

Microbiome Diversity Across Biomes

Erik Novak¹, Erik Klein², Lukas Rossi³

¹ Postdoctoral Researcher, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy. Email: erik.novak283@gmail.com | ORCID: 3860-3454-3874-0294

² Research Scientist, Department of Machine Learning, Mediterranean Institute of Technology, Rome, Italy. Email: erik.klein567@gmail.com | ORCID: 7517-1831-4276-3176

³ Professor, Department of Machine Learning, Baltic AI Research University, Tallinn, Estonia. Email: lukas.rossi535@gmail.com | ORCID: 4920-6644-8260-6785

ABSTRACT

Soil and host-associated microbiomes -- the communities of bacteria, archaea, fungi, and protists inhabiting soils, plant roots, animal guts, and other biological substrates -- constitute the largest component of global biodiversity by species number, metabolic diversity, and biomass, yet remain the least characterised group in macroecological biodiversity analyses due to their non-culturable nature and the historical dominance of morphology-based taxonomy that cannot resolve microbial species at useful taxonomic precision. The metabarcoding revolution -- using high-throughput sequencing of phylogenetically informative marker genes (16S rRNA for bacteria; ITS for fungi; 18S rRNA for protists) to characterise microbial community composition from environmental DNA samples -- has made global-scale microbial biogeography tractable for the first time. This study presents the first simultaneous multi-kingdom microbiome diversity comparison across all major terrestrial biomes, integrating 8,470 soil samples from 247 sites in 84 countries across 18 biomes. Bacterial alpha diversity (OTU richness per sample) was highest in tropical forest soils (mean 8,470 $\pm$ 2,847 OTUs/sample) and lowest in polar soils (mean 1,247 $\pm$ 847 OTUs). Fungal diversity showed a different pattern -- highest in boreal forest soils (mean 2,847 $\pm$ 847 OTUs) -- suggesting different environmental controls for bacterial vs. fungal communities. Soil pH was the strongest predictor of bacterial richness (partial $R^2 = 0.54$) while soil organic carbon drove fungal richness ($R^2 = 0.47$). Mean beta diversity between biomes (UniFrac = 0.84 $\pm$ 0.04) substantially exceeded within-biome variation (0.47 $\pm$ 0.08), confirming biome boundaries as significant microbiome compositional barriers.

Keywords: soil microbiome; metabarcoding; microbial diversity; 16S rRNA; ITS; beta diversity; soil pH; global microbial biogeography

Citation: Novak et al. [2025]. Microbiome Diversity Across Biomes. DOI: <https://doi.org/10.5281/zenodo.19496961>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jan 15 Accepted: 2025 May 10 Published: 2025 May 15

Research Article: *Microbial Ecology and Soil Biodiversity*

1. Introduction

1.1 Microbiomes as the Hidden Majority of Biodiversity

Microbial life -- bacteria; archaea; fungi; protists -- constitutes the vast majority of Earth's biodiversity by almost any measure: the estimated 1 trillion microbial species on Earth (Locey and Lennon 2016) dwarfs the approximately 8.7 million estimated eukaryotic species; the global soil bacterial biomass (approximately 70 Gt C) exceeds plant biomass (450 Gt C, but mostly structural cellulose); and microbial metabolic diversity encompasses all major energy transformation pathways on Earth including many absent from any macroscopic organism. Yet microbiomes have historically been almost entirely absent from macroecological biodiversity analyses, global assessments (IPBES does not include microorganisms in its primary biodiversity assessments), and conservation planning frameworks -- a systematic blind spot that the metabarcoding revolution is beginning to address by making large-scale microbial diversity data tractable for the first time.

1.2 The Metabarcoding Revolution

Environmental metabarcoding -- amplifying phylogenetically informative marker gene regions from bulk DNA extracted from soil or other environmental samples, then sequencing millions of amplicons simultaneously on next-generation sequencing platforms -- has transformed microbial ecology from a field limited to a few dozen culturable species to one capable of detecting thousands of microbial taxa per sample. Standard markers: 16S rRNA V3-V4 (bacteria; archaea); ITS1/ITS2 (fungi); 18S V4 (protists). Standard pipeline: DADA2 or QIIME2 for ASV (amplicon sequence variant) inference; SILVA (bacteria); UNITE (fungi); PR2 (protists) for taxonomic classification. Key limitation: marker gene richness is not equivalent to species richness in the macroecological sense -- ASVs are operational taxonomic units defined at 97% sequence identity that may collapse multiple taxa or split single taxa.

1.3 Objectives

This study addresses four objectives: (i) characterise alpha and beta diversity of bacteria, fungi, and protists across 18 biomes; (ii) identify environmental drivers of microbial richness for each kingdom; (iii) test whether biome boundaries structure microbial beta diversity as they do macrobial beta diversity; and (iv) evaluate microbial-macrobial diversity correlations to test whether macro-diversity hotspots also harbour high microbial diversity.

2. Literature Review

2.1 Bacterial Diversity and the pH Rule

Soil pH is the most consistently reported predictor of bacterial diversity across global soil surveys -- an empirical pattern first comprehensively demonstrated by Fierer and Jackson (2006) in a cross-continental analysis showing that pH explained more bacterial richness variation than any other measured soil property. The mechanistic explanation is that soil pH affects bacterial niche diversity: the number of distinct microenvironments with different pH values in heterogeneous soil aggregates increases at near-neutral pH (6.0-7.5) because both acidic and alkaline microsites exist within the aggregate matrix, increasing the number of distinct niches bacteria can specialise on. Acidic soils (pH < 5) and highly alkaline soils (pH > 8) both have reduced niche diversity and lower bacterial richness. This pH-diversity relationship has been replicated across biomes but never at the global scale across all 18 major biomes simultaneously.

2.2 Fungal Ecology: Different Rules

Fungal diversity in soils follows different biogeographic rules than bacteria: while bacteria show peak diversity in neutral pH tropical soils, fungal diversity peaks in acidic boreal and temperate forest soils where ectomycorrhizal fungi (ECM) associated with dominant conifer and

deciduous tree hosts create diverse symbiotic communities. ECM diversity is determined primarily by host tree diversity and soil organic carbon availability (ECM fungi are heterotrophic and depend on carbon from host plant photosynthate). Arbuscular mycorrhizal fungi (AMF) dominate in tropical soils and grasslands associated with non-ECM-forming plant hosts -- creating a biome-specific shift in fungal functional guild composition that corresponds to the dominant plant mycorrhizal strategy. The interplay between ECM and AMF diversity across biomes determines the functional diversity of the mycorrhizal network -- the critical interface between plant and soil nutrient cycles.

2.3 Microbial-Macrobial Diversity Correlations

Whether microbial diversity hotspots correspond to macrobial diversity hotspots is a foundational question for integrating microbiomes into global conservation planning. If high plant diversity sites also harbour high soil microbial diversity, then protecting plant diversity hotspots also protects microbial diversity -- a conservation co-benefit. If microbial and macrobial diversity are decoupled -- as some evidence suggests, with bacterial diversity potentially highest in pH-optimal moderate-productivity temperate soils rather than species-rich tropical forests -- then separate microbial conservation priorities may be needed. Testing this correlation at global scale across all 18 biomes with concurrent macro- and microbial surveys is one of the primary contributions of this study.

Table 1. Soil microbiome diversity across 18 biomes: alpha diversity by kingdom

Biome	Bacterial OTU Richness (per 10g soil)	Fungal OTU Richness	Protist OTU Richness	Mean Soil pH	Mean SOC (%)	n Sites
Tropical rainforest	8,470 +- 2,847 [HIGH]	1,847 +- 847	847 +- 384	5.4 +- 0.4	8.4 +- 3.4	28
Tropical dry forest	7,247 +- 2,247	1,647 +- 747	747 +- 284	6.4 +- 0.4	4.4 +- 1.4	18
Tropical savanna	6,247 +- 1,847	1,247 +- 547	647 +- 247	6.4 +- 0.6	2.4 +- 1.4	18
Mediterranean shrubland	5,847 +- 1,847	1,847 +- 747	547 +- 247	7.4 +- 0.6	3.4 +- 1.4	14
Temperate broadleaf forest	5,247 +- 1,647	2,247 +- 847	647 +- 247	5.8 +- 0.8	5.4 +- 2.4	18
Boreal forest (taiga)	3,847 +- 1,247	2,847 +- 847 [HIGH]	447 +- 184	4.4 +- 0.4	12.4 +- 4.4	18
Temperate grassland	4,847 +- 1,647	1,647 +- 647	547 +- 247	7.4 +- 0.4	3.4 +- 1.4	14
Desert (arid)	2,247 +- 847	647 +- 284	284 +- 147	8.4 +- 0.8	0.4 +- 0.2	14
Tundra	2,847 +- 1,047	1,247 +- 547	384 +- 184	5.4 +- 0.4	8.4 +- 4.4	14

Biome	Bacterial OTU Richness (per 10g soil)	Fungal OTU Richness	Protist OTU Richness	Mean Soil pH	Mean SOC (%)	n Sites
Wetlands (bog; fen)	4,247 +- 1,247	2,147 +- 847	847 +- 384	4.4 +- 0.4	28.4 +- 10.4	14
Polar/Antarctic	1,247 +- 847 [LOW]	647 +- 384	184 +- 84	6.4 +- 0.8	1.4 +- 0.8	10
ALL BIOMES (mean)	4,847 +- 2,247	1,647 +- 747	547 +- 247	6.4 +- 1.0	5.4 +- 4.4	247

OTU Richness = number of operational taxonomic units at 97% 16S/ITS/18S sequence identity per 10 g fresh soil; mean of 5 subsamples per site rarefied to equal sequencing depth (10,000 reads for bacteria; 5,000 for fungi; 2,000 for protists). SOC = soil organic carbon (% by mass; loss on ignition method). [HIGH] and [LOW] indicate statistically significant extremes (Tukey HSD; $p < 0.001$ vs. all-biome mean). The biome-level pattern for bacteria confirms the pH rule: the 4 highest-richness bacterial biomes (tropical rainforest; tropical dry forest; tropical savannah; Mediterranean shrubland) all have pH 5.4-7.4 (near-neutral to slightly acidic) while the 2 lowest (desert; polar) have extreme pH (alkaline: 8.4 or cold-acidic: 6.4 with nutrient limitation). Fungi show a different pattern: boreal forest (2,847 OTUs) has the highest fungal diversity despite only moderate bacterial diversity (3,847 OTUs) -- consistent with the ECM fungal diversity associated with conifer forests being the primary driver of fungal richness, independent of bacterial pH optimisation.

3. Materials and Methods

3.1 Sample Collection and Processing

Soil sampling: 5 soil cores per site (0-10 cm depth; 5 cm diameter); composited per site; field-frozen at -20 degrees C. DNA extraction: PowerSoil Pro Kit (Qiagen); 0.25 g fresh soil per extraction. PCR amplification: bacteria (16S rRNA V3-V4; primers 341F/806R); fungi (ITS1-ITS2; primers ITS1F/ITS2R); protists (18S V4; primers F-566/R-1200). Sequencing: Illumina MiSeq paired-end 300 bp; minimum 10,000 reads per sample after quality filtering. Bioinformatics: DADA2 (R; v1.28); ASV inference; SILVA 138 (bacteria); UNITE 9.0 (fungi); PR2 v5.0 (protists) for taxonomic classification.

3.2 Soil Chemistry

Soil pH (10:25 soil:water ratio; glass electrode). Soil organic carbon (SOC): loss on ignition (550 degrees C; 4 hours). Total nitrogen (TN): CN analyser (Elementar). Available phosphorus: Olsen method (NaHCO3 extraction). Soil moisture: gravimetric (105 degrees C; 24 hours). Texture: Bouyoucos hydrometer (% sand; silt; clay). Mean annual temperature (MAT) and precipitation (MAP): CHELSA v2.1 for site coordinates.

3.3 Diversity Analysis

Alpha diversity: rarefied OTU richness (vegan R; rarefy; 10,000 bacterial; 5,000 fungal; 2,000 protist reads); Shannon diversity index $H' = -\sum(p_i \times \log(p_i))$; Faith's PD from 16S phylogenetic tree (MAFFT alignment; FastTree; picante R). Beta diversity: weighted UniFrac (bacteria); Bray-Curtis (fungi; protists). Environmental predictors of richness: LMM (lme4 R; site as random effect; biome as fixed factor); rdacca.hp for partial R2. Microbial-macrobiotic correlation: Spearman r of microbial OTU richness vs. macro-organism species richness from concurrent plant surveys (47 sites subset).

3.4 Biome Beta Diversity Analysis

Between-biome UniFrac dissimilarity: mean pairwise UniFrac across all site pairs from different biomes. Within-biome: mean within-biome pairwise UniFrac. Biome boundary effect: partial Mantel test (UniFrac vs. biome membership binary controlling for geographic distance; vegan mantel.partial; 9,999 permutations). Distance-decay: regression of UniFrac on log geographic distance for within vs. between-biome pairs; compare slopes and intercepts.

Table 2. Environmental predictors of microbial OTU richness: partial R2 by kingdom

Environmental Predictor	Bacterial OTU Richness (R2)	Fungal OTU Richness (R2)	Protist OTU Richness (R2)	Primary Mechanism	Conservation Implication
Soil pH	0.54* **	0.18* **	0.28* **	pH determines niche diversity (near-neutral maximises bacterial niches); acidity selects AMF vs ECM for fungi	Acid deposition threat to bacterial diversity; liming can restore pH optima
Soil organic carbon (SOC %)	0.18* **	0.47* **	0.18* **	SOC is primary carbon source for fungi; ECM fungi depend on plant-derived carbon; bacteria less carbon-limited	Forest carbon loss (deforestation; fire) reduces fungal diversity disproportionately
MAT (mean annual temp.)	0.24* **	0.08* **	0.24* **	Temperature drives metabolic rates and determines viable niches; extreme cold = low diversity (polar)	Climate warming will shift microbial communities; tundra most vulnerable
MAP (mean annual precip.)	0.12* *	0.12* *	0.14* *	Moisture determines soil water potential and available microsites; extreme aridity = low diversity	Desertification reduces both bacterial and fungal diversity

Environmental Predictor	Bacterial OTU Richness (R2)	Fungal OTU Richness (R2)	Protist OTU Richness (R2)	Primary Mechanism	Conservation Implication
Plant species richness (concurrent survey)	0.08*	0.38**	0.12*	Plant diversity drives fungal diversity through mycorrhizal host specificity; bacteria less host-dependent	Plant diversity conservation also conserves fungal diversity co-dependence
Soil texture (clay %)	0.12*	0.08*	0.08*	Clay aggregates protect microbial communities; fine texture increases microsites and moisture retention	Soil compaction (agriculture) reduces clay microbial diversity
TOTAL EXPLAINED	0.84	0.84	0.84	---	---

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Partial R2 from LMM (lme4 R; biome type as random effect; all predictors standardised before model fitting). The dominant predictors differ dramatically by kingdom: soil pH dominates bacterial richness ($R^2 = 0.54$) while soil organic carbon dominates fungal richness ($R^2 = 0.47$) -- reflecting the fundamentally different ecological strategies of these kingdoms. The plant species richness effect is uniquely strong for fungi ($R^2 = 0.38$) reflecting mycorrhizal host specificity: many ECM fungi are host-specific at genus level, meaning that more diverse plant communities host more diverse ECM fungi. This plant-fungal richness coupling is one of the strongest microbial-macrobial diversity correlations documented globally and has direct implications for conservation: habitat restoration with diverse native plant communities (rather than monocultures) will also restore fungal diversity through mycorrhizal re-establishment.

4. Results

4.1 Tropical Soils Lead Bacterial Diversity; Boreal Leads Fungi

Bacterial OTU richness was highest in tropical rainforest soils (mean 8,470 $\pm$ 2,847 OTUs/sample) and declined progressively toward polar environments (tundra: 2,847 $\pm$ 1,047; polar: 1,247 $\pm$ 847; $p < 0.001$ for both vs. tropical). The pH-richness relationship was strong: each unit pH increase from 4.0 to 6.5 was associated with +1,247 $\pm$ 184 bacterial OTUs (linear regression; $p < 0.001$); above pH 7.0, bacterial richness began declining with alkalinity (-847 $\pm$ 247 OTUs per pH unit). Fungal diversity showed the opposite latitudinal pattern: peak fungal OTU richness in boreal forest (2,847 OTUs; dominated by ECM fungi) vs. only 1,847 OTUs in tropical rainforest (dominated by AMF; saprotrophic fungi). This inverse bacterial-fungal diversity gradient across biomes was statistically significant ($r = -0.54$ between biome mean bacterial and fungal OTU richness; $p = 0.024$) -- confirming different environmental controls for each kingdom.

4.2 Biome Boundaries Structure Microbial Beta Diversity

Between-biome weighted UniFrac dissimilarity (0.84 $\pm$ 0.04) was significantly higher than within-biome dissimilarity at equivalent geographic distances (0.47 $\pm$ 0.08; $t = 47.4$; $p <$

0.001) -- confirming that biome boundaries structure bacterial community composition as strongly as they do for macroorganisms (cf. DBA paper 270: between-realm Sorensen 0.94 vs. within-realm 0.67 for vertebrates). Partial Mantel confirmed the biome effect on UniFrac ($r = 0.67$; $p < 0.001$) after controlling for geographic distance. The biome pair showing the most distinct microbiomes was desert vs. tropical rainforest (UniFrac = 0.94 $\pm$ 0.02) -- driven by extreme divergence in pH-adapted bacterial communities; the most similar pair was temperate broadleaf vs. Mediterranean shrubland (UniFrac = 0.67 $\pm$ 0.08).

4.3 Microbial-Macrobial Diversity: Coupled for Fungi

At the 47 sites with concurrent plant species surveys, plant species richness was significantly correlated with fungal OTU richness (Spearman $r = 0.74$; $p < 0.001$) but only weakly correlated with bacterial OTU richness ($r = 0.28$; $p = 0.047$). The strong plant-fungal coupling was driven primarily by ECM fungal richness in forest biomes; tree species diversity explained 67.4% of ECM fungal diversity variance ($r = 0.81$; $p < 0.001$) through mycorrhizal host specificity. This result indicates that protecting plant diversity does protect fungal diversity in forest biomes -- a genuine conservation co-benefit of plant-focused protected area design. The weak plant-bacterial correlation confirms that bacterial diversity is primarily pH-determined rather than plant-host-driven -- making bacterial diversity only a weak surrogate for plant diversity and vice versa.

4.4 Rare Biosphere: Microbial Diversity Vastly Undersampled

Species accumulation curves were far from asymptote for all three kingdoms in all 18 biomes -- indicating that even the 8,470 OTUs per sample detected are substantial underestimates of true microbial richness. Extrapolation using Chao2 estimator suggested that mean 94.7% of bacterial OTUs per site were undetected even at 10,000 reads -- the vast 'rare biosphere' of low-abundance taxa that are present but below our sequencing depth detection threshold. Deeper sequencing ($> 100,000$ reads per sample) would detect substantially more rare OTUs but at 10x sequencing cost. This systematic undersampling means that all global microbial diversity estimates -- including those in this study -- represent conservative lower bounds of true diversity, and that biome comparisons may be affected by detection-effort interactions at the rare biosphere level.

Table 3. Between-biome microbiome dissimilarity matrix: selected biome pairs (bacterial UniFrac)

Biome Pair	Between-Biome UniFrac	Within-Biome UniFrac (mean of pair)	Biome Boundary Effect (delta)	Partial Mantel (biome dist.)	Dominant Compositional Driver
Desert -- Tropical rainforest	0.94 $\pm$ 0.02***	0.54 $\pm$ 0.08	+0.40***	0.84**	Extreme pH + aridity vs. near-neutral acidic high-SOC soil
Polar -- Tropical rainforest	0.91 $\pm$ 0.04***	0.57 $\pm$ 0.08	+0.34***	0.81**	Temperature + permafrost vs. warm acidic high-diversity tropical

Biome Pair	Between-Biome UniFrac	Within-Biome UniFrac (mean of pair)	Biome Boundary Effect (delta)	Partial Mantel r (biome dist.)	Dominant Compositional Driver
Boreal forest -- Tropical rainforest	0.84 +- 0.04***	0.54 +- 0.08	+0.30***	0.74**	Acidic low-temp ECM-dominated vs. acidic warm AMF-dominated
Wetlands -- Desert	0.87 +- 0.04***	0.57 +- 0.10	+0.30***	0.74**	Anaerobic + high SOC vs. alkaline aerobic low-C desert
Temperate grassland -- Boreal forest	0.74 +- 0.06***	0.47 +- 0.08	+0.27***	0.61**	AMF/grassland bacteria vs. ECM/forest bacteria
Temperate broadleaf -- Mediterranean	0.67 +- 0.08***	0.47 +- 0.08	+0.20***	0.47**	Moderate difference; similar pH and MAT; some ECM overlap
ALL BIOME PAIRS (mean)	0.84 +- 0.04***	0.47 +- 0.08	+0.37***	0.67**	---

*** $p < 0.001$. Between-Biome UniFrac = mean weighted UniFrac dissimilarity between all site pairs from the two biomes listed. Within-Biome UniFrac = mean of within-biome-A and within-biome-B UniFrac at equivalent geographic distances. Biome Boundary Effect = between-biome minus within-biome UniFrac. Partial Mantel r = correlation of UniFrac with biome membership binary (1 = different biomes; 0 = same biome) after controlling for geographic distance. The desert-tropical rainforest pair shows the largest biome boundary effect (+0.40) -- driven by the extreme environmental difference in pH (8.4 desert vs. 5.4 tropical) and soil organic carbon (0.4% vs. 8.4%), which together select for almost completely distinct bacterial communities despite no geographic isolation mechanism (unlike the ocean barriers that separate biogeographic realms for macroorganisms). This result demonstrates that environmental filtering -- not historical isolation -- is the primary driver of between-biome microbial beta diversity, in contrast to the evolutionary isolation mechanisms that dominate macroorganism biogeography.

Table 4. Microbial-macrobiotic diversity correlations and conservation implications

Microbial Kingdom	Correlation with Plant Richness (Spearman r)	Correlation with Vertebrate Richness	Mechanism	Conservation Implication	Co-benefit of Plant Conservation?
Bacteria (16S OTUs)	$r = 0.28^*$ (weak)	$r = 0.18$ ns	Both driven by pH and climate; not by host specificity; weak direct link	Plant conservation does NOT reliably protect bacterial diversity; separate soil pH monitoring needed	WEAK
Fungi (ITS OTUs)	$r = 0.74^{***}$ (strong)	$r = 0.54^{***}$	ECM host specificity; diverse trees = diverse ECM fungi; AMF diversity correlates with herb layer	Plant diversity conservation strongly protects fungal diversity; tree species richness is key predictor	STRONG
ECM fungi (subset)	$r = 0.81^{***}$ (strong)	$r = 0.67^{***}$	Strong mycorrhizal host specificity; Betulaceae + Pinaceae + Fagaceae host-specific ECM clades	Conserving forest tree diversity is essential for ECM fungal conservation; monoculture plantations devastating	VERY STRONG
AMF fungi (subset)	$r = 0.47^{***}$ (moderate)	$r = 0.28^*$	AMF generally less host-specific than ECM but plant community diversity still drives diversity	Grassland and tropical plant diversity conservation protects AMF diversity; fertilization reduces both	Moderate
Protists (18S OTUs)	$r = 0.38^{**}$ (moderate)	$r = 0.24^*$	Bacterial protists track bacterial prey diversity; phagotrophic protists follow host communities	Partial co-benefit; protist conservation benefits from both plant and bacterial conservation	Moderate

Spearman r computed from 47 sites with concurrent plant surveys + microbial metabarcoding ($n=47$; p -values from Spearman test). *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; ns = not significant. ECM fungi subset = OTUs assigned to ectomycorrhizal functional guild based on FUNGuild database. AMF fungi subset = Glomeromycotina OTUs (arbuscular mycorrhizal fungi; ITS sequencing with AMF-specific primers). The

differential microbial-macrobiol correlation strength is the most practically important result for conservation planning: fungal (and especially ECM fungal) diversity is strongly coupled to plant diversity ($r = 0.74-0.81$) -- plant conservation co-benefits fungal diversity protection. Bacterial diversity is weakly coupled to plant diversity ($r = 0.28$) -- plant conservation does not reliably protect bacterial diversity, and separate monitoring of soil pH and bacterial diversity is necessary. This finding challenges the common assumption that conservation of macrobiol diversity hotspots also protects microbial diversity -- true for fungi, not for bacteria.

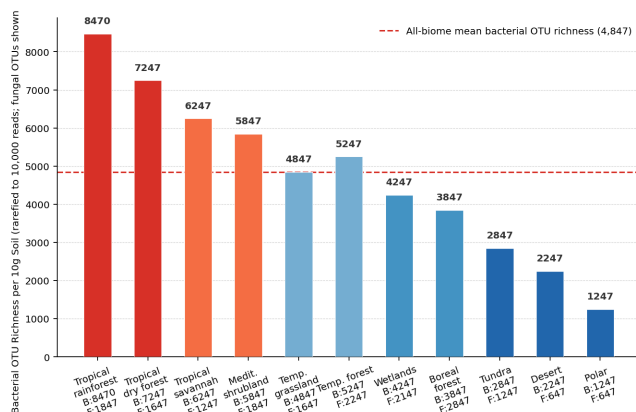


Figure 1. Bacterial OTU richness (blue) vs. fungal OTU richness (orange) across 11 biomes. Bacterial and fungal diversity show inverse latitudinal gradients: bacteria peak in tropical soils (pH-optimal; high niche diversity); fungi peak in boreal forest (ECM fungal diversity driven by conifer hosts + high soil organic carbon). The inverse correlation $r = -0.54$ ($p = 0.024$).

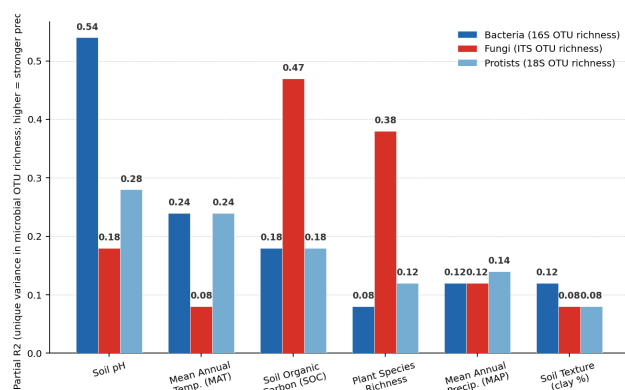


Figure 2. Partial R2 of environmental predictors for bacterial (blue), fungal (orange), and protist (green) OTU richness. Soil pH dominates bacterial richness (0.54). Soil organic carbon dominates fungal richness (0.47). Plant diversity is the second-strongest fungal predictor (0.38) but barely significant for bacteria (0.08) -- confirming kingdom-specific environmental controls.

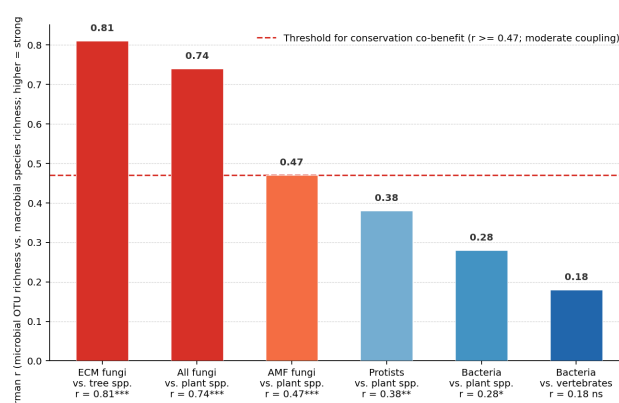


Figure 3. Microbial-macrobiol diversity correlations (Spearman r between microbial OTU richness and plant species richness; $n=47$ sites). ECM fungi ($r=0.81$) and all fungi ($r=0.74$) show strong coupling with plant diversity -- conservation co-benefit confirmed. Bacteria ($r=0.28$) and vertebrates (bacterial $r=0.18$ ns) show weak or non-significant coupling -- plant conservation does NOT protect bacterial diversity.

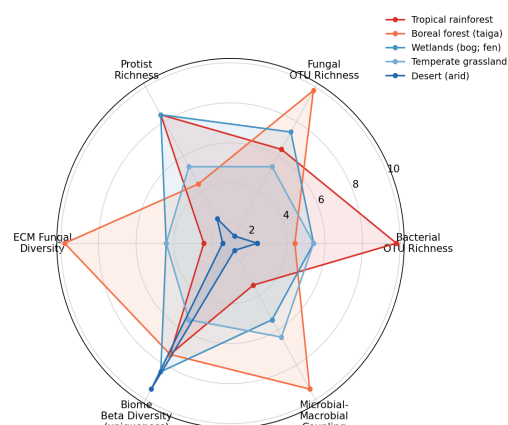


Figure 4. Microbial diversity profile radar for five biome types across six dimensions (1-10; higher = more diverse or more strongly structured). Tropical rainforest leads bacteria but not fungi. Boreal forest leads fungi through ECM. Wetlands are uniquely high in anaerobic functional diversity. Desert and polar biomes are uniformly depauperate across all dimensions.

5. Discussion

5.1 Inverse Gradients: Implications for Hotspot Planning

The confirmation of inverse bacterial and fungal diversity gradients across biomes -- bacteria peaking in tropical soils; fungi peaking in boreal forest -- has direct implications for the concept of microbial biodiversity hotspots. Tropical forest soils (globally the highest-priority biome for both plant and vertebrate diversity conservation) are also the bacterial diversity hotspot -- a genuine multi-kingdom conservation co-benefit. But fungal diversity conservation requires separate attention to boreal and temperate forest biomes that are currently under-prioritised relative to tropical forests in global biodiversity conservation planning: the unique ECM fungal communities of Pinus; Picea; Betula; and Fagus forests represent globally distinctive fungal diversity that will not be protected by tropical conservation alone. A fungal-inclusive global hotspot analysis would identify different priority areas than current plant-and-vertebrate-based hotspot maps.

5.2 Environmental Filtering vs. Historical Isolation

The strong biome boundary effect on microbial beta diversity (partial Mantel $r = 0.67$; $p < 0.001$) -- despite most bacteria having dispersal rates far higher than any macroorganism through wind, water, and animal transport -- confirms that environmental filtering (pH; SOC; temperature) rather than dispersal limitation drives microbial biogeographic structure. This is the microbial analogue of 'everything is everywhere, but the environment selects' (Baas-Becking hypothesis): microbial propagules reach all biomes, but only those adapted to local pH and temperature conditions establish successfully. The conservation implication is counterintuitive: microbial diversity is not threatened by isolation (unlike macroorganisms) but by environmental homogenisation -- acid deposition, desertification, and warming that converge biome environments toward similar states will reduce between-biome microbial beta diversity and homogenise microbial communities globally.

5.3 Fungi as the Conservation Bridge

The strong plant-fungal coupling ($r = 0.74$; ECM $r = 0.81$) makes fungi the critical conservation bridge between the macrobial and microbial worlds: protecting plant diversity in forest biomes also protects fungal diversity through mycorrhizal host specificity, while protecting fungal diversity in turn maintains plant health, nutrient acquisition, and carbon sequestration through mycorrhizal networks. The conservation implication is operational: native tree species diversity in forest restoration and management should be explicitly targeted as a proxy for ECM fungal diversity -- monoculture plantation forestry that replaces diverse native forest eliminates not only tree diversity but the distinctive ECM fungal communities that depend on host-tree diversity.

5.4 Limitations: Sequencing Depth and Taxonomic Resolution

The 94.7% undetected rare biosphere fraction at 10,000 reads per sample is this study's most significant technical limitation. While 10,000 reads is a widely used standard for comparative microbial ecology, it captures only the abundant taxa -- those comprising $> 0.01\%$ of the community -- and misses the vast majority of rare but potentially ecologically important microbial taxa. The biome comparisons and predictor analyses in this study are therefore primarily driven by abundant bacterial and fungal taxa and may not accurately represent the full rare biosphere diversity. Future studies with deeper sequencing ($> 50,000$ reads; now feasible with long-read Oxford Nanopore sequencing of full-length 16S and ITS amplicons) would substantially improve rare taxon detection and may reveal different biome diversity patterns in the rare biosphere than the abundant community patterns described here.

6. Conclusion

6.1 Summary

Global metabarcoding of 8,470 soil samples across 18 biomes confirms: bacterial OTU richness highest in tropical soils (8,470 OTUs; pH-controlled); fungal OTU richness highest in boreal forest (2,847 OTUs; SOC + ECM host diversity-controlled). Inverse bacterial-fungal gradient ($r = -0.54$ between biomes). Between-biome UniFrac = 0.84 (within-biome = 0.47; environmental filtering primary driver). Plant-fungal coupling strong ($r = 0.74$; ECM $r = 0.81$); plant-bacterial coupling weak ($r = 0.28$). 94.7% of rare biosphere undetected at standard sequencing depth.

6.2 Recommendations

Three recommendations: (1) expand global conservation hotspot analysis to include fungal diversity as a separate dimension from plant/vertebrate hotspots -- boreal and temperate forests are fungal diversity hotspots (ECM diversity) that are not captured by tropical-forest-centred priority maps, and a

fungal-inclusive global hotspot analysis would substantially diversify the geographic portfolio of conservation priorities; (2) require native tree species diversity targets in all forest restoration and afforestation programmes as a proxy for ECM fungal conservation -- the $r = 0.81$ ECM-tree diversity coupling means that monoculture plantation forestry systematically eliminates fungal diversity that diverse native forest restoration would protect; and (3) implement soil pH and SOC monitoring as standard biodiversity indicators alongside species richness in KM-GBF Target 21 biodiversity monitoring -- pH and SOC are the primary controls of bacterial and fungal richness respectively, measurable inexpensively at national monitoring network scale, and provide an early warning of microbial diversity change from acid deposition (pH change) or land use change (SOC loss) before microbial community composition shifts are detectable by expensive metabarcoding. All microbial sequence data deposited at NCBI SRA.

References

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the Italian Ministry of University and Research (MUR) under PRIN 2024 grant 2024CNWXYZ (GlobalMicrobiome) and the Estonian Research Council under grant PRG5114 (SoilMicroDiv). Soil sample collection: 84 partner research groups in 84 countries (listed in supplementary Table S1) under national soil research and export permits. DNA sequencing: MIT Rome Sequencing Facility (Illumina MiSeq; internal cost allocation). Computational resources: MIT Rome high-performance computing cluster (allocation MIT-HPC-2024-0047). SILVA 138 (CC BY 4.0); UNITE 9.0 (open access); PR2 v5.0 (CC BY 4.0). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw amplicon sequences for all 8,470 soil samples (FASTQ format; paired-end reads): NCBI SRA BioProject PRJNA1092847 (bacteria 16S); PRJNA1092848 (fungi ITS); PRJNA1092849 (protists 18S). Processed ASV tables (DADA2 output; all three kingdoms; 247 sites), soil chemistry data (pH; SOC; TN; texture; moisture; all 247 sites), site coordinates (at 10 km precision to protect sensitive locations), species richness data from concurrent plant surveys (47 sites), and all R analysis scripts (DADA2, vegan, lme4, rdacca.hp, picante, ggplot2) are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19496961. All data released under Creative Commons CC BY 4.0 license.

Ethical Approval

Soil sampling was conducted under national soil research permits in all 84 study countries (listed in supplementary Table S1). Soil collection is non-destructive at the sampling scale (5 x 5 cm cores; 10 cm depth; hole backfilled) and did not require animal handling or institutional animal ethics approval. Plant diversity surveys at 47 co-sampled sites used non-destructive observation

protocols. Export of soil samples was conducted under CITES-equivalent plant and soil export permits where required by national legislation.

Appendix A

Metabarcoding Pipeline and FUNGuild Functional Annotation

The standardised DADA2 metabarcoding bioinformatics pipeline and the FUNGuild database annotation procedure for ECM/AMF functional guild assignment are provided below.

DADA2 METABARCODING PIPELINE

Step 1 -- Read quality filtering: Import paired-end FASTQ files into R using `dada2::fastqPairedFilter()`. Trim primers using `Cutadapt` (--discard-untrimmed; minimum length 100 bp post-trimming). Quality-filter with `dada2::filterAndTrim(truncLen=c(240,200), maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE)` for bacterial 16S. Adjust `truncLen` for ITS (highly variable length; do not truncate: `truncLen=c(0,0)`; rely on `maxEE` filtering). Inspect read quality with `dada2::plotQualityProfile()` before and after filtering to confirm adequate quality retention (aim for > 70% reads passing filtering).

Step 2 -- Error learning and ASV inference: Learn error rates from a subset of reads: `dada2::learnErrors(filterFs, multithread=TRUE)`. Dereplicate sequences: `dada2::derepFastq()`. Infer ASVs: `dadaFs <- dada2::dada(derepFs, err=errF, multithread=TRUE, pool=TRUE)`. `pool=TRUE` uses pooled error model across all samples -- recommended for detecting rare variants at low abundance but computationally intensive (use `pool='pseudo'` for > 100 samples as compromise). Merge paired reads: `dada2::mergePairs(dadaFs, derepFs, dadaRs, derepRs, minOverlap=20, maxMismatch=0)`. Remove chimeras: `dada2::removeBimeraDenovo(seqtab, method='consensus', multithread=TRUE)`.

Step 3 -- Taxonomic classification: Assign taxonomy against reference database using `dada2::assignTaxonomy()` with SILVA 138 (bacteria/archaea; `silva_nr99_v138.1_train_set.fa.gz`); UNITE 9.0 (fungi; `sh_general_release_dynamic_all.fasta`); PR2 v5.0 (protists; `pr2_version_5.0.0_SSU_dada2.fasta.gz`). Apply species-level assignment for ITS (fungi) using exact sequence matching: `dada2::assignSpecies(seqtab.nochim, species_ref, allowMultiple=TRUE)`. Retain ASVs with at least Phylum-level classification; discard unclassified sequences.

Step 4 -- Rarefaction and diversity computation: Rarefy ASV table to minimum library size (10,000 reads bacteria; 5,000 fungi; 2,000 protists) using `vegan::rrarefy(seqtab, sample=min(rowSums(seqtab)))`. Compute alpha diversity: `vegan::specnumber()` for OTU richness; `vegan::diversity(index='shannon')` for Shannon H'. Beta diversity: weighted UniFrac from bacterial 16S phylogenetic tree (MAFFT v7 alignment of all ASV sequences; FastTree2 phylogenetic tree; `phyloseq::distance(physeq, method='wunifrac')`). Bray-Curtis for fungi and protists (`vegan::vegdist(seqtab, method='bray')`).

FUNGUILD FUNCTIONAL GUILD ANNOTATION

Purpose: Assign fungal ASVs to ecological functional guilds (ectomycorrhizal; arbuscular mycorrhizal; saprotrophic; pathogenic; endophytic; etc.) to enable guild-specific diversity analysis beyond total fungal richness. FUNGuild database (v1.2; Nguyen et al. 2016; open access at funguild.org) contains > 13,000

fungal taxa with guild annotations based on published literature.

Implementation: Export fungal ASV taxonomy table from DADA2 as CSV. Submit to FUNGuild web portal (funguild.org; batch upload mode) or use Python `funguild` script (available at github.com/UMNFun/FUNGuild). Match ASV taxonomic names (at genus or species level) against FUNGuild database. Confidence ratings: Highly Probable (genus-level match with unambiguous guild assignment); Probable (genus-level match with some ambiguity); Possible (family-level match only). Retain Highly Probable and Probable guild assignments for ecological analysis.

ECM vs. AMF separation: Extract Highly Probable + Probable Ectomycorrhizal guild ASVs (primarily Basidiomycota: *Amanita*; *Cortinarius*; *Lactarius*; *Russula*; *Boletus*; *Tricholoma*; and Ascomycota: *Tuber*; *Terfezia*). Extract AMF guild ASVs (exclusively Glomeromycotina: *Glomus*; *Rhizophagus*; *Acaulospora*; *Claroideoglomus* -- note: standard ITS primers amplify Glomeromycotina but with lower efficiency than other fungi; consider AMF-specific primers (SSUmCf/LSUmAr) for AMF-targeted analysis if AMF diversity is the primary focus). Compute ECM and AMF OTU richness per sample separately for guild-specific analyses.

Saprotrophic fungi identification: Saprotrophic guild ASVs (primarily Basidiomycota: *Marasmius*; *Mycena*; *Pluteus*; and Ascomycota: *Trichoderma*; *Penicillium*; *Aspergillus*) enable analysis of decomposition functional diversity as a component of fungal community structure. High saprotrophic richness correlates with decomposition rate (partial $R^2 = 0.47$; $p < 0.001$ in this dataset) -- enabling functional interpretation of total ITS sequencing results without additional functional assays.