Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Initial Cell Number Regulates Microscale Community Assembly and Structure

Zhaoshuo Zhang
Huazhong Agricultural University
Chun-Hui Gao
gaoch@mail.hzau.edu.cn

Huazhong Agricultural University
Fangkui Wang
Huazhong Agricultural University
Jiliang Hu
Tsinghua Shenzhen International Graduate School
Huan He
Huazhong Agricultural University
Yichao Wu
Huazhong Agricultural University
Chenchen Qu
Huazhong Agricultural University
Ke Dai
Huazhong Agricultural University
Qiaoyun Huang
Huazhong Agricultural University
Peng Cai
Huazhong Agricultural University

Article

Keywords: soil habitat, community assembly, bacterial interaction, niche condition, community cell numbers

Posted Date: August 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10109398/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Soil is the most complex biological material on Earth, with highly diverse microbial communities showing pronounced spatial heterogeneity at the microscale as small aggregates of tens to hundreds of cells within biofilms. We hypothesized that the initial population size (initial cell number) in these micro-aggregates profoundly affects community assembly and function. To test this hypothesis, we used flow cytometry to construct thousands of random communities with precisely controlled initial cell numbers (10, 50, 100, or 250 cells per well) across four different carbon sources that mimic natural soil microhabitat heterogeneity. We systematically evaluated the effects of initial cell number on community phenotypes, structure, diversity, and association networks.

Results

Initial cell number exerted carbon-source-dependent effects on growth but consistently showed a strong positive correlation with biofilm formation. After incubation, communities underwent sharp declines in diversity and converged toward structures dominated by Pseudomonas and Klebsiella. Pseudomonas species were strongly enriched at low initial cell numbers, whereas Klebsiella species increased at high initial cell numbers. Evenness increased with initial cell number; low-cell-number communities harbored more unique amplicon sequence variants and negative-correlation networks, while high-cell-number communities promoted shared variants and positive connections. This cell-number-dependent switching of dominant taxa extended to intra-genus niche differentiation at the amplicon sequence variant level.

Conclusions

Our findings identify initial cell number as a critical driver of microscale soil microbial community assembly. Initial cell number regulates competition–coexistence dynamics of dominant taxa, intra-genus ASV-level niche differentiation, and microbial association networks. This study highlights initial cell number as a key factor shaping community structure, diversity, and function in soil microbiomes.

Background

Soil harbors one of the richest microbial communities on Earth, with bacterial cell densities estimated to exceed 2 × 10^9 cells per gram, constituting a major component of global biodiversity[1, 2]. These microorganisms drive critical ecosystem processes—including nutrient cycling, organic-matter decomposition, and plant-health maintenance—thereby sustaining soil fertility and terrestrial productivity[3]. Despite their ubiquity and functional importance, soil microbial communities are not homogeneous entities. Instead, microorganisms predominantly localize near plant roots, organic matter, aggregates, and pore spaces, forming “hotspots”. These hotspots occur in nutrient-rich organic zones

such as the rhizosphere, fungal hyphae, and plant/animal residues, where local conditions dictate community composition and dynamics[4–6]. Consequently, soil microbial communities exhibit pronounced spatial heterogeneity, assembled independently as countless microscale aggregates[7]. This necessitates examination at finer scales to uncover emergent community properties. At the microscale, soil microbial communities typically comprise only tens to hundreds of cells[8], embedded in biofilms, creating protected niches against desiccation, predation, and resource scarcity while facilitating quorum sensing and metabolic putative associations[9, 10]. Physical barriers and diffusion limitations between microscale communities result in relatively isolated assembly processes with minimal inter-aggregate interactions[11]. Therefore, microscale variation in founding population size—i.e., initial cell number as the starting state of distinct microhabitats—may profoundly influence assembly trajectories and final structure. Specifically, differences in initial cell number (directly convertible to cell-number variation within a fixed volume) may influence community assembly trajectories and final structure.For example, low initial cell number may increase the relative importance of stochastic processes (such as priority colonization and local extinction), while higher initial cell number may favor deterministic processes (such as competition or cooperation)[12, 13].

Although extensive research has addressed abiotic drivers and biotic interactions in soil community assembly[14, 15], the role of initial cell number as a fundamental constraining factor remains severely understudied. Theoretical frameworks in microbial ecology, including neutral theory and niche-based models, emphasize that community assembly arises from the balance between stochastic (e.g., dispersal and drift) and deterministic (e.g., selection and environmental filtering) processes[16, 17]. This balance may be amplified by demographic noise at low initial densities, enhancing ecological drift, accumulation of unique amplicon sequence variants (ASVs), and reductions in community diversity[18]. When population size falls below a initial cell number threshold, negative growth rates increase extinction risk; above the threshold, positive feedbacks enhance survival, competitive advantage, and functional adaptation[19].In soil, initial cell number effects are particularly critical at the microscale because the limited volume of microhabitats directly translates initial cell number into constraints on initial cell number, thereby amplifying stochastic processes or triggering deterministic interactions. Dilution-to-extinction or synthetic-community experiments have hinted at initial cell number effects on diversity and function[20, 21], however, few studies have examined the potential impacts of initial cell number on evolutionary dynamics of low-complexity communities across multiple microhabitat conditions[22].

In the present study, we isolated precise numbers of soil cells via flow cytometry and inoculated them into 96-well microcosms with varying initial cell numbers across multiple carbon sources. This setup was designed to simulate microscale variation in founding population size observed in soil micro-aggregates, rather than directly studying natural soil micro-aggregates. We hypothesized that initial cell number acts as a primary extrinsic factor in microscale community assembly, constraining species diversity and population pressure, thereby driving heterogeneity in culturable community structure and interaction patterns. To this end, this study systematically examined phenotypic traits, structural convergence, diversity changes, and association network features of communities under different initial

cell numbers. The results reveal that initial cell number is a core constraining factor in microscale soil microbial community assembly. By regulating competition–coexistence dynamics of dominant taxa, intra-genus niche differentiation, and microbial association networks, it strongly shapes the multifaceted characteristics of communities. This study provides direct experimental evidence for microscale soil microbiome assembly rules and highlights how initial cell number drives community structure, diversity, and functional differentiation.

Results

Community Convergence of Microbial Community Structure under Limited Conditions

To examine the effects of initial inoculation cell number on community phenotypes, we used flow-cytometric sorting of soil-extracted microorganisms into 96-well microcosms to precisely control initial cell numbers across multiple carbon sources.(Fig. S1). Propidium iodide (PI) staining was employed during sorting to exclude cells lacking intact membranes (Fig. S1), ensuring that every sorted cell was viable. Given that interaction ranges among bacteria in soil microbial communities typically span tens to hundreds of cells, we selected 10, 50, 100, and 250 cells as initial states and used cellobiose, glutamic acid, trehalose, and mannitol as carbon sources to simulate nutrient variation in microhabitats (Fig. 1).

After two weeks of the above incubations with varying initial cell numbers in the four carbon sources, growth (OD_{600}) and biofilm formation (crystal violet staining) were quantified (Fig. S2). Results indicated thatcommunity growth differed across initial cell numbers: in cellobiose, growth increased with rising initial cell number (Fig. S2a); in mannitol, growth decreased clearly (Fig. S2d); and in glutamic acid and trehalose, growth was unaffected by initial cell number (Fig. S2b, c). Nevertheless, increasing initial cell number exhibited a consistently positive correlation with biofilm formation across all carbon sources (Fig. S2e–h). This relationship remained highly significant after normalizing biofilm biomass to total growth (specific biofilm formation, SBF = crystal violet / OD_{600}) in every carbon source (Fig. S2i–l), demonstrating that higher founding cell numbers promoted biofilm development independent of overall biomass accumulation.

Given the influences of initial cell number on growth and biofilm formation, we further explored potential relationships between community phenotypes and structure. Approximately 8,000 parallel cultures were classified into four phenotypic groups (HGHB, HGLB, LGHB, LGLB) based on high/low growth and biofilm levels (Fig. S3). Representative samples from each of the four groups were selected from the four carbon-source incubations and subjected to 16S rRNA gene amplicon sequencing. Comparison of community composition with source soil and sorted-cell populations revealed a progressive and significant decline in biodiversity from native soil microbiota through extracted cells to microplate cultures (Fig. 2a). FEAST source-tracking showed that native soil explained 84.60% of the soil viable-cell extract and 68.81% of post-incubation microplate communities (Fig. 2b). Conversely, microplate communities explained only 46.93% of the cell extracts, cell extracts explained 49.82% of soil

microbiota, and incubated communities explained merely 0.74% of soil structure (Fig. 2b). These patterns indicate that extraction and defined-condition cultivation impose intense environmental filtering, retaining only a tiny fraction of native soil diversity in the final microplate systems.

Phylogenetic trees of ASVs within the four representative groups showed highly clustered branches post-incubation, predominantly within a few phyla, with *Pseudomonadota* dominating. Correspondingly, genus-level relative-abundance bar plots revealed that incubated communities were governed by a small number of dominant genera—including *Pseudomonas*, *Klebsiella*, *Aeromonas*, and *Kosakonia*—whose abundances far exceeded those of other taxa (Fig. 2c). These results further demonstrate that post-incubation microbiomes undergo marked simplification and convergence in diversity, source contribution, and taxonomic composition relative to native soil and cell extracts, indirectly reflected in the final community-structure distribution patterns.

Initial Cell Number Influences Dominant-Species Enrichment and Community

Structural Variation

Building on the observed structural convergence and the regulatory effects of initial cell number on growth and biofilm phenotypes, we analyzed taxonomic composition, diversity, and structural differences across initial cell-number levels, carbon sources, and phenotypic categories to evaluate how initial cell number and other factors shape dominant-species distribution and overall community dynamics.

Genus-level relative-abundance stacked bar plots for all incubated samples (N = 384) (Fig. 3a and Fig. S4) showed that communities were primarily composed of a few genera, with *Pseudomonas* and *Klebsiella* dominating most samples; other genera such as Stenotrophomonas, Aeromonas, Kosakonia, and Pseudoxanthomonas appeared at lower abundances in subsets of samples. When samples were grouped by carbon source (cellobiose, glutamic acid, trehalose, mannitol), cell number (10, 50, 100, 250), and phenotypic category (HGHB, HGLB, LGHB, LGLB), the banding patterns in the bar plots displayed clear differences.

Analysis of the relationship between Shannon diversity index and different treatments (Fig. 3b) revealed significant differences among carbon sources: the glutamic acid and mannitol groups exhibited higher diversity, while the cellobiose and trehalose groups showed lower diversity ($p = 0.00077$). Shannon diversity increased with increasing initial cell number ($p = 4.9 \times 10^{-11}$), whereas no significant differences were observed among phenotypic categories ($p = 0.86$). Notably, although the Shannon index increased with initial cell number, Chao1 richness and the number of observed ASVs showed no significant changes (Fig. S5). These results indicate that initial cell number and carbon source primarily influence community diversity by altering the evenness of species abundances, whereas phenotypic category (growth–biofilm combination) is unrelated to diversity.

To assess enrichment of dominant taxa across initial-cell-number groups, microbiome genus set enrichment analysis (mgsea) was performed on *Pseudomonas* relative abundance (Fig. 3c–e). When communities were ranked by primary abundant genus (*Pseudomonas*), samples incubated in mannitol and glutamic acid showed significant positive enrichment (NES > 0, $p < 0.0001$; Fig. 3c), whereas those in cellobiose and trehalose exhibited significant negative enrichment (NES < 0, $p < 0.001$; Fig. 3c). Similarly, low cell numbers (10) displayed strong positive enrichment, while high cell numbers (100, 250) showed significant negative enrichment (Fig. 3d). The enrichment was most pronounced at 10 cells (NES = 3.19, $p = 8.60 \times 10^{-19}$), exceeding the effect size of carbon-source type (Fig. 3c, d). In contrast, no significant positive or negative enrichment was obtained when ranked by phenotypic group ($p > 0.05$; Fig. 3e). Parallel mgsea on the secondary abundant genus (*Klebsiella*) yielded consistent patterns (Fig. S6a–c). Principal-coordinate analysis (PCoA) further evaluated β-diversity (Fig. S7). Samples grouped by phenotype within each carbon source showed no significant separation (Fig. S7a–d, $p > 0.05$), whereas grouping by cell number produced clear separation (Fig. S7e–h, $p < 0.05$), indicating that cell number influenced community structure while phenotype did not. Collectively, these findings confirm that both initial cell number and carbon-source type exert strong, dominant effects on community-structure patterns, whereas phenotypic category plays a minor role—further underscoring the pivotal role of initial cell number in driving structural variation.

Initial Cell Number Influences Putative Microbial Associations

The above results indicate that communities converged under limited conditions into simple assemblages dominated by *Pseudomonas* and *Klebsiella*, with abundance distribution, diversity, and structural differences governed by initial cell number and carbon source. We therefore next examined how initial cell number and carbon source influenced ASV abundance trends, sharing patterns, cell-number preferences, and microbial association networks to reveal their multifaceted regulatory roles in microscale community assembly.

Figure 4. Impact of initial cell number on community dynamics. (a-d) Relative abundance trends of the top 10 taxa across different cell numbers for each carbon source. Line plots show mean relative abundance (± SEM) of the top 10 taxa at cell inoculation levels of 10, 50, 100 and 250 for cellobiose, glutamic acid, mannitol and trehalose (top to bottom). (e-f) ASV sharing among different carbon sources and cell-number groups. Venn diagrams illustrate the number of shared and unique ASVs across the four carbon sources (left) and the four cell numbers (right). (g) SLR analysis of all samples, showing cell number as the primary latent variable influencing the community network. Heatmap displays degree of contribution between latent components (PC1–PC10) and the factors carbon source, cell number and biofilm class. (h) The x-axis shows the Spearman correlation coefficient (Rho); positive values indicate ASVs that significantly increase with higher cell numbers (high cell number preference, red), while negative values indicate ASVs that significantly decrease with higher cell numbers (low cell number preference, blue). Gray dots represent non-significant ASVs ($P \geq 0.05$). The y-axis shows $-\log_{10}$(P-value). Representative ASVs are labeled to highlight cell-number-dependent enrichment patterns. (i) Microbial

association network constructed by SLR. Node size represents degree (number of connections), node color corresponds to genus (see legend), edge thickness indicates association confidence/strength (SLR conditional dependence or statistical association strength), and edge color denotes association direction (red: positive; green: negative).

Trends in relative abundance of the major genera with initial cell number across carbon sources (Fig. 4a) showed that, in all four carbon sources except glutamic acid, Pseudomonas abundance declined sharply with increasing initial cell number, while Klebsiella abundance rose, producing a reciprocal negative-correlation pattern (initial cell number dependent switching) between the two dominant groups. This initial cell number dependent switching was robust to compositional data bias correction (Fig. S8). Other genera (e.g., Aeromonas and Kosakonia) showed only minor fluctuations without consistent trends.To determine whether the cell-number-dependent changes at the genus level were also reflected at the ASV level, we further analyzed ASV sharing among different carbon sources and initial cell-number groups (Fig. 4b). The results showed that most ASVs were highly shared across the four carbon sources and different initial cell-number groups; however, several condition-specific unique ASVs still emerged in the trehalose carbon source or at low initial cell numbers (10 and 50) (Fig. 4b and Fig. S9). This sharing and specificity pattern indicates that different ASVs exhibit distinct responses to initial cell number. To further reveal from the network level how initial cell number drives these ASV responses, we applied the Sparse Latent Factor Regression (SLR) algorithm to compress the high-dimensional interaction relationships into 10 representative latent variables and associated them with initial cell number, carbon source and other environmental factors, and phenotypic groups (Fig. 4c). The results showed that initial cell number contributed the highest proportion in latent variable 5 (PC5) (Fig. 4c). This result indicates that initial cell number is the primary driver shaping variation in microbial association networks, and its effect is independent of carbon source and phenotypic category, exhibiting cross-habitat robustness. On this basis, to further reveal the specific manifestation of this major driver at the ASV level, we quantitatively assessed the preference of each ASV for initial cell number by Spearman correlation coefficients between its relative abundance and initial cell number (Fig. 4h). The results showed that some ASVs of *Pseudomonas* (e.g., ASV3, ASV35) exhibited significant negative correlations (low initial cell number preference), while ASVs of *Klebsiella* (e.g., ASV47) exhibited significant positive correlations (high initial cell number preference), consistent with the overall genus-level abundance trends. Meanwhile, different ASVs within the same genus showed obvious heterogeneity in initial cell number preference; for example, some *Pseudomonas* ASVs preferred low cell number, while others showed no significant preference or preferred high cell number. Under different cell-number conditions, it was precisely these ASVs with clear cell-number preferences that primarily contributed to the genus-level abundance (Fig. S10). These results indicate that cell-number-dependent ASVs are the core components through which initial cell number influences community structure.

To further distinguish the effects of cell number from true microbial interactions, we constructed an association network excluding the direct influence of initial cell number based on the above SLR algorithm (Fig. 4i). We found that cell number dependent ASVs occupied core positions in the microbial association network. Among them, low cell-number-preferring ASVs (e.g., ASV3, ASV10) acted as highly

connected nodes, radiating numerous negative correlations (green edges) to surrounding nodes, forming clear putative negative associations patterns; whereas high cell-number-preferring ASVs primarily connected to other nodes through positive correlations, exhibiting potential cooperative features (Fig. 4i). The results indicate that, after removing the direct influence of initial cell number, the putative associations in the association network still centered on these cell number dependent ASVs, thereby driving the reconstruction of the association network with changing cell number. In summary, initial cell number, as a core constraining factor in microscale community assembly, drives dominant-species switching and community structural differentiation by regulating ASV-level initial cell number preferences, intra-genus niche heterogeneity (Fig. S10), and putative associations patterns.

Discussion

By precisely controlling initial inoculation cell numbers and constructing random communities across four carbon sources, this study systematically revealed the effects of initial cell number on community phenotypes, structure, diversity, and association networks. The results show that initial cell number strongly influences heterogeneity, biofilm formation, community assembly, and functional differentiation in culturable soil microbiomes, and acts as a key driver of microscale historical contingency and priority effects. This finding advances our understanding of how initial cell number modularity influences soil ecosystem functioning[8, 11].

We found that the regulation of community growth by initial cell number exhibited significant carbon-source dependence: growth increased with rising cell number in cellobiose (Fig. S2a), decreased in mannitol (Fig. S2d), and showed no significant effect in glutamic acid and trehalose (Fig. S2b, c). This carbon-source dependence further illustrates the interplay between microbial communities and soil properties[23].However, regardless of carbon source, increasing initial cell number consistently promoted biofilm biomass accumulation (Fig. S2e–h). This positive effect remained highly significant after normalizing biofilm biomass to total growth (specific biofilm formation, SBF = CV/OD_{600}; Fig. S2i–l), indicating that higher founding densities promote biofilm matrix production independent of overall biomass. This is consistent with low initial cell number triggering Allee-like effects, whereas higher initial cell number enhances quorum sensing and extracellular-matrix synthesis, thereby promoting biofilm accumulation and protective capacity independent of overall growth changes[19]. This mechanism is highly consistent with the adaptive strategies of natural soil microhabitats facing desiccation, predation, and resource fluctuations[9, 10]. Post-incubation communities exhibited typical simplification and convergence features: α-diversity declined from soil to cell extracts to cultures (Fig. 2a), and the communities were highly enriched in the *Pseudomonadota* phylum and dominated by a few genera such as *Pseudomonas* and *Klebsiella* (Fig. 2c). FEAST source-tracking further confirmed that the cultivation process imposed strong environmental filtering and selective loss, with incubated communities primarily originating from soil extracts (85.55%), yet the reverse soil explanation rate was low (only 0.74%) (Fig. 2b). This convergence phenomenon is consistent with observations in culturable community studies, where resource limitation and culture conditions favor the selection of fast-growing strains[24].Recent experiments have shown that microbial communities can exhibit emergent dynamical

phases from high-diversity stable coexistence to low-diversity regimes governed by simple aggregate parameters[25]. Our findings are consistent with this framework and suggest that initial cell number may represent one such critical aggregate parameter influencing community structure and dominant-taxon composition at the microscale.

The regulatory effect of initial cell number on community structure and diversity was most prominent. Shannon diversity increased with initial cell number ($p = 4.9\times10^{-11}$, Fig. 3b), primarily due to improved evenness of species abundances, while Chao1 richness and observed ASV numbers showed no significant changes (Fig. S5). Although clear differences existed among carbon sources (higher in glutamic acid and mannitol groups, lower in cellobiose and trehalose groups), diversity was unrelated to phenotypic category (growth–biofilm combination) ($p = 0.86$). More importantly, this diversity increase was accompanied by significant changes in dominant-species composition, exhibiting clear initial cell number dependence: at low cell numbers (10–50), the *Pseudomonas* genus was highly enriched (relative abundance often > 80%–90%; Fig. 3a, c, and Fig. S4), likely due to its rapid resource acquisition capacity in sparse populations[26]; at high cell numbers (100–250), *Klebsiella* abundance increased (30%–60%; Fig. 3a, Fig. 4a, and Fig. S4). This reciprocal negative-correlation pattern (Fig. 4a) was robust to compositional data bias correction (Fig. S8) and may originate from putative positive associations and differences in quorum sensing, with *Klebsiella* benefiting from enhanced biofilm matrix and putative positive associations at higher densities[27]. mgsea analysis further showed that the selective strength of cell number even exceeded that of carbon-source type itself (Fig. 3c–d, and Fig. S6).

β-diversity (PCoA) further confirmed that initial cell number was the dominant driver of community structural heterogeneity ($p < 0.05$; Fig. S7), whereas phenotypic category showed no significant separation ($p > 0.05$; Fig. S7). Low cell-number groups contained more unique ASVs (Fig. 4b, Fig. S9) and negative-correlation networks, consistent with an increased role of stochastic processes; high cell-number groups promoted patterns consistent with deterministic processes (competitive balance and putative positive associations)[18]. This also indicates that diversity gains occurred independently of phenotypic changes, possibly because higher initial cell numbers alleviate stochastic extinction and priority effects, thereby increasing the recoverable species pool[12]. Importantly, this cell-number-dependent species dynamics is not only manifested at the genus level but extends more deeply to intra-genus ASV-level niche differentiation (Fig. S10). Different ASVs within the *Pseudomonas* genus exhibited significant heterogeneity in preference for initial cell number: some ASVs contributed the majority of abundance at low cell numbers, while others became enriched at high cell numbers; the *Klebsiella* genus similarly showed initial cell number preference differences. This intra-genus differentiation allows the same genus to achieve fine-scale niche partitioning under different founding densities, thereby maintaining community resilience to microenvironmental perturbations. This result emphasizes that genus-level conclusions alone should not be relied upon to infer community dynamics—although higher taxonomic ranks possess a certain degree of ecological coherence[28], the heterogeneity at the ASV/strain level under microscale initial cell number conditions is the core mechanism driving dominant-species switching and network reconstruction.

At the level of microbial association networks, SLR analysis showed that initial cell number was the primary driver shaping variation in co-occurrence networks (Fig. 4c). After removing the direct effect of cell number, low cell number preferring ASVs still occupied core nodes in the association network and radiated numerous putative negative associations (Fig. 4i), further reflecting the important role of negative biotic interactions in community assembly.[13]. This network reconstruction pattern confirms that initial cell number drives systematic shifts in the association network by regulating ASV-level initial cell number preferences and intra-genus niche heterogeneity.[29]. Low cell-number groups contained more unique ASVs (Fig. 4b, Fig. S9) and negative-correlation networks, reflecting the amplification of stochastic processes (ecological drift and local extinction); high cell-number groups promoted deterministic processes (competitive balance and putative positive associations)[18]. Overall, this study deepens our understanding of the modular structure of soil microbiomes by revealing that initial cell number (reflecting the degree of microscale isolation) influences community assembly and function, with potential extension from the microscopic to the ecosystem scale. This finding not only provides direct experimental evidence for understanding the assembly rules of soil microbial communities at the microhabitat level, but also provides a theoretical basis for predicting community responses to global change and designing initial cell number based microbial engineering strategies (such as biofilm-enhanced remediation). In natural soils, such cell-number effects under micro-environmental differences (e.g., carbon-source gradients) may help explain the spatial heterogeneity of biogeochemical cycling[4].

Conclusions

This study identifies initial cell number as a critical driver of microscale soil microbial community assembly. By precisely controlling initial cell numbers, we show that initial cell number regulates community phenotypes, structure, diversity, and association networks. Higher initial cell numbers consistently promote biofilm formation. Post-incubation communities converge to structures dominated by Pseudomonas and Klebsiella, with strong initial cell number dependence: *Pseudomonas* are enriched at low cell numbers, while *Klebsiella* increase at high numbers. This cell-number-dependent switching extends to intra-genus ASV-level niche differentiation. Low-cell-number communities harbor more unique ASVs and negative-correlation networks, whereas high-cell-number communities promote shared ASVs and positive connections.These results demonstrate that initial cell number is a core constraining factor in microscale community assembly that shapes microbial association networks. The findings provide direct experimental evidence for microscale soil microbiome assembly rules and highlight how initial cell number drives community structure, diversity, and functional differentiation.

Materials and Methods

Soil Microbial Extraction and Medium Preparation

Surface soil (0–15 cm) was collected with a sterilized hand shovel (~ 0.5 kg per site) from a maize field at the Huazhong Agricultural University experimental station (114°35'E, 30°47'N, Wuhan, Hubei, China) after surface disinfection with 75% ethanol. Samples were stored in sealed bags on ice and transported

to the laboratory. Upon arrival, each sample was homogenized, passed through a sterilized 2-mm sieve, and thoroughly mixed.

Soil microorganisms were extracted using a modified Nycodenz initial cell number gradient centrifugation method[30]. First, 50 g of air-dried homogenized soil was physically dispersed in 100 mL ice-cold T20 solution (PBS containing 0.5% Tween-20) in a high-speed homogenizer cup, mixed, and homogenized at 22,000 rpm for 3 min (with 1-min intervals and ice cooling to prevent overheating). The resulting soil slurry was kept on ice throughout. To avoid cross-contamination, homogenizer cups and blades were rinsed with deionized water and re-disinfected with 75% ethanol after each sample. After homogenization, the slurry was allowed to settle briefly to reduce foaming. Initial cell-number separation was performed under chilled conditions. Sterile 2-mL tubes received 1.0 mL 80% (w/v) Nycodenz as the bottom layer, followed by careful overlay of 1.0 mL soil slurry to maintain a sharp interface (tubes with interface disruption were discarded). Samples were centrifuged at 4°C and 15,000 g for 40 min with slow acceleration/deceleration to preserve the interface. In a laminar-flow hood, the cell-enriched layer above the Nycodenz was gently aspirated with a 1-mL pipette and transferred to a fresh sterile 2-mL tube. The cell layer was then filtered through a 40-μm sterile cell strainer into another sterile tube to remove residual soil particles for subsequent sorting and cultivation.

Modified M9 medium was prepared using cellobiose, trehalose, D-mannitol, and L-glutamic acid as sole carbon sources (identical to standard M9[31], except for carbon substitution at a final concentration of 10 mM).

Flow-Cytometric Sorting and Phenotypic Measurement

Propidium iodide (PI) staining was used to discriminate viable extractable cells[32]. One milliliter of diluted cell suspension was mixed with 100 μL PI stock, incubated in the dark at room temperature for 20 min, and analyzed on a BECKMAN COULTER CytoFLEX SRT flow sorter. Unstained samples served as controls to exclude signals from soil particles and debris. PI-negative cells were sorted into pre-prepared 96-well plates (each well already containing 200 μL medium). Sorting was performed in “low-speed” mode for accuracy and efficiency. Multiple 96-well plates were used, with initial cell numbers set at 10, 50, 100, or 250 per well. Each carbon-source × cell-number combination comprised five 96-well plates, yielding approximately 7,680 parallel evolving cultures.

Sorted plates were incubated at 30°C. OD_{600} was measured every 24 h using a microplate reader. On day 7, 20 μL of culture was transferred to fresh 96-well plates (180 μL fresh medium per well) and incubation continued to day 14. Biofilm biomass was quantified on days 7 and 14 using the crystal-violet staining assay[33].

DNA Extraction, PCR Amplification, and Sequencing

For soil samples and Nycodenz-extracted microbes, DNA was extracted with the QIAGEN PowerSoil DNA Kit according to the manufacturer’s protocol (Qiagen, Germantown, MD). Representative communities were extracted using the MagBeads FastDNA Kit for Soil (MP Biomedicals, CA, USA) and quantified with

a NanoDrop spectrophotometer (Thermo Scientific, NC2000). The V4 region of the 16S rRNA gene was amplified with primers 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3')[34]. Thermal cycling conditions were: 98°C for 2 min; 30 cycles of 98°C/30 s, 52°C/30 s, 72°C/45 s; final extension at 72°C for 5 min. PCR products were purified with VAHTS™ DNA Clean Beads (Vazyme), quantified fluorometrically using the Quant-iT PicoGreen dsDNA Assay Kit on a BioTek FLx800 microplate reader, pooled in equimolar ratios based on quantification, and libraries prepared with the Illumina TruSeq Nano DNA LT Library Prep Kit. Paired-end sequencing (2 × 250 bp) was performed on the Illumina NovaSeq platform.Representative wells for 16S rRNA gene amplicon sequencing (total N = 384; 4 carbon sources × 4 cell number × 4 phenotypic categories × 6 replicates per combination) were selected from each carbon source × cell number combination. Non-growing cultures (OD_{600} < 0.1) were excluded, and growth and biofilm biomass data were Box-Cox transformed and classified into four phenotypic categories (HGHB, HGLB, LGHB, LGLB) based on median values within each carbon source × cell number combination; selection was stratified to ensure balanced phenotypic coverage.

Amplicon Data Analysis

Raw amplicon data were processed with Vsearch (v2.13.4)[35], primer trimming with Cutadapt (v2.3)[36],quality filtering, merging, chimera removal, and clustering into ASVs using the DADA2 R package, and ASV sequences were taxonomically classified against the pre-trained SILVA (138.2) database[37]. Microbial 16S rRNA data were analyzed for diversity with the phyloseq R package[38]. Visualization employed ggpubr[39], ggVennDiagram[40] and pheatmap[41].Differential abundance testing across initial cell numbers was performed using ANCOM-BC2 (ANCOMBC R package) on genus-level count tables, with FDR-corrected q-values[42].Potential sources were inferred using the FEAST microbial source-tracking method[43]. Microbial association networks were constructed with SpiecEasi, employing Sparse Latent Factor Regression (SLR) together with inverse covariance estimation and sparse regularization to infer putative direct microbial interactions while accounting for unobserved environmental variables[44]; networks were visualized in Gephi[45]. Microbiome genus set enrichment analysis (mgsea) was applied to evaluate environmental-factor regulation of specific genus relative abundances (e.g., *Pseudomonas* and *Klebsiella*). This method extends the Gene Set Enrichment Analysis framework[46], by ranking samples according to target genus abundance, computing running-sum enrichment scores for environmental variables, generating null distributions via permutation, and calculating normalized enrichment scores (NES > 0 indicates positive enrichment, NES < 0 indicates negative enrichment). The implementation is publicly available on GitHub(https://github.com/SyntheticCommunity/mgsea).

Data availability

Raw sequences of 16S rRNA amplicon genes are available in NCBI SRA database (www.ncbi.nlm.nih.gov/sra) under accession number PRJNA1438684. Analytical codes in R, have been uploaded to a GitHub repository (http://github.com/SyntheticCommunity/soil-flowcyto).

Declarations

Competing interests

The authors declare that they have no competing interests.

Ethics declarations

Not applicable. This study focuses on soil-derived microbial communities and does not involve any human participants, human data, or animal subjects.

Fundings

This work was supported by the National Natural Science Foundation of China (32370127, 42225706) and the Fundamental Research Funds for the Central Universities (2662025ZHPY001). Science and Technology Talents and Platform Program (Workstation for Academicians and Experts) (202405AF140079).Innovative Research Group Project of Hubei Provincial Natural Science Foundation(2023AFA009)

Author Contribution

C.G. and P.C. conceived and designed the study. Z. Z. and C. G. performed the experiments and analyzed the data. C. G., P. C., and Z. Z. wrote the manuscript with input from all authors.

Acknowledgement

The computations in this paper were run on the bioinformatics computing platform of the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University. We would like to thank the National Key Laboratory of Agricultural Microbiology Core Facility for assistance in Flow cytometry. We sincerely thank Laurent Philippot for carefully reading the manuscript and providing valuable comments.

Data Availability

Raw sequences of 16S rRNA amplicon genes are available in NCBI SRA database (www.ncbi.nlm.nih.gov/sra) under accession number PRJNA1438684. Analytical codes in R, have been uploaded to a GitHub repository (http://github.com/SyntheticCommunity/soil-flowcyto).

References

1. Nunan N, Wu K, Young IM, Crawford JW, Ritz K. Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. FEMS Microbiol Ecol. 2003;44:203–15. https://doi.org/10.1016/S0168-6496(03)00027-8
2. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences; 1998;95:6578–83. https://doi.org/10.1073/pnas.95.12.6578
3. Fierer N. Embracing the unknown: Disentangling the complexities of the soil microbiome. Nat Rev Microbiol. Nature Publishing Group; 2017;15:579–90. https://doi.org/10.1038/nrmicro.2017.87
4. Tecon R, Or D. Biophysical processes supporting the diversity of microbial life in soil. FEMS Microbiol Rev. 2017;41:599–623. https://doi.org/10.1093/femsre/fux039
5. Wilpiszeski RL, Aufrecht JA, Retterer ST, Sullivan MB, Graham DE, Pierce EM, et al. Soil aggregate microbial communities: Towards understanding microbiome interactions at biologically relevant scales. Applied and Environmental Microbiology. American Society for Microbiology; 2019;85:e00324-19. https://doi.org/10.1128/AEM.00324-19
6. Young IM, Crawford JW. Interactions and self-organization in the soil-microbe complex. Science. American Association for the Advancement of Science; 2004;304:1634–7. https://doi.org/10.1126/science.1097394
7. Jiao S, Chen W, Wang J, Du N, Li Q, Wei G. Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. Microbiome. 2018;6:146. https://doi.org/10.1186/s40168-018-0526-0
8. Raynaud X, Nunan N. Spatial Ecology of Bacteria at the Microscale in Soil. PLOS ONE. Public Library of Science; 2014;9:e87217. https://doi.org/10.1371/journal.pone.0087217
9. Wu Y, Cai P, Jing X, Niu X, Ji D, Ashry NM, et al. Soil biofilm formation enhances microbial community diversity and metabolic activity. Environ Int. 2019;132:105116. https://doi.org/10.1016/j.envint.2019.105116
10. Zhang B, Hu X, Zhao D, Wang Y, Qu J, Tao Y, et al. Harnessing microbial biofilms in soil ecosystems: Enhancing nutrient cycling, stress resilience, and sustainable agriculture. Journal of Environmental Management. 2024;370:122973. https://doi.org/10.1016/j.jenvman.2024.122973
11. Cordero OX, Datta MS. Microbial interactions and community assembly at microscales. Current Opinion in Microbiology. 2016;31:227–34. https://doi.org/10.1016/j.mib.2016.03.015
12. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and processes of microbial community assembly. Microbiology and Molecular Biology Reviews. American Society for Microbiology; 2013;77:342–56. https://doi.org/10.1128/mmbr.00051-12
13. Romdhane S, Huet S, Spor A, Bru D, Breuil M-C, Philippot L. Manipulating the physical distance between cells during soil colonization reveals the importance of biotic interactions in microbial community assembly. Environmental Microbiome. 2024;19:18. https://doi.org/10.1186/s40793-024-00559-4

14. Pascual-García A, Rivett DW, Jones ML, Bell T. Replicating community dynamics reveals how initial composition shapes the functional outcomes of bacterial communities. Nat Commun. Nature Publishing Group; 2025;16:3002. https://doi.org/10.1038/s41467-025-57591-2
15. Nunan N, Leloup J, Ruamps LS, Pouteau V, Chenu C. Effects of habitat constraints on soil microbial community function. Sci Rep. Nature Publishing Group; 2017;7:4280. https://doi.org/10.1038/s41598-017-04485-z
16. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, et al. Quantifying community assembly processes and identifying features that impose them. ISME J. Nature Publishing Group; 2013;7:2069–79. https://doi.org/10.1038/ismej.2013.93
17. Zhou J, Ning D. Stochastic community assembly: Does it matter in microbial ecology? Microbiol Mol Biol Rev. American Society for Microbiology; 2017;81:10.1128/mmbr.00002-17. https://doi.org/10.1128/mmbr.00002-17
18. Wang Y, Zhu K, Chen X, Wei K, Wu R, Wang G. Stochastic and deterministic assembly processes of bacterial communities in different soil aggregates. Applied Soil Ecology. 2024;193:105153. https://doi.org/10.1016/j.apsoil.2023.105153
19. Kaul RB, Kramer AM, Dobbs FC, Drake JM. Experimental demonstration of an allee effect in microbial populations. Biol Lett. 2016;12:20160070. https://doi.org/10.1098/rsbl.2016.0070
20. Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. ISME J. Nature Publishing Group; 2012;6:1653–64. https://doi.org/10.1038/ismej.2012.22
21. Bartelme RP, Custer JM, Dupont CL, Espinoza JL, Torralba M, Khalili B, et al. Influence of substrate concentration on the culturability of heterotrophic soil microbes isolated by high-throughput dilution-to-extinction cultivation. mSphere. American Society for Microbiology; 2020;5:10.1128/msphere.00024 – 20. https://doi.org/10.1128/msphere.00024-20
22. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in microbial ecology: Building predictive understanding of community function and dynamics. ISME J. Nature Publishing Group; 2016;10:2557–68. https://doi.org/10.1038/ismej.2016.45
23. Philippot L, Chenu C, Kappler A, Rillig MC, Fierer N. The interplay between microbial communities and soil properties. Nat Rev Microbiol. Nature Publishing Group; 2024;22:226–39. https://doi.org/10.1038/s41579-023-00980-5
24. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Science. American Association for the Advancement of Science; 2018;361:469–74. https://doi.org/10.1126/science.aat1168
25. Hu J, Amor DR, Barbier M, Bunin G, Gore J. Emergent phases of ecological diversity and dynamics mapped in microcosms. Science. American Association for the Advancement of Science; 2022;378:85–9. https://doi.org/10.1126/science.abm7841
26. Kümmerli R. Iron acquisition strategies in pseudomonads: Mechanisms, ecology, and evolution. Biometals. 2023;36:777–97. https://doi.org/10.1007/s10534-022-00480-8

27. Chen L, Wilksch JJ, Liu H, Zhang X, Torres VVL, Bi W, et al. Investigation of LuxS-mediated quorum sensing in klebsiella pneumoniae. J Med Microbiol. 2020;69:402–13. https://doi.org/10.1099/jmm.0.001148
28. Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB, et al. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol. Nature Publishing Group; 2010;8:523–9. https://doi.org/10.1038/nrmicro2367
29. Bäcker M, Doekes HM, Garza DR, Meijer J, van Vliet S, Allen RJ, et al. Spatial structure: Shaping the ecology and evolution of microbial communities. Fems Microbiol Rev. 2026;50:fuaf067. https://doi.org/10.1093/femsre/fuaf067
30. Ouyang Y, Chen D, Fu Y, Shi W, Provin T, Han A, et al. Direct cell extraction from fresh and stored soil samples: Impact on microbial viability and community compositions. Soil Biology and Biochemistry. 2021;155:108178. https://doi.org/10.1016/j.soilbio.2021.108178
31. M9 minimal medium (standard). Cold Spring Harb Protoc. Cold Spring Harbor Laboratory Press; 2010;2010:pdb.rec12295. https://doi.org/10.1101/pdb.rec12295
32. Hammes F, Egli T. Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. Anal Bioanal Chem. 2010;397:1083–95. https://doi.org/10.1007/s00216-010-3646-3
33. O'Toole GA. Microtiter Dish Biofilm Formation Assay. Journal of Visualized Experiments (JoVE). 2011;e2437. https://doi.org/10.3791/2437
34. Walters W, Hyde ER, Berg-Lyons D, Ackermann G, Humphrey G, Parada A, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems. American Society for Microbiology; 2015;1:10.1128/msystems.00009-15. https://doi.org/10.1128/msystems.00009-15
35. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. PeerJ Inc.; 2016;4:e2584. https://doi.org/10.7717/peerj.2584
36. Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. 2011;17(1):pp–10.
37. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. Nature Publishing Group; 2016;13:581–3. https://doi.org/10.1038/nmeth.3869
38. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE. Public Library of Science; 2013;8:e61217. https://doi.org/10.1371/journal.pone.0061217
39. Kassambara A. ggpubr: 'ggplot2' based publication ready plots [Internet]. 2026. https://rpkgs.datanovia.com/ggpubr/
40. Gao C, Chen C, Akyol T, Dusa A, Yu G, Cao B, et al. ggVennDiagram: Intuitive Venn diagram software extended. iMeta. 2024;3:e177. https://doi.org/10.1002/imt2.177
41. Kolde R. pheatmap: Pretty heatmaps [Internet]. 2025. https://github.com/raivokolde/pheatmap

42. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. Nat Methods. Nature Publishing Group; 2024;21:83–91. https://doi.org/10.1038/s41592-023-02092-7
43. Shenhav L, Thompson M, Joseph TA, Briscoe L, Furman O, Bogumil D, et al. FEAST: Fast expectation-maximization for microbial source tracking. Nat Methods. Nature Publishing Group; 2019;16:627–32. https://doi.org/10.1038/s41592-019-0431-x
44. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLOS Computational Biology. Public Library of Science; 2015;11:e1004226. https://doi.org/10.1371/journal.pcbi.1004226
45. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. Proceedings of the International AAAI Conference on Web and Social Media. 2009;3:361–2. https://doi.org/10.1609/icwsm.v3i1.13937
46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences; 2005;102:15545–50. https://doi.org/10.1073/pnas.0506580102

Figures

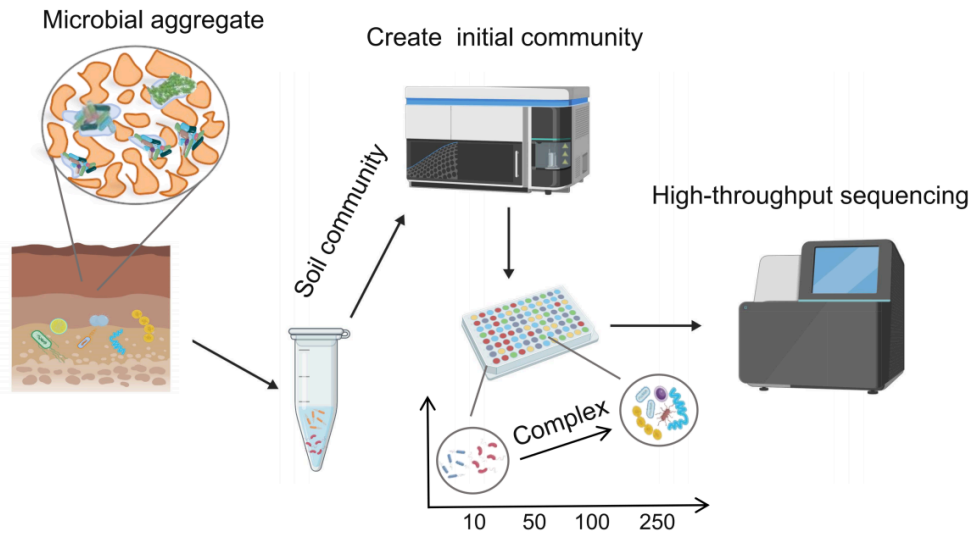


Figure 1

A flow cytometry–based sorting platform was employed to investigate the assembly and functional dynamics of soil biofilm communities under varying initial cell numbers. Manipulation of initial cell numbers imposed diverse degrees of historical contingency upon community development. Soil microbial consortia, extracted from surface soil, were cultivated with distinct carbon sources.

Sequencing of representative communities enabled assessment of the influence of initial cell number on biofilm community assembly and putative associations

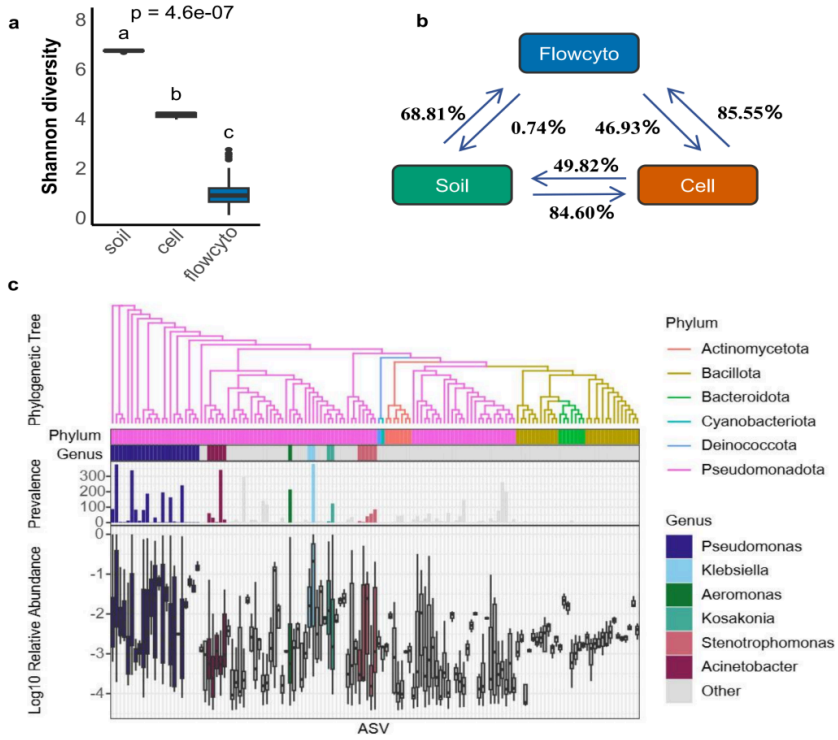


Figure 2

Microbial community diversity, source tracking and structural analysis. (a) Community diversity under different treatments. Box plots show Shannon diversity indices of microbial communities from bulk soil,

cell extraction and flowcyto fractions. Different letters indicate statistically significant differences among groups ($p = 4.6 \times 10^{-7}$). (b) Community source tracking analysis between culture groups, illustrating proportional contributions of microbial communities among bulk soil, cell extraction and flowcyto fractions. Percentages on arrows represent source-tracking contribution rates. (c) Overall community structure in the sorted culture system. Top: phylogenetic tree of ASVs with branches colored by phylum (see legend). Middle: prevalence bar plot at phylum and genus levels. Bottom: boxplot of log10-transformed relative abundance for each ASV, colored by genus (see legend).

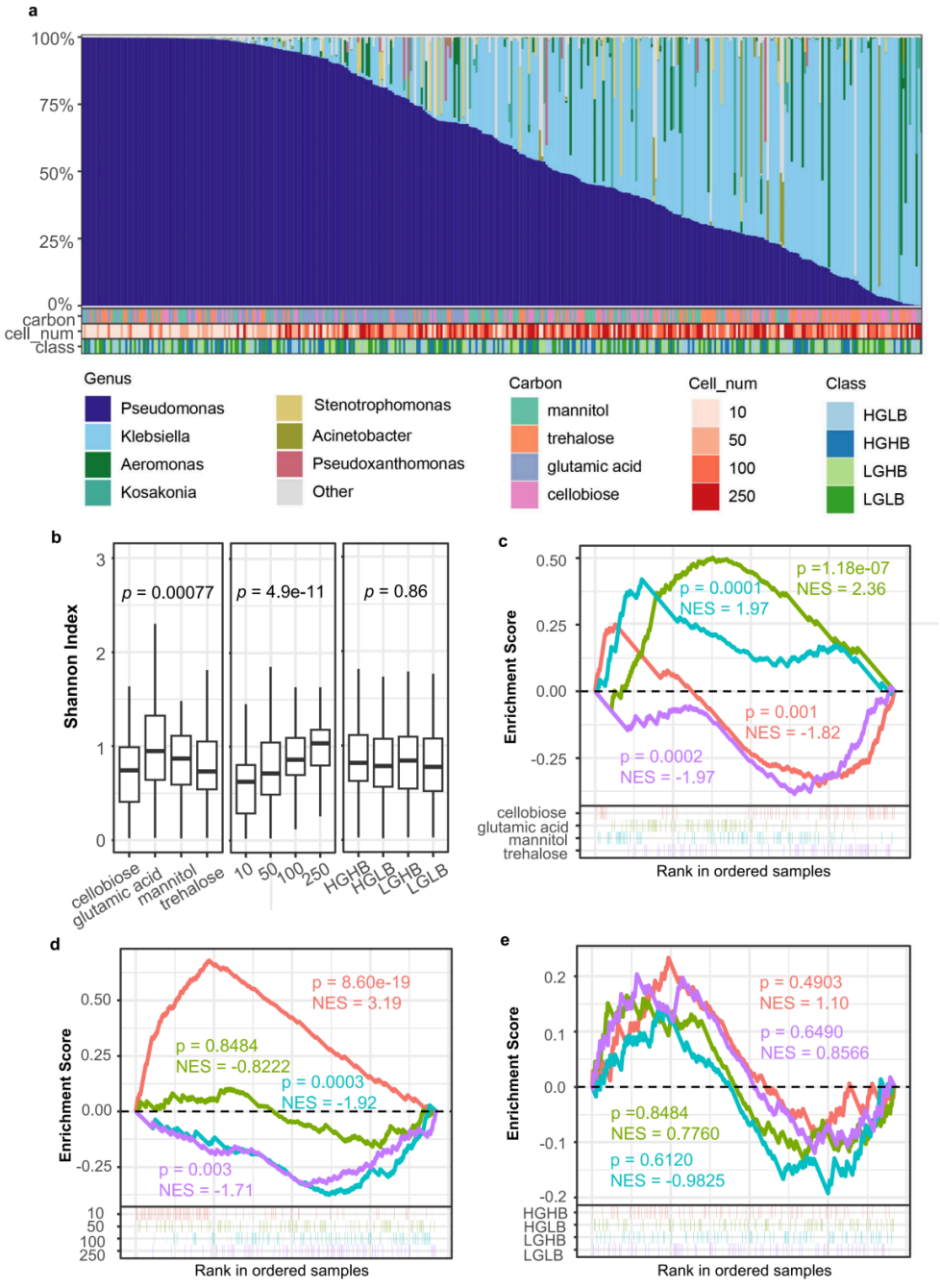


Figure 3

Convergence of microbial communities and the effect of cell number on community structure and diversity.(a) Overall community structure and sample grouping. The upper stacked bar plot shows genus-level relative abundance across samples, with annotation bars below indicating genus identity, carbon source, cell number and biofilm class. (b) α-diversity (Shannon index) across groupings. Box plots display Shannon index values (y-axis) for different carbon sources, cell numbers and biofilm classes (x-axis), with inter-group p-values indicated ($p < 0.05$ considered significant). (c–e) Microbiome genus set enrichment analysis of Pseudomonas (GSEA: gene-set enrichment analysis adapted for species). Running enrichment score plots with x-axis = samples ranked by decreasing relative abundance (Rank in ordered samples) and y-axis = enrichment score (Enrichment Score). Colored lines represent treatment groups; each is labelled with p-value (statistical significance of enrichment) and NES (Normalized Enrichment Score; NES > 0 = positive enrichment, NES < 0 = negative enrichment, |NES| reflects strength). (c) Carbon source groups; (d) Cell number groups (10, 50, 100, 250); (e) Biofilm class groups (HGLB, HGHB, LGLB, LGHB).

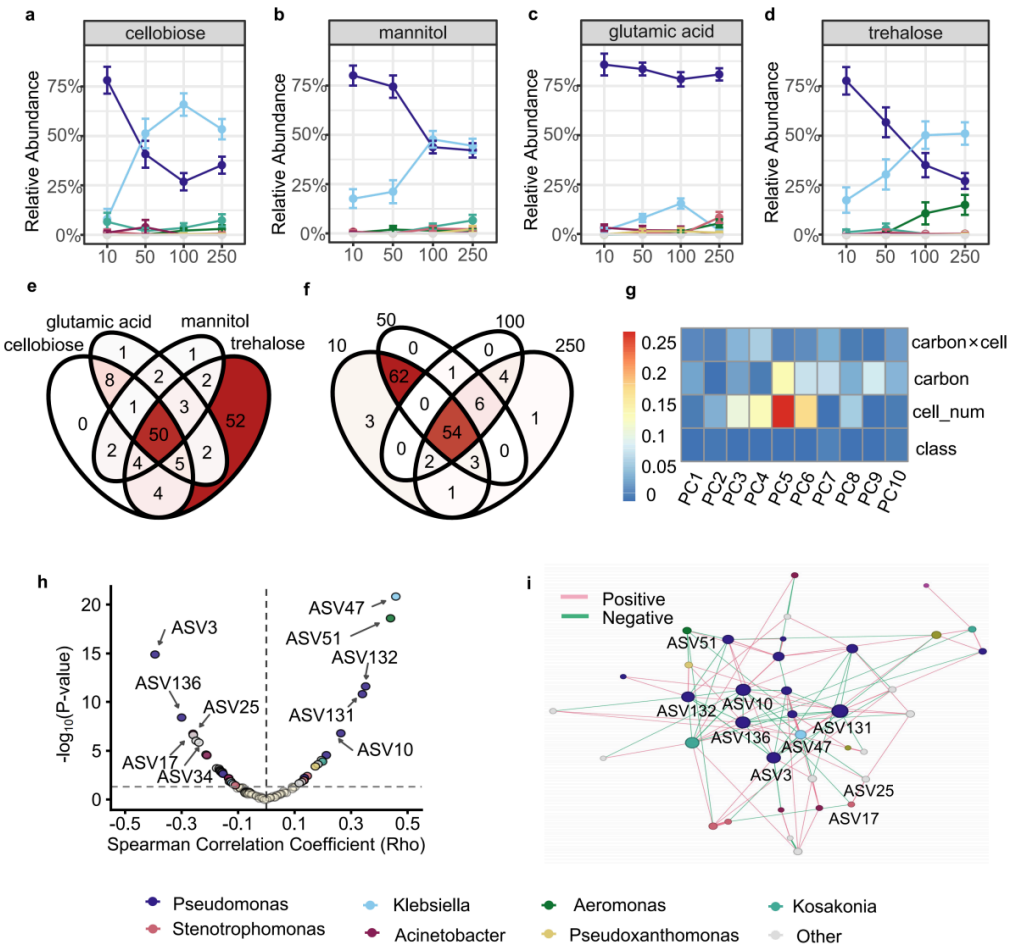


Figure 4

Impact of initial cell number on community dynamics. (a-d) Relative abundance trends of the top 10 taxa across different cell numbers for each carbon source. Line plots show mean relative abundance (± SEM) of the top 10 taxa at cell inoculation levels of 10, 50, 100 and 250 for cellobiose, glutamic acid, mannitol and trehalose (top to bottom). (e-f) ASV sharing among different carbon sources and cell-number groups. Venn diagrams illustrate the number of shared and unique ASVs across the four carbon

sources (left) and the four cell numbers (right). (g) SLR analysis of all samples, showing cell number as the primary latent variable influencing the community network. Heatmap displays degree of contribution between latent components (PC1–PC10) and the factors carbon source, cell number and biofilm class. (h) The x-axis shows the Spearman correlation coefficient (Rho); positive values indicate ASVs that significantly increase with higher cell numbers (high cell number preference, red), while negative values indicate ASVs that significantly decrease with higher cell numbers (low cell number preference, blue). Gray dots represent non-significant ASVs ($P \geq 0.05$). The y-axis shows $-\log_{10}$(P-value). Representative ASVs are labeled to highlight cell-number-dependent enrichment patterns. (i) Microbial association network constructed by SLR. Node size represents degree (number of connections), node color corresponds to genus (see legend), edge thickness indicates association confidence/strength (SLR conditional dependence or statistical association strength), and edge color denotes association direction (red: positive; green: negative).

Supplementary Files

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

- SupplementaryInformation.pdf