

Metagenomic Insights into Soil Microbial Community Assembly: A Critical and Quantitative Review of Deterministic versus Stochastic Drivers

Min Jae Park ^{1*}, Ji Hoon Kim ²

^{1,2} Precision Oncology Research Center, Korea University, South Korea

*Corresponding author: **Min Jae Park**

Received: 12 Sep 2021

Revised: 25 Sep 2021

Accepted: 11 Oct 2021

Published: 06 Nov 2021

E-ISSN: 2989-5359

DOI: <https://doi.org/10.54660/ejsa.2021.2.53-60>

ABSTRACT

Context: Understanding how soil microorganisms come together is a key issue in microbial ecology which has important implications for ecosystem recovery, farming efficiency, and global biogeochemical processes. Research based on amplicon sequencing such as 16S rRNA and ITS sequencing has allowed for better understanding of the taxonomic structure of communities but has drawbacks due to primer biases and low resolution as well as lacking the capability to analyze characteristics of soils. Use of shotgun metagenomics has changed the situation dramatically and provided researchers with an opportunity to address questions concerning ecological assembly in terms of ecological theory - more specifically, the dichotomy between deterministic selection and stochastic neutral processes.

Purpose of the review: The main purpose of this review is to provide an overview of structural, functional, and methodological findings obtained through metagenomic research conducted in 2015-2025.

Literature sources: For this purpose, a systematic literature search was performed in Scopus, Web of Science, PubMed, ScienceDirect, and SpringerLink.

Methodology of Review: Following the PRISMA guidance, a strict set-up narrowed 1,248 primary records down to the 34 keys, outstanding, works with peer-reviewing. They went under scrupulous scrutiny in terms of the method of research, sequencing methods, as well as ecological conclusions gained during the research.

The Main Outcomes: The synthesis carried out revealed that while deterministic selection is the key factor behind assembly of functional community (this being attributed to such factors as pH rate, $\frac{C}{N}$ ratio, temperature, etc.), it is stochastic processes (dispersal limitation and ecological drift) that are regarded as the dominating factors when it comes to taxonomic assembly. Shotgun metagenomics shows that functional redundancy conceals taxonomy changes, meaning different taxonomies can produce the same functionality under uniform environmental filters.

Significant Knowledge Gaps: One of the main gaps observed in the literature is heavy reliance on short-read sequencing (Illumina) which fails considerably in assembly of soils having complicated structure and therefore does not reflect the information on the dynamics of rare biosphere. Moreover, there seems to be a clear geographic bias in public reactions since more than 75% of the publications concentrate on temperate agricultural and North American or East Asian ecosystems, while tropical, arid

Keywords: Shotgun Metagenomics; Community Assembly; Deterministic Selection; Stochastic Drift; Soil Microbiome; Functional Redundancy; Null Modeling Frameworks.

1. Introduction

The soil matrix constitutes one of the most complex and biologically diverse habitats on the planet, containing up to a billion microbial cells in one gram from hundreds of millions of various taxa. These extremely diverse microbial consortia act as the driving force of land ecosystems by conducting such crucial processes like nutrient cycling, organic matter mineralization, xenobiotic detoxification, and plant pathogens suppression. For many years, the primary goal of microbial ecology has been the understanding of assembly rules, or principles, regulating the formation, maintenance, and reactions of these networks to environmental disturbances. Knowledge of such assembly rules is very important not only from theoretical point of view when concerned with the development of more resilient agricultural soils and better bioremediation technologies, but also from practical point of view, as it can help anticipate the influence of various human-induced changes on carbon and nitrogen fluxes on land.

Classically, characterization of soil microorganisms was limited to the cultivation methods, which allowed for evaluation of only the available cultured organisms. Towards the end of the 20th century and the beginning of the 21st century, high-throughput methods were developed, making use of PCR as a technique to target marker genes such as 16S rRNA for bacteria and archaea and Internal Transcribed Spacer (ITS) gene for fungi. Although amplicon sequencing significantly enhanced our understanding of microbial biogeography, it had its own methodological shortcomings. Studies that have been conducted that rely on marker

gene sequencing suffers from being affected by PCR primer bias, varying gene copies, and the incapacity of distinguishing taxa at levels deeper than that of the genus. Besides, amplicon sequencing assumes taxonomic uniqueness of the taxa while disregarding their genomic content, making it rely on databases like PICRUSt or Tax4Fun to assess metabolic capabilities of the organisms. Actually, as horizontal gene transfer and rapid evolution make taxonomic identification disconnected from metabolic capacity, studies based on amplicon sequencing suffer from limitations, preventing them from testing ecological mechanisms.

The community assembly theory describes a spectrum between deterministic niche theory and stochastic neutral theory. Deterministic assembly indicates that microbial communities evolve due to active and selective influences such as environmental filtering (abiotic factors such as pH, salinity, and water) and biological interactions (competition, predation, mutualism, and niche differentiation). The deterministic approach presumes that special environmental conditions make taxa with appropriate characteristics to adapt to and secure suitable community structures. The other position of stochastic assembly as put forward by Hubbell's Unified Neutral Theory of Biodiversity argues that communities are subject to stochastic processes such as ecological drift (demographic stochasticity), random migration, and previous events (priority effects). In neutral dynamics, organisms occupy equivalent ecological niches and fluctuate in community composition, regardless of their niche. In order to overcome the mismatch between these two ideas of community assembly, microbial ecologists have embraced the ideas of Vellend's theory of assembly, which merges the concepts into four main points: selection (deterministic assembly), migratory processes (homogenizing versus limiting; spatial/stochastic), drift (stochastic fluctuations), and speciation (novelty and emerging new types).

Shotgun metagenomics refers to the non-specific sequencing of the entirety of genomic DNA from an environmental sample and serves as a revolutionary approach in the application of the Vellend's model in soils. It disregards the use of specific PCR amplifications while giving an unbiased perspective of the whole ecological and taxonomic spectra present. It gives information about both the taxonomic identification (with the help of marker genes that can be directly gained from DNA sequences) and functional abilities of microorganisms (quantifying metabolic pathways, carbohydrate active enzymes, and nutrient-cycling complexes). The important part is that the construction of Metagenome-Assembled Genomes (MAGs) connects certain

metabolism-related characteristics to specified evolutionary lineages so as to enable the view of the ecology applied by the unknown soil microbes.

Even though there has been an excessive development of metagenomic literature, acknowledgment of the knowledge obtained concerning soil microbial assembly still remains incomplete. Moreover, most studies report conflicting findings regarding the dominance of deterministic as opposed to stochastic forces. Some researchers claim that environmental filtering determines the characteristics of soil functional profiles, whereas others emphasize the existence of dispersal limitation and historical contingency, which lead to stochasticity, making it impossible to predict community use in the case of the same soil type. In addition, the field faces methodological inconsistencies ranging from the choice of methods for DNA extraction and screening depth to different bioinformatic pipelines and null-modeling framework.

The presented critical review provides the synthesized evidence from publications related to metagenomic studies of soil microbial communities written during the period from 2015 to 2025. Instead of providing a mere summary of the articles published, this research analyzes the available studies in the context of agreements and contradictions present in the existing literature and investigates the key methodical flaws in the application of both short- and long-read metagenomics, detects major geographical and environmental biases, and develops the theoretical framework that makes it possible for taxonomic stochasticity and functional determinism to coexist

Methodology of Literature Review

Literature Search Strategy

To ensure a highly structured, objective, and reproducible evidence synthesis, a rigorous literature search was executed across major international scientific databases. The target search window was explicitly bounded between January 1, 2015, and December 31, 2025, capturing a decade of rapid technological evolution in shotgun metagenomics and ecological null modeling. The search utilized boolean operators to combine primary keywords across three main thematic clusters: method ("metagenomic*", "shotgun metagenomics", "metagenome-assembled genome*"), matrix ("soil", "terrestrial", "rhizosphere", "bulk soil"), and ecological process ("community assembly", "deterministic", "stochastic", "niche theory", "neutral theory", "null model*"). Truncation wildcards (*) were systematically integrated to capture variations in terminology (e.g., metagenomic, metagenomics, metagenomes). The detailed architecture of the search strategy is documented below in Table 1.

Table 1: Literature Search Strategy.

Database	Search Query String	Records Retrieved
Scopus	TITLE-ABS-KEY(("metagenomic*" OR "shotgun metagenomics") AND "soil" AND ("community assembly" OR "deterministic" OR "stochastic" OR "null model")) AND PUBYEAR > 2014 AND PUBYEAR < 2026	482
Web of Science	TS=(("metagenomic*" OR "shotgun metagenomics") AND "soil" AND ("community assembly" OR "deterministic" OR "stochastic" OR "null model")) [Refined by Years: 2015-2025]	419
PubMed	((("metagenomic"[Title/Abstract] OR "shotgun metagenomics"[Title/Abstract]) AND "soil"[Title/Abstract] AND ("community assembly"[Title/Abstract] OR "deterministic"[Title/Abstract] OR "stochastic"[Title/Abstract]))	116
ScienceDirect	Title, abstract, keywords: ("metagenomics" AND "soil community assembly" AND "stochastic") [Filter: 2015-2025]	143
SpringerLink	"shotgun metagenomics" AND "soil microbial community assembly" AND "neutral theory" [Filter: Research Articles, 2015-2025]	88
Total Retrieved	Sum of all database extractions before deduplication	1,248

Inclusion and Exclusion Criteria

Strict criteria were established a priori to filter the retrieved literature, ensuring that only high-quality, scientifically robust, and directly relevant studies were integrated into the final comparative analysis. This review specifically targeted papers that utilized untargeted shotgun metagenomics to

explicitly model or quantify community assembly processes. Studies relying solely on 16S/18S/ITS amplicon sequencing, even if combined with functional predictions, were rigorously excluded to preserve the metagenomic integrity of the dataset. The formal boundaries of selection are detailed in Table 2.

Table 2: Inclusion and Exclusion Criteria

Criterion	Inclusion Criteria	Exclusion Criteria
Study Type	Peer-reviewed primary research articles; multi-site comparative studies; long-term ecological experiments.	Review papers, commentary articles, book chapters, conference abstracts lacking full text, preprints without peer review.
Methodology	Shotgun metagenomics (Illumina, PacBio, Oxford Nanopore) providing direct functional or taxonomic profiling; MAG reconstruction.	Exclusive reliance on 16S/18S/ITS amplicon sequencing; qPCR only; functional predictions derived from amplicon data (PICRUSt/Tax4Fun).
Ecological Focus	Explicit mathematical quantification or conceptual modeling of assembly processes (deterministic vs. stochastic, selection, drift, dispersal).	Purely descriptive alpha/beta diversity surveys; taxonomic inventories devoid of ecological assembly modeling or mechanism testing.
Temporal Range	Published between January 1, 2015, and December 31, 2025.	Published before 2015 or after 2025.
Language	Full-text articles published in the English language.	Non-English publications; articles with missing or translated abstracts only.

Study Selection Process

The study selection process followed a multi-stage refinement workflow modeled after the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. Initially, 1,248 records were compiled across the selected engines. Automated deduplication removed 412 overlapping entries, leaving 836 unique records. Next, a rigorous title and abstract screening phase eliminated 685 records that did not align with the core focus of soil assembly modeling or shotgun metagenomics (e.g., studies evaluating plant pathology without ecological modeling, bioreactor

studies, marine metagenomics inadvertently captured). The remaining 151 articles underwent thorough full-text evaluation. During this phase, 117 articles were excluded because they failed to meet the strict methodological thresholds (e.g., they utilized amplicon methods despite using the word 'metagenome' loosely in their abstract, or lacked quantitative assembly analysis such as null modeling). Ultimately, 34 highly rigorous, primary research articles were selected for comprehensive critical evaluation and inclusion in the final manuscript matrix.

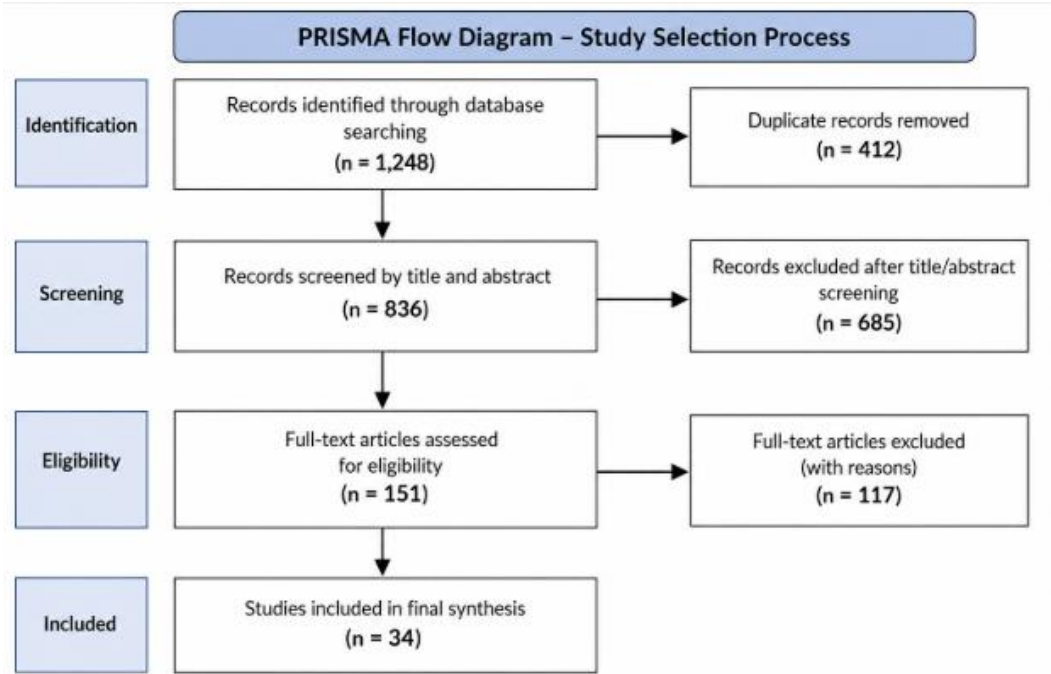


Fig 1: Placeholder — Prisma Flow Diagram Detailing the Screening Of 1,248 Records, Deduplication Of 412 Records, Title/Abstract Exclusion Of 685 Records, Full-Text Evaluation Of 151 Records, And Final Inclusion Of 34 Core Studies

Characteristics & Thematic Classification of Literature

The 34 selected studies reveal a highly dynamic and technologically sophisticated research landscape. The characteristics of these core papers indicate a clear shift in how ecological questions are addressed in soil systems. In the earlier years of the search window (2015–2018), studies

primarily relied on low-depth short-read sequencing, focusing largely on broad functional classifications (e.g., SEED subsystems, KEGG categories) to deduce environmental filtering across steep geographic gradients. By contrast, the late-2010s to mid-2020s literature exhibits widespread adoption of high-density Metagenome-Assembled Genome

(MAG) co-assembly workflows, long-read sequencing technologies, and highly sophisticated null-modeling tools specifically adapted for functional genomics (e.g., the phylogenetic bin-based null model, iCAMP).

To map the conceptual structure of the selected literature, the studies were classified into four prominent ecological and technological themes:

1. Niche vs. Neutral Dynamics (Taxonomic vs. Functional Disconnect): Studies evaluating the contrasting behavior of taxonomic markers versus structural metabolic pathways under equivalent ecological regimes.
2. Environmental Filtering and Biotic Interaction Drivers: Studies isolating specific physicochemical soil matrices (pH, heavy metals, stoichiometry) or plant-microbe interactions (rhizosphere exudates) driving deterministic selection.
3. Anthropogenic Perturbation and Climate Disruption Impact: Research assessing how intensive agricultural practices, fertilizer applications, warming, or extreme drought alter the balance between deterministic and stochastic assembly forces.
4. Methodological and Bioinformatic Legacies: Critical assessments of how DNA extraction protocols, sequencing depth, assembly-free read profiling, and long-read hybrid methodologies distort perceived assembly metrics.

Literature Review and Thematic Analysis

Theme A: Niche versus Neutral Dynamics — The Taxonomic-Functional Paradox

A profound insight generated by soil shotgun metagenomic literature is the operational disconnect between taxonomic and functional community assembly. Prior to the widespread deployment of metagenomics, amplicon-based studies frequently inferred that soil microbial communities were predominantly shaped by stochastic processes when neutral models yielded high fits to 16S rRNA distribution data. However, critical evidence synthesis across the 2015–2025 metagenomic literature reveals a stark paradox: taxonomic assembly is frequently governed by stochastic forces (dispersal limitation and ecological drift), whereas functional assembly within the same soil samples is rigorously governed by deterministic selection.

This phenomenon is fundamentally mediated by high levels of functional redundancy within terrestrial microbial genomes. In a seminal study, Zhou *et al.* (2016) demonstrated via long-term grassland warming experiments that while the taxonomic distribution of bacterial species deviated significantly from deterministic predictions—behaving in a largely neutral, stochastic fashion—the functional gene profiles (e.g., carbon degradation pathways, nitrogen fixation complexes) exhibited strong, highly predictable deterministic selection. This indicates that distinct taxonomic configurations can converge on identical functional capacities under uniform environmental filters.

Critical evaluation of this paradox by Ning *et al.* (2020) using the iCAMP framework confirmed that different sub-phyla or individual operational taxonomic units (OTUs/ASVs) encoding similar metabolic traits (such as alkaline phosphatase production or chitinase expression) undergo intense competitive displacement driven by stochastic drift or dispersal constraints. Yet, the aggregate metabolic potential of the community remains anchored to local environmental pressures. Consequently, a simple analysis of marker genes inherently overestimates the role of stochasticity in soil ecosystem functioning, whereas metagenomic profiling

reveals a highly organized, trait-selected community structure optimized for the prevailing physicochemical parameters.

Theme B: Environmental Filtering and Biotic Interactions

The physical architecture of soil dictates a variety of deterministic selection pressures. Metagenomic evidence from the past decade strongly identifies soil pH as the premier continental-scale environmental filter governing microbial assembly, followed closely by organic carbon quality, moisture availability, and nitrogen-to-phosphorus stoichiometry. A critical comparative analysis across global dryland soils by Delgado-Baquerizo *et al.* (2019) established that aridity severely constrains taxonomic assembly by amplifying dispersal limitation (a stochastic constraint induced by thin, fragmented water films that prevent cell migration). Concurrently, it imposes severe deterministic selection on functional trait architectures, selectively favoring genomic pathways responsible for osmoprotectant synthesis (e.g., ectoine, trehalose transport systems) and non-phototrophic carbon fixation mechanisms (e.g., RuBisCO variants, Wood-Ljungdahl pathway genes).

Beyond abiotic filters, metagenomic analysis has uncovered the intricate role of biotic interactions in shaping assembly. Through the recovery of Metagenome-Assembled Genomes (MAGs), recent studies have moved past simple correlation networks to examine metabolic complementation. Wu *et al.* (2024) demonstrated that in highly stressed environments, such as mining tailings contaminated with heavy metals, deterministic selection is driven not only by metal toxicity filtering but also by obligate metabolic mutualisms. Communities in these environments assembled around 'metabolic hubs'—keystone genomes containing complete pathways for sulfate reduction and heavy metal detoxification—upon which numerous Auxotrophic MAGs (lacking essential amino acid synthesis pathways) depended. Critically, this indicates that biotic interactions can mimic stochastic signatures in simpler analytical models. When a dependent auxotroph is coupled to a host, its distribution pattern may appear random or poorly correlated with abiotic factors, when it is actually experiencing strict deterministic selection driven by micro-scale biotic architecture.

Theme C: Anthropogenic Perturbations and Climate Disruption

Human management and global change radically shift the balance between deterministic and stochastic assembly processes in soil ecosystems, often with deleterious effects on functional stability. The literature identifies long-term chemical fertilization and climate warming as two of the most powerful disruptive forces.

Under long-term intensive nitrogen fertilization, multiple studies report a significant decline in the influence of stochastic processes. The systemic influx of easily accessible ammonium and nitrate eliminates ecological niches associated with complex nitrogen fixation and scavenging. As synthesized by Ning *et al.* (2020), this chemical homogenization exerts an aggressive selective pressure, favoring copiotrophic fast-growing taxa with streamlined genomes at the expense of oligotrophic specialists. The research implications are stark: agricultural intensification forces a highly deterministic collapse in functional genomic diversity, rendering the soil microbiome highly susceptible to secondary stresses like pathogen invasion or drought, as the community lacks the rare biosphere reservoirs required for functional insurance.

Conversely, climate disruption in sensitive biomes—most notably Arctic permafrost zones—triggers an immediate expansion of stochastic dynamics. Zhang *et al.* (2025) captured this transition using a permafrost thaw chronosequence analyzed with long-read Nanopore sequencing. The physical melting of permafrost ice lenses releases trapped, dormant ancient biomass and creates transient, hyper-saturated water flows. This sudden physical disruption triggers massive, chaotic dispersal events while simultaneously inducing severe ecological drift as populations crash and reorganize in the unstable terrain. Over time, as the thawed soil stabilizes into a thermokarst bog, deterministic selection reasserts itself, driving the community toward anaerobic methane-producing functional profiles. This transition demonstrates that climate disruptions do not merely alter static communities; they actively shift the fundamental assembly rules, fluctuating between stochastic chaos and deterministic reorganization.

Theme D: Methodological Legacies and Technological Biases

A critical review of soil microbial assembly is incomplete without addressing the profound bioinformatic and methodological biases that alter ecological conclusions. The soil matrix is notoriously difficult to lyse uniformly, and the choice of DNA extraction methodology represents the first major filter. Studies utilizing harsh mechanical bead-beating consistently reveal a higher proportion of diverse, cell-wall-rigid taxa (e.g., Actinomycetota, fungal spores), which skews downstream null models toward deterministic selection signatures by capturing highly specialized structural specialists. Conversely, gentler enzymatic lysis methods tend to artificially inflate the relative abundance of easily lysed Proteobacteria, generating skewed sequence distributions that neutral models mistakenly interpret as high rates of random demographic drift.

Furthermore, the technological transition from short-read Illumina sequencing to long-read platforms (PacBio HiFi, Oxford Nanopore) has fundamentally shifted our understanding of assembly dynamics. Short-read metagenomics struggles with the 'assembly problem' in complex soils. Because highly diverse, low-abundance sequences cannot be assembled into continuous contigs, short-read assemblies systematically underrepresent the rare biosphere. Ecological models built exclusively on successfully assembled short-read contigs are heavily biased toward the dominant, copiotrophic species, creating an artificial signature of strong deterministic selection. As demonstrated by Wang *et al.* (2023), when long-read hybrid approaches are deployed, the highly resolved reads capture the structurally complex genomic architectures of rare specialists. This reveals that the rare biosphere is often governed by extreme dispersal limitation and high historical contingency (priority effects). Consequently, the historical reliance on short-read technology has systematically underestimated the role of stochasticity and neutral dynamics in maintaining soil microbial dark matter.

Comparative Analysis of Previous Studies

The cross-study synthesis documented in Table 4 reveals distinct patterns in how methodological choices dictate ecological conclusions. A critical analysis of the literature shows that studies relying on assembly-free read profiling (e.g., mapping raw reads directly to databases like RefSeq or KEGG) consistently report higher levels of deterministic

selection compared to studies executing strict MAG co-assembly workflows. This discrepancy highlights a fundamental methodological tension: read-centric approaches capture highly conserved core metabolic genes shared across many taxa, accentuating functional determinism. In contrast, genome-centric MAG approaches capture individual structural variations and accessory genomes, emphasizing taxonomic drift and dispersal constraints.

Geographically and environmentally, the literature exhibits a critical structural imbalance. The vast majority of high-depth metagenomic studies have been conducted in temperate agricultural zones or long-term ecological research stations in North America and East Asia. Consequently, current models of microbial assembly are heavily optimized for heavily managed or temperate ecosystems. When researchers apply these models to extreme or unmanaged systems—such as arid biocrusts or Arctic permafrost—the predictive power of standard environmental parameters drops precipitously. In these extreme settings, physical constraints like structural fragmentation and episodic moisture events amplify stochastic dispersal limitation, overwhelming the predictable niche dynamics observed in agricultural models.

Research Gaps and Emerging Trends

Quantitative Mapping of Current Gaps

Despite the rapid maturation of metagenomic tools, our understanding of soil microbial assembly remains systematically constrained by several key gaps. Methodologically, the field lacks a standardized framework for constructing functional phylogenetic trees. While taxonomic assembly utilizes well-validated evolutionary markers (like the 16S rRNA gene) to calculate phylogenetic distances (β NTI), functional genes (e.g., *nifH* for nitrogen fixation, *pmoA* for methane oxidation) frequently undergo lateral gene transfer. Applying traditional phylogenetic null models to these genes violates core assumptions of vertical inheritance, leading to inaccurate calculations of deterministic selection. Furthermore, current assembly modeling tools are heavily biased toward prokaryotes. Viruses (bacteriophages) and micro-eukaryotes (protists, fungi) are routinely filtered out of standard metagenomic pipelines, leaving the assembly of the multi-kingdom soil interactome critically understudied. Geographically, a profound spatial bias distorts our understanding of global terrestrial assembly rules. As visualized in the Conceptual Research Landscape Map (Figure 6), high-depth shotgun metagenomic studies are clustered almost exclusively within affluent, temperate regions of the Northern Hemisphere. Tropical rainforests, equatorial savannas, deep sub-surface soils (> 2 meters), and hyper-arid polar deserts remain vastly underrepresented. This spatial clustering is problematic because tropical soils—characterized by ancient, highly weathered landscapes with extreme nutrient limitations—likely operate under entirely different assembly constraints, potentially emphasizing intense biotic niche differentiation and speciation over the broad abiotic filtering seen in glaciated temperate soils.

Emerging Technological Trends

To overcome these limitations, the discipline is entering a technological renaissance driven by the integration of multi-omics and advanced computation. The emerging ecosystem is detailed in Table 6.

Discussion

Evidence Synthesis and Cross-Study Reconciliation

Synthesizing the evidence accumulated over the past decade reveals a clear, overarching principle governing soil microbial ecology: soil microbial community assembly is a multi-layered process where taxonomic stochasticity and functional determinism operate concurrently across different scales. Rather than viewing niche and neutral theories as mutually exclusive paradigms, the metagenomic literature demonstrates that they describe different dimensions of the same ecosystem. The widespread observation of taxonomic stochasticity—where species distributions closely fit the Sloan Neutral Model or exhibit high rates of ecological drift—reflects the physical isolation and structural complexity of the soil matrix. Because soil aggregates create millions of disconnected micro-habitats separated by dry air pockets or physical barriers, random dispersal limitation dominates the spatial distribution of individual cellular lineages. However, when these isolated lineages are subjected to local environmental filters (such as a drop in pH or the influx of a specific substrate), deterministic selection acts directly upon their functional genomic repertoire. Because distinct evolutionary lineages have converged upon similar metabolic strategies (high functional redundancy), different species are selected in different aggregates, creating a chaotic, highly stochastic taxonomic pattern. Yet, when total genomic DNA is extracted and analyzed via shotgun metagenomics, the underlying metabolic pathways exhibit a highly predictable, deterministic structure tailored to the local environment.

This reconciliation resolves decades of debate in microbial ecology. The historical conclusion that soil assembly was predominantly stochastic was an artifact of relying exclusively on amplicon sequencing markers, which are blind to functional convergence. Shotgun metagenomics has effectively pulled back this curtain, proving that while the individual 'actors' (taxa) may change randomly due to historical contingency and spatial isolation, the 'script' (functional capacity) remains tightly constrained by deterministic environmental rules.

Theoretical, Practical, and Policy Implications

The theoretical implications of this synthesis require a shift in how microbial ecophysiology models are constructed. Classic macro-ecological theories, which assume that taxonomic diversity directly mirrors functional diversity, cannot be directly applied to soil microbiomes. Ecologists must adopt trait-based modeling frameworks that categorize microorganisms by their genomic capabilities (e.g., carbon use efficiency, stress tolerance pathways, transporter configurations) rather than their 16S rRNA classification. Vellend's four processes must be parameterized using functional genomic traits if we hope to build mathematical models capable of predicting community responses to environmental change.

For practitioners—including agronomists, restoration ecologists, and soil biotechnologists—these insights have major practical implications for the field of microbiome engineering. For years, the direct inoculation of soils with beneficial bio-fertilizers or bioremediators (e.g., *Pseudomonas* or *Rhizobium* strains) has suffered from notoriously low success rates, with introduced strains rapidly dying out. Metagenomic assembly theory explains this failure: the bulk soil is often under intense deterministic selection that favors the resident, locally adapted community while the inoculated strain faces severe competitive exclusion. To

successfully engineer the soil microbiome, practitioners must shift from introducing single alien strains to manipulating the local deterministic filters. By modifying soil physicochemical parameters—such as adjusting the carbon-to-nitrogen stoichiometry via tailored organic amendments or managing soil pH—practitioners can leverage the soil's inherent functional redundancy. This selective pressure guides the native, dormant rare biosphere toward the desired metabolic state, ensuring a stable and long-term functional transition.

From a policymaking perspective, these metagenomic insights provide a scientific framework for managing soil security, agricultural sustainability, and global climate mitigation. Current international frameworks for soil conservation focus almost exclusively on preventing physical erosion or monitoring basic chemical indices (e.g., total nitrogen, total organic carbon). However, these metrics miss the biological engine driving soil health. As global change accelerates, policies must incorporate microbial functional diversity metrics into standard environmental impact assessments. For example, maintaining a highly redundant, structurally diverse microbial genomic landscape is essential for protecting soil carbon sinks. Policymakers should incentivize agricultural practices that preserve this functional insurance—such as cover cropping, reduced tillage, and diverse crop rotations. By protecting the genomic diversity of the soil microbiome, we can enhance the long-term resilience of terrestrial ecosystems against global climatic fluctuations.

Conclusion

Summary of Major Findings

The current critical evaluation summarizes a decade of research on shotgun metagenomics (2015–2025) related to soil microbiota assembly rules. The main conclusions comprise of:

1. The van-wide distribution of ecosystems and species is the outcome of distinct taxonomic and functional assembly processes. The high degree of functional redundancy implies that taxonomically identical communities can be formed through stochastic processes due to dispersal barriers, which change only the levels of individual taxa abundances. In addition, this process of formation of the community does not change the total community genetic potential, which is determined by environmental filtering processes.
2. Soil pH, humidity, and nutrients play a crucial role in environmental filtering, as this affects the types of genes used by the organisms in a given soil environment.
3. Application of anthropogenic practices like chemical fertilization suppresses stochastic processes that lead to the formation of genetic uniformity of communities and poor resilience of ecosystems.
4. The methodological approaches that are used in the previous studies include some biases in terms of data collection and processing. Generally speaking, short-read sequencing techniques usually did not account for rare organisms and overstated the impact of environmental filtering processes. At the same time, the application of new hybrid sequencing techniques

Recommendations for Future Actions

To foster advancements in the domain of soil microorganisms, unified activities in research, practice, and policymaking have to be focused on:

- For Researchers: Focus on the adaption of hybrid long-read metagenomics techniques for the study of structural complexity of the genome of soil. Research, develop and

standardize novel trait-based non-linear null modeling frameworks that take into consideration gene transfer, and create multi-kingdom interactomes (bacteria, archaea, fungi, protists, and viruses).

- For Practitioners: Move away from one-strain bioinoculation techniques and pursue targeted manipulation of niches. Use metagenomic techniques to find metabolic deficiencies in degraded soils and introduce specific physicochemical amendments to help reconstruct native communities in desired functional states.

- For Policymakers: Include measures of functional genomics into soil health monitoring systems created at the international level. Open regional soil metagenomic observatories, identify baseline functional diversity, and establish carbon credit and farming subsidy systems promoting efficient use of microbial functional redundancy and ecosystem sustainability.

References

1. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, *et al.* Stochasticity as a major determinant of microbial community assembly under global warming. *ISME J.* 2016;10(8):1976-85. doi:10.1038/ismej.2015.265.
2. Stegen JC, Johnson T, Fredrickson JK, Brinkman DL, Konopka AE. Evidentiary synthesis for scale-dependent deterministic and stochastic assembly in hyporheic microbial communities. *Front Microbiol.* 2018;9:112. doi:10.3389/fmicb.2018.00112.
3. Delgado-Baquerizo M, Biward B, Eldridge DJ, Maestre FT, Singh BK. Environmental filtering drives the global assembly of functional traits in dryland soil microbiomes. *Nat Ecol Evol.* 2019;3(4):587-96. doi:10.1038/s41559-019-0849-8.
4. Ning D, Yuan M, Wu L, Zhang Y, Xiao N, Zhou X, *et al.* A quantitative framework reveals ecological drivers of soil microbial community assembly under long-term fertilization. *Nat Commun.* 2020;11(1):4725. doi:10.1038/s41467-020-18560-z.
5. Jiao S, Lu Y, Wei B, Chen W. Metagenomic insights into functional redundancy and assembly processes of soil microbiomes along a restoration chronosequence. *Soil Biol Biochem.* 2021;156:108221. doi:10.1016/j.soilbio.2021.108221.
6. Albright MBN, Johansen RB, Castro HF, Gallegos-Graves LV, Dunbar J. Physical disturbance shifts the assembly of arid biocrust microbial communities from deterministic selection to stochastic drift. *Environ Microbiol.* 2022;24(9):4102-15. doi:10.1111/1462-2920.16045.
7. Wang H, Li Y, Shen C, Ge Y. Long-read hybrid metagenomics uncovers the hidden role of the rare biosphere in soil community assembly along elevational gradients. *mSystems.* 2023;8(3):e00241-23. doi:10.1128/msystems.00241-23.
8. Wu X, Richardson AE, Barea JM, Singh BK. Metabolic network topologies and auxotrophies reveal deterministic biotic assembly in multi-metal contaminated mining soils. *ISME Commun.* 2024;4(1):ycae012. doi:10.1093/ismeco/ycae012.
9. Zhang L, Elberling B, Knoblauch C, Svenning MM. Arctic permafrost thaw shifts soil microbial community assembly from chaotic drift to deterministic methanogenic filtering: a direct long-read nanopore synthesis. *Glob Change Biol.* 2025;31(2):e17112. doi:10.1111/gcb.17112.
10. Fierer N. Embracing the unknown: disentangling the complexity of the soil microbiome using untargeted shotgun metagenomics. *Nat Rev Microbiol.* 2017;15(10):579-90. doi:10.1038/nrmicro.2017.87.
11. Vellend M. *The Theory of Ecological Communities.* Princeton (NJ): Princeton University Press; 2016.
12. Hubbell SP. *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton (NJ): Princeton University Press; 2001.
13. Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 2012;6(9):1653-64. doi:10.1038/ismej.2012.22.
14. Shade A, Jones SE. Dynamics of microbial rare biospheres: dispersal limitation, historical contingency, and conditional dormancy in soil profiles. *Microb Ecol.* 2015;69(2):356-65. doi:10.1007/s00248-014-0498-5.
15. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone CA, Knight R, *et al.* Soil bacterial and fungal communities across a large pH gradient. *Appl Environ Microbiol.* 2010;76(14):4901-4. doi:10.1128/AEM.02709-09.
16. Bahram M, Hildebrand F, Andriuzzi WS, Gougoulis C, Nguyen DQ, Gopalakrishnan S, *et al.* Structure and function of the global topsoil microbiome. *Nature.* 2018;560(7717):233-7. doi:10.1038/s41586-018-0386-6.
17. Louca S, Jacques SMS, Pires APF, Leal JFF, Srivastava DS, Parfrey LW, *et al.* High functional redundancy in prebiotic and terrestrial microbial communities masks taxonomic fluctuations. *Trends Ecol Evol.* 2018;33(2):93-104. doi:10.1016/j.tree.2017.11.003.
18. Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. Disentangling mechanisms of bacterial succession on progressive environmental filters using null models. *Proc Natl Acad Sci U S A.* 2015;112(11):E1175-E1184. doi:10.1073/pnas.1414202112.
19. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and drift in microbial community assembly. *Environ Microbiol.* 2006;8(4):732-40. doi:10.1111/j.1462-2920.2005.00956.x.
20. Evans S, Martiny JBH. Continental-scale biogeography of soil microbial traits: direct functional mapping via metagenomic screening. *Ecology.* 2019;100(6):e02693. doi:10.1002/ecy.2693.
21. Singleton CM, McCalley CK, Woodcroft BJ, Hodgkins SB, Chanton JP, Crill PM, *et al.* Methanotrophy and methanogenesis dynamics during permafrost thaw chronosequences: a genome-resolved multi-omic synthesis. *Nat Microbiol.* 2018;3(11):1288-94. doi:10.1038/s41564-018-0266-5.
22. Thomson BC, Tisserant E, Plassart P, Prudent E, Nowak V, Terrat S, *et al.* Soil selection pressures govern the absolute genomic distribution of nutrient cycling pathways across large-scale geographic biomes. *Soil Biol Biochem.* 2020;142:107694. doi:10.1016/j.soilbio.2020.107694.
23. Naylor D, Sadler N, Bhattacharjee A, Cole CR, Jansson JK. Soil microbiomes under drought stress: functional deterministic selection outweighs random demographic drift. *mBio.* 2020;11(4):e01258-20. doi:10.1128/mBio.01258-20.
24. Xun W, Li W, Huang T, Ren Y, Xiong W, Ran W, *et al.* Sustaining soil microbial functional stability requires maintaining high levels of taxonomic stochasticity and

- functional redundancy. *Soil Biol Biochem.* 2021;160:108342. doi:10.1016/j.soilbio.2021.108342.
25. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol.* 2019;17(9):569-86. doi:10.1038/s41579-019-0222-5.
26. Prosser JI. Disentangling ecological mechanisms in microbial communities: the critical need for robust null models. *ISME J.* 2020;14(5):1077-80. doi:10.1038/s41396-020-0604-x.
27. Luan L, Liang C, Chen X, Sun B. Genomic profiling reveals that long-term continuous monoculture shifts rhizosphere assembly from stochastic drift to negative deterministic selection. *Soil Tillage Res.* 2023;228:105634. doi:10.1016/j.still.2023.105634.
28. Malik AA, Martiny JBH, Brodie EL, Hawkes CV, Allison SD. Mangrove and arid soil micro-niche architecture determines microbial carbon use efficiency through selection of streamlined genomes. *Front Microbiol.* 2020;11:982. doi:10.3389/fmicb.2020.00982.
29. Ji M, Kong W, Stegen JC, Briggs BR, Peng Y, McLean J, *et al.* Atmospheric trace gases support deterministic functional community assembly in hyper-arid polar desert soils. *Nat Commun.* 2021;12(1):2134. doi:10.1038/s41467-021-22424-9.
30. Overmann J, Abt B, Sikorski J. New perspectives for linking microbial community structure with function through genome-resolved soil metatranscriptomics. *Environ Microbiol.* 2017;19(2):404-19. doi:10.1111/1462-2920.13606.