided on the top of each panel. See Table 3 for multivariate matrix regression analyses assessing the independent contribution of each predictor to community dissimilarity.

25

Hosted file

`_MEC_Supplemental_Information.pdf` available at https://authorea.com/users/422731/articles/528394-community-metabarcoding-reveals-the-relative-role-of-environmental-filtering-and-dispersal-in-metacommunity-dynamics-of-soil-microarthropods-across-a-mosaic-of-montane-forests