# Heterogeneity in Microbial Systems: Scales, Drivers, and Consequences

Zhanshan (Sam) Ma
Computational Biology and Medical Ecology
Chinese Academy of Sciences
Harvard Forest, Faculty of Arts and Sciences
Harvard University, Cambridge, MA, USA
ma@vandals.uidaho.edu

## Table of Contents

## Abstract

Heterogeneity is a fundamental, multi-scale organizing principle of microbial life, transcending its historical treatment as mere statistical noise. This review synthesizes recent advances to construct a unified framework for understanding microbial heterogeneity across hierarchical scales—from molecular and cellular phenotypic diversity, through strain-level genomic micro-diversity, to interpersonal variation and landscape-level spatial mosaics. We articulate how this heterogeneity is generated by the dynamic interplay of stochastic and deterministic forces, including rapid microbial evolution and phage activity, and reframe it as a functional, adaptive trait enabling bet-hedging, division of labor, and ecosystem stability. The field's conceptual evolution is traced from descriptive phenomenology to a predictive, mechanism-driven phase, now supported by integrative frameworks like medical ecology. A recent synthesis (Ma & Ellison, 2026) provides a dual-thread paradigm distinguishing environmental heterogeneity (the abiotic template) from ecological heterogeneity (biotic variation expressed on that template)—a distinction that anchors this review. To navigate this complexity, we outline a methodological synthesis centered on multi-dimensional metagenomics, integrating strain-resolved genomics, spatial omics, AI/ML analytics, and causal inference to move from correlation to mechanism. Finally, we detail the translational roadmap enabled by this heterogeneity-aware perspective, including precision microbial therapeutics based on ecological competition principles, personalized nutrition via strain-matched synbiotics, and robust clinical translation frameworks. We also address critical gaps in sampling representation, ethics, as well as embracing cross-disciplinary conceptual and methodological developments. Embracing heterogeneity is not an analytical obstacle but the essential path to a predictive, impactful microbiome science.

**Keywords:** heterogeneity; microbiome; medical ecology; bet-hedging; strain-level diversity; precision therapeutics; template; environmental heterogeneity; ecological heterogeneity

## 1. Introduction: Heterogeneity as the Organizing Principle of Microbial Life

Heterogeneity is not merely a statistical artifact or a confounding variable in microbiome research; it is a fundamental, multi-dimensional, and adaptive property that structures microbial life across all scales of biological organization (Lidstrom & Konopka, 2010). From stochastic fluctuations in gene expression within isogenic populations to continental-scale gradients in microbial community composition, heterogeneity pervades microbial systems and dictates their ecological dynamics, evolutionary trajectories, and functional outputs. Historically, such variation was often treated as "noise" to be averaged out, but a paradigm shift is now underway. Contemporary research increasingly frames heterogeneity as a core explanatory principle—a source of resilience, a driver of complexity, and a scaffold for adaptation (Heckman 2001, Shavit & Ellison 2021; Eisenhauer et al. 2023; Elliott-Graves 2025; Ma & Ellison 2025, 2026). This shift reflects the maturation of microbiome science from descriptive cataloguing toward mechanistic, predictive modeling underpinned by integrative frameworks like medical ecology, which cross-disciplinarily studies human microbiomes to understand their implications for health and disease from an ecological perspective and is supported by computational biology and bioinformatics (Ma & Zhang, 2022).

A foundational conceptual distinction, articulated by Shavit & Ellison (2021), separates **diversity**—variation within non-interacting populations—from **heterogeneity**—variation within interacting collectives whose constituent entities integrate into larger, functioning wholes. This distinction is crystallized by their aphorism: "a zoo is diverse, whereas an ecosystem is heterogeneous." A zoo aggregates non-interacting populations; an ecosystem comprises interacting populations whose relationships generate emergent properties. Building on this foundation, a recent synthesis by Ma & Ellison (2026) consolidates current understanding of heterogeneity across ecology and related fields, articulating a dual-thread framework distinguishing **environmental heterogeneity**—the abiotic template of physical structure, resource gradients, and disturbance regimes—from **ecological heterogeneity**—the biotic variation in organismal distributions, traits, and interactions expressed *on* that template. This distinction maps precisely onto the template concept we develop here: the host environment (e.g., gut, soil, forest, ocular surface) constitutes environmental heterogeneity; the microbial community constitutes ecological heterogeneity. Understanding microbial variation requires understanding the templates that shape it, a principle that unifies human and environmental microbiome research.

In this review, we synthesize a broad and integrative body of recent literature to construct a coherent conceptual framework for understanding microbial heterogeneity. We argue that heterogeneity operates across hierarchical scales—molecular, cellular, strain, host, community, and ecosystem—and is generated by the dynamic interplay of stochastic and deterministic forces, including rapid microbial evolution. Crucially, we emphasize that heterogeneity is not passive but

functional: it enables bet-hedging against environmental uncertainty, facilitates division of labor in structured collectives, stabilizes ecosystems, and shapes host health outcomes (Ackermann, 2015; Martins & Locke, 2015). Drawing on advances in single-cell genomics, population genetics, spatial ecology, machine learning, and causal inference, we examine the methodologies required to measure and model this complexity. Finally, we explore the translational implications of a heterogeneity-aware perspective for precision medicine, sustainable agriculture, biodiversity conservation, and pandemic preparedness. By weaving together these threads, this review aims to demonstrate that embracing heterogeneity is not an analytical challenge to overcome, but a conceptual necessity for advancing a predictive and impactful microbiome science.

It should be noted that heterogeneity itself is a concept—and at times, a framework or paradigm—that has permeated disciplines across the social and natural sciences and the humanities (Heckman 2001; Shavit & Ellison, 2021; Ma & Ellison, 2025; Web Link#1). Extending the logic of Vienni-Baptista et al. (2022), who argued that "the heterogeneity of understandings in interdisciplinarity and transdisciplinarity constitutes an asset," we argue that the heterogeneity *of the concept of heterogeneity itself* should similarly be seen as a strength. Thus, while a unified definition can be meaningful, we argue it is both more productive and practical for each discipline to retain its own evolving conception and trajectory of the term. Conversely, we also contend that interdisciplinary exchange on this very plurality is critical for enriching research within individual fields (Ma & Ellison, 2024, 2025). For example, in macrobial (plant and animal) ecology, the study of environmental (habitat) heterogeneity is as prevalent as, if not more prevalent than, the study of ecological heterogeneities of the organisms, assemblages, or ecosystems themselves. A similar approach—distinguishing between the heterogeneity of the microbial community and that of the host environment—is fully justified in microbiome ecology, yet this conceptual shift has not been widely adopted. To further illustrate this position, Boxes 1 and Box 2 offer comparative perspectives and suggestions for applying the concept of heterogeneity in microbiome ecology, and in the additional contexts of macrobial ecology/evolution (Box 1) and cross-disciplinary settings (Box 2), respectively. However, the remainder of this article will focus specifically on microbial heterogeneity.

## 2. The Hierarchical Architecture (Scale) of Microbial Heterogeneity

### 2.1. Molecular and Cellular Scales: Phenotypic Noise as a Regulated, Functional Trait

Within clonal populations, genetically identical cells exhibit striking phenotypic diversity in growth rate, gene expression, and stress tolerance. This **phenotypic heterogeneity** arises from stochastic biochemical reactions involving low-copy-number molecules—a phenomenon often termed "noise." Lidstrom & Konopka (2010) established the functional significance of this variation, most notably in bacterial persistence, where a small subpopulation of slow-growing cells

survives antibiotic treatment and regenerates the population—a phenomenon not explained by genetic resistance alone.

Subsequent research reframed this noise as a regulated, evolvable trait. Martins & Locke (2015) and Ackermann (2015) synthesized evidence showing that specific genetic circuit architectures have evolved to generate and control phenotypic diversity. Bistable switches can drive populations into distinct expression states, whereas excitable circuits generate transient, pulsed heterogeneity. These enable two key ecological strategies: bet-hedging, which uses probabilistic switching to spread risk in unpredictable environments, and division of labor, which allows clonal groups to perform complex, multicellular-like tasks through specialization.

External biotic agents further amplify this cellular-level heterogeneity. Liang et al. (2025) argued that bacteriophages are fundamental drivers of phenotypic diversity. Through lytic infection, phages hijack host metabolism, creating distinct "virocells"; through lysogeny, prophage integration can disrupt host gene regulation and activate new phenotypes. Because infection outcomes are heterogeneous across a clonal population, phages generate a functional mosaic within otherwise uniform groups. This **phage-driven heterogeneity** scales to influence ecosystem processes, most notably by modulating the "viral shunt," which redirects carbon flow from grazing food webs to dissolved organic matter, thereby altering global biogeochemical cycles.

### 2.2. Genomic and Strain-Level Scales: The Functional Consequences of Genetic Microdiversity

Beneath species taxonomy lies a critical layer of **strain-level heterogeneity**. Microbial species are not monolithic, but instead are collections of genetically distinct strains that vary in core genomes and in accessory gene content—the pan-genome. Garud & Pollard (2020) elucidated population genetics principles: individuals are typically oligocolonized by one or a few strains; these strains evolve *in situ* via mutation and recombination on timescales relevant to host life; and horizontal gene transfer is pervasive. Large-scale meta-analyses confirm this landscape: phylogenetic modeling of over 32,000 human gut metagenomes revealed strain-level signals related to host age, disease, lifestyle, and geography (Andreu-Sánchez et al., 2025). Longitudinal tracking shows persistent strains undergo accelerated evolution during major life transitions like weaning, nearly doubling their mutation rate as they adapt to new dietary regimes (Sawhney et al., 2025).

The functional implications changes challenge species-level descriptors. Arzamasov et al. (2025) reconstructed 68 carbohydrate utilization pathways across 3,083 *Bifidobacterium* genomes, revealing extensive strain-level variation structured by host ecology. Infant strains were enriched for human milk oligosaccharide pathways; adult strains carried unique gene clusters for digesting

local dietary polysaccharides. Notably, isolates from Bangladeshi infants possessed novel clusters for degrading xyloglucan, a plant hemicellulose common in local weaning foods. This finding illustrates that **functional heterogeneity** is a mechanism of local adaptation, enabling microbial populations to fine-tune metabolic repertoires to specific host diets. Consequently, predicting an isolate's ecological role requires resolution down to the strain or genotype level. This strain-level heterogeneity extends to other body sites. Kang et al. (2021) showed that *Staphylococcus epidermidis* and *Streptococcus pyogenes* exhibit phylogenetically distinct strains across individuals, with competitive interactions structuring community assembly on the ocular template.

### 2.3. Interpersonal and Population Scales: Challenging the "Healthy Microbiome" Paradigm

Perhaps the most conspicuous manifestation of heterogeneity is the dramatic variation in microbiome composition between individual human hosts. However, composition alone offers an incomplete picture—true heterogeneity must also consider interactions between entities (species or strains), which shape community dynamics and functional outcomes (Ma & Ellison 2026). Shanahan et al. (2023) argued that the perceived extent of human microbiome variance is systematically underestimated due to sampling bias: over 70% of published data originate from Westernized, industrialized populations in Europe and North America, representing only a fraction of human genetic, cultural, and dietary diversity. Ghosh & O'Toole (2023) further documented this bias, showing that ageing, migration, and ethnicity shape microbiomes in ways Western-centric research misses. Thus, current literature captures only a partial view of heterogeneity, overlooking the interaction-based components that are essential for understanding ecological and functional specialization.

A primary axis is the global industrialization gradient. Non-industrialized or ancestral microbiomes—characterized by higher alpha diversity, greater *Prevotella* abundance, and the presence of taxa like *Treponema*—differ starkly from industrialized Western microbiomes dominated by *Bacteroides*. Migration studies demonstrate that these patterns are malleable; upon moving to industrialized settings, microbiomes "westernize," a transition correlated with increased risk for metabolic and immune-mediated diseases.

This population-level heterogeneity undermines the concept of a singular "healthy microbiome." Joos et al. (2025) argued that defining health *a priori* using small, homogeneous control groups is inherently flawed, excluding healthy individuals who deviate from narrow norms. Instead, they advocated for large, diverse cohorts where definitions of microbiome health emerge from data as statistical constructs accounting for the spectrum of healthy configurations across ages, geographies, and lifestyles.

This data-driven approach is exemplified by Gupta et al. (2020), who developed the Gut Microbiome Health Index from over 4,000 metagenomes across 34 studies. Using a prevalence-based strategy—identifying species consistently more or less frequent in healthy individuals—rather than comparing average abundances, they produced a 50-species signature that robustly distinguished healthy from non-healthy states in independent cohorts. Network-based approaches further reveal that clinically meaningful signals reside in co-occurring modules rather than single taxa. Chin Fatt et al. (2023) identified a butyrate-producing module in depression whose collective abundance correlated with anxiety severity—a signal undetectable by single-taxon analysis. Ma and Li (2017) quantified vaginal community state types using species specificity indices, and Ma (2023) extended this to a four-type framework, namely complex versus simple, and healthy versus BV (bacterial vaginosis), resolving paradoxes of high-diversity communities in healthy women and low-diversity in BV patients. From a heterogeneity perspective, these findings suggest that diversity alone is insufficient to define community state; rather, the nature of species interactions and trait specialization matters (Ma & Ellison 2026). High diversity in healthy women may reflect cooperative interactions and functional redundancy, whereas low diversity in BV may indicate competitive or disruptive interactions among pathobionts. Thus, heterogeneity, defined not merely by compositional variation but by interaction-based ecological structure, provides a more nuanced understanding of vaginal ecosystem states.

### 2.4. Community and Ecosystem Scales: Assembly, Rare Biospheres, and Spatial Mosaics

At the community level, heterogeneity governs assembly, stability, and function. Huber et al. (2020) provided a nuanced empirical resolution in a floodplain river system, where hydrological cycles create a gradient from environmental heterogeneity to homogeneity. They found that heterogeneous selection prevailed under variable conditions, driving high beta-diversity, while homogeneous selection dominated during floods, reducing beta-diversity. Stochastic processes peaked at intermediate heterogeneity. This study shows that environmental heterogeneity directly dictates the operative ecological rules, with cascading effects on biodiversity. These principles extend to host systems: along the human intestinal tract, mucosal communities exhibit significant spatial heterogeneity structured by competitive interactions, a pattern masked by inter-individual variation in standard analyses (Zhang et al., 2014).

The rare biosphere—the long tail of low-abundance taxa—is integral to community heterogeneity. Chakraborty et al. (2025) introduced “core rare taxa,” defined by low abundance but high occupancy, hypothesizing they constitute a functional “seed bank” critical for ecosystem resilience. This functional reserve is not static: Shade and Gilbert (2015) showed that “conditionally rare taxa”—organisms usually at low abundance that occasionally bloom—account for up to 97% of

temporal variability in community structure, revealing rarity as functional responsiveness, not passivity.

**Spatial heterogeneity** adds a critical dimension frequently overlooked in microbial research. In soil, Nunan et al. (2020) argued that physical pore structure and resource chemistry favor a "waiting game" strategy—cells relatively immobile, investing in biofilms, relying on diffusing resources—contributing to soil organic carbon persistence. In biofilms, Dal Co et al. (2020) demonstrated that metabolic interactions are extremely short-ranged (micrometers); a cell's growth rate depends solely on its immediate neighbors, making spatial structure a primary generator of functional heterogeneity.

Experimental manipulation provides causal evidence for the role of species interactions in shaping heterogeneity. Chen et al. (2025) isolated individual clusters from anaerobic digesters, showing thermophilic conditions reduced spatial heterogeneity by eliminating low-abundant species. This intensified interspecies competition while simultaneously enabling efficient syntrophic consortia around *Clostridiales*, ultimately enhancing methanogenic performance—a cascading effect from temperature to ecosystem function lost in bulk samples.

Spatial omics technologies now enable ecosystem-level analysis. In oncology, spatial transcriptomics applies landscape ecology principles to map functional niches and infer cell-cell communication networks (Liu et al., 2026). These principles are increasingly being adapted for microbiome studies. Cross-kingdom comparisons reveal conserved assembly principles. Garrido-Sanz & Keel (2025) showed that wheat rhizosphere assembly is deterministically driven by seed-borne bacteria exerting priority effects and cross-feeding. This mirrors human infant gut succession, where weaning acts as a selective bottleneck accelerating strain evolution (Sawhney et al., 2025). In both systems, heterogeneity in founding inoculum and host-mediated selection dictates long-term community structure.

## 3. Functional and Ecological Consequences: The "Why" of Heterogeneity

Understanding the mechanisms that generate heterogeneity is essential, but its true significance lies in its functional and ecological consequences. We delineate five major categories of impact, now supported by advanced theoretical and modeling frameworks:

### 3.1. Risk Management and Population Resilience via Bet-Hedging

In fluctuating or unpredictable environments, phenotypic heterogeneity acts as a form of evolutionary bet-hedging. By stochastically generating a diversity of phenotypes in advance of an

environmental change, a clonal population ensures that some individuals will survive a sudden stress. The canonical example is antibiotic persistence. Metabolic bet-hedging is equally important; a subpopulation pre-expressing catabolic pathways for an alternate nutrient source can ensure growth when the primary source is depleted. This strategy enhances population-level fitness and stability and is a key reason why heterogeneity is subject to positive selection in variable environments (Ackermann, 2015). However, direct evidence for bet-hedging in natural microbial communities remains limited, and alternative explanations—such as non-adaptive stochasticity—are often difficult to exclude. At the community level, the "core rare biosphere" (Chakraborty et al., 2025) functions as a hypothesized reservoir of functional diversity, providing genetic and metabolic redundancy that buffers the ecosystem against perturbations. This aligns with and is formalized by ecological frameworks such as the diversity-disease relationship (Civitello et al., 2015; Keesing et al., 2010; Ma et al. 2019) and dominance-stability relationship (Ma & Ellison, 2019), in which greater biodiversity (a form of heterogeneity) can inhibit pathogen spread and stabilize ecosystem functions. The core rare taxa may act as stress tolerators within established life-history frameworks, possessing resilience from both innate genetic diversity and adaptive plasticity gained through selection or horizontal gene transfer.

### 3.2. Enabling Division of Labor and Social Complexity

In spatially structured environments like biofilms, host tissues, or microbial aggregates, phenotypic heterogeneity enables functional specialization and division of labor. Subpopulations can perform complementary tasks that are incompatible or inefficient within a single cell. Ackermann (2015) reviewed classic examples: in *Salmonella* infections, one subpopulation invades host intestinal cells (a costly action that induces inflammation), while another subpopulation proliferates in the altered luminal environment. In some cyanobacteria, the oxygen-sensitive process of nitrogen fixation is segregated into specialized cells (heterocysts) that are distinct from those performing oxygenic photosynthesis. This division of labor that leads to the emergent property of complex, multicellular-like behaviors in certain groups is an exemplar of biotic heterogeneity. Theoretical models using percolation theory from statistical physics further illustrate the systems-level consequences of such interdependencies. Clegg and Gross (2025) modeled cross-feeding networks as bipartite hypergraphs and demonstrated that they possess inherent tipping points and exhibit hysteresis. This result explained phenomena such as the "great plate-count anomaly," which refers to the long-observed discrepancy between the high number of microbial cells visible under a microscope in an environmental sample and the much lower number that will grow into colonies on a laboratory petri dish—often less than 1% of the total cells. The uncultivability of many microbes is not a technical failure but rather a consequence of disrupting the fragile web of metabolic interdependencies that sustains diversity in natural environments.

### 3.3. Driving Ecosystem Stability and Biogeochemical Function

Heterogeneity at the community level is a cornerstone of ecosystem stability. High functional redundancy, in which multiple taxa can perform a given ecosystem process, buffers the system against species loss. Tian et al. (2020) provided a mechanistic framework for understanding functional redundancy through the Genomic Content Network (GCN), a bipartite graph linking microbes to genes in their genomes. Analysis of the human microbiome GCN revealed three features favoring high functional redundancy: nested structure, unimodal functional distance distribution, and heterogeneous gene degree distribution. Crucially, Tian et al. (2020) showed that high functional redundancy of the recipient's pre-FMT (fecal microbiota transplantation) microbiota predicted poor donor strain engraftment, demonstrating that functional redundancy acted as a barrier to invasion—a measure of template resilience with direct clinical implications. The "core rare biosphere" may be a critical reservoir of this redundancy (Chakraborty et al., 2025). The relationship between functional redundancy and ecosystem stability is well supported theoretically, but empirical tests in complex natural communities have yielded mixed results. Furthermore, heterogeneity in traits like substrate affinity, growth rate, and stress tolerance allows for more complete and efficient exploitation of diverse resource pools, enhancing overall ecosystem productivity. The stability of these heterogeneous systems can be diagnosed and predicted through complex ecological network analysis (e.g., Ma & Ellison, 2019). Analyzing properties like core/periphery structures, high-salience skeletons (the backbone of strongest interactions), and the ratio of positive-to-negative interactions (P/N ratios) in microbial networks can effectively diagnose dysbiotic states and network fragility. These analytical approaches move beyond simple diversity metrics to identify keystone species and potential intervention targets within complex communities. Perhaps the most dramatic ecosystem-scale impact is via the viral shunt. Liang et al. (2025) synthesized how phage-induced lysis of heterogeneous bacterial subpopulations shunts carbon and nutrients away from the classic grazing food chain into dissolved organic matter, fundamentally altering microbial loop efficiency and global ocean carbon cycling. Thus, cellular-level heterogeneity directly scales to planetary biogeochemistry.

### 3.4. Influencing Disease Emergence and Pathogen Spread

Heterogeneity in animal reservoir populations is a critical factor in zoonotic spillover. Pekar et al. (2025) employed recombination-aware phylogenetics to reconstruct the origins of SARS-CoV-1 and SARS-CoV-2. They found that the immediate bat virus ancestors of these viruses circulated remarkably recently, within years of emergence, and in specific regions far from the sites of human outbreaks. The dispersal velocity required for these progenitors to reach the outbreak sites far exceeded natural bat movement rates, strongly implicating human-mediated wildlife trade as the accelerant. This study illustrated how landscape heterogeneity in viral distribution interacted with

heterogeneous human activity networks to bridge ecological gaps and facilitate pandemic emergence. Notably, however, the study did not provide a realistic quantitative measure of heterogeneity, limiting its ability to rigorously assess the role of environmental heterogeneity in spillover dynamics.

The role of biodiversity in reducing zoonotic spillover risk (the "dilution effect") has been extensively reviewed elsewhere (Wilkinson et al., 2018). These models quantitatively examine how host and microbial biodiversity and heterogeneity modulate disease risk through dilution effects (where high diversity reduces pathogen transmission) or amplification effects, providing a predictive lens for epidemiology (Keesing et al., 2010). In human microbiome-associated diseases, the relationship between microbial diversity and host health are routinely investigated, especially in the early days of the Human Microbiome Project (HMP), perhaps because diversity is a de facto standard metric to compute in those studies. Nevertheless, it was found that in only approximately 1/3 of the studies cases, the diversity-disease relationship (DDR) was significant (Ma et al. 2019). The heterogeneity-disease relationship was even lower (Ma 2020). Nevertheless, it should be noted that the DDR/DHR in microbiome-associated diseases should have totally different mechanisms from those studied in zoonosis (amplification/dilution effects), given that the transmission/transfer media in both fields are totally different (e.g., air vs. metabolic pathways).

### 3.5. Revealing Convergent Solutions Across Disparate Systems

Despite immense heterogeneity in host biology and environments, deep functional convergence exists in how hosts manage their microbiomes. Ilyaskina et al. (2025) proposed the elegant "inside-out" analogy, comparing the human gut to an internalized plant rhizosphere. Both are interfaces in which the host secretes a complex cocktail of chemicals (mucus/bile acids vs. root exudates) to selectively recruit and shape a microbial community. Both systems use conserved metabolite-based dialogues for communication; for example, microbial metabolism of tryptophan produces indole derivatives that modulate host immune responses in the gut (e.g., aryl hydrocarbon receptor ligands) and act as phytohormones (*e.g.*, indole-3-acetic acid) in the rhizosphere. This convergence suggests that differences in host form create parallel selective pressures, leading to the evolution of analogous microbial solutions for establishing and maintaining beneficial symbioses. This insight suggests a transformational strategy: engineering plant microbiomes (e.g., using synthetic communities) can inform approaches to human microbiome restoration, and vice versa.

## 4. An Evolving Conceptual Framework: From Descriptive Noise to Predictive, Mechanism-Driven Ecology

The conceptualization of heterogeneity in microbial systems from observation to mechanism, and from ecology to translational reflects a maturation in biological thinking: from treating variation as a nuisance to recognizing it as a fundamental, exploitable property of living systems. Four phases mark this conceptual maturation. The initial descriptive phase established that heterogeneity is ubiquitous—from bacterial persistence to strain variation—and cannot be dismissed as experimental noise (Lidstrom & Konopka, 2010). A subsequent functional and evolutionary phase reframed heterogeneity as an adaptive strategy, showing that stochastic phenotypic variation enables bet-hedging and division of labor, traits shaped by natural selection rather than biochemical imperfection (Ackermann, 2015; Martins & Locke, 2015; Amir & Balaban, 2018). The spatial and eco-evolutionary phase grounded this heterogeneity in physical reality, demonstrating that micron-scale spatial structure dictates metabolic interactions and that ecological and evolutionary processes operate on commensurate, rapid timescales (Dal Co et al., 2020; Castledine et al., 2020). We are now in a fourth, predictive and mechanism-driven phase, characterized by unifying frameworks in medical ecology, information theory, and network science that integrate across scales and disciplines to forecast dynamics, explain health outcomes, and design interventions (Sherwin et al. 2017; Ma & Zhang, 2022; Varley, 2025; Clegg & Gross, 2025; Caminero et al., 2025). Central to this phase is the dual-thread paradigm articulated by Ma & Ellison (2026), which distinguishes **environmental heterogeneity** (the abiotic template) from **ecological heterogeneity** (the biotic processes expressed on that template)—a distinction that anchors this review and unifies human and environmental microbiome research.

### 4.1. The Descriptive and Phenomenological Phase

The initial phase, exemplified by Lidstrom and Konopka (2010), was descriptive and phenomenological. It established the ubiquity of physiological heterogeneity—such as bacterial persistence—and its clear functional consequences. This work was crucial for shifting the perception of non-genetic variation from mere "noise" to an observable biological phenomenon with survival implications. However, it primarily asked *what* exists and *how* it manifests, not yet *why*.

### 4.2. The Functional and Evolutionary Phase

The second phase, synthesized by Martins and Locke (2015) and Ackermann (2015), was functional and evolutionary. It posed the critical question: *why* does heterogeneity exist? This phase repositioned heterogeneity as an evolvable, adaptive strategy shaped by natural selection. The central explanatory concepts became bet-hedging and division of labor that together enable complex, multicellular-like tasks in clonal groups. This represented a profound intellectual shift: variability and variation diversity were no longer seen as costs or biochemical imperfections, but

as beneficial traits, a product of evolved genetic circuit architectures that generate heterogeneity. This phase laid the groundwork for understanding heterogeneity as a life-history strategy. Amir and Balaban (2018) contributed to this phase by demonstrating that stochastic fluctuations at the single-cell level contain mechanistic information, reinforcing that variance itself is signal, not noise, and providing tools for extracting biological insight from observed heterogeneity.

### 4.3. The Spatial and Eco-Evolutionary Integration Phase

The third phase integrated spatial and eco-evolutionary dimensions, grounding heterogeneity in physical and dynamic reality. Experimental work like that of Dal Co et al. (2020) quantified the micrometer scale of metabolic interactions, demonstrating that a cell's fate is determined by its immediate neighbors, making spatial structure a primary generator of functional heterogeneity. Concurrently, the eco-evolutionary perspective matured, emphasizing that for microbes, ecological and evolutionary processes operate on similar, rapid timescales and interact dynamically. Castledine et al. (2020) applied this perspective to "community coalescence"—the mixing of whole communities. They theorized that the outcome depends not just on taxonomic composition but on the dominant interaction types (competitive, trophic, mutualistic) and, crucially, their coevolutionary histories. This framework began to treat communities as units with functional and evolutionary integrity, moving beyond a simple mixing bowl model.

### 4.4. The Current Phase: Predictive, Mechanism-Driven Integration

We are now in the fourth and current phase, which focuses on predictive, mechanism-driven modeling through the integration of population genetics, epidemiology, and, most importantly, unifying theoretical frameworks that bridge fundamental ecology with clinical and translational science. This phase is characterized by several convergent strands:

#### 4.4.1. The Rise of Medical Ecology as a Unifying Paradigm

This phase is being consolidated under frameworks like Medical Ecology (Ma & Zhang, 2022), which provides the essential cross-disciplinary lens to unify the study of heterogeneity. Medical Ecology explicitly integrates ecological theory, computational biology, and clinical medicine to understand how multi-scale heterogeneity in host-microbe-environment interactions drives health and disease outcomes. It actively incorporates powerful evolutionary concepts: (i) Niche Construction explains the reciprocal feedback where hosts actively shape their microbial environment through immune function, diet, and behavior (e.g., root exudates, bile acids, mucosa). This altered environment, in turn, imposes new selection pressures on the microbial community and on future host generations—a process termed ecological inheritance (Week et al., 2025). (ii)

Multilevel Selection Theory provides the crucial analytical framework for examining potential conflicts. Selection can act on microbial strains within a host (favoring fast growth, resource exploitation) versus on the host-microbiome assemblage, or holobiont, as a functional unit (favoring stability, host health). Understanding this tension is key for designing sustainable interventions. (iii) Theories of Inheritance are expanded to include the microbiome as a potent form of non-genetic inheritance, whose evolutionary impact depends on its vertical and horizontal transmission fidelity between hosts (Week et al., 2025). (iv) Ecological Filters and Trait-Based Approaches, as articulated by Violle et al. (2012), provide a framework for understanding how environmental heterogeneity selects for microbial traits. Their T-statistics partition variance across hierarchical levels—individual, population, community, regional pool—demonstrating that variance itself contains the signature of ecological filters operating on templates. (v) Movement and Dispersal Across Template Landscapes, as developed by Wisnoski and Lennon (2023), apply movement ecology to microorganisms, arguing that microbial movement—whether active swimming, passive dispersal, or transport by hosts—must be understood across scales, from individual cells to populations to communities. The template provides the landscape across which movement occurs, and heterogeneity in template structure fundamentally shapes movement patterns and their ecological consequences.

### 4.4.2. Information Theory and Complex Systems Science

The analysis of heterogeneous systems requires advanced mathematical frameworks. **Information theory** (Eliazar 2021; Varley, 2025) is increasingly applied to dissect complex dependencies in microbiome data. Tools like Partial Information Decomposition can disentangle redundant, unique, and synergistic information flows among species, moving beyond pairwise correlations to capture the higher-order interactions that likely govern community stability and function.

### 4.4.3. From Topology to Mechanism in Network Science

The field is moving from inferring co-occurrence networks to modeling **mechanistic interaction networks**. Theoretical work like that of Clegg and Gross (2025), which applies percolation theory to cross-feeding hypergraphs, demonstrates how network structure creates tipping points and hysteresis, explaining systemic behaviors like the "great plate count anomaly." This shift from descriptive network topology to predictive, dynamics-aware models is fundamental. Ma & Ellison (2019) further advanced network analysis by developing frameworks for identifying critical network structures—core/periphery organizations, high-salience skeletons, and positive-to-negative interaction ratios—that can diagnose dysbiotic states and network fragility in microbial communities.

#### 4.4.4. Spatial Explicitness and Clinical Integration

The conceptual framework now mandates spatial explicitness. Recognizing that heterogeneity has a physical structure, technologies like spatial transcriptomics are being used to apply landscape ecology principles to clinical settings. For example, in oncology, spatial omics enables analysis of the tumor microenvironment as an ecosystem, mapping functional niches and inferring communication networks (Liu et al., 2026). This represents the full conceptual circle: from recognizing microscopic spatial heterogeneity in biofilms to analyzing spatial ecosystem pathology in human disease.

#### 4.4.5. Closed-Loop, AI-Informed Translation

The final strand of the current phase is the conceptualization of a closed-loop translation framework (Caminero et al., 2025). Here, AI analyzes high-dimensional, heterogeneous clinical and multi-omic data to generate mechanistic hypotheses. These hypotheses are tested in reductionist experimental models (gnotobiotic mice, synthetic communities, organ-on-a-chip), and the results feedback to refine the computational models and clinical applications. This framework explicitly links pattern discovery to causal mechanism and therapeutic design.

#### 4.4.6. The Forecasting Potential of Microbiomes

If microbial communities are displayed on templates, then changes in templates should predict changes in communities. Correa-Garcia et al. (2022) articulated this "forecasting power" of the microbiome, arguing that because microbiomes integrate past and present environmental conditions and are directly involved in ecosystem processes, they can be used to predict future states. This framework has been applied to forecast crop yields from soil microbiomes, disease susceptibility from gut microbiomes, and ecosystem responses to perturbation—extending the predictive ambitions of the field.

In summary, the conceptual framework for microbial heterogeneity has evolved from description to functional explanation, and now to predictive, mechanism-driven integration. The goal is no longer merely to catalog variation but to build a predictive science that can forecast microbiome dynamics, explain health outcomes, and rationally design interventions. This is achieved by viewing heterogeneity through the integrated lenses of evolutionary ecology, information theory, spatial analysis, and translational medicine—a synthesis embodied by the growing framework of Medical Ecology. This maturing framework posits that understanding and harnessing heterogeneity is the central challenge and opportunity for achieving precision in microbiome

science. Nevertheless, significant challenges remain in implementing this agenda and vision at the current stage.

## 5. Methodological Synthesis: Navigating and Modeling Multi-Dimensional Data

### 5.1 Multi-dimensional metagenomics framework

The complexity of heterogeneity demands equally sophisticated methodological approaches. Peng et al. (2025) proposed a comprehensive **multi-dimensional metagenomics framework**, arguing that integration across four dimensions is essential for mechanistic insight:

1: **Composition & Abundance:** The foundational layer, but limited by **compositionality** — the inherent constraint that relative abundance data are interdependent (i.e., an increase in one taxon necessarily decreases others), which can obscure true biological signals and lead to spurious correlations — and lack of resolution. Tools such as quantitative microbiome profiling and careful compositional data analysis are crucial starting points.

2: **Genetic Variation**: This dimension requires strain-resolved metagenomics. Tools like inStrain and StrainPhlAn track strains across samples, while pangenome analysis reveals accessory gene content. High-quality metagenome-assembled genomes (MAGs) are the currency of this dimension, enabling the genotype-to-phenotype mapping demonstrated by Arzamasov et al. (2025).

3: **Protein Structure & Function**: A vast amount of microbial genomic "dark matter" (like “junk” DNA) encodes proteins of unknown function. AI-driven structure prediction tools like AlphaFold and ESMFold can infer 3D structures from metagenomic sequences, and tools like Foldseek can use structural homology to suggest functional annotations, using genetic inventories to suggest mechanistic hypothesis.

4: **Spatiotemporal Dynamics**: Capturing heterogeneity across space and time is challenging. This requires longitudinal sampling and spatial profiling techniques such as multiplexed imaging (e.g., CLASI-FISH), spatial transcriptomics, and mass spectrometry imaging. These technologies are transforming fields like oncology, where spatial ‘omics enables true ecosystem-level analysis of tumor microenvironments, applying landscape ecology principles to identify functional niches and infer cell-cell communication networks using spatial graph theory and optimal transport models (Liu et al., 2026). Technologies like Microbiome single-cell Transcriptomics (MscT), developed by Jia et al. (2024), represent a complementary breakthrough, allowing profiling of active gene

expression in individual microbial cells within complex communities, revealing functional clusters that cross taxonomic boundaries.

### 5.2 Quantifying Heterogeneity: Metrics, Models and Frameworks

To navigate these multi-dimensional data, robust analytical frameworks are required. Information theory (van der Lubbe 1986; Varley, 2025) provides essential mathematical tools for analyzing complex systems like microbiomes. By quantifying uncertainty and dependencies, methods like Partial Information Decomposition allow researchers to distinguish between redundant, unique, and synergistic information components in multivariate datasets. This addresses significant limitations of traditional pairwise analyses that often fail to capture higher-order interactions characteristic of complex microbial communities. Information-theoretic measures such as transfer entropy and active information storage enable sophisticated analysis of temporal processes, whereas network inference methods based on information theory facilitate the construction of functional and effective connectivity networks from complex microbiome data streams.

Medical ecology analytics (e.g., Ma, 2019, 2020, Ma & Zhang 2022) also provide a suite of models and metrics specifically designed to measure key ecological properties of microbiomes, including diversity, heterogeneity, specificity, stochasticity, and stability. For example, the Species Specificity and Specificity Diversity (SSD) framework (Ma, 2024a, 2024b) can identify treatment- or disease-specific unique species (US) or enriched species (ES) that may serve as diagnostic biomarkers or as candidates for probiotic development. Furthermore, Taylor's Power Law has been extended from quantifying population aggregation to community heterogeneity and ecological networks to provide a general methodology for heterogeneity scaling analysis across different levels of organization (Ma, 2015, 2025; Ma & Taylor, 2025). These extensions reveal that power-law distributions of heterogeneity are common across systems, with the Gaussian distribution being the exception rather than the rule—a pattern consistent with the broader principles on heterogeneity articulated by Ma & Ellison (2026).

Network analysis has emerged as a particularly powerful toolkit for investigating microbial heterogeneity. Hu et al. (2022) introduced the Microbial Interaction Network Database (MIND), a platform for integrating heterogeneous network types, including co-occurrence networks, metabolic influence networks, horizontal gene transfer networks. By allowing comparisons across datasets and taxonomic levels, MIND enables researchers to identify robust interaction patterns that transcend study-specific variation. Chin Fatt et al. (2023) applied weighted correlation network analysis (WGCNA) to depression and demonstrated that the community, not a single taxon, is the functional unit.

### 5.3 Artificial Intelligence and Machine Learning in Microbiome Science

AI and ML are indispensable for finding patterns and making predictions in high-dimensional, heterogeneous datasets. Asnicar et al. (2024) provided a practical primer, emphasizing that model choice must match the question (supervised for prediction, unsupervised for discovery) and that rigorous validation, particularly Leave-One-Dataset-Out (LODO) validation, is best for estimating real-world performance across heterogeneous studies because it helps overcome pervasive pitfalls like data leakage and confounding batch effects. Specialized ML applications are tackling previously intractable ecological questions. The Data-driven Keystone Identification (DKI) framework of Wang et al. (2024) uses composition-aware neural ordinary differential equations to learn community assembly rules directly from observational abundance data, enabling *in silico* experiments of species removal and revealing that traditional network centrality metrics show no correlation with true keystone status. Similarly, Graph Neural Networks (GNNs) can learn inter-species interaction graphs from longitudinal abundance data and accurately predict microbial community dynamics months in advance, with the counterintuitive finding that clustering taxa by inferred interaction strength optimizes prediction accuracy more than taxonomy or function (Andersen et al., 2025). AI approaches can be further enhanced by integrating established ecological analytical frameworks, such as dominance-stability and heterogeneity-stability relationships (Ma & Ellison, 2019), to ensure models optimize for biologically meaningful stability properties. Nevertheless, a persistent limitation is the black-box nature of deep learning models, which complicates biological interpretation and limits clinical translation.

### 5.4 From Correlation to Causation—Closing the Inference Gap

A critical methodological frontier is moving from observing correlations to establishing causal relationships within microbial communities. Traditional co-occurrence network analysis, while computationally scalable, has proven inadequate for causal inference, often mistaking shared environmental preferences or technical artifacts for genuine biological interactions (Oña et al., 2025). This has catalyzed a shift toward hybrid experimental-computational frameworks. Experimental approaches now employ sophisticated full factorial *n*-wise designs and drop-out/knockout experiments to systematically identify keystone species and higher-order interactions, though they face cultivability and combinatorial challenges. Predictive constraint-based approaches like flux balance analysis predict metabolic interactions from genome-scale models but overlook non-metabolic interactions. The emerging consensus advocates for iterative validation cycles where computational predictions guide targeted experiments, and experimental results refine computational models. This framework is essential for translational development, moving from pattern recognition to the identification of actionable therapeutic targets and mechanistic pathways.

A particularly promising development is the QuanT (Quantile Thresholding) method introduced by Lu et al. (2026). This non-parametric approach identifies unmeasured technical heterogeneity in sparse, overdispersed microbiome data by applying quantile regression across multiple quantiles and using singular value decomposition on thresholded residuals. When applied to integrated HIV and colorectal cancer datasets, QuanT improved false discovery rate control and recovered true biological signals better than methods developed for genomic data, demonstrating that accounting for unmeasured heterogeneity is essential for reliable inference.

Finally, metabolic modeling and experimental tracing are critical for linking genomic potential to ecosystem function. Noecker and Turnbaugh (2024) reviewed emerging toolkits, from constructing genome-scale metabolic models (GEMs) for individual organisms (COBRA) to community-level models (MICOM), and coupling these with experimental validation via high-throughput phenotyping (BacterAI) and stable isotope probing to trace nutrient flows. These integrated approaches may help researchers move on from correlative analysis to mechanistic, causal understanding of how heterogeneous communities function.

### 5.5 The Need for Standards

The iHMP Consortium (2019) demonstrated best practices through multi-omic integration across three longitudinal cohorts. Their approach revealed coordinated molecular changes during disease flares and showed that baseline individual variation often exceeded disease-associated signals—underscoring the necessity of within-subject longitudinal designs and standardized methods. The power of these methods depends on their transparent application. Mirzayi et al. (2021) developed the STORMS (Strengthening the Organization and Reporting of Microbiome Studies) checklist, a 17-item guideline spanning study design, laboratory methods, bioinformatics, and statistics. STORMS provides a common grammar for reporting that enables comparison and synthesis across studies. The STREAMS guidelines (STREAMS Working Group, 2025) extend this with a detailed 67-item checklist enforcing FAIR data principles and precise computational reporting.

These standards offer opportunities to reduce unwanted technical heterogeneity. The consequences of ignoring them are stark. Pichon et al. (2025) sent an identical stool sample to five French direct-to-consumer testing laboratories and received wildly divergent results: alpha diversity estimates ranged from 3.64 to 6.11, clinical interpretations varied from “dysbiotic” to “eubiotic,” and health recommendations were contradictory and often unsubstantiated. This empirical demonstration underscores the urgent need for methodological standardization. Strictly speaking, however, what Pichon et al. measured was microbial diversity, and the heterogeneity they documented reflects variability in sequencing platforms and bioinformatics pipelines, rather than true microbial heterogeneity, which should consider species interactions in the first place.

## 6. Translational Implications, Ethics, and a Roadmap for Future Research

Embracing a heterogeneity-aware perspective has profound implications for translating microbiome science into real-world applications across health, agriculture, and environmental stewardship.

### 6.1. Precision Medicine and Nutrition

The era of one-size-fits-all probiotics is ending. Strain-level functional heterogeneity enables rational design of precision synbiotics, matching probiotic strains to prebiotics based on genomic signatures of glycan metabolism tailored to host age and diet (Arzamasov et al., 2025). Large-scale nutritional epidemiology links specific microbes to cardiometabolic health (Asnicar et al., 2025), providing evidence for personalized dietary recommendations. Future trials must stratify participants by baseline microbiome composition to identify responders. Host genetics further refines this approach; replicable loci such as *LCT* and *ABO* modulate microbial effects depending on diet (Ferretti et al., 2025).

Network-based signatures offer clinical biomarkers. A butyrate-producing module linked to anxiety severity in depression (Chin Fatt et al., 2023) and vaginal community state typing (Ma, 2023) exemplify how community-level patterns can guide intervention. Airway microbiome profiling similarly distinguishes asthma phenotypes (Huang, 2025).

### 6.2. Precision Microbial Therapeutics: Ecological Design Principles

Therapeutic design increasingly follows ecological principles. Successful strain displacement requires a two-stage process: invasion via private nutrients, followed by displacement through interference competition (Bakkeren et al., 2025). Community types emerge from heterogeneity in interspecies interaction strengths; manipulating "strongly interacting species" can steer ecosystem states (Gibson et al., 2016).

Data-driven keystone identification (Wang et al., 2024) enables context-specific targeting. Therapeutic success depends on community stability, formalized in diversity-disease and heterogeneity-disease frameworks (Ma et al., 2019; Ma, 2020). Functional redundancy acts as a barrier to invasion: high pre-FMT (fecal microbiota transplantation) redundancy predicts poor donor engraftment (Tian et al., 2020), turning a community property into a predictive biomarker.

### 6.3. Microbiome Epidemiology and Public Health Diagnostics

The Gut Microbiome Health Index (Gupta et al., 2020) demonstrates how a quantifiable health metric can emerge from heterogeneous meta-data. Robust microbiome-disease associations require large, diverse cohorts with analysis at multiple biological resolutions—strains, genes, metabolites (Joos et al., 2025; VanEvery et al., 2023).

### 6.4. Clinical Translation Frameworks: Standards, Validation, and Evolution

The STREAMS guidelines (STREAMS Working Group, 2025) enforce FAIR data principles and precise computational reporting. Regulatory frameworks must adapt to personalized live biotherapeutics, addressing trial design for consortia and the need for explainable AI (Caminero et al., 2025). Therapeutic durability requires evolutionary forethought; interventions must consider long-term trajectories and transmission fidelity (Week et al., 2025).

### 6.5. Sustainable Agriculture, Climate Mitigation, and Pandemic Preparedness

Rhizosphere assembly principles enable next-generation biofertilizers—seed coatings with facilitative consortia that enhance crop resilience (Garrido-Sanz & Keel, 2025). Soil spatial heterogeneity is key to managing organic carbon for climate mitigation (Nunan et al., 2020). The "core rare biosphere" concept argues for conserving microbial diversity, including low-abundance taxa, as a component of ecosystem health (Chakraborty et al., 2025).

Pandemic risk emerges at the interface of viral heterogeneity and human activity networks. Effective surveillance must monitor viral diversity in wildlife reservoirs *and* the trade networks that transport pathogens globally (Pekar et al., 2025).

### 6.6. The Unseen Dimensions: Bias, Representation, and Ethics

Any framework for translation must confront biases in how templates are selected and studied. For examples, which templates are studied? How are templates framed and who sets the research agenda? as briefly interpreted below:

Over 70% of human microbiome samples originate from Europe and North America—regions representing only ~15% of the global population (Ghosh & O'Toole, 2023). This sampling bias distorts our understanding of global microbial variation. The "industrialized" microbiome is not a universal baseline but a particular configuration shaped by specific environmental and lifestyle factors.

Núñez Casal (2024) shows that non-Western and indigenous populations are often treated as microbial "reservoirs"—their microbiomes bioprospected as ancestral baselines, then

commodified in personalized medicine platforms serving Western consumers. Template selection is not neutral; who is studied and how they are framed carries ethical weight.

Greenhough et al. (2020) ask foundational questions: who benefits from microbiome research? Who decides which questions matter? How can research be conducted equitably? These questions remind us that the templates we study are embedded in histories, cultures, and power relations.

### 6.7. Future Research Agenda

Advancing the field requires progress along interconnected pathways:

(i) Causal Integration: Tightly coupling observational studies with experimental models—gnotobiotic animals, synthetic communities—in iterative cycles to move from association to mechanism.

(ii) Global Representation: Building cohorts that represent the full spectrum of human genetic, cultural, and lifestyle diversity to overcome sampling bias (Shanahan et al., 2023; Ghosh & O'Toole, 2023).

(iii) Eco-Evolutionary Forecasting: Developing dynamical models that incorporate rapid microbial evolution to predict responses to antibiotics, dietary shifts, and climate change.

(iv) Theoretical Unification: Expanding integrative frameworks like Medical Ecology that unify concepts of heterogeneity across scales, building on the dual-thread paradigm (Ma & Ellison, 2026).

(v) Integrated Diagnostic-Therapeutic Platforms: Developing closed-loop systems where AI-driven discovery, ecological design of interventions, and rigorous validation are seamlessly integrated.

(vi) Cross-disciplinary, conceptual and methodological developments (*see* the next section).

## 7. Conclusion and Challenges

Heterogeneity is the heartbeat of microbial systems. On ecological timescales, it is a multi-scale, emergent property generated by ecological interactions—particularly among species—embedded within a physical environment that is itself heterogeneous, shaped largely by habitat connectedness and mosaics. On evolutionary timescales, heterogeneity is a dynamic and adaptive phenomenon driven by the interplay of stochasticity and deterministic selection, continuously reshaped by rapid evolution and coevolutionary dynamics between hosts and microbes, and conferring critical

functional redundancies across all levels of biological organization. From enabling bacterial persistence to stabilizing global carbon cycles, from shaping personalized health outcomes and preventing dysbiosis to influencing pandemic emergence, heterogeneity is not a peripheral detail—it is the central plot.

The framework advocated in this review—grounded in the distinction between environmental heterogeneity (the template) and ecological heterogeneity (the biotic variability emerged from interactions expressed on that template) (Ma & Ellison, 2026)—provides a unified lens for understanding microbial heterogeneity across systems. Whether the template is a human gut, an ocular surface, a forest soil, or a floodplain river, the same principles apply: spatial structure creates niches, temporal dynamics generate functional reserves, and interactions among heterogeneous components produce emergent properties that cannot be predicted from average measurements alone.

The review presented here, drawing on molecular microbiology, ecology, evolution, epidemiology, and clinical translation frameworks, reveals a coherent narrative: the path to a predictive, impactful microbiome science lies not in controlling for heterogeneity, but in embracing, measuring, and modeling it in all its rich dimensions. The methodological toolkit for doing so is increasingly sophisticated—from single-cell genomics and spatial omics to AI-driven causal inference and ecological network analysis. The conceptual frameworks for interpreting what we measure are increasingly mature—from Medical Ecology (Ma & Zhang 2022) and information theory (van der Lubbe 1986; Chen et al. 2025) to the dual-thread ecological vs. environmental heterogeneity (Ma & Ellison 2026), and statistical methods for heterogeneity analysis (Ma, Liu & Ellison 2026). The translational applications are increasingly tangible—from precision synbiotics and keystone-directed therapies to microbiome-based diagnostics and climate-resilient agriculture.

Yet the work is far from complete. The severe sampling bias that distorts our understanding of human microbial heterogeneity must be corrected. The ethical dimensions of template selection and bioprospecting must be addressed. The causal mechanisms linking observed heterogeneity to functional outcomes must be rigorously established through iterative cycles of computational prediction and experimental/clinical validation. The forecasting power of microbiomes must be harnessed to predict responses to environmental change, therapeutic interventions, and emerging threats.

In embracing heterogeneity, we embrace the true nature of the microbial world. We acknowledge that heterogeneity is not noise to be averaged away but signal to be interpreted. We recognize that the template matters as much as the community displayed upon it. And we commit to a science

that is not only predictive and precise, but also equitable and inclusive—attending to the full diversity of templates, human and environmental, that constitute the microbial planet we inhabit.

Finally, we acknowledge that the two text boxes (Box 1 and Box 2) were not fully aligned in their framing. Box 1 captures the classical, often reductionist perspectives on heterogeneity — rooted in diversity metrics and compositional thinking. Box 2, in contrast, represents our attempt to reframe heterogeneity through a cross-disciplinary lens, drawing from ecology (Ma & Ellison 2026), network science (Ma & Ellison 2025), statistics (Ma, Liu & Ellison 2026), and philosophy (Shavit & Ellison 2021) so far.

Their biggest distinction lies in the treatment of fundamental concepts: heterogeneity, diversity, variation (variability), and difference. Classically, heterogeneity is often conflated with diversity (Shavit & Ellison 2021). We emphasize that diversity measures count discrete objects and ignore their relationships (interactions), whereas heterogeneity measures emergent properties from interactions at minimum, and ideally captures both objects and interactions.

One way to define diversity is to recall that it has two components: varieties and abundances, for example, number of species and their relative abundances. Hill numbers (Hill 1973), or equivalently Rényi's entropy, capture this duality ideally, and all can be expressed as discrete Hill numbers. Heterogeneity, in contrast, focuses on interactions in the case of ecological heterogeneity, and on connectedness (mosaics) in the case of environmental heterogeneity. Varieties (types), variability (or variation) and/or differences are too general to represent heterogeneity, although they may arise as an outcome of interactions or connectedness. Such measures are related to, and may even serve as building blocks of, heterogeneity, but they are not heterogeneity per se.

Measuring diversity is now a largely settled problem, particularly following the rediscovery and unification of Hill numbers (Chao et al. 2014). Measuring heterogeneity, however, remains far from settled. It is an open and active frontier, and entropy alone may not be the most illuminating path forward. Instead, variance scaling approaches, such as power laws and network analysis, are beginning to reveal deeper organizational principles that entropy-based metrics often miss (Ji et al. 2019, 2020; Chen et al. 2025; Ma & Ellison 2025; Ma et al. 2026; Ma 2025).

**Box 1**. Relationship between General Ecological and Microbial Heterogeneity

| Aspect | General Ecological Heterogeneity | Microbial Heterogeneity | Relationship / Overlap |
|---|---|---|---|
| **Definition** | The non-uniform distribution and variability of biotic and abiotic factors across spatial, temporal, and | The multi-scale variation in microbial traits, genotypes, and community composition, arising from stochastic molecular processes, rapid evolution, and | Microbial heterogeneity is a specific manifestation of ecological heterogeneity, extending its principles to finer biological scales (molecular, |

| Aspect | General Ecological Heterogeneity | Microbial Heterogeneity | Relationship / Overlap |
|---|---|---|---|
| | organizational scales, driving biodiversity patterns, community assembly, and ecosystem function (Levin, 1992; Turner, 2005; Huber et al., 2020; Nunan et al., 2020). | environmental selection, which serves as a functional adaptive strategy influencing host and ecosystem outcomes (Ackermann, 2015; Lidstrom & Konopka, 2010; Garud & Pollard, 2020; Ma & Ellison, 2024). | strain) and incorporating unique microbial drivers (e.g., horizontal gene transfer, phage dynamics). The two are united by shared ecological theory but differ in resolution, mechanisms, and translational applications. |
| **Scale** | Landscape, ecosystem, community, population, individual. | Molecular, cellular, strain, population (host), community, ecosystem. | **Nested hierarchy:** Microbial heterogeneity operates within and across all traditional ecological scales, adding finer molecular-to-strain layers. |
| **Drivers** | Abiotic gradients, disturbance, biotic interactions, dispersal, evolution (long-term). | Stochastic molecular noise, rapid evolution (mutation/HGT), phage activity, host selection, environmental gradients. | **Shared:** Environmental gradients, biotic interactions, dispersal, evolution. **Distinct:** Microbial systems add intrinsic stochasticity and rapid eco-evolutionary feedbacks as primary drivers. |
| **Manifestation** | Species diversity, trait variation, spatial patchiness, genetic diversity. | Phenotypic noise (persistence), strain-level genomic diversity, interpersonal microbiome variation, spatial micro-niches, phage-host mosaics. | **Shared:** Diversity, spatial structure. **Distinct:** Microbial systems show non-genetic phenotypic heterogeneity and strain-level functional microdiversity beneath species level. |
| **Functional Role** | Stability, resilience, niche partitioning, coexistence, ecosystem functioning. | Bet-hedging, division of labor, functional redundancy, viral shunt, host adaptation, disease modulation. | **Shared:** Promotes stability, resilience, and ecosystem function. **Distinct:** Microbial heterogeneity enables proactive survival strategies (bet-hedging) and sub-cellular division of labor. |
| **Research Methods (Experimental)** | Field surveys, transplant experiments, mesocosms, isotope tracing, long-term monitoring. | Gnotobiotic models, synthetic microbial communities, microfluidics, phage-host assays, stable isotope probing, longitudinal host sampling. | **Shared:** Manipulative experiments, tracer studies. **Distinct:** Microbial methods enable reductionist, high-throughput host-associated experiments impossible in macro-ecology. |
| **Research Methods (Computational)** | GIS, spatial statistics, population genetics models, species distribution modeling. | Strain-resolved metagenomics, multi-omics integration, AI/ML for pattern discovery, information theory, network inference, genome-scale metabolic modeling. | **Shared:** Spatial & genetic analysis. **Distinct:** Microbial informatics requires compositional data analysis, high-dimensional ML, and causal inference from sparse, dynamic data. |
| **Theoretical Framework** | Classical theoretical ecology (e.g., island biogeography, metacommunity theory, niche theory). | Draws from classical theoretical ecology and is supported by computational biology, bioinformatics, medical microbiology, and clinical | **Medical Ecology applies** macrobial ecological theory to human-microbe systems, creating a translational bridge informed by computational and clinical fields. |

| Aspect | General Ecological Heterogeneity | Microbial Heterogeneity | Relationship / Overlap |
|---|---|---|---|
| | | medicine—forming Medical Ecology. | |
| **Translational Impact** | Conservation, climate change mitigation, ecosystem management. | Precision medicine, microbiome therapeutics, pandemic preparedness, sustainable agriculture. | **Shared:** Biodiversity conservation, ecosystem management. **Distinct:** Direct personalized health applications and microbiome engineering. |

**Box 2**. Cross-disciplinary Heterogeneity: From Conceptual Distinction to Analytical Methodology (Ma & Ellison 2026; Ma, Liu & Ellison 2026; Ma *et al*. 2026)

| Section | Description & Core Logic |
|---|---|
| **Heterogeneity vs. Diversity: A Philosophical Argument** (Shavit & Ellison, 2021, *J. of Philosophy*) | **Definition:** Heterogeneity is a property of a *collective*—a group defined by joint processes and a common structure—that arises from the interactions and integration of its different constituent parts. In contrast, diversity is a property of a *population*—a subset within a larger group with no prerequisite for internal structure—that quantifies only the number and abundance of different types while ignoring their relationships. **Core Distinction:** The distinction is crystallized by the analogy: **"a zoo is diverse, whereas an ecosystem is heterogeneous."** A zoo is an aggregated collection of non-interacting populations, while an ecosystem constitutes a collective where populations interact. This demonstrates that heterogeneity = diversity + interaction + integration. **Practical Implication:** Conflating these concepts or applying diversity-based measurements to a collective can lead to a conflict between diversity and heterogeneity, resulting in flawed models and ineffective policies for managing complex, interactive systems. |
| **Working Definition** | Heterogeneity (H) is a multi-faceted, systemic property of groups of objects, formally definable as a tuple of six core elements: Scale (S), Objects (O), Quantity (Q), Dimension (D), Interaction (I), and Environment (E). A seventh element, Method (M), is required for its operational measurement, H = (S, O, Q, D, I, E, M). This definition stipulates that heterogeneity is not a simple metric but an emergent property arising from: Scale-dependent observations of interacting groups of objects (O & I), whose multidimensional quantities (Q & D) exist within a specific environmental context (E), and must be quantified using a fit-for-purpose analytical method (M). |
| **Scale (S)** | **Description:** The indispensable observational lens, parameterized by grain, extent, and context. **Core Logic:** Serves as the foundational constraint for measurability. The *a priori* definition of scale is a prerequisite for rendering heterogeneity a quantifiable property rather than an indeterminate phenomenon. It inherently couples heterogeneity to spatial (or analogous) frameworks and determines the manifestation of scale-dependent attributes, including scale-invariance and critical transitions. |
| **Objects (O) & Interaction (I)** | **Description:** Heterogeneity is a property of groups of objects (O) that emerges from the interactions between them (I). **Core Logic:** Interaction is the defining feature that separates heterogeneity from diversity, which only counts groups. Measuring these interactions is complex, which is why network analysis is essential. Furthermore, the scale (S) at which a system is observed directly determines how these interactions are measured. |
| **Quantity (Q) & Dimension (D)** | **Description:** Quantity (Q) is what is measured (group quantity), and Dimension (D) describes its structure (from unidimensional to hierarchical). **Core Logic:** While quantity is the starting point (shared with diversity), heterogeneity is more complex because it is inherently multidimensional. Time can be treated as one dimension, relating to concepts of stability, though research often employs cross-sectional snapshots. |
| **Environment (E)** | **Description:** The contextual matrix within which objects are situated, with its definition and treatment varying in explicitness across disciplines. **Core Logic:** Environment (E) |

| Section | Description & Core Logic |
|---|---|
| | represents a methodological and conceptual spectrum. This spans from its treatment as a distinct and explicit analytical component (*e.g.*, in ecology, where environmental heterogeneity is studied alongside community heterogeneity) to being ignored or operationalized as a statistical covariate to its integration with objects on a shared, inseparable continuum in other research frameworks. |
| **Method (M)** | **Description:** The pragmatic toolkit for quantifying heterogeneity, encompassing conceptual, quantitatively descriptive, metric, modeling, and integrated framework-based approaches. **Core Logic:** No single methodological approach is universally superior; the selection is contingent upon the specific research question and available resources, necessitating a pragmatic trade-off between parsimony and comprehensive analytical power. Cross-disciplinary methodological exchange is a critical conduit for advancing the overall efficacy of heterogeneity research. |
| **The 3-LHA (Three-layer heterogeneity analysis) Methodology (Ma & Ellison 2025, 2026)** | **Description:** A modular analytical framework inspired by the Open Systems Interconnection (OSI) reference model—the architectural blueprint for modern internet protocols—consisting of three discrete layers: (1) **Definitions/Metrics Layer:** Specifies heterogeneity measures at both node- and network-levels, along with scaling metrics. (2) **Statistical Tests Layer:** Incorporates bootstrapping procedures to partition heterogeneous from random clusters, and permutation tests for comparative analysis of heterogeneity. (3) **Applications Layer:** Supports specific analytical tasks such as identifying keystone species via "rich club" detection, biomarker discovery, and analyzing heterogeneity-stability relationships. **Core Logic:** The 3-LHA provides a coherent, scalable, and cross-domain interoperable analytical stack designed to operationalize the abstract conceptual elements of heterogeneity. It functions as the translational engine that converts the theoretical tuple (S, O, Q, D, I, E, M) into a concrete, actionable workflow for tackling complex, cross-disciplinary problems. |

## References


Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology, 13*(8), 497–508.

Amir, A., & Balaban, N. Q. (2018). Learning from noise: how observing stochasticity may aid microbiology. *Trends in Microbiology, 26*(5), 376-385.

Andersen, K. S., Zhao, K., de Linde Agerskov, A., Sørensen, C. B., Holmager, T. J., Nierychlo, M., Peces, M., Guo, C., & Nielsen, P. H. (2025). Predicting microbial community structure and temporal dynamics by using graph neural network models. *Nature Communications, 16*(1), 9124.

Andreu-Sánchez, S., Blanco-Míguez, A., Wang, D., Golzato, D., Manghi, P., Heidrich, V., Fackelmann, G., Zhernakova, D. V., Kurilshikov, A., Valles-Colomer, M., Weersma, R. K., Zhernakova, A., Fu, J., & Segata, N. (2025). Global genetic diversity of human gut microbiome species is related to geographic location and host health. *Cell, 188*(16), 3942–3959.

Arzamasov, A. A., Rodionov, D. A., Hibberd, M. C., Guruge, J. L., Kent, J. E., Kazanov, M. D., Leyn, S. A., Elane, M. L., Sejane, K., Furst, A., Bode, L., Barratt, M. J., Gordon, J. I., & Osterman, A. L. (2025). Integrative genomic reconstruction reveals heterogeneity in carbohydrate utilization across human gut bifidobacteria. *Nature Microbiology, 10*(8), 2031–2047.

Asnicar, F., Manghi, P., Fackelmann, G., Baldanzi, G., Bakker, E., Ricci, L., Piccinno, G., Piperni, E., Mladenovic, K., Amati, F., Arrè, A., Ganesh, S., Giordano, F., Davies, R., Wolf, J., Bermingham, K. M., Berry, S. E., Spector, T. D., & Segata, N. (2025). Gut micro-organisms associated with health, nutrition and dietary interventions. *Nature*.

Asnicar, F., Thomas, A. M., Passerini, A., Waldron, L., & Segata, N. (2024). Machine learning for microbiologists. *Nature Reviews Microbiology, 22*(4), 191–205.

Bakkeren, E., Piskovsky, V., Lee, M. N. Y., Jahn, M. T., & Foster, K. R. (2025). Strain displacement in microbiomes via ecological competition. *Nature Microbiology, 10*, 3122–3135.

Caminero, A., Tropini, C., Valles-Colomer, M., Shung, D. L., Gibbons, S. M., Surette, M. G., Sokol, H., Tomeo, N. J., Scientific Advisory Board of the Center for Gut Microbiome Research and Education of the American Gastroenterological Association, & Tarr, P. I. (2025). Credible inferences in microbiome research: Ensuring rigour, reproducibility and relevance in the era of AI. *Nature Reviews Gastroenterology & Hepatology*.

Castledine, M., Sierocinski, P., Padfield, D., & Buckling, A. (2020). Community coalescence: an eco-evolutionary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1798), 20190252.

Chakraborty, D., Jousset, A., Wei, Z., & Banerjee, S. (2025). Rare taxa in the core microbiome. *Trends in Microbiology, 33*(7), 727–737.

Chao, A., Chiu, C. H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45(1), 297–324.

Chen, L., Zhao, B., Zhang, M., Yan, Y., Nie, C., Yu, K., Tu, Z., & Xia, Y. (2025). Micron-scale heterogeneity reduction leads to increased interspecies competition in thermophilic digestion microbiome. *Water Research, 274*, 123419.

Chen, XJ, WL Qiao, S Zheng, ZT Dong, GJ Pan (2025) The complexity measure of complex networks based on the information structure heterogeneity, *Chinese Journal of Physics*, Vol. 97, 375-384, https://doi.org/10.1016/j.cjph.2025.03.035.

Chin Fatt, C. R., Asbury, S., Jha, M. K., Minhajuddin, A., Sethuram, S., Mayes, T., Kennedy, S. H., Foster, J. A., & Trivedi, M. H. (2023). Leveraging the microbiome to understand clinical heterogeneity in depression: findings from the T-RAD study. *Translational Psychiatry, 13*(1), 139.

Civitello, D. J., Cohen, J., Fatima, H., Halstead, N. T., Liriano, J., McMahon, T. A., Ortega, C. N., Sauer, E. L., Sehgal, T., Young, S., & Rohr, J. R. (2015). Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Proceedings of the National Academy of Sciences of the United States of America, 112*(28), 8667–8671.

Clegg, T., & Gross, T. (2025). Cross-feeding creates tipping points in microbiome diversity. *Proceedings of the National Academy of Sciences, 122*(19), e2425603122.

Correa-Garcia, S., Constant, P., & Yergeau, E. (2022). The forecasting power of the microbiome. *Trends in Microbiology, 31*(1), 4-8.

Dal Co, A., van Vliet, S., Kiviet, D. J., Schlegel, S., & Ackermann, M. (2020). Short-range interactions govern the dynamics and functions of microbial communities. *Nature Ecology & Evolution, 4*(3), 366–375.

Eisenhauer, N., Angst, G., Asato, A. E. B., Beugnon, R., Bönisch, E., Cesarz, S., Dietrich, P., Jurburg, S. D., Madaj, A.-M., Reuben, R. C., Ristok, C., Sünnemann, M., Yi, H., Guerra, C. A., & Hines, J. (2023). The heterogeneity–diversity–system performance nexus. *National Science Review, 10*(8), nwad109. https://doi.org/10.1093/nsr/nwad109

Eliazar, I. (2021). Five degrees of randomness. *Physica A: Statistical Mechanics and its Applications, 568*, 125662.

Elliott-Graves, A. (2025). Reevaluating Heterogeneity in Evidence Synthesis. *Philosophy of Science*, 1–20. https://doi.org/10.1017/psa.2025.10176

Ferretti, P., Johnson, K., Priya, S., & Blekhman, R. (2025). Genomics of host–microbiome interactions in humans. *Nature Reviews Genetics*.

Garrido-Sanz, D., & Keel, C. (2025). Seed-borne bacteria drive wheat rhizosphere microbiome assembly via niche partitioning and facilitation. *Nature Microbiology, 10*, 1130–1144.

Garud, N. R., & Pollard, K. S. (2020). Population genetics in the human microbiome. *Trends in Genetics, 36*(1), 53–67.

Ghosh, T. S., & O'Toole, P. W. (2023). Human microbiome variance is underestimated. *Current Opinion in Microbiology, 73*, 102288.

Gibson, T. E., Bashan, A., Cao, H. T., Weiss, S. T., & Liu, Y. Y. (2016). On the origins and control of community types in the human microbiome. *PLoS Computational Biology, 12*(2), e1004688.

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine, 24*(4), 392-400.

Greenhough, B., Jervis Read, C., Lorimer, J., Lezaun, J., McLeod, C., Benezra, A., ... & Wills, J. (2020). Setting the agenda for social science research on the human microbiome. *Palgrave Communications, 6*(1), 1-11.

Gupta, V. K., Kim, M., Bakshi, U., Cunningham, K. Y., Davis, J. M., III, Lazaridis, K. N., Nelson, H., Chia, N., & Sung, J. (2020). A predictive index for health status using species-level gut microbiome profiling. *Nature Communications, 11*(1), 4635.

Hayes, C. L., Peters, B. J., & Foster, J. A. (2020). Microbes and mental health: Can the microbiome help explain clinical heterogeneity in psychiatry? *Frontiers in Psychiatry, 11*, 134.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4), 673–748.

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432.

Hu, Z., Wang, Y., Kishore, D., Birzu, G., DeLisi, C., Korolev, K. S., & Segrè, D. (2022). A resource for the comparison and integration of heterogeneous microbiome networks. *bioRxiv*, 2022-08. [Preprint]

Huang, Y. J. (2025). The microbiome in asthma heterogeneity: The role of multi-omic investigations. *Immunological Reviews*, 1-15. Advance online publication.

Huber, P., Metz, S., Unrein, F., Mayora, G., Sarmento, H., & Devercelli, M. (2020). Environmental heterogeneity determines the ecological processes that govern bacterial metacommunity assembly in a floodplain river system. *The ISME Journal, 14*(12), 2951–2964.

Ilyaskina, D., Altveş, S., Dong, L., Bouwmeester, H., & El Aidy, S. (2025). Adaptive and metabolic convergence in rhizosphere and gut microbiomes. *Microbiome, 13*(1), 173.

Ji, B. W., Sheth, R. U., Dixit, P. D., et al. (2019). Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling. *Nature Methods*, 16, 731–736.

Ji, B. W., Sheth, R. U., Dixit, P. D., et al. (2020). Macroecological dynamics of gut microbiota. *Nature Microbiology*, 5, 768–775.

The Integrative HMP (iHMP) Research Network Consortium. (2019). The integrative human microbiome project. *Nature, 569*(7758), 641-648.

Jia, M., Zhu, S., Xue, M.-Y., Chen, H., Xu, J., Song, M., Tang, Y., Liu, X., Tao, Y., Zhang, T., Liu, J.-X., Wang, Y., & Sun, H.-Z. (2024). Single-cell transcriptomics across 2,534 microbial species reveals functional heterogeneity in the rumen microbiome. *Nature Microbiology*.

Joos, R., Boucher, K., Lavelle, A., Arumugam, M., Blaser, M. J., Claesson, M. J., Clarke, G., Cotter, P. D., De Sordi, L., Dominguez-Bello, M. G., Dutilh, B. E., Ehrlich, S. D., Ghosh, T. S., Hill, C., Junot, C., Lahti, L., Lawley, T. D., Licht, T. R., ... Ross, R. P. (2025). Examining the healthy human microbiome concept. *Nature Reviews Microbiology, 23*(3), 192–205.

Kang, Y., Lin, S., Ma, X., Che, Y., Chen, Y., Wan, T., Zhang, D., Shao, J., Xu, J., Xu, Y., Lou, Y., & Zheng, M. (2021). Strain heterogeneity, cooccurrence network, taxonomic composition and functional profile of the healthy ocular surface microbiome. *Eye and Vision, 8*(1), 6.

Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., Myers, S. S., Bogich, T., & Ostfeld, R. S. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature, 468*(7324), 647–652.

Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology, 73*(6), 1943–1967.

Liang, X., Yang, S., Radosevich, M., Wang, Y., Duan, N., & Jia, Y. (2025). Bacteriophage-driven microbial phenotypic heterogeneity: ecological and biogeochemical importance. *npj Biofilms and Microbiomes, 11*(1), 82.

Lidstrom, M. E., & Konopka, M. C. (2010). The role of physiological heterogeneity in microbial population behavior. *Nature Chemical Biology, 6*(10), 705–712.

Liu, Y., Dai, Y., & Wang, L. (2026). Spatial omics at the forefront: emerging technologies, analytical innovations, and clinical applications. *Cancer Cell, 44*(1), 24–41.

Llorens-Rico, V., Simcock, J. A., Huys, G. R., & Raes, J. (2022). Single-cell approaches in human microbiome research. *Cell, 185*(15), 2725-2738.

Lu, J., Satten, G., Meyer, K., Launer, L., Ling, W., & Zhao, N. (2026). Identifying unmeasured heterogeneity in microbiome data via quantile thresholding (QuanT). *Microbiome*, in press.

Ma, Z. S. (2015). Power law analysis of the human microbiome. *Molecular Ecology, 24*(21), 5428–5445.

Ma, Z. S. (2019). A new DTAR (diversity–time–area relationship) model demonstrated with the indoor microbiome. *Journal of Biogeography, 46*(1), 189–202.

Ma, Z. S., & Ellison, A. M. (2019). Dominance network analysis provides a new framework for studying the diversity–stability relationship. *Ecological Monographs, 89*(2), e01358.

Ma, Z. S., Li, L. W., & Gotelli, N. J. (2019). Diversity-disease relationships and shared species analyses for human microbiome-associated diseases. *The ISME Journal, 13*, 1911–1919.

Ma, Z. S. (2020). Heterogeneity-disease relationship in the human microbiome associated diseases. *FEMS Microbiology Ecology, 96*(7), fiaa093.

Ma, Z. S., & Zhang, Y. (2022). Ecology of human medical enterprises: from disease ecology of zoonoses, cancer ecology through to medical ecology of human microbiomes. *Frontiers in Ecology and Evolution, 10*, 879130.

Ma, Z. S. (2023). A new hypothesis on BV etiology: dichotomous and crisscrossing categorization of complex versus simple on healthy versus BV vaginal microbiomes. *mSystems, 8*(4), e00049-23.

Ma, Z. S., & Ellison, A. M. (2024). Cross-scale scaling-law analyses for the heterogeneity and diversity of animal gut microbiomes from community to landscape. *Oikos,* 2024: e10598.

Ma, Z. S. (2024a). Species specificity and specificity diversity (SSD) framework: a novel method for detecting the unique and enriched species associated with disease by leveraging the microbiome heterogeneity. *BMC Biology*. Vol. *22*(1), 283. https://link.springer.com/article/10.1186/s12915-024-02024-7

Ma, Z. S. (2024b). Revisiting microgenderome: detecting and cataloguing sexually unique and enriched species in human microbiomes. *BMC Biology*. Vol. *22*(1), 284. https://link.springer.com/article/10.1186/s12915-024-02025-6

Ma, Z. S. (2025). Taylor's power law meets ecological networks: introducing TPLoN as a novel framework for heterogeneity scaling analysis. *Oikos*, 2025: e10956.

Ma, Z. S., & Taylor, R. A. J. (2025). Six decades post-discovery of Taylor's Power Law: From ecological and statistical universality, through prime number distributions and tipping-point signals, to heterogeneity and stability of complex networks. *arXiv*. https://doi.org/10.48550/arXiv.2506.18154
***Physics of Life Review*** (in Press)

Ma, Z. S., & Ellison, A. M. (2025). A Novel Three-Layer Methodology for Heterogeneity Analysis: from metrics, through statistical tests and modeling to applications. *bioRxiv* 2025.12.18.695221.
***Methods in Ecology and Evolution*** (in Revision).

Ma, Z. S., & Ellison, A. M. (2026a). *Heterogeneity: Meaning and Measurement, Causes and Consequences*. *EcoEvoRxiv*. https://doi.org/10.32942/X2Z66N

Ma, Z. S., Liu, S & Ellison, A. M. (2026b). Heterogeneity in Statistics: A Conceptual and Methodological Review. EcoEvoRxiv. https://ecoevorxiv.org/repository/view/12583/
***J of Statistical Theory and Practice*** (in Revision).

Ma ZS, Li LW, Liu S, and Ellison, A (2026) Revisiting Taylor's law, rediscovering Smith's law: A dual framework for scaling heterogeneity. *EcoEvoRxiv*. https://ecoevorxiv.org/repository/view/14350/

Martins, B. M. C., & Locke, J. C. W. (2015). Microbial individuality: how single-cell heterogeneity enables population level strategies. *Current Opinion in Microbiology, 25*, 1–7.

Mirzayi, C., Renson, A., Genomic Standards Consortium, Massive Analysis and Quality Control Society, Zohra, F., Elsafoury, S., ... & Waldron, L. (2021). Reporting guidelines for human microbiome research: the STORMS checklist. *Nature Medicine, 27*(11), 1885-1892.

Neuenkamp, L., & Saiz, H. (2025). Space-for-time substitution can predict grassland biodiversity dynamics. *Nature Ecology & Evolution*.

Noecker, C., & Turnbaugh, P. J. (2024). Emerging tools and best practices for studying gut microbial community metabolism. *Nature Metabolism, 6*(7), 1225–1236.

Nunan, N., Schmidt, H., & Raynaud, X. (2020). The ecology of heterogeneity: soil bacterial communities and C dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences, 375*(1798), 20190249.

Núñez Casal, A. (2024). Race and indigeneity in human microbiome science: microbiosmisation and the historiality of otherness. *History and Philosophy of the Life Sciences, 46*(2), 17.

Oña, L., Shreekar, S. K., & Kost, C. (2025). Disentangling microbial interaction networks. *Trends in Microbiology, 33*(6), 619–632.

Pekar, J. E., Lytras, S., Ghafari, M., Magee, A. F., Parker, E., Wang, Y., Ji, X., Havens, J. L., Katzourakis, A., Vasylyeva, T. I., Suchard, M. A., Hughes, A. C., Hughes, J., Rambaut, A., Robertson, D. L., Dellicour, S., Worobey, M., Wertheim, J. O., & Lemey, P. (2025). The recency and geographical origins of the bat viruses ancestral to SARS-CoV and SARS-CoV-2. *Cell, 188*(13), 3167–3183.

Peng, H., Ruiz-Moreno, A. J., & Fu, J. (2025). Multi-dimensional metagenomics. *Nature Reviews Bioengineering, 3*, 289–308.

Pichon, M., Bouleti, C., MICMAC Study Group SFM, Hery-Arnaud, G., & Buruoca, C. (2025). Heterogeneity and lack of standardisation in gut microbiome testing: a comparative assessment of French medical biology laboratories. *Gut*, Published Online First.

Sawhney, S. S., Thänert, R., Thänert, A., Hall-Moore, C., Ndao, I. M., Mahmud, B., Warner, B. B., Tarr, P. I., & Dantas, G. (2025). Gut microbiome evolution from infancy to 8 years of age. *Nature Medicine, 31*, 2004–2015.

Shade, A., & Gilbert, J. A. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. *Trends in Microbiology, 23*(6), 335-340.

Shanahan, F., Ghosh, T. S., & O'Toole, P. W. (2023). Human microbiome variance is underestimated. *Current Opinion in Microbiology, 73*, 102288.

Shavit, A., & Ellison, A. M. (2021). Diverse Populations are Conflated with Heterogeneous Collectives. *The Journal of Philosophy, 118*(10), 525-548.

Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2017). Information Theory Broadens the Spectrum of Molecular Ecology and Evolution. *Trends in Ecology & Evolution*, 32(12)

STREAMS Working Group. (2025). STREAMS guidelines: standards for technical reporting in environmental and host-associated microbiome studies. *Nature Microbiology, 10*, 3059–3068.

Tian, L., Wang, X. W., Wu, A. K., Fan, Y., Friedman, J., Dahlin, A., Waldor, M. K., Weinstock, G. M., Weiss, S. T., & Liu, Y. Y. (2020). Deciphering functional redundancy in the human microbiome. *Nature Communications, 11*(1), 6217.

Turner, M. G. (2005). Landscape ecology: What is the state of the science? *Annual Review of Ecology, Evolution, and Systematics, 36*, 319–344.

van der Lubbe, J. C. A. (1986). An axiomatic theory of heterogeneity and homogeneity. *Metrika*, 33, 223-245. DOI: 10.1007/BF01894751

VanEvery, H., Franzosa, E. A., Nguyen, L. H., & Huttenhower, C. (2023). Microbiome epidemiology and association studies in human health. *Nature Reviews Genetics, 24*(2), 109–121.

Varley, T. F. (2025). Information theory for complex systems scientists: What, why, and how. *Physics Reports, 1148*, 1–55.

Vienni-Baptista, B., Fletcher, I., Lyall, C., & Pohl, C. (2022). Embracing heterogeneity: Why plural understandings strengthen interdisciplinarity and transdisciplinarity. *Science and Public Policy, 49*(6), 865–877. https://doi.org/10.1093/scipol/scac034

Violle, C., Enquist, B. J., McGill, B. J., Jiang, L., Albert, C. H., Hulshof, C., ... & Messier, J. (2012). The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution, 27*(4), 244-252.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T.-Y.,

Manrai, A., ... Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature, 620*(7974), 47–57.

Wang, X.-W., Sun, Z., Jia, H., Michel-Mata, S., Angulo, M. T., Dai, L., He, X., Weiss, S. T., & Liu, Y.-Y. (2024). Identifying keystone species in microbial communities using deep learning. *Nature Ecology & Evolution, 8*(1), 22–31.

Week, B., Russell, S. L., Schulenburg, H., Bohannan, B. J. M., & Bruijning, M. (2025). Applying evolutionary theory to understand host–microbiome evolution. *Nature Ecology & Evolution*.

Wilkinson, D. A., Marshall, J. C., French, N. P., & Hayman, D. T. S. (2018). Habitat fragmentation, biodiversity loss and the risk of novel infectious disease emergence. *Journal of the Royal Society Interface*, 15(149), 20180403.

Wille, M., Shi, M., Klaassen, M., Hurt, A. C., & Holmes, E. C. (2019). Virome heterogeneity and connectivity in waterfowl and shorebird communities. *The ISME Journal, 13*(10), 2603–2614.

Wisnoski, N. I., & Lennon, J. T. (2023). Scaling up and down: movement ecology for microorganisms. *Trends in Microbiology, 31*(11), 1122-1131.

Zhang, Z., Geng, J., Tang, X., Fan, H., Xu, J., Wen, X., Ma, ZS., & Shi, P. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *The ISME Journal, 8*(4), 881-893.

Zhernakova, A., Yassour, M., Hall, L. J., & Collado, M. C. (2025). Unlocking the power of human milk and infant feeding: Understanding how nutrition and early microbiota interaction shapes health programming. *Cell Host & Microbe, 33*(6), 820–829.

Web Link #1: https://harvardforest.fas.harvard.edu/notes/bullard-spotlight-comparative-investigation-of-heterogeneity-across-the-natural-sciences-social-sciences-and-humanities/ (Accessed on Sept. 1st, 2026)

## Acknowledgements

I thank Dr. Aaron Ellison for his thoughtful comments and suggestions on revising this manuscript. I am deeply indebted to him for our collaborations on the heterogeneity project during my Bullard Fellowship at Harvard University. Besides Bullard Fellowship, this study was also supported by the National Natural Science Foundation of China (grant no. 72274192) and the National Key Research and Development Program of China (grant no. 2021YFF1000602).

During the preparation of this manuscript, the AI tool Deepseek® was used for language polishing and reference formatting and summarization. The author takes full responsibility for the content and accuracy of the manuscript.