

Review

Current State-of-the-Art of NGS in Soil Microbial Ecology Interpreted Through the Hierarchical Environmental Filtering (HEF) Framework

Larysa Kucher ¹, Liudmyla Kava ² , Liliana Cepoi ³ , Olena Myronycheva ⁴ , Burkhard Kirchhoff ⁵,
Kateryna Davydenko ⁶  and Andrii P. Gryganskyi ^{7,*} 

- ¹ Department of Agrobiolgy, National University of Life and Environmental Sciences of Ukraine (NUBiP), 03041 Kyiv, Ukraine; lora_kucher@ukr.net
² Department of Plant Protection, National University of Life and Environmental Sciences of Ukraine (NUBiP), 03041 Kyiv, Ukraine; kavalyuda@ukr.net
³ Institute of Microbiology and Biotechnology, Technical University of Moldova, 2028 Chişinău, Moldova; liliana.cepoi@imb.utm.md
⁴ Department of Engineering Sciences and Mathematics, Luleå University of Technology, 931 87 Skellefteå, Sweden; olena.myronycheva@ltu.se
⁵ Kirchhoff GmbH & Co. KG, 31863 Coppenbrügge-Bisperode, Germany; b.kirchhoff@biogasgutachter.de
⁶ Department of the Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden; kateryna.davydenko@slu.se
⁷ AeroVironment Inc., Dayton, OH 45458, USA
* Correspondence: andrii.gryganskyi@gmail.com

Abstract

Next-generation sequencing (NGS) has transformed soil microbial ecology by revealing taxonomic and functional diversity that was largely inaccessible through cultivation-based approaches. However, greater sequencing resolution alone does not explain why particular microbial assemblages repeatedly emerge under specific soil and environmental conditions. This structured narrative conceptual review examines the current state of NGS-based soil microbiome characterization and introduces Hierarchical Environmental Filtering (HEF) as a framework for interpreting microbial community assembly. We discuss advances from amplicon sequencing to shotgun metagenomics, long-read sequencing, multi-omics, and computational analysis, together with limitations related to sampling, DNA extraction, primer selection, sequencing depth, bioinformatics, reference databases, and relic DNA. Within HEF, pedogenesis establishes a historically contingent physicochemical template, whereas contemporary climate, vegetation, rhizosphere processes, biotic interactions, dispersal, and stochastic processes modify assembly within that template. Anthropogenic disturbances can alter several levels simultaneously and may partially override inherited soil constraints. Soil microbiomes should therefore be interpreted as dynamic outcomes of environmental selection across spatial and temporal scales. Integrating standardized NGS workflows with functional multi-omics and predictive computation should advance soil microbiome research from descriptive inventories toward a mechanistic understanding of community assembly.

Keywords: microbiome assembly; pedogenesis; metagenomics; edaphic gradients; anthropogenic disturbance; multi-omics; soil health



Academic Editor: Ipek Kurtboke

Received: 11 August 2026

Revised: 2 September 2026

Accepted: 7 September 2026

Published: 10 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

For more than a century, soil classification has relied primarily on morphological, physical, and chemical characteristics that reflect the long-term processes of pedogenesis. These properties remain the foundation of modern soil taxonomy because they integrate the cumulative effects of parent material, climate, organisms, topography, and time during soil development [1]. However, soils are also among the most biologically diverse ecosystems on Earth. Their microbial communities regulate organic matter decomposition, nutrient cycling, soil aggregation, greenhouse gas fluxes, plant productivity, and numerous other ecosystem functions [2–4]. Until relatively recently, this biological dimension remained only partially understood because a large proportion of soil microorganisms could not be readily cultivated or characterized using conventional microbiological techniques [2].

The development of next-generation sequencing (NGS) fundamentally transformed soil microbial ecology by enabling the characterization of entire microbial communities directly from environmental DNA. Amplicon sequencing, shotgun metagenomics, and related high-throughput approaches have revealed extraordinary taxonomic and functional diversity across soils and have substantially expanded understanding of microbial biogeography, community structure, and functional potential [5–8]. These studies have also demonstrated that soil microbial communities are neither randomly assembled nor uniformly distributed across landscapes. Instead, their composition reflects reproducible associations with soil physicochemical properties, climate, vegetation, land use, geographic location, and other environmental conditions [2–4,6].

Nevertheless, increasing sequencing resolution does not by itself explain why particular microbial assemblages repeatedly emerge under comparable environmental conditions. Taxonomically different communities may possess overlapping functional capabilities, whereas communities with apparently similar taxonomic structures may differ in their ecological functions and responses to environmental change [9]. Furthermore, the ecological patterns inferred from sequencing data can be affected by sampling design, DNA extraction, primer selection, sequencing depth, bioinformatic processing, reference-database coverage, contamination, and the persistence of extracellular relic DNA [7,10–12]. Such methodological and conceptual limitations become particularly important in large-scale comparisons, where technical variability may obscure or be mistaken for genuine ecological variation [13]. Consequently, the principal challenge is no longer simply to catalogue the microorganisms present in soil but to determine how environmental constraints, biological interactions, historical processes, and anthropogenic disturbances jointly generate reproducible patterns of community organization and ecosystem functioning [2,4,14].

Addressing this challenge requires the integration of molecular approaches with concepts from soil science, community ecology, and ecosystem theory. In this review, we introduce Hierarchical Environmental Filtering (HEF) as a conceptual framework for synthesizing independent observations concerning soil microbial community assembly. The framework considers pedogenesis as establishing the fundamental physicochemical template upon which additional climatic, vegetational, biological, and anthropogenic filters operate. Rather than focusing solely on sequencing technologies, we examine how modern NGS approaches have transformed soil microbiome characterization, evaluate their principal methodological and interpretive limitations, and explore how hierarchical environmental filtering can provide an ecological context for interpreting NGS-derived patterns.

The objectives of this review are therefore fourfold: (i) to evaluate how next-generation sequencing has transformed the characterization of soil microbial communities; (ii) to examine the ecological mechanisms responsible for the assembly, persistence, and variability of these communities; (iii) to assess the limitations imposed by environmental variability, anthropogenic disturbance, sampling design, and methodological uncertainty; and (iv) to

explore how integrating NGS with ecological theory, functional approaches, and computational analysis can improve the interpretation and monitoring of soil microbial systems.

Environmental filtering is an established concept in community ecology and generally refers to the deterministic exclusion of organisms whose ecological traits do not permit their establishment or persistence under particular abiotic conditions [15,16]. In microbial ecology, environmental selection is similarly recognized as one of several processes governing community assembly, operating together with dispersal, ecological drift, diversification, and historical contingency [17–19]. Evidence of compositional differences among environments therefore does not, by itself, demonstrate environmental filtering, because comparable patterns may also arise through biotic interactions, dispersal limitation, stochastic processes, or methodological effects [16]. In the present review, we consequently use environmental filtering in its restricted sense for abiotic constraints and distinguish it from subsequent community modification through vegetation, host associations, microbial interactions, dispersal, and disturbance.

The Hierarchical Environmental Filtering framework proposed here does not introduce environmental filtering as a new ecological mechanism, nor does it assume that soil microbial communities are assembled through a rigid universal sequence. Instead, HEF is a soil-specific interpretive framework that organizes established assembly processes according to their causal precedence, spatial scale, temporal persistence, and potential reversibility. It treats pedogenesis and soil-horizon development as relatively persistent boundary conditions that define the physicochemical habitat and restrict the range of microbial assemblages that can become established. Within this constrained ecological space, climate and vegetation operate across broad to intermediate scales, whereas rhizosphere processes, biological interactions, dispersal, and short-term environmental variation further modify community composition at finer spatial and temporal scales. Anthropogenic disturbances constitute cross-cutting forces that may modify individual filters or, when sufficiently intense, partially override the pedogenic template and redirect community assembly toward alternative states.

Climate operates within HEF through historically accumulated and contemporary pathways. Over pedogenic timescales, climate influences weathering, horizon development, organic-matter accumulation and nutrient availability and is therefore embedded in the inherited soil template. At ecological timescales, current temperature, precipitation and soil-moisture regimes act directly on microbial growth, dormancy, dispersal and activity. Climatic extremes may produce short-term community reassembly, whereas persistent directional climate change can gradually modify vegetation, hydrology and soil properties and thereby alter the template itself. These pathways should be distinguished analytically to avoid treating climate simultaneously as both the origin and an independent component of the same environmental effect.

The relationships among the components of HEF are summarized in Figure 1. Although the framework is described hierarchically, it should not be understood as a rigid or strictly unidirectional sequence. Historically accumulated pedogenic constraints delimit the range of contemporary soil conditions, but current climate, vegetation, microbial activity and anthropogenic disturbance can modify those conditions through feedbacks operating over different timescales. In parallel, dispersal, priority effects, ecological drift and biotic interactions determine which members of the regional species pool reach, establish and persist within individual soils. HEF therefore describes nested and scale-dependent constraints whose relative importance varies among ecosystems, microbial groups, spatial scales and periods of observation.

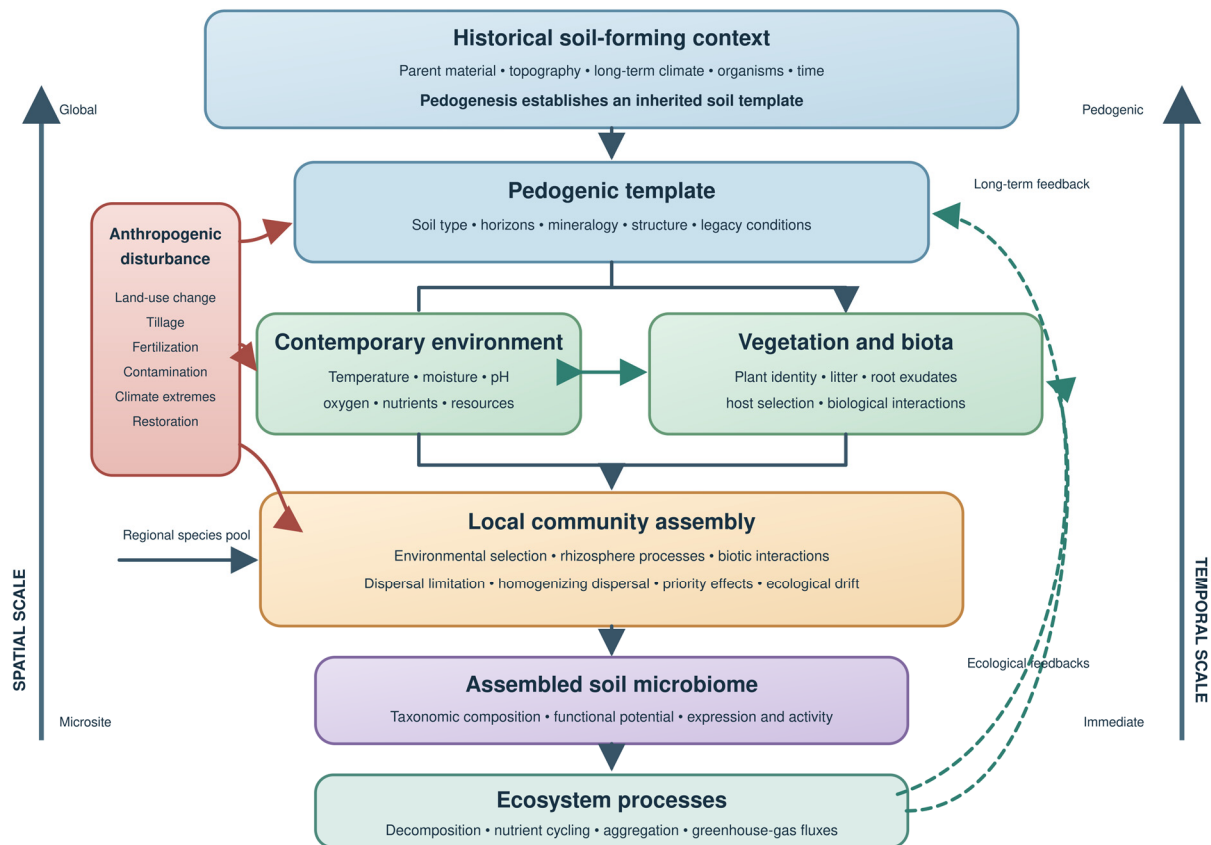


Figure 1. Conceptual representation of Hierarchical Environmental Filtering (HEF) in soil microbial community assembly. Pedogenesis generates a historically contingent soil template through the long-term interaction of parent material, topography, climate, organisms, and time. Contemporary climate, vegetation, resource availability, and soil physicochemical conditions subsequently operate within—but can also modify—this inherited template. At finer scales, rhizosphere processes and biotic interactions influence local establishment and persistence, while dispersal limitation, homogenizing dispersal, priority effects, and ecological drift generate community patterns that are not attributable solely to environmental selection. Anthropogenic disturbances can affect several levels simultaneously by modifying soil structure, resource availability, vegetation, dispersal, and local biological interactions. The assembled microbiome influences ecosystem processes, which feed back to vegetation, contemporary soil conditions, and, over longer periods, pedogenesis. The framework therefore represents a nested and interactive system rather than a strictly unidirectional sequence. The surrounding gradients indicate that different processes predominate across spatial scales, from global and regional constraints to soil horizons and rhizosphere microsites, and across temporal scales, from pedogenic history to seasonal variation and short-term disturbance.

HEF therefore differs from a simple sequential list of environmental drivers in four respects. First, the filters are nested rather than merely additive: the effect of a finer-scale driver is conditional on the physicochemical and biological context established at broader or more persistent levels. Second, the framework distinguishes persistent habitat constraints from dynamic ecological modifiers and transient disturbances. Third, it accommodates both deterministic and stochastic assembly by proposing that dispersal and ecological drift operate within, and may sometimes obscure, the range of states permitted by environmental constraints. Fourth, it explicitly connects this hierarchy to the interpretation of NGS data, helping distinguish methodological variation and short-term community fluctuations from reproducible microbial patterns associated with soil formation and ecosystem development.

This organization generates testable expectations. Microbial communities occurring in comparable soil horizons and pedogenetic settings should exhibit greater taxonomic or

functional convergence than communities compared without accounting for soil context; the explanatory importance of individual drivers should change with spatial and temporal scale; and sufficiently strong disturbances should weaken associations between microbiome composition and pedogenic characteristics. HEF is thus presented as an integrative and hypothesis-generating framework for interpreting sequencing-derived patterns rather than as a replacement for existing theories of microbial community assembly.

2. Materials and Methods

2.1. Review Design and Scope

This study was conducted as a structured critical narrative review of current next-generation sequencing (NGS) approaches in soil microbial ecology and their interpretation within the proposed Hierarchical Environmental Filtering (HEF) framework. The review was designed to integrate methodological, ecological, and pedological evidence rather than to estimate a pooled effect size. Accordingly, no statistical meta-analysis was performed. The literature search and selection process was documented using relevant reporting principles from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework to improve transparency [20]; however, the study was not designed as a formal systematic review or meta-analysis.

The review addressed three principal questions: (i) how have NGS technologies changed the characterization of soil microbial communities; (ii) which environmental and biological factors explain the assembly, persistence, and variability of these communities; and (iii) how can the available evidence be integrated into a hierarchical conceptual model connecting pedogenesis, climate, vegetation, biological interactions, and anthropogenic disturbance? The scope included studies of bacterial, archaeal, and fungal communities in soils, as well as methodological studies evaluating sampling, DNA extraction, marker selection, sequencing platforms, bioinformatic processing, reference databases, and functional interpretation.

2.2. Information Sources and Literature Search

The primary structured search was conducted in the Web of Science Core Collection. PubMed and Google Scholar were used as complementary sources to identify methodological publications and relevant studies not retrieved by the primary query. Reference lists of included publications were examined to locate foundational studies, and forward citation searching was used to identify later methodological developments and applications.

The search strategy combined terms describing soil microbial communities with terms related to sequencing technologies and microbial community assembly. The principal combinations included: (“soil microbiome” OR “soil microbial community” OR “soil bacteria” OR “soil fungi” OR “soil archaea”) AND (“next-generation sequencing” OR “high-throughput sequencing” OR NGS OR metabarcoding OR “amplicon sequencing” OR metagenomics OR “shotgun metagenomics” OR “long-read sequencing” OR metatranscriptomics OR metaproteomics OR multi-omics). Complementary searches combined the soil microbial terms with ecological and methodological concepts, including “environmental filtering”, pedogenesis, “soil type”, “soil horizon”, “sampling depth”, climate, moisture, vegetation, rhizosphere, “land use”, agriculture, disturbance, “DNA extraction”, “primer bias”, “sequencing depth”, bioinformatics, “reference database”, “relic DNA”, “core microbiome”, “soil health”, and “machine learning”.

Search terms were adapted to the syntax and search capabilities of each database. Searches in Web of Science were applied to the Topic field, encompassing titles, abstracts, author keywords, and Keywords Plus, whereas PubMed searches were applied to titles, abstracts, and indexed terms. Google Scholar was searched using shorter combinations

of the principal terms because of limitations in the length and reproducibility of complex Boolean queries. The primary search was completed on 1 September 2026 and covered publications from 2015 through 2026. Earlier foundational studies were eligible when identified through complementary searches or citation chaining and were required to define major concepts or methodological developments.

2.3. Eligibility Criteria and Study Selection

Publications were considered eligible when they met at least one of the following criteria: (i) applied NGS or related molecular approaches to characterize soil bacterial, archaeal, or fungal communities; (ii) evaluated methodological factors influencing soil microbiome analysis; (iii) investigated the effects of soil properties, pedogenesis, climate, vegetation, rhizosphere processes, biological interactions, land use, or anthropogenic disturbance on microbial community assembly; or (iv) contributed conceptual or analytical approaches relevant to the ecological interpretation of soil microbiome data. Primary research articles were prioritized, although authoritative reviews and database or software publications were included when they provided essential methodological context or supported the development of the conceptual framework.

Publications were excluded when they: (i) did not examine soil or a directly soil-associated compartment; (ii) focused exclusively on cultivation-based microbiology without relevance to molecular community characterization; (iii) reported sequencing results without sufficient ecological or methodological information for the objectives of this review; (iv) were unrelated to microbial community assembly, sequencing methodology, soil properties, or ecosystem functioning; or (v) duplicated information reported in another version of the same study. Studies of sediments, aquatic microbiomes, animal microbiomes, and built environments were excluded unless they contained methodological evidence directly applicable to soil microbiome analysis.

The Web of Science search returned 632 records. Titles and abstracts were screened against the eligibility criteria, producing a focused shortlist of 65 publications for detailed assessment and citation chaining. Potentially eligible publications were subsequently evaluated from their full texts. Additional publications identified through PubMed, Google Scholar, reference-list screening, and forward citation searches were subjected to the same eligibility criteria (Table S1). Uncertainties concerning eligibility or interpretation were discussed among the authors and resolved by consensus.

2.4. Data Extraction and Critical Evaluation

For each eligible publication, information relevant to the objectives of the review was extracted and organized according to the study context, microbial group, sequencing or analytical approach, principal environmental variables, major ecological findings, and reported methodological limitations. When available, particular attention was given to soil type, soil horizon and sampling depth, climate, moisture regime, vegetation, rhizosphere or bulk-soil origin, land use, disturbance history, DNA extraction procedure, molecular marker, sequencing platform, sequencing depth, taxonomic database, and bioinformatic workflow.

Methodological publications were evaluated according to their relevance to the major sources of uncertainty in soil microbiome characterization, including sampling design, laboratory processing, amplification bias, sequencing coverage, taxonomic assignment, computational analysis, and the contribution of relic DNA. Ecological studies were evaluated according to whether they provided evidence for reproducible relationships between microbial communities and pedogenetic, climatic, vegetation-related, biological, or anthropogenic factors. Because the included publications differed substantially in experimental design, geographic scale, sequencing platform, molecular targets, and reported outcomes,

formal quantitative comparison and meta-analysis were not considered appropriate. Evidence was therefore synthesized qualitatively and critically.

2.5. Conceptual Synthesis and Development of the HEF Framework

The selected evidence was organized through an iterative conceptual synthesis. Reported drivers of soil microbial community composition were first grouped according to their spatial and temporal scales and their relationship to soil formation and ecosystem functioning. Evidence was subsequently classified into four principal levels of environmental filtering: (i) pedogenesis and the resulting soil physicochemical template; (ii) broad-scale climatic and biogeographic constraints; (iii) contemporary biological controls, including vegetation, rhizosphere processes, and microbial interactions; and (iv) anthropogenic modification or overriding of natural environmental filters through land use, agricultural management, nutrient enrichment, and other disturbances.

Methodological sources of variation, including sampling depth, sampling period, DNA extraction, marker choice, sequencing depth, bioinformatic processing, and reference-database selection, were evaluated separately because they influence the observed representation of microbial communities rather than community assembly itself. The final HEF framework was developed by integrating recurring patterns across the selected studies and identifying areas of agreement, uncertainty, and methodological limitation. The framework is therefore presented as a conceptual interpretation of the available evidence rather than as a statistically validated predictive model.

2.6. Limitations of the Review Method

The literature-selection strategy may have been affected by database coverage and search-term sensitivity. Relevant studies may not have been retrieved when they used highly specific terminology without explicitly referring to soil microbiomes, NGS, microbial community assembly, or environmental filtering. Google Scholar improved the breadth of retrieval but offers less reproducible search results than curated bibliographic databases. The inclusion of heterogeneous observational, experimental, methodological, and conceptual publications also prevented formal meta-analysis and standardized risk-of-bias assessment. Finally, the categorization of environmental drivers within the HEF framework involved interpretive synthesis; therefore, boundaries among pedogenetic, climatic, biological, and anthropogenic filters should be regarded as conceptually useful but not mutually exclusive.

3. Sequencing Strategies and Interpretive Limits in Soil Microbiome Research

3.1. Comparison of Sequencing Strategies

The choice of sequencing strategy determines which components of microbial community structure can be resolved and, consequently, which levels of environmental filtering can be evaluated. Targeted amplicon sequencing supports cost-effective comparison of large sample sets but provides limited functional information and taxonomic resolution. Short-read shotgun metagenomics captures both taxonomic and functional profiles, although fragmented assemblies can restrict genome reconstruction and the assignment of genes to particular organisms. Long-read PacBio HiFi and Oxford Nanopore sequencing can improve taxonomic resolution, genome assembly and linkage of functional genes to their genomic contexts, but they differ in accuracy, read length, throughput, sample requirements and analytical workflows [21–23]. These distinctions are summarized in Table 1.

Table 1. Comparison of principal sequencing strategies used in soil microbial community studies.

Characteristic	Short-Read Amplicon Sequencing	Short-Read Shotgun Metagenomics	PacBio HiFi Sequencing	Oxford Nanopore Sequencing
Typical target	Selected taxonomic marker, commonly bacterial and archaeal 16S rRNA genes, fungal ITS regions or eukaryotic 18S rRNA genes	Total extracted community DNA	Full-length marker genes, long metagenomic fragments or reconstructed microbial genomes	Full-length marker genes, long to ultra-long metagenomic fragments, native DNA or RNA
Typical read length	Commonly 150–300 bp from each end; paired reads may cover one or more selected marker regions	Commonly 100–300 bp from each end	Long, highly accurate consensus reads, commonly several kilobases and potentially extending to approximately 15–25 kb depending on library preparation	Read length is largely determined by input-fragment length; reads may range from several kilobases to more than 100 kb, with ultra-long reads possible
Relative sequencing depth	Very high numbers of marker reads per sample; allows extensive sample multiplexing	High total read numbers, although much of the data may originate from abundant organisms or non-target DNA	Generally fewer reads than short-read platforms, but each read contains substantially more contiguous sequence information	Scalable according to device and flow-cell format; sequencing can continue until the desired depth is obtained
Taxonomic resolution	Usually genus- or family-level for short 16S/18S regions; species-level assignments are possible only for some taxa and markers	Potential species- and strain-level resolution when reference coverage and sequencing depth are adequate	Higher resolution than short-region amplicon sequencing; species-level assignment is possible for some taxa but is not guaranteed and depends on marker variability, sequence quality and reference-database coverage.	Full-length markers and long shotgun reads can provide species- or strain-level resolution, depending on read accuracy, depth and database quality
Functional information	Primarily indirect and prediction-based; restricted to functions inferred from taxonomic composition	Direct detection of genes and metabolic pathways, although short reads may not preserve linkage among genes, taxa and mobile elements	Direct functional profiling with improved genomic context; long reads can span multiple genes, operons and repetitive regions	Direct functional profiling with strong physical linkage among genes, genomes, plasmids and other mobile elements
Error characteristics	Low per-base error after quality filtering, but susceptible to PCR, primer, chimera and marker-copy-number biases	Low per-base error; susceptible to library-composition bias, uneven genome coverage and assembly ambiguity	High single-read consensus accuracy, typically above Q30 under appropriate conditions; residual errors and extraction bias remain relevant	Raw-read accuracy depends on chemistry, base-calling model and read mode; consensus polishing can substantially improve assembled-sequence accuracy
Major analytical requirements	Primer removal, denoising, chimera detection, ASV or OTU inference and taxonomic classification	Quality control, removal of host or plant DNA, taxonomic profiling, assembly, binning, gene prediction and functional annotation	HiFi-read processing, long-read assembly, metagenome binning, polishing and comprehensive taxonomic and functional annotation	Base calling, read-quality filtering, long-read classification or assembly, polishing and detection of structural variants or base modifications
Principal strengths	Economical; well established; suitable for large replicated surveys and comparison of many environmental treatments	Simultaneous taxonomic and functional characterization without targeting a single marker gene	Combines long reads with high consensus accuracy; supports full-length amplicons and high-quality metagenome-assembled genomes	Very long reads, real-time analysis, flexible run duration, scalable instrumentation and potential field deployment; can directly sequence native DNA or RNA
Principal limitations	Primer-dependent detection; compositional data; limited functional inference; variable taxonomic resolution; different marker regions may be poorly comparable	Greater cost and computational demand; incomplete recovery of rare populations; fragmented assemblies in highly complex soils	Higher DNA-quality requirements and generally higher cost per sample than routine short-read amplicon sequencing; fewer samples may be multiplexed at equivalent depth	Raw-read errors may complicate fine-scale taxonomic and gene-level inference without adequate depth or polishing; results depend strongly on DNA integrity and base-calling workflow

Table 1. Cont.

Characteristic	Short-Read Amplicon Sequencing	Short-Read Shotgun Metagenomics	PacBio HiFi Sequencing	Oxford Nanopore Sequencing
Relevance to HEF testing	Particularly suitable for replicated assessment of compositional turnover across hierarchical environmental gradients	Allows environmental filters to be related to both taxonomic composition and functional potential	Improves recognition of closely related taxa and functional adaptations responding differently to particular filters	Enables resolution of genomic context, strain variation and mobile elements that may mediate rapid responses to environmental or anthropogenic filters
Abundance interpretation	Primarily compositional; read proportions do not directly represent cell numbers	Primarily compositional and affected by genome size, extraction efficiency and database representation	Compositional unless calibrated using quantitative measurements or internal standards	Compositional unless calibrated; real-time read accumulation does not itself provide absolute biological abundance

Note: Read length, throughput, accuracy, and cost depend on instrument, chemistry, library preparation, and bioinformatic workflow and should be interpreted as general characteristics rather than fixed platform specifications. PacBio HiFi and Oxford Nanopore technologies can support both targeted amplicon sequencing and shotgun metagenomics.

No single sequencing strategy is optimal for every HEF study. Short-read amplicon sequencing remains particularly useful when statistical power depends on the inclusion of many independent samples across nested spatial or environmental levels. Shotgun metagenomics is more appropriate when HEF predictions concern functional potential in addition to taxonomic composition. Long-read approaches offer important advantages when closely related taxa, complete metabolic pathways, operons, mobile genetic elements or metagenome-assembled genomes must be resolved. Hybrid designs that combine deep short-read coverage with long-read genomic continuity may provide the most comprehensive characterization, but their increased expense and analytical complexity can reduce the number of biological replicates. Because robust testing of hierarchical filtering depends primarily on appropriate sampling, replication and measurement of environmental covariates, greater sequence resolution cannot compensate for an inadequate ecological design.

3.2. Distinguishing Presence, Functional Potential, and Activity

The ecological interpretation of soil microbiome data depends on the biological layer measured. Detection of a taxon or gene does not demonstrate that the corresponding organism is viable, metabolically active, or contributing to an ecosystem process at the time of sampling. Amplicon sequencing primarily describes taxonomic composition, and functions inferred from marker-gene data remain predictions based on reference genomes. Shotgun metagenomics directly detects community genes and pathways but describes functional potential rather than their expression or activity [7,21].

Metatranscriptomics provides evidence of recent transcription, whereas metaproteomics identifies expressed protein machinery and metabolomics measures the integrated pool of compounds produced, consumed, transported, or degraded within soil. These layers are complementary but not interchangeable: transcript abundance need not predict protein abundance or reaction rate, proteins may persist after expression, and metabolites may have microbial, plant, or abiotic origins. Strong attribution of an in situ function therefore requires complementary process measurements, enzyme assays, isotope tracing, or stable-isotope probing that links substrate use to particular taxa or molecules.

Relic DNA further separates molecular detection from biological activity because soil sequences may originate from extracellular DNA or non-intact cells. Its removal can improve characterization of the viable community, but DNA from an intact cell still does not establish metabolic activity because that cell may be dormant [10]. Throughout this review, we therefore distinguish taxonomic presence, functional potential, molecular expression,

biochemical activity, and ecosystem process rates. An extended comparison of sequencing and complementary analytical layers is provided in Table S2.

3.3. Compositionality and Quantitative Interpretation of Sequencing Data

Microbiome sequencing generally provides compositional rather than absolute-abundance data [24]. Because the number of reads generated for a sample is determined by sequencing depth and bioinformatic processing, taxon counts are constrained by an arbitrary library total and are usually interpreted as proportions of the recovered community. Consequently, an increase in the relative abundance of one taxon does not necessarily indicate an increase in its absolute population size; it may instead result from a decline in other community members or a change in total microbial biomass. Likewise, a taxon can decrease in relative abundance even when its absolute abundance remains constant or increases. These limitations are particularly important when HEF is used to interpret microbial responses to disturbance, because drought, contamination, nutrient enrichment or land management may alter total microbial biomass as well as community composition.

Analytical approaches should therefore account explicitly for compositionality. Appropriate options include log-ratio methods based on Aitchison geometry, such as centered log-ratio or isometric log-ratio transformations, analysis of taxon ratios or balances, and calculation of Aitchison distances for multivariate comparisons. Differential-abundance methods designed for compositional and sparse microbiome data, including ALDEx2, ANCOM, ANCOM-BC, or comparable validated approaches [25], can also reduce misleading inferences produced by direct comparison of proportions. The selection of zero-handling procedures, prevalence filters and reference denominators should be reported because these decisions can materially affect the results. Rarefaction or conversion to percentages may standardize library size for particular purposes, but neither procedure resolves the fundamental compositional constraint nor converts sequence counts into absolute biological abundance.

Where absolute abundance is ecologically important, sequencing should be complemented by independent quantitative measurements. Total or taxon-specific gene abundance can be estimated using quantitative PCR or digital PCR, whereas flow cytometry, phospholipid fatty-acid analysis or direct cell counting may provide complementary estimates of microbial biomass or cell density. Known quantities of exogenous cells or synthetic DNA added before extraction can serve as internal standards and permit sequence data to be calibrated against sample mass or volume [26]. Standards added before DNA extraction can additionally account for some variation in extraction recovery, whereas standards introduced only during library preparation cannot correct for losses occurring during extraction. Quantitative estimates nevertheless require careful interpretation because marker-gene copy number varies among taxa, DNA extraction efficiency differs among soil types, extracellular DNA may be retained, and gene-copy abundance is not necessarily equivalent to the abundance of viable or metabolically active cells. Accordingly, compositional analyses are appropriate for evaluating relative community restructuring, whereas claims concerning population growth, decline or microbial load require complementary quantitative evidence. Distinguishing compositional turnover from changes in absolute microbial abundance is essential for testing HEF, because an environmental filter may alter the relative structure of a community, its total microbial load, or both. Moreover, quantitative calibration does not eliminate taxon-specific extraction, primer or amplification biases; compositional and technical sources of uncertainty must therefore be evaluated independently.

3.4. Technical Bias, Quality Control, and Reproducibility

Soil microbiome profiles reflect not only ecological variation but also biases introduced during sampling, DNA extraction, amplification, sequencing and bioinformatic processing. These effects become particularly important when microbiome signatures are proposed as indicators of soil condition or as diagnostic tools, because a technically generated difference may be mistaken for an environmental response. Standardized procedures, appropriate controls and transparent reporting are therefore required throughout the analytical workflow.

Marker-gene copy number varies substantially among microbial taxa [27]. Bacterial and archaeal genomes may contain different numbers of 16S rRNA gene copies, while fungi vary in the number of ribosomal operons and ITS copies. Consequently, sequence proportions do not necessarily correspond directly to organismal or cellular proportions, and taxa with high marker-copy numbers may appear overrepresented. Copy-number correction based on reference genomes may reduce this effect for well-characterized groups but remains uncertain for poorly represented soil microorganisms and should not be treated as an exact reconstruction of cell abundance.

PCR and library construction introduce systematic biases through differential primer affinity, template competition, amplification efficiency, chimera formation and the number of amplification cycles. No primer set captures all microbial lineages equally, and the selected 16S rRNA gene region or fungal ITS region influences both the taxa detected and the taxonomic resolution obtained. Primer choice should therefore be matched to the target microbial group and documented in sufficient detail. Technical replicates, minimized amplification-cycle numbers and validated high-fidelity polymerases can improve reproducibility but cannot completely eliminate amplification bias [28,29].

Soils also contain humic acids, phenolic compounds, clay-associated substances, metals and other compounds that may inhibit DNA polymerases or interfere with DNA extraction and library preparation. Inhibition can vary systematically among soil types and treatments, potentially creating an apparent ecological pattern. DNA purity measurements alone may not reliably reveal inhibition; dilution tests, amplification controls or the addition of known internal amplification controls can provide more direct evidence. DNA purification procedures should be optimized carefully because additional purification may remove inhibitors while also reducing DNA yield or altering recovery among microbial groups.

Batch effects can arise from differences among sampling dates, storage conditions, extraction kits, operators, reagent lots, PCR plates, library-preparation runs, sequencing runs and bioinformatic versions. Ecological treatments should therefore be randomized or balanced across extraction, amplification and sequencing batches wherever possible. Batch identity should be recorded and evaluated statistically rather than being confounded with ecosystem type, treatment or sampling location. Reprocessing all samples through a consistent bioinformatic workflow is also important when datasets generated at different times are compared.

Extraction blanks, PCR-negative controls and library blanks are necessary to detect contamination originating from reagents, laboratory handling or sequencing. Their importance increases for low-biomass samples, in which contaminant DNA can represent a substantial fraction of the recovered sequences. Negative controls should be sequenced and analyzed alongside biological samples rather than assessed only by the absence of a visible PCR product. Positive controls and mock communities with known composition provide complementary information about extraction efficiency, primer coverage, amplification distortion, taxonomic misclassification and sequencing error. However, mock communities should include organisms or DNA characteristics representative of the intended samples

because a simple commercial standard may not reproduce the extraction difficulty and taxonomic complexity of soil.

For diagnostic applications, quality-control criteria should be defined before samples are classified. These may include minimum DNA yield, inhibition thresholds, sequencing depth, control performance, reproducibility among technical replicates and rules for identifying or removing contaminants. Candidate microbial indicators should subsequently be validated using independent sample sets, different laboratory batches and, preferably, different laboratories. Without such validation, a microbiome signature may reflect a study-specific extraction protocol, primer pair, sequencing run or bioinformatic pipeline rather than a reproducible response to an environmental filter. Within the HEF framework, these technical processes can be considered an observation layer rather than ecological filters: they determine which components of the underlying community are recovered and may modify the apparent strength or order of environmental associations. Ecological and molecular sources of variation, their interpretive risks, and corresponding methodological responses are summarized in Table S3.

4. Hierarchical Environmental Controls on Soil Microbial Assembly

4.1. Pedogenic Context and Contemporary Climate

Although pedogenesis establishes the physicochemical template upon which soil microbial communities assemble, it does not fully determine community composition. Even within soils sharing a common pedogenetic origin, microbial communities may differ substantially across soil horizons because ecological selection continues under local physicochemical conditions after soil formation [30]. Once the structural environment has been established, microbial communities are subsequently shaped by a second level of ecological filtering imposed by climate, vegetation, rhizosphere processes, and biotic interactions. These secondary filters do not replace pedogenetic constraints; rather, they refine the ecological space created during soil formation, thereby shaping the dynamic component of community assembly.

Secondary ecological controls operate across different spatial and temporal scales and differ in their mechanisms, persistence, and expected effects on sequencing-derived microbial patterns. Their principal characteristics and their conditional relationships with the pedogenic template are summarized in Table 2.

Table 2. Secondary ecological controls refining soil microbial community assembly within the physicochemical boundaries established by pedogenesis.

Secondary Ecological Control	Principal Spatial and Temporal Scale	Main Mechanisms	Expected Microbial or NGS-Derived Pattern	Relationship to the Pedogenic Template	Representative References
Regional climate	Regional to continental; long-term climatic regimes with seasonal variation	Temperature and precipitation regulate soil moisture, microbial metabolism, decomposition, nutrient turnover, plant productivity, and resource availability	Biogeographic differentiation of communities; enrichment of taxa and functions adapted to particular temperature or moisture regimes; shifts in carbon- and nutrient-cycling potential	Climate acts partly through soil formation and partly through contemporary environmental conditions; its effects are conditioned by soil texture, mineralogy, organic matter, drainage, and local climatic history	[3,4,8]
Seasonal variability and climatic extremes	Days to years; local to regional	Wetting and drying, warming, freezing, drought, flooding, and seasonal changes alter microbial activity, dormancy, mortality, substrate availability, and plant–microbe interactions	Rapid functional responses; temporary changes in relative abundance; activation of dormant populations; resistance or recovery trajectories following extreme events	Usually modifies microbial activity and the dynamic fraction of community composition without immediately altering the underlying pedogenic template; responses depend on previous climatic exposure and soil buffering capacity	[31–33]

Table 2. *Cont.*

Secondary Ecological Control	Principal Spatial and Temporal Scale	Main Mechanisms	Expected Microbial or NGS-Derived Pattern	Relationship to the Pedogenic Template	Representative References
Vegetation and litter inputs	Ecosystem to local scales; seasonal to decadal	Plant identity and community composition influence litter chemistry, organic matter inputs, nutrient uptake, shading, soil moisture, and belowground carbon allocation	Differences in fungal and bacterial composition among vegetation types; enrichment of decomposers adapted to particular litter substrates; altered functional potential related to carbon and nutrient cycling	Vegetation modifies resource availability within constraints imposed by soil fertility, mineralogy, moisture, aeration, and rooting depth; identical vegetation may therefore support different microbiomes on contrasting soils	[34–36]
Rhizosphere and host-associated selection	Millimetres to metres; plant developmental and seasonal timescales	Root exudation, rhizodeposition, oxygen gradients, nutrient depletion, symbiosis, and host immune or signalling processes selectively recruit microorganisms	Strong differentiation between rhizosphere and bulk-soil communities; host-associated taxa; increased interaction-network complexity; enrichment of plant-beneficial or host-adapted functions	Represents a localized biological refinement of the soil species pool; recruitment is constrained by the organisms available in the surrounding soil and by pedogenetically determined nutrient and moisture conditions	[34,37–39]
Microbial interactions and community self-organization	Microsite to soil-aggregate scales; hours to generations	Competition, facilitation, syntrophy, cross-feeding, antagonism, quorum signalling, and metabolic complementarity influence establishment and persistence	Non-random co-occurrence patterns; microbial modules or guilds; functional redundancy; alternative community configurations capable of similar ecosystem functions	Interactions operate among organisms able to reach and tolerate the habitat; their outcomes depend on resource distributions, pore structure, pH, redox conditions, and other soil properties	[9,14,40]
Predation, parasitism, and viral control	Microsite to local scales; rapid and episodic	Protistan grazing, bacteriophage infection, fungal parasites, and density-dependent mortality regulate population abundance and nutrient release	Turnover of dominant populations; maintenance of diversity; strain-specific fluctuations; changes in interaction networks and nutrient cycling that may not be apparent from taxonomic inventories alone	These processes modify relative abundance and activity within the environmentally permissible community but do not independently define the physicochemical habitat	[14,40]
Ecological drift, priority effects, and historical contingency	Microsite to landscape; generations to decades	Demographic stochasticity, local extinction, chance colonization, niche pre-emption, niche modification, and environmental legacies	Divergence among replicate communities under similar conditions; alternative community states; persistent effects of colonization order or previous land use	These processes determine which of several communities compatible with the pedogenic and environmental constraints becomes established	Note: In accordance with the restricted ecological definition, environmental filtering refers primarily to abiotic constraints on establishment or persistence. Vegetation, host selection, biotic interactions, predation, dispersal, and historical contingency are presented here as secondary ecological controls that refine community assembly but should not all be interpreted as environmental filtering in the strict sense. NGS, next-generation sequencing [19,41,42]
Dispersal limitation and homogenizing dispersal	Local to biogeographic; episodic to continuous	Restricted connectivity limits colonization, whereas high exchange rates introduce and establish organisms across communities	Dispersal limitation increases differentiation among environmentally similar sites; homogenizing dispersal increases similarity among environmentally contrasting sites	The pedogenic template influences whether immigrants can persist, but dispersal controls their opportunity to encounter the habitat; observed convergence or divergence may therefore differ from that predicted by environmental conditions alone	[43–45]

Note: NGS, next-generation sequencing.

Table 2 emphasizes that secondary ecological controls differ not only in identity but also in spatial extent, temporal persistence, and mechanistic relationship with soil properties. Regional climate and vegetation generate relatively broad and persistent differences among microbial communities, whereas rhizosphere processes, species interactions, predation, and short-term climatic variability operate at finer scales and may produce rapid compositional or functional responses. Dispersal and historical contingency introduce additional variation by determining which organisms are available for environmental selection. Consequently, HEF should not be interpreted as a rigid linear sequence in which each filter replaces the preceding one. Rather, community assembly occurs through nested and interacting constraints, with finer-scale processes operating conditionally within the habitat boundaries and species pools established at broader scales.

Among the secondary ecological filters proposed in the HEF framework, climate operates at the broadest spatial scale, influencing microbial communities through regional temperature and moisture regimes [4]. Temperature and moisture regulate microbial metabolism, decomposition rates, nutrient turnover, seasonal ecosystem processes, and microbial responses to climatic extremes [31,32]. Seasonal climatic variability can substantially influence microbial community dynamics, with bacterial communities often responding more strongly than fungal communities to seasonal environmental fluctuations [32,33]. Unlike pedogenesis, which operates over centuries to millennia, climatic conditions fluctuate continuously and therefore primarily influence microbial activity rather than the long-term physicochemical template of the soil. Jansson and Hofmockel [3] emphasized that climatic variability influences both the composition and functioning of soil microbiomes through changes in temperature, moisture, resource availability, and plant–soil interactions, although functional responses often occur more rapidly than detectable shifts in community composition. Recent experimental evidence further demonstrates that soil microbiomes exhibit consistent and predictable responses to climatic extremes across diverse ecosystems, with the magnitude of these responses largely determined by local climatic history and soil properties [31]. Within the HEF framework, climate should therefore be regarded as a modulator of ecological filtering rather than an independent driver of microbial assembly.

4.2. Vegetation, Rhizosphere Processes, and Biotic Interactions

Vegetation introduces an additional layer of biological selection by continuously modifying the chemical environment surrounding plant roots [34,38]. Through root exudation, litter deposition, nutrient uptake, and symbiotic interactions, plants selectively enrich microorganisms capable of exploiting specific carbon substrates or establishing mutualistic associations [38]. This rhizosphere-driven selection creates localized microbial communities that may differ substantially from those inhabiting adjacent bulk soil despite sharing the same pedogenetic origin [34,39]. The rhizosphere therefore represents one of the strongest biological filters operating within terrestrial ecosystems, generating highly specialized microbial assemblages associated with particular host species and exhibiting greater network complexity than adjacent bulk soils [38,39].

Importantly, plant-mediated filtering does not occur independently of soil properties. The influence of vegetation is itself constrained by the physicochemical environment established during pedogenesis. Root architecture, exudation patterns, nutrient acquisition strategies, and microbial recruitment differ among soils because mineralogy, water availability, aeration, and nutrient status regulate both plant physiology and microbial responses. Thus, vegetation should not be viewed as an alternative explanation for microbial community assembly but rather as a biological filter acting within the environmental boundaries defined by soil formation.

Biological interactions among microorganisms provide an additional level of ecological filtering. Competition for limited resources, metabolic cooperation, syntrophic relationships, predation by protists, viral infections, and antagonistic interactions collectively determine which microorganisms successfully establish stable populations after initial colonization. Community assembly therefore reflects not only environmental selection but also self-organization emerging from microbial interaction networks. Increasing evidence suggests that microbial communities possess emergent properties arising from these interactions, including functional redundancy, ecological resilience, and network stability, none of which can be explained by the ecological characteristics of individual taxa alone [14,40].

The role of biological filtering is particularly evident in fungal communities. Compared with bacteria, fungi often exhibit stronger host specificity, greater dependence on plant-derived carbon, and more pronounced responses to changes in litter chemistry and root exudation. Tedersoo et al. [35] demonstrated that vegetation represents one of the principal ecological drivers of global soil fungal diversity, acting together with climate, soil properties, and geographic factors to structure fungal communities. These findings indicate that fungal communities are often more responsive to vegetation-driven ecological filtering than bacterial communities because their assembly reflects the combined influence of pedogenetic constraints and ongoing plant–microbe interactions.

Despite the strong influence of climate and vegetation, secondary ecological filters rarely eliminate the underlying influence of soil formation. Instead, they modify microbial assemblages within the physicochemical constraints established during pedogenesis. This nested organization helps explain why similar vegetation may support different microbial communities when growing on contrasting soils, whereas microbial assemblages occurring within comparable pedogenetic environments may retain recognizable compositional patterns despite differences in vegetation. Community assembly therefore reflects the sequential action of multiple ecological filters operating across spatial and temporal scales rather than the dominance of any single environmental variable [3,4].

This hierarchical organization has important implications for the interpretation of microbial assemblies. Stable microbial communities emerge not because they remain unchanged through time but because successive ecological filters repeatedly constrain assembly toward similar community states. Consequently, they should be viewed as probabilistic ecological outcomes rather than fixed taxonomic identities. Their apparent stability depends on the persistence of the environmental filters that generate them, whereas their variability reflects the strength of secondary ecological selection acting on the pedogenetic template [3].

The action of these secondary filters also explains why microbial communities exhibit considerable ecological plasticity. Under natural conditions, climatic fluctuations and vegetation dynamics continuously reshape microbial community composition while preserving the broader community structure imposed by pedogenesis. However, sufficiently intense environmental disturbances may override the underlying pedogenetic signal. Understanding how anthropogenic disturbances modify or override pedogenetic influences is therefore essential for interpreting NGS-derived microbial patterns and distinguishing natural ecological variation from disturbance-driven community reassembly.

4.3. Dispersal, Stochasticity, and Historical Contingency

Environmental selection alone cannot account for all patterns observed in soil microbial communities. Community assembly also reflects dispersal, ecological drift, diversification, priority effects, and historical contingency, whose relative influence may vary with spatial scale, habitat connectivity, disturbance, population size, and the strength of environ-

mental selection [15,19,43,44]. These processes are not additional levels of environmental filtering in the strict sense. Instead, they interact with the environmental constraints represented by HEF and influence which of the potentially compatible microbial assemblages becomes established.

Dispersal can influence microbial patterns in contrasting ways. Dispersal limitation restricts the movement of organisms among sites and may cause compositionally different communities to develop under similar environmental conditions. In such cases, geographic turnover or distance–decay relationships should not automatically be attributed to environmental heterogeneity. Conversely, homogenizing dispersal can increase similarity among communities exposed to different environmental conditions by continuously introducing and establishing organisms across sites. High rates of dispersal may therefore weaken or obscure the microbial differentiation otherwise expected from environmental filtering [44,45]. Consequently, similarity among soil microbiomes does not necessarily demonstrate homogeneous environmental selection, just as dissimilarity does not necessarily indicate contrasting environmental filters.

Ecological drift produces stochastic changes in the presence and relative abundance of populations through demographic variation, local extinction, and chance colonization. Its influence is expected to increase when microbial populations are small, habitats are fragmented, selection is weak, or disturbance reduces community size. Drift can therefore generate divergence among replicate communities exposed to similar environmental conditions, particularly when combined with dispersal limitation [19,43,44]. In NGS datasets, part of the unexplained variation among environmentally similar samples may consequently reflect stochastic turnover rather than an unmeasured environmental filter.

Priority effects and historical contingency further influence assembly because the identity and order of early colonists can alter subsequent resource availability, habitat conditions, and opportunities for later-arriving organisms. Through niche pre-emption or niche modification, early-established populations may lead communities exposed to comparable environmental conditions toward alternative taxonomic or functional states [41]. Previous land use, disturbance, vegetation, climatic exposure, and colonization history may therefore leave persistent microbial legacies even after contemporary environmental conditions become similar. Such historical effects are particularly relevant when interpreting restoration chronosequences, land-use transitions, disturbed soils, and microbial inoculation experiments.

Within HEF, environmental filters should therefore be understood as defining a probabilistic range of ecologically permissible community states rather than determining a single inevitable taxonomic composition. Dispersal determines which members of the regional species pool reach a habitat; abiotic environmental selection determines which organisms can persist; biotic interactions and priority effects modify establishment and coexistence; and ecological drift produces stochastic variation among the resulting communities. Strong and persistent environmental constraints may generate convergence despite differences in arrival history, whereas weak selection, restricted dispersal, or strong priority effects may permit multiple community configurations under apparently similar conditions.

Accordingly, HEF should not be used to attribute all residual microbial variation to environmental selection. Distinguishing among selection, dispersal, drift, and historical contingency requires spatially and temporally replicated sampling, explicit consideration of geographic distance and habitat connectivity, environmental measurements at appropriate scales, and, where justified, null-model or process-partitioning approaches [19,42,44]. These methods also depend on assumptions concerning phylogenetic signal, taxonomic resolution, sampling completeness, and the definition of the regional species pool. Their results should

therefore be interpreted as evidence concerning the relative contributions of assembly processes rather than as direct proof of individual mechanisms.

5. Anthropogenic Overrides and Environmental Plasticity

The reproducibility of microbial communities depends on the long-term stability of the ecological filters governing community assembly. As discussed in the preceding sections, pedogenesis establishes the structural template of the soil, while climatic conditions, vegetation, and biological interactions progressively refine microbial communities within this environmental framework. Under relatively undisturbed conditions, these nested filters repeatedly generate microbial communities with recognizable ecological signatures associated with particular soil types. However, this hierarchical organization is not immutable. Environmental disturbances capable of modifying the strength or sequence of ecological filters may partially disrupt community assembly and thereby reduce the distinctiveness of microbial communities.

Among all anthropogenic disturbances, agricultural management represents the most pervasive driver of microbial community reassembly. Intensive tillage, mineral fertilization, irrigation, liming, pesticide application, and simplified crop rotations profoundly alter the physical, chemical, and biological properties of agricultural soils, reshaping microbial community assembly on timescales that are orders of magnitude shorter than those of pedogenesis [46,47]. Rather than creating entirely new microbial communities, these practices impose a common set of selective pressures that increasingly outweigh the natural environmental constraints responsible for soil-specific community assembly. Hartmann et al. [48] demonstrated that intensive agricultural management rapidly restructures microbial communities, shifting them away from the long-term equilibrium established under natural ecosystem development. Consequently, microbial assemblages become increasingly structured by management practices rather than by intrinsic pedogenetic characteristics [49–51].

Nitrogen fertilization provides one of the clearest examples of anthropogenic ecological override within the HEF framework. Long-term nitrogen enrichment consistently favors fast-growing copiotrophic microorganisms capable of exploiting nutrient-rich conditions while suppressing oligotrophic taxa adapted to nutrient-limited environments [52]. Across diverse ecosystems, nitrogen enrichment promotes bacterial groups associated with rapid resource acquisition at the expense of taxa characteristic of undisturbed soils. Within the HEF framework, these changes should not be interpreted merely as shifts in microbial diversity. Rather, they weaken the natural resource gradients established during soil development, redirecting community assembly toward alternative ecological states imposed by anthropogenic nutrient enrichment.

Collectively, recent studies suggest that intensive land use may increase similarity among microbial communities occurring in contrasting soils [53,54]. In natural ecosystems, different pedogenetic pathways generate distinct physicochemical environments that contribute to the assembly of different microbial communities. Under prolonged cultivation, however, these differences may become increasingly obscured because tillage, fertilization, irrigation, and crop management impose similar selective pressures across otherwise contrasting soils. Consequently, microbial community composition increasingly reflects contemporary management conditions in addition to the longer-term influence of pedogenesis. Within the HEF framework, this represents an anthropogenic reorganization of environmental filtering rather than simply a reduction in microbial diversity (Table 3).

Table 3. Anthropogenic drivers, mechanisms, and microbial responses within the Hierarchical Environmental Filtering framework.

Anthropogenic Driver	Proximal Environmental Condition Altered	Mechanism of Community Reassembly	Expected Microbial Response	Consequence for HEF Interpretation	Evidence
Mineral fertilization or other nutrient amendment	Availability and stoichiometric balance of nitrogen, phosphorus and other nutrients; frequently soil pH	Relaxation or intensification of resource limitation changes competitive relationships among microbial taxa	Enrichment of copiotrophic and nutrient-responsive taxa; decline of some oligotrophic taxa; changes in nutrient-cycling genes	Contemporary resource enrichment may weaken the correspondence between the microbiome and the antecedent nutrient regime or pedogenic template	[52]
Tillage and repeated soil cultivation	Aggregate structure, aeration, moisture retention, organic-matter distribution and vertical organization of soil horizons	Physical disturbance redistributes substrates and microorganisms, disrupts microsites and changes resource accessibility	Community turnover; altered bacterial-to-fungal relationships; reduced abundance of disturbance-sensitive taxa; possible community homogenization	Management may override or obscure microbial differentiation associated with natural soil structure and horizon development	[48,50]
Crop selection, crop rotation or replacement of natural vegetation	Quantity, quality and seasonality of litter, root exudates and rhizodeposition	Altered plant-derived substrates and host associations favour microorganisms adapted to the cultivated vegetation	Changes in rhizosphere and bulk-soil community composition, diversity and functional potential	Vegetation-associated microbial signatures may primarily reflect recent management rather than the long-term pedogenic environment	[53,54]

Note: HEF, Hierarchical Environmental Filtering.

Collectively, these observations demonstrate that soil microbial communities are neither immutable taxonomic entities nor temporally invariant ecological systems. Their composition reflects the combined influence of hierarchical environmental filters operating across different spatial and temporal scales. When pedogenetic constraints remain dominant, microbial communities may repeatedly converge toward comparable ecological states. Conversely, sufficiently strong anthropogenic disturbances can redirect community assembly toward alternative configurations increasingly governed by contemporary land use and resource conditions. Interpretation of NGS-derived microbial patterns must therefore consider the relative strength of pedogenetic, climatic, biological, and anthropogenic filters rather than attributing observed community composition to any single environmental factor.

6. Future Perspectives: From Microbial Inventories to Ecological Interpretation

The rapid development of sequencing technologies and computational ecology has fundamentally expanded our ability to characterize soil microbial communities. Nevertheless, the ultimate value of microbial sequencing extends beyond increasingly detailed descriptions of microbial diversity. The central challenge is no longer to determine which microorganisms inhabit particular soils, but to understand how ecological assembly generates reproducible biological signatures that can be interpreted, predicted, and ultimately applied in soil science. Future research should therefore shift from descriptive microbiome inventories toward mechanistic models capable of linking community assembly with soil genesis and ecosystem functioning.

Building on recent frameworks for soil bioindicator assessment and best-practice recommendations for microbiome research [11,12], we summarize six minimum criteria for defensible microbiome-based soil diagnostics in Table S4.

6.1. Functional and Multi-Omics Integration

One of the most promising directions is the transition from taxonomic profiling to functional ecology. Although amplicon sequencing has revolutionized our understanding of microbial diversity, taxonomic composition alone provides only indirect information about ecosystem processes. Functionally similar communities may differ substantially in taxonomic composition, whereas closely related microorganisms may occupy contrasting ecological niches. Shotgun metagenomics simultaneously characterizes taxonomic composition and the repertoire of functional genes present in microbial communities. It therefore provides information about functional potential rather than direct evidence of gene expression, metabolic activity or process rates. Linking the metagenomic repertoire to realized ecosystem functioning requires complementary transcriptomic, proteomic, metabolomic or process-level measurements. Such multi-omics approaches could transform soil microbiome analysis from taxonomic description into comprehensive ecological signatures of soil functioning [55].

6.2. Machine Learning: Predictive Potential and Validation Requirements

The high dimensionality and non-linearity of microbiome datasets have encouraged the application of machine-learning methods to classify soil conditions and predict environmental or functional variables. Random forests, support-vector machines, gradient boosting, nearest-neighbour methods and neural networks can identify multivariate patterns that may not be apparent from individual taxa or conventional univariate analyses. However, high apparent predictive performance does not necessarily demonstrate that a model has learned a generalizable ecological relationship. Microbiome datasets commonly contain spatially, temporally or experimentally correlated samples, creating a substantial risk of data leakage and overly optimistic estimates of model performance.

Randomly dividing individual samples between training and testing sets is inappropriate when multiple samples originate from the same site, plot, soil core, season or experimental unit. Under such conditions, the model may recognize site-specific, temporal or batch-associated signatures that occur in both subsets rather than learn relationships transferable to new environments. Training and testing partitions should therefore be defined at the highest relevant independent level. Suitable approaches include grouped cross-validation, in which all observations from a site, plot or experimental unit remain within the same fold; leave-one-site-out or leave-one-region-out validation; spatially blocked cross-validation; and temporally separated validation in which earlier observations are used to predict later ones. The validation strategy should reflect the intended application. For example, random grouped cross-validation may evaluate prediction of unsampled plots within a studied region, whereas leave-one-region-out validation provides a more demanding test of geographic transferability.

Data leakage can also occur before model fitting. Feature filtering, taxon selection, zero replacement, normalization, compositional transformation, batch correction, imputation, scaling and hyperparameter selection must be performed using the training data within each cross-validation iteration. Selecting microbial predictors from the complete dataset before cross-validation exposes the model to information from the test folds and can markedly inflate performance. When feature selection or hyperparameter optimization is required, nested cross-validation should be used: the inner loop performs model and feature selection, whereas the outer loop provides an unbiased estimate of predictive performance. A final independent dataset that was not used during preprocessing, feature selection, model selection or parameter tuning remains the strongest test of generalizability.

Performance should be reported using metrics appropriate to the prediction task and outcome distribution. Classification studies should provide a confusion matrix and report

sensitivity, specificity, balanced accuracy, precision, recall, F1 score or Matthews correlation coefficient as appropriate. Receiver-operating-characteristic area under the curve (ROC–AUC) can be informative, but precision–recall AUC may be more revealing when classes are strongly imbalanced. Overall accuracy alone can be misleading when one class predominates. Regression studies should report the coefficient of determination (R^2) together with an error measure expressed on the scale of the response, such as root-mean-square error (RMSE) or mean absolute error (MAE). Calibration, uncertainty intervals and performance variability among validation folds should also be presented. Comparisons with simple baseline models are necessary to establish whether the microbiome contributes meaningful predictive information beyond class prevalence or readily measured soil properties.

Published applications illustrate both the potential and limitations of these approaches. Using bacterial communities from more than 3000 soil samples representing 606 sites in New Zealand, Hermans et al. [56] reported 85% accuracy for land-use classification, up to 83% accuracy for categories based on soil physicochemical properties, R^2 values ranging from 0.35 to 0.79 for individual soil variables, and 50–95% accuracy in assigning soil-quality scores. More recently, Aouabed et al. [57] compared traditional machine-learning and deep-learning methods using two publicly available soil-microbiome datasets. Their best reported predictions reached $R^2 = 0.57$ for bacterial composition at the phylum level and $R^2 = 0.45$ for fungal functional groups, while performance generally declined at finer taxonomic levels. Random forest and k-nearest-neighbour methods frequently outperformed the multilayer perceptron, illustrating that increased model complexity does not necessarily improve prediction when the available sample size is limited. These values are illustrative rather than directly comparable because the studies addressed different outcomes, taxonomic levels, datasets and validation designs.

For HEF-based applications, the most informative validation tests are those that progressively withhold entire levels of the proposed hierarchy. Models may, for example, be trained in some plots and tested in new plots within the same soil type, then tested in independent sites, climatic regions or land-use systems. A large decline in performance during geographic or temporal transfer would indicate that the learned microbial signature is context-dependent rather than a universal diagnostic indicator. Conversely, reproducible performance in independent environments would support the proposition that hierarchical environmental filtering generates transferable associations between microbiome structure and soil condition. Therefore, machine learning should be viewed as a tool for testing the generality and predictive value of HEF-derived patterns, not as evidence of ecological mechanism by itself.

Representative examples of predictive performance reported in soil microbiome studies are summarized in Table S5 [56–58]. These values should be regarded as study-specific illustrations rather than universal benchmarks because the underlying prediction tasks, datasets, taxonomic resolutions and validation procedures differ substantially.

6.3. Core Microbiomes and Integrated Soil Assessment

Another major challenge concerns the identification of stable core microbiomes associated with particular soil types. The concept of a core microbiome generally focuses on microorganisms that consistently occur across samples or environments under defined ecological conditions [59]. However, from the perspective of hierarchical environmental filtering, ecological persistence may be more informative than taxonomic persistence alone. Future studies should therefore identify microbial assemblages that are repeatedly assembled under similar pedogenetic conditions. Such ecological core communities would represent biological manifestations of persistent environmental filtering rather than fixed

taxonomic inventories. This distinction may prove fundamental for developing reliable microbiome-based indicators of soil identity.

The integration of microbial information with conventional soil measurements represents another promising research direction. Physical and chemical soil properties provide essential information about pedogenetic history and present environmental conditions, whereas microbial communities provide a dynamic biological dimension reflecting ecosystem functioning and contemporary ecological responses. Combining these complementary information layers may improve monitoring of soil degradation, restoration trajectories, land-use effects, and changes in ecosystem functioning. Evidence that microbial community composition responds systematically to soil properties, land use, environmental gradients, and ecological restoration supports the value of such integrative approaches [60–62].

However, successful integration of NGS-based microbiome analysis into soil ecological assessment will require rigorous methodological standardization. Differences in sampling strategy, sequencing platforms, DNA extraction protocols, bioinformatic pipelines, and reference databases continue to limit direct comparisons among studies [11,12]. Perhaps even more importantly, ecological interpretation requires standardized consideration of sampling depth, seasonality, land-use history, and disturbance intensity, all of which influence microbial community assembly [3,30,63]. Without harmonized analytical protocols, methodological variability may exceed biological variation, substantially reducing the reproducibility of microbial sequencing across laboratories and geographic regions.

6.4. Empirical Testing and Testable Predictions of HEF

For HEF to function as more than a descriptive organization of environmental drivers, it must generate predictions that can be evaluated against alternative models of community assembly. Its central proposition is that inherited soil properties establish persistent but non-exclusive constraints on microbial community assembly, while contemporary climate, vegetation, biological interactions, dispersal, disturbance, and stochastic processes determine variation within or across those constraints. HEF does not predict that pedogenic characteristics will explain all microbial variation or that environmentally similar soils will contain identical communities. It predicts that explicitly accounting for the temporal persistence, scale, and conditional relationships among environmental variables should explain microbial community structure more effectively than treating all potential drivers as independent and additive.

A primary prediction is that inherited soil characteristics, including horizon identity, mineralogy, texture, organic matter distribution, drainage, and buffering capacity, will explain a statistically detectable component of microbial taxonomic or functional variation after contemporary climate, vegetation, geographic distance, and management have been considered. This prediction can be evaluated using variation partitioning, partial distance-based redundancy analysis (db-RDA), hierarchical mixed models, or related multivariate approaches. Support for HEF would be provided by a reproducible unique contribution of inherited soil properties and by convergence of at least some taxonomic or functional attributes among comparable soil horizons across independent locations. Conversely, repeated failure of inherited soil properties to explain any unique or transferable variation, particularly when contemporary and spatial variables explain nearly all observed structure, would contradict the proposed importance of the pedogenic template in the systems examined.

A second prediction is that the effects of contemporary climate and vegetation will be conditional on inherited soil properties rather than universally additive. For example, microbial responses to drought should differ among soils with contrasting texture, drainage, organic matter, or water-holding capacity, while rhizosphere effects should vary with soil

fertility, pH, and mineralogy. Such conditionality can be tested using interaction terms in PERMANOVA or multivariate generalized models, stratified db-RDA, hierarchical models with soil-specific slopes, and structural equation models representing direct and indirect pathways. HEF would be supported if models incorporating interactions or mediated pathways consistently outperform models containing only independent additive effects. Similar contemporary responses across contrasting soils, with no detectable interaction with inherited properties, would weaken this aspect of the framework.

A third prediction concerns temporal scale. Short-term climatic variation should initially produce stronger changes in microbial activity, transcription, protein expression, or metabolite production than in the persistent component of community composition, whereas long-term environmental change may progressively modify both microbial composition and the inherited soil template. Repeated seasonal sampling, time-series analysis, and integration of DNA-based profiles with metatranscriptomic, metaproteomic, or metabolomic measurements could test this prediction. HEF would be supported by differential temporal responses among molecular layers and by greater persistence of the soil-associated component of community structure than of microbial activity. Equivalent instability across all taxonomic and functional layers, unrelated to the persistence of environmental variables, would be inconsistent with this expectation.

HEF further predicts that strong anthropogenic disturbance may reduce the correspondence between microbial communities and inherited soil properties. This could be tested by comparing undisturbed and intensively managed soils across multiple soil types using factorial PERMANOVA, variation partitioning, multigroup structural equation modeling, and analyses of beta diversity and multivariate dispersion. Evidence of decreasing unique variation attributable to soil type, combined with increasing similarity among microbial communities subjected to comparable management, would support the proposed anthropogenic override. Persistence or strengthening of soil-specific microbial patterns across increasing disturbance intensity would contradict this prediction or indicate that the inherited template remains more resistant than proposed.

Environmental associations must also be distinguished from dispersal and stochastic assembly. Phylogenetic null models and β -nearest taxon index (β NTI) analyses can assess whether observed phylogenetic turnover differs from null expectations. Commonly applied frameworks interpret strongly positive β NTI values as evidence consistent with variable selection and strongly negative values as evidence consistent with homogeneous selection. For comparisons not assigned to strong selection, taxonomic null models, such as abundance-weighted Raup–Crick approaches, may help distinguish patterns consistent with dispersal limitation, homogenizing dispersal, or an undominated fraction incorporating drift, weak selection, and weak dispersal [42,44]. HEF would be supported where the relative contribution of inferred selection changes predictably across pedogenic contexts, environmental gradients, or disturbance levels. Predominance of spatial or stochastic processes with little reproducible association with environmental variables would limit the applicability of HEF in the system examined.

No individual statistical method can demonstrate the causal hierarchy proposed by HEF. PERMANOVA can test differences in multivariate composition but cannot alone distinguish environmental selection from differences in within-group dispersion. Tests of multivariate dispersion should therefore accompany PERMANOVA. Variation partitioning and db-RDA can quantify unique and shared fractions of explained variation, but these fractions remain sensitive to collinearity, variable selection, spatial structure, and unmeasured covariates. Structural equation modeling can assess whether observed data are consistent with hypothesized direct and indirect pathways, but cross-sectional models do not by themselves establish causality. Likewise, β NTI and related null-model classifications de-

pend on phylogenetic signal, the selected regional species pool, taxonomic resolution, and null-model assumptions. The strongest tests of HEF would therefore combine replicated observational studies with temporal sampling, reciprocal soil or microbiome transplants, controlled manipulation of moisture, vegetation or management, and validation across independent soils and geographic regions (Table 4).

Table 4. Testable predictions of the Hierarchical Environmental Filtering framework and analytical approaches for their evaluation.

HEF Prediction	Observation Supporting HEF	Observation Contradicting or Weakening HEF	Suitable Approaches
Inherited soil properties retain an independent association with microbial community structure	Soil horizon, mineralogy, texture, drainage, or persistent organic matter explain unique variation after contemporary and spatial factors are controlled	Inherited properties explain no reproducible unique variation, while contemporary or spatial variables account for nearly all structure	Variation partitioning; partial db-RDA; hierarchical models; cross-site validation
Contemporary effects are conditional on the inherited soil context	Climate, drought, vegetation, or rhizosphere effects differ predictably among soil types or horizons	Contemporary drivers have uniform effects and models containing soil-context interactions do not improve fit or prediction	Factorial PERMANOVA; interaction models; stratified db-RDA; structural equation modelling
The relative importance of controls changes with temporal scale	Microbial activity and expression respond rapidly, whereas persistent community structure changes more slowly	Taxonomic and functional layers display equivalent instability unrelated to environmental persistence	Repeated sampling; time-series models; DNA–RNA–protein–metabolite integration
Strong disturbance can override the inherited pedogenic signal	Increasing management intensity reduces soil-associated differentiation and promotes convergence under similar management	Soil-specific microbial patterns remain unchanged or become stronger with disturbance	Factorial designs; variation partitioning; PERMANOVA with dispersion tests; multigroup structural equation modelling
Comparable persistent environmental constraints promote probabilistic convergence	Independent sites with comparable horizons exhibit greater taxonomic or functional similarity than expected after accounting for distance	Environmentally comparable sites are no more similar than expected under spatial or null models	db-RDA; matched-site comparisons; variance decomposition; phylogenetic and taxonomic null models
Selection, dispersal, and drift contribute differently across HEF contexts	β NTI or related analyses show predictable changes in inferred selection across soil types, gradients, or disturbance levels	Turnover is consistently dominated by dispersal or stochastic processes and lacks reproducible environmental associations	β NTI; Raup–Crick metrics; phylogenetic null models; iCAMP or comparable process-partitioning frameworks
A hierarchical representation performs better than a flat driver model	Hierarchical or mediated models improve out-of-sample prediction, information criteria, or explained variation	Flat additive, spatial, or stochastic models perform equally well or better across independent datasets	Model comparison; cross-validation; SEM; hierarchical Bayesian or mixed models

Note: db-RDA, distance-based redundancy analysis; PERMANOVA, permutational multivariate analysis of variance; SEM, structural equation modeling.

At the framework level, HEF would be weakened if hierarchical models consistently failed to improve explanatory or predictive performance relative to flat additive, purely spatial, or stochastic models; if inherited soil properties provided no reproducible unique contribution; and if the effects of contemporary drivers showed no dependence on soil context. HEF should therefore be considered a falsifiable, context-dependent framework whose relevance may vary among microbial groups, spatial scales, ecosystem types, and levels of disturbance.

Taken together, these studies support several central expectations of HEF (Table 5). First, globally and regionally distributed surveys indicate that broad climatic and biogeographic constraints delimit the community pool available for assembly, whereas soil pH, nutrient availability, horizon and other edaphic properties explain substantial differentiation at finer scales. Second, experimental studies demonstrate that contemporary disturbances, nutrient enrichment and land management can reorganize communities within—or, under sufficiently strong disturbance, partly override—the constraints imposed by the antecedent soil environment. Third, the relative importance of individual filters differs among bacteria, archaea and fungi, indicating that the proposed hierarchy is conditional rather than universal. Finally, methodological effects, including relic DNA and DNA-extraction procedures, can alter the apparent magnitude of environmental associations and should therefore be distinguished from ecological filtering itself. The collected evidence is consistent with HEF, but it does not yet constitute a direct test of the complete framework; such validation will require studies designed to quantify multiple filter levels simultaneously while accounting for dispersal, stochasticity and historical contingency.

Table 5. Representative empirical studies providing evidence consistent with different components of the Hierarchical Environmental Filtering framework.

Study	Soil or Ecosystem and Geographic Scope	Microbial Group	Molecular or Sequencing Method	Sampling Scale	Main Environmental Driver or Contrast	Principal Statistical Approach	Reported Magnitude or Strength of Association	HEF Level Represented
Fierer et al. [63]	Temperate soil profiles sampled from the surface to approximately 2 m depth	Bacteria	Phospholipid fatty-acid profiling and culture-independent community characterization	Two contrasting soil profiles sampled at multiple depths	Soil depth and associated changes in organic carbon and soil conditions	Principal component and multivariate community analyses	Strong compositional differentiation between surface and subsurface communities; no single standardized effect size reported	Pedogenic template and vertical soil-horizon filtering
Tedersoo et al. [35]	Natural ecosystems distributed globally	Fungi	ITS metabarcoding	365 soil samples	Climate, edaphic properties, vegetation and geographic context	Multivariate, spatial and regression-based modelling	Fungal richness and composition showed strong but taxon-dependent environmental and geographic relationships; richness was not globally coupled to plant diversity	Global climate, biogeographic context and vegetation-associated filtering
Hartmann et al. [48]	Long-term agricultural field experiment comparing organic and conventional management	Bacteria and fungi	16S rRNA gene and fungal marker-gene pyrosequencing	Replicated plots exposed to more than two decades of contrasting management	Farming system and long-term management history	Ordination, diversity analysis and permutational multivariate testing	Organic management increased richness, decreased evenness and shifted community composition; no common standardized effect size was reported	Anthropogenic modification of the local edaphic filter
Leff et al. [52]	Grasslands distributed across several continents	Bacteria, archaea and fungi	Marker-gene sequencing, supplemented by metagenomic analysis of a subset	25 grassland sites subjected to standardized nutrient additions	Nitrogen and phosphorus enrichment	Cross-site multivariate and mixed-effects analyses	Nutrient addition produced geographically consistent taxonomic and functional shifts, including enrichment of copiotrophic taxa; no single cross-domain effect size reported	Contemporary resource filter and anthropogenic override

Table 5. Cont.

Study	Soil or Ecosystem and Geographic Scope	Microbial Group	Molecular or Sequencing Method	Sampling Scale	Main Environmental Driver or Contrast	Principal Statistical Approach	Reported Magnitude or Strength of Association	HEF Level Represented
Carini et al. [10]	Soils spanning contrasting physicochemical conditions	Prokaryotes and fungi	16S rRNA gene and ITS sequencing combined with propidium-monoazide viability treatment	Broad soil collection with paired untreated and relic-DNA-depleted samples	Relic DNA abundance, partly associated with soil pH	Paired community comparisons, richness analysis and multivariate dissimilarity testing	On average, approximately 40% of amplifiable prokaryotic and fungal DNA originated from extracellular DNA or non-intact cells; observed richness was inflated by as much as 55%	Methodological observation filter affecting inference across HEF levels
Ma et al. [64]	Natural forests across four vegetation zones in eastern China	Archaea, bacteria and fungi	Archaeal, bacterial and fungal amplicon sequencing	110 forest sites	Vegetation zone, edaphic properties, climate and spatial separation	Ordination, PERMANOVA, distance–decay analysis and variation partitioning	The relative importance of environmental and spatial variables differed among microbial domains; no common standardized effect size reported	Regional climate and vegetation filters acting on local soil conditions
Bahram et al. [6]	Global topsoil survey encompassing major terrestrial ecosystems	Bacteria, archaea, fungi and functional genes	Metabarcoding and shotgun metagenomics	189 sites comprising 7560 local subsamples	Precipitation, soil pH and other edaphic properties	Multivariate and spatial analyses and predictive modelling	Environmental variables were more strongly associated with microbial taxonomic and functional composition than geographic distance; no single comparable effect size reported	Global climatic constraint followed by regional and local edaphic filtering
Cole et al. [53]	Contrasting land-use systems	Bacteria, archaea and fungi	Metabarcoding, metaproteomics and stable-isotope assessment of litter decomposition	Multiple soils representing contrasting land-use conditions	Soil pH and land use	Multivariate community analysis and models linking microbiome characteristics with carbon cycling	Soil pH emerged as a major correlate of microbial community characteristics and carbon-processing responses; no single standardized effect size reported	Local edaphic filtering and anthropogenic modification

Table 5. Cont.

Study	Soil or Ecosystem and Geographic Scope	Microbial Group	Molecular or Sequencing Method	Sampling Scale	Main Environmental Driver or Contrast	Principal Statistical Approach	Reported Magnitude or Strength of Association	HEF Level Represented
Knight et al. [31]	Grasslands distributed across Europe	Soil prokaryotic and fungal communities	Marker-gene sequencing following experimental climatic disturbances	Soils from 30 grasslands exposed to drought, flooding, freezing or heat treatments	Climatic extremes and the climatic history of the source soils	Mixed-effects, ordination and phylogenetic analyses	Responses were generally small but highly consistent and phylogenetically conserved; heat produced the strongest disturbance response	Contemporary climatic filtering conditioned by historical environmental exposure Sample size refers to the highest-level independent sampling units reported by each study—such as sites, soil profiles or experimental plots—rather than sequencing reads or technical replicates. “No single standardized effect size reported” indicates that the study demonstrated significant or consistent associations but did not express the overall result using one effect-size statistic directly comparable among studies. Assignment to an HEF level represents the interpretation proposed in the present review and not terminology used in the original publication. The underlying claims in the strongest quantitative rows are supported particularly well by the global surveys of Tedersoo et al., Leff et al. and Bahram et al.; the methodological magnitude reported by Carini et al. is especially useful because it supplies the kind of numerical association requested by the reviewer. Knight et al.’s experimental study provides a recent test of how climatic history conditions microbial responses to contemporary extremes [31]. Supplemental Materials Table S1. Inclusion and exclusion criteria applied during literature selection.

Note: Sample size refers to the highest-level independent sampling units reported by each study rather than sequencing reads or technical replicates. Where no single standardized effect size was reported, the table describes the direction or strength stated by the original authors. Assignments to HEF levels are interpretations made in the present review; the cited studies did not explicitly test HEF under that name.

7. Conclusions

The rapid expansion of sequencing technologies has generated an unprecedented amount of information describing soil microbial communities. Yet the central challenge of contemporary soil microbial ecology is no longer exclusively technological. A fundamental ecological question remains: why do reproducible patterns of microbial community organization emerge within soils despite immense biological complexity and environmental variability?

Throughout this review, we propose that these patterns can be interpreted through the hierarchical organization of environmental filtering. Soil microbial communities should not be considered passive collections of microorganisms responding independently to individual environmental variables. Rather, they represent ecological systems assembled through successive constraints operating during soil formation and ecosystem development. Pedogenesis establishes the fundamental physicochemical architecture of the habitat; resource availability defines metabolic opportunities; vegetation and biological interactions further refine community composition; climate modulates microbial activity and ecosystem functioning; and anthropogenic disturbance may modify or override these natural environmental filters.

This perspective provides an ecological context for interpreting NGS-derived observations. Soil microbial assemblages are dynamic outcomes of community assembly rather than static taxonomic entities. Their apparent reproducibility reflects the persistence of environmental constraints, whereas their temporal and spatial variability reflects ecological plasticity and responsiveness to changing environmental conditions. Intensive anthropogenic disturbance can alter this organization by modifying soil physicochemical properties and resource gradients, thereby redirecting microbial assembly toward alternative community states.

The Hierarchical Environmental Filtering framework proposed here does not seek to replace established theories of microbial community assembly. Instead, it provides an integrative framework for connecting observations from pedology, microbial ecology, ecosystem science, and environmental genomics. Within this perspective, sequencing technologies are powerful observational tools: they reveal the biological consequences of environmental filtering but do not themselves explain the ecological processes responsible for microbial community organization.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/d18090556/s1>, Figure S1: Literature-identification and selection workflow used for the structured critical narrative review. Table S1: Inclusion and exclusion criteria applied during literature selection. Table S2: Sequencing and complementary analytical approaches used in soil microbiome research and their relevance to HEF interpretation. Table S3: Ecological and molecular sources of variation affecting NGS-based interpretation and corresponding methodological responses. Table S4: Minimum criteria and research priorities for defensible NGS-based soil microbiome assessment. Table S5: Representative examples of machine-learning performance reported in soil microbiome studies. References [65–71] are cited in the Supplementary Materials.

Author Contributions: Conceptualization, K.D. and A.P.G.; methodology, L.C., O.M., B.K., K.D. and A.P.G.; investigation, L.K. (Larysa Kucher) and L.K. (Liudmyla Kava); literature curation, L.K. (Larysa Kucher) and L.K. (Liudmyla Kava); validation, L.C., O.M. and B.K.; writing—original draft preparation, L.K. (Larysa Kucher), L.K. (Liudmyla Kava), K.D. and A.P.G.; writing—review and editing, L.C., O.M., B.K., K.D. and A.P.G.; supervision, K.D. and A.P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this review. Data sharing is not applicable to this article.

Conflicts of Interest: Author Andrii P. Gryganskyi was presently employed by AeroVironment Inc. and Burkhard Kirchhoff was presently employed by Kirchhoff GmbH & Co. KG. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ma, Y.; Minasny, B.; Malone, B.P.; McBratney, A.B. Pedology and digital soil mapping (DSM). *Eur. J. Soil Sci.* **2019**, *70*, 216–235. [\[CrossRef\]](#)
2. Fierer, N. Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **2017**, *15*, 579–590. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Jansson, J.K.; Hofmockel, K.S. Soil microbiomes and climate change. *Nat. Rev. Microbiol.* **2020**, *18*, 35–46. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Philippot, L.; Chenu, C.; Kappler, A.; Rillig, M.C.; Fierer, N. The interplay between microbial communities and soil properties. *Nat. Rev. Microbiol.* **2024**, *22*, 226–239. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Thompson, L.R.; Sanders, J.G.; McDonald, D.; Amir, A.; Ladau, J.; Locey, K.J.; Prill, R.J.; Tripathi, A.; Gibbons, S.M.; Ackermann, G.; et al. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* **2017**, *551*, 457–463. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Bahram, M.; Hildebrand, F.; Forslund, S.K.; Anderson, J.L.; Soudzilovskaia, N.A.; Bodegom, P.M.; Bengtsson-Palme, J.; Anslan, S.; Coelho, L.P.; Harend, H.; et al. Structure and function of the global topsoil microbiome. *Nature* **2018**, *560*, 233–237. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Schöler, A.; Jacquiod, S.; Vestergaard, G.; Schulz, S.; Schloter, M. Analysis of soil microbial communities based on amplicon sequencing of marker genes. *Biol. Fertil. Soils* **2017**, *53*, 485–489. [\[CrossRef\]](#)
8. Vasar, M.; Davison, J.; Sepp, S.K.; Mucina, L.; Oja, J.; Al-Quraishy, S.; Anslan, S.; Bahram, M.; Bueno, C.G.; Cantero, J.J.; et al. Global soil microbiomes: A new frontline of biome-ecology research. *Glob. Ecol. Biogeogr.* **2022**, *31*, 1120–1132. [\[CrossRef\]](#)
9. Chen, H.H.; Ma, K.Y.; Lu, C.Y.; Fu, Q.; Qiu, Y.B.; Zhao, J.Y.; Huang, Y.; Yang, Y.C.; Schadt, C.W.; Chen, H. Functional redundancy in soil microbial community based on metagenomics across the globe. *Front. Microbiol.* **2022**, *13*, 878978. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Carini, P.; Marsden, P.J.; Leff, J.W.; Morgan, E.E.; Strickland, M.S.; Fierer, N. Relic DNA is abundant in soil and inflates estimates of microbial diversity. *Nat. Microbiol.* **2016**, *2*, 16242. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Riggi, L.G.A.; Pérès, G.; Aponte, C.; Bispo, A.; Bragato, G.; Cluzeau, D.; D’Avino, L.; Edlinger, A.; Gronchi, N.; Guegan, G.; et al. Establishing assessment criteria for soil bioindicators: Insights from case studies in Europe. *Ecol. Indic.* **2025**, *178*, 114063. [\[CrossRef\]](#)
12. Knight, R.; Vrbanc, A.; Taylor, B.C.; Aksenov, A.; Callewaert, C.; Debelius, J.; Gonzalez, A.; Kosciolk, T.; McCall, L.I.; McDonald, D.; et al. Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **2018**, *16*, 410–422. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Legeay, J.; Hijri, M. A comprehensive insight of current and future challenges in large-scale soil microbiome analyses. *Microb. Ecol.* **2023**, *86*, 75–85. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Widder, S.; Allen, R.; Pfeiffer, T.; Curtis, T.P.; Wiuf, C.; Sloan, W.T.; Cordero, O.X.; Brown, S.P.; Momeni, B.; Shou, W.; et al. Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *ISME J.* **2016**, *10*, 2557–2568. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Vellend, M. Conceptual synthesis in community ecology. *Q. Rev. Biol.* **2010**, *85*, 183–206. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kraft, N.J.B.; Adler, P.B.; Godoy, O.; James, E.C.; Fuller, S.; Levine, J.M. Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **2015**, *29*, 592–599. [\[CrossRef\]](#)
17. Stegen, J.C.; Lin, X.; Konopka, A.E.; Fredrickson, J.K. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* **2012**, *6*, 1653–1664. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Dini-Andreote, F.; Stegen, J.C.; van Elsland, J.D.; Salles, J.F. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1326–E1332. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Zhou, J.; Ning, D. Stochastic community assembly: Does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* **2017**, *81*, e00002-17. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Quince, C.; Walker, A.W.; Simpson, J.T.; Loman, N.J.; Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **2017**, *35*, 833–844. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Winand, R.; Bogaerts, B.; Hoffman, S.; Lefevre, L.; Delvoye, M.; Van Braekel, J.; Fu, Q.; Roosens, N.H.; De Keersmaecker, S.C.; Vanneste, K. Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (Illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. *Int. J. Mol. Sci.* **2020**, *21*, 298. [[CrossRef](#)] [[PubMed](#)]
23. Aja-Macaya, P.; Conde-Pérez, K.; Trigo-Tasende, N.; Buetas, E.; Nasser-Ali, M.; Nión, P.; Rumbo-Feal, S.; Ladra, S.; Bou, G.; Mira, Á.; et al. Nanopore full length 16S rRNA gene sequencing increases species resolution in bacterial biomarker discovery. *Sci. Rep.* **2025**, *15*, 26486. [[CrossRef](#)] [[PubMed](#)]
24. Gloor, G.B.; Macklaim, J.M.; Pawlowsky-Glahn, V.; Egozcue, J.J. Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **2017**, *8*, 2224. [[CrossRef](#)] [[PubMed](#)]
25. Lin, H.; Peddada, S.D. Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* **2020**, *11*, 3514. [[CrossRef](#)] [[PubMed](#)]
26. Vandeputte, D.; Kathagen, G.; D'Hoe, K.; Vieira-Silva, S.; Valles-Colomer, M.; Sabino, J.; Wang, J.; Tito, R.Y.; De Commer, L.; Darzi, Y.; et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* **2017**, *551*, 507–511. [[CrossRef](#)] [[PubMed](#)]
27. Stoddard, S.F.; Smith, B.J.; Hein, R.; Roller, B.R.K.; Schmidt, T.M. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* **2015**, *43*, D593–D598. [[CrossRef](#)] [[PubMed](#)]
28. Thijs, S.; Op De Beeck, M.; Beckers, B.; Truyens, S.; Stevens, V.; Van Hamme, J.D.; Weyens, N.; Vangronsveld, J. Comparative evaluation of four bacteria-specific primer pairs for 16S rRNA gene surveys. *Front. Microbiol.* **2017**, *8*, 494. [[CrossRef](#)] [[PubMed](#)]
29. McLaren, M.R.; Willis, A.D.; Callahan, B.J. Consistent and correctable bias in metagenomic sequencing experiments. *eLife* **2019**, *8*, e46923. [[CrossRef](#)] [[PubMed](#)]
30. Du, C.; Geng, Z.; Wang, Q.; Zhang, T.; He, W.; Hou, L.; Wang, Y. Variations in bacterial and fungal communities through soil depth profiles in a *Betula albosinensis* forest. *J. Microbiol.* **2017**, *55*, 684–693. [[CrossRef](#)] [[PubMed](#)]
31. Knight, C.G.; Nicolitch, O.; Griffiths, R.I.; Goodall, T.; Jones, B.; Weser, C.; Langridge, H.; Davison, J.; Dellavalle, A.; Eisenhauer, N.; et al. Soil microbiomes show consistent and predictable responses to extreme events. *Nature* **2024**, *636*, 690–696. [[CrossRef](#)] [[PubMed](#)]
32. Duque-Zapata, J.D.; Vélez-Martínez, G.A.; Reyes-Ardila, W.L.; Florez, J.E.M.; López-Álvarez, D. Seasonal dynamics of soil bacterial and fungal communities in Colombian tropical forests and páramo ecosystems. *Soil Ecol. Lett.* **2025**, *7*, 250348. [[CrossRef](#)]
33. Pold, G.; Schimel, J.P.; Sistla, S.A. Soil bacterial communities vary more by season than with over two decades of experimental warming in Arctic tussock tundra. *Elem. Sci. Anthr.* **2021**, *9*, 00116. [[CrossRef](#)]
34. Berg, G.; Smalla, K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol. Ecol.* **2009**, *68*, 1–13. [[CrossRef](#)] [[PubMed](#)]
35. Tedersoo, L.; Bahram, M.; Pölme, S.; Kõljalg, U.; Yorou, N.S.; Wijesundera, R.; Villarreal Ruiz, L.; Vasco-Palacios, A.M.; Thu, P.Q.; Suija, A.; et al. Fungal biogeography: Global diversity and geography of soil fungi. *Science* **2014**, *346*, 1256688. [[CrossRef](#)] [[PubMed](#)]
36. Habtewold, J.Z.; Helgason, B.L.; Yanni, S.F.; Janzen, H.H.; Ellert, B.H.; Gregorich, E.G. Litter composition has stronger influence on the structure of soil fungal than bacterial communities. *Eur. J. Soil Biol.* **2020**, *98*, 103190. [[CrossRef](#)]
37. Naylor, D.; DeGraaf, S.; Purdom, E.; Coleman-Derr, D. Drought and host selection influence bacterial community dynamics in the grass root microbiome. *ISME J.* **2017**, *11*, 2691–2704. [[CrossRef](#)] [[PubMed](#)]
38. Philippot, L.; Raaijmakers, J.M.; Lemanceau, P.; van der Putten, W.H. Going back to the roots: The microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* **2013**, *11*, 789–799. [[CrossRef](#)] [[PubMed](#)]
39. Wang, D.; Deng, S.; Yang, H.; Li, N.; Feng, Q.; Liu, J.; Yin, H. The microbial network exhibits higher complexity in the rhizosphere than in bulk soils along elevational gradients in the alpine forests. *Appl. Soil Ecol.* **2025**, *213*, 106264. [[CrossRef](#)]
40. Faust, K.; Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **2012**, *10*, 538–550. [[CrossRef](#)] [[PubMed](#)]
41. Fukami, T. Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Syst.* **2015**, *46*, 1–23. [[CrossRef](#)]
42. Ning, D.; Yuan, M.; Wu, L.; Zhang, Y.; Guo, X.; Zhou, X.; Yang, Y.; Arkin, A.P.; Firestone, M.K.; Zhou, J. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **2020**, *11*, 4717. [[CrossRef](#)] [[PubMed](#)]
43. Nemergut, D.R.; Schmidt, S.K.; Fukami, T.; O'Neill, S.P.; Bilinski, T.M.; Stanish, L.F.; Knelman, J.E.; Darcy, J.L.; Lynch, R.C.; Wickey, P.; et al. Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.* **2013**, *77*, 342–356. [[CrossRef](#)] [[PubMed](#)]
44. Stegen, J.C.; Lin, X.; Fredrickson, J.K.; Chen, X.; Kennedy, D.W.; Murray, C.J.; Rockhold, M.L.; Konopka, A. Quantifying community assembly processes and identifying features that impose them. *ISME J.* **2013**, *7*, 2069–2079. [[CrossRef](#)] [[PubMed](#)]
45. Evans, S.; Martiny, J.B.H.; Allison, S.D. Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME J.* **2017**, *11*, 176–185. [[CrossRef](#)] [[PubMed](#)]

46. Gupta, A.; Singh, U.B.; Sahu, P.K.; Paul, S.; Kumar, A.; Malviya, D.; Singh, S.; Kuppasamy, P.; Singh, P.; Paul, D.; et al. Linking soil microbial diversity to modern agriculture practices: A review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 3141. [[CrossRef](#)] [[PubMed](#)]
47. Kong, W.; Qiu, L.; Ishii, S.; Jia, X.; Su, F.; Song, Y.; Hao, M.; Shao, M.; Wei, X. Contrasting response of soil microbiomes to long-term fertilization in various highland cropping systems. *ISME Commun.* **2023**, *3*, 81. [[CrossRef](#)] [[PubMed](#)]
48. Hartmann, M.; Frey, B.; Mayer, J.; Mäder, P.; Widmer, F. Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J.* **2015**, *9*, 1177–1194. [[CrossRef](#)] [[PubMed](#)]
49. Dunn, L.; Lang, C.; Marilleau, N.; Terrat, S.; Biju-Duval, L.; Lelièvre, M.; Perrin, S.; Chemidlin Prévost-Bouré, N. Soil microbial communities in the face of changing farming practices: A case study in an agricultural landscape in France. *PLoS ONE* **2021**, *16*, e0252216. [[CrossRef](#)] [[PubMed](#)]
50. Fox, A.; Widmer, F.; Lüscher, A. Soil microbial community structures are shaped by agricultural systems revealing little temporal variation. *Environ. Res.* **2022**, *214*, 113915. [[CrossRef](#)] [[PubMed](#)]
51. La Porta, C.A.M.; Marchi, E.; Chiaffarelli, G.; Vagge, I.; Vaglia, V.; De Marinis, P.; Zapperi, S.; Bocchi, S. Soil health and microbial diversity across land-use types: Evidence for agroecological management in peri-urban areas. *Open Res. Eur.* **2026**, *6*, 29. [[CrossRef](#)]
52. Leff, J.W.; Jones, S.E.; Prober, S.M.; Barberán, A.; Borer, E.T.; Firn, J.L.; Harpole, W.S.; Hobbie, S.E.; Hofmockel, K.S.; Knops, J.M.H.; et al. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10967–10972. [[CrossRef](#)] [[PubMed](#)]
53. Cole, L.; Goodall, T.; Jehmlich, N.; Griffiths, R.I.; Gleixner, G.; Gubry-Rangin, C.; Malik, A.A. Land use effects on soil microbiome composition and traits with consequences for soil carbon cycling. *ISME Commun.* **2024**, *4*, ycae116. [[CrossRef](#)] [[PubMed](#)]
54. Hua, H.; Sui, X.; Liu, Y.; Liu, X.; Chang, Q.; Xu, R.; Li, M.; Mu, L. Effects of land use type transformation on the structure and diversity of soil bacterial communities. *Life* **2024**, *14*, 252. [[CrossRef](#)] [[PubMed](#)]
55. Fierer, N.; Wood, S.A.; de Mesquita, C.P.B. How microbes can, and cannot, be used to assess soil health. *Soil Biol. Biochem.* **2021**, *153*, 108111. [[CrossRef](#)]
56. Hermans, S.M.; Buckley, H.L.; Case, B.S.; Curran-Cournane, F.; Taylor, M.; Lear, G. Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome* **2020**, *8*, 79. [[CrossRef](#)] [[PubMed](#)]
57. Aouabed, Z.; Therrien, V.; Bouaoune, M.A.; Bakhtyari, M.; Hijri, M.; Makarenkov, V. Soil microbiome prediction using traditional machine learning and deep learning models. *Sci. Rep.* **2026**, *16*, 11069. [[CrossRef](#)] [[PubMed](#)]
58. Wilhelm, R.C.; van Es, H.M.; Buckley, D.H. Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biol. Biochem.* **2022**, *164*, 108472. [[CrossRef](#)]
59. Toju, H.; Peay, K.G.; Yamamichi, M.; Narisawa, K.; Hiruma, K.; Naito, K.; Fukuda, S.; Ushio, M.; Nakaoka, S.; Onoda, Y.; et al. Core microbiomes for sustainable agroecosystems. *Nat. Plants* **2018**, *4*, 247–257. [[CrossRef](#)] [[PubMed](#)]
60. Stephanou, C.; Omirou, M.; Philippot, L.; Zissimos, A.M.; Christoforou, I.C.; Trajanoski, S.; Oulas, A.; Ioannides, I.M. Land use in urban areas impacts the composition of soil bacterial communities involved in nitrogen cycling. A case study from Lefkosia (Nicosia) Cyprus. *Sci. Rep.* **2021**, *11*, 8198. [[CrossRef](#)] [[PubMed](#)]
61. Sun, S.; Li, S.; Avera, B.N.; Strahm, B.D.; Badgley, B.D. Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration. *Appl. Environ. Microbiol.* **2017**, *83*, e00966–17. [[CrossRef](#)] [[PubMed](#)]
62. Xue, P.; McBratney, A.B.; Minasny, B.; O'Donnell, T.; Pino, V.; Fajardo, M.; Ng, W.; Wilson, N.; Deaker, R. Soil bacterial depth distribution controlled by soil orders and soil forms. *Soil Ecol. Lett.* **2022**, *4*, 57–68. [[CrossRef](#)]
63. Fierer, N.; Schimel, J.P.; Holden, P.A. Variations in microbial community composition through two soil depth profiles. *Soil Biol. Biochem.* **2003**, *35*, 167–176. [[CrossRef](#)]
64. Ma, B.; Dai, Z.; Wang, H.; Dsouza, M.; Liu, X.; He, Y.; Wu, J.; Rodrigues, J.L.M.; Gilbert, J.A.; Brookes, P.C.; et al. Distinct biogeographic patterns for archaea, bacteria, and fungi along the vegetation gradient at the continental scale in eastern China. *mSystems* **2017**, *2*, e00174–16. [[CrossRef](#)] [[PubMed](#)]
65. Lynch, M.D.; Neufeld, J.D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **2015**, *13*, 217–229. [[CrossRef](#)] [[PubMed](#)]
66. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)] [[PubMed](#)]
67. Garg, D.; Patel, N.; Rawat, A.; Rosado, A.S. Cutting edge tools in the field of soil microbiology. *Curr. Res. Microb. Sci.* **2024**, *6*, 100226. [[CrossRef](#)] [[PubMed](#)]
68. Sorensen, P.O.; Karaoz, U.; Beller, H.R.; Bill, M.; Bouskill, N.J.; Banfield, J.F.; Chu, R.K.; Hoyt, D.W.; Eder, E.; Eloee-Fadrosch, E.; et al. Multi-omics reveals nitrogen dynamics associated with soil microbial blooms during snowmelt. *Nat. Microbiol.* **2026**, *11*, 359–374. [[CrossRef](#)] [[PubMed](#)]

69. Samrat, R.; Salas, E.; Fuchslueger, L.; Schmidt, H.; Gorfer, M.; Schagerl, M.; Eichorst, S.A.; Wanek, W. High-resolution lipidomics for decoding the soil biome: Improved lipid annotation, quantitation, and response to climate stress. *Soil Biol. Biochem.* **2025**, *209*, 109892. [[CrossRef](#)]
70. Pace, R.; Monti, M.M.; Cuomo, S.; Affinito, A.; Ruocco, M. Machine Learning Approaches to Assess Soil Microbiome Dynamics and Bio-Sustainability. *Physiol. Plant.* **2026**, *178*, e70719. [[CrossRef](#)] [[PubMed](#)]
71. López-Angulo, J.; de la Cruz, M.; Chacón-Labela, J.; Illuminati, A.; Matesanz, S.; Pescador, D.S.; Pías, B.; Sánchez, A.M.; Escudero, A. The role of root community attributes in predicting soil fungal and bacterial community patterns. *New Phytol.* **2020**, *228*, 1070–1082. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.