Still, the environment selects: Disentangling spatial and environmental effects on soil prokaryotes

Niklas Neisse[1]*, Maria Kostakou[1], Sophia N. Meyer[2], Anja Linstädter[2], Antonis Chatzinotas[1,3,4], Stephanie D. Jurburg [1,3]*

[1] Department of Applied Microbial Ecology, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany

[2] Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

[3] Institute of Biology, Leipzig University, Leipzig, Germany

[4] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

*Corresponding authors: neisse.n@protonmail.com; s.d.jurburg@gmail.com

Abstract

Soil prokaryotes play a crucial role in terrestrial ecosystems, yet the relative importance of dispersal limitation, environmental selection, and biotic interactions in shaping their assembly remains unresolved. To better understand the underlying drivers of soil prokaryotic communities in temperate grasslands, we sampled grassland transects across Germany, and examined how geographic distance, soil properties, plant community composition, and vegetation characteristics influence prokaryotic communities. We found a clear distance decay relationship (DDR): community similarity decreased by approximately 5% with each doubling of geographic distance. To disentangle the contribution of geographic distance from spatial changes in soil properties and vegetation, we combined variation partitioning and commonality analysis. Physicochemical heterogeneity in soil, which accounted for over ~30% of the total variation in prokaryotic community composition, explained more than 52% of the observed DDR and had a larger unique contribution to the DDR than geographic distance. In contrast, geographic distance alone only explained 2% of the variation in community composition. Small-scale variations in soil properties within individual sites resulted in high turnover over short distances, suggesting strong abiotic filtering at the local level. Plant community composition explained ~36% of the variation in community composition, but only 6.4% of the DDR, while vegetation characteristics

had only marginal explanatory power. Finally, we found rare taxa that were either locally restricted or widespread, indicating that rarity does not always restrict dispersal. Dominant taxa, particularly from Proteobacteria and Firmicutes, were consistently widespread, aligning with generalist strategies. Interestingly, a smaller group of highly abundant taxa exhibited intermediate spatial ranges, suggesting that dominance does not arise solely from dispersal limitation. Overall, our findings indicate that environmental filtering and plant-soil interactions are the primary factors driving DDRs among soil prokaryotes, while geographic distance plays a lesser role. Our study emphasizes the importance of sampling methods that explicitly account for spatial patterns and integrate soil and vegetation data, enhancing our understanding of the intricate patterns of dispersal limitation in microbial communities.

Introduction

Soil prokaryotes are essential to terrestrial ecosystem functioning, nutrient cycling, organic matter decomposition, and plant growth (Van Der Heijden *et al.* 2008; Wagg *et al.* 2019), but knowledge of their spatial distribution is limited by their high diversity, wide distributions, and considerable spatiotemporal turnover (Fierer & Jackson 2006; Lemoine *et al.* 2023). Understanding the processes that structure microbial communities (i.e., their assembly) is crucial for predicting ecosystem responses to environmental change. Community assembly is shaped by four key mechanisms: selection, drift, speciation, and dispersal (Vellend 2010). Among these, selection and dispersal are central to shaping microbial diversity at local and regional scales (Nemergut *et al.* 2013). While selection and dispersal have received considerable attention in microbial ecology, most work to date has focused on large scales (i.e., km; see Griffiths *et al.* 2011; Hazard *et al.* 2013; Ranjard *et al.* 2013), which likely exceed the dispersal capacity of most soil microbes and the local factors that drive their selection (Barbour *et al.* 2022; Hanson *et al.* 2012; but see Richter-Heitmann *et al.* 2020).

The distance-decay relationship (DDR) describes the observation that the similarity of community composition decreases with geographic distance, and has been well-documented for plants and animals (Graco‐Roza *et al.* 2022; Nekola & McGill 2014; Nekola & White 1999), as well as for microbes (Clark *et al.* 2021). DDRs arise from the interaction between dispersal limitation, environmental gradients, and biotic interactions (Nekola & White 1999). Dispersal limitation contributes to DDRs when a lack of connectivity between habitat patches prevents organisms from dispersing across them (Clark *et al.* 2021). Furthermore, environmental heterogeneity and biotic interactions, both drivers of selection, contribute indirectly to DDRs as communities in close proximity are likely to encounter more similar environmental conditions and interact with similar organisms, respectively.

Existing research into DDRs of soil prokaryotes has found mixed evidence for this pattern (Fierer & Jackson 2006; Rousk *et al.* 2010), and generally weaker DDRs than for other communities. Differences in the strength of the relationship have been attributed to experimental designs (Barbour *et al.* 2022), but whether properties of soil communities contribute to the observed DDRs is unclear. First, the extent to which microbes are limited by dispersal remains unresolved. On the one hand, passive dispersal mechanisms, high population densities, and the ability to enter dormancy suggest a high dispersal potential (Clark *et al.* 2021). On the other hand, mounting evidence indicates that microbial dispersal can indeed be limited (Barbour *et al.* 2022; Hanson *et al.* 2012), but the spatial extent of this dispersal limitation is unclear (Barbour *et al.* 2022). Second, microbial community assembly in soils is influenced by plant-microbe associations, which play a significant role in structuring soil prokaryotic communities by exerting selective pressures on the local microbiota (Eisenhauer *et al.* 2010; Trivedi *et al.* 2020). However, since plant distributions themselves follow DDRs (Nekola & White 1999), these interactions introduce further complexity into microbial assembly patterns. Furthermore, a wide

range of work has demonstrated the role of small-scale environmental differences, such as pH (Fierer & Jackson 2006; Rousk *et al.* 2010; Zhou *et al.* 2024) and soil moisture (Delgado-Baquerizo *et al.* 2018; Frindte *et al.* 2019), in shaping soil bacterial communities, and steep environmental gradients in soil can occur over very short distances (Dumbrell *et al.* 2009; Vos *et al.* 2013), even over millimeters.

A more complete understanding of soil microbial community assembly requires disentangling the role of dispersal from the selective biotic and abiotic environment across spatial scales. Grasslands are ideal ecosystems to study soil microbial DDRs due to their high microbial diversity (Labouyrie *et al.* 2023), spatial continuity at small scales, and variable environmental conditions across different spatial scales. Although stronger DDRs are expected in systems with low connectivity or limited dispersal (i.e., soil; Hanson *et al.* 2012), grassland soil microbial communities exhibit weaker DDRs than those in aquatic environments. This suggests that additional factors like complex dispersal patterns (Walters *et al.* 2022) or uneven environmental gradients may influence microbial community turnover in these systems (Clark *et al.* 2021). We investigated the drivers of soil prokaryotic community assembly in grassland soils across two regions in Germany. We hypothesized that: 1) prokaryotic communities in grassland soils exhibit distance-decay relationships from the local (m) to the regional (km) scales; 2) the spatial distribution of selective factors including plant communities and environmental dissimilarity contribute to prokaryotic DDRs, accounting for a considerable portion of the observed variation in microbial community composition, and that 3) dispersal limitation plays a significant role in the assembly of soil prokaryotic communities.

Methods

Study area and sampling design

We sampled managed grasslands in 18 plots across the Hainich-Dün and Schorfheide-Chorin regions of Germany as part of the Biodiversity Exploratories (Fischer *et al.* 2010). These regions are separated by approximately 300 km. Schorfheide-Chorin is mainly shaped by sandy and loamy soils which originate from young glacial sediments, whereas Hainich-Dün is characterized by loamy and clayey soils formed from calcareous bedrocks (Fischer *et al.* 2010). All grassland plots are located along a land use intensity gradient and are managed by local farmers (Blüthgen *et al.* 2012).

Soil and vegetation data were collected in June 2023 in nine permanently installed research plots per region. Within each plot, three 1 m² subplots were positioned along a 50 m south-north transect. The first subplot was located 13.5 m from the second, and the second 18 m from the third (see Muro *et al.* 2022). Within each subplot, plant species inventories and cover estimations were conducted, and biomass and soil samples were taken. For ground cover types, we recorded the cover of bare ground, litter, senescent vegetation, green vegetation, and moss in each subplot, summing up to 100%. For plant functional trait measurements, all plant species with a cover exceeding 1% were sampled with one individual per subplot, following standard protocols (Pérez-Harguindeguy *et al.* 2016). Individuals were sampled in the vicinity of subplots to not interfere with biomass measurements. Leaf area and leaf dry mass were measured to calculate specific leaf area (SLA). Finally, aboveground biomass was clipped at 2 cm stubble height from a 0.6 x 0.6 m area per subplot, oven-dried (55 °C, > 48 h) and weighted.

From the center of each subplot, the top 10 cm of soil was collected using a 2.5 cm diameter auger, and the sample was then homogenized. An aliquot of 5 g was stored at -20°C for subsequent the analysis of soil prokaryotes. The remaining soil was sieved using a 1-2 mm mesh and subsequently dried (at least at 40°C for 24 h) to assess soil carbon and nitrogen content using a CN analyser (Eurovector EA3100). Soil moisture was measured as (fresh weight [g] - dry weight [g]) / dry weight [g]. For pH measurements, 10±0.2 g moist soil was mixed with 20 ml of deionized water, briefly shaken, and an equilibrium period of 30 min was allowed before measuring the pH using a pH meter.

DNA extraction and sequencing

DNA was extracted from 0.25 g of soil for each sample using the NucleoSpin Soil kit (Macherey Nagel), according to the manufacturer's instructions, and DNA quality and concentration were assessed with gel electrophoresis and Qubit assays, respectively. The 16S rRNA gene was amplified using primers targeting the V4 region (515F-806R), following the standard Earth

Microbiome Project protocols (Thompson *et al.* 2017) and primers 16S_Illu_515F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGYCAGCMGCCGCGGTAA-3') and 16S_Illu_806R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACNVGGGTWTCTAAT-3'). Briefly, The first PCR incorporated locus-specific primers with Illumina overhang adapter sequences, and a second PCR was performed to attach sample-specific indices using the Nextera XT Index Kit (Illumina). All PCRs were run with negative controls to detect potential contamination. PCR products were purified using AMPure XP magnetic beads (Beckman Coulter), quantified again with the Qubit system, and pooled. Final libraries were sequenced on an Illumina MiSeq platform using 2 × 250 bp paired-end reads (MiSeq Reagent Kit v3, 600 cycles).

Bioinformatics

All bioinformatics and statistics analyses were conducted in R (version 4.3.2) (R. Core Team 2025). R code for all analyses is available in GitHub (https://github.com/NeisseN/BEO_DDR). The *dada2* pipeline was used to infer amplicon sequence variants (ASVs) from the raw sequencing data according to standard protocols (Callahan *et al.* 2016). Reads were trimmed to exclude the first 10 nucleotides and truncated at 230 and 220 bp for the forward and reverse reads respectively, with a maximum of 3 expected errors per read, resulting in an average read length of 272 bp. Taxonomic assignment was performed with the SILVA 16S rRNA reference database v138.1 (Quast *et al.* 2012). Prior to statistical analyses, all samples were standardized to 23,942 observations per sample and thereby excluding one sample using the *rarefy_even_depth* function of the *phyloseq* package (McMurdie & Holmes 2013).

Statistical analyses

To test the strength of the soil prokaryotic DDR and assess the contributions of selective biotic and abiotic factors such as plant and environmental dissimilarity across space, we performed a commonality analysis based on a generalized linear model (GLM) with a Gamma error distribution and a log link function, using the *glmm.hp* function of the *glmm.hp* package (Lai *et al.* 2022). We considered four groups of explanatory variables, all measured at the sample level. First, we represented geographic distance as the pairwise Haversine distance on a log scale based on natural logarithms between sample coordinates, calculated using the *distm* function from the *geosphere* package (version 1.5-20) (Hijmans *et al.* 2024). Second, we quantified abiotic environmental variation as the Euclidean distance based on scaled soil carbon and nitrogen concentrations, moisture content, and pH, which previous studies consistently identified as key drivers of prokaryotic communities across ecosystems (Bell 2010; Delgado-Baquerizo *et al.* 2018; Fierer & Jackson 2006). Third, we assessed variation in plant community composition through Bray-Curtis dissimilarities with the *vegdist* function. Finally, we captured vegetation characteristics by calculating Euclidean distances of aboveground biomass, the

community-weighted mean of specific leaf area (SLA mean), and the percentage cover of four ground cover types (bare ground, litter, moss, and senescent vegetation).

To build the GLM, we first addressed potential collinearity among the explanatory distance matrices by calculating pairwise Pearson correlations. We then performed bidirectional stepwise selection using the full model, including all interaction terms, as the upper scope. We identified only a significant interaction between physico-chemical Euclidean distance and plant compositional Bray-Curtis similarity for model performance. Although vegetation characteristics did not improve model fit, we retained them in the final model to ensure consistency with the variables used in the variation partitioning analysis. Model assumptions were checked with the *DHARMa* R package (Hartig *et al.* 2024). We performed commonality analysis with a lognormal approximation in our GLM to disentangle the unique and shared effects of each ecological factor on prokaryotic community composition. We quantified the individual and overlapping contributions as proportions of explained deviance, analogous to marginal R^2, based on commonality coefficients. To assess statistical significance, we performed permutation tests with 10,000 iterations, where the response variable was randomly permuted while maintaining the structure of explanatory variables. The total model fit was assessed using a pseudo-R^2 calculated as the deviance explained by the final model relative to the null model. Mean values are presented throughout as mean ± standard deviation (SD).

To further disentangle the interactions between drivers of soil prokaryotic β-diversity, we performed variation partitioning using the *varpart* function from the *vegan* R package (version 2.6-10) (Oksanen *et al.* 2007) on the Bray-Curtis dissimilarities of prokaryotic communities. We included four potential drivers of prokaryotic β-diversity: 1) geographic location (spatial coordinates of the samples), 2) abiotic variables, 3) the relative abundances of a total of 106 plants, and 4) vegetation characteristics variables. To select only relevant variables for each group, we applied the *ordiR2step* function on a distance based RDA (*dbRDA*) to perform stepwise model selection with 999 permutations. The significance of each fraction of the variation partitioning analysis was assessed with the *test_vp4* function from the *comecol* package in R (https://github.com/jgmv/comecol). This function performs permutation-based significance testing for each explanatory component using the *anova.cca* function from the *vegan* package.

Spatial range

To explore how the spatial ranges of soil prokaryotes drive the observed DDRs, we investigated the relationship between the spatial distribution between samples and relative abundance of ASVs by calculating the convex hull area and mean relative abundance for each ASV across all samples where it was present. The convex hull area (i.e., the smallest convex shape that encloses a set of observations), representing the geographic extent of each ASV, was calculated using the *CHullArea* function from the *GeoRange* R packages (Boyle 2017), based on the longitude and latitude coordinates (WGS84; UTM Zone 33N, Berlin) of the samples in which the ASV was

present. ASVs with less than 3 locations were assigned a convex hull area of 0, and all values were converted to the natural log scale (+1 m^2). ASVs were classified according to their relationship of occurrence (within plot, within site, and between site), and by phylum. To examine whether growth rate was related to dispersal range, we retrieved predicted 16S rRNA gene copy numbers from the rrnDB (version 5.10; Stoddard *et al.* 2015) and assigned them to ASVs based on their taxonomy. Approximately 20.15% of ASVs could be matched to rrnDB on a genus level and were included in this analysis. We then assessed the relationship between convex hull area and predicted copy number using Kendall's rank correlation.

Results

Distance decay in prokaryotic communities

We decomposed the distance-decay relationships (DDR) of soil prokaryotes to assess the contributions of geographic distance, the similarity of soil physicochemical properties, the similarity of plant communities, and the accompanying gradient in vegetative characteristics. The fitted GLM captured 53% of the variation in prokaryotic community dissimilarities (Fadden's pseudo-$R^2 = 0.53$). As expected, geographic distance had a significant, negative effect on prokaryotic community similarity ($\beta = -0.074$, SE = 0.005, $t(1269) = -14.03$, $p < 0.001$; Fig. 1). Doubling the distance between sites was associated with a 5% decrease in prokaryotic community similarity, holding other variables constant (Supplemental Equation S1-S4).

The highest observed prokaryotic community similarity was 59.8% and occurred between two samples collected within the same plot, though not at the closest recorded distances. Nevertheless, some pairs of samples, despite originating from the same plot, exhibited very low similarity (minimum 11.7%). The average similarity among sample pairs from the same plot was 40.4% ± 13.4. At the minimum geographic distance of 12.96 m, a prokaryotic community similarity of 35.2% ± 3 was estimated (holding all other variables at their mean values). At average short range (20.13 m), mid range (5.7 km), and long range (309.4 km) distances, community similarities of 34.2% ± 2.8, 22.6% ± 0.7, and 16.9% ± 0.5 were observed, respectively. Notably, within-site comparisons included the 22 pairs with lowest similarity (minimum similarity of 3%), and exhibited high variation in physicochemical properties (Supplemental Fig. S1).

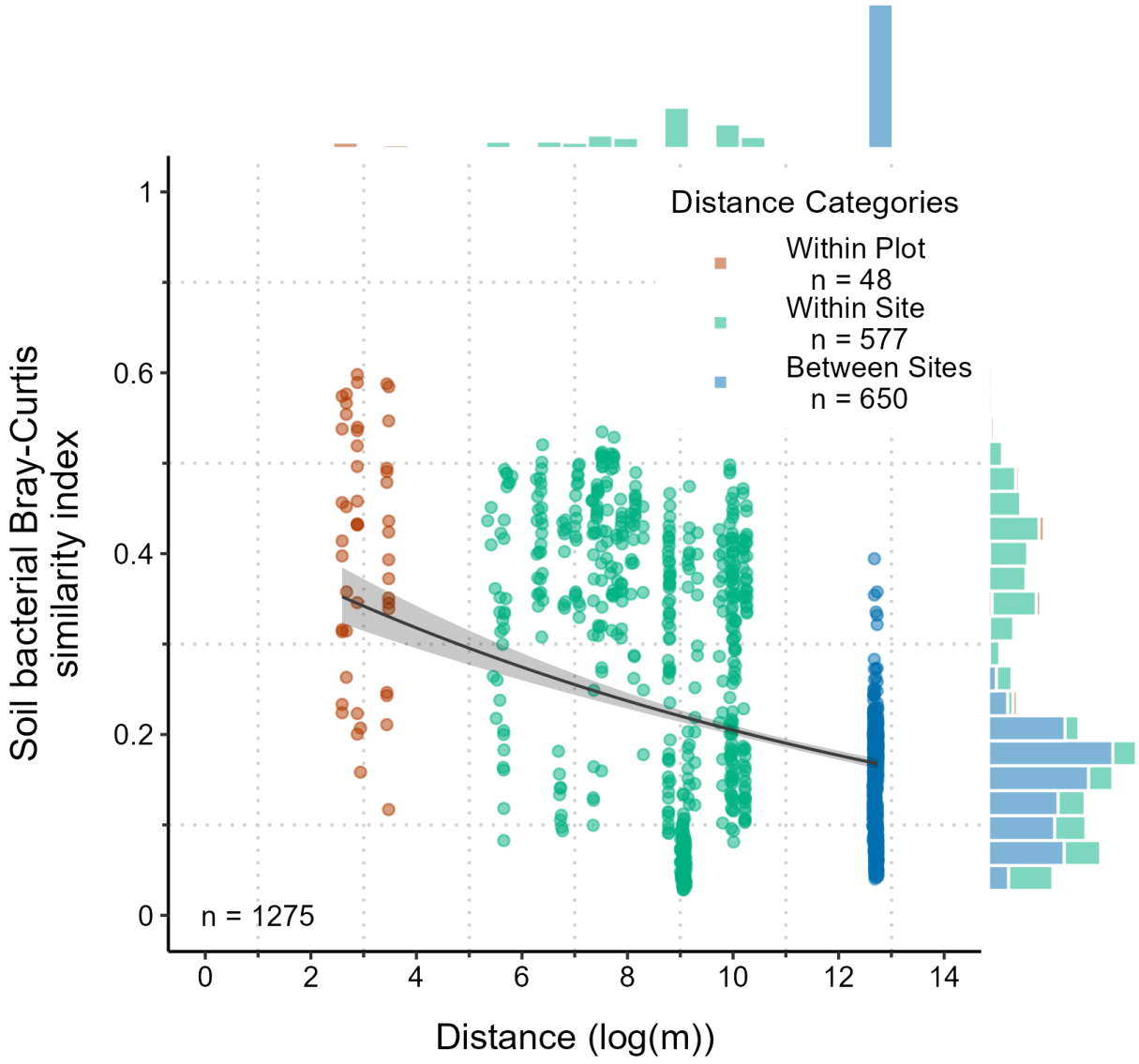

Fig. 1 Distance-decay relationship of soil prokaryotes within and across grasslands plots. Distance between samples (log-transformed Haversine distance in meters), modeled using a GLM with a log link and the difference in physicochemical properties, plant communities, vegetation characteristics set to their mean ($p < 0.001$). Points represent raw pairwise similarity data; the line and shaded ribbon show the GLM-predicted relationship with 95% confidence intervals. Stacked bar charts along the X and Y axes show the distribution of pairwise distances, colored to match the comparison scale.

To evaluate the relative importance of dissimilarity of soil physicochemical parameters, plant community composition, and their vegetation characteristics, we performed a commonality analysis on the GLM (total marginal adjusted R^2= 57.7%; Table 1). Geographic distance, and changes in physicochemical soil properties and plant community composition, as well as their interaction were all highly significant ($p < 0.001$), while vegetation characteristics showed no significant contribution. Changes in physicochemical parameters accounted for the largest share of the explained variance, contributing 52.1% to the total R^2. This included the highest unique contribution (15.4%), along with a shared contribution of 13.5%, primarily with geographic distance and in three-way combinations involving either plant similarity or the interaction between physicochemical properties and geographic distance. Geographic distance was the second most influential predictor, explaining 38.7% of the total R^2, with a unique contribution of 9.6%. Changes in plant beta diversity accounted for 6.4% of the total R^2, largely through shared variance with other predictors.

Table 1 Contribution shares of selective processes to prokaryotic community similarity in grasslands, based on commonality analysis using a lognormal approximation. The table presents the unique, average shared, and total (individual) contributions of each predictor to the model's adjusted R^2 (total R^2 = 0.577), along with their percentage contributions. The interaction term (E) reflects the conditional effect of physicochemical properties and geographic distance. Significance codes: *** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1, ns = not significant (based on 1000 permutations).

		a. Unique		b. Average share		Individual (a+b)		Individual (%)
A.	Changes in soil properties	0.154	***	0.135	***	0.289	***	52.1
B.	Geographic distance	0.096	***	0.119	***	0.215	***	38.7
C.	Beta diversity of plant community	0.007	*	0.028	***	0.035	***	6.4
D.	Changes in vegetation characteristics	-0.001	ns	0.001	***	0.000	ns	0.1
E.	Interaction of A. and B.	0.012	***	0.004	***	0.016	***	2.8

Drivers of prokaryotic beta diversity

To assess the direct role of a) location, b) physicochemical parameters, c) plants, and d) their vegetation characteristics to in modulating the soil prokaryotic community (i.e., rather than the effect of their DDRs), we first selected the relevant drivers for compartments b-d through stepwise selection. The variables selected for b-d were b) soil carbon concentration, pH, and the carbon-to-nitrogen (C:N) ratio; c) the abundance of *Poa trivialis* L.*, Achillea millefolium* L.*, Lolium perennes* L.*, Anthoxanthum odoratum* L.*, Bromus hordeaceus* L.*, Medicago sativa* L.*, Geranium rotundifolium* L.*, Daucus carota* L.*, Crataegus* L. *spec., Cynosurus cristatus* L.*, Phleum pratense* L.*, and Centaurea jacea* L.; and d) bare ground cover, total plant biomass, and ground cover of litter. Then, we conducted a variation partitioning analysis, which explained 42% of the total variation in prokaryotic β-diversity (Fig. 2). All predictor groups contributed significantly to the explained variation (Supplemental Table S1). Plant community composition accounted for the largest proportion of the variance (~36%), followed by soil physicochemical properties (~30%), and location (~19%). Vegetation characteristics were the weakest drivers, explaining only about 10% of the variation. Among the unique fractions, only plant community composition (7.4%) and location (1.8%) showed statistically significant effects. The unique contributions of soil physicochemical properties and vegetation characteristics were not significant (0.9%, and 0.8%, respectively). Plant community composition and soil properties together accounted for about 41% of the total variance.

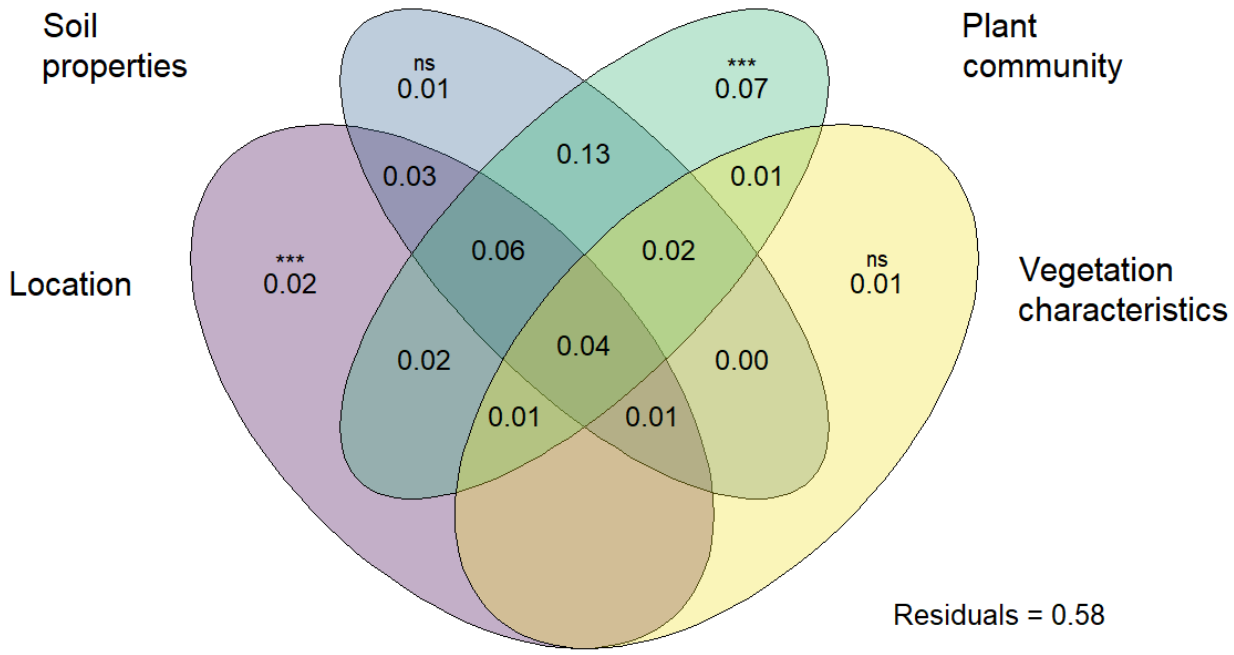

Fig. 2 Venn diagram summarizing the results of a variation partitioning analysing the effect of 1. Location, 2. Soil properties, 3. Plant community composition, 4. Vegetation characteristics on the Bray-Curtis similarity of prokaryotic communities. The values displayed represent the adjusted R^2 of both unique and overlapping contributions of each predictor group. Blank areas indicate no explained or shared variance. Significance codes: *** $p < 0.001$, ns = not significant, blank = not testable.

Spatial range

We investigated the spatial distribution of soil prokaryotes by calculating the convex hull area as a proxy for range and determining the average relative abundance of each ASV across all samples (Fig. 3). Based on their spatial extent, we classified ASVs into three categories: (1) local distribution (within plot), with a convex hull area of less than approximately 10 m^2, (2) intermediate distribution (within site), ranging between roughly 150 m^2 and 170,000,000 m^2, (3) and broad distribution, exceeding around 5,000,000,000 m^2.

We observed a strong positive relationship between ASV relative abundance and range size (Kendall's $\tau = 0.578$ $p < 0.001$). Most ASVs (99%) exhibited average relative abundances below 0.1% (mean: 0.008%). Dominant prokaryotes (> 0.1% average relative abundance) primarily belonged to Proteobacteria (n = 57) and Actinobacteriota (n = 30). Acidobacteriota, Bacteroidota, Verrucomicrobiota, Firmicutes, Myxococcota, and Crenarchaeota included ≤ 10 dominant ASVs each. Of these, 21 ASVs belonging to Actinobacteriota (8 ASVs), Proteobacteria (6 ASVs), Crenarchaeota (3 ASVs), Verrucomicrobiota (2 ASVs), Acidobacteriota and Myxococcota (1 ASV each) were restricted to intermediate spatial ranges (Table S1). All dominant Firmicutes belonged to Bacillales, including one unclassified *Planococcaceae*, and were detected at broad ranges. Notably, no dominant Firmicutes or Bacteroidota were detected at intermediate distributions. Across broad spatial scales, we observed consistently strong positive correlations between ASV relative abundance and range size for Proteobacteria ($\tau = 0.547$, $p < 0.001$), Myxococcota ($\tau = 0.632$, $p < 0.001$), Firmicutes ($\tau = 0.567$, $p < 0.001$), Bacteroidota ($\tau = 0.545$, $p < 0.001$), Acidobacteriota ($\tau = 0.525$, $p < 0.001$), Verrucomicrobiota ($\tau = 0.538$, $p < 0.001$), Actinobacteriota ($\tau = 0.493$, $p < 0.001$), and Crenarchaeota ($\tau = 0.357$, $p = 0.034$).

Highly dominant ASVs (>0.25% average relative abundance) were distributed among Proteobacteria (15 ASVs), Firmicutes (4 ASVs), Actinobacteriota (3 ASVs), and Crenarchaeota (3 ASVs). Only three ASVs belonging to Proteobacteria (unclassified *Sutterellaceae*), Actinobacteriota (unclassified *MB-A2-108*), and Crenarchaeota (unclassified *Nitrososphaeraceae*), occurred at intermediate scales. The rest were broadly distributed, including ASVs from Proteobacteria (14 ASVs), Firmicutes (3 ASVs), two ASVs each from Acidobacteriota and Actinobacteriota, and one each from Bacteroidota, Crenarchaeota, Myxococcota. The two most abundant taxa were a *Bradyrhizobium* and an unclassified *Bacillales*, each with a ~1% average relative abundance and spanning the broad spatial scale. We detected no dominant ASVs at local ranges (Fig. 3).

At intermediate distributions, we observed positive between abundance and range for Proteobacteria ($\tau = 0.456$, $p < 0.001$), Myxococcota ($\tau = 0.504$, $p < 0.001$), Actinobacteriota ($\tau = 0.483$, $p < 0.001$), Bacteroidota ($\tau = 0.471$, $p < 0.001$), Acidobacteriota ($\tau = 0.431$, $p < 0.001$), Verrucomicrobiota ($\tau = 0.498$, $p < 0.001$), Crenarchaeota ($\tau = 0.562$, $p = 0.00036$), and

Firmicutes ($\tau = 0.362$, $p = 0.0031$), but these were generally weaker than those observed for broadly-distributed taxa. Few significant relationships between abundance and range were observed for locally distributed taxa, including Actinobacteriota ($\tau = 0.091$, $p = 0.0034$), Acidobacteriota ($\tau = 0.095$, $p = 0.0017$), Proteobacteria ($\tau = 0.069$, $p = 0.00072$).

To examine whether growth strategies were linked to spatial distributions, we estimated bacterial growth potential using mean 16S rRNA operon copy number as a proxy. We detected no significant global association across taxa (Kendall's $\tau = 0.008$, $p = 0.624$, Supplemental Fig. S2). Actinobacteriota at the intermediate scale showed a small positive relationship ($\tau = 0.14$, $p = 0.026$), whereas Bacteroidota and Myxococcota showed weak negative relationships at local scales ($\tau = -0.23$, $p = 0.025$; $\tau = -0.24$, $p = 0.080$, respectively). Most relationships were non-significant, and several phyla lacked sufficient data to evaluate correlations.

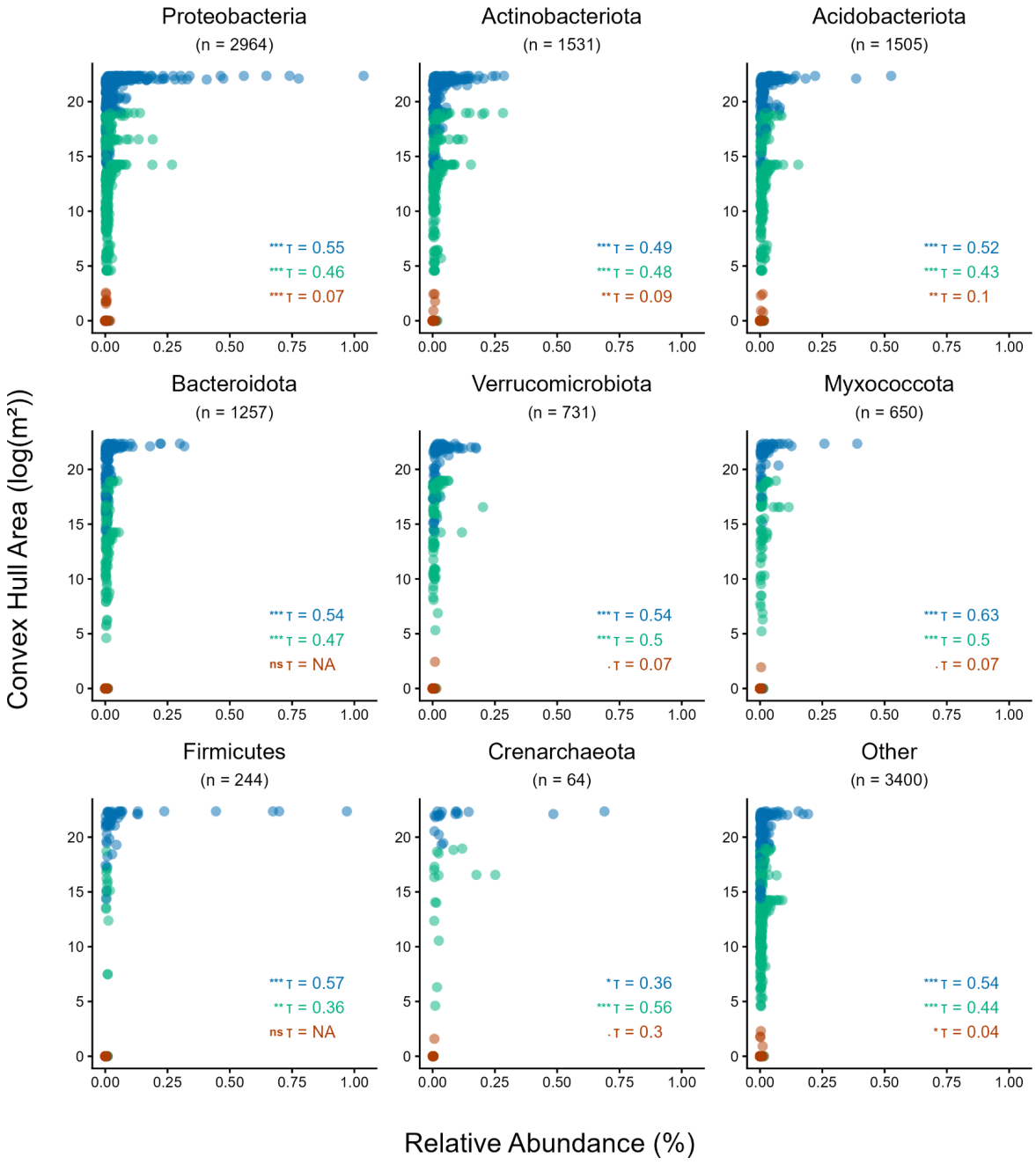

Fig. 3 Relationship between average relative abundance (%) and spatial range (convex hull area in ln(m²)) of soil prokaryotic ASVs in grasslands, faceted by phylum. Each point represents a single ASV. For faceting, the eight most abundant prokaryotic phyla, and the rest of the taxa are shown separately. Colors indicate the convex hull area relative to the sampling design: red is within plot, green is within sites, and blue is across sites. Kendall's rank estimation depicts the correlation. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$, ns = not significant.

Discussion

Understanding the factors governing the distribution of soil prokaryotes is essential to advance microbial biogeography and inform strategies for sustainable management of belowground diversity in grasslands. While previous research has established that bacterial communities often exhibit distance-decay relationships (Martiny *et al.* 2011), the extent to which these patterns reflect spatial constraints, environmental filtering, biotic interactions or combinations of these factors remains insufficiently resolved–particularly across multiple spatial scales (Barbour *et al.* 2022). Using variation partitioning together with commonality analysis, we distinguished between (i) the direct influence of geographic, abiotic, and biotic factors and (ii) the indirect influence of spatially structured abiotic and biotic changes on prokaryotic communities. Our findings shed light on the soil prokaryotic distance-decay relationships (DDR), which emerges from the interaction between dispersal limitation and abiotic and biotic selective factors, and their respective DDRs.

We found a negative effect of geographic distance on community similarity, with a ~5% decrease in similarity with each doubling of distance while controlling for soil and plant-related characteristics. This aligns with existing research into microbial DDRs (Clark *et al.* 2021; Martiny *et al.* 2011), and underscores the role of spatial separation in microbial community assembly, likely through dispersal and historical contingencies. Indeed, according to our commonality analysis, geographic distance was the second most important driver of prokaryotic communities. However, our variation partitioning analysis found a small but significant individual effect of geographic location, suggesting that while spatial processes do affect community composition, much of their influence is confounded with biotic and abiotic gradients driving gradual changes in prokaryotic communities. This aligns with findings from a synthesis by Hanson *et al.* (2012), which showed that environmental variables explained more variation in bacterial composition (26.9%) than geographic distance (10.3%), with studies more likely to detect significant effects from environmental than from spatial factors. Importantly, the lack of influence of geographic location may have been partially due to our sampling design, which had a minimum distance between samples of 12.96 m. While it boasted high spatial resolution relative to the existing literature (Griffiths *et al.* 2011), this distance may have still been too large to capture dispersal ranges (Barbour *et al.* 2022).

When considered individually, soil properties (soil carbon concentration, pH, and the carbon-to-nitrogen ratio) had only minor direct effects on prokaryotic community structure, which were marginally non-significant (1%; $p = 0.08$). Nevertheless, when considering their interaction with plants and distance, soil properties were the most important direct of soil prokaryotic communities, explaining nearly 30% of the total variance in their composition. Changes in soil properties accounted for over half of the variation in the prokaryotic DDR and contributed the largest unique share. This suggests that abiotic factors act in concert with biotic interactions to structure prokaryotic communities (Drenovsky *et al.* 2009) pointing to synergistic

interactions between abiotic and biotic filters. For instance, soil pH has long been recognized as a key determinant of microbial diversity, but its effect may be amplified or mitigated by plant species composition (Fierer & Jackson 2006; Rousk *et al.* 2010). Taken together, these findings support the notion that environmental conditions are important, but represent just one component of the broader set of selective factors influencing microbial community assembly, including active biotic interactions and spatial gradients in the abiotic environment (Vellend 2010).

Prokaryotic communities varied considerably, even within the same plot over distances of several meters in response to changes in physicochemical properties, highlighting the role of fine-scale environmental heterogeneity in driving soil prokaryotes. The low prokaryotic similarity observed between nearby samples–even within a relatively homogeneous grassland environment–underscores the need to reconsider standard sampling designs for soil microbiota. Popular sampling strategies often rely on pooling samples to characterize microbial diversity at the plot scale. For example, multiscale national assessments of soil bacterial and fungal diversity in France, UK, and Ireland randomly sampled and homogenized between 5-25 soil cores from each of the 20 - 900 m2 plots, respectively (Griffiths *et al.* 2011; Hazard *et al.* 2013; Ranjard *et al.* 2013). Our findings suggest that these designs are likely unable to disentangle the role of local dispersal limitation and biotic interactions in prokaryotic community assembly, and likely introduce further noise to the scales at which selection and microbial dispersal occur in soil ecosystems (Clark *et al.* 2021).

We found a strong contribution of plant community composition, but not vegetation characteristics, to the change in soil prokaryotic communities. The strong, unique contribution of plant composition likely reflects the direct two-way interactions between plants and soil prokaryotes (e.g., through root exudates, litter quality, and rhizosphere dynamics; Philippot *et al.* 2013; Trivedi *et al.* 2020), while the shared contribution of plant composition and abiotic parameters likely highlights the role of plants in modulating the microbial environment (e.g., soil pH, nutrient availability, and organic matter content). The lack of influence of vegetation characteristics further suggests that trait-based filtering of soil prokaryotes may be weak or indirect, potentially mediated through plant effects on soil chemistry that were already captured by other measured variables.

Our dual analytical strategy allowed us to quantify the contribution of abiotic and biotic DDRs to soil prokaryotic DDRs. We found that the interaction between geographic location and physicochemical processes explained a small, but significant portion of the variance in prokaryotic DDRs. This suggests that spatial changes in soil properties may modulate the effects of distance in a multiplicative manner, adding further nuance to the role of the abiotic environment in structuring soil prokaryotes, both locally, and across spatial gradients (Legendre *et al.* 2005). Similarly, the influence of the plant DDR was primarily observed through shared effects with changes in geographic distance and soil physicochemical properties and not through unique contributions. This suggests that shifts in plant composition across sites do not

necessarily translate into corresponding prokaryotic community changes at the spatial scale examined, likely because their effects are context-dependent and mediated through co-varying environmental conditions. Previous studies have shown that spatial heterogeneity in soil properties and stochastic dispersal processes can outweigh deterministic selection by plant communities in structuring microbial communities (Regan *et al.* 2017, 2014; Richter-Heitmann *et al.* 2020). In our study, abiotic filtering and dispersal limitation likely dominated prokaryotic community assembly, consistent with findings from previous large-scale grassland surveys (Delgado-Baquerizo *et al.* 2018).

We further explored the potential contribution of dispersal limitation to community assembly by examining the relationship between prokaryotic ASV abundance and geographic range. Most taxa exhibited low abundances and were restricted to local ranges, consistent with the notion of a rare biosphere (Jousset *et al.* 2017; Nemergut *et al.* 2011). We also observed the accumulation of dominant ASVs (especially Proteobacteria) at the largest spatial scale, highlighting the presence of successful generalists in the community, consistent with established ecological classifications (Fierer *et al.* 2007). These dominant phyla, which include Proteobacteria, Actinobacteriota, and Bacteroidota, exhibited wide geographic distributions and often increased in abundance at broader spatial scales, suggesting high dispersal capacity and competitive success across heterogeneous soil environments (Shade & Gilbert 2015). The consistent dominance of Proteobacteria across scales likely reflects their metabolic versatility, rapid growth, and ability to exploit diverse root-derived carbon sources (Fierer *et al.* 2007; Philippot *et al.* 2013; Spain *et al.* 2009). Interestingly, we found a small and diverse group of taxa that were dominant at intermediate scales, likely indicating that dominant taxa can also be dispersal limited. These taxa were found across all phyla except for Firmicutes and Bacteroidota, suggesting that dispersal limitation (e.g. the ability to form stress-resistant spores or cyst-like structures; Mandic-Mulec *et al.* 2015) plays a key role in shaping these patterns.

Notably, we found a subset of rare microbes with broad and intermediate distributions, suggesting the existence of rare generalists-low-abundance taxa that can disperse broadly or persist in diverse habitats. While the existence of rare microbes is well established (Lynch & Neufeld 2015; Sogin *et al.* 2006), less is known about the distribution of these taxa, and it is generally assumed thatrare taxa are locally distributed (Lindh *et al.* 2017; Meyer *et al.* 2018; Shade & Stopnisek 2019; Thompson *et al.* 2017). This may be due to analytical choices, as similarity analyses, which fail to capture rare taxa, can overestimate compositional similarity across communities, thus weakening distance-decay relationships under the assumption that rare taxa are also spatially restricted (Clark *et al.* 2021). Our observation that rare ASVs can persist at intermediate and broad distributions(an area exceeding 5,000 km^2) challenges this assumption. The presence of both narrowly and broadly distributed rare taxa across all phyla suggests diverse dispersal strategies, potentially shaped by microbial life-history traits that facilitate survival and transport - such as dormancy in vegetative cells, cysts, or spores (Hanson *et al.* 2012; Locey *et*

al. 2020). However such strategies, and their influence on community assembly, remain poorly characterized (Barbour *et al.* 2022) and require further study.

Conclusion

Our study paints a more nuanced picture of how the biotic and abiotic environment result in the spatial distribution of soil microbes. We demonstrate that soil prokaryotes follow a distance-decay relationship, but that this relationship is driven primarily by spatial gradients in soil physicochemical parameters and the composition of plant communities aboveground (i.e., selective factors), or selection, rather than dispersal limitationTaken together, our findings suggest a modification to the hypothesis that "everything is everywhere, but the environment selects" (Baas Becking & Nicolai 1934): "everything is *not* everywhere. Still, the environment selects".

Acknowledgments

We thank the managers of the three Exploratories, Julia Bass, Max Müller, Sven Pompe, Anna K. Franke, Melissa Jüds, and all former managers for their work in maintaining the plot and project infrastructure; Victoria Grießmeier for giving support through the central office, Andreas Ostrowski for managing the central database, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze, Wolfgang W. Weisser and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. We thank the administration of the Hainich national park, the UNESCO Biosphere Reserve Swabian Alb and the UNESCO Biosphere Reserve Schorfheide-Chorin as well as all land owners for the excellent collaboration. The work has been funded by the DFG Priority Program 1374 "Biodiversity- Exploratories" (#512281171). Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg.

Data Availability

This work is based on data elaborated by several projects of the Biodiversity Exploratories program (DFG Priority Program 1374). The datasets underlying this study are publicly available in the Biodiversity Exploratories Information System (BExIS): http://doi.org/10.17616/R32P9Q. The datasets are listed in the references section. Specifically, we used the following datasets: the position of sampling quadrats in SeBAS 2020-2021 (dataset 31386; Muro 2022), plant trait measurements (dataset 31971; Meyer *et al.* 2024b), plant species inventories (dataset 31976; Meyer *et al.* 2024a), soil carbon and nitrogen content, soil moisture, and pH in three quadrats (dataset 32112; Meyer & Linstädter 2025a), aboveground biomass (dataset 32211; Meyer *et al.* 2025), and vegetation cover (dataset 32212; Meyer & Linstädter 2025b). In accordance with the Biodiversity Exploratories' data and publication policy, newly collected data are subject to an

embargo period of three years after completion of data collection/assembly, during which access is restricted to data owners and collectors. This embargo applies to datasets 31386, 31971, 31976, 32112, 32211, and 32212. Raw sequence data are available in NCBI's Sequence Read Archive under accession number PRJNA1284051. The processed sequence data and corresponding metadata are available in BExIS under accession numbers 32155 and 32156 Neisse & Jurburg (2025a, b).

References

Baas Becking, L.G.M. & Nicolai, E. (1934). On the ecology of a Sphagnum bog. *Blumea: Biodiversity, Evolution and Biogeography of Plants*, 1, 10–45.

Barbour, K.M., Barrón-Sandoval, A., Walters, K.E. & Martiny, J.B.H. (2022). Towards quantifying microbial dispersal in the environment. *Environmental Microbiology*, 25, 137–142.

Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *The ISME Journal*, 4, 1357–1365.

Blüthgen, N., Dormann, C.F., Prati, D., Klaus, V.H., Kleinebecker, T., Hölzel, N., *et al.* (2012). A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. *Basic and Applied Ecology*, 13, 207–220.

Boyle, J. (2017). *Package 'GeoRange': Calculating Geographic Range from Occurrence Data* (No. 0.1.0).

Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J. & Holmes, S.P. (2016). Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research*, 5, 1492.

Clark, D.R., Underwood, G.J.C., McGenity, T.J. & Dumbrell, A.J. (2021). What drives study-dependent differences in distance–decay relationships of microbial communities? *Global Ecology and Biogeography*, 30, 811–825.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., *et al.* (2018). A global atlas of the dominant bacteria found in soil. *Science*, 359, 320–325.

Drenovsky, R.E., Steenwerth, K.L., Jackson, L.E. & Scow, K.M. (2009). Land use and climatic factors structure regional patterns in soil microbial communities. *Global Ecology and Biogeography*, 19, 27–39.

Dumbrell, A.J., Nelson, M., Helgason, T., Dytham, C. & Fitter, A.H. (2009). Relative roles of niche and neutral processes in structuring a soil microbial community. *The ISME Journal*, 4, 337–345.

Eisenhauer, N., Beßler, H., Engels, C., Gleixner, G., Habekost, M., Milcu, A., *et al.* (2010). Plant diversity effects on soil microorganisms support the singular hypothesis. *Ecology*, 91, 485–496.

Fierer, N., Bradford, M.A. & Jackson, R.B. (2007). Toward an ecological classification of soil bacteria. *Ecology*, 88, 1354–1364.

Fierer, N. & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103, 626–631.

Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., *et al.* (2010). Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. *Basic and Applied Ecology*, 11, 473–485.

Frindte, K., Pape, R., Werner, K., Löffler, J. & Knief, C. (2019). Temperature and soil moisture control microbial community composition in an arctic–alpine ecosystem along elevational and micro-topographic gradients. *The ISME Journal*, 13, 2031–2043.

Graco-Roza, C., Aarnio, S., Abrego, N., Acosta, A.T.R., Alahuhta, J., Altman, J., *et al.* (2022). Distance decay 2.0 – A global synthesis of taxonomic and functional turnover in ecological communities. *Global Ecology and Biogeography*, 31, 1399–1421.

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M. & Whiteley, A.S. (2011). The bacterial biogeography of British soils. *Environmental Microbiology*, 13, 1642–1654.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B.H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10, 497–506.

Hartig, F., Lohse, L. & de Souza leite, M. (2024). *Package 'DHARMa': Residual Diagnostics for*

Hierarchical (Multi-Level / Mixed) Regression Models (No. 0.4.7).
Hazard, C., Gosling, P., Van Der Gast, C.J., Mitchell, D.T., Doohan, F.M. & Bending, G.D. (2013). The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *The ISME Journal*, 7, 498–508.
Hijmans, R.J., Karney, C., Williams, E. & Vennes, C. (2024). *Package 'geosphere': Spherical Trigonometry* (No. 1.5-20).
Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., *et al.* (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. *The ISME Journal*, 11, 853–862.
Labouyrie, M., Ballabio, C., Romero, F., Panagos, P., Jones, A., Schmid, M.W., *et al.* (2023). Patterns in soil microbial diversity across Europe. *Nat Commun*, 14, 3311.
Lai, J., Zou, Y., Zhang, S., Zhang, X. & Mao, L. (2022). glmm.hp: an R package for computing individual effect of predictors in generalized linear mixed models. *Journal of Plant Ecology*, 15, 1302–1307.
Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005). Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, 75, 435–450.
Lemoine, N.P., Adams, B.J., Diaz, M., Dragone, N.B., Franco, A.L.C., Fierer, N., *et al.* (2023). Strong Dispersal Limitation of Microbial Communities at Shackleton Glacier, Antarctica. *mSystems*, 8.
Lindh, M.V., Sjöstedt, J., Ekstam, B., Casini, M., Lundin, D., Hugerth, L.W., *et al.* (2017). Metapopulation theory identifies biogeographical patterns among core and satellite marine bacteria scaling from tens to thousands of kilometers. *Environmental Microbiology*, 19, 1222–1236.
Locey, K.J., Muscarella, M.E., Larsen, M.L., Bray, S.R., Jones, S.E. & Lennon, J.T. (2020). Dormancy dampens the microbial distance–decay relationship. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375, 20190243.
Lynch, M.D.J. & Neufeld, J.D. (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, 13, 217–229.
Mandic-Mulec, I., Stefanic, P. & van Elsas, J.D. (2015). Ecology of Bacillaceae. *Microbiology Spectrum*, 3.
Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D. & Horner-Devine, M.C. (2011). Drivers of bacterial β-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences*, 108, 7850–7854.
McMurdie, P.J. & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8, e61217.
Meyer, K.M., Memiaghe, H., Korte, L., Kenfack, D., Alonso, A. & Bohannan, B.J.M. (2018). Why do microbes exhibit weak biogeographic patterns? *The ISME Journal*, 12, 1404–1413.
Meyer, S.N. & Linstädter, A. (2025a). Soil carbon and nitrogen content, soil moisture and pH value (0-10cm) in three quadrats along N-S grassland plot margin in spring and autumn 2023.
Meyer, S.N. & Linstädter, A. (2025b). Vegetation cover in three 1 m^2 quadrats of grassland plots in 2023 (spring & autumn).
Meyer, S.N., Muro, J., Schwarz, L.-M., Männer, F.A., Niedersen, J. & Linstädter, A. (2025). Aboveground biomass in three 0.6 x 0.6 m2 quadrats of grassland plots in 2020, 2021, 2023 (spring & autumn).
Meyer, S.N., Schwarz, L.-M., Männer, F.A., Niedersen, J. & Linstädter, A. (2024a). Plant species inventories in experimental plots (not all) of grasslands for trait measurements.
Meyer, S.N., Schwarz, L.-M., Muro, J., Niedersen, J. & Linstädter, A. (2024b). Plant trait measurements in experimental plots (not all) of grasslands.
Muro, J. (2022). Position of sampling quadrats in SeBAS 2020-2021.

Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F.A., Schwarz, L.-M., *et al.* (2022). Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sensing of Environment*, 282, 113262.

Neisse, N. & Jurburg, S. (2025a). Amplicon sequence variants, their taxonomic classification, and relative abundance per sample in grassland experimental plots (not all).

Neisse, N. & Jurburg, S. (2025b). Geographic location, soil physicochemical properties, plant community composition and vegetation characteristic in grassland experimental plots (not all).

Nekola, J.C. & McGill, B.J. (2014). Scale dependency in the functional form of the distance decay relationship. *Ecography*, 37, 309–320.

Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, 26, 867–878.

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., *et al.* (2011). Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology*, 13, 135–144.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F., *et al.* (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews*, 77, 342–356.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., *et al.* (2007). *vegan: Community Ecology Package*. *CRAN: Contributed Packages*. The R Foundation.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., *et al.* (2016). Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.*, 64, 715.

Philippot, L., Raaijmakers, J.M., Lemanceau, P. & van der Putten, W.H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology*, 11, 789–799.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41, D590–D596.

R. Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N.P.A., Lelievre, M., *et al.* (2013). Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat Commun*, 4, 1434.

Regan, K., Stempfhuber, B., Schloter, M., Rasche, F., Prati, D., Philippot, L., *et al.* (2017). Spatial and temporal dynamics of nitrogen fixing, nitrifying and denitrifying microbes in an unfertilized grassland soil. *Soil Biology and Biochemistry*, 109, 214–226.

Regan, K.M., Nunan, N., Boeddinghaus, R.S., Baumgartner, V., Berner, D., Boch, S., *et al.* (2014). Seasonal controls on grassland microbial biogeography: Are they governed by plants, abiotic properties or both? *Soil Biology and Biochemistry*, 71, 21–30.

Richter-Heitmann, T., Hofner, B., Krah, F.-S., Sikorski, J., Wüst, P.K., Bunk, B., *et al.* (2020). Stochastic Dispersal Rather Than Deterministic Selection Explains the Spatio-Temporal Distribution of Soil Bacteria in a Temperate Grassland. *Frontiers in Microbiology*, 11.

Rousk, J., Bååth, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., *et al.* (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal*, 4, 1340–1351.

Shade, A. & Gilbert, J.A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology*, 23, 335–340.

Shade, A. & Stopnisek, N. (2019). Abundance-occupancy distributions to prioritize plant core microbiome membership. *Current Opinion in Microbiology*, 49, 50–58.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences*, 103, 12115–12120.

Spain, A.M., Krumholz, L.R. & Elshahed, M.S. (2009). Abundance, composition, diversity and novelty of soil Proteobacteria. *The ISME Journal*, 3, 992–1000.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K. & Schmidt, T.M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, 43, D593–D598.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., *et al.* (2017). A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature*, 551, 457–463.

Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T. & Singh, B.K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology*, 18, 607–621.

Van Der Heijden, M.G.A., Bardgett, R.D. & Van Straalen, N.M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, 11, 296–310.

Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85, 183–206.

Vos, M., Wolf, A.B., Jennings, S.J. & Kowalchuk, G.A. (2013). Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiology Reviews*, 37, 936–954.

Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E.E. & Van Der Heijden, M.G.A. (2019). Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nat Commun*, 10, 4841.

Walters, K.E., Capocchi, J.K., Albright, M.B.N., Hao, Z., Brodie, E.L. & Martiny, J.B.H. (2022). Routes and rates of bacterial dispersal impact surface soil microbiome composition and functioning. *The ISME Journal*, 16, 2295–2304.

Zhou, X., Tahvanainen, T., Malard, L., Chen, L., Pérez-Pérez, J. & Berninger, F. (2024). Global analysis of soil bacterial genera and diversity in response to pH. *Soil Biology and Biochemistry*, 198, 109552.

Supplementary materials

Formula to calculate the effect of doubling x in a GLM with a log link.

$$\Delta = \beta * ln(2x) - \beta * ln(x) = \beta * [ln(x) + ln(2) - ln(x)] = \beta * ln(2) \quad (S1)$$

$$\beta * ln(2) = -0.736 * 0.6931 = -0.051 \quad (S2)$$

To back transform the log transformed geographic predictor, we use the exponential.

$$exp(-0.051) = 0.95 \quad (S3)$$

$$1 - 0.95 = 0.05 \quad (S4)$$

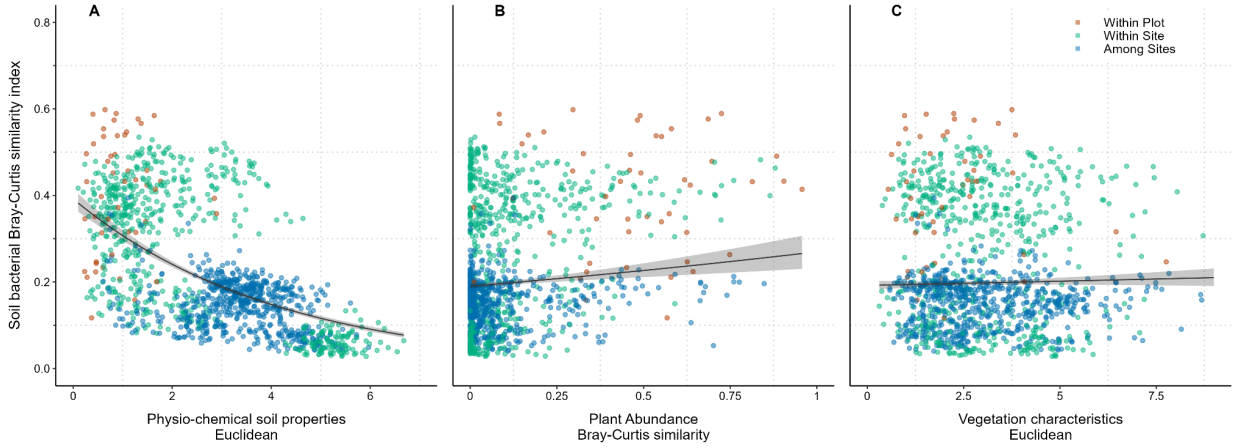

Fig. S1 Effect plots from a Gamma generalized linear model (GLM) with a log link, four predictors, and a single two-way interaction, predicting soil prokaryotic community similarity (Bray-Curtis index). Each panel shows the modeled relationship between Bray-Curtis similarity and one predictor: (A) physicochemical similarity, (B) plant community similarity, and (C) vegetation characteristic similarity, with geographic distance and the remaining predictors held at their mean values. Physicochemical differences and plant community similarity significantly influenced prokaryotic similarity ($p < 0.001$), whereas vegetation characteristic similarity had no significant effect. Points represent observed pairwise similarity values; solid lines show model predictions, and shaded ribbons represent 95% confidence intervals.

Table S1 Taxonomic identification, spatial distribution, and relative abundance of dominant (> 0.1% average relative abundance) ASVs that occur within the same sampling site, but not within the same plot. Each ASV is listed with its corresponding phylum, family (representing the last known taxonomic identification), genus, log-transformed convex hull area (CH; m²) as a proxy for spatial extent, and average relative abundance (%) across samples. Family depicts last known identification at that level. Unspecified genera indicate unresolved taxonomic classification at that level.

ID	Phylum	Family	Genus	Convex hull area	Abundance
ASV23	Actinobacteriota	MB-A2-108	unspecified	17.193	0.282
ASV28	Proteobacteria	*Sutterellaceae*	unspecified	15.441	0.268
ASV31	Crenarchaeota	*Nitrososphaeraceae*	unspecified	15.52	0.252
ASV40	Actinobacteriota	MB-A2-108	unspecified	17.193	0.208
ASV44	Verrucomicrobiota	*Xiphinematobacteraceae*	*Candidatus Xiphinematobacter*	15.52	0.202
ASV46	Proteobacteria	*Geminicoccaceae*	unspecified	15.52	0.19
ASV48	Proteobacteria	*Sutterellaceae*	unspecified	15.237	0.189
ASV55	Crenarchaeota	*Nitrososphaeraceae*	unspecified	15.52	0.176
ASV59	Actinobacteriota	*Nocardioidaceae*	Kribbella	15.298	0.199
ASV71	Actinobacteriota	*Actinomarinales*	unspecified	15.476	0.154
ASV73	Acidobacteriota	*Vicinamibacterales*	unspecified	15.376	0.153
ASV80	Proteobacteria	*Geminicoccaceae*	unspecified	15.52	0.134
ASV87	Proteobacteria	SC-I-84	unspecified	17.204	0.139
ASV90	Actinobacteriota	*Nocardioidaceae*	Kribbella	17.198	0.148
ASV94	Proteobacteria	TRA3-20	unspecified	17.188	0.116
ASV100	Actinobacteriota	*Gaiellales*	unspecified	17.204	0.134
ASV103	Actinobacteriota	*Gaiellaceae*	*Gaiella*	15.52	0.122
ASV104	Verrucomicrobiota	*Xiphinematobacteraceae*	*Candidatus Xiphinematobacter*	15.44	0.118

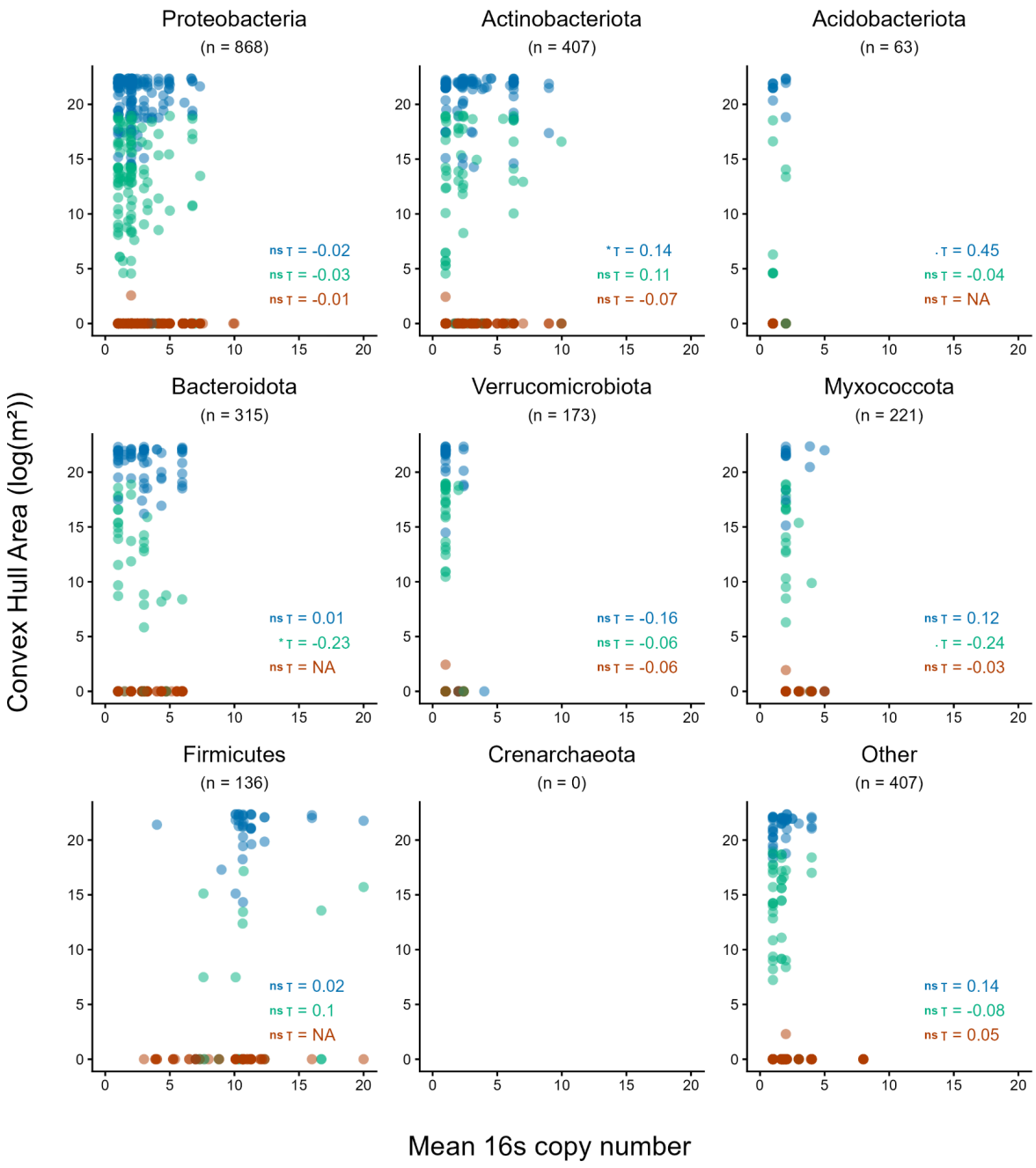

Fig. S2 Relationship between mean 16s copy number and spatial range (convex hull area in ln(m²)) of soil prokaryotic ASVs in grasslands, faceted by phylum. Each point represents a single ASV. For faceting, the eight most abundant prokaryotic phyla, and the rest of the taxa are shown separately. Colors indicate the convex hull area relative to the sampling design: red is within plot, green is within sites, and blue is across sites. Kendall's rank estimation depicts the correlation. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$, ns = not significant.