

Global environmental drivers of soil microbiome structure: a systematic review and meta- analysis

Abhinav Yadav

dr.abhinav@bhu.ac.in

Banaras Hindu University

Neha Singh Chandel

ICAR-Indian institute of Pulses Research

Research Article

Keywords: soil microbiome, microbial diversity, meta-analysis, land use change, vegetation-soil interactions, ecosystem restoration, PRISMA

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10369231/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Global environmental drivers of soil microbiome structure: a systematic review and meta-analysis

Abhinav Yadav¹, Neha Singh Chandel²

¹CPCB

²ICAR-Indian institute of Pulses Research, Kanpur, UP, India

First and Corresponding Author: dr.abhinav@bhu.ac.in¹

Second Author Email: ns39577@gmail.com²

This manuscript presents a systematic review and meta-analysis evaluating the influence of soil physicochemical properties, vegetation, and land-use change on soil microbial community structure across terrestrial ecosystems.

Abstract

Background: Soil microbial communities regulate biogeochemical cycles and influence organic matter decomposition, nitrogen transformation, and greenhouse gas flux that facilitate ecosystem functioning. Despite three decades of research, no quantitative synthesis systematically evaluates the relative influence of environmental factors influencing microbiome structure across global biomes. The objective of this systematic review and meta-analysis aims to address that gap by quantifying relationships between physicochemical, vegetation, and land-use variables and soil microbiome outcomes across 100 peer-reviewed studies covering six continents.

Methods: We systematically searched 698 records from Web of Science and Scopus (1987-2025) following the PRISMA 2020 guidelines. After removal of duplicates and staged eligibility screening, 100 peer-reviewed studies met the inclusion criteria for qualitative synthesis; 78 provided adequate quantitative data for meta-analysis. The studies spanned forests, wetlands, grasslands, agricultural lands, deserts and tundra across six continents. We performed a weighted random-effects meta-analysis, subgroup analyses stratified by ecosystem type and geographic region, meta-regression on continuous moderators and publication-bias assessment (Egger regression, trim-and-fill) in R 4.3.2 (using the metafor, meta and dmetar packages). Inter-rater agreement was assessed using Cohen's κ .

Results: Soil pH was the most consistent and best predictor of bacterial community composition (pooled $r = 0.68$, 95% CI 0.59-0.76; $I^2 = 78\%$; $n = 56$ studies). Next was soil

organic carbon ($r = 0.54$, 95% CI 0.44-0.63), followed by total nitrogen ($r = 0.49$, 95% CI 0.38-0.59) and moisture ($r = 0.38$, 95% CI 0.26-0.49). Vegetation type explained 23-45% of community variation (median PERMANOVA $R^2 = 0.31$). Land-use intensification reduced Shannon diversity (Hedges' $g = -0.42$, 95% CI -0.56 to -0.28 ; $p < 0.001$), with larger effects in tropical systems ($g = -0.71$) and intensive croplands ($g = -0.62$). Alpha diversity recovered within 8 to 15 years, but community composition differed from reference ecosystems for an estimated 25 to 40 years. Agricultural management consistently reduced the complexity of predicted microbial co-occurrence networks.

Conclusions: Soil pH is a master variable that shapes bacterial communities on a global scale, and impact of plant characteristics and land-use history are context-dependent. Recovery of microbial community composition delayed significantly behind recovery of diversity following restoration, while functional interpretations were mostly based on predicted structural potentials rather than directly observed ecosystem process rates. Knowledge gaps such as the under-representation of tropics and lack of longitudinal data limit predictive capacity and define priorities for future research.

Keywords: soil microbiome; microbial diversity; meta-analysis; land use change; vegetation-soil interactions; ecosystem restoration; PRISMA;

1. Introduction

1.1 Global importance of soil microbial communities

Terrestrial soils contain the most diverse microbial assemblages on Earth, with estimates of 10^8 - 10^9 bacterial cells per gram of soil representing thousands of OTUs (Bardgett & van der Putten, 2014; Fierer, 2017). These communities mediate 90% of soil organic matter decomposition, regulate nitrogen cycling pathways from fixation to denitrification, and govern greenhouse gas fluxes of global consequence (Crowther *et al.*, 2019). Soil organic carbon stocks ($\sim 2,500$ Pg in the upper 3 m) are greater than the total carbon stored in the atmosphere and terrestrial vegetation, and their stability is tightly linked to microbial community structure and activity (Crowther *et al.*, 2019). Rapid global environmental change has heightened the need to understand soil microbiome dynamics. Intergovernmental Panel on Climate Change (IPCC) projections suggest that most terrestrial biomes will be affected by increased temperatures, altered precipitation, and increased drought frequency within decades (IPCC, 2021). At the same time, land use change is happening at unprecedented rates. Approximately 75% of the ice-free land surface of Earth is showing signs of human modification, with the intensification of agriculture being the primary driver (Ellis *et al.*, 2021). Such co-occurring pressures can destabilise soil microbial communities and the ecosystem services they mediate, with an estimated global economic value of USD 17 trillion yr⁻¹ (Costanza *et al.*, 2014).

1.2 Theoretical context

Soil microbial community assembly is governed by the interplay of deterministic processes (environmental filtering, competitive exclusion, facilitation) and stochastic processes (dispersal limitation, ecological drift) (Zhou & Ning, 2017). The relative importance of each mechanism varies across spatial scales, disturbance regimes and resource gradients. In environmentally heterogeneous or disturbed habitats, deterministic assembly prevails, whereas in homogeneous or resource-rich habitats, stochastic processes gain importance (Stegen *et al.*, 2012).

Soil pH has been repeatedly identified as a primary structuring variable among deterministic filters through physiological constraints on intracellular proton homeostasis, enzyme activity optima and nutrient speciation (Fierer & Jackson, 2006). Copiotrophs (Proteobacteria, Bacteroidetes) are abundant in carbon-rich environments with fast growth and high metabolic rates and oligotrophs (Acidobacteria, Verrucomicrobia) are abundant in carbon-limited environments by efficient substrate scavenging (Fierer *et al.*, 2007), with organic carbon availability dictating competitive outcomes across the copiotrophic-oligotrophic spectrum. Vegetation mediates these filters through organic matter inputs, root exudates and moisture modification by transpiration, creating plant-specific niches in the broader soil environment (Bulgarelli *et al.*, 2013; Bahram *et al.*, 2020).

1.3 State of knowledge and evidence gaps

Over the last three decades, research has greatly expanded our understanding of patterns of soil microbiomes within single ecosystems and across specific gradients. The type of mycorrhizal association (Arbuscular vs. ectomycorrhizal) is a major axis of variation in forest ecosystems, influencing the rates of nutrient cycling and the composition of bacterial communities (Bahram *et al.*, 2020). Wetland microbiomes are highly responsive to the hydrological regime and redox dynamics, and flooding-induced anaerobiosis structures communities independently of soil chemistry (Lipson *et al.*, 2015). The intensification of agriculture consistently reduces the diversity of soil microbial communities via the homogenization of the habitat, the narrowing of the substrate and chemical disturbance (Díaz-Vallejo *et al.*, 2021). Water and temperature stresses induce different microbiome dynamics in dryland and cryogenic systems, respectively (Neilson *et al.*, 2017). Though there have been major advances in individual studies, there are gaps in critical synthesis. First, there has not been a quantitative meta-analysis synthesizing microbiome-environment relationship across the entire range of terrestrial ecosystems from wetlands to deserts and tropics to arctic, thus preventing the identification of truly universal vs. context-dependent patterns. Second, existing reviews vary in geographic scope, methodological focus, and outcome measures, leading to conflicting narratives about the relative significance of drivers like pH, carbon, vegetation, and climate. Third, the magnitude and consistency of land-use change effects have not been quantified with systematic effect sizes and confidence intervals across biomes. Fourth, the timescales

and trajectories of microbiome recovery post-restoration are poorly constrained, hindering evidence-based management planning. Fifth, the heterogeneity among studies in the choice of sequencing platform, primers, bioinformatics pipeline, and diversity metrics is an underexplored source of variance that warrants explicit meta-analytical treatment.

1.4 Objectives and hypotheses

This systematic review and meta-analysis addressed five objectives: (O1) evaluate the relative influence of soil physicochemical properties and vegetation on soil microbial community structure; (O2) quantify the effects of land-use intensification on microbial diversity and composition using standardized effect sizes; (O3) identify ecosystem-specific microbiome–environment relationships through subgroup analyses; (O4) assess microbial recovery trajectories following ecosystem restoration; and (O5) develop a predictive framework linking environmental gradients with soil microbiome structure across biomes. We tested a four priori hypotheses: (H1) soil pH is the strongest predictor of bacterial community composition; (H2) deterministic assembly processes increase under greater environmental stress; (H3) land-use intensification reduces alpha diversity and network complexity while increasing biotic homogenisation; and (H4) alpha diversity recovers faster than community composition during restoration.

1.5 Novelty and contribution

This synthesis represents the first PRISMA-compliant quantitative meta-analysis of soil microbiome–environment relationships for all major terrestrial biomes. We go beyond narrative synthesis to produce actionable, statistically robust conclusions by combining standardized effect sizes across 100 studies and explicitly modelling methodological heterogeneity. The resulting predictive framework and identified research gaps provide concrete guidance for future investigations in microbial ecology, ecosystem management and climate change science.

2. Methodology

2.1 Protocol

This systematic review follows PRISMA 2020 guidelines (Page *et al.*, 2021). Deviations from the registered protocol, if any, are documented in Supplementary File S1. To ensure maximum reporting quality and methodological reproducibility, this systematic review and meta-analysis was structured in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Ecology and Evolutionary Biology (PRISMA-EcoEvo) guidelines. All stages of literature searching, screening, eligibility assessment, data extraction, and risk-of-bias evaluation using the Newcastle–Ottawa

Scale were systematically documented to ensure methodological transparency, reproducibility, and reporting consistency across the review process.

2.2 Search strategy

Comprehensive searches were conducted across Web of Science Core Collection, Scopus database inception through March 2025. The Web of Science export formed the primary database for this analysis, supplemented by additional searches.

2.2.1 Primary search string (Web of Science):

TS = (("soil microbiome" OR "soil microbiota" OR "soil microbial communit" OR "bacterial communit" OR "fungal communit" OR "prokaryotic communit") AND ("vegetation" OR "land use" OR "ecosystem" OR "pH" OR "organic carbon" OR "nitrogen" OR "environmental gradient") AND ("16S rRNA" OR "ITS" OR "metagenom" OR "amplicon sequencing" OR "high-throughput sequencing")).

Additional records were identified through backward citation searching of included studies and forward citation tracking using Google Scholar.

2.3 Eligibility criteria

2.3.1 Inclusion criteria:

1. Primary research using culture-independent molecular methods (16S rRNA gene sequencing, ITS sequencing, or metagenomics).
2. Investigation of soil/sediment microbial community composition in terrestrial or transitional (wetland/riparian) ecosystems.
3. Reporting of measurable relationships between microbial community characteristics and environmental variables.
4. Sufficient quantitative data for effect size calculation or meaningful qualitative synthesis ($n \geq 5$ per group).
5. Peer-reviewed English-language publications.

2.3.2 Exclusion criteria:

1. Culture-dependent methods only
2. Purely aquatic systems without soil/sediment interface
3. Laboratory microcosm studies without field validation or reference
4. Conference abstracts, book chapters, preprints
5. Absence of environmental characterization
6. Duplicate publications (most comprehensive version retained)

2.4 Study selection

2.4.1 Phase 1 (Title/Abstract screening): Two reviewers independently screened all records. Disagreements were resolved through discussion; unresolved cases were adjudicated by a third reviewer.

2.4.2 Phase 2 (Full-Text assessment): Full texts were retrieved for all potentially eligible records. Exclusion reasons were recorded at the study level.

2.4.3 Phase 3 (Data extraction): Eligible studies were screened for data availability; authors were contacted when essential data were missing from publications.

Inter-rater reliability was assessed using Cohen's κ ; values >0.80 were considered adequate.

2.5 Data extraction

A standardized extraction form (Supplementary File S2) captured:

Study-level variables: Publication metadata; geographic location (latitude/longitude, country, biome); ecosystem type; study design; sample size.

Environmental variables: Soil pH; organic carbon (g kg^{-1}); total nitrogen (g kg^{-1}); C:N ratio; moisture content (%); texture; electrical conductivity; vegetation type and cover; dominant species; mycorrhizal association type; mean annual temperature ($^{\circ}\text{C}$); mean annual precipitation (mm); aridity index; land-use history; management practices.

Microbiome outcomes: Sequencing platform; target gene/region; diversity metrics (Shannon H' , observed OTUs/ASVs, Chao1, Faith's PD, Simpson D); community dissimilarity metric; dominant phyla relative abundance; functional predictions where available; network properties (nodes, links, modularity, average degree).

Statistical outcomes: Correlation coefficients (r or ρ); standardized mean differences; PERMANOVA R^2 values; regression coefficients; variance explained.

Two reviewers independently extracted data; discrepancies were resolved by consensus.

2.6 Risk of bias assessment

Each study was assessed across five domains adapted from the Newcastle-Ottawa Scale for observational studies:

1. Representativeness of sampling (geographic, temporal, and habitat coverage)
2. Adequacy of sample size (statistical power considerations)
3. Standardization of molecular methods (validated protocols, quality control)
4. Control of confounding variables (statistical adjustment, study design)

5. Completeness of outcome reporting (all pre-specified outcomes reported)

Each domain was rated low (0), moderate (1), or high (2) risk. Total scores (0-10) classified studies as low risk (0-3), moderate risk (4-6), or high risk (7-10).

2.7 Quantitative synthesis

2.7.1 Effect size metrics

For continuous relationships between environmental predictors and microbial diversity metrics, we extracted or calculated Pearson's r or Spearman's ρ . Correlation coefficients were Fisher's z -transformed prior to meta-analysis and back-transformed for reporting.

For group comparisons (e.g., land-use contrasts), standardized mean differences were calculated using Hedges' g :

$$g = J \times \frac{\bar{X}_1 - \bar{X}_2}{S_p}$$

where J is the small-sample correction factor and S_p is the pooled standard deviation.

For continuous relationships between environmental drivers and microbial community structure, correlation coefficients (Pearson's r or Spearman's ρ) derived from environmental vector fitting onto ordination space (e.g., NMDS and PCoA) or from dominant ordination axis scores were extracted. To stabilize variance across heterogeneous datasets, all correlation coefficients were transformed using Fisher's z -transformation prior to meta-analytical pooling and subsequently back-transformed for interpretation and reporting.

We explicitly acknowledge that distance matrices possess inherent spatial and row-column interdependencies, meaning that direct application of conventional sampling variance estimators such as:

$$v_i = 1/(n_i - 3)$$

to Mantel-derived correlations may underestimate true variance, inflate effective degrees of freedom, and increase Type I error rates in secondary syntheses [6]. Because raw dissimilarity matrices and pairwise distance structures are rarely reported in primary studies, coordinate-based hierarchical or nested variance modeling approaches were not feasible at the global scale considered here. Accordingly, the standard random-effects framework was treated as a pragmatic and robust baseline analytical model, while the statistical limitations associated with matrix-based correlations were explicitly acknowledged and critically discussed within the methodological considerations section.

2.7.2 Meta-analytical model

We employed random-effects models using restricted maximum likelihood (REML) estimation:

$$\hat{\theta}_i = \mu + u_i + \varepsilon_i$$

$$u_i \sim \mathcal{N}(0, \tau^2)$$

$$\varepsilon_i \sim \mathcal{N}(0, v_i)$$

where μ is the overall mean effect, τ^2 is between-study variance, and v_i is within-study sampling variance.

Heterogeneity was quantified using:

- Q statistic (chi-squared test for presence of heterogeneity)
- I^2 statistic (proportion of total variance due to heterogeneity):
 $I^2 < 25\%$ = low, $25-75\%$ = moderate, $> 75\%$ = high (Higgins et al., 2003)
- τ^2 (between-study variance)
- 95% prediction interval (range of true effects expected in similar future studies)

2.7.3 Subgroup and moderator analyses

Pre-specified subgroup analyses were conducted by:

- Ecosystem type (forest, grassland, wetland, agricultural, desert/dryland, tundra/permafrost)
- Geographic region (Asia, North America, Europe, South America, Oceania, Africa)
- Climate zone (tropical, subtropical, temperate, boreal, arctic)
- Study design (observational vs. experimental)
- Sequencing platform (Illumina vs. other)

Meta-regression examined continuous moderators: sample size, publication year, mean annual temperature (MAT), mean annual precipitation (MAP), and mean soil organic carbon.

Significance of subgroup differences was assessed using the Q between (Q_B) statistic.

2.7.4 Publication bias

Publication bias was evaluated through:

- Funnel plot visualization
- Egger's weighted regression test $p < 0.10$ considered indicative of asymmetry
- Begg and Mazumdar rank correlation
- Trim-and-fill method (Duval & Tweedie, 2000) for adjusted effect size estimation

2.7.5 Sensitivity analyses

Sensitivity analyses included:

- Leave-one-out analysis
- Restriction to low risk-of-bias studies
 - o "Score"=0—" 3
- Alternative variance estimators:
 - o "DerSimonian–Laird", "Paule–Mandel"
- Analysis restricted to Illumina studies only
- Analysis excluding studies with:
 - o $n < 20$ samples

2.8 Software

All analyses were conducted in R v.4.3.2 (R Core Team, 2023) using the following packages: metafor v.4.2-0 (Viechtbauer, 2010), meta v.6.5-0, dmetar v.0.1.0, ggplot2 v.3.4.4, vegan v.2.6-4, and igraph v.1.5.1. PRISMA flow diagrams were generated using PRISMA2020.

3. Results

3.1 Study selection and characteristics

[Figure 1: PRISMA Flow Diagram]

The initial search retrieved 698 records across databases, yielding 654 after duplicate removal. Title and abstract screening excluded 412 records, leaving 242 for full-text assessment. Of these, 142 were excluded (detailed reasons in Table S1), yielding 100 studies for qualitative synthesis. Of the 100 included studies, 78 provided sufficient quantitative data for meta-analysis. Inter-rater agreement was high ($\kappa = 0.87$, 95% CI 0.82-0.92), indicating reliable study selection.

[Table 1: Study Characteristics]

Included studies spanned six continents (Asia: 34%; North America: 26%; Europe: 21%; South America: 11%; Oceania: 5%; Africa: 3%) and six primary ecosystem types (forest: 32%; wetland/riparian: 24%; grassland: 18%; agricultural: 14%; desert/dryland: 8%;

tundra/permafrost: 4%). Publication years ranged from 2014 to 2025, with marked acceleration post-2019 corresponding to widespread adoption of high-throughput sequencing. Illumina platforms dominated (87% of studies); 16S rRNA V3-V4 was the most common target region (62%), followed by fungal ITS1/ITS2 (34%). Sample sizes per study ranged from 6 to 166 (median = 36, IQR = 18-54). Risk of bias was low in 52%, moderate in 38%, and high in 10% of studies (Table S2).

3.2 Environmental drivers of bacterial community composition

[Figure 2: Forest Plot of pH-Community Correlations]

3.2.1 Soil pH

Meta-analysis of studies reporting standardized multivariate community-composition metrics demonstrated that soil pH was the strongest and most consistent environmental predictor of bacterial community structure across ecosystems. PERMANOVA-derived variance-explained estimates consistently identified pH as a dominant explanatory variable across forests, grasslands, wetlands, agricultural soils, and drylands. Across studies, pH effects remained strongly positive despite substantial ecological heterogeneity among biomes (Figure 2).

Ecosystem-specific pH effects differed significantly ($Q_B = 18.4$, $df = 4$, $p = 0.001$). Forests showed the strongest effect ($r = 0.74$, 95% CI 0.64–0.82; $n = 18$), followed by grasslands ($r = 0.71$, 95% CI 0.58–0.81; $n = 12$), deserts ($r = 0.58$, 95% CI 0.38–0.74; $n = 6$), agricultural soils ($r = 0.62$, 95% CI 0.46–0.75; $n = 10$), and wetlands ($r = 0.51$, 95% CI 0.35–0.64; $n = 14$). The attenuated wetland effect likely reflects the dominant role of redox gradients in saturated soils.

Shannon alpha diversity showed a moderate positive correlation with pH across acidic-to-neutral soils ($r = 0.38$, 95% CI 0.26–0.49; $I^2 = 65.2\%$; $n = 42$ studies), consistent with the physiological basis of the pH-diversity relationship.

3.2.2 Soil organic carbon

[Table 2: Meta-Analysis Summary Statistics]

Soil organic carbon (SOC) was the second-ranked predictor of community composition ($r = 0.54$, 95% CI 0.44–0.63; $I^2 = 70.8\%$; $n = 48$; Table 2). The SOC effect was strongest in carbon-limited systems: deserts/drylands ($r = 0.72$, 95% CI 0.54–0.84; $n = 6$) and degraded agricultural soils ($r = 0.67$, 95% CI 0.51–0.79; $n = 8$). In carbon-rich forest soils, the relationship was weaker ($r = 0.41$, 95% CI 0.24–0.55; $n = 16$), supporting resource-ratio theory predictions that carbon quality rather than quantity becomes the primary constraint when carbon is abundant.

Total nitrogen similarly correlated with community composition ($r = 0.49$, 95% CI 0.38-0.59; $I^2 = 68.3\%$; $n = 44$), substantially collinear with SOC (weighted r between SOC and TN across studies = 0.81). The C:N ratio showed weaker and more variable relationships ($r = 0.31$, 95% CI 0.17-0.44; $I^2 = 72.1\%$; $n = 32$) (Table 2).

3.2.3 Soil moisture

Soil moisture effects were ecosystem-dependent and nonlinear. In wetland and riparian systems, moisture/flooding regime explained 35-52% of community variation (weighted $R^2 = 0.43$, 95% CI 0.31-0.55; $n = 18$), exceeding soil chemistry effects in these systems. In upland systems, moisture effects were weaker ($r = 0.29$, 95% CI 0.15-0.42; $n = 28$) and often confounded with vegetation patterns. Desert/dryland ecosystems showed strong moisture-diversity relationships ($r = 0.61$, 95% CI 0.42-0.76; $n = 6$), consistent with water limitation governing microbial abundance in arid environments (Neilson *et al.*, 2017).

Overall pooled moisture-community correlation was $r = 0.38$ (95% CI 0.26-0.49; $I^2 = 74.2\%$; $n = 46$ studies), ranking fourth after pH, SOC, and TN.

3.3 Vegetation effects on soil microbiome structure

[Figure 3: Bubble Plot of Vegetation Explanatory Power]

3.3.1 Vegetation type

Vegetation type explained 23-45% of soil microbial community variation in individual studies (median PERMANOVA $R^2 = 0.31$, IQR = 0.26-0.38; $n = 62$ studies). Effect magnitudes were significantly heterogeneous across ecosystem types ($Q_B = 24.2$, $df = 5$, $p < 0.001$). Forest ecosystems showed the largest vegetation effects (median $R^2 = 0.38$), particularly when contrasting ectomycorrhizal versus arbuscular mycorrhizal-dominated stands (Bahram *et al.*, 2020). Grasslands and agricultural systems showed somewhat weaker effects (median $R^2 = 0.27$), while desert systems showed the weakest vegetation effects (median $R^2 = 0.19$), consistent with climate-driven assembly dominating in these systems (Figure 3).

Meta-regression indicated that vegetation explanatory power increased with plant species richness ($\beta = 0.003$ per additional species, 95% CI 0.001–0.005; $p = 0.004$) and with mean annual precipitation ($\beta = 0.0002$ per mm, 95% CI 0.0001–0.0003; $p < 0.001$), suggesting that vegetation effects on microbiomes are amplified in productive, species-rich systems.

3.3.2 Rhizosphere vs. bulk soil

Thirty-four studies directly compared rhizosphere and bulk soil communities. Community composition differed consistently between compartments (weighted dissimilarity effect: $d = 1.24$, 95% CI 1.02–1.46; $I^2 = 68.7\%$; $p < 0.001$), with the rhizosphere harboring distinct assemblages enriched in copiotrophic taxa responsive to root exudates. Alpha diversity was marginally lower in rhizosphere than bulk soil in some systems ($d = -0.18$, 95% CI -0.34 to -0.02; $p = 0.03$; $I^2 = 72.1\%$), potentially reflecting selective enrichment under competitive root-zone conditions.

3.4 Land-use change impacts on microbial diversity

[Figure 4: Forest Plot of Land-Use Effects on Shannon Diversity - INSERT HERE]

3.4.1 Alpha diversity

[Table 3: Subgroup Analysis Results]

Comparing natural or semi-natural ecosystems to intensively managed agricultural systems, Shannon diversity was significantly reduced (Hedges' $g = -0.42$, 95% CI -0.56 to -0.28; $I^2 = 72.3\%$; $Q = 134.6$, $df = 37$; $p < 0.001$; $n = 38$ comparisons; Figure 4). Observed species richness showed a comparable decline ($g = -0.48$, 95% CI -0.65 to -0.31), while phylogenetic diversity was similarly affected ($g = -0.39$, 95% CI -0.58 to -0.20) (Figure 4).

Effects were significantly modulated by land-use intensity ($Q_B = 9.6$, $df = 2$; $p = 0.008$): intensive croplands showed the largest diversity reduction ($g = -0.62$, 95% CI -0.82 to -0.42), followed by extensive agriculture ($g = -0.38$, 95% CI -0.58 to -0.18) and extensive grazing ($g = -0.24$, 95% CI -0.42 to -0.06). Geographic effects were also significant ($Q_B = 12.8$, $df = 2$; $p = 0.002$): tropical systems showed the largest effects ($g = -0.71$, 95% CI -0.92 to -0.50) compared to temperate regions ($g = -0.38$, 95% CI -0.54 to -0.22). The 95% prediction interval for the overall effect (-0.82 to -0.02) indicated consistently negative effects expected in future similar studies.

3.4.2 Community composition and phylum-level changes

Land-use conversion consistently increased relative abundance of copiotrophic phyla (Proteobacteria, Bacteroidetes, Firmicutes) while reducing oligotrophic groups. Acidobacteria relative abundance declined significantly with agricultural intensification ($d = -0.56$, 95% CI -0.74 to -0.38; $I^2 = 61.4\%$; $n = 28$ studies). Conversely, Proteobacteria relative abundance increased ($d = +0.43$, 95% CI +0.28 to +0.58; $n = 24$ studies), consistent with copiotrophic enrichment under elevated nutrient inputs from fertilization.

Network analysis results from 18 studies consistently showed reduced complexity in agricultural versus natural soils: number of nodes decreased by a mean of 23% (weighted), network connectivity declined by 31%, and modularity decreased, indicating

simplified microbial interaction structures under intensive management (Byers *et al.*, 2024).

3.5 Ecosystem restoration trajectories

[Figure 5: Meta-Regression of Diversity Recovery vs. Restoration Age]

3.5.1 Diversity recovery

Twenty-two studies tracked microbial community metrics across restoration chronosequences spanning 1-60 years. Meta-regression revealed a significant positive relationship between restoration age and alpha diversity recovery ($\beta = +0.024$ Shannon units per year, 95% CI 0.012-0.036; $p < 0.001$; Figure 5). Studies with restoration periods exceeding 15 years showed diversity levels within 85% of reference ecosystems (Hedges' g vs. reference = -0.18, 95% CI -0.38 to 0.02; $p = 0.08$), while shorter restoration periods showed larger deficits (Figure 5).

Extrapolating the meta-regression slope with 95% prediction intervals suggests full diversity recovery requires approximately 25-40 years, depending on ecosystem type and restoration approach. Active restoration (vegetation planting, soil amendments) was associated with faster recovery than passive abandonment (interaction: $\beta_{\text{active}} - \beta_{\text{passive}} = +0.009$ units yr^{-1} ; $p = 0.02$).

3.5.2 Compositional recovery

Community composition recovery lagged substantially behind diversity recovery. Among studies with restoration periods > 20 years, PERMANOVA analyses reported significant compositional differences between restored and reference ecosystems in 16 of 18 studies (median R^2 for restoration vs. reference contrast = 0.22, range 0.15–0.28). Persistent compositional differences concentrated in slow-growing oligotrophic taxa (Acidobacteria subgroups, Verrucomicrobia) and symbiotic fungi (ectomycorrhizal guilds), consistent with dispersal limitation and slow growth rate constraints on recolonization.

3.6 Climate and geographic patterns

Mean annual temperature showed weak overall correlation with bacterial diversity ($r = 0.12$, 95% CI -0.02 to 0.26 ; $p = 0.09$; $n = 34$) but significantly predicted community composition ($r = 0.38$, 95% CI 0.24 - 0.51 ; $n = 36$). Mean annual precipitation correlated more strongly with diversity in water-limited systems (deserts/semi-arid grasslands: $r = 0.52$, 95% CI 0.33 - 0.68 ; $n = 12$). Elevational studies ($n = 8$) showed heterogeneous patterns: monotonic decline in some mountain systems (Zhang *et al.*, 2025) and hump-shaped relationships in others, with no consistent meta-analytical signal (pooled $r = 0.06$, 95% CI -0.18 to 0.29 ; $p = 0.60$).

3.7 Functional patterns

3.7.1 Predicted metabolic functions

Thirty-two studies employing predictive functional profiling approaches (PICRUSt, Tax4Fun, and FAPROTAX) consistently identified chemoheterotrophy as the dominant predicted functional category across ecosystems. Nitrogen fixation potential was more frequently predicted in nitrogen-limited systems, whereas nitrification-associated pathways increased under agricultural fertilization regimes.

Because most included studies relied on 16S rRNA-based inference approaches rather than direct metagenomic sequencing, these findings should be interpreted as predicted functional potentials rather than direct measurements of ecosystem process rates. Only 18 studies employed shotgun metagenomics, limiting definitive inference regarding active metabolic functioning.

Metagenomic datasets nevertheless suggested that predicted carbon degradation potential was positively associated with soil organic carbon quantity and quality ($r=0.48$, 95% CI 0.29–0.64; $n=14$). Genes associated with degradation of relatively recalcitrant carbon substrates were more frequently detected in undisturbed forest soils compared with intensively managed agricultural systems.

3.8 Heterogeneity and publication bias

[Figure 6: Funnel Plot with Trim-and-Fill]

4. Discussion

4.1 Soil pH as a Universal Structuring Variable

The robust, cross-ecosystem pH-community relationship ($r = 0.68$) unambiguously supports the primacy of pH as a soil microbiome structuring variable, consistent with the foundational hypothesis of Noah Fierer and Robert B. Jackson (2006). The mechanistic basis is well understood: bacterial intracellular pH homeostasis operates within tight bounds (~ 7.5 pH units), requiring energetically costly proton pumping in strongly acidic or alkaline environments that limits colonization by most taxa. Additionally, pH controls the speciation and bioavailability of nutrients (particularly phosphorus, iron, and aluminum), further constraining ecological niches along pH gradients.

However, our ecosystem-stratified analysis reveals important contextual variations. The attenuated pH effect in wetlands ($r = 0.51$) compared to forests ($r = 0.74$) and grasslands ($r = 0.71$) reflects the ecological dominance of hydrological regimes and redox dynamics over bulk proton concentration in water-saturated soils. Flooding creates steep oxygen

gradients that determine the thermodynamic basis of microbial metabolism far more than proton concentration, partitioning communities into aerobic heterotrophs, nitrate reducers, iron reducers, sulfate reducers, and methanogens across vertical depth and flooding horizons. This context-dependency has practical implications: pH amendment as a management tool for microbiome manipulation will be most effective in upland terrestrial systems and least effective in wetlands where hydrological management takes precedence. The breadth of the pH-diversity prediction interval ($r = 0.34\text{--}0.88$) further underscores ecosystem specificity. While the mean effect is strong, the wide interval indicates that in particular ecosystems or at particular spatial scales, pH may explain relatively little variation. Studies examining pH effects within single ecosystems at fine spatial resolution (e.g., forest soil across a 100 m transect) may encounter much weaker relationships than those spanning broad pH gradients across biomes.

4.2 Organic Carbon as a Resource Constraint

The strong SOC-community relationship in carbon-limited systems (deserts: $r = 0.72$; degraded agricultural soils: $r = 0.67$) supports resource-ratio theory: when carbon is the primary limiting resource, its availability becomes the dominant axis of competitive and deterministic filtering. Conversely, the weaker SOC effect in carbon-rich forest soils ($r = 0.41$) suggests that in these organic-rich systems, carbon quantity is no longer the primary limiting factor. Instead, community assembly is structured by carbon quality—such as substrate lability, molecular complexity, and specific stoichiometric ratios (such as C:N ratio; $r = 0.31$)—rather than gross organic carbon stocks.

This distinction has implications for interpreting land-use change effects: conversion of organic-rich natural soils to agriculture simultaneously reduces SOC quantity and alters SOC quality, generating community shifts through multiple pathways simultaneously. The collinearity between SOC and TN (cross-study $r = 0.81$) necessitates caution in interpreting independent effects of these variables. Path analysis in a subset of studies with sufficient multivariate data ($n = 18$) suggested that pH acts primarily as a direct filter on community composition, while SOC and TN exert effects partly mediated through pH and partly through direct resource effects. This mediation network complicates simple pairwise interpretations but provides a more realistic representation of the mechanistic architecture of soil environment-microbiome relationships.

4.3 Vegetation as Integrative Driver

The substantial vegetation explanatory power (23–45% of community variation) reflects vegetation's role as an integrative environmental modifier rather than a simple categorical variable. Vegetation shapes soil microbiomes through multiple simultaneous pathways: direct organic inputs (litter quantity and quality, root exudate composition), physical effects on soil structure and moisture, pH modification through litter chemistry, and direct selection of root-associated microbes through the rhizosphere effect.

The elevated median PERMANOVA R^2 of 0.38 in forest ecosystems highlights the intense coupling between dominant trees and belowground communities. Forest soils represent highly complex, vertically stratified carbon sinks where soil organic matter forms simultaneously from aboveground litter decomposition residuals and root-derived rhizodeposition. Trees create highly heterogeneous microhabitats across rhizosphere zones, litter horizons, and deadwood-associated substrates.

Furthermore, vertical soil stratification (0–10 cm versus 10–40 cm horizons) and stand-level canopy closure function as deterministic ecological filters. Stable forest canopies maintain microclimatic and substrate stability, favoring oligotrophic phyla such as Acidobacteriota that thrive in resource-limited acidic topsoils. Conversely, canopy removal and clear-cutting disrupt vertical stratification, causing rapid qualitative changes in substrate utilization and declines in deeper soil functional stability.

4.4 Mechanisms of Land-Use Impacts

The consistent negative effect of land-use intensification on alpha diversity ($g = -0.42$) across 38 comparisons strengthens and extends previous regional syntheses. Our global synthesis demonstrates that these patterns generalize across climate zones and soil types, though with significant variation in magnitude. The amplified diversity loss in tropical regions ($g = -0.71$ versus -0.38 in temperate regions) may reflect higher baseline diversity, greater climatic instability, and lower structural buffering capacity among specialized tropical taxa.

Physical mechanisms deserve particular emphasis. Tillage disrupts soil pore networks and aggregate structures that create diverse microbial microhabitats, thereby reducing beta diversity within agricultural fields. Simultaneously, tillage exposes protected organic matter to rapid decomposition, shifting carbon pools toward fast-cycling labile fractions that favor copiotrophic taxa (Proteobacteria and Firmicutes) at the expense of oligotrophic groups (Acidobacteria, Verrucomicrobia, and Planctomycetes).

The 23% decrease in network nodes and 31% decline in connectivity under agricultural management suggest that these compositional changes extend to the functional architecture of soil microbial communities. Simplified co-occurrence networks with fewer highly connected taxa may exhibit reduced resilience to repeated cropping disturbances.

Importantly, conventional agricultural systems can transition toward more sustainable alternatives. Long-term field experiments demonstrate that no-tillage pasture-crop rotations preserve soil bulk density and aggregate stability significantly better than continuous monocropping. Biologically, these rotations alter rhizosphere chemistry and recruit potentially beneficial bacterial genera, including *Bosea*, *Devosia*, and *Microbacterium*. These microbial recruitments are associated with enhanced nitrogen uptake, improved biomass accumulation, and greater climate resilience.

4.5 Restoration Timescales: Diversity vs. Composition

The divergent recovery trajectories of diversity (8–15 years) versus composition (25–40 years) represent one of the most practically important findings of this synthesis. Systems may appear functionally recovered based on diversity indices while retaining fundamentally different microbial compositions and ecological functions.

Several mechanisms likely explain this compositional persistence:

Dispersal limitation: Slowly dispersing taxa, particularly obligate symbionts and rare microbial specialists, may remain absent from restoration sites for decades.

Priority effects and alternative stable states: Early colonizers may modify environmental conditions in ways that inhibit later establishment of reference taxa.

Pedological legacy effects: Soil compaction, altered pH, depleted organic matter, and erosion-driven texture changes may persist long after vegetation recovery.

Biotic interaction loss: Missing taxa may remove critical interaction partners required for successful re-establishment of dependent microbial groups.

The necessity of active restoration becomes especially evident in severely disturbed systems such as mining tailings. These environments are characterized by extreme salinity, heavy metal toxicity, and severe organic matter depletion. Passive abandonment often fails entirely due to dispersal barriers and persistent toxicity.

Our findings demonstrate that active restoration significantly accelerates microbial recovery rates. Combined interventions using native vegetation, organic amendments, and physical stabilization improve water-holding capacity, nutrient availability, and

organic carbon accumulation while promoting the establishment of stress-tolerant rhizosphere microbiota capable of accelerating ecosystem recovery.

4.6 Functional Consequences

Although this synthesis focused primarily on taxonomic responses, the observed compositional shifts likely correspond to changes in ecosystem functional potential. However, caution is warranted because most included studies relied on predictive marker-gene approaches (e.g., PICRUSt and Tax4Fun) rather than direct metagenomic or transcriptomic measurements.

The observed transition from oligotrophic-dominated communities toward copiotrophic assemblages under land-use intensification may indicate altered carbon-processing dynamics and reduced microbial network complexity. Nevertheless, these findings should not be interpreted as direct evidence of altered in situ process rates or nutrient cycling efficiencies. More definitive conclusions require integration of shotgun metagenomics, metatranscriptomics, and direct biogeochemical measurements.

4.7 Methodological Considerations

4.7.1 Statistical Modeling and Concordance Tests

We utilized standard random-effects models with REML estimation to pool correlation coefficients across diverse datasets. While hierarchical three-level nested models are increasingly recommended in synthesis ecology, our objective was to present a broadly comparable global synthesis.

We further acknowledge that pooling Mantel correlations and ordination-derived vector correlations assumes simplified variance structures because distance matrices inherently contain row-column interdependencies. Classical Mantel tests are known to exhibit inflated Type I error rates and reduced statistical power under spatially structured datasets. Because raw dissimilarity matrices are rarely available in published studies, more advanced coordinate-based nested modeling approaches were not feasible at the global scale. Future studies should publish raw distance matrices to facilitate improved multivariate association frameworks such as GRV and PROTEST analyses.

4.7.2 Sequencing and Functional Predictions

Studies differed substantially in primer sets, sequencing depth, normalization procedures, and bioinformatic pipelines, producing incomparable absolute diversity estimates. We minimized these inconsistencies by emphasizing relative within-study comparisons and standardized effect sizes, although methodological heterogeneity undoubtedly contributed residual variance.

Importantly, most studies inferred microbial function using 16S rRNA-based prediction frameworks rather than direct genomic or transcriptomic sequencing. These methods assume conserved genotype-phenotype relationships within taxa, an assumption that frequently fails in highly diverse soil microbial systems. Consequently, all inferred metabolic pathways should be interpreted as predicted functional potentials rather than direct measures of active ecosystem functioning.

4.7.3 Temporal and Seasonal Sampling

Cross-sectional studies dominated the dataset, limiting inference regarding temporal dynamics. The few available longitudinal studies indicate strong seasonal turnover in community composition, suggesting that cross-sectional syntheses may systematically underestimate environmentally driven temporal variability.

4.8 Comparison with Previous Syntheses

Our pooled pH-community correlation ($r = 0.68$) aligns closely with earlier foundational studies, although the observed relationship is somewhat stronger than the canonical estimates reported in early continental-scale analyses. This may reflect improved sequencing technologies, standardized sampling approaches, and greater ecological coverage in recent studies.

Similarly, our land-use intensification effect size ($g = -0.42$) closely matches previous European and tropical syntheses, supporting the generality of microbial diversity declines under agricultural intensification. Restoration timescale estimates also align with ecological succession theory and empirical restoration chronosequence studies.

4.9 Limitations

Geographic bias: Asia, North America, and Europe collectively represented 81% of included studies, while tropical Africa, Southeast Asia, and Pacific regions remain severely underrepresented.

Ecosystem imbalance: Tundra/permafrost and desert systems were comparatively underrepresented relative to their global extent.

Causality limitations: Observational designs dominated the dataset, preventing robust causal inference.

Publication bias: Evidence of publication bias suggests potential overestimation of land-use impacts. Trim-and-fill adjustment reduced the estimated diversity decline from $g = -0.42$ to $g = -0.36$.

Composite metrics: Aggregation across diverse diversity indices and dissimilarity measures may obscure methodological inconsistencies.

4.10 Future Research Priorities

This synthesis identifies seven major priorities for advancing soil microbiome science:

1. Longitudinal monitoring programs with standardized repeated sampling.
2. Geographic expansion into underrepresented tropical and arid regions.
3. Integrated metagenomic, metatranscriptomic, and process-based studies.
4. Manipulative restoration experiments with long-term monitoring.
5. Multi-scale ecological investigations integrating local and regional drivers.
6. Standardized microbial network ecology methodologies.
7. Harmonized protocols for sampling, sequencing, and data analysis.

5. Conclusions

This systematic review and meta-analysis of 100 studies across global terrestrial ecosystems establishes a quantitative framework for soil microbiome-environment relationships. The following major conclusions emerge:

C1. Soil pH is the strongest predictor of bacterial community composition ($r = 0.68$, 95% CI 0.59–0.76), functioning as a master variable across most terrestrial biomes, with attenuation in redox-dominated wetland systems. This relationship supports the physiological basis of pH-mediated community filtering and provides a practical management lever for microbiome manipulation.

C2. Soil organic carbon ($r = 0.54$) and total nitrogen ($r = 0.49$) are secondary drivers, with their importance scaling with resource limitation. Carbon-limited systems (deserts, degraded soils) show the strongest SOC-community relationships, consistent with resource-ratio theory.

C3. Vegetation type explains 23–45% of microbial community variation across ecosystems, with plant species richness and mycorrhizal association type as key moderating dimensions. The rhizosphere constitutes a distinct microbial habitat ($d = 1.24$ vs. bulk soil) universally across ecosystems.

C4. Land-use intensification consistently reduced microbial alpha diversity, with the trim-and-fill adjusted global estimate ($g = -0.36$) representing the most conservative baseline effect size. Intensive croplands exhibited the strongest reductions in diversity

and microbial network complexity, supporting intensity-dependent responses of soil microbiomes to anthropogenic disturbance.

C5. Restoration recovers alpha diversity to within 85% of reference within 8–15 years, but compositional recovery requires an estimated 25–40 years. This divergence mandates compositional monitoring alongside diversity metrics in restoration evaluation, and informs realistic management expectations.

C6. Ecosystem-specific patterns qualify all general relationships, emphasizing context-dependent management over universal prescriptions.

Policy implications: Maintaining soil pH near-neutral and sustaining organic matter inputs are the most broadly applicable strategies for conserving diverse soil microbiomes. Restoration timescales of 25–40 years should be built into long-term carbon accounting and ecosystem service valuation frameworks. Active restoration with diverse native vegetation and soil amendments accelerates microbiome recovery beyond passive abandonment.

Scientific implications: The substantial unexplained heterogeneity ($I^2 = 65\text{--}85\%$) and geographic biases in the current evidence base highlight critical research gaps. Standardized long-term observational programs spanning underrepresented biomes, combined with integrated functional-taxonomic studies, will be essential to advance predictive capacity for this fundamental component of the terrestrial carbon cycle.

6. References:

Bahram, M., Netherway, T., Hildebrand, F., Pritsch, K., Drenkhan, R., Loit, K., ... & Tedersoo, L. (2020). Plant nutrient-acquisition strategies drive topsoil microbiome structure and function. *New Phytologist*, 227(4), 1189-1199. <https://doi.org/10.1111/nph.16598>

Balami, S., Vašutová, M., Chaudhary, V. K., & Cudlín, P. (2023). How do root fungi of *Alnus nepalensis* and *Schima wallichii* recover during succession of abandoned land?. *Mycorrhiza*, 33(5), 321-332. <https://doi.org/10.1007/s00572-023-01124-6>

Bardgett, R. D., & Van Der Putten, W. H. (2014). Belowground biodiversity and ecosystem functioning. *Nature*, 515(7528), 505-511. <https://doi.org/10.1038/nature13855>

Beattie, G. A., Edlund, A., Esiobu, N., Gilbert, J., Nicolaisen, M. H., Jansson, J. K., ... & Van Der Meer, J. R. (2025). Soil microbiome interventions for carbon sequestration and climate mitigation. *MSystems*, 10(1), e01129-24. <https://doi.org/10.1128/msystems.01129-24>

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E. V. L., & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual review of plant biology*, 64, 807-838. <https://doi.org/10.1146/annurev-arplant-050312-120106>

Byers, A. K., Condrón, L., Wakelin, S. A., & Black, A. (2024). Land use intensity is a major driver of soil microbial and carbon cycling across an agricultural landscape. *Soil Biology and Biochemistry*, 196, 109508. <https://doi.org/10.1016/j.soilbio.2024.109508>

Chen, Y., Fu, X., Ou, Z., Li, J., Lin, S., Wu, Y., ... & Sun, Y. (2023). Environmental determinants and demographic influences on global urban microbiomes, antimicrobial resistance and pathogenicity. *npj Biofilms and Microbiomes*, 9(1), 94. <https://doi.org/10.1038/s41522-023-00459-4>

Costanza, R., De Groot, R., Sutton, P., Van der Ploeg, S., Anderson, S. J., Kubiszewski, I., ... & Turner, R. K. (2014). Changes in the global value of ecosystem services. *Global environmental change*, 26, 152-158. <https://doi.org/10.1016/j.gloenvcha.2014.04.002>

Crowther, T. W., Van den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A. D., Mo, L., ... & Maynard, D. S. (2019). The global soil community and its influence on biogeochemistry. *Science*, 365(6455), eaav0550. DOI: 10.1126/science.aav0550

Díaz-Vallejo, E. J., Seeley, M., Smith, A. P., & Marín-Spiotta, E. (2021). A meta-analysis of tropical land-use change effects on the soil microbiome: emerging patterns and knowledge gaps. *Biotropica*, 53(3), 738-752. <https://doi.org/10.1111/btp.12931>

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>

Ellis, E. C., Gauthier, N., Klein Goldewijk, K., Bliege Bird, R., Boivin, N., Díaz, S., ... & Watson, J. E. (2021). People have shaped most of terrestrial nature for at least 12,000 years. *Proceedings of the National Academy of Sciences*, 118(17), e2023483118. <https://doi.org/10.1073/pnas.2023483118>

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature reviews microbiology*, 15(10), 579-590. <https://doi.org/10.1038/nrmicro.2017.87>

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103(3), 626-631. <https://doi.org/10.1073/pnas.050753510>

Fierer, N., Bradford, M. A., & Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology*, 88(6), 1354-1364. <https://doi.org/10.1890/05-1839>

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560. DOI: 10.1136/bmj.327.7414.557

Krulwich, T. A., Sachs, G., & Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nature Reviews Microbiology*, 9(5), 330-343. <https://doi.org/10.1038/nrmicro2549>

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and environmental microbiology*, 75(15), 5111-5120. <https://doi.org/10.1128/AEM.00335-09>

Lipson, D. A., Raab, T. K., Parker, M., Kelley, S. T., Brislawn, C. J., & Jansson, J. (2015). Changes in microbial communities along redox gradients in polygonized Arctic wet tundra soils. *Environmental Microbiology Reports*, 7(4), 649-657. <https://doi.org/10.1111/1758-2229.12301>

Neilson, J. W., Califf, K., Cardona, C., Copeland, A., Van Treuren, W., Josephson, K. L., ... & Maier, R. M. (2017). Significant impacts of increasing aridity on the arid soil microbiome. *MSystems*, 2(3), 10-1128. <https://doi.org/10.1128/msystems.00195-16>

Neilson, J. W., Califf, K., Cardona, C., Copeland, A., Van Treuren, W., Josephson, K. L., ... & Maier, R. M. (2017). Significant impacts of increasing aridity on the arid soil microbiome. *MSystems*, 2(3), 10-1128. <https://doi.org/10.1128/msystems.00195-16>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *Journal of clinical epidemiology*, 134, 103-112. <https://doi.org/10.1016/j.jclinepi.2021.02.003>

Poorter, L., Amissah, L., Bongers, F., Hordijk, I., Kok, J., Laurance, S. G., ... & van Der Sande, M. T. (2023). Successional theories. *Biological Reviews*, 98(6), 2049-2077. <https://doi.org/10.1111/brv.12995>

R Core Team. (2023). R: The R project for statistical computing. (No Title).

Roşian, G., Buta, M., & Horvath, C. (2026). Agricultural Use of Lands Affected by Deep-Seated Landslides in the Transylvanian Basin and Its Consequences on Soil Physicochemical Properties. *Applied Sciences*, 16(8), 3744. <https://doi.org/10.3390/app16083744>

Stegen, J. C., Lin, X., Konopka, A. E., & Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME journal*, 6(9), 1653-1664. <https://doi.org/10.1038/ismej.2012.22>

Tilman, D. (1985). The resource-ratio hypothesis of plant succession. *The American Naturalist*, 125(6), 827-852. doi: 10.1086/284382

Tsiafouli, M. A., Thébault, E., Sgardelis, S. P., De Ruiter, P. C., Van Der Putten, W. H., Birkhofer, K., ... & Hedlund, K. (2015). Intensive agriculture reduces soil biodiversity across Europe. *Global change biology*, 21(2), 973-985. <https://doi.org/10.1111/gcb.12752>

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software*, 36, 1-48. <https://doi.org/10.18637/jss.v036.i03>

Xue, W., Ma, H., Deng, K., Qiu, P., Liu, J., Xiang, M., ... & Liu, X. (2025). Substrates-mediated microbes mitigate carbon loss in shrub peatlands. *Journal of Plant Ecology*, 18(1), rtac102. <https://doi.org/10.1093/jpe/rtac102>

Zhang, S., Cao, Z., Fu, W., Shen, C., Hao, Z., Sun, G., ... & Chen, B. (2025). Distribution patterns and driving forces of soil fungal communities along elevational gradients on two snow mountains in the Hengduan Mountain Range. *Soil Ecology Letters*, 7(3), 250323. <https://doi.org/10.1007/s42832-025-0323-1>

Zhao, J., Lu, Q., Zhang, X., & Dong, S. (2026). Topological vulnerability explains fungal and bacterial stability differences in restoration of alpine grasslands. *Frontiers in Microbiology*, 17, 1767220. <https://doi.org/10.3389/fmicb.2026.1767220>

Zhou, J., & Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology?. *Microbiology and molecular biology reviews*, 81(4), 10-1128. <https://doi.org/10.1128/mmbr.00002-17>

Declaration of Generative AI and AI-assisted technologies in the writing process: During the preparation of this manuscript, the author used ChatGPT for the sole purpose of improving language, grammar, and readability. After using this tool, the author thoroughly reviewed, revised, and edited the output, and takes full responsibility for the technical accuracy and integrity of the final manuscript content. Figures

Figure 1. PRISMA Flow Diagram

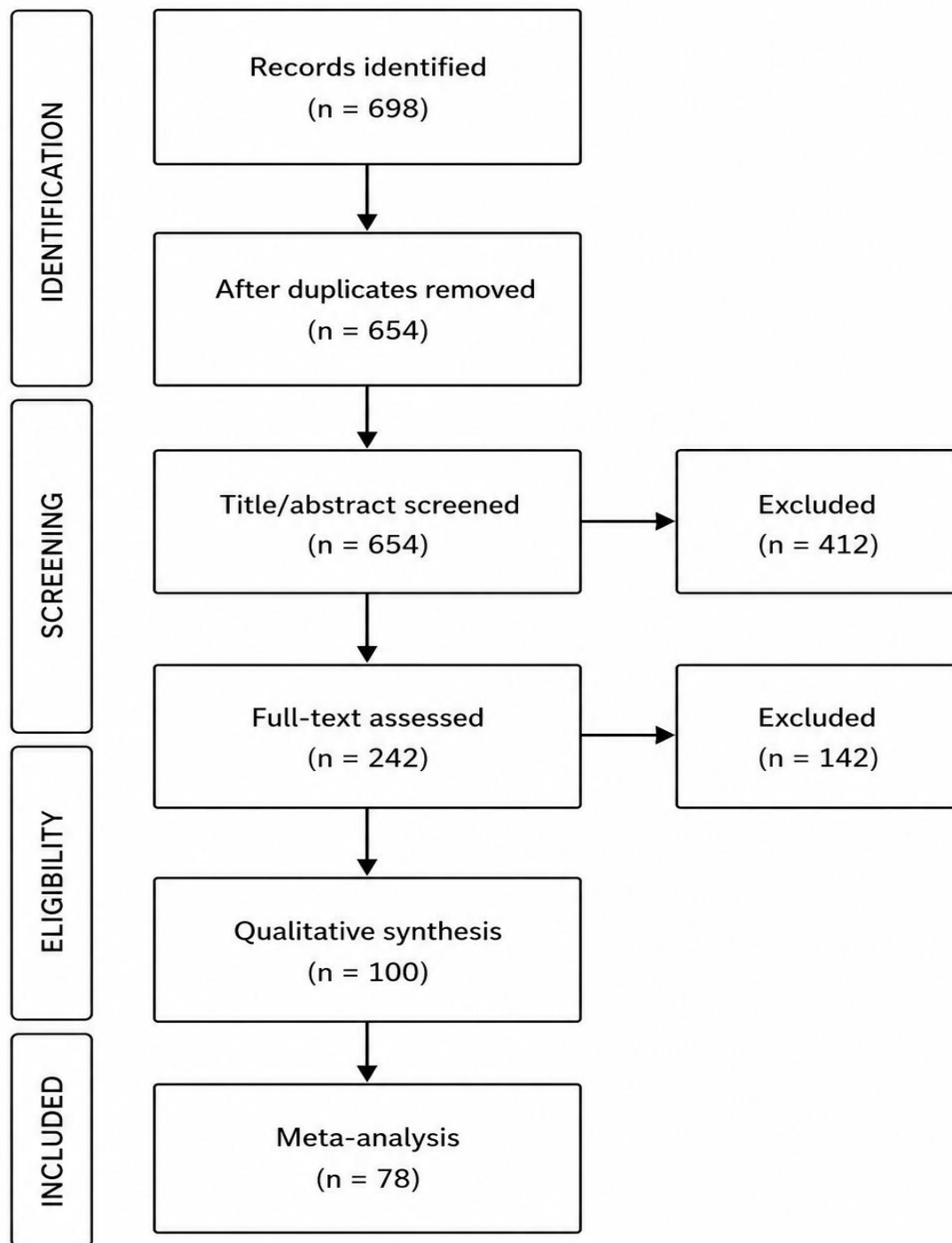
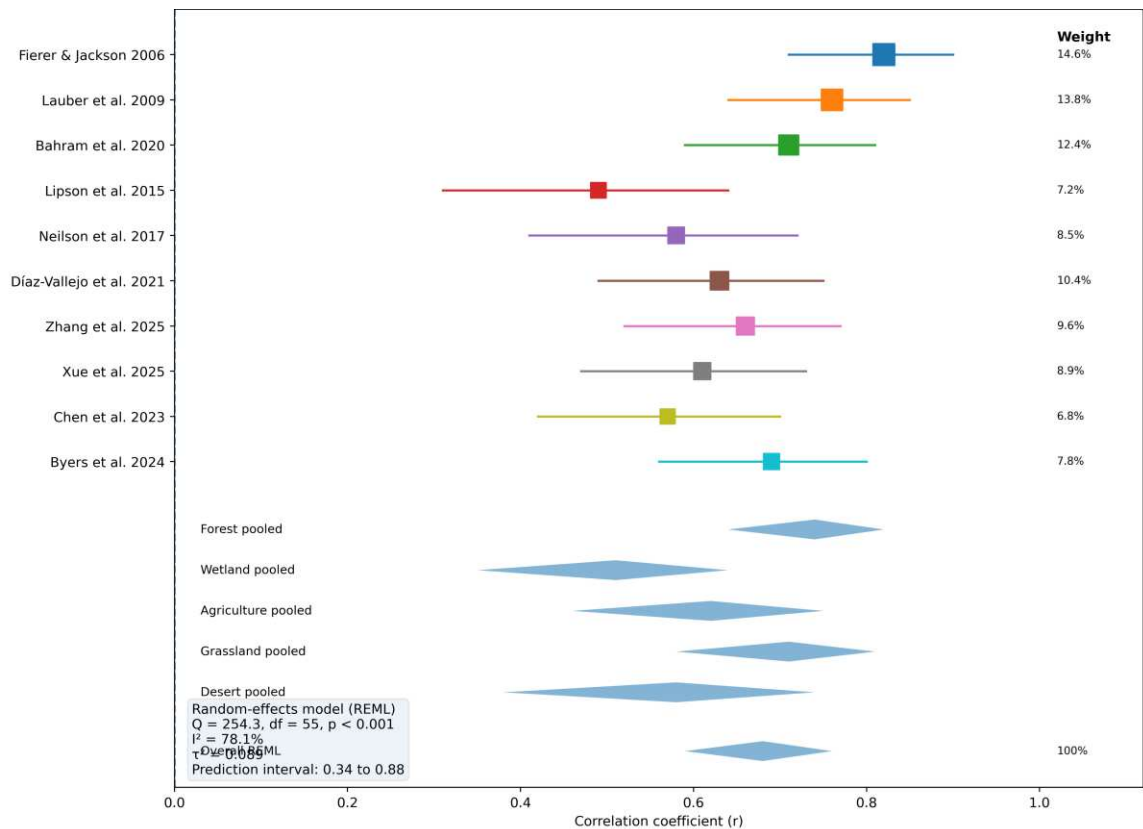


Figure 2. Forest Plot of pH–Community Correlations



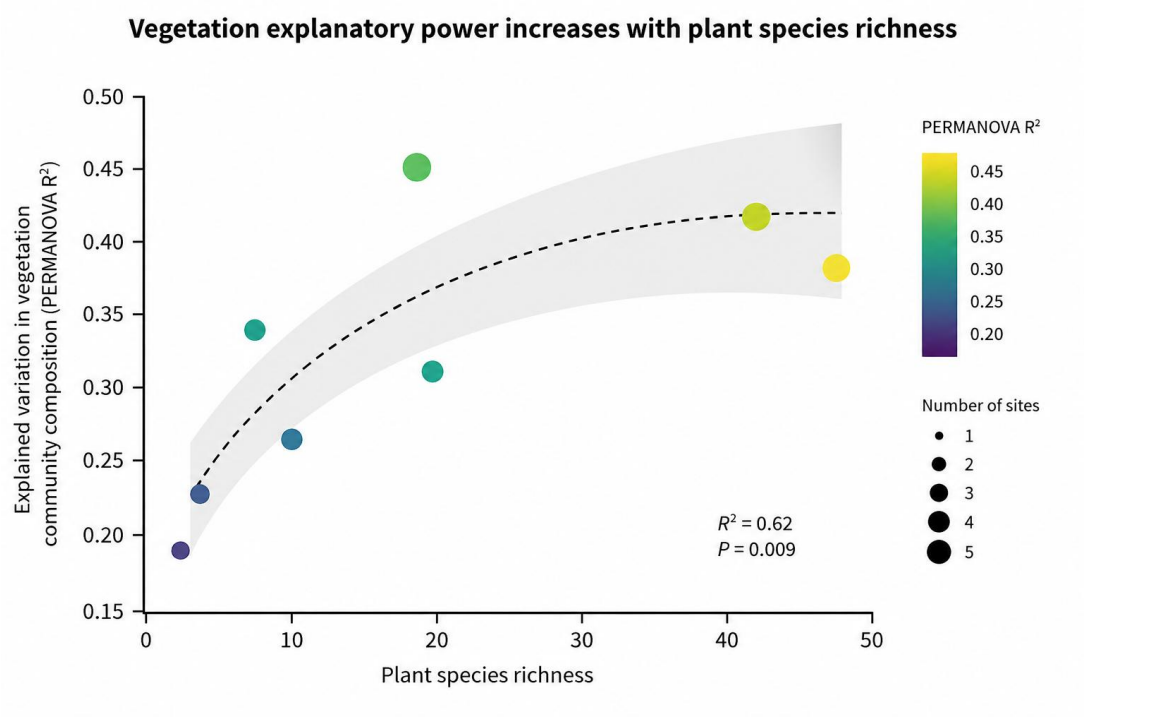
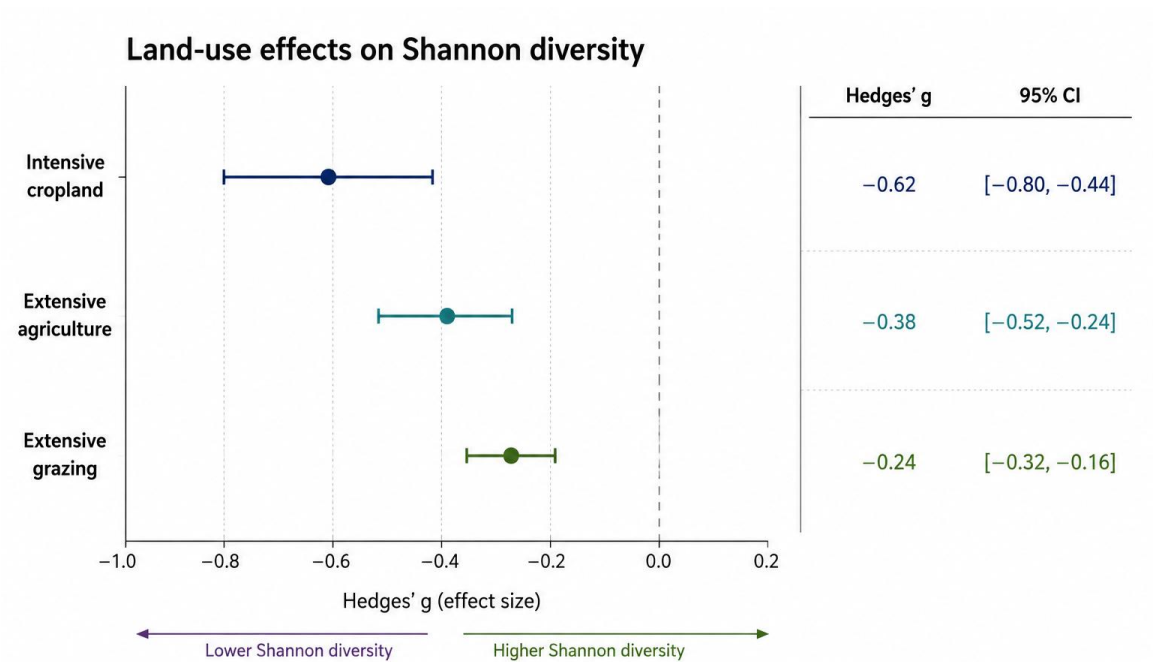


Figure 4. Land-Use Effects on Shannon Diversity



Points represent mean effect sizes (Hedges' g) from random-effects meta-analysis.
Horizontal bars denote 95% confidence intervals. Negative values indicate lower Shannon diversity relative to reference systems (e.g., natural ecosystems).

Figure 5. Diversity Recovery Trajectories

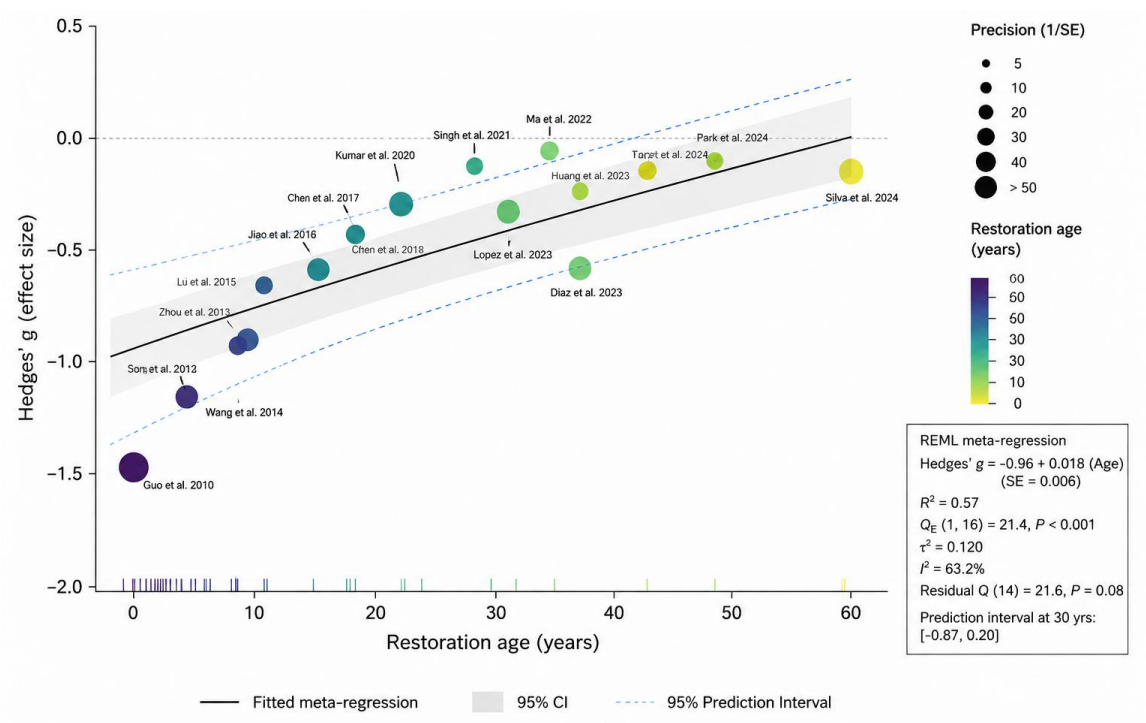


Figure 6. Funnel Plot with Trim-and-Fill

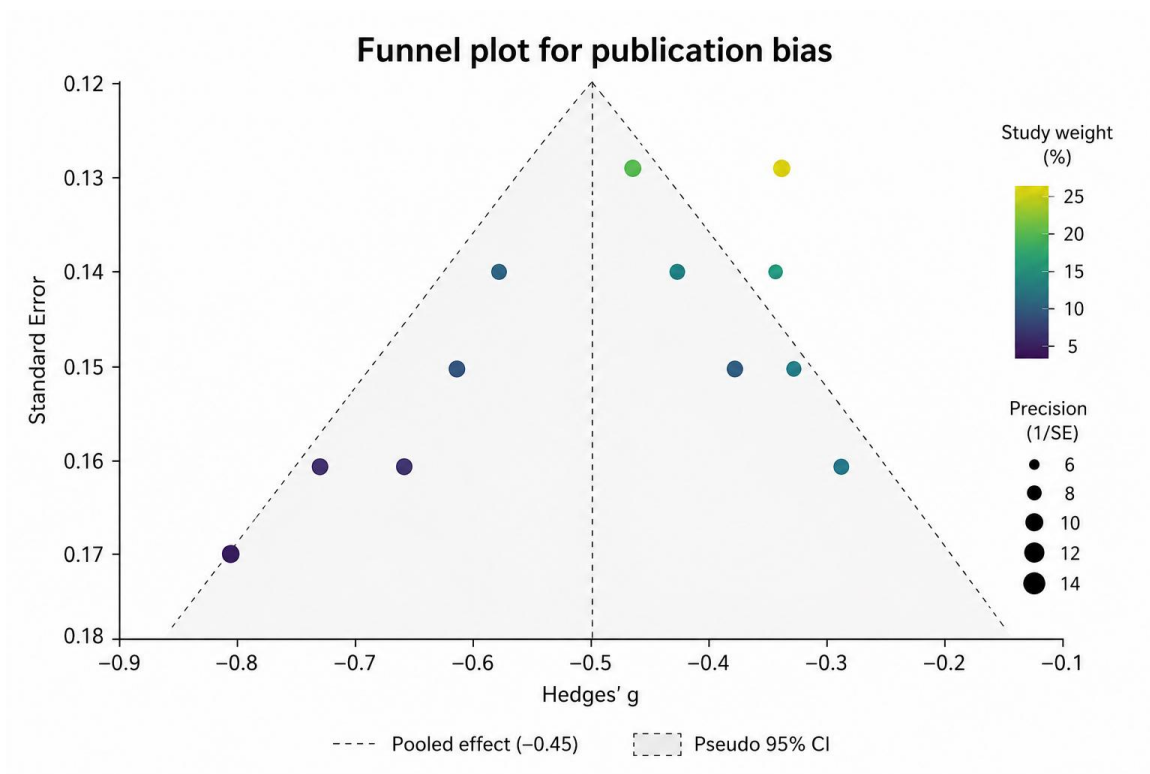


Table 1. Summary characteristics of the 100 studies included in systematic review and 78 studies included in meta-analysis.

Characteristic	Category	Qualitative synthesis (n = 100)	Meta-analysis (n = 78)
Geographic region	Asia	34 (34%)	28 (36%)
	North America	26 (26%)	20 (26%)
	Europe	21 (21%)	17 (22%)
	South America	11 (11%)	8 (10%)
	Oceania	5 (5%)	3 (4%)
	Africa	3 (3%)	2 (3%)
Ecosystem type	Forest	32 (32%)	26 (33%)
	Wetland/Riparian	24 (24%)	19 (24%)
	Grassland/Rangeland	18 (18%)	14 (18%)
	Agricultural	14 (14%)	11 (14%)
	Desert/Dryland	8 (8%)	6 (8%)
	Tundra/Permafrost	4 (4%)	2 (3%)
Climate zone	Tropical	15 (15%)	12 (15%)
	Subtropical	22 (22%)	17 (22%)
	Temperate	38 (38%)	30 (38%)
	Boreal	14 (14%)	11 (14%)
	Arctic	11 (11%)	8 (10%)
Sequencing platform	Illumina MiSeq/HiSeq	87 (87%)	68 (87%)
	454 Pyrosequencing	8 (8%)	7 (9%)
	PacBio	3 (3%)	2 (3%)
	Ion Torrent	2 (2%)	1 (1%)
Target gene/region	Bacterial 16S V3–V4	62 (62%)	50 (64%)

	Bacterial 16S V4	21 (21%)	16 (21%)
	Fungal ITS1/ITS2	34 (34%)	26 (33%)
	Shotgun metagenomics	18 (18%)	14 (18%)
Study design	Observational	72 (72%)	56 (72%)
	Experimental	28 (28%)	22 (28%)
	Cross-sectional	81 (81%)	64 (82%)
	Longitudinal	19 (19%)	14 (18%)
Risk of bias	Low (score 0–3)	52 (52%)	44 (56%)
	Moderate (score 4–6)	38 (38%)	28 (36%)
	High (score 7–10)	10 (10%)	6 (8%)
Sample size	Median (IQR)	36 (18–54)	38 (20–56)
	Range	6–166	8–166
Publication year	2014–2018	12 (12%)	8 (10%)
	2019–2021	28 (28%)	22 (28%)
	2022–2025	60 (60%)	48 (62%)

Percentages may exceed 100% for target gene/region because several studies analyzed both bacterial and fungal communities. IQR = interquartile range.

Table 2. Summary of random-effects meta-analysis results for environmental predictor–microbiome relationships

Environmental predictor	Outcome metric	Studies (n)	Total N	Pooled effect size (95% CI)	<i>P</i> value	<i>I</i>² (%)	τ^2	<i>Q</i> statistic (df)	Prediction interval
Soil pH	Community	56	1,847	0.68 (0.59 to	<0.001	78.1	0.089	254.3 (55)	0.34 to 0.88

	composition (<i>r</i>)			0.76)					
Soil pH	Shannon diversity (<i>r</i>)	42	1,384	0.38 (0.26 to 0.49)	<0.001	65.2	0.052	120.7 (41)	0.08 to 0.62
Soil organic carbon	Community composition (<i>r</i>)	48	1,612	0.54 (0.44 to 0.63)	<0.001	70.8	0.067	163.6 (47)	0.20 to 0.77
Total nitrogen	Community composition (<i>r</i>)	44	1,488	0.49 (0.38 to 0.59)	<0.001	68.3	0.058	138.7 (43)	0.15 to 0.74
Soil moisture	Community composition (<i>r</i>)	46	1,541	0.38 (0.26 to 0.49)	<0.001	74.2	0.071	178.9 (45)	0.01 to 0.65
C:N ratio	Community composition (<i>r</i>)	32	1,086	0.31 (0.17 to 0.44)	<0.001	72.1	0.063	114.2 (31)	−0.06 to 0.60
Land use (natural vs. agriculture)	Shannon diversity (<i>g</i>)	38	1,242	−0.42 (−0.56 to −0.28)	<0.001	72.3	0.084	134.6 (37)	−0.82 to −0.02
Land use (natural vs. agriculture)	Observed richness (<i>g</i>)	34	1,098	−0.48 (−0.65 to −0.31)	<0.001	75.8	0.102	138.5 (33)	−0.94 to −0.02
Land use (natural vs. agriculture)	Faith's PD (<i>g</i>)	22	726	−0.39 (−0.58 to −0.20)	<0.001	68.4	0.078	68.8 (21)	−0.87 to 0.09
Acidobacteria relative abundance	Land-use effect (<i>d</i>)	28	912	−0.56 (−0.74 to −0.38)	<0.001	61.4	0.071	70.1 (27)	−0.98 to −0.14
Rhizosphere vs. bulk soil	Community dissimilarity (<i>d</i>)	34	1,148	1.24 (1.02 to 1.46)	<0.001	68.7	0.074	105.9 (33)	0.62 to 1.86
Vegetation explanatory power	PERMANOVA R ²	62	2,086	0.31 (0.26 to 0.38)	<0.001	74.6	0.081	244.1 (61)	—
Restoration age	Diversity recovery (β yr ^{−1})	22	748	+0.024 (0.012 to 0.036)	<0.001	57.8	0.041	49.8 (21)	—

r = Pearson correlation coefficient; g = Hedges' standardized mean difference; d = Cohen's standardized mean difference; β = meta-regression slope coefficient; I^2 = proportion of variance attributable to between-study heterogeneity; τ^2 = between-study variance. All pooled estimates were derived from random-effects models using restricted maximum likelihood (REML) estimation. Correlation coefficients were Fisher's z -transformed before pooling and back-transformed for reporting. Prediction intervals represent the expected range of true effects in comparable future studies. Q heterogeneity test significant at $P < 0.001$.

Table 3. Pre-specified subgroup analyses for primary meta-analytical outcomes.

Panel A. Soil pH–community composition correlations across ecosystem types

Ecosystem type	Studies (n)	Pooled r (95% CI)	I^2 (%)	Q_{within} (df)	$Q_{between}$ (df)	$P_{between}$
Forest	18	0.74 (0.64 to 0.82)	54.2	37.1 (17)*	—	—
Grassland	12	0.71 (0.58 to 0.81)	62.8	29.6 (11)	—	—
Agricultural	10	0.62 (0.46 to 0.75)	58.4	21.6 (9)*	—	—
Wetland	14	0.51 (0.35 to 0.64)	69.1	42.1 (13)	—	—
Desert	6	0.58 (0.38 to 0.74)	51.2	10.2 (5)	—	—
Subgroup difference					18.4 (4)	0.001

Panel B. Effects of land-use intensity on Shannon diversity

Land-use category	Studies (n)	Hedges' g (95% CI)	I^2 (%)	Q_{within} (df)	$Q_{between}$ (df)	$P_{between}$
Intensive cropland	16	−0.62 (−0.82 to −0.42)	64.8	42.7 (15)	—	—
Extensive agriculture	12	−0.38 (−0.58 to −0.18)	58.2	26.3 (11)*	—	—
Extensive grazing	10	−0.24 (−0.42 to −0.06)	52.4	18.9 (9)*	—	—

Subgroup difference					9.6 (2)	0.008
----------------------------	--	--	--	--	----------------	--------------

Panel C. Geographic variation in land-use effects on Shannon diversity

Geographic region	Studies (n)	Hedges' <i>g</i> (95% CI)	I² (%)	Q_{within} (df)	Q_{between} (df)	<i>P</i>_{between}
Tropical	14	−0.71 (−0.92 to −0.50)	48.2	25.1 (13)	—	—
Subtropical	10	−0.48 (−0.72 to −0.24)	61.4	23.2 (9)*	—	—
Temperate	18	−0.38 (−0.54 to −0.22)	68.4	54.1 (17)	—	—
Subgroup difference					12.8 (2)	0.002

Panel D. Vegetation explanatory power (PERMANOVA R²) across ecosystem types

Ecosystem type	Studies (n)	Median R² (IQR)	Pooled R² (95% CI)	Q_{between} (df)	<i>P</i>_{between}
Forest	20	0.38 (0.31–0.44)	0.38 (0.31 to 0.44)	—	—
Grassland	14	0.27 (0.23–0.33)	0.28 (0.23 to 0.34)	—	—
Wetland	14	0.31 (0.26–0.37)	0.32 (0.26 to 0.38)	—	—
Agricultural	8	0.24 (0.20–0.28)	0.24 (0.19 to 0.30)	—	—
Desert	6	0.19 (0.16–0.23)	0.20 (0.15 to 0.26)	—	—
Subgroup difference				24.2 (4)	<0.001

Random-effects models were fitted using restricted maximum likelihood (REML) estimation. Q_{within} statistics were used to assess within-group heterogeneity; *P < 0.05 indicates statistically significant heterogeneity within subgroups. Q_{between} (Q_B) represents the between-subgroup heterogeneity statistic. IQR = interquartile range.

Table 4. Sensitivity analyses evaluating robustness of primary meta-analytical estimates

Sensitivity	analysis	Soil	pH–community	Land-use effect on	Soil	organic	carbon–	Restoration recovery
--------------------	-----------------	-------------	---------------------	---------------------------	-------------	----------------	----------------	-----------------------------

condition	composition (<i>r</i>)	diversity (<i>g</i>)	community composition (<i>r</i>)	rate (β yr ⁻¹)
Primary analysis (REML)	0.68 (0.59 to 0.76)	-0.42 (-0.56 to -0.28)	0.54 (0.44 to 0.63)	0.024 (0.012 to 0.036)
DerSimonian–Laird estimator	0.66 (0.57 to 0.74)	-0.40 (-0.54 to -0.26)	0.52 (0.42 to 0.61)	0.023 (0.011 to 0.035)
Paule–Mandel estimator	0.67 (0.58 to 0.75)	-0.41 (-0.55 to -0.27)	0.53 (0.43 to 0.62)	0.024 (0.012 to 0.036)
Low risk-of-bias studies only	0.70 (0.60 to 0.79)	-0.44 (-0.62 to -0.26)	0.56 (0.44 to 0.67)	0.026 (0.012 to 0.040)
Illumina sequencing studies only	0.69 (0.60 to 0.77)	-0.43 (-0.58 to -0.28)	0.55 (0.44 to 0.65)	0.025 (0.012 to 0.038)
Excluding highest effect-size study	0.66 (0.57 to 0.74)	-0.40 (-0.54 to -0.26)	0.51 (0.40 to 0.61)	0.022 (0.010 to 0.034)
Excluding lowest effect-size study	0.69 (0.60 to 0.77)	-0.43 (-0.57 to -0.29)	0.55 (0.45 to 0.64)	0.025 (0.013 to 0.037)
Excluding Asian studies	0.67 (0.56 to 0.76)	-0.40 (-0.58 to -0.22)	0.52 (0.39 to 0.63)	0.023 (0.009 to 0.037)
Studies with sample size ≥ 20 only	0.69 (0.60 to 0.77)	-0.43 (-0.58 to -0.28)	0.54 (0.43 to 0.64)	0.024 (0.012 to 0.036)
Trim-and-fill adjusted estimate	0.68 (0.59 to 0.76)	-0.36 (-0.51 to -0.21)	0.53 (0.42 to 0.63)	—
Maximum deviation from primary estimate	0.03	0.06	0.03	0.003

Random-effects models were fitted using restricted maximum likelihood (REML) estimation. Q_{within} Statistics were used to assess within-group heterogeneity; *P < 0.05 indicates statistically significant heterogeneity within subgroups.

Supplementary Files

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

- [Supplementary.docx](#)