**Supplementary Information**

# Initial Cell Number Regulates Microscale Community Assembly and Structure

Zhaoshuo Zhang[1,2,3,#], Chun-Hui Gao[1,2,3,#,*], Fangkui Wang[1], Jiliang Hu[4,5], Huan He[3], Yichao Wu[1,2,3,] Chenchen Qu[2,3], Ke Dai[3], Qiaoyun Huang[1,2,3], Peng Cai[1,2,3,*]

[1]National Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China.
[2]Key Laboratory of Farmland Remediation in Central China, Ministry of Agriculture and Rural Affairs, Wuhan 430000, China.
[3]College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China.
[4]Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen, Guangdong 518000, China.
[5]Shenzhen X-Institute, Tsinghua Shenzhen International Graduate School, Shenzhen, Guangdong 518000, China.

#These authors contributed equally to this work.
*Correspondence should be addressed to C-H. Gao, E-mail: gaoch@mail.hzau.edu.cn, or Peng Cai, E-mail: cp@mail.hzau.edu.cn

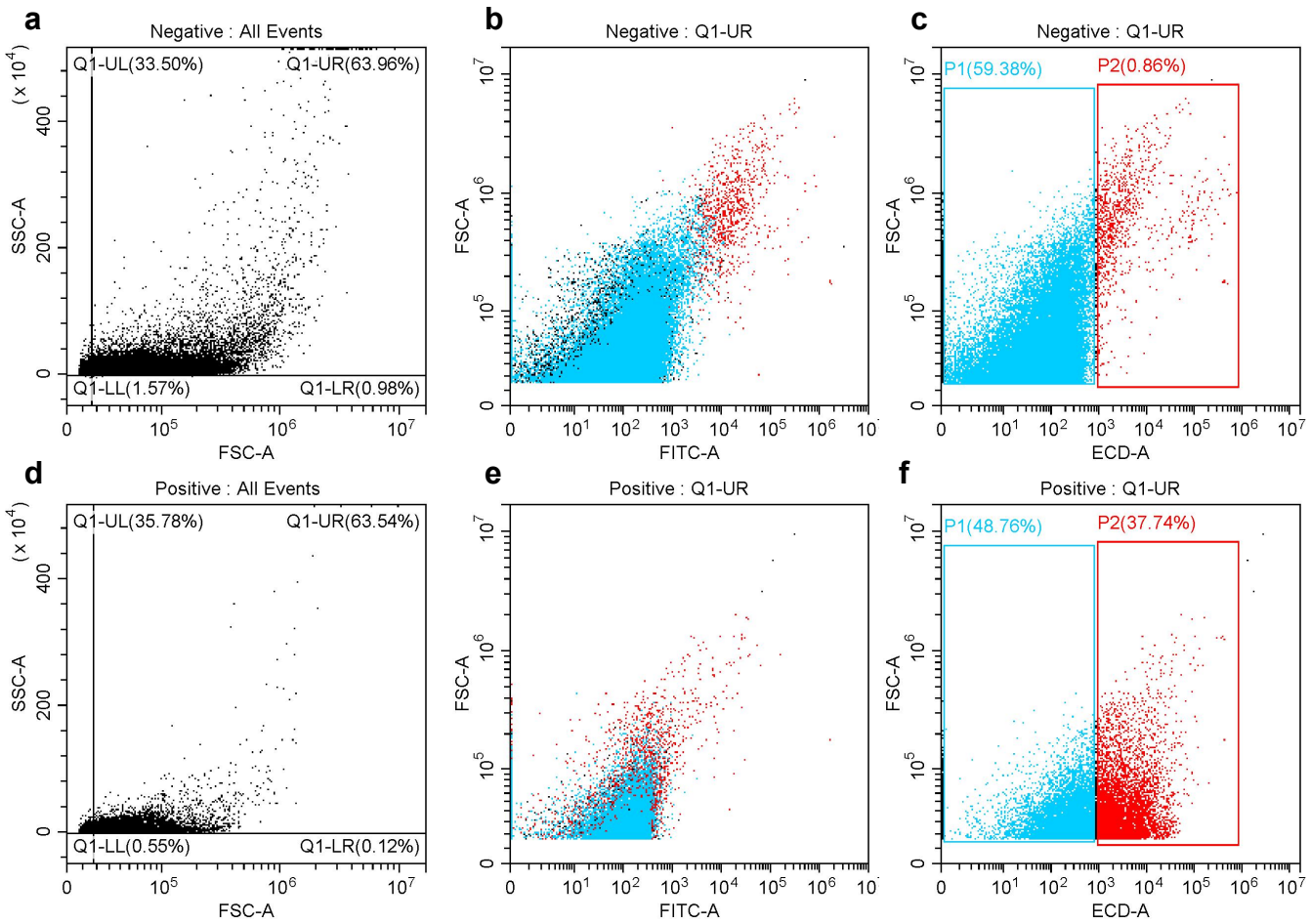


**Fig. S1. Flow cytometry–based construction of initial inocula with varying cell numbers.** (a–c) represent negative controls, whereas (d–f) depict positively sorted samples for establishing differential initial cell numbers. In (a) and (b), gating strategies were employed to exclude small and complex particles. (d) and (e) illustrate detection of green fluorescence via the FITC channel, with no substantial enhancement observed in either group. (c) and (f) demonstrate detection of propidium iodide (PI)-stained dead cells using the ECD channel, wherein marked fluorescence intensification is evident in the P2 gate of positive samples, and unstained live cells in the P1 gate were isolated through sorting.

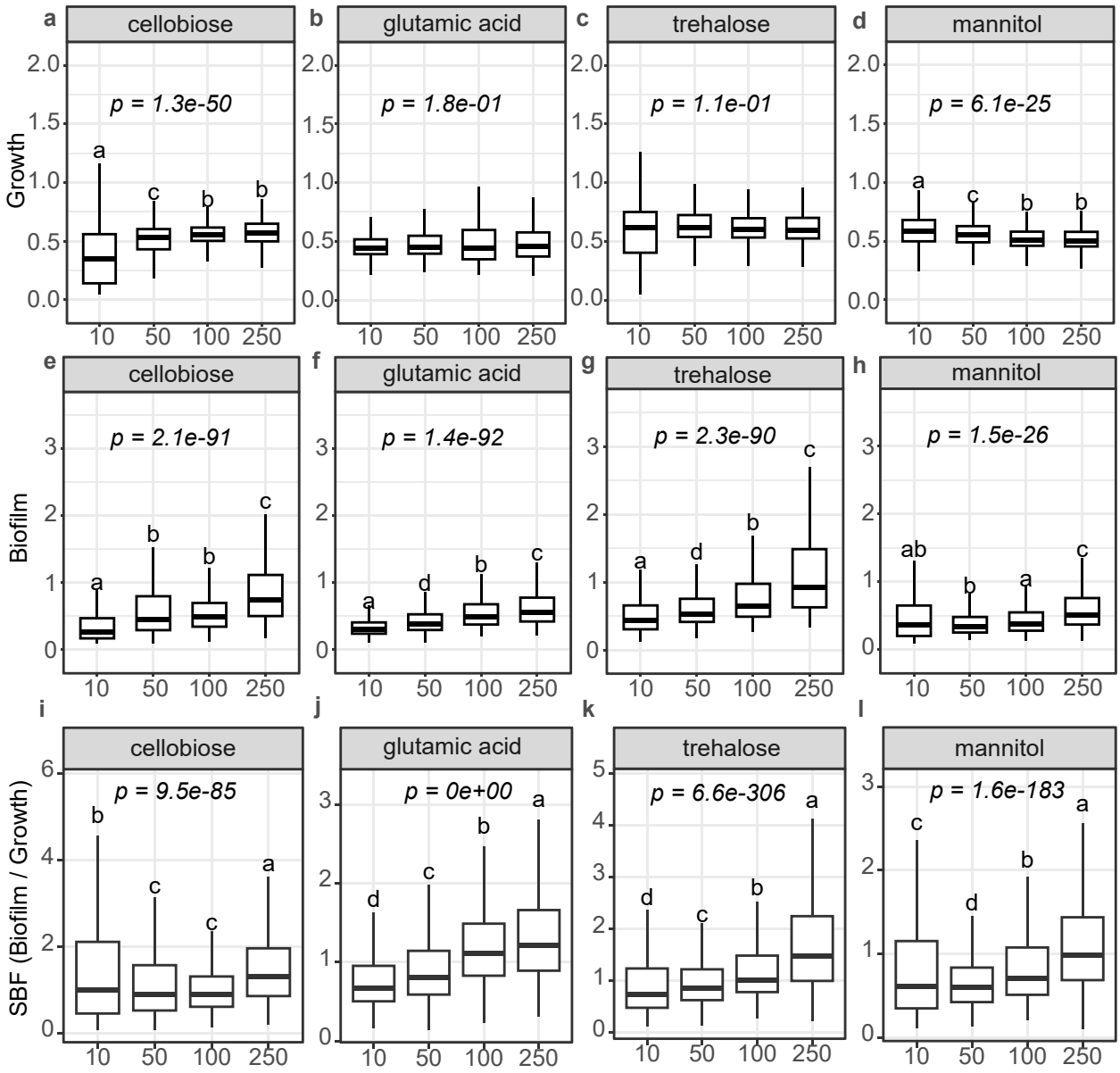


**Fig. S2. Growth (a–d), biofilm formation (e–h), and specific biofilm formation (i–l) of soil-derived microbial subpopulations under various carbon sources, as specified in the subplot titles.** The x-axis denotes the initial cell number in all panels. Panels (a–d) show biomass measured as OD600; panels (e–h) show biofilm biomass quantified via crystal violet staining; panels (i–l) show specific biofilm formation (SBF = biofilm biomass normalized to total growth, CV/OD600), with extreme outliers removed for visualization. Statistical significance was assessed using the Kruskal–Wallis test, with post-hoc pairwise comparisons conducted via Dunn's test (a–h) or the kruskal function (agricolae package) with Bonferroni correction (i–l); distinct letters signify significant differences among cell number groups ($P < 0.05$).

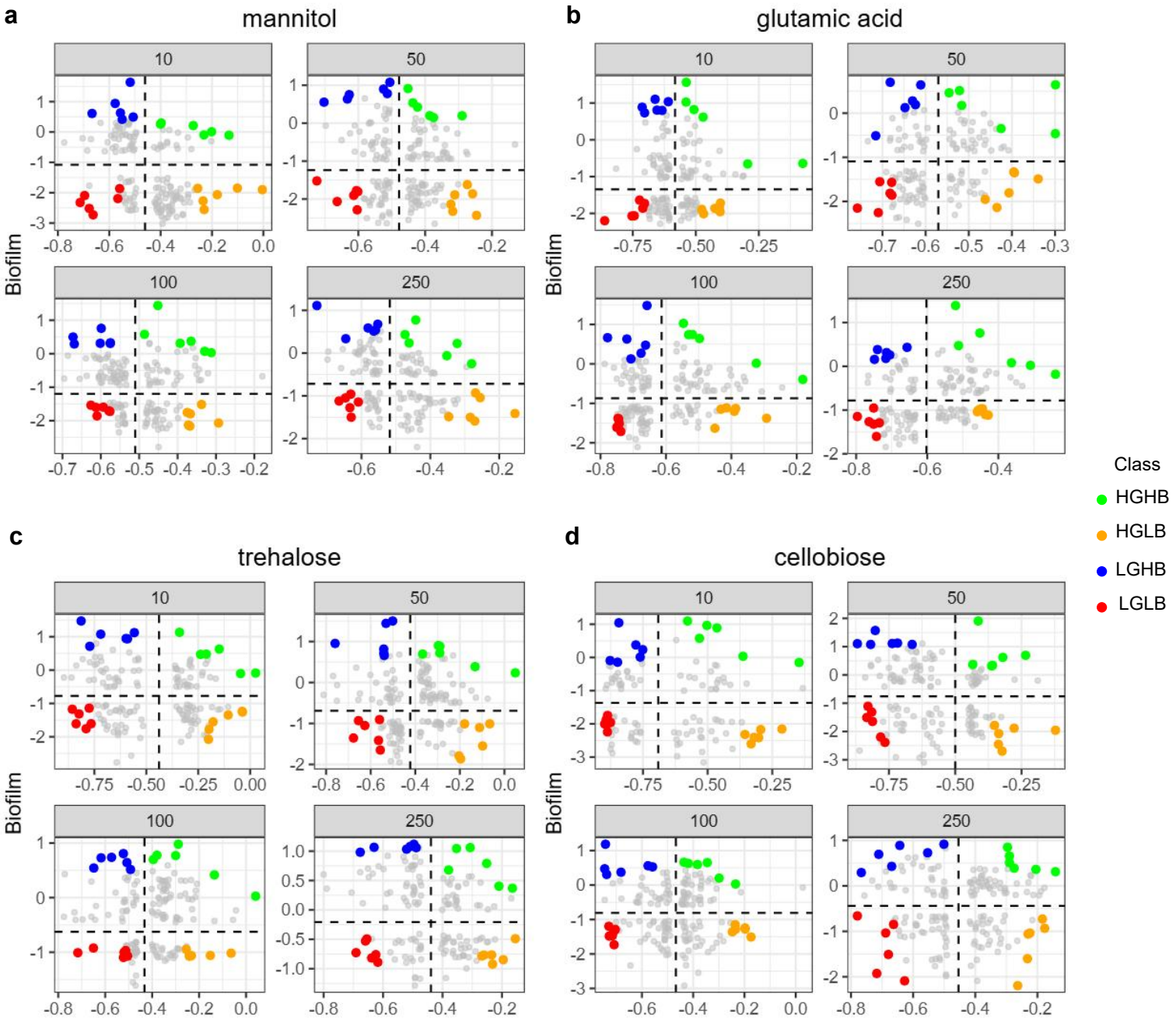


**Fig. S3. Representative distribution patterns and sample grouping across microbial communities.** (a–d) depict overall community growth dynamics under varying carbon sources, stratified by initial cell number. The x- and y-axes represent Box–Cox-transformed growth performance and biofilm biomass, respectively. Gray points denote non-selected community data, whereas representative samples within each group are highlighted in accordance with the legend (HGHB: high growth-high biofilm; HGLB: high growth-low biofilm; LGHB: low growth-high biofilm; LGLB: low growth-low biofilm).

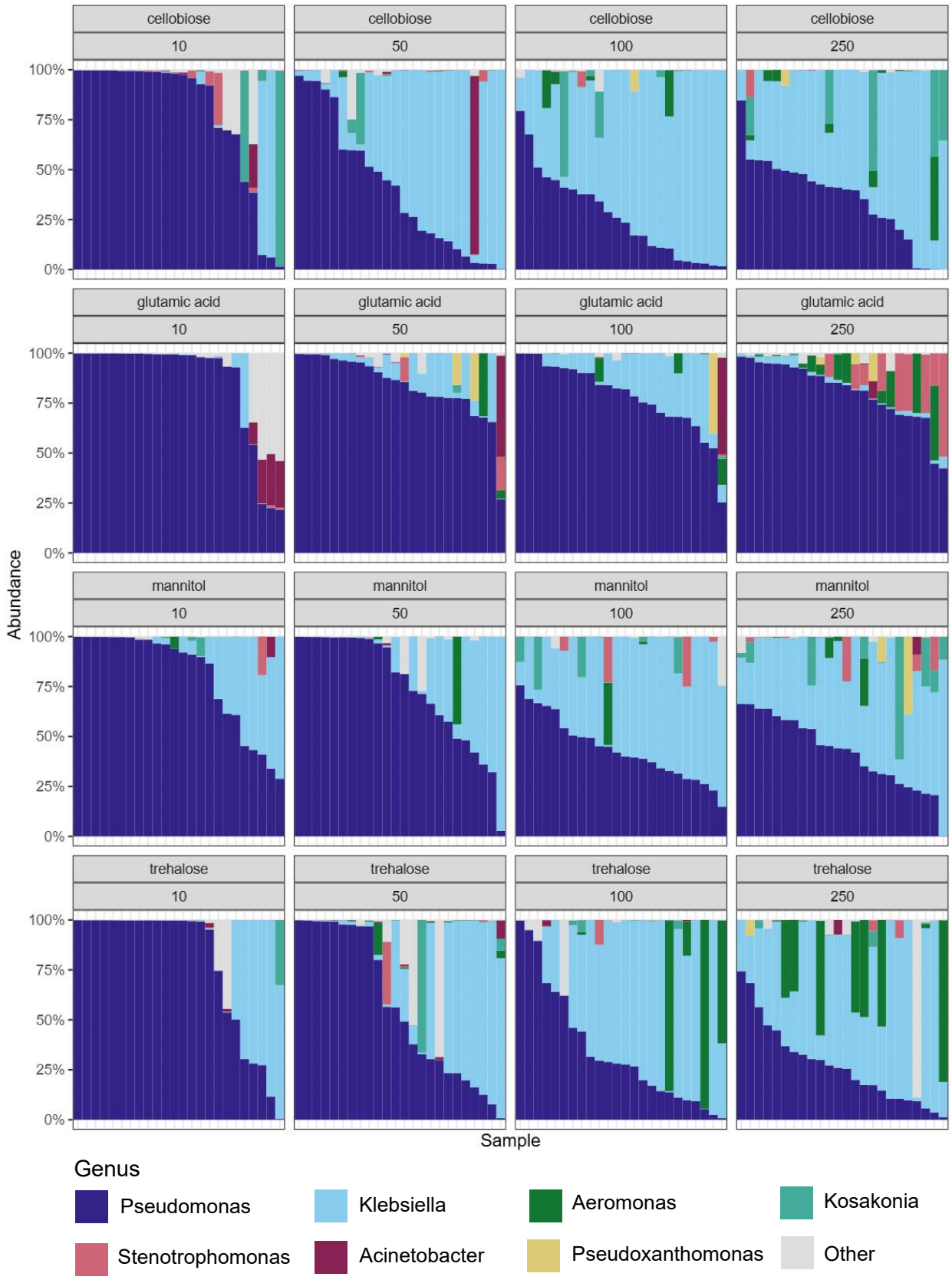


**Fig. S4. Microbial community composition of all samples under different carbon sources and cell numbers.**Stacked bar plots showing genus-level relative abundance (0–100%) across samples for four carbon sources and four cell inoculation numbers . Each panel represents one carbon source–cell number combination; different colours indicate different microbial genera.

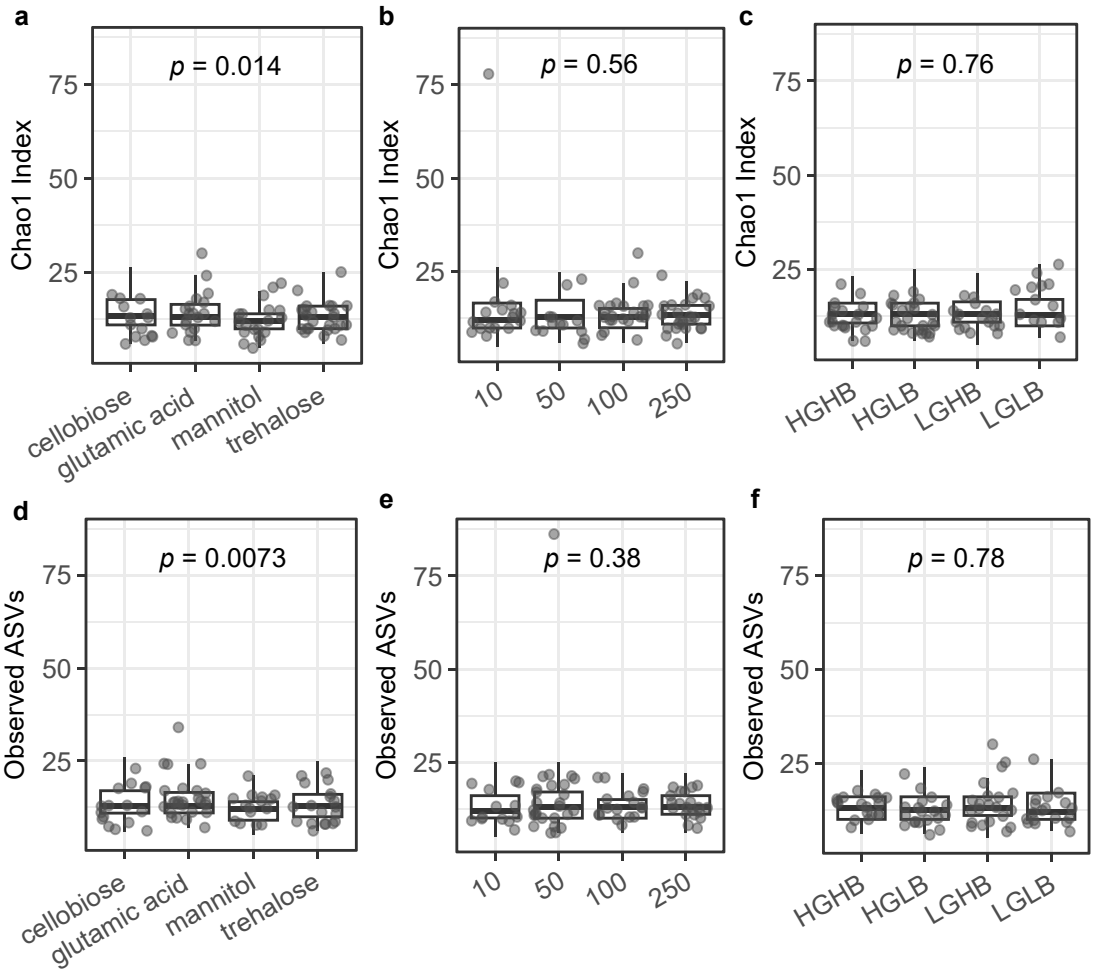


**Fig S5. Alpha-diversity richness of cultured microbial communities**. (a–c) Chao1 richness index and (d–f) observed ASV number in response to (a, d) different carbon sources , (b, e) initial cell numbers , and (c, f) phenotypic groups . P-values above each panel indicate overall significance among groups (Kruskal–Wallis test)

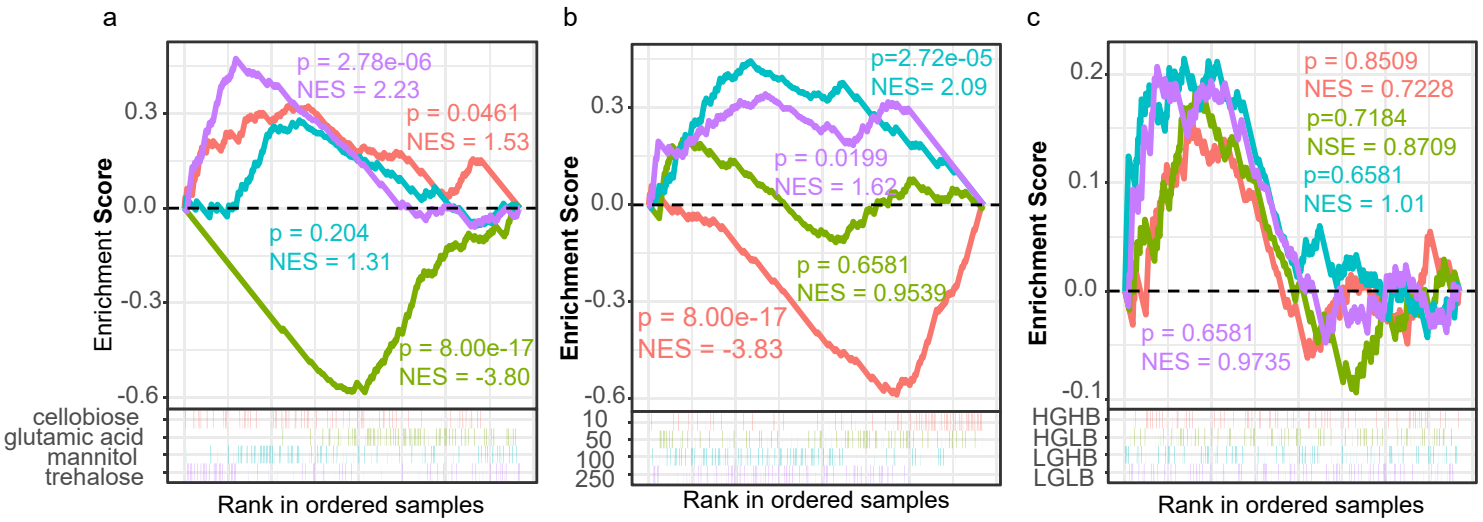


**Fig. S6 Microbiome genus set enrichment analysis of Klebsiella (GSEA: gene-set enrichment analysis adapted for species).** (a–c) Running enrichment score plots with x-axis = samples ranked by decreasing relative abundance (Rank in ordered samples) and y-axis = enrichment score (Enrichment Score). Coloured 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)(a) Carbon source groups; (b) Cell number groups ; (c) Biofilm class groups.

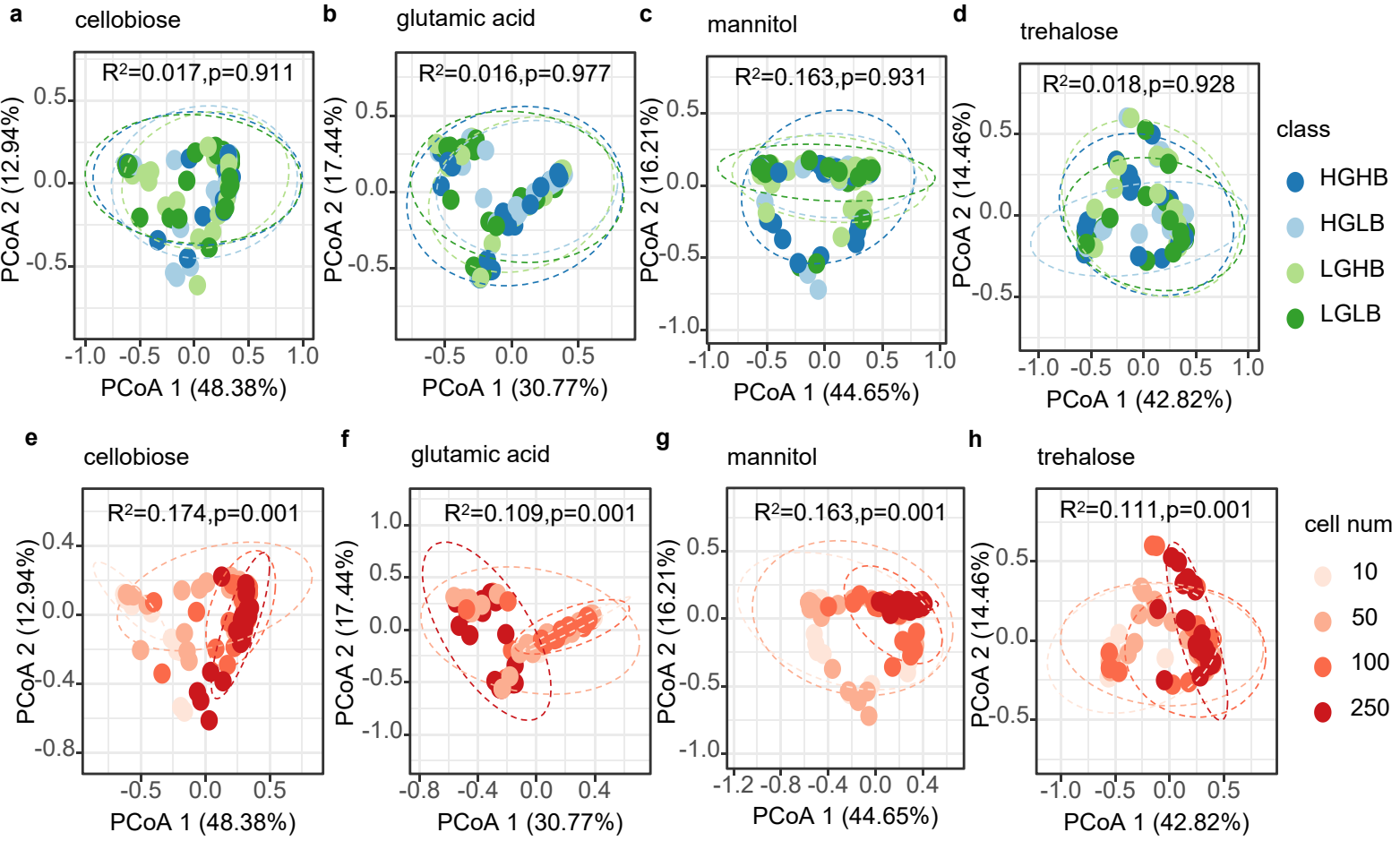


**Fig. S7 β-diversity of microbial communities under different carbon sources, grouped by biofilm class and cell number.**(a–d) Principal coordinate analysis (PCoA) plots showing community β-diversity for different biofilm groupings under four carbon sources. Points are coloured by biofilm class, with 95% confidence ellipses. $R^2$ and p-values are indicated in each panel.(e–h) Principal coordinate analysis plots showing community β-diversity for different cell-number groupings under the same four carbon sources. Points are coloured by cell number , with 95% confidence ellipses. $R^2$ and p-values are indicated in each panel.

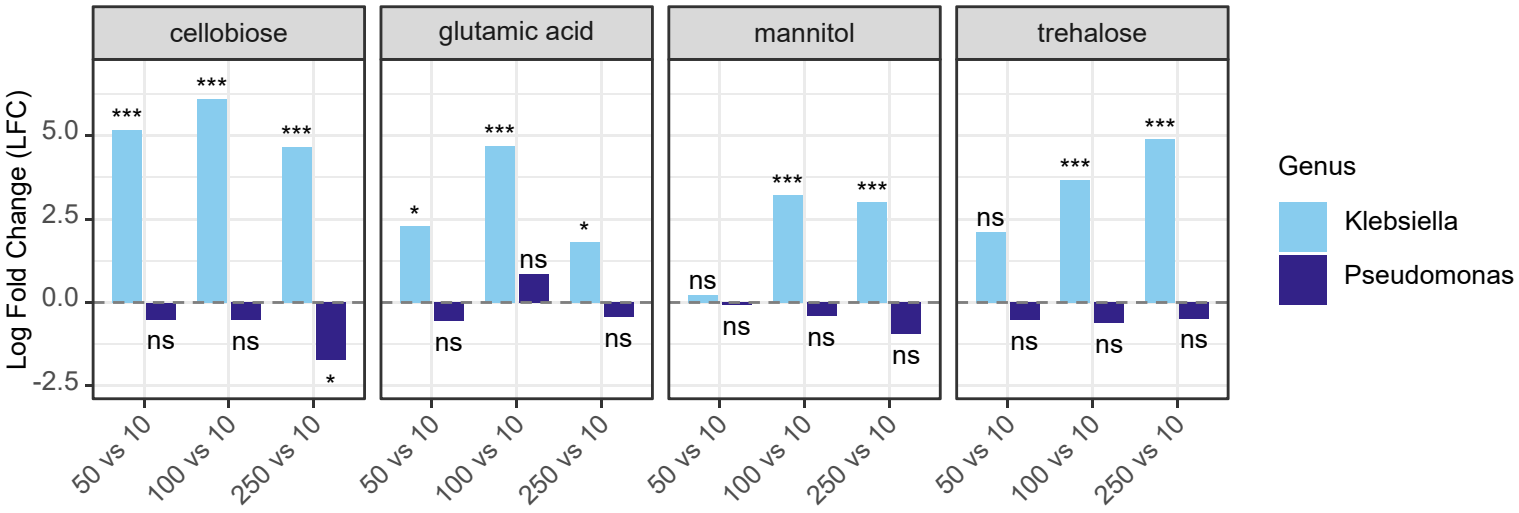


**Fig. S8. ANCOM-BC2 analysis of density-dependent true abundance shifts of Klebsiella and Pseudomonas.**Bars represent log fold change (LFC) relative to the 10-cell group. Light blue bars indicate Klebsiella and dark purple bars indicate Pseudomonas. ANCOM-BC2 was performed independently for each carbon source using cell number as the fixed factor (FDR-corrected q-values). Asterisks denote statistical significance (* P < 0.05, ** P < 0.01, *** P < 0.001, ns = not significant).

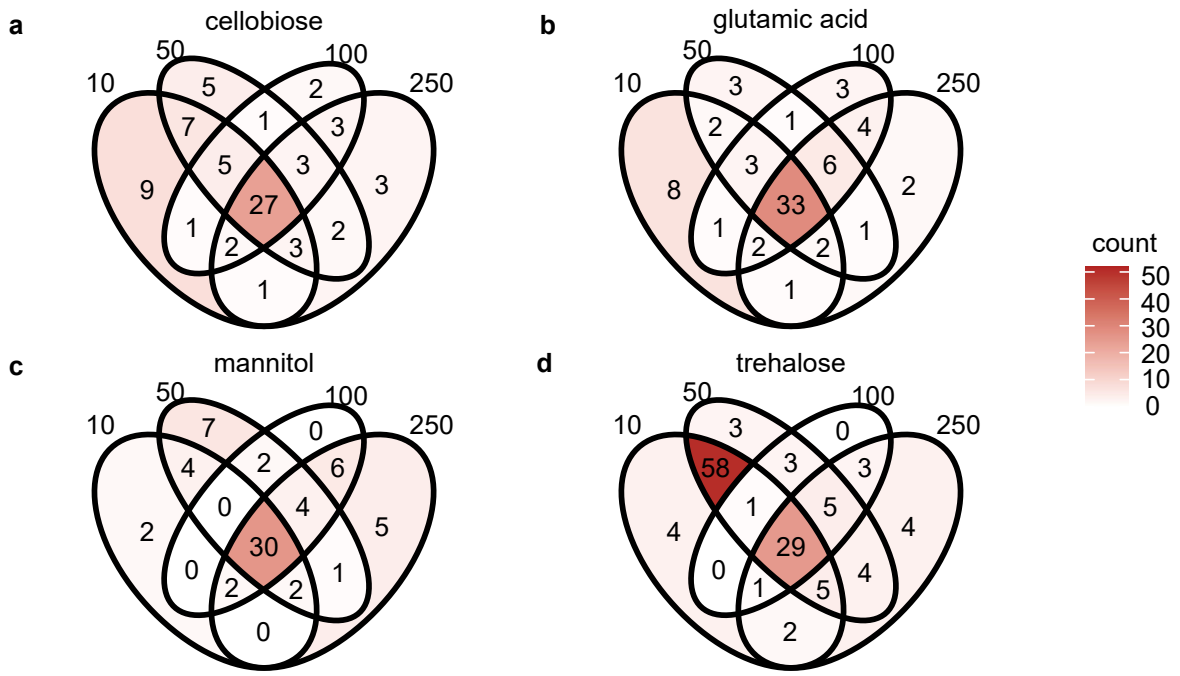


**Fig. S9. ASV sharing across varying initial cell numbers under various carbon sources.**(a-d) Venn diagrams illustrating the overlap of ASVs among communities subjected to diverse initial cell numbers. Numerals in each sector denote the count of ASVs shared or unique to respective inoculum groups, with color intensity corresponding to the relative abundance of ASVs.

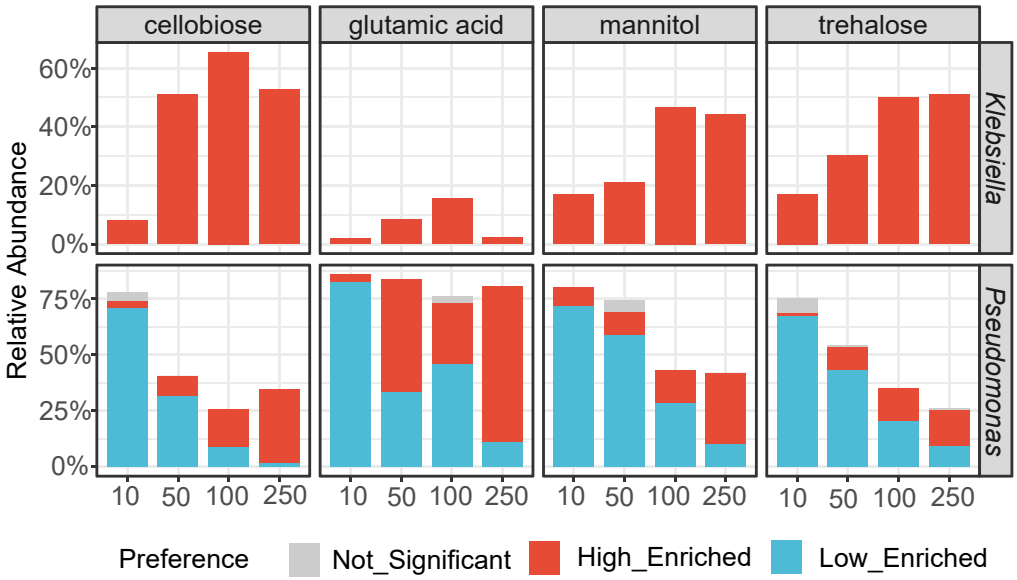


**Fig. S10. Relative abundance composition of the two dominant genera and intra-genus differentiation in cell number preference.**
Stacked bar plots show the relative abundances of Klebsiella and Pseudomonas across four carbon sources and different initial cell numbers . Blue bars represent low cell number-preferring ASVs, red bars represent high cell number-preferring ASVs, and grey bars represent ASVs with no significant preference.