Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uncovering depth and spatial patterns of soil environmental DNA in arid regions

Madeleine P. Malmfeldt
U.S. Army Engineer Research and Development Center
Denise L. Lindsay
U.S. Army Engineer Research and Development Center
Carina M. Jung
U.S. Army Engineer Research and Development Center
Mark A. Chappell
U.S. Army Engineer Research and Development Center
N. Paige Fowler
U.S. Army Engineer Research and Development Center
Autumn Acree
Oak Ridge Institute for Science and Education
Michelle Wynter
U.S. Army Engineer Research and Development Center
Yoko Masue-Slowey
U.S. Army Engineer Research and Development Center
Adam J. Norris
U.S. Army Engineer Research and Development Center
George M. Ferguson
University of Arizona
Richard F. Lance
Richard.F.Lance@erdc.dren.mil

U.S. Army Engineer Research and Development Center

Research Article

Keywords: bioindicators, eDNA metabarcoding, geoforensics, high-throughput sequencing, taxon richness, terrestrial eDNA

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11651412/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Uncovering depth and spatial patterns of soil environmental DNA in arid regions

Madeleine P. Malmfeldt[1], Denise L. Lindsay[1], Carina M. Jung[1], Mark A. Chappell[1], N. Paige Fowler[1], Autumn Acree[2,†] , Michelle Wynter[1], Yoko Masue-Slowey[1], Adam J. Norris[1,‡] , George M. Ferguson[3], and Richard F. Lance[1*]

[1] Environmental Laboratory, U.S. Army Engineer Research and Development Center, 3909 Halls Ferry Road, Vicksburg, MS, USA, 39180
[2] Oak Ridge Institute, 1299 Bethel Valley Rd, Oak Ridge, TN, USA 37830
[3] University of Arizona Herbarium, Herring Hall, 1130 E. South Campus Drive, Tucson, AZ, USA 85721

[*]*Corresponding Author*

Corresponding author email: Richard.F.Lance@erdc.dren.mil

ORCiD
MPM: 0009-0004-1938-9287
DLL: 0000-0002-4079-6061
CMJ: 0000-0002-3171-7341
MAC: 0000-0003-4989-6533
NPF: 0009-0006-4321-7885
AA: 0000-0002-0212-8783
MW: 0009-0005-2276-5105
YM-S: 0000-0001-7227-9467
AJN: 0000-0003-1532-0456
GMF: 0009-0004-1576-4203
RFL: 0000-0002-9727-4782

[†] Current Address: Geosyntec Consultants, Inc., NE Ste. 1050, 1360 Peachtree St., Atlanta, GA, USA, 30309
[‡] Current Address: Ergon Refining, Inc., 2829 Lakeland Drive, Flowood, MS, USA 39215

Abstract

Environmental DNA (eDNA) holds promise as a bioindicator for determining the habitat and geographic sources of anonymous samples. We investigated the patterns of animal, plant, fungal, protist, and algal eDNA in soil to identify indicators suitable for such provenance analysis. We characterize amplicon sequence variant (ASV) similarity (Jaccard Index) between samples taken at different soil depths and at increasing distances across four distinct habitat types from southeastern Arizona, USA. The contribution diversity ($\boldsymbol{C_\gamma}$) of samples from various depths to overall ASV richness were also measured and compared. Parallel analyses were conducted on sets of "key ASVs" with high bioindicator potential commonly found in samples from a single habitat type. Overall, samples showed limited overlap in ASV sets. PERMANOVA tests of factors influencing Jaccard distance indicated both habitat type and soil depth significantly influence Jaccard distance among samples and among key ASVs. With few exceptions, different soil depth classes did not significantly differ in their contribution to habitat-level ASV richness. There was a general trend toward increasing dissimilarity in both total ASV and key ASV sets with greater distance between sample points, although this pattern varied among habitats and transects. ASV accumulation curves indicated that considerable ASV richness was unaccounted for in all habitat types, suggesting that some observed dissimilarity was attributed to incomplete sampling. Detecting ASV indicators habitats similar to those studied here would benefit from sampling designs that focus on maximizing sample numbers, increase distance between sampling points, and represent as much soil depth as feasible.

Keywords: bioindicators, eDNA metabarcoding, geoforensics, high-throughput sequencing, taxon richness, terrestrial eDNA

Introduction

Since the late 2000s, the term “environmental DNA” or “eDNA” has come to be primarily associated with DNA shed by eukaryotic organisms into environmental matrices such as water, sediment, soil, and air. These DNA-bearing traces are largely deposited through urination, defecation, sloughing of epidermal cells., etc. These traces are used to address important ecological issues such as detecting elusive endangered species (Feist et al., 2022), monitoring biodiversity (Hermans et al., 2022), and tracking the spread of invasive species (Jerde et al., 2013). While early eukaryotic eDNA applications focused on the detecting species in water samples (Ficetola et al., 2008; Goldberg et al., 2011; Jerde et al., 2013), the application range quickly grew to span multiple different environmental matrices for targets ranging from single taxa to entire biological communities, and to address increasingly complex ecological questions (Bienert et al., 2012; Deagle et al., 2005; Jerde et al., 2011; Johnson et al., 2019b, 2019a; Taberlet et al., 2012; Yoccoz et al., 2012). Continued exploration of localized dispersion, transport, and fate of eDNA within different environmental matrices and anticipated local patterns of eDNA should prove useful in developing increasingly powerful eDNA-based capabilities. The natural history of eDNA in aquatic environments (e.g., deposition, transport, decay) has been frequently studied (Andruszkiewicz et al., 2017, 2019; Buxton et al., 2018; Carraro et al., 2022; Fremier et al., 2019; Jane et al., 2015; Jerde et al., 2016; Nevers et al., 2020; Sakata et al., 2020; Shogren et al., 2017; Turner et al., 2015; Wacker et al., 2019). However, local dynamics and eDNA patterns in air and soil matrices, are not as well understood (Bienert et al., 2012;

77 Donhauser et al., 2023; Guthrie et al., 2024; Kirse et al., 2021; Pansu et al., 2015; Rota et
78 al., 2020; Ruppert et al., 2019; Sirois & Buckley, 2019; Taberlet et al., 2012).

79 Many soil eDNA studies to date have focused on soil-dwelling invertebrates (Bienert
80 et al., 2012; Clarke et al., 2014; Elbrecht et al., 2019; Kirse et al., 2021; Lilja et al., 2023;
81 Lunghi et al., 2022; Pansu et al., 2015). However, eDNA from surface-dwelling
82 invertebrates, soil and surface-dwelling vertebrates, and plants can also be locally
83 recovered, providing insights into local surface and subterranean biological communities.
84 One potentially important application for soil eDNA is in geoforensics, where such data may
85 be useful for predicting the origins or provenance of samples with unknown geographic or
86 ecological origins (Ariza et al., 2023; Fløjgaard et al., 2019; Frøslev et al., 2023; Ruffell &
87 McKinley, 2008). Different eukaryotic taxa, from species to ecotypes, and their eDNA are
88 natural indicators of the geographic areas and ecological domains where they occur. Taxa
89 that have ranges that are more narrowly endemic or that are very habitat specific would
90 provide increasingly fine-scale predictions of soil provenance. Furthermore, combinations
91 of taxa with known ranges and/or habitat associations provide even greater weight of
92 evidence than a single indicator taxon. Methods for identifying sets of taxa as
93 complementary indicators, and for characterizing the predictive power of such sets, have
94 been well-developed in ecology (De Cáceres et al., 2012; Dufrêne & Legendre, 1997). The
95 use of indicator species for predicting soil origins has been touched upon in ecology
96 (Fløjgaard et al., 2019; Frøslev et al., 2023), including models based on characterization of
97 pollen grains found in soil using eDNA metabarcoding (Tong et al., 2021).

One approach to predicting the origin of a soil sample using eDNA traces would be to review the many various sequences produced by metabarcoding using a high-throughput sequencing platform, match or assign those sequences to taxa, and then review the known ranges or habitat affinities of all or select subset of species. This methodology is beginning to be employed, particularly with genotyped pollen found in environmental samples, or with plant eDNA found in samples (Tong et al., 2021). Such an approach requires robust, comprehensive databases of the DNA barcode locus (or loci) sequences of key taxa found in the regions of interest. Unfortunately, both species range and species barcode DNA sequence records are likely to be incomplete for many taxa in those regions. Additional challenges, such as shared or ambiguous sets of DNA barcode sequences among phylogenetically-close taxa, or simple mistakes or inaccuracies in the databases, may also confound this approach.

An alternative, though related approach, is to conduct soil sampling surveys in geographic regions or ecological domains of interest to identify eDNA that might serve as effective indicators of soil provenance. Those sequences that appear to be diagnostic for a geographic locale and/or habitat could be subsequently identified to a taxon if a reliable sequence record exists (Fløjgaard et al., 2019; Frøslev et al., 2023), but this would not be a necessary step. Such a taxon-agnostic approach is attractive because archived DNA barcode sequences exist for likely less than 25% of the world's plants, fungi, and animals (Gostel & Kress, 2022). Another advantage of such an approach is that the search for effective indicator eDNAs would not be limited to any particular taxonomic level, as distinct DNA barcode sequences associated with subspecies, populations, or even ecotypes that are

underrepresented or absent in available databases could serve as indicators. An obvious disadvantage of such an approach is that it would generally require large numbers of soil eDNA samples from a wide array of sites and ecological domains within the region of interest in order to robustly predict soil provenance. Thorough characterization and cataloging of soil eDNA indicators across even moderate extents or suites of ecological domains (e.g., habitats) within a geographic region represents a daunting prospect, particularly where wide-ranging systematic characterizations of soils have been undertaken. For example, the U.S. Department of Agriculture's Natural Resources Conservation Service (NRCS) has led the effort to systematically classify and characterize soils across the entire United States of America for the better part of a century.

For animals, eDNA deposition in soil occurs via urination, defecation, epidermal cell sloughing, and predator defecation, while plants release soil eDNA via root exudates, leaf litter, pollen and seeds, and herbivore defecation (Ariza et al., 2023; Ruppert et al., 2019). Live and dead fungi are generally found in high numbers in soil, but because they are primarily saprophytes, fungal densities will be concomitant with the availability of soil organic carbon (Gordon & Van Norman, 2021; Rosa et al., 2020). Though we expect that soil eDNA will experience comparatively little spatial dispersion and to be more patchily distributed than eDNA in more fluid or liquid-saturated environments (Taberlet et al., 2012; Villacorta-Rath et al., 2022), there are several processes that may transport and disperse soil eDNA. Both water (e.g., rain run off) and wind can transport surface and near-surface soil eDNA (Clare et al., 2021, 2022; Littlefair et al., 2023). For example, water percolation (e.g., seepage) transports eDNA vertically within the soil, including relatively extensive

transport via soil macropores (Haile et al., 2007; Lunghi et al., 2022; Luo et al., 2010; Pietramellara et al., 2009). It is also likely that eDNA bound to soil particulates may also be transported through soils via bioturbation (e.g., mixing of soil layers by burrowing animals or growing plant roots; Davidson et al., 1999; Prosser & Hedgpeth, 2018). These processes all influence the detectability and reliability of soil eDNA as provenance indicators.

Another consideration with soil eDNA is that some or all of that eDNA will exhibit lower degradation and decay rates than typically observed in aquatic systems. Under select conditions soil eDNA has been found to be preserved for thousands of years (van der Heyde et al., 2020). Abiotic adsorption and complexation with the surfaces of soil particles protects eDNA by increasing the activation energy required for degradation by microbial exonucleases (Lorenz & Wackernagel, 1994; Pietramellara et al., 2009). The complex influence of environmental factors driving soil eDNA fate and persistence, such as temperature, soil moisture, pH, microbial load, etc. (Cai et al., 2006; Lorenz & Wackernagel, 1994; Ogram et al., 1994; Trevors, 1996), are as variable as those factors driving soil formation and genesis, but vary across space, season, and habitat. However, the combined effects of these factors on soil eDNA presence and persistence remain poorly understood. Research that explores soil eDNA patterns in particular geographic regions and among multiple ecological domains is needed to fully realize the potential of soil eDNA for provenance and ecological inference.

Study objectives

Our study aimed to identify patterns of animal, plant, and fungal eDNA in the soils from a semi-arid region, with a particular emphasis on eDNA traces that could serve as potential bioindicators for different desert and montane habitat types. We investigated the influence of sampling depth on soil eDNA detection and diversity, hypothesizing that increased depth (Davidson et al., 2002; O'Brien et al., 2005) would yield more diverse eDNA sets and greater detection of indicator ASVs. However, some natural forces such as bioturbation and water percolation could obscure such patterns (Andersen et al., 2012; Davidson et al., 1999; Haile et al., 2007; Hebsgaard et al., 2009; Lunghi et al., 2022; Pietramellara et al., 2009; Prosser & Hedgpeth, 2018). The optimum sampling depth for soil ASV characterizations will be a function of these dynamics, along with practical consideration of the time and effort required for digging to different depths in different soils. We also sought to characterize horizontal patterns in soil eDNA, which could inform optimal sampling across a plot or transect. These patterns likely emerge as a result of multiple complex factors including, but not limited to, plant dispersal and growth determinants, wind transport, intermittent surface flows from rain, and patterns in animal spatial use (e.g., game trails).

Materials and methods

Sampling locale and transects

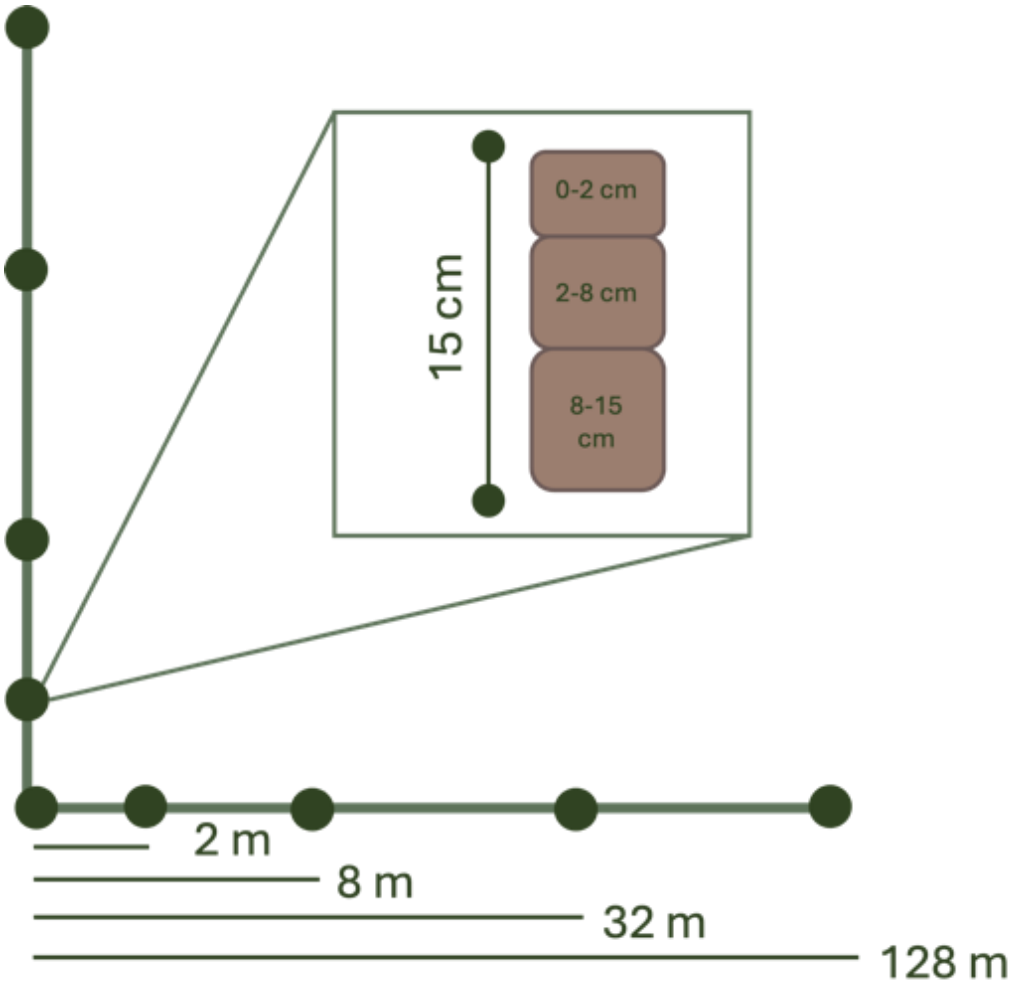
Soil samples were collected from four distinct habitat types (hereafter, HT) at Fort Huachuca, AZ, USA. This US Army installation is located within portions of the "sky island"

Huachuca Mountain Range and adjacent foothills and lowlands. The highest peak of the Huachuca Mountains is 2885 m above sea level, approximately 1550 m higher in elevation than the surrounding desert. This elevational gradient created diverse HTs within a concentrated geographic region, including the four HTs included in this study: desert scrubland ("Desert"), desert grassland ("Grassland"), open oak-juniper savannah/encinal woodland ("Encinal"), and denser pine-oak forest ("Pine-Oak"; Online Resource 1). Overall, this region receives on average 360.7 mm of precipitation annually, with the majority of rainfall (avg. 96.5 mm) occurring during the monsoon season in mid- to late summer. This monsoonal period represents an important peak in biological activity in the region.

Three L-shaped transects, each comprised of two 128 m segments arranged at right angles, were established within each HT in February and April 2022. HTs were identified using existing vegetation survey maps and visual assessments of plant species cover within the 128 m x 128 m plot (Online Resource 1). Each transect included a center point and four sampling points in each direction at 2, 8, 32, and 128 m from the center. This plot design enabled evaluation of a variety of lateral (n = 20) and diagonal (n = 16) interpoint distance pairings ranging from 2 –181 m. Center points of all transects were at least ≥ 0.9 km apart, with no two points from any two transects < 0.4 km apart (Figure 1).

Depth sampling

In April 2022, we randomly selected three sample points along each L-shaped transect within each HT. At each sample point, a 15 cm diameter hole was excavated to collect soil



from three depths: surface (0-2 cm), middle (2-8 cm), and bottom (8-15 cm) (Figure 1). Loose organic matter (e.g., leaf litter, large roots) was removed prior to sampling along with any observed fauna (e.g., beetle grubs, etc.). Each sample collected at a different depth was poured into a separate 4 L sealable storage bag and vigorously shaken by hand in an effort to physically homogenize the material. An approximately 30 mL volume of that soil was then subsampled and transferred to a 50 mL conical centrifuge tube containing 30 mL of

RNAlater stabilizing solution (Invitrogen™ by ThermoFisher Scientific, Waltham, MS, USA). The tubes were then shaken by hand to fully saturate the soil. Afterwards, the tubes were stored in a cooler on ice until transferred to a 4°C refrigerator at the end of each field day. Prior to this study, we verified that this method was sufficient to maintain DNA integrity (data not published). All samples were kept in sealed storage bags, grouped by transect and continuously refrigerated. Disposable gloves were worn while taking soil samples and changed for each new sample collected. Soil knives or hand trowels were sterilized between uses by scrubbing with soap and water to remove residual soil particles, rinsing in tap water to remove soap residue, soaking in 15% bleach solution for 10 minutes, and finally rinsing in distilled water to remove the bleach. Clean and used knives were stored separately. At the conclusion of fieldwork, all soil eDNA samples were shipped in a cooler with ice packs overnight to USACE – ERDC, Vicksburg, MS for storage at 4°C until further processing.

Distance sampling

In August 2022, surface (0-2 cm) soil eDNA samples were collected at all designated sampling points (n = 9 samples per transect) along the previously established transects, following the same methods used previously. Where a point had been sampled previously, a new sample hole was taken adjacent to the earlier hole.

DNA extraction

DNA was extracted from 0.25 g of sample soil using the DNeasy 96 PowerSoil Pro QIAcube HT Kit with the QIAcube HT liquid handling machine (Qiagen, Maryland, USA). To maximize the amount of eDNA isolated for each sample, we performed six replicate DNA extractions per sample and pooled the extracts before performing PCR (Hermans et al., 2022). The concentration of the eDNA samples was validated as sufficient for amplification on the NanoDrop One (ThermoScientific).

Sequencing library preparation

We employed four primer sets: MICOlintF + PolyShortCoiR (metazoans; Carr et al., 2011; Leray et al., 2013; Robeson et al., 2018), UniplantF + UniplantR (plants; Moorhouse-Gann et al., 2018), fwhF2 + fwhR2n (macro-arthropods; Vamos et al., 2017), and Earth Microbiome Project primers ITS1F + ITS2R (fungi; Gilbert et al., 2014). PCR amplifications were carried out using AccuStart™ II PCR ToughMix (Quantbio, Beverly, MA) following published protocols, except for changing the initial Taq activation temperature to 94° C for 3 min to accommodate the recommended protocol for AccuStart. Initial PCR was performed to amplify the target region and to append the Illumina overhangs to the amplicon. This reaction was performed in triplicate, then 10 μl subaliquots for each amplified product were pooled. We included one no-template control (NTC) on each 96-well plate, then pooled the NTCs for sequencing. Samples were indexed with 5 μl of IDT® for Illumina® DNA/RNA UD Indexes sets A-D (IDT, Coralville, IA, USA), 12.5 μl of 2x KAPA HiFi HotStart ReadyMix (Roche, Basel, CH), and 7.5 μl of pooled amplified product. The reaction was

performed at an initial 95°C for 3 min, then eight cycles of 95° C for 30 s, 55° C for 30 s and 72° C for 30 s, followed by an elongation step of 72° C for 5 min. Amplified products were cleaned following both amplicon PCR and indexing PCR using Select-a-Size DNA Clean & Concentrator MagBead Kit (Zymo Research Corp., Irvine, CA, USA), removing fragments below 200 bp. Libraries were denatured and diluted according to the Illumina protocol for the MiSeq (Illumina, 2013) with modification to add 5 μl of 200mM Tris-HCL pH 7.0 during initial library and Phi control v3 (Illumina) denaturation. eDNA samples from the April and August trips were sequenced separately, except for the fungal assay, for which all samples were run together. Libraries were diluted to between 10-12 pM with a 30% PhiX spike. The fungal samples were sequenced using the 500-cycle v2 MiSeq reagent kit (Illumina), all others, were sequenced using the 600-cycle v3 MiSeq reagent kit (Illumina). Each sequencing run included the library from one assay, all 112 soil samples, and one NTC.

Bioinformatics

Demultiplexed samples were analyzed with Qiime2-amplicon-2023.9 (Bolyen et al., 2019) The amplicon primers were removed using the Qiime2 cut-adapt plugin (Martin, 2011) and the resulting sequences were denoised using the DADA2 plug in (via q2-dada2) (Callahan et al., 2016) with default settings, except that sequences were trimmed and truncated to eliminate cycles with a quality score < 30. ASVs produced from soil samples that were also produced in an NTC were identified as “suspect reads” and were eliminated from the final sample ASV sets if their read count was less than double the read count found for the same

ASV in the NTC sequence set. We did not cluster ASVs into operational taxonomic units (OTUs, i.e. 97% or greater similarity) as we reasoned that interesting patterns might be found at finer phylogenetic scales than would typically correspond to a species-level OTU.

Statistical analyses

For both the April 2022 and August 2022 samples Venn diagrams depicting ASV overlap within and between HTs were generated using the R package '*ggVennDiagram'* (Gao et al., 2021, 2024).

Depth sampling statistical analyses

For each HT included in the April 2022 depth sampling effort, Venn diagrams showing ASV overlap found within and shared between sampling depth classes were generated as described above. The difference in the ASV richness between the depths was assessed via analysis of variance (ANOVA) with richness as the dependent factors and HT and depth as the independent factors. The q2-diversity function in Qiime2 (Legendre & Legendre, 2012) was used to calculate Jaccard Distance values of ASV set dissimilarity between all samples and to derive a Principal Coordinate Analysis plot for HTs and depth classes (PCoA; Halko et al., 2011). A permutational multivariate analysis of variance (PERMANOVA) with the Jaccard Distance values (i.e., 1-Jaccard Index) as the dependent variable, and sample depth class, transect, and HT as independent factors, was calculated using the 'adonis' function in the R

296 package vegan. The package pairwiseAdonis (Martinez Arbizu, 2017) was used to execute
297 post-hoc pairwise analyses (with a Bonferroni correction of the critical value for multiple
298 comparisons (i.e., adjusted *p*-value). Additionally, the differing contributions of soil depth
299 classes to overall ASV richness for an HT (i.e., $\boldsymbol{\gamma}$-diversity) were characterized by the relative
300 contribution diversity metric $\boldsymbol{C_\gamma}$ described by Lu et al. (2007). $\boldsymbol{C_\gamma}$, in the case of
301 presence/absence of ASVs, reflects the proportional contribution of a sample to the overall
302 ASV richness in a transect, while the mean depth class ($\overline{C_\gamma}$) values for across all transects in
303 an HT describe the contribution of that depth class to ASV richness in the HT. The
304 contribution of each sample to total ASV richness, or $\boldsymbol{\gamma}_{s(k)}$, was calculated using the
305 contridiv function in the R package "vegan" (Oksanen et al., 2001). $\boldsymbol{C_\gamma}$ was then calculated
306 as described in Lu et al. (2007). Because the $\boldsymbol{C_\gamma}$ data for some HTs significantly deviated
307 from key assumptions for parametric tests, namely homogeneity of variance and normal
308 distribution, differences in $\boldsymbol{C_\gamma}$ between depth classes within each HT was assessed using
309 nonparametric aligned rank transform ANOVA (ART-ANOVA) with factor interactions.
310 Transect was included as an additional "nuisance factor" in the analyses. The R package
311 "ARTool" was used to conduct the ART-ANOVAs (Elkin et al., 2021; Kay et al., 2025;
312 Wobbrock et al., 2011). Post-hoc pairwise comparisons of differences among depth classes
313 were performed using the R package "emmeans" as described in the ARTool package (Lenth
314 et al., 2025). $\boldsymbol{C_\gamma}$ as a function of soil sample depth was similarly tested for the key ASV
315 dataset.

316

317 **Distance sampling statistical analyses**

To visualize general patterns in ASV constituencies among HTs in the August 2022 samples, a PCoA plot based on JI values was constructed as described above. ASV accumulation curves were generated using the specaccum function in the R package vegan. Accumulation curves were generated both for the entire HT and for the individual transects within the HTs. Both Pearson correlation and a nonparametric alternative, Spearman correlation, were calculated between JI values and interpoint distances for the overall dataset and within each HT. A linear regression analysis of JI values as the dependent variable, interpoint distances and HTs as the independent factors, and with factor interactions, was performed for the overall dataset using base R. Post-hoc tests of between-HT differences in regression slope (or "trend") were performed using emtrends function in the "emmeans" package. This function calculates within and compares estimated marginal means of linear trends between different factor classes. We next fit a linear regression model within each HT with JI value as the dependent variable, interpoint distance as a factor, and transect as a (nuisance) factor. Post-hoc tests of between-transect differences in regression slope were tested as described above. Because the datasets obtained from our study violated multiple assumptions underlaying linear regression analysis (e.g., independence of observations, homoscedasticity) we employed nonparametric tools to further assess the model. The statistical significance of the adjusted R^2 value for each model and the absolute t-values for model coefficients were compared to ranges of these values generated from 10,000 bootstrapped datasets in which JI values were resampled with replacement among all observations (i.e., within transect interpoint pairings). Similarly, significant contrasts in slope between HTs and between transects, as detected in the estimated marginal means of

linear trends analysis, were further validated by comparing absolute value of the t-value ratios for these contrasts to the absolute value of ratios generated from the same 10,000 bootstrapped datasets.

ASVs that were encountered in all transects of an HT, and only in the transects of that HT, were identified using the "multipatt" function in R package indicspecies 1.7.15 (De Cáceres & Legendre, 2009). These ASVs were termed "key ASVs" for this study and were considered to have the potential to emerge as HT-indicator ASVs, should a more sampling-intensive study than ours be conducted in the region.

For the April 2022 dataset, the effect of sampling at increasing soil depth on the detection of key ASVs was assessed by 1) determining the number and relative frequency of key ASVs that were found in only a single soil depth class and 2) the change in cumulative number of key ASVs identified for each HT with increasing sampling depth.

For the August 2022 distance sampling set, the effect of number of sampling points and distance between sampling points on the detection of key ASVs was assessed by 1) calculating mean frequency of individual key ASV encounters within transects within each HT (minimum frequency = 1 sample point, maximum frequency = 9 sample points), 2) fitting and testing correlation and linear regression models to the data as described for the full dataset, and 3) assessing the change in numbers of key ASVs detected for each HT as the number of sample points per transect increased. For this last objective an R script was developed to randomly select a user-determined number of sample points per transect for each transect in each HT and to identify key ASVs as described above, and then to iterate

these sets of analytical steps 100 times each for 1 – 8 randomly selected sample points per transect. The mean number of key ASVs, with 95% confidence intervals, were then plotted for each sample number class for each HT.

Taxonomic Classification of ASVs

Merged and cleaned sequences were exported from Qiime2 and assigned to a taxonomic classification level using command line blast (Blast+) (Camacho et al., 2009) with the MIDORI2 reference database (Leray et al., 2022) for assays targeting the COI loci ('Genbank 259, 12-17-2023'). The ITS targeted loci were compared against the UNITE database (Nilsson et al., 2019). Any classification that had an alignment length between the subject and query of less than 200 bp was removed, and all singleton reads were removed. ASVs provided by primer sets targeting other taxa (e.g., fungal taxa detected using the metazoan-focused set MICOlintF + PolyShortCoiR) were incorporated in all datasets and analyses.

Results

General sequencing and classification results

The four universal primers sets and 216 combined samples produced > 8.6 x 10^7 raw sequence reads, with largely consistent outputs per primer set between the two sets of samples (Online Resource 2). The two samples sets produced similar numbers of raw sequences, and similar proportions of ASVs determined to be high quality and of ASVS that

met our criteria for classification (Table 1). Likewise, the orders of magnitude of protist ASVs assigned to animals, plants, fungi and protists were similar for both sampling events (Table 1).

Table 1. Numbers of raw paired unmerged sequencing reads generated for each set of soil samples obtained for assessing different components of eukaryotic DNA distribution patterns in soil, as well as numbers merged reads meeting quality filtering criteria, and the numbers of quality sequences and unique amplified sequence variants (ASVs) assigned to each of several broad taxonomic groups. "Algae" refers here only to eukaryotic algae. Protists refers broadly to microscopic eukaryotes (e.g., Phylum Amoebozoa, Phylum Oomycota). "Invertebrates" refers broadly to multicellular animals lacking notochords or vertebral columns (e.g. Phylum Arthropoda, Phylum Annelida). "Vertebrates" refers here to Subphylum Vertebrata (Phylum Chordata). Unclassified reads were those quality reads that either lacked strong sequence matches or were identified as bacteria.

2022 Event	Raw Reads Unmerged	Quality Reads Merged	Classification Outcomes (Reads)		ASVs
April	$3.79e^7$	$1.65e^7$	Animalia: Invertebrates	$4.93e^5$	2,886
Sampling Depth			Animalia: Vertebrates	$2.50e^1$	2
			Plants	$4.48e^6$	2,485
			Fungi	$3.09e^6$	10,652
			Protists	$2.57e^5$	1,807
			Algae	$5.24e^4$	554
			Unclassified	$6.26e^6$	30,901
August	$4.82e^7$	$1.62e^7$	Animal: Invertebrates	$5.47e^5$	4,053
Sample Spacing and Numbers			Animal: Vertebrates	$5.40e^1$	3
			Plants	$4.43e^6$	3,583
			Fungi	$3.23e^6$	17,569
			Protists	$1.80e^5$	2,418
			Algae	$2.59e^4$	769
			Unclassified	$5.85e^6$	45,492

Habitat and general soil eDNA patterns

A total of 123,174 distinct ASVs were encountered in our two sets of samples, with 45,458 that appeared to originate from target taxa (animals, plants, fungi, and protists; Table 1). There were 4,046 ASVs from the total ASV pool that appeared in both the April and the August sampling trips. Across both sampling events, the majority of ASVs that were encountered within an HT were particular to that HT (Table 2). Compared to the April samples, the August samples resulted in 143%, 65%, 31%, and 16% more total ASVs in Desert, Grassland, Encinal, and Pine-Oak, respectively (Table 2). Also, in comparison to the April samples, the August 2022 samples contained 12%, 12%, 13%, and 4% more HT-particular ASVs in the Desert, Grassland, Encinal, and Pine-Oak HTs (Table 2). Pairwise similarities among samples within the same transect (i.e., JI values) were generally low overall (mean JI = 0.071, SD = 0.060, Table 2), with apparently higher pairwise similarities in the April samples (mean JI = 0.103, SD = 0.065, min = 0.008, max = 0.303) than in August samples (mean JI = 0.037, SD = 0.026, min = 0, max = 0.122). Samples from the August 2022 field event exhibited a lower degree of overlap between HTs than in the April 2022 samples (Table 3; Fig. 2).

Table. 2. Numbers of amplified sequence variants (ASVs) of targeted taxa identified within DNA metabarcoding sequencing read datasets for soil eDNA samples taken from four habitat types (HT) during each of two 2022 field sampling events. Additionally, the percentages of ASVs particular to each HT (HTP-ASV) and mean Jaccard Index values (with standard deviation) for those HTs are provided.

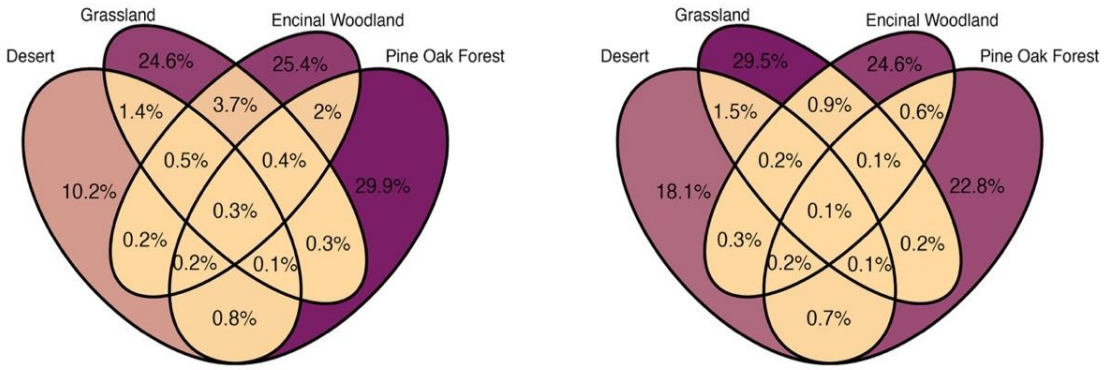
HT	Apr Total ASVs	Apr % HTP-ASVs	Mean JI	Aug Total ASVs	Aug % HTP-ASVs	Mean JI
Desert	6,089	73%	0.070 (0.041)	14,823	85%	0.039 (0.033)
Grassland	13,808	78%	0.135 (0.069)	22,841	90%	0.060 (0.020)
Encinal	14,412	78%	0.098 (0.071)	18,869	91%	0.021 (0.016)

Pine-Oak	14,978	88%	0.110 (0.058)	17,354	92%	0.030 (0.012)

418

419 Habitat and eDNA patterns by depth

420 With the April 2022 soil depth samples most of the HTs exhibited similar numbers of ASVs
421 — except Desert (Table 2), which provided lower number of ASVs and significantly lower
422 ASV richness ($p < 0.001$). The lowest proportion of shared ASVs between HTs was between
423 Grassland and Pine-Oak and the highest proportion between Grasslands and Encinal (Table
424 3; Fig. 2). A small percentage of ASVs were encountered in all HTs (0.2%; Fig. 2).
425 PERMANOVA results from the April 2022 trip indicated that ASV composition differed
426 significantly between each HT (p =0.001), except between Desert and Encinal, and between
427 Desert and Pine-Oak Forest (Table 3; Figure 3) and between the depths within each HT
428 (adjusted p = 0.05).



429 **Fig. 2.** Proportions of all ASVs found in particular HTs and in the different
430 combinations of HTs for the April (left) soil depth-focused and August interpoint
431 distance-focused (right) sampling events in 2022. Darker colors on the color gradient
432 indicate larger percents and lighter colors indicate smaller percents.

433

Table 3. Percent overlap in ASV constituency among habitat types (HT) from April and August 2022 sampling events and statistical significance of pairwise comparisons of ASV dissimilarity in PERMANOVA tests.

		Apr 2022		Aug 2022	
HT		**Overlap**	**Adj. P value**	**Overlap**	**Adj. P value**
Grassland	Desert	3.7%	0.024	3.1%	0.012
Grassland	Pine-Oak	1.1%	0.006	0.9%	0.006
Grassland	Encinal	5.4%	0.006	2.0%	0.006
Desert	Pine-Oak	2.1%	0.450	2.2%	0.006
Desert	Encinal	2.1%	0.060	1.6%	0.006
Pine-Oak	Encinal	3.0%	0.006	1.8%	0.006

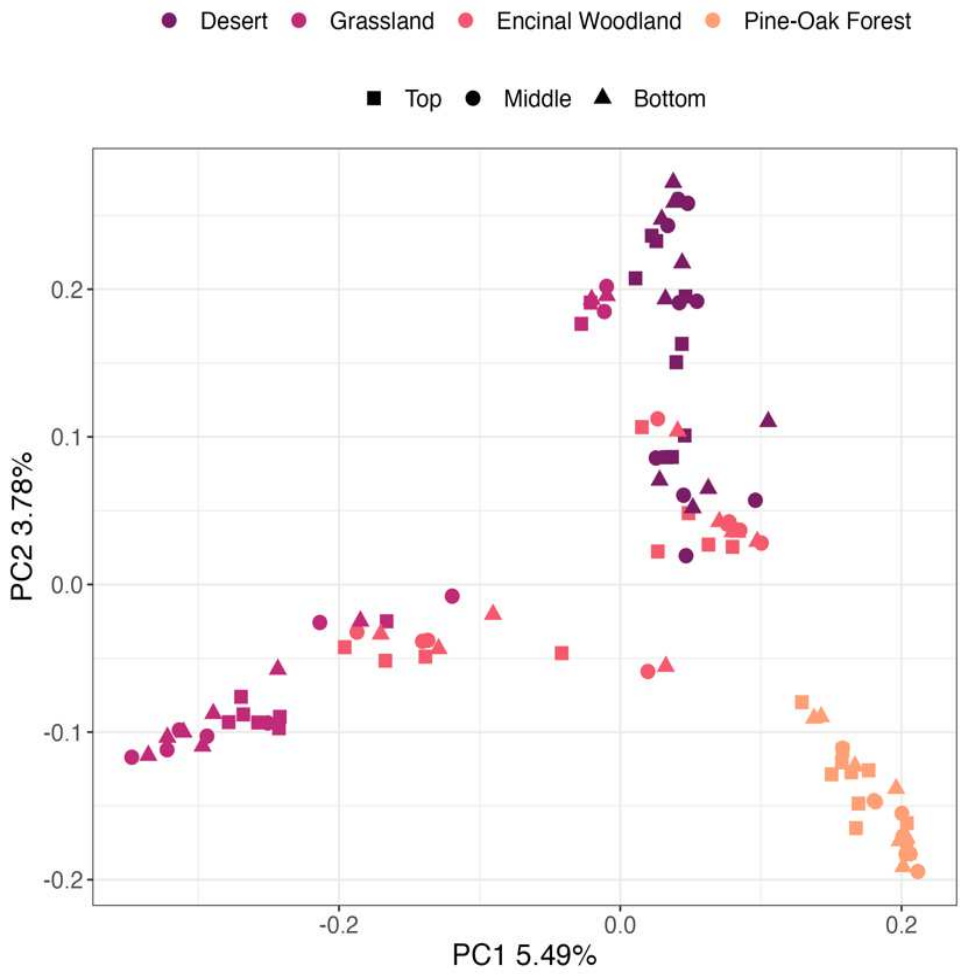


Fig. 3. Principle coordinate analysis (PCoA) of differences in sets of eDNA amplicon sequence variant (ASVs) taken from four habitat types (HTs) in April

2022. In this sampling event, soil samples were taken from three depth classes from three randomly selected points in three plots per HT. Depth classes included top (surface down to 2 cm), middle (2 cm down to 8 cm), and bottom (8 cm down to 15 cm).

Habitat and key ASVs by depth

Only very small portions of the ASVs that were encountered in the April 2022 samples were categorized as key ASVs, with the percentage ranging from 0.0004% in the Desert samples to 0.01% in the Pine-Oak samples (Tables 2, 4). The number of key ASVs encountered varied widely among HTs and did not reflect the relative numbers of high-quality ASVs encountered within each of the HTs (Tables 2, 4).

Table 4. The number (N) of key ASVs (or potential habitat type indicator ASVs) encountered in April 2022 field collections at three depth classes for each of three randomly selected sample points in each of four habitat types (HT). Also, key ASVs encountered in only a single soil depth class (SDC), and the relative frequency (RF) of such key ASVs compared to all key ASVs for each HT. The total number of classified ASVs particular to each HT is found in Table 3.

HT	Key ASVs (N)	Key ASVs SDC (N)	Key ASVs SDC (RF)
Desert	6	2	33.3%
Grassland	189	14	7.4%
Encinal	24	3	12.5%
Pine-Oak	451	18	4.0%

Habitat and eDNA patterns with interpoint distance

With the August 2022 interpoint distance samples, Desert again produced the lowest number of ASVs, but not to the degree as observed in the April samples (Table 2). Grassland samples from the August 2022 effort produced a higher number of ASVs than the Encinal and Pine-Oak to about the same degree as Desert produced fewer ASVs (Table 2). The lowest overlap in ASV constituencies was between Grassland and Pine-Oak and the highest overlap between Desert and Grassland (Table 3; Fig. 2). Again, a small percentage of ASVs, including plants, animals, and fungi were encountered in all HTs (0.1%; Fig. 2). PERMANOVA results indicated that ASV composition differed significantly between each HT (adjusted $p < 0.05$; Fig. 4).

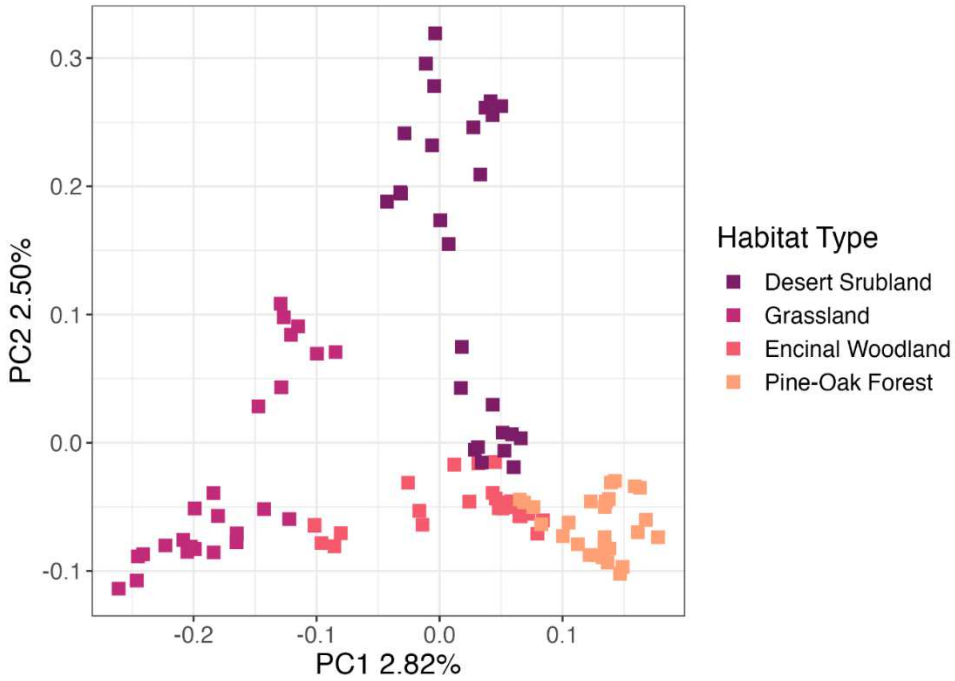


Fig. 4. Principle coordinate analysis (PCoA) of differences in eDNA amplicon sequence variant (ASV) sets taken from four habitat types (HTs) in August 2022. In this sampling event near-surface soil samples were taken from nine points within each of three plots per HT.

Habitat and key ASVs by interpoint distance

As with the April 2022 samples, only very small portions of ASVs were categorized as key ASVs (Tables 2, 5), with the percentage ranging from 0.0002% in the Encinal samples to 0.01% in the Pine-Oak samples (Tables 2, 5). The number of key ASVs encountered in samples from the August 2022 sample set varied widely among HTs, and did not reflect the relative numbers of ASVs encountered within each of the HTs (Tables 2, 5).

Table 5. The number (N) of key ASVs found within each habitat type (HT) for the August 2022 sample set. The mean (and standard deviation) of the number of sample points at which key ASVs were encountered within a transect.

Plant Community	Key ASVs (N)	Mean (SD) Frequency of Key ASVs
Desert	28	2.58 (1.32)
Grassland	78	2.38 (1.31)
Encinal	4	1.67 (0.33)
Pine-Oak	139	2.52 (1.42)

eDNA patterns by soil depth

Though there were some differences in the percent of ASVs particular to each depth class in each HT, depth class was not a significant factor contributing to differences in ASV richness in the April 2022 dataset (p = 0.685; Fig. 5). The relative percentages of ASVs classified to different taxonomic groups within each depth class were broadly similar across HTs (Fig. 6).

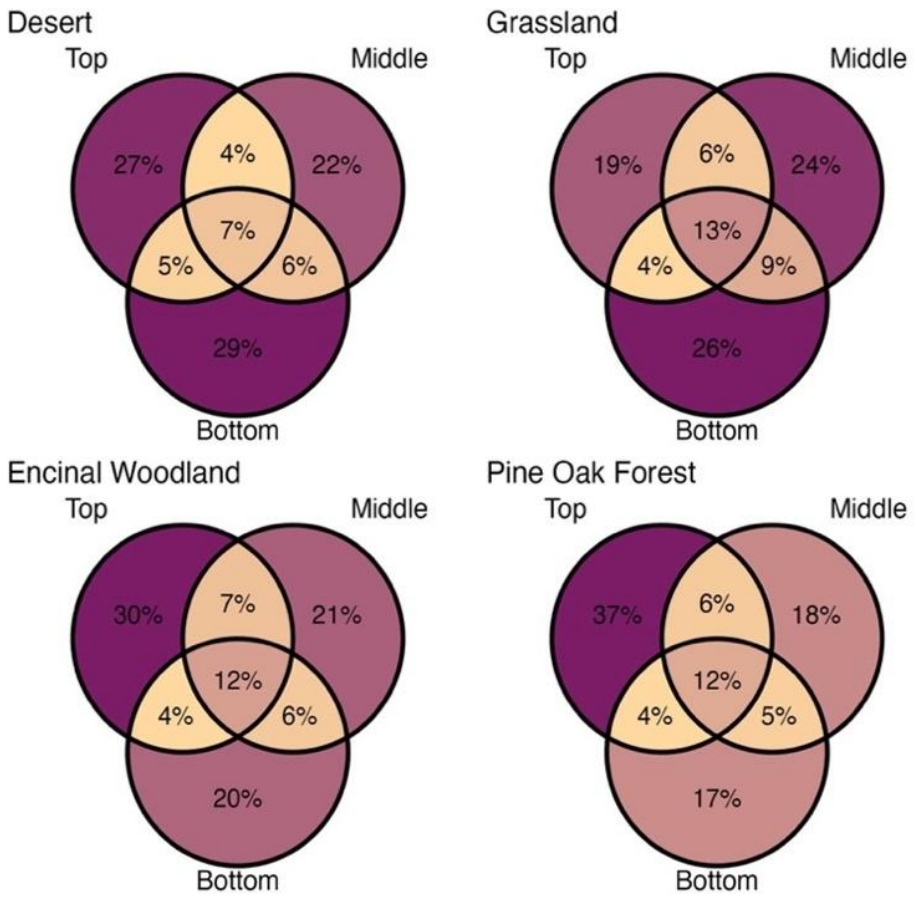


Fig. 5. Venn Diagrams displaying the percent of shared ASVs between each of the sampled depth classes by habitat type. Darker colors on the color gradient indicate larger percents and lighter colors indicate smaller percents.

Depth and ASV contribution diversity

The relative contribution diversity values ($\boldsymbol{C_\gamma}$) for different sampling depth classes (Online Resource 3) were not significantly different ($p > 0.106$) within any of the HTs, apart from Pine-Oak ($p = 0.034$; Fig. 7). Within Pine-Oak transect ($p = 0.030$) and the depth class-transect interaction ($p = 0.010$) were also statistically significant factors (Fig. 7). In particular, there was a significant pairwise difference in $\boldsymbol{C_\gamma}$ values between the top and bottom depth classes in Pine-Oak, with the top

depth class samples having greater contributions to ASV richness than bottom depth class samples (adjusted $p = 0.029$; Fig. 7).

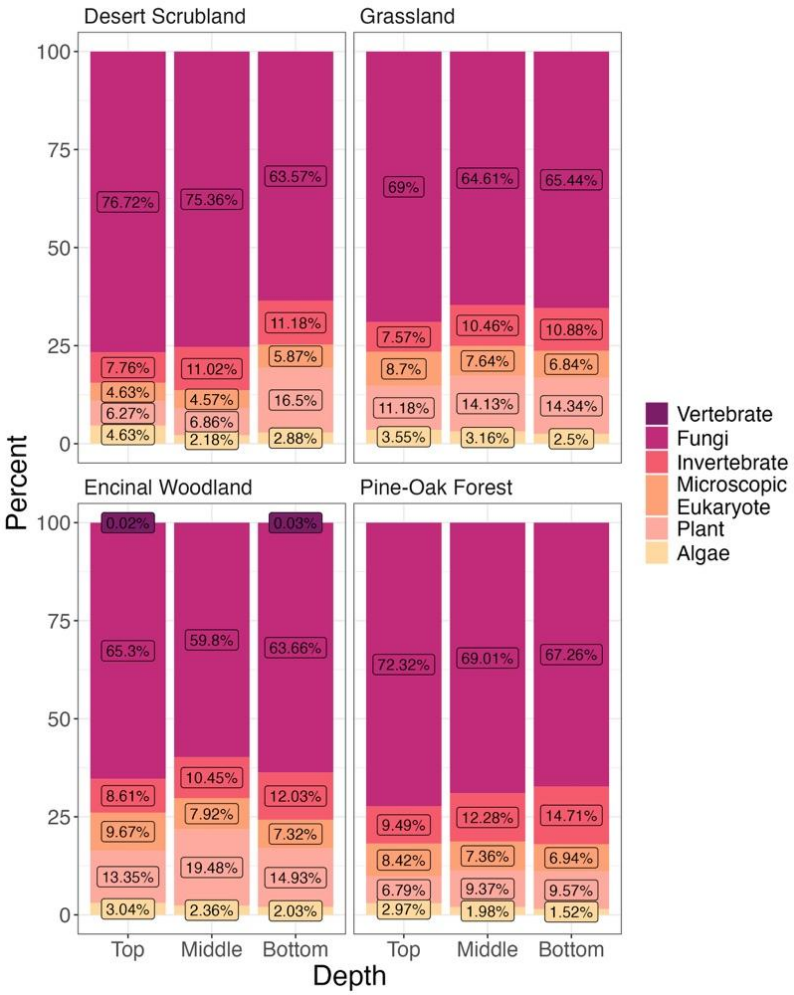
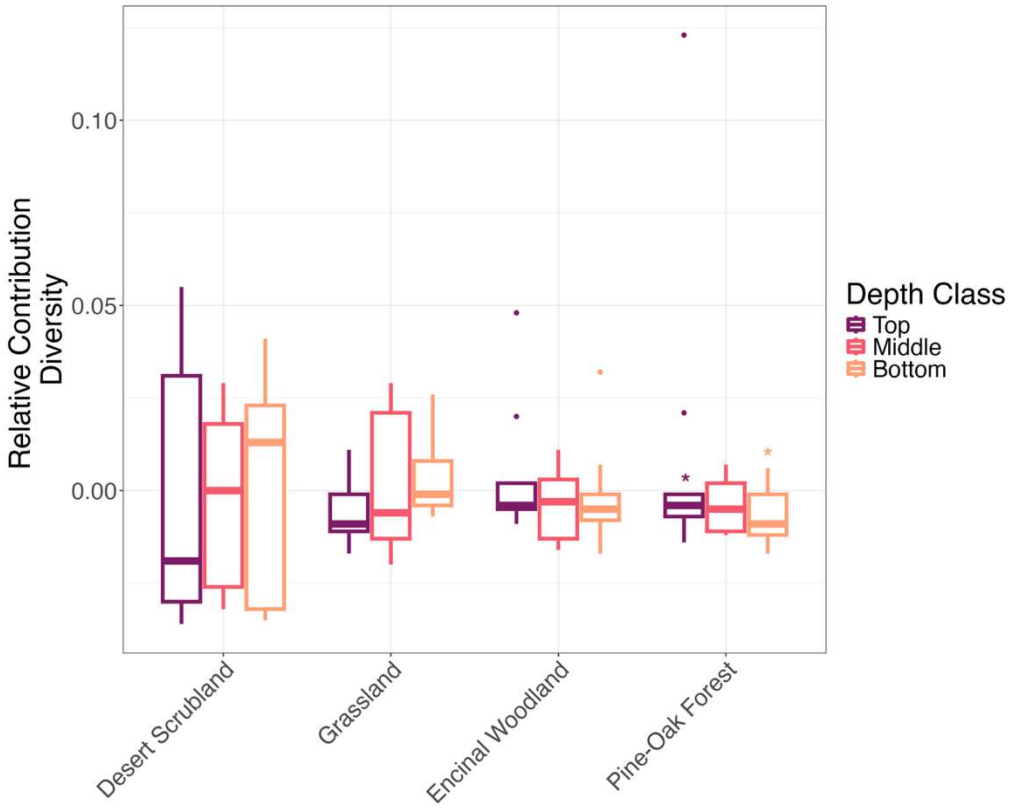


Fig. 6. Percent contribution for each taxonomic classification to the top, middle, and bottom soil depths by habitat type. Boxed percentages indicate the percent contribution of each taxon to each soil depth.

There was also a significant effect of transect on $\boldsymbol{C_\gamma}$ values within the Grassland samples (adjusted $p = 0.024$). Depth classes did not significantly influence key ASV $\boldsymbol{C_\gamma}$ (Online Resource 3) in the Grassland ($p = 0.778$), Encinal ($p = 0.064$), or Pine-Oak ($p = 0.265$) HTs, nor were any of the pairwise contrasts statistically significant (adjusted $p > 0.05$). The

520 Desert HT produced too few key ASVs, with too many samples lacking any key ASV, for the

521 relative contribution diversity modeling and tests.



522 **Fig. 7.** Range of contribution diversity values (C_γ) associated with depth
523 classes within each habitat type. Significant differences between the relative
524 contribution diversity of depth within HT is denoted with '*'.
525

526 **Depth and key ASVs**

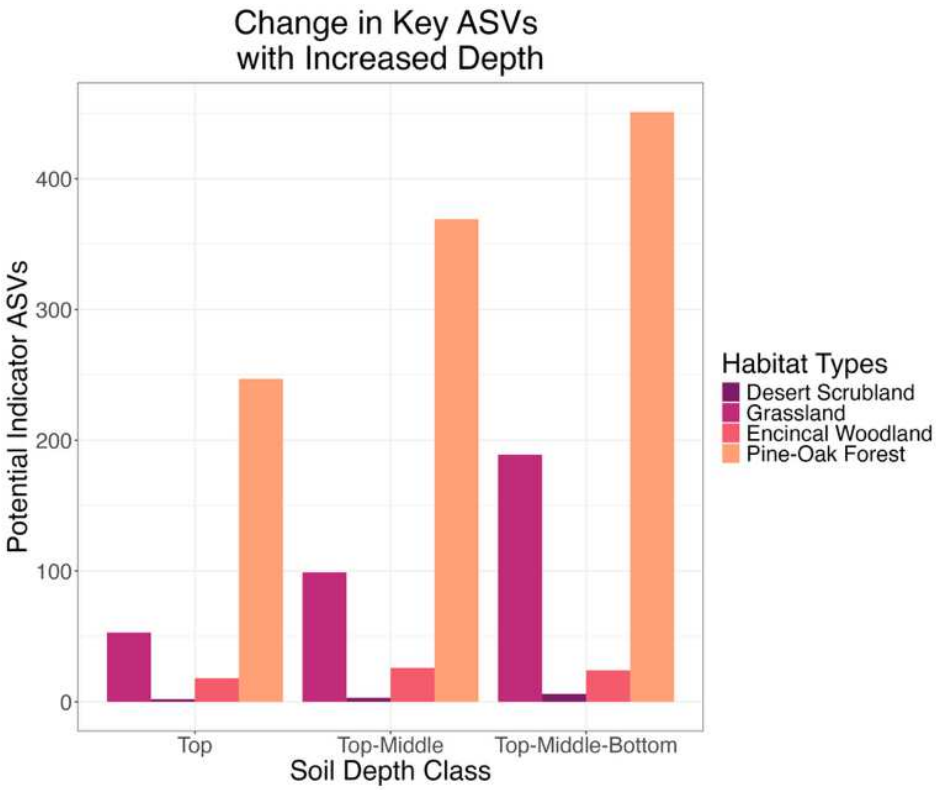
527 The majority of key ASVs were found in multiple soil class depths as opposed to occurring

528 within a single depth class (Table 4). The number of key ASVs encountered in each HT

529 increased with added sampling depth, except for Encinal (Fig. 8; Online Resource 4).

530

531



532 **Fig. 8.** Cumulative numbers of key ASVs encountered with added depth
533 of soil samples for different habitat types in southern Arizona, USA.
534

535 **eDNA patterns with interpoint distance**

536 JI similarities for each interpoint distance class (ranging from 2 – 181 m) were generally low
537 across all HTs (Fig. 9). There was a significant negative correlation between JI similarity and
538 interpoint distance across all samples combined from all HTs (Pearson r = -0.256, $p < 0.001$;
539 Spearman r = -0.271, $p < 0.001$). A linear regression model of JI similarity as a function of
540 interpoint distance, with HT as a co-factor, was significant (adjusted R^2 = 0.376, $p < 0.001$)
541 and inverse (coefficient = -0.00015; $p <$ bootstrap $p < 0.0001$), though the interaction
542 between these factors was not ($p > 0.170$). The effect of HT on this relationship was
543 significantly different between all HTS ($p < 0.001$; bootstrap $p < 0.0001$), though there were

no significant pairwise differences in slopes (adjusted $p > 0.423$; Fig. 9).

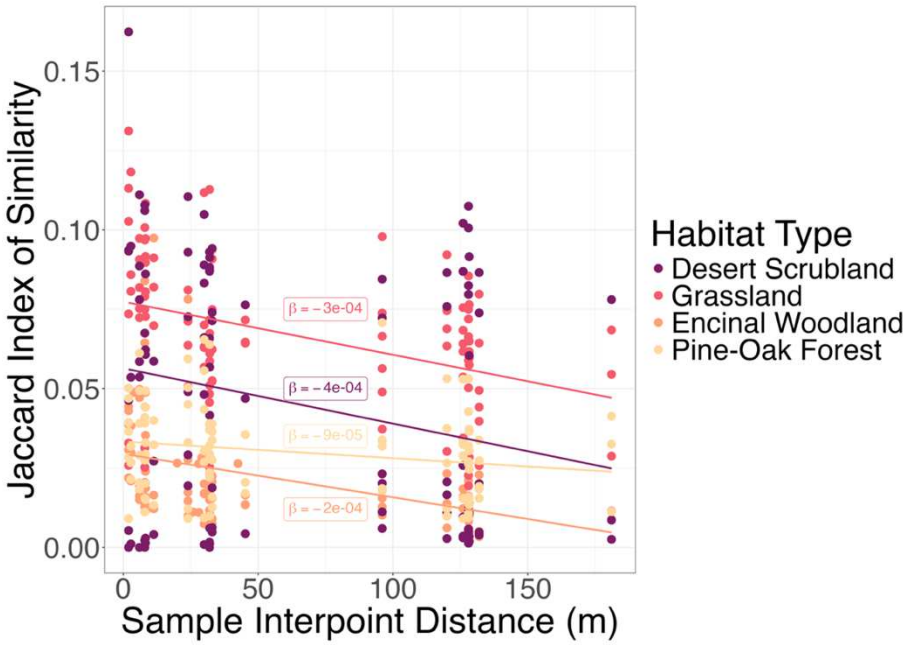


Fig. 9. Plot of Jaccard Index (JI) values against interpoint distances for ASVs encountered in four different habitat types (HT). Slope lines derived from statistically significant linear model coefficients are provided.

Interpoint distance and ASV similarity within habitats

Within each HT there were also significant negative correlations between JI similarity and interpoint distance (Fig. 9): Desert, $r = -0.270$ ($p = 0.027$); Grassland, $r = -0.409$ ($p < 0.001$); Encinal, $r = -0.417$ ($p < 0.001$); Pine-Oak, $r = -0.378$ ($p < 0.001$). Spearman correlation values were similar (± 30%) to Pearson correlation values in each test (Online Resource 5). There were significant linear relationships between JI and interpoint distance in each HT ($p <$ 0.007; bootstrap $p < 0.009$; Fig. 9). There we also significant effects of transect and the transect:interpoint distance interaction on the relationship between JI and interpoint distance for three HTs ($p < 0.001$; bootstrap $p < 0.001$). The one exception was Pine-Oak,

559 where neither transects ($p > 0.053$) nor the interaction effect were significant ($p > 0.199$).
560 The effect size of interpoint distance on ASV similarity varied by HT: Desert, adjusted R^2 =
561 0.750; Grassland, adjusted R^2 = 0.332; Encinal, adjusted R^2 = 0.401; Pine-Oak, adjusted R^2 =
562 0.183. In some cases, Desert and Grassland, there were significant differences in the
563 modeled relationship between transects within the same HT ($p < 0.005$; bootstrap $p <$
564 0.001; (Online Resource 6). These between-transect differences were not observed for the
565 Encinal nor Pine-Oak habitats ($p > 0.092$; Online Resource 6).

566

567 **Interpoint distance and key ASVs**

568 In the sampling distance data set, key ASVs were encountered at similar frequencies within
569 the transects of each HT ($\bar{x} \sim 2.5$ sample points), except for Encinal, for which very few key
570 ASVs identified (Table 5; Online Resource 7). In general, there was a significant negative
571 correlation between JI values and interpoint distance for key ASVs (Pearson r = -0.177, $p <$
572 0.001; Spearman r = -0.159, p = 0.001). Within individual HTs, significant negative
573 correlations were found within Desert (Pearson r = -0.299; p = 0.002) and Pine-Oak
574 (Pearson r = -0.301; p = 0.002). There were no significant correlations within the Grassland
575 and Encinal HTs ($p > 0.273$). Spearman correlations were within 13% of Pearson correlations
576 for the two HTs with significant correlations (Online Resource 5). The linear regression
577 model of key ASV JI values as a function of interpoint distances, with HT as a co-factor, was
578 not significant (p = 0.560). Within HTs, significant linear regression models for key ASV JI
579 values against interpoint distances, with transect as a cofactor, were observed for the

following: Desert, adjusted R^2 = 0.677 (p < 0.001; bootstrap p < 0.001), Grassland, adjusted R^2 = 0.122 (p < 0.003; bootstrap p = 0.001), and Pine-Oak, adjusted R^2 = 0.326 (p < 0.001; bootstrap p < 0.0001). The same linear regression model for Encinal was not statistically significant (p = 0.479). For the three HTs with significant models, interpoint distance was a significant factor in all cases (p < 0.005; bootstrap p < 0.005; Fig. 10). The regression coefficients for transects and interaction effects (between transect and interpoint distance) were also significant in many cases (p < 0.001; bootstrap p < 0.026), with significant differences in the linear slope among transects (adjusted p < 0.048; bootstrap p < 0.021; Online Resource 6). The number of key ASVs encountered showed a general increase with increasing numbers of sample points across all HTs (Fig. 11; Online Resource 8).

Discussion

General Patterns

This study examined patterns of eDNA from a broad swath of eukaryotic taxa in soil samples in arid and semi-arid habitats. The objective was to inform future applications for identifying eDNA indicators from unknown geographic and/or habitat soil sources. Our

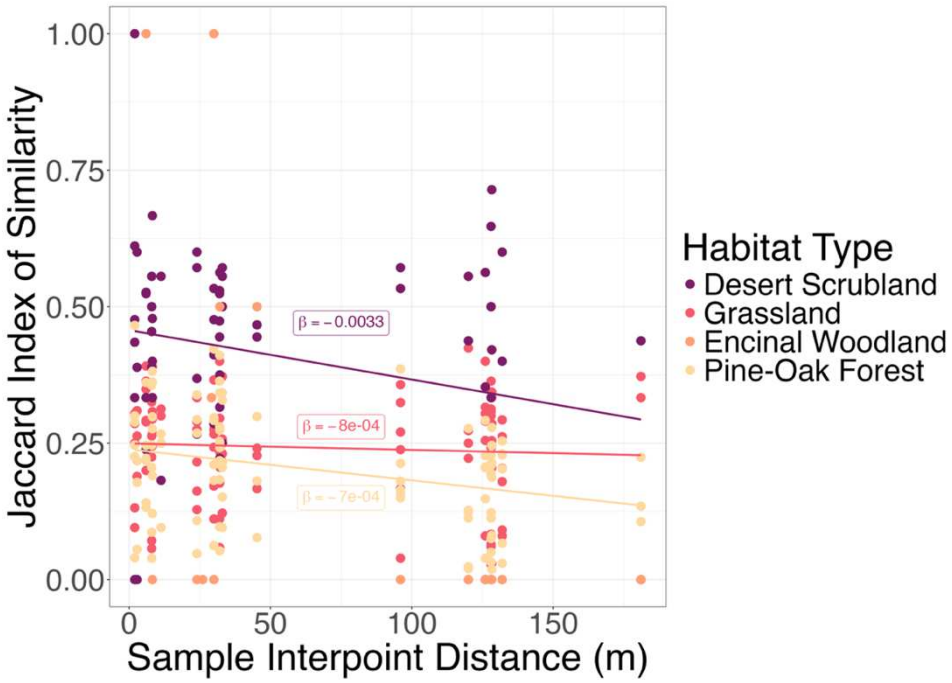


Fig. 10. Plot of Jaccard Index (JI) values against interpoint distances for *key* ASVs encountered in four different habitat types (HT). Slope lines derived from statistically significant linear model coefficients are provided.

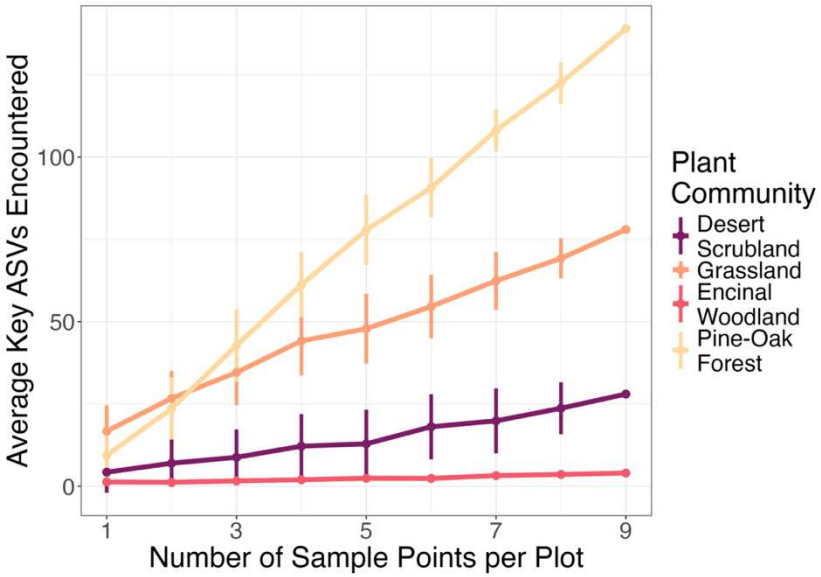


Fig. 11. Mean number (± 1 SD) of key ASVs encountered with increasing number of randomly-selected points within each transect by habitat type. All transects contained nine sample points.

findings add to the body of scientific information supporting the large diversity and nonuniform spatial distribution of eDNA in soil, both from environmental DNA sensu stricto and DNA from intact bodies (e.g., miniscule arthropods, invertebrate eggs, pollen, spores) that could functionally be considered eDNA. These eDNAs are not uniformly distributed in space, either vertically or horizontally. These factors alone make the kind of application described in this paper challenging (Ryan et al., 2022; Taberlet et al., 2012). As one way to reduce the size and complexity of the dataset and, perhaps, the effort required for effective pattern detection, we secondarily conducted analyses on a subset of the data comprised only of key ASVs.

Our first set of samples were taken during the spring pre-monsoon period that was both dry and preceded by very little to no rainfall over the course of several months. Our second set of samples were taken at the end of the summer monsoon season, preceded by several weeks of occasional rain events. Bioactivity in the region might be expected to be relatively low during the first sampling effort, and at its peak during the latter effort. We expected that there would be substantial differences in overall ASV richness and diversity between the two sampling events as a consequence of those differences in biological biomass and activity. While we did not observe a substantial increase in raw sequence reads or quality paired reads between sampling events (Table 1), we did observe a general increase in the number of ASVs in August samples compared to the April samples (Table 2). These differences were particularly pronounced for the Desert and Grassland HTs, and less pronounced for the Pine-Oak HT. Another potential factor that would increase diversity in the August samples compared to the April samples would be that sample distance had a

greater positive effect on inter-sample-diversity than depth. The August samples were all separated by some degree of spatial distance, while 25% of the April sample pairs came from different depths within the same sampling point.

The richness and nonuniform distribution of soil eDNAs were evident in the generally low JI values for all HTs (Table 2). Additionally, the ASV accumulation curves for all HTs in the August 2022 sample set exhibited no indication of impending saturation (Fig. 9). Likewise, in nearly all cases, the number of key AVS climbed steadily with increasing depth of soil samples or with increasing numbers of samples (Figs. 7, 9). The exception was in the April 2022 Encinal dataset, where ASV richness did not always increase with added depth classes. This HT produced a particularly low relative richness of key ASVs, and stochastic sampling error across small true differences in richness may account for this pattern. It is apparent then that sampling effort required to obtain comprehensive or near-comprehensive surveys of soil ASVs, even in these arid regions, would require either a much more intensive and extensive sampling effort than we employed (Lin et al., 2021). When conducting soil eDNA surveys to identify indicator ASVs, non-exhaustive coverage of the ASVs within the HT or region of interest would not be of much concern as the survey would not target rare ASVs, but only those ASVs that were common enough to be reliably encountered.

Total ASV diversity across all HTs, as well as ASV diversity within each HT, was numerically dominated by fungal ASVs (Table 1, Fig. 6). Kingdom Fungi was represented by up to three times the number of ASVs as encountered for animals (vertebrates and invertebrates) or plants. This outcome was not unexpected due to the large number of

651 fungal cells typically found in the soil (i.e., 105-6 per g soil; Dindal, 1991; Frøslev et al.,
652 2019; Nielsen et al., 2007). Many of these fungal ASVs were obtained from DNA
653 sequencing libraries that had been constructed using PCR amplicons generated from
654 “universal” primers that were designed to selectively amplify plants, invertebrate animals,
655 and vertebrate animals. This is not an uncommon outcome for metabarcoding surveys of
656 soil (Lin et al., 2021; Roger et al., 2022). The next most abundantly identified taxonomic
657 classifications after fungi were plants and invertebrates. Again, these groups include taxa or
658 life stages that exist as intact organisms in the soil and that are small enough to escape
659 observation (e.g., springtails (Collembola)), or ubiquitous enough (e.g., roots) that they
660 remain in samples even after larger and obvious organisms (e.g., insect grubs) or leaf litter
661 have been removed (Ariza et al., 2023; Davidson et al., 1999; Villacorta-Rath et al., 2022).
662 Notably, there were only two vertebrate animal ASVs detected within our samples (and
663 that were not eliminated as likely derived from contaminant DNA): mourning dove (*Zenaida*
664 *macroura*) and white-tailed deer (*Odocoileus virginianus*). Both species are common in the
665 study area. The relative scarcity of vertebrate eDNA can likely be ascribed to there being
666 only a relatively minute fraction of vertebrate eDNA at one point in soil compared to eDNA
667 from high-biomass soil dwelling organisms, including bacteria, fungi, and invertebrates.
668 Additionally, some portion of this non-vertebrate eDNA will include concentrated, intact
669 DNA found in living cells (up until the sample is chemically and/or thermodynamically
670 preserved and processed). Focusing soil sampling on points where terrestrial vertebrates
671 are more frequently active or present (e.g., game trails, water sources, nests, etc.) would
672 likely result in more vertebrate eDNA detections (Andersen et al., 2012; Neice & McRae,

673 2021; Ryan et al., 2022). We did not select sites that would likely have high vertebrate
674 traffic or other vertebrate eDNA “hotspots” in order to minimize bias in our overall
675 observations on patterns of eukaryotic eDNA in soil.

676 Little more than one-third of the ASVs encountered in our study could be assigned
677 with confidence to a taxonomic group (Table 1), and even fewer to family, genus, or
678 species. The lack of taxonomic precision is unsurprising because, in large part, much of the
679 recovered eDNA appeared to originate from fungi, a very diverse taxon for which majority
680 of groups fall into the category of “Dark Taxa” (taxonomic groups that have not been
681 thoroughly catalogued or classified, and that are underrepresented in DNA sequence
682 databases like GenBank; Bhunjun et al., 2022; Bienert et al., 2012; Lücking et al., 2021;
683 Page, 2011). Another 20 –25% of ASVs encountered in the soil samples were assigned to
684 protists and invertebrates, two groups that also contain many dark taxa (Beaudin et al.,
685 2025; Grattepanche et al., 2018; Jafari et al., 2023). It is also possible that for some of those
686 taxa that are detected using primers that were not designed for use with those taxa that
687 the barcode loci are not diagnostic at lower taxonomic levels. Again, we note that for
688 efforts focused on identifying indicator ASVs that taxonomic classification of ASVs is not an
689 absolute requirement.

690 The number of key ASVs encountered varied widely among HTs (Tables 4, 5) and did
691 not proportionally reflect differences among HTs in the total numbers of ASVs encountered
692 (Table 2). For example, Pine-Oak HT samples produced many more key ASVs than samples
693 from any other HT for both sampling events, despite exhibiting similar total ASV richness as
694 Grassland and Encinal samples in the April 2022 dataset and fewer total ASVs than the

Grassland and Encinal HTs in the August 2022 dataset. Only small numbers of key ASVs were encountered in the April Desert and Encinal transects, which reflected the low total number of ASVs encountered in Desert samples, but did not reflect the relatively high total number of ASVs in the Encinal samples (Table 3). In the August samples, there is practically no correspondence between total ASV and key ASV numbers (Table 4). This lack of correspondence in both datasets seems to indicate that there are factors influencing the frequency of key ASVs, and potentially indicator ASVs, other than simple ASV (i.e., biological) richness. These factors might include patchiness of habitat distribution, inter-site variance in key habitat features (e.g., slope, elevation, dominant vegetation), intensity and frequency of eDNA mixing events (e.g., windstorms, precipitation runoff), levels of eDNA decay factors (e.g., pH, mineral ion content, solar exposure), etc.

Soil eDNA Patterns and Sampling Depths

We anticipated that there might be differences in ASV constituencies related to soil depth, as some soil organisms have preferred soil conditions that might vary by depth (e.g., temperature, moisture), and because some soils have different water (and surface eDNA) penetration horizons. The different depth classes within each HT contained varying percentages of ASVs (19 – 37%) that were not found in the other depth classes (Fig. 5), and the different depth classes did not differ in contributions to overall ASV richness in most cases (Fig. 7). These patterns indicate that sampling to greater depths would likely then result in an increased richness in ASVs encountered at a sampling point. This pattern was

716 also observed for key ASVs HT (Online Resource 3). The degree to which the added ASV
717 richness is a function of simply sampling a greater volume of soil or a function of sampling
718 soil-stratified communities is difficult to conclude. Only a small percentage of key ASVs we
719 encountered occurred within only one depth class, a percentage which shrunk across HTs
720 with increasing numbers of key ASVs (Table 4). This pattern would argue for a random
721 subsampling effect. Another observation indicating that differences among depth classes
722 may have been in part due to subsampling effect was that the distributions of ASVs from
723 different taxonomic groups were strikingly similar across depth classes in all HTs (Fig. 6).
724 However, PERMANOVA results indicated that depth class was a significant factor in
725 dissimilarities among sample ASV sets. This finding indicates that to some degree
726 stratification of soil communities influenced ASV richness. Prior studies have observed
727 apparent depth stratification of soil biota (Andersen et al., 2012; Davidson et al., 2002;
728 O'Brien et al., 2005), but these studies have tended to sample to greater soil depths than in
729 our study (e.g., 80 cm; Andersen et al., 2012). An important practical consideration is that,
730 with the exception of the sandy soils of the Desert HT, soils occurring in the other HTs were
731 very rocky, such that digging to depths below 15 cm would require considerable effort.

732

733 **Interpoint Sampling Distance**

734 In general, ASV sets became increasingly dissimilar as the space between sample
735 points increased, though correlations and linear model effect sizes were weak to moderate
736 (R^2 = 0.2 – 0.4) – with the exception of the linear model for the Desert HT (R^2 = 0.75). Thus,

much of the variance in ASV set similarity between sample points was either stochastic or driven by unmeasured factors. Given that our sampling effort did not approach ASV saturation, much of the unexplained variance was likely stochastic due to underrepresentation of the ASV constituencies within the transects. There was often significant variance in slope coefficients and interactions between slope and transect in these models. The same general patterns were observed for key ASVs in the August sample set, except for the Encinal HT, which produced very few key ASVs. In the case of indicator ASVs (for which key ASVs are proxies) we would therefore expect that the number of indicator ASVs encountered would decrease with the diminishing of overlap in ASVs between samples that comes with increasing distance between points. However, we would also expect that indicator ASVs that co-occur at relatively distant sample points would be more likely to be highly useful and robust indicators, having a broad distribution within the HT or region of interest.

Conclusions

There were large differences in the numbers of sequencing reads and ASVs recovered among taxonomic groups (Table 2). While these differences may reflect real differences in the relative frequencies of occurrence of these ASV in the soil at the sampled locale, different taxa may be of relatively greater interest. For example, in our study, plant ASV richness was much lower than fungal richness, but plants generally would be much more effective indicators because their ranges and habitat associations are much better

758 known. Similarly, both sequence reads and ASVs for vertebrate animals were very sparse in
759 our dataset, but these species are likely the best characterized group in terms of range and
760 habitat, with the possible exception of plants. One strategy to expand the representation of
761 a desired taxonomic group in a metabarcoding survey would be to employ more than one
762 assay for that target group (Miya et al., 2015; Schenekar et al., 2020; Zhang et al., 2020).
763 Single universal primer pairs were utilized for each taxonomic group of interest in our study
764 and all assays in this study were chosen based on PCR band intensity in initial sequencing
765 library enrichment optimization trials. While band intensity reflects comparatively strong
766 levels of PCR amplification, intensity does not necessarily reflect amplicon sequence
767 richness or level of specificity to the target group. Pairing a general assay with more
768 narrowly targeted and designed assays could increase the representation of more difficult
769 to detect taxa as opposed to an approach relying solely on a single universal primer set.
770 Site selection is another strategic option for prioritizing different taxonomic groups. As
771 noted above, strategic decisions regarding season and point locations for sampling can be
772 used to achieve different soil eDNA objectives, such as maximizing ASV richness sampled or
773 targeting priority taxa.

774 In general, whether attempting to characterize biological diversity using soil eDNA,
775 surveying for the eDNA of particular terrestrial taxa, or seeking indicator ASVs, obtaining as
776 many samples as feasible should be an objective (Lin et al., 2021; Hermans et al. 2022). The
777 effective number of samples required likely will vary by habitat, and clearly by season.
778 Another potential strategy, given limitations to sampling effort, is to focus surveys on a
779 finer scale, spatially limited habitat type or geographic locale. In addition to maximizing

sample numbers, maximizing distance between sample points would benefit efforts to characterize diversity by minimizing redundancy in the ASVs encountered. In surveys for target taxa, especially in early exploratory stages, we can extrapolate from the patterns we observed that if a target's eDNA is not detected at one point, it is more probable to be encountered at distant points than at closer points. As noted earlier, ASVs that co-occur at distant points may be more promising as potential indicators, and surveys focused on identifying indicators likely would also want to maximize distances between samples within the region or habitat of interest. Finally, additional sampling depth increased the richness of ASVs and key ASVs encountered. As has been noted elsewhere (Taberlet et al., 2012), maximizing the volume of well-mixed soil that is subsampled for DNA extraction and analysis should increase the diversity of ASVs and potential indicator ASVs encountered, as well as increase the likelihood of encountering the eDNA of targeted taxa.

Acknowledgements

The authors wish to thank D. Brewer with the US Army Garrison Fort Huachuca for assistance with site consultation and access, permitting, and numerous other points of logistical support. The authors also wish to thank P. Landin, B. Clemens, M. Narup and the Range Control office with Fort Huachuca for assistance with planning and executing work on the installation. We thank J. Fuller (US Army Engineer Research and Development Center) for assistance with field sampling. This work funded by the U.S. Army Installations and Operational Environments research program administered by the U.S. Army Engineer Research and Development Center. The funding body had no discretionary involvement in design, execution, or communication of the research or its results. Publication permission was granted by the Chief of Engineers. The views expressed are those of the authors and do not necessarily represent those of the U.S. Army Corps of Engineers. The use of trade, product, or firm names in this article is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contribution

C.J., M.C., Y.MS., and R.L. conceptualized the study. C.J., M.C., and R.L. acquired funding, while C.J., M.C., A.A., Y.MS., G.F., and R.L. developed the methodology. D.L., M.C., Y.MS., and R.L. provided project administration and supervision. Investigations were conducted by M.M., D.L., C.J., M.C., N.F., A.A., M.W., Y.MS., A.N., G.F., and R.L. Data curation was performed by M.M., D.L., N.F., A.A., M.W., G.F., and R.L., and M.M., A.N., and R.L. performed formal data analysis. M.M. and R.L. prepared the visualizations. M.M., M.C.,

N.F., M.W., Y.MS., G.F., and R.L. wrote the original draft of the manuscript. M.M. and R.L. were primarily responsible for manuscript revision. All authors reviewed the manuscript.

Declarations

Funding

This work funded by the U.S. Army Installations and Operational Environments research program administered by the U.S. Army Engineer Research and Development Center. The funding body had no discretionary involvement in design, execution, or communication of the research or its results.

Data availability

All DNA metabarcoding sequence data generated by this project, including sample location data, may be found in the National Center for Biotechnology Information Sequence Read Archive PRJNA1380462.

Ethics approval

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors.

Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

837

References

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjær, K. H., Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, *21*(8), 1966–1979. https://doi.org/10.1111/j.1365-294X.2011.05261.x

Andruszkiewicz, E. A., Koseff, J. R., Fringer, O. B., Ouellette, N. T., Lowe, A. B., Edwards, C. A., & Boehm, A. B. (2019). Modeling Environmental DNA Transport in the Coastal Ocean Using Lagrangian Particle Tracking. *Frontiers in Marine Science*, *6*. https://www.frontiersin.org/articles/10.3389/fmars.2019.00477

Andruszkiewicz, E. A., Sassoubre, L. M., & Boehm, A. B. (2017). Persistence of marine fish environmental DNA and the influence of sunlight. *PLOS ONE*, *12*(9), e0185043. https://doi.org/10.1371/journal.pone.0185043

Ariza, M., Fouks, B., Mauvisseau, Q., Halvorsen, R., Alsos, I. G., & de Boer, H. J. (2023). Plant biodiversity assessment through soil eDNA reflects temporal and local diversity. *Methods in Ecology and Evolution*, *14*(2), 415–430. https://doi.org/10.1111/2041-210X.13865

Beaudin, M., Höcherl, A., Boudreault, C., Cullingham, C. I., Lapen, D., & Fernandez-Triana, J. (2025). A research hotspot for Microgastrinae parasitoid wasps (Hymenoptera, Braconidae) in North America: DNA barcoding reveals the need for increased taxonomic efforts in dark taxa. *Frontiers in Sustainable Food Systems*, *9*. https://doi.org/10.3389/fsufs.2025.1457055

859 Bhunjun, C. S., Niskanen, T., Suwannarach, N., Wannathes, N., Chen, Y.-J., McKenzie, E. H.
860 C., Maharachchikumbura, S. S. N., Buyck, B., Zhao, C.-L., Fan, Y.-G., Zhang, J.-Y.,
861 Dissanayake, A. J., Marasinghe, D. S., Jayawardena, R. S., Kumla, J., Padamsee, M.,
862 Chen, Y.-Y., Liimatainen, K., Ammirati, J. F., … Lumyong, S. (2022). The numbers of
863 fungi: Are the most speciose genera truly diverse? *Fungal Diversity*, *114*(1), 387–
864 462. https://doi.org/10.1007/s13225-022-00501-4
865 Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J.-J., & Taberlet, P. (2012).
866 Tracking earthworm communities from soil DNA. *Molecular Ecology*, *21*(8), 2017–
867 2030. https://doi.org/10.1111/j.1365-294X.2011.05407.x
868 Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A.,
869 Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K.,
870 Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M.,
871 Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible
872 microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), Article 8.
873 https://doi.org/10.1038/s41587-019-0209-9
874 Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2018). Seasonal variation in
875 environmental DNA detection in sediment and water samples. *PLOS ONE*, *13*(1),
876 e0191737. https://doi.org/10.1371/journal.pone.0191737
877 Cai, P., Huang, Q.-Y., & Zhang, X.-W. (2006). Interactions of DNA with Clay Minerals and Soil
878 Colloidal Particles and Protection against Degradation by DNase. *Environmental*
879 *Science & Technology*, *40*(9), 2971–2976. https://doi.org/10.1021/es0522985

880 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P.
881 (2016). DADA2: High-resolution sample inference from Illumina amplicon data.
882 *Nature Methods*, *13*(7), Article 7. https://doi.org/10.1038/nmeth.3869
883 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.
884 L. (2009). BLAST+: Architecture and applications | BMC Bioinformatics. *BMC*
885 *Bioinformatics*, *10*, 421.
886 Carr, C. M., Hardy, S. M., Brown, T. M., Macdonald, T. A., & Hebert, P. D. N. (2011). A Tri-
887 Oceanic Perspective: DNA Barcoding Reveals Geographic Structure and Cryptic
888 Diversity in Canadian Polychaetes. *PLOS ONE*, *6*(7), e22232.
889 https://doi.org/10.1371/journal.pone.0022232
890 Carraro, L., Blackman, R., & Altermatt, F. (2022). *Modelling eDNA transport in river*
891 *networks reveals highly resolved spatio-temporal patterns of freshwater biodiversity*
892 *| bioRxiv*. https://doi.org/10.1101/2022.01.25.475970
893 Clare, E. L., Economou, C. K., Bennett, F. J., Dyer, C. E., Adams, K., McRobie, B., Drinkwater,
894 R., & Littlefair, J. E. (2022). Measuring biodiversity from DNA in the air. *Current*
895 *Biology*, *32*(3), 693-700.e5. https://doi.org/10.1016/j.cub.2021.11.064
896 Clare, E. L., Economou, C. K., Faulkes, C. G., Gilbert, J. D., Bennett, F., Drinkwater, R., &
897 Littlefair, J. E. (2021). eDNAir: Proof of concept that animal DNA can be collected
898 from air sampling. *PeerJ*, *9*, e11030. https://doi.org/10.7717/peerj.11030
899 Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes
900 for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology*
901 *Resources*, *14*(6), 1160–1170. https://doi.org/10.1111/1755-0998.12265

902 Davidson, D. A., Bruneau, P. M. C., Grieve, I. C., & Young, I. M. (2002). Impacts of fauna on
903 an upland grassland soil as determined by micromorphological analysis. *Applied Soil*
904 *Ecology*, *20*(2), 133–143. https://doi.org/10.1016/S0929-1393(02)00017-3
905 Davidson, D. A., Carter, S., Boag, B., Long, D., Tipping, R., & Tyler, A. (1999). Analysis of
906 pollen in soils: Processes of incorporation and redistribution of pollen in five soil
907 profile types. *Soil Biology and Biochemistry*, *31*(5), 643–653.
908 https://doi.org/10.1016/S0038-0717(98)00123-0
909 De Cáceres, M., & Legendre, P. (2009). Associations between species and groups of sites:
910 Indices and statistical inference. *Ecology*, *90*(12), 3566–3574.
911 https://doi.org/10.1890/08-1823.1
912 De Cáceres, M., Legendre, P., Wiser, S. K., & Brotons, L. (2012). Using species combinations
913 in indicator value analyses. *Methods in Ecology and Evolution*, *3*(6), 973–982.
914 https://doi.org/10.1111/j.2041-210X.2012.00246.x
915 Deagle, B. E., Tollit, D. J., Jarman, S. N., Hindell, M. A., Trites, A. W., & Gales, N. J. (2005).
916 Molecular scatology as a tool to study diet: Analysis of prey DNA in scats from
917 captive Steller sea lions. *Molecular Ecology*, *14*(6), 1831–1842.
918 https://doi.org/10.1111/j.1365-294X.2005.02531.x
919 Dindal, D. L. (1991). *Soil Biology Guide*. John Wiley & Sons.
920 Donhauser, J., Briones, M. J. I., Mikola, J., Jones, D. L., Eder, R., Filser, J., Frossard, A., Krogh,
921 P. H., Sousa, J. P., Cortet, J., Desie, E., Domene, X., Djuric, S., Hackenberger, D.,
922 Jimenez, J. J., Iamandei, M., Rissmann, C., Schmidt, O., Shanskiy, M., … Frey, B.
923 (2023). Extracting DNA from soil or directly from isolated nematodes indicate

924 dissimilar community structure for Europe-wide forest soils. *Soil Biology and*
925 *Biochemistry*, *185*, 109154. https://doi.org/10.1016/j.soilbio.2023.109154
926 Dufrêne, M., & Legendre, P. (1997). Species Assemblages and Indicator Species: the Need
927 for a Flexible Asymmetrical Approach. *Ecological Monographs*, *67*(3), 345–366.
928 https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2
929 Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright,
930 M., Zakharov, E. V., Hebert, P. D. N., & Steinke, D. (2019). Validation of COI
931 metabarcoding primers for terrestrial arthropods. *PeerJ*, *7*, e7745.
932 https://doi.org/10.7717/peerj.7745
933 Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An Aligned Rank Transform
934 Procedure for Multifactor Contrast Tests. *The 34th Annual ACM Symposium on User*
935 *Interface Software and Technology*, 754–768.
936 https://doi.org/10.1145/3472749.3474784
937 Feist, S. M., Guan, X., Malmfeldt, M. P., & Lance, R. F. (2022). Two novel qPCR assays to
938 enhance black rail (Laterallus jamaicensis) eDNA surveys in the United States.
939 *Conservation Genetics Resources*, *14*(3), 321–329. https://doi.org/10.1007/s12686-
940 022-01279-y
941 Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using
942 environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425.
943 https://doi.org/10.1098/rsbl.2008.0118
944 Fløjgaard, C., Frøslev, T. G., Brunbjerg, A. K., Bruun, H. H., Moeslund, J., Hansen, A. J., &
945 Ejrnæs, R. (2019). Predicting provenance of forensic soil samples: Linking soil to

946 ecological habitats by metabarcoding and supervised classification. *PLOS ONE*,
947 *14*(7), e0202844. https://doi.org/10.1371/journal.pone.0202844
948 Fremier, A. K., Strickler, K. M., Parzych, J., Powers, S., & Goldberg, C. S. (2019). Stream
949 Transport and Retention of Environmental DNA Pulse Releases in Relation to
950 Hydrogeomorphic Scaling Factors. *Environmental Science & Technology*, *53*(12),
951 6640–6649. https://doi.org/10.1021/acs.est.8b06829
952 Frøslev, T. G., Ejrnæs, R., Hansen, A. J., Bruun, H. H., Nielsen, I. B., Ekelund, F., Vestergård,
953 M., & Kjøller, R. (2023). Treated like dirt: Robust forensic and ecological inferences
954 from soil eDNA after challenging sample storage. *Environmental DNA*, *5*(1), 158–
955 174. https://doi.org/10.1002/edn3.367
956 Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Hansen, A. J., Læssøe, T., & Heilmann-
957 Clausen, J. (2019). Man against machine: Do fungal fruitbodies and eDNA give
958 similar biodiversity assessments across broad environmental gradients? *Biological*
959 *Conservation*, *233*, 201–212. https://doi.org/10.1016/j.biocon.2019.02.038
960 Gao, C.-H., Chen, C., Akyol, T., Dusa, A., Yu, G., Cao, B., & Cai, P. (2024). ggVennDiagram:
961 Intuitive Venn diagram software extended. *Imeta*, *3*(1), e177.
962 https://doi.org/10.1002/imt2.177
963 Gao, C.-H., Yu, G., & Cai, P. (2021). ggVennDiagram: An Intuitive, Easy-to-Use, and Highly
964 Customizable R Package to Generate Venn Diagram. *Frontiers in Genetics*, *12*.
965 https://doi.org/10.3389/fgene.2021.706907

966 Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome project: Successes
967 and aspirations. *BMC Biology*, *12*(1), 69. https://doi.org/10.1186/s12915-014-0069-
968 1
969 Goldberg, C. S., Pilliod, D. S., Arkle, R. S., & Waits, L. P. (2011). Molecular Detection of
970 Vertebrates in Stream Water: A Demonstration Using Rocky Mountain Tailed Frogs
971 and Idaho Giant Salamanders. *PLOS ONE*, *6*(7), e22746.
972 https://doi.org/10.1371/journal.pone.0022746
973 Gordon, M., & Van Norman, K. (2021). Mycelial DNA persistence in a forest soil.
974 *Environmental DNA*, *3*(6), 1208–1213. https://doi.org/10.1002/edn3.242
975 Gostel, M. R., & Kress, W. J. (2022). The Expanding Role of DNA Barcodes: Indispensable
976 Tools for Ecology, Evolution, and Conservation. *Diversity*, *14*(3), Article 3.
977 https://doi.org/10.3390/d14030213
978 Grattepanche, J.-D., Walker, L. M., Ott, B. M., Paim Pinto, D. L., Delwiche, C. F., Lane, C. E., &
979 Katz, L. A. (2018). Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the
980 Darkness Between Morphology and Molecular Data. *BioEssays*, *40*(4), 1700198.
981 https://doi.org/10.1002/bies.201700198
982 Guthrie, A. M., Cooper, C. E., Bateman, P. W., van der Heyde, M., Allentoft, M. E., & Nevill, P.
983 (2024). A quantitative analysis of vertebrate environmental DNA degradation in soil
984 in response to time, UV light, and temperature. *Environmental DNA*, *6*(4), e581.
985 https://doi.org/10.1002/edn3.581
986 Haile, J., Holdaway, R., Oliver, K., Bunce, M., Gilbert, M. T. P., Nielsen, R., Munch, K., Ho, S. Y.
987 W., Shapiro, B., & Willerslev, E. (2007). Ancient DNA Chronology within Sediment

Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? *Molecular Biology and Evolution*, *24*(4), 982–989. https://doi.org/10.1093/molbev/msm016

Halko, N., Martinsson, P.-G., Shkolnisky, Y., & Tygert, M. (2011). An Algorithm for the Principal Component Analysis of Large Data Sets. *SIAM Journal on Scientific Computing*, *33*(5), 2580–2594. https://doi.org/10.1137/100804139

Hebsgaard, M. B., Gilbert, M. T. P., Arneborg, J., Heyn, P., Allentoft, M. E., Bunce, M., Munch, K., Schweger, C., & Willerslev, E. (2009). 'The Farm Beneath the Sand' – an archaeological case study on ancient 'dirt' DNA. *Antiquity*, *83*(320), 430–444. https://doi.org/10.1017/S0003598X00098537

Hermans, S. M., Lear, G., Buckley, T. R., & Buckley, H. L. (2022). Environmental DNA sampling detects between-habitat variation in soil arthropod communities, but is a poor indicator of fine-scale spatial and seasonal variation. *Ecological Indicators*, *140*, 109040. https://doi.org/10.1016/j.ecolind.2022.109040

Illumina. (2013). *16S Metagenomic Sequencing Library Preparation*. Illumina. https://emea.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Jafari, S., Müller, B., Rulik, B., Rduch, V., & S. Peters, R. (2023). Another crack in the Dark Taxa wall: A custom DNA barcoding protocol for the species-rich and common Eurytomidae (Hymenoptera, Chalcidoidea). *Biodiversity Data Journal*, *11*, e101998. https://doi.org/10.3897/BDJ.11.e101998

Jane, S. F., Wilcox, T. M., McKelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., Letcher, B. H., & Whiteley, A. R. (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources*, *15*(1), 216–227. https://doi.org/10.1111/1755-0998.12285

Jerde, C. L., Chadderton, W. L., Mahon, A. R., Renshaw, M. A., Corush, J., Budny, M. L., Mysorekar, S., & Lodge, D. M. (2013). Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(4), 522–526. https://doi.org/10.1139/cjfas-2012-0478

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*(2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jerde, C. L., Olds, B. P., Shogren, A. J., Andruszkiewicz, E. A., Mahon, A. R., Bolster, D., & Tank, J. L. (2016). Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environmental Science & Technology*, *50*(16), 8770–8779. https://doi.org/10.1021/acs.est.6b01761

Johnson, M. D., Cox, R. D., & Barnes, M. A. (2019a). Analyzing airborne environmental DNA: A comparison of extraction methods, primer type, and trap type on the ability to detect airborne eDNA from terrestrial plant communities. *Environmental DNA*, *1*(2), 176–185. https://doi.org/10.1002/edn3.19

Johnson, M. D., Cox, R. D., & Barnes, M. A. (2019b). The detection of a non-anemophilous plant species using airborne eDNA. *PLOS ONE*, *14*(11), e0225262. https://doi.org/10.1371/journal.pone.0225262

Kay, M., Elkin, L. A., Higgins, J. J., & Wobbrock, J. O. (2025). *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs* (R package version 0.11.2) [R]. https://github.com/mjskay/ARTool

Kirse, A., Bourlat, S. J., Langen, K., & Fonseca, V. G. (2021). Unearthing the Potential of Soil eDNA Metabarcoding—Towards Best Practice Advice for Invertebrate Biodiversity Assessment. *Frontiers in Ecology and Evolution*, *9*. https://www.frontiersin.org/articles/10.3389/fevo.2021.630560

Legendre, P., & Legendre, L. (2012). *Numerical Ecology*. Elsevier.

Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Piaskowski, J., Riebl, H., & Singmann, H. (2025). *Package 'emmeans'* (Version 1.11.1) [Computer software]. CRAN repository. https://rvlenth.github.io/emmeans/

Leray, M., Knowlton, N., & Machida, R. J. (2022). MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. *Environmental DNA*, *4*(4), 894–907. https://doi.org/10.1002/edn3.303

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34. https://doi.org/10.1186/1742-9994-10-34

Lilja, M. A., Buivydaitė, Ž., Zervas, A., Krogh, P. H., Hansen, B. W., Winding, A., & Sapkota, R. (2023). Comparing earthworm biodiversity estimated by DNA metabarcoding and morphology-based approaches. *Applied Soil Ecology*, *185*, 104798. https://doi.org/10.1016/j.apsoil.2022.104798

Lin, M., Simons, A. L., Harrigan, R. J., Curd, E. E., Schneider, F. D., Ruiz-Ramos, D. V., Gold, Z., Osborne, M. G., Shirazi, S., Schweizer, T. M., Moore, T. N., Fox, E. A., Turba, R., Garcia-Vedrenne, A. E., Helman, S. K., Rutledge, K., Mejia, M. P., Marwayana, O., Munguia Ramos, M. N., … Meyer, R. S. (2021). Landscape analyses using eDNA metabarcoding and Earth observation predict community biodiversity in California. *Ecological Applications*, *31*(6), e02379. https://doi.org/10.1002/eap.2379

Littlefair, J. E., Allerton, J. J., Brown, A. S., Butterfield, D. M., Robins, C., Economou, C. K., Garrett, N. R., & Clare, E. L. (2023). Air-quality networks collect environmental DNA with the potential to measure biodiversity at continental scales. *Current Biology*, *33*(11), R426–R428. https://doi.org/10.1016/j.cub.2023.04.036

Lorenz, M. G., & Wackernagel, W. (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiological Reviews*, *58*(3), 563–602.

Lu, H.-P., Wagner, H. H., & Chen, X.-Y. (2007). A contribution diversity approach to evaluate species diversity. *Basic and Applied Ecology*, *8*(1), 1–12.

Lücking, R., Aime, M. C., Robbertse, B., Miller, A. N., Aoki, T., Ariyawansa, H. A., Cardinali, G., Crous, P. W., Druzhinina, I. S., Geiser, D. M., Hawksworth, D. L., Hyde, K. D., Irinyi, L., Jeewon, R., Johnston, P. R., Kirk, P. M., Malosso, E., May, T. W., Meyer, W., …

Schoch, C. L. (2021). Fungal taxonomy and sequence-based nomenclature. *Nature Microbiology*, *6*(5), 540–548. https://doi.org/10.1038/s41564-021-00888-x

Lunghi, E., Valle, B., Guerrieri, A., Bonin, A., Cianferoni, F., Manenti, R., & Ficetola, G. F. (2022). Environmental DNA of insects and springtails from caves reveals complex processes of eDNA transfer in soils. *Science of The Total Environment*, *826*, 154022. https://doi.org/10.1016/j.scitotenv.2022.154022

Luo, L., Lin, H., & Schmidt, J. (2010). Quantitative Relationships between Soil Macropore Characteristics and Preferential Flow and Transport. *Soil Science Society of America Journal*, *74*(6), 1929–1937. https://doi.org/10.2136/sssaj2010.0062

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

Martinez Arbizu, P. (2017). *pairwiseAdonis: Pairwise Multilevel Comparison using Adonis* (Version R package version 0.4.1) [Computer software].

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7), 150088. https://doi.org/10.1098/rsos.150088

Moorhouse-Gann, R. J., Dunn, J. C., de Vere, N., Goder, M., Cole, N., Hipperson, H., & Symondson, W. O. C. (2018). New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-26648-2

1096 Neice, A. A., & McRae, S. B. (2021). An eDNA diagnostic test to detect a rare, secretive
1097 marsh bird. *Global Ecology and Conservation*, *27*, e01529.
1098 https://doi.org/10.1016/j.gecco.2021.e01529
1099 Nevers, M. B., Przybyla-Kelly, K., Shively, D., Morris, C. C., Dickey, J., & Byappanahalli, M. N.
1100 (2020). Influence of sediment and stream transport on detecting a source of
1101 environmental DNA. *PLOS ONE*, *15*(12), e0244086.
1102 https://doi.org/10.1371/journal.pone.0244086
1103 Nielsen, K. M., Johnsen, P. J., Bensasson, D., & Daffonchio, D. (2007). Release and
1104 persistence of extracellular DNA in the environment. *Environmental Biosafety*
1105 *Research*, *6*(1–2), 37–53. https://doi.org/10.1051/ebr:2007031
1106 Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D.,
1107 Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov,
1108 K. (2019). The UNITE database for molecular identification of fungi: Handling dark
1109 taxa and parallel taxonomic classifications. *Nucleic Acids Research*, *47*(D1), D259–
1110 D264. https://doi.org/10.1093/nar/gky1022
1111 O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J.-M., & Vilgalys, R. (2005). Fungal
1112 Community Analysis by Large-Scale Sequencing of Environmental Samples. *Applied*
1113 *and Environmental Microbiology*, *71*(9), 5544–5550.
1114 https://doi.org/10.1128/AEM.71.9.5544-5550.2005
1115 Ogram, A. V., Mathot, M. L., Harsh, J. B., Boyle, J., & Pettigrew, C. A. (1994). Effects of DNA
1116 Polymer Length on Its Adsorption to Soils. *Applied and Environmental Microbiology*,
1117 *60*(2), 393–396. https://doi.org/10.1128/aem.60.2.393-396.1994

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2001). *vegan: Community Ecology Package* (p. 2.6-8) [Dataset]. https://doi.org/10.32614/CRAN.package.vegan

Page, R. (2011). *Dark taxa: GenBank in a post-taxonomic world*. https://doi.org/10.59350/xhvv2-xjt24

Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J.-M., Gielly, L., Cordonnier, T., Zinger, L., Brun, J.-J., Choler, P., Taberlet, P., & Cécillon, L. (2015). Landscape-scale distribution patterns of earthworms inferred from soil DNA. *Soil Biology and Biochemistry*, *83*, 100–105. https://doi.org/10.1016/j.soilbio.2015.01.004

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils*, *45*(3), 219–235. https://doi.org/10.1007/s00374-008-0345-8

Prosser, C. M., & Hedgpeth, B. M. (2018). Effects of bioturbation on environmental DNA migration through soil media. *PLOS ONE*, *13*(4), e0196430. https://doi.org/10.1371/journal.pone.0196430

Robeson, M. S., Khanipov, K., Golovko, G., Wisely, S. M., White, M. D., Bodenchuck, M., Smyser, T. J., Fofanov, Y., Fierer, N., & Piaggio, A. J. (2018). Assessing the utility of metabarcoding for diet analyses of the omnivorous wild pig (Sus scrofa). *Ecology and Evolution*, *8*(1), 185–196. https://doi.org/10.1002/ece3.3638

Roger, F., Ghanavi, H. R., Danielsson, N., Wahlberg, N., Löndahl, J., Pettersson, L. B., Andersson, G. K. S., Boke Olén, N., & Clough, Y. (2022). Airborne environmental DNA metabarcoding for the monitoring of terrestrial insects—A proof of concept from the field. *Environmental DNA*, *4*(4), 790–807. https://doi.org/10.1002/edn3.290

Rosa, L. H., da Silva, T. H., Ogaki, M. B., Pinto, O. H. B., Stech, M., Convey, P., Carvalho-Silva, M., Rosa, C. A., & Câmara, P. E. A. S. (2020). DNA metabarcoding uncovers fungal diversity in soils of protected and non-protected areas on Deception Island, Antarctica. *Scientific Reports*, *10*(1), 21986. https://doi.org/10.1038/s41598-020-78934-7

Rota, N., Canedoli, C., Ferrè, 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), Article 7. https://doi.org/10.3390/f11070738

Ruffell, A., & McKinley, J. (2008). *Geoforensics*. John Wiley & Sons.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Ryan, E., Bateman, P., Fernandes, K., van der Heyde, M., & Nevill, P. (2022). eDNA metabarcoding of log hollow sediments and soils highlights the importance of substrate type, frequency of sampling and animal size, for vertebrate species detection. *Environmental DNA*, *4*(4), 940–953. https://doi.org/10.1002/edn3.306

Sakata, M. K., Yamamoto, S., Gotoh, R. O., Miya, M., Yamanaka, H., & Minamoto, T. (2020). Sedimentary eDNA provides different information on timescale and fish species composition compared with aqueous eDNA. *Environmental DNA*, *2*(4), 505–518. https://doi.org/10.1002/edn3.75

Schenekar, T., Schletterer, M., Lecaudey, L. A., & Weiss, S. J. (2020). Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. *River Research and Applications*, *36*(7), 1004–1013. https://doi.org/10.1002/rra.3610

Shogren, A. J., Tank, J. L., Andruszkiewicz, E., Olds, B., Mahon, A. R., Jerde, C. L., & Bolster, D. (2017). Controls on eDNA movement in streams: Transport, Retention, and Resuspension. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-05223-1

Sirois, S. H., & Buckley, D. H. (2019). Factors governing extracellular DNA degradation dynamics in soil. *Environmental Microbiology Reports*, *11*(2), 173–184. https://doi.org/10.1111/1758-2229.12725

Taberlet, P., Prud'homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., & Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology*, *21*(8), 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Tong, D., Grubesic, T. H., Mu, W., Miller, J. A., Helderop, E., Jha, S., Brosi, B. J., & Bienenstock, E. J. (2021). Identifying the spatial footprint of pollen distributions using the Geoforensic Interdiction (GOFIND) model. *Computers, Environment and Urban Systems*, *87*, 101615. https://doi.org/10.1016/j.compenvurbsys.2021.101615

Trevors, J. T. (1996). DNA in soil: Adsorption, genetic transformation, molecular evolution and genetic microchip. *Antonie van Leeuwenhoek*, *70*(1), 1–10. https://doi.org/10.1007/BF00393564

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

Vamos, E., Elbrecht, V., & Leese, F. (2017). Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, *1*, e14625. https://doi.org/10.3897/mbmg.1.14625

van der Heyde, M., Bunce, M., Wardell-Johnson, G., Fernandes, K., White, N. E., & Nevill, P. (2020). Testing multiple substrates for terrestrial biodiversity monitoring using environmental DNA metabarcoding. *Molecular Ecology Resources*, *20*(3), 732–745. https://doi.org/10.1111/1755-0998.13148

Villacorta-Rath, C., Lach, L., Andrade-Rodriguez, N., Burrows, D., Gleeson, D., & Trujillo-González, A. (2022). *Invasive terrestrial invertebrate detection in water and soil using a targeted eDNA approach* (p. 2022.11.29.518289). bioRxiv. https://doi.org/10.1101/2022.11.29.518289

Wacker, S., Fossøy, F., Larsen, B. M., Brandsegg, H., Sivertsgård, R., & Karlsson, S. (2019). Downstream transport and seasonal variation in freshwater pearl mussel (Margaritifera margaritifera) eDNA concentration. *Environmental DNA*, *1*(1), 64–73. https://doi.org/10.1002/edn3.10

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 143–146. https://doi.org/10.1145/1978942.1978963

Yoccoz, N. G., Bråthen, K. A., Gielly, L., Haile, J., Edwards, M. E., Goslar, T., Von Stedingk, H., Brysting, A. K., Coissac, E., Pompanon, F., Sønstebø, J. H., Miquel, C., Valentini, A., De BELLO, F., Chave, J., Thuiller, W., Wincker, P., Cruaud, C., Gavory, F., … Taberlet, P. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, *21*(15), 3647–3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x

Zhang, S., Zhao, J., & Yao, M. (2020). A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. *Methods in Ecology and Evolution*, *11*(12), 1609–1625. https://doi.org/10.1111/2041-210X.13485

Supplementary Files

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

- UncoveringdepthandspatialpatternssoileDNAaridregionsOR1.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR4.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR5.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR2.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR8.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR3.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR7.pdf
- UncoveringdepthandspatialpatternssoileDNAaridregionsOR6.pdf