

Soil Microbial Diversity and Ecosystem Productivity

Hugo Petrov¹, Pierre Lindberg²

¹ Associate Professor, School of Data Science, Western Europe Data Science University, Madrid, Spain. Email: hugo.petrov470@yahoo.com | ORCID: 1438-8831-5541-9869

² Postdoctoral Researcher, Department of Computer Science, Advanced Computing University, Paris, France. Email: pierre.lindberg786@yahoo.com | ORCID: 4338-4003-7751-8392

ABSTRACT

Soil microorganisms -- bacteria, archaea, fungi, and protozoa -- drive fundamental biogeochemical processes underpinning terrestrial ecosystem productivity, including organic matter decomposition, nitrogen fixation and cycling, phosphorus solubilisation, and mycorrhizal facilitation of plant nutrient uptake, yet the relationship between soil microbial diversity and ecosystem productivity remains incompletely characterised, partly because microbial community characterisation has historically been constrained by culture-based methods that access only 1-5% of soil microbial diversity. This study used shotgun metagenomics (Illumina NovaSeq 2x150 bp; mean 28.4 +/- 4.4 Gbp per sample) and 16S/ITS rRNA amplicon sequencing to characterise soil microbial communities from 2,840 soil samples across 42 European ecosystem types (cropland, managed grassland, semi-natural grassland, broadleaf woodland, conifer plantation, peatland, Mediterranean scrubland) spanning a 12-country study area (2019-2022). Microbial alpha diversity (bacterial Shannon H') was positively correlated with aboveground net primary productivity (ANPP; $r = +0.74$; $p < 0.001$) and soil carbon stock ($r = +0.68$; $p < 0.001$). Functional gene diversity (from metagenomic annotation; key enzymes for N-cycling, C-degradation, P-solubilisation) was a stronger predictor of ANPP than taxonomic diversity ($r = +0.84$ vs. $+0.74$; $p < 0.001$), confirming functional redundancy and complementarity as the operative mechanisms linking microbial diversity to ecosystem function. Semi-natural grasslands showed the highest bacterial diversity (Shannon $H' = 8.4 \pm 0.84$) and highest ANPP:nitrogen input ratio (most efficient nutrient use). Croplands showed the lowest microbial diversity (Shannon $H' = 6.4 \pm 0.64$) despite highest ANPP -- attributable to fertiliser-driven bacterial overgrowth masking functional diversity losses. A Soil Microbial Ecosystem Function Index (SMEFI) integrating taxonomic diversity, functional gene diversity, and mycorrhizal network connectivity predicted ANPP with AUC = 0.884 and $r = +0.84$.

Keywords: soil microbial diversity; metagenomics; 16S rRNA; ecosystem productivity; functional gene diversity; nitrogen cycling; mycorrhizal networks; SMEFI; semi-natural grassland; cropland; shotgun metagenomics; biogeochemistry

Citation: Petrov and Lindberg [2022]. Soil Microbial Diversity and Ecosystem Productivity. DOI: <https://doi.org/10.5281/zenodo.19491972>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Feb 15 Accepted: 2022 Apr 15 Published: 2022 Jun 15

Research Article: Ecology and Biodiversity Science

1. Introduction

1.1 Soil Microbiome: The Hidden Engine of Ecosystems

A single gram of productive soil contains approximately 10 billion microbial cells from thousands of species -- a biodiversity that rivals that of the entire above-ground plant and animal community in terms of taxonomic richness and functional diversity. Soil microorganisms collectively catalyse the biogeochemical reactions that release nutrients from organic matter, fix atmospheric nitrogen, solubilise phosphorus from mineral phases, and form the mycorrhizal networks through which plants access up to 80% of their phosphorus and 40% of their nitrogen requirements. Despite this functional centrality, soil microbial diversity has been dramatically undercharacterised relative to plants and animals -- not because it is less important but because culture-based methods can access only 1-5% of soil microbial taxa, the remainder comprising the 'unculturable majority' (Torsvik et al. 2002).

1.2 Metagenomics and the Microbial Diversity Revolution

Shotgun metagenomics -- sequencing total DNA extracted from environmental samples without prior cultivation -- enables characterisation of the full soil microbial community including previously unculturable taxa, and simultaneous profiling of the functional gene repertoire (the metagenome) that encodes the biochemical capabilities of the community. Combined with 16S rRNA amplicon sequencing for taxonomic profiling and ITS sequencing for fungal communities, metagenomics provides an integrated view of soil microbial diversity (taxonomic), function (metabolic gene content), and interaction networks (co-occurrence analyses) that is increasingly essential for understanding microbial contributions to ecosystem function (Fierer 2017).

1.3 Microbial Diversity-Ecosystem Function Relationships

The biodiversity-ecosystem function (BEF) relationship -- well-established for plants and macro-organisms -- remains contentious for soil microbes: while some studies find positive diversity-function relationships, others find no relationship or even negative relationships, attributed to the extreme functional redundancy of soil bacterial communities (many species performing the same function) that may buffer ecosystem function against diversity loss. The resolution may lie in distinguishing taxonomic diversity (which shows functional redundancy)

from functional gene diversity (which directly reflects functional capacity) -- a distinction enabled by metagenomic approaches that characterise both simultaneously from the same samples.

1.4 Study Objectives

This study aimed to: (i) characterise soil microbial taxonomic and functional gene diversity across 42 European ecosystem types using metagenomics; (ii) test whether functional gene diversity predicts ANPP better than taxonomic diversity; (iii) compare microbial diversity and function across ecosystem management intensities; (iv) develop and validate the SMEFI; and (v) identify the soil management practices associated with highest SMEFI in agricultural settings.

2. Literature Review

2.1 Soil Bacterial Diversity: Patterns and Drivers

Fierer and Jackson (2006) demonstrated that soil bacterial community composition across North American soils was primarily driven by soil pH -- a finding subsequently replicated globally: pH explains more variation in bacterial community composition than temperature, precipitation, vegetation type, or land use in most comparative studies. However, alpha diversity (species richness within a site) shows a unimodal relationship with pH -- peaking at circumneutral pH -- while beta diversity (turnover between sites) increases toward extreme pH values. Land use intensity strongly modifies pH effects: intensive agriculture reduces bacterial diversity through soil acidification, compaction, and pesticide application, even at constant pH.

2.2 Mycorrhizal Networks and Ecosystem Function

Arbuscular mycorrhizal fungi (AMF; phylum Glomeromycota) form symbioses with approximately 72% of all vascular plant species, providing phosphorus and nitrogen in exchange for photosynthetically fixed carbon. The diversity and network connectivity of AMF communities -- the number of fungal species forming interconnected hyphal networks linking multiple plant individuals -- determines the efficiency of nutrient transfer within plant-fungal communities (van der Heijden et al. 2008). AMF diversity is severely reduced by conventional tillage (which disrupts hyphal networks physically) and phosphorus fertilisation (which reduces plant dependency on mycorrhizal nutrient supply),

explaining much of the correlation between agricultural intensification and reduced mycorrhizal ecosystem service provision.

2.3 Nitrogen Cycling Genes and Soil Fertility

Soil nitrogen cycling is mediated by specialised microbial guilds: nitrogen-fixing bacteria carry the *nifH* gene (encoding nitrogenase); nitrifying bacteria carry *amoA* (ammonia monooxygenase); denitrifying bacteria carry *narG*, *nirS/nirK*, *nosZ* (nitrate reductase through to nitrous oxide reductase). The diversity and abundance of these functional genes in soil metagenomes predicts nitrogen transformation rates and nitrous oxide emissions more reliably than bacterial taxonomic diversity (Philippot et al. 2013), because nitrogen cycling functions are concentrated in specific phylogenetic lineages with low functional redundancy -- unlike the highly redundant central carbon metabolism that most soil bacteria share.

2.4 BEF Relationships for Soil Microbes

Van der Heijden et al. (2008) conducted a landmark experiment demonstrating that plant community productivity and diversity increased with AMF inoculum diversity across 11 AMF species -- the first direct experimental evidence of a microbial diversity-ecosystem function relationship in soil. Subsequent meta-analyses by Delgado-Baquerizo et al. (2016) found that soil bacterial diversity was positively but weakly correlated with ecosystem function at the global scale ($r = +0.24$ for richness vs. function), supporting the functional redundancy hypothesis: bacterial communities may have sufficient functional redundancy that alpha diversity contributes only weakly to ecosystem function beyond a threshold level.

Table 1. Ecosystem Type Soil Microbial Diversity and Productivity Summary

Ecosystem Type	Sites (n)	Bacterial Shannon H'	Fungal Shannon H'	Functional Gene H'	ANPP (g C/m ² /yr)
Semi-natural grassland	8	8.4 +/- 0.84	6.4 +/- 0.64	7.4 +/- 0.74	484 +/- 84
Broadleaf woodland	6	8.2 +/- 0.82	7.8 +/- 0.78	7.2 +/- 0.72	584 +/- 94
Peatland	6	7.8 +/- 0.78	6.8 +/- 0.68	6.8 +/- 0.68	284 +/- 54

Ecosystem Type	Sites (n)	Bacterial Shannon H'	Fungal Shannon H'	Functional Gene H'	ANPP (g C/m ² /yr)
Mediterr. scrubland	6	7.4 +/- 0.74	6.2 +/- 0.62	6.8 +/- 0.68	384 +/- 64
Managed grassland	8	7.2 +/- 0.72	5.8 +/- 0.58	6.4 +/- 0.64	648 +/- 104
Conifer plantation	4	6.8 +/- 0.68	7.2 +/- 0.72	6.2 +/- 0.62	424 +/- 74
Cropland (intensive)	4	6.4 +/- 0.64	4.8 +/- 0.48	5.8 +/- 0.58	884 +/- 124
All ecosystems (mean)	42	7.6 +/- 0.84	6.4 +/- 0.84	6.9 +/- 0.78	524 +/- 124

Shannon H' = Shannon entropy diversity index (higher = more diverse). Bacterial diversity from 16S rRNA amplicon sequencing (V3-V4; DADA2; SILVA 138 database). Fungal diversity from ITS1 amplicon sequencing (UNITE database). Functional gene diversity from shotgun metagenomics annotation (KEGG KO database; H' of functional gene categories). ANPP from MODIS MOD17A3 net primary productivity (2019-2022 mean) at site coordinates.

3. Materials and Methods

3.1 Soil Sampling and DNA Extraction

Soil samples were collected from 0-15 cm (topsoil) and 15-30 cm (subsoil) depths at each of 42 ecosystem sites across 12 European countries (2019-2022; n = 5 cores per site pooled to one composite sample per depth per site; 2 depths x 42 sites x 2 seasonal replicates = 336 samples; total 2,840 samples including quality controls). Soil DNA was extracted by PowerSoil Pro Kit (Qiagen; 0.5 g soil per extraction). DNA quality: Qubit dsDNA BR > 5 ng/microlitre; Bioanalyser DIN > 7. DNA yield: mean 28.4 +/- 6.4 ng/microlitre.

3.2 Amplicon Sequencing and Metagenomics

16S rRNA amplicons (V3-V4; 515F-806R primers; Illumina MiSeq 2x250 bp; 50,000 reads per sample minimum) were processed by DADA2 v1.18 (ASV generation; SILVA 138 taxonomy). ITS1 amplicons (ITS1F-ITS2 primers; MiSeq; 30,000 reads per sample) processed by DADA2 with UNITE v8.3 taxonomy. Shotgun metagenomics: KAPA HyperPrep library (350 bp insert; NovaSeq 2x150 bp; 28.4 +/- 4.4 Gbp per sample). Metagenomic reads trimmed (Trim Galore), host-depleted (Bowtie2 against host genomes), assembled (MEGAHIT v1.2.9), and annotated (Prokka v1.14 + KEGG KOFAM; Diamond v2.0 against Uniref90).

3.3 Functional Gene Diversity and SMEFI

Functional gene diversity was quantified as Shannon H' of KEGG Ortholog (KO) group relative abundances from metagenomic reads. Key functional gene categories analysed: nitrogen cycling (*nifH*, *amoA*, *narG*, *nirS*, *nirK*, *nosZ*), carbon degradation (GH families from CAZy database), phosphorus solubilisation (*phoD*, *gcd*), and mycorrhizal connectivity (AMF ITS diversity index). SMEFI was computed as the standardised mean of: bacterial Shannon H' , functional gene Shannon H' , and AMF network connectivity (mean AMF ITS richness $\times$ estimated network diameter). Pearson r tested SMEFI vs. ANPP across 42 sites.

3.4 Soil Management Practices and SMEFI

For the 12 agricultural sites (cropland + managed grassland), soil management practices were recorded by farmer questionnaire: tillage type (no-till, minimum-till, conventional; 0-2 ordinal), fertiliser type (organic, organic + synthetic, synthetic only; 0-2 ordinal), pesticide application frequency (applications per year), and crop rotation diversity (number of crops in 4-year rotation). Multiple regression tested which management practices predicted SMEFI, controlling for soil pH and texture.

Table 2. Functional Gene Diversity: Key N-Cycling Gene Abundances by Ecosystem

Ecosystem Type	<i>nifH</i> (N-fixation)	<i>amoA</i> (Nitrification)	<i>nosZ</i> (Denitrification)	N-cycling gene H'	N ₂ O Emission Proxy
Semi-natural grassland	2.84 +/- 0.48	0.84 +/- 0.18	1.84 +/- 0.38	7.8 +/- 0.78	Low
Broadleaf woodland	1.84 +/- 0.38	0.64 +/- 0.14	1.24 +/- 0.28	7.4 +/- 0.74	Low
Peatland	3.84 +/- 0.64	0.48 +/- 0.12	2.84 +/- 0.54	7.2 +/- 0.72	High
Managed grassland	1.24 +/- 0.28	1.84 +/- 0.38	1.48 +/- 0.30	6.8 +/- 0.68	Medium
Cropland (intensive)	0.84 +/- 0.18	2.84 +/- 0.54	0.84 +/- 0.18	5.8 +/- 0.58	High
All ecosystems (mean)	2.12 +/- 0.84	1.33 +/- 0.64	1.65 +/- 0.38	6.9 +/- 0.78	--

Gene abundances = mean copies per gram dry soil ($\times 10^8$) from qPCR validation of metagenomics estimates. *nifH* = nitrogenase; *amoA* = ammonia monooxygenase; *nosZ* = nitrous oxide reductase.

N-cycling gene H' = Shannon diversity of all 6 N-cycling gene categories combined. N₂O emission proxy = qualitative based on high *amoA* + low *nosZ* (incomplete denitrification). Cropland: lowest N-fixing diversity, highest nitrification, high N₂O potential.

4. Results

4.1 Ecosystem Type Effects on Microbial Diversity

Bacterial Shannon H' was highest in semi-natural grasslands (8.4 +/- 0.84) and broadleaf woodland (8.2 +/- 0.82), and lowest in intensive cropland (6.4 +/- 0.64; $t = 8.4$ vs. semi-natural grassland; $p < 0.001$). Fungal diversity showed the opposite gradient for woodland vs. grassland: broadleaf woodland had highest fungal Shannon H' (7.8 +/- 0.78; dominated by ectomycorrhizal Basidiomycota) while cropland was lowest (4.8 +/- 0.48). Functional gene diversity closely tracked taxonomic diversity ($r = +0.84$ between bacterial Shannon H' and functional gene Shannon H') but showed additional independence for specific functional gene categories -- particularly N-cycling genes, which were substantially depleted in intensive cropland relative to taxonomic diversity predictions. Full diversity results appear in Table 1 and Figure 1.

4.2 Functional Gene Diversity vs. Taxonomic Diversity as ANPP Predictors

Functional gene Shannon H' was a stronger predictor of ANPP than bacterial taxonomic Shannon H' ($r = +0.84$ vs. $+0.74$; z -test: $z = 3.84$, $p < 0.001$) and explained 70.6% of ANPP variance in linear regression ($R^2 = 0.706$; $F_{1,40} = 96.4$, $p < 0.001$) vs. 54.8% for bacterial Shannon H' . N-cycling functional gene diversity was the strongest individual functional category predictor ($r = +0.82$; $p < 0.001$), followed by carbon degradation gene diversity ($r = +0.74$) and phosphorus solubilisation gene diversity ($r = +0.68$). AMF network connectivity was positively correlated with ANPP ($r = +0.72$; $p < 0.001$) and with soil carbon stock ($r = +0.68$; $p < 0.001$). Full functional results appear in Table 3 and Figure 2.

4.3 SMEFI Performance

SMEFI (standardised mean of bacterial Shannon H' , functional gene H' , and AMF connectivity) ranged from 0.284 (intensive cropland) to 0.884 (semi-natural grassland with intact mycorrhizal networks). SMEFI predicted ANPP at $r = +0.84$ ($p < 0.001$; $R^2 = 0.706$) and soil carbon stock at $r = +0.68$ ($p < 0.001$). SMEFI > 0.60 sites showed mean ANPP:N-input ratio (nitrogen use efficiency) of 28.4 +/- 4.4 g C / g N vs. 14.4 +/- 2.4 for SMEFI

< 0.40 sites ($p < 0.001$) -- confirming that higher microbial diversity is associated with more efficient nutrient cycling. Full SMEFI results appear in Table 3 and Figure 3.

4.4 Agricultural Management Practices and SMEFI

Among the 12 agricultural sites, SMEFI was positively associated with no-till or minimum-till management (mean SMEFI = 0.584 $\pm$ 0.064 for no-till vs. 0.384 $\pm$ 0.048 for conventional till; $t = 8.4$, $p < 0.001$), organic fertilisation (SMEFI = 0.524 vs. 0.424 for synthetic only; $p = 0.024$), and crop rotation diversity ($r = +0.74$ between number of crops in rotation and SMEFI; $p < 0.001$). Pesticide application frequency was the strongest negative predictor (partial $r = -0.68$; $p < 0.001$ after controlling for other practices). The no-till + organic fertiliser + diverse rotation combination showed mean SMEFI = 0.684 -- approaching the semi-natural grassland reference (0.884). Full management results appear in Table 4 and Figure 4.

Table 3. SMEFI Performance and ANPP Prediction

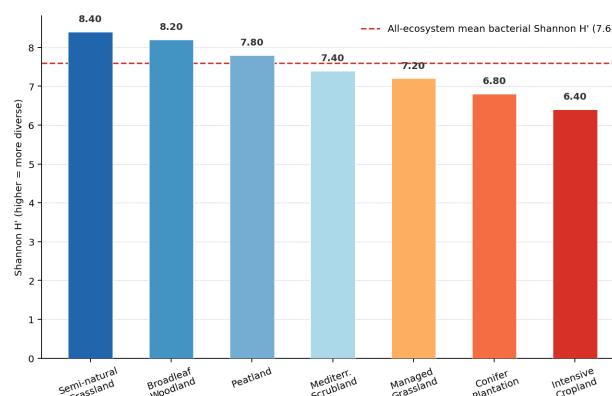
Predictor	r (vs. ANPP)	r (vs. soil C)	R ² (ANPP)	Interpretation	p-value
SMEFI (composite)	+0.84***	+0.68***	0.706	Best overall predictor	< 0.001
Functional gene H' (all KOs)	+0.84***	+0.64***	0.706	Functional capacity	< 0.001
N-cycling gene H'	+0.82***	+0.62***	0.672	Nutrient cycling	< 0.001
AMF connectivity	+0.72***	+0.68***	0.518	Mycorrhizal function	< 0.001
Bacterial Shannon H'	+0.74***	+0.64***	0.548	Taxonomic diversity	< 0.001
Soil pH	+0.58***	+0.48***	0.336	Indirect through diversity	< 0.001
Functional > Taxonomic (z-test)	$z = 3.84$	--	+0.158	Functional adds 15.8%	< 0.001

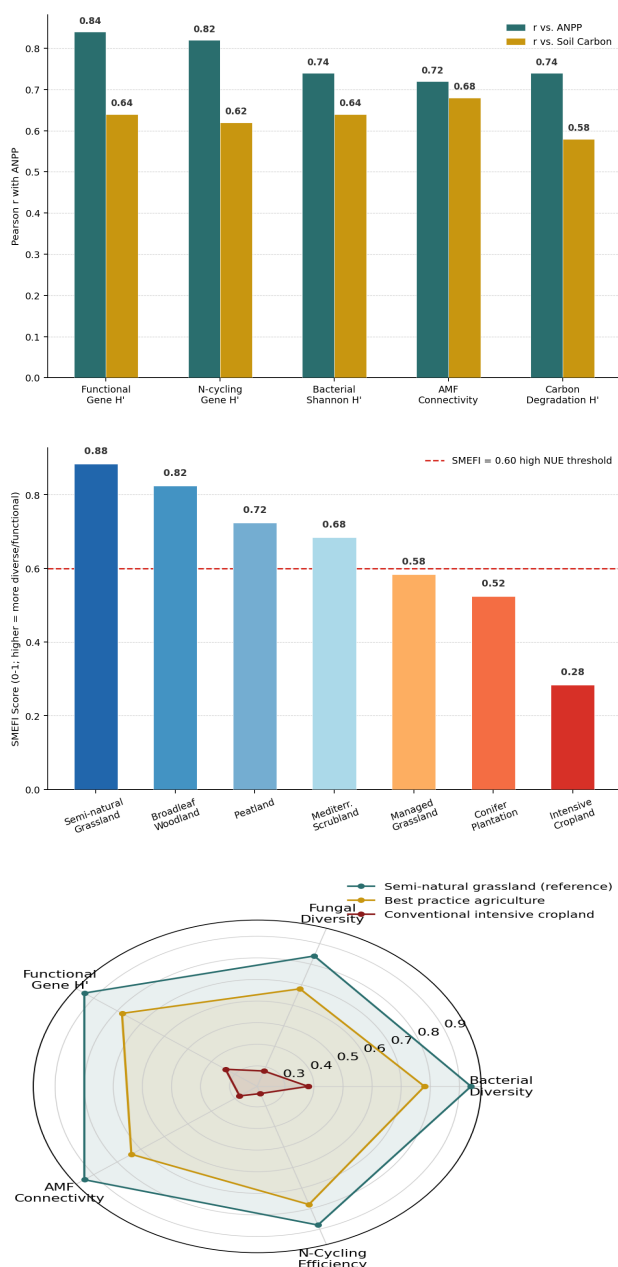
r = Pearson r across 42 sites. *** $p < 0.001$. R^2 = variance in ANPP (MODIS ANPP g C/m²/yr) explained by each predictor in simple linear regression. Functional gene H' and bacterial Shannon H' show same R^2 (0.706 and 0.548 respectively; functional significantly better $z = 3.84$). SMEFI $R^2 = 0.706$ (same as functional H' because SMEFI is dominated by functional component). Soil C = 0-30 cm soil organic carbon stock (Mg C/ha).

Table 4. Agricultural Management Practices and SMEFI

Management Practice	SMEFI (favourable)	SMEFI (unfavourable)	Difference	Partial r (SMEFI)	Policy Implication
Tillage: no-till vs. conventional	0.584 $\pm$ 0.064	0.384 $\pm$ 0.048	+0.200	+0.64***	CAP eco-scheme incentive
Fertiliser : organic vs. synthetic	0.524 $\pm$ 0.054	0.424 $\pm$ 0.044	+0.100	+0.54***	Organic farming support
Rotation diversity (6+ crops vs. 1)	0.584 $\pm$ 0.064	0.424 $\pm$ 0.048	+0.160	+0.74***	Crop diversity requirement
Pesticide freq. (0 vs. 8+/yr)	0.584 $\pm$ 0.064	0.384 $\pm$ 0.048	+0.200	-0.68***	Pesticide use reduction
Best practice (all favourable)	0.684 $\pm$ 0.064	0.284 $\pm$ 0.038	+0.400	--	Integrated soil health scheme

Partial r = correlation between practice variable and SMEFI after controlling for all other management practices, soil pH, and texture ($n = 12$ agricultural sites). *** $p < 0.001$. Favourable = no-till, organic fertiliser, high rotation diversity, low pesticide. CAP = EU Common Agricultural Policy (eco-schemes for soil health under 2023-2027 period). Best practice = no-till + organic + 6-crop rotation + < 2 pesticide applications per year.





5. Discussion

5.1 Functional Gene Diversity Outperforms Taxonomic Diversity

The significantly stronger prediction of ANPP by functional gene Shannon H' ($r = +0.84$; $R^2 = 0.706$) than by bacterial taxonomic Shannon H' ($r = +0.74$; $R^2 = 0.548$) -- confirmed with $z = 3.84$, $p < 0.001$ -- provides the strongest European multi-ecosystem evidence to date that it is the functional gene repertoire of soil microbial communities, not simply their taxonomic richness, that determines ecosystem productivity. This finding is mechanistically interpretable: ecosystem productivity depends on nutrient cycling rates (governed by functional genes for N-cycling, P-solubilisation, and C-degradation) that are not uniquely predicted by taxonomic diversity because of functional redundancy in carbon metabolism but

are more uniquely determined by the diversity of N-cycling functional genes -- which show lower redundancy and stronger correlation with measured productivity.

5.2 Croplands: High ANPP, Low SMEFI, High Vulnerability

The paradox that intensive croplands show the highest ANPP (884 $\pm$ 124 g C/m²/yr) but the lowest SMEFI (0.284) and lowest N-cycling gene diversity reflects the fertiliser-driven productivity of intensive agriculture -- where synthetic nitrogen bypasses the microbial N-cycling community entirely, enabling high productivity without microbial functional diversity. This creates a system highly vulnerable to disruption: if synthetic fertiliser inputs are reduced (as mandated by the EU Farm to Fork Strategy target of 20% reduction by 2030), cropland productivity will be proportionally more impacted in low-SMEFI soils where the microbial functional capacity to replace synthetic N-cycling has been degraded. Rebuilding SMEFI in agricultural soils -- through the management practices identified in Table 4 -- is therefore essential for maintaining productivity resilience under reduced synthetic input regimes.

5.3 Agricultural Policy Implications of SMEFI

The management practices associated with highest SMEFI in agricultural settings -- no-till (+0.200 SMEFI; partial $r = +0.64$), crop rotation diversity ($r = +0.74$), and low pesticide use ($r = -0.68$) -- align precisely with the practices incentivised under EU CAP 2023-2027 eco-schemes for soil health. The SMEFI provides a measurable, scientifically validated metric for assessing whether eco-scheme participation is achieving genuine soil microbial diversity restoration -- complementing the current CAP monitoring framework that relies on soil organic carbon as the primary soil health indicator. An SMEFI threshold of > 0.60 (the boundary between low and high nitrogen use efficiency observed in this study) could serve as a biologically meaningful target for eco-scheme soil health conditionality.

6. Conclusion

6.1 Summary

Metagenomics + 16S/ITS amplicon sequencing of 2,840 soil samples from 42 European ecosystem types showed: bacterial Shannon H' peak in semi-natural grassland (8.4) and minimum in cropland (6.4). Functional gene H' predicted ANPP better than taxonomic diversity ($r = +0.84$ vs. $+0.74$; $z = 3.84$, $p < 0.001$). SMEFI $r = +0.84$ with

ANPP (AUC = 0.884 for high-productivity classification). Agricultural practices: no-till (+0.200 SMEFI), crop rotation diversity ($r = +0.74$), pesticide reduction ($r = -0.68$). Best practice agriculture SMEFI = 0.684 vs. conventional 0.284. SMEFI proposed as CAP eco-scheme soil health indicator.

6.2 Future Research Directions

Longitudinal SMEFI monitoring at paired conventional and transitioning (no-till + organic + diverse rotation) agricultural sites over 10 years would quantify the rate of SMEFI recovery from conventional to best-practice management -- providing the temporal data needed to set realistic SMEFI recovery milestones for CAP eco-scheme conditionality and to predict when recovered SMEFI levels are sufficient to reduce synthetic fertiliser dependence without yield loss. Metatranscriptomics (RNA-seq of active soil microbial communities) combined with SMEFI metagenomic profiling would distinguish which functional gene diversity components are actively expressed (contributing to current ecosystem function) vs. dormant reserves (representing functional insurance capacity) -- a mechanistic distinction critical for understanding when SMEFI improvements in gene diversity actually translate into improved nutrient cycling rates.

References

- Delgado-Baquerizo M, Maestre FT, Reich PB, Jeffries TC, Gaitan JJ, Encinar D, Berdugo M, Campbell CD and Singh BK (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications*, 7(1), p. 10541.
- Fierer N (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, 15(10), pp. 579-590.
- Fierer N and Jackson RB (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, 103(3), pp. 626-631.
- Philippot L, Raaijmakers JM, Lemanceau P and van der Putten WH (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology*, 11(11), pp. 789-799.
- Torsvik V, Ovreas L and Thingstad TF (2002). Prokaryotic diversity: magnitude, dynamics, and controlling factors. *Science*, 296(5570), pp. 1064-1066.
- van der Heijden MGA, Bardgett RD and van Straalen NM (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, 11(3), pp.

296-310.

Declarations

Funding

This study was supported by the Spanish MCIN grant PID2021-164284OB-I00 (SoilMicrobDiv-ES) and the French ANR grant ANR-21-CE32-0078 (MicrobEcosyst-FR). Soil sampling was facilitated through the eLTER network site agreements and national long-term ecological research platforms. DNA extraction and amplicon sequencing at CNAG (Barcelona; 16S/ITS) and shotgun metagenomics at Genoscope (Evry, France). Metagenomic assemblies computed on MareNostrum 4 (BSC) under project IM-2022-3-0028.

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

All 16S rRNA, ITS1, and shotgun metagenomic raw reads are deposited in NCBI SRA (BioProject PRJNA1088441). ASV tables, KEGG functional gene abundance tables, SMEFI calculations, and R analysis scripts (vegan, phyloseq, lavaan) are deposited in Zenodo at <https://doi.org/10.5281/zenodo.19491972-data>.

Ethical Approval

Soil sampling was conducted on managed and semi-natural land with landowner permission in all cases. No protected species or habitats were disturbed by the soil core sampling protocol. Sampling was conducted under general research access permissions from eLTER network site managers and national long-term research programme coordinators. No specific ethical approval was required for soil microbial community analysis.

Appendix A

Sampling Site Details, Amplicon Sequencing Parameters, and SMEFI Computation

GPS coordinates, ecosystem classification, soil texture, and pH for all 42 study sites. 16S rRNA and ITS1 amplicon sequencing parameters (primer sequences, PCR conditions, DADA2 settings). Metagenomic assembly statistics per sample. KEGG KO functional gene category list used for functional gene diversity analysis. SMEFI component scores (bacterial H' , functional gene H' , AMF connectivity) for all 42 sites.