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.

From assembly to yield: kingdom-specific rules, keystone networks and a Bayesian structural equation framework for the maize rhizosphere microbiome

Yixiang Ban
Guofen Zhao
Shilin Feng
Xiaodong Han
Qiushi Lü
Yanhua You
Tiantian Huang
Guihua Wang

wangguiuaz@sohu.com

Inner Mongolia Agriculture University: Inner Mongolia Agricultural University https://orcid.org/0009-0009-0622-1457

Research Article

Keywords: Bayesian structural equation modeling, co-occurrence network, maize (Zea mays L.), rhizosphere microbiome, yield-associated microbial signatures

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11207504/v1

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

From assembly to yield: kingdom-specific rules, keystone networks and a Bayesian structural equation framework for the maize rhizosphere microbiome

Yixiang Ban[1]†,Guofen Zhao[1]†, Shilin Feng[1], Xiaodong Han[1],Qiushi Lü[2], Yanhua You[2], Tiantian Huang[2],*Guihua Wang[1]*

[1] College of Life Sciences, Inner Mongolia Agricultural University, Hohhot 010018, China; [2] Xing'an League Institute of Agricultural and Animal Husbandry Sciences, Ulanhot 137400, China

* Corresponding authors. E-mails: wangguiuaz@sohu.com (G.W.); 15024851009@163.com (T.H.)

† These authors contributed equally to this work.

Running title: From assembly to yield: maize rhizosphere framework

Abstract

Background and aims. The rhizosphere microbiome is an extension of the crop genome, yet the rules governing its assembly and its statistical links to yield remain unresolved. We asked whether bacteria and fungi follow kingdom-specific assembly rules across genotypes and stages (RQ1), which taxa form network backbones (RQ2), and through which statistical pathways microbial features connect to measured grain yield (RQ3).

Methods. Bacterial and fungal communities of seven maize cultivars were profiled at seedling (V6) and tasseling (VT) stages, integrating diversity analyses, robustness-validated co-occurrence networks, FAPROTAX/FUNGuild annotation, BSEM, and machine-learning signature identification; grain yield was measured and no forecasting was attempted.

Results. Bacterial assembly was dominated by developmental stage (ANOSIM R = 0.90), shifting from copiotrophic Proteobacteria toward oligotrophic lineages, whereas fungi retained stronger genotype structure at tasseling (R up to 0.74). Phenotypic and community distances were decoupled at the genotype level. Proteobacteria and Ascomycota formed network backbones whose removal eliminated 81.5 % and 93.6 % of edges. BSEM reproduced 68–69 % of yield variance and resolved coherent phyla → functional guilds → agronomic traits → yield pathways (β = 0.55 for agronomic traits → yield), with cultivar effects ranging from +0.50 (Nonghua 106) to −0.49 (Dr. King 825).

Machine learning confirmed complementary yield-associated signatures (Rhodanobacter ASV1138, $\rho = 0.91$).

Conclusion. Bacteria and fungi obey kingdom-specific assembly rules; Proteobacteria and Ascomycota are structurally indispensable network backbones; and a Bayesian structural equation framework traces interpretable pathways from microbiome features to measured yield, offering testable hypotheses for microbiome-assisted maize breeding.

Keywords: Bayesian structural equation modeling; co-occurrence network; maize (Zea mays L.); rhizosphere microbiome; yield-associated microbial signatures

Abbreviations: ACC, 1-aminocyclopropane-1-carboxylate; ANOSIM, analysis of similarities; ASV, amplicon sequence variant; BSEM, Bayesian structural equation modeling; CFI, comparative fit index; CI, credible interval; FAPROTAX, Functional Annotation of Prokaryotic Taxa; FUNGuild, Fungal Functional Guild; HSD, honestly significant difference; IAA, indole-3-acetic acid; ITS, internal transcribed spacer; LDA, linear discriminant analysis; LEfSe, linear discriminant analysis effect size; MCMC, Markov chain Monte Carlo; NMDS, non-metric multidimensional scaling; PCA, principal component analysis; PCoA, principal coordinates analysis; PERMANOVA, permutational multivariate analysis of variance; RMSEA, root mean square error of approximation; RQ, research question; SD, standard deviation; SRMR, standardized root mean square residual; SynCom, synthetic community; V6, six-leaf stage; VT, tasseling stage; wUF, weighted UniFrac.

Introduction

The global population is projected to reach approximately 9.7 billion by 2050 (United Nations 2022), placing unprecedented pressure on food security. Maize (Zea mays L.) is among the world's highest-yielding cereal crops and serves as staple food, animal feed, and industrial raw material, so its stable production is essential for agricultural sustainability (Ren et al. 2022). However, decades of intensive fertilization and conventional breeding have been accompanied by diminishing marginal yield returns, soil degradation, non-point source pollution, and biodiversity loss in many major production regions (Das et al. 2022; Li et al. 2024), making the transition toward sustainable crop production an urgent priority.

The rhizosphere—the narrow soil zone influenced by root activity—harbors microbial communities that are intimately linked to soil fertility, stress tolerance, plant development, and yield formation (Berendsen et al. 2012; Philippot et al. 2013). Plants allocate 11–40 % of photosynthetically fixed carbon to the rhizosphere through root exudates, creating a specialized nutritional niche that drives directed microbial assembly (Sasse et al. 2018; Zhalnina et al. 2018). In return, microorganisms supply the host with nitrogen fixation, phosphate solubilization, pathogen suppression, and phytohormone regulation (Feng et al. 2025; Liu-Xu et al. 2024; Sagar et al. 2021; Vandenkoornhuyse et al. 2015; Wang et al. 2023). The rhizosphere microbiome is therefore widely viewed as a core functional extension of the host plant and one of the most promising levers for ecologically sustainable yield improvement (Trivedi et al. 2020).

Host genotype is a well-documented driver of rhizosphere microbiome variation. Landmark studies in *Arabidopsis*, rice, and maize have shown that root-associated communities are structured by host-driven selection and are partly heritable under field conditions (Bulgarelli et al. 2012; Edwards et al. 2015; Garrido-Oter et al. 2018; Peiffer et al. 2013; Walters et al. 2018). In maize specifically, rhizosphere taxa are heritable and repeatable across seasons (Peiffer et al. 2013; Walters et al. 2018), and recent work has extended this picture in three directions: locally adapted maize accessions exert genotype- and environment-dependent genetic control over their root microbiomes, down to host variants associated with individual keystone bacteria (He et al. 2024); intraspecific genetic variation modulates even narrow functional groups such as nitrifiers and denitrifiers under conventional agriculture (Favela et al. 2025); and comparisons of inbred and hybrid maize indicate that non-genetic factors often outweigh genotype effects, leaving the strength

of host control context-dependent (Schultz et al. 2023). Developmental stage is an equally powerful filter: as plants grow, root exudate chemistry shifts from readily assimilable sugars and organic acids toward more recalcitrant compounds, reordering microbial substrate preferences and community succession (Zhalnina et al. 2018). Despite this progress, four gaps remain. First, bacteria and fungi are usually profiled separately, so whether the two kingdoms obey distinct assembly rules under an identical field design is rarely quantified side by side. Second, co-occurrence networks are routinely used to nominate keystone taxa, yet their structural indispensability is seldom validated by systematic removal or attack–tolerance analyses (Banerjee et al. 2018; Faust and Raes 2012). Third, taxonomic composition does not map one-to-one onto function, and the functional guilds that plausibly connect microbes to host performance are often left unexamined (Louca et al. 2016). Fourth—and most consequential for breeding—most studies remain at the level of descriptive composition or static trait–community correlations, leaving the statistical pathways from host factors through microbial features to measured yield largely uncharted; indeed, whether phenotypic differences among genotypes track community differences at all is rarely tested explicitly. Modern maize cultivars, having undergone intensive artificial selection, may have inadvertently altered their capacity to recruit beneficial soil microorganisms (Trivedi et al. 2020; Vandenkoornhuyse et al. 2015); quantifying such cultivar-dependent microbial effects is a prerequisite for microbiome-assisted breeding.

In this study, we used seven divergent maize cultivars grown in a replicated field trial and sampled rhizosphere soils at the seedling (V6) and tasseling (VT) stages, profiling bacterial and fungal communities in parallel. We built an integrated analytical framework with four components: (i) kingdom-resolved diversity analysis to quantify the relative contributions of genotype and developmental stage (RQ1); (ii) co-occurrence network analysis combined with phylum-level removal and attack–tolerance simulations to identify keystone taxa and verify their structural role (RQ2) (Banerjee et al. 2018; Barberán et al. 2012); (iii) FAPROTAX and FUNGuild functional annotation to bridge taxonomy and function (Louca et al. 2016; Nguyen et al. 2016); and (iv) a Bayesian structural equation modeling (BSEM) framework that specifies ecologically informed a priori structures and resolves directional statistical pathways from source variables through microbial features and functional guilds to measured grain yield (RQ3) (Merkle and Rosseel 2018; Rosseel 2012). Grain yield was measured directly in the field. Machine learning was deliberately restricted to two supporting tasks—identifying yield-associated microbial signatures and cross-

checking the coherence of BSEM-derived pathways—because the number of genotype-level observations does not support predictive validation (Topçuoğlu et al. 2020); throughout this article, terms such as causal path refer strictly to statistical pathways estimated under the a priori BSEM structure, not to experimental demonstrations of the underlying biological processes.

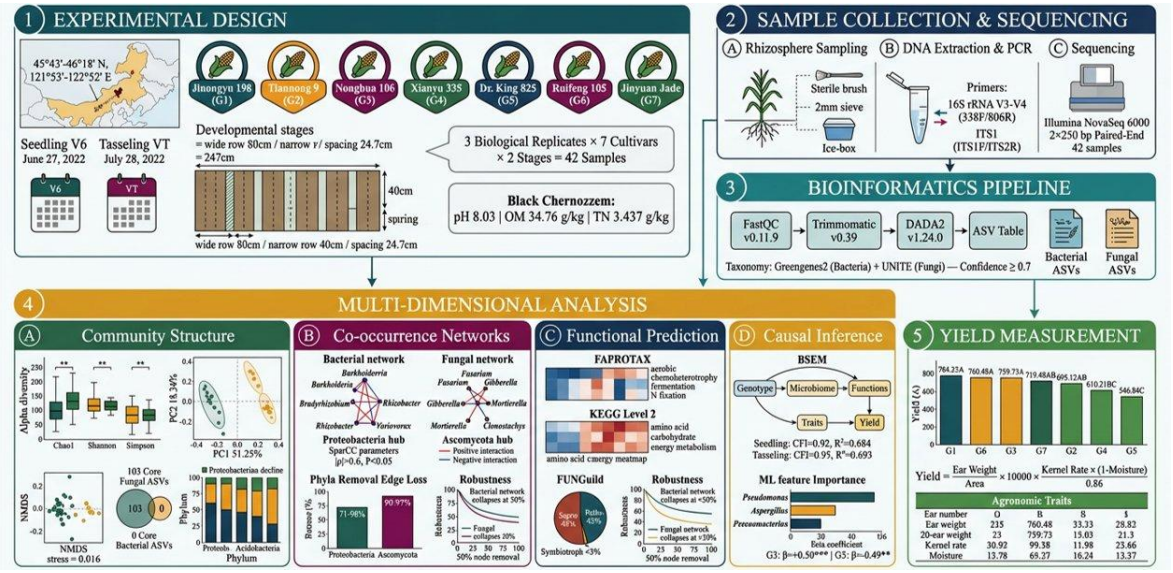


Fig. 1. Experimental design and integrative analytical framework. Seven maize cultivars were grown in a replicated field trial and rhizosphere-sampled at the seedling (V6) and tasseling (VT) stages (42 samples); bacterial (16S rRNA) and fungal (ITS) communities were profiled in parallel and integrated through diversity analyses, co-occurrence networks with robustness validation, functional annotation, Bayesian structural equation modeling (BSEM), and machine-learning-based signature identification. G1–G7 denote BSEM cultivar codes.

Materials and methods

Study site and plant materials

The field experiment was established in 2022 at the experimental station of the Xing'an League Institute of Agricultural and Animal Husbandry Sciences, Inner Mongolia, China (45°43′–46°18′ N, 121°53′–122°52′ E). Seven maize (Zea mays L.) cultivars with divergent genetic backgrounds and maturity groups were selected: Jinongyu 198 (G1), Tiannong 9 (G2), Nonghua 106 (G3), Xianyu 335 (G4), Dr. King 825 (G5), Ruifeng 105 (G6), and Jinyuan Jade 177 (G7) (Table S1). Cultivars were grown in a wide–narrow row configuration (wide row 80 cm, narrow row 40 cm, intra-row spacing 24.7 cm) under uniform field management, including basal fertilization (compound fertilizer N–P_2O_5–K_2O = 15–15–15, 750 kg ha^{-1}) and a single topdressing of urea (300 kg ha^{-1}); irrigation, pest, and weed control followed local high-yield practice and were identical

across cultivars and replicates. Each cultivar was represented by three biological replicate plots (0.20 ha each), and the same plots were sampled at both developmental stages. Rhizosphere soil was collected on 27 June (seedling stage, V6) and 28 July (tasseling stage, VT) 2022, yielding 42 samples in total (7 cultivars × 2 stages × 3 replicates). Cultivar codes G1–G7 correspond to sequencing sample prefixes WY2, WY3, WY8, WY9, WY13, WY14, and WY15, respectively (Table S1).

Rhizosphere soil collection

A five-point sampling scheme was applied within each replicate plot at both stages. Five plants were selected at random, excavated to 10–20 cm depth with intact root systems, and gently shaken to dislodge loosely adhering bulk soil. Rhizosphere soil tightly bound to the root surface was brushed off with sterile brushes and pooled across the five plants into one composite sample per plot–stage combination. Samples were sealed in sterile bags, immediately flash-frozen in liquid nitrogen in the field, transported on dry ice, and stored at −80 °C until DNA extraction; sterile gloves and brushes were changed between samples to prevent cross-contamination.

DNA extraction, sequencing, and bioinformatics

Total community DNA was extracted from rhizosphere soil using a standard commercial soil DNA extraction kit, treated with RNase, and quantified; extraction blanks were included as negative controls. The bacterial 16S rRNA V3–V4 region (primers 338F/806R) and the fungal ITS1 region (primers ITS1F/ITS2R) were amplified with barcoded primers and sequenced on the Illumina NovaSeq 6000 platform. Raw reads were quality-filtered, adapter- and primer-trimmed, merged, and denoised, and amplicon sequence variants (ASVs) were inferred with DADA2 (Callahan et al. 2016), which removes chimeras. Taxonomic classification was performed against the Greengenes2 database for bacteria and the UNITE database for fungi. QIIME 2 (Bolyen et al. 2019) was used to calculate alpha-diversity (Chao1, Shannon, Simpson, ACE, observed features, and Faith's phylogenetic diversity) and beta-diversity (Bray–Curtis and weighted UniFrac) indices. ASVs with relative abundance < 0.01 % were filtered from the feature table. Differentially abundant taxa were identified using LEfSe with an LDA score threshold of 3.0 (Segata et al. 2011).

Yield and agronomic trait measurement

At physiological maturity, ear number, ear weight, 20-ear weight, cob weight, kernel percentage (grain ratio), and grain moisture were recorded for each plot, and standardized grain yield was

calculated as kg per mu (1 mu = 1/15 ha) following national standards for maize variety trials. Grain yield was therefore an observed field measurement used throughout as the response variable; no yield forecasting or prediction of any kind was performed.

Diversity statistics and Mantel tests

Differences in alpha-diversity between stages were assessed with Welch's two-sample t-tests, and genotype effects within each stage with Kruskal–Wallis tests followed by Dunn's post hoc comparisons. Community differentiation was quantified with ANOSIM and PERMANOVA (9,999 permutations each) on Bray–Curtis and weighted UniFrac distance matrices, testing the factor developmental stage across all samples and the factor genotype within each stage; non-metric multidimensional scaling (NMDS) provided ordination. To test whether phenotypic differentiation tracks community differentiation, Mantel tests (9,999 permutations) were computed between z-scored agronomic-trait distance matrices and community distance matrices. Because agronomic traits were measured once per cultivar, the primary Mantel analyses were performed at the genotype level using 7 × 7 mean matrices (2 kingdoms × 2 distance metrics × 2 stages = 8 tests), which avoids pseudo-replication; sample-level tests with genotype-assigned traits (a further 8 tests) are reported for completeness and interpreted with caution. P-values across the 16 tests were adjusted with the Benjamini–Hochberg (BH) procedure, and only BH-adjusted P values are reported as significant.

Co-occurrence network analysis and robustness validation

Co-occurrence networks were constructed separately for each cultivar × stage × kingdom combination (7 × 2 × 2 = 28 networks). Pairwise associations were inferred with SparCC (Friedman and Alm 2012) and edges were retained at $|\rho| > 0.6$ and $P < 0.05$; the 50 most abundant ASVs in each network were kept as nodes, and edge sign was assigned by the sign of the correlation (green, positive; red, negative). Networks were built and analyzed with the igraph package (Csardi and Nepusz 2006). Topology (nodes, edges, positive/negative edge ratio, density, average degree, clustering coefficient, and modularity) was recomputed directly from the edge tables; communities were detected with a greedy modularity-optimization algorithm. Keystone nodes were defined as the top five nodes by degree centrality in each network, and hub consistency was quantified as the number of networks in which an ASV recurred in the degree top five. Robustness was assessed in two complementary ways. First, in phylum-level removal tests, all nodes of a focal phylum with ≥

3 nodes were deleted and the resulting changes in edge count, node count, and modularity were recorded; modularity before and after removal was recomputed de novo from the edge tables for every network to guarantee numerical validity. Second, attack–tolerance curves were generated for the functional co-occurrence networks by removing nodes either randomly or in decreasing order of degree (targeted attack) and tracking the relative size of the largest connected component (Bastian et al. 2009; Faust and Raes 2012). Network visualization used a force-directed layout with nodes colored by phylum.

Functional annotation and functional co-occurrence networks

Bacterial functions were predicted with FAPROTAX (Louca et al. 2016) and fungal trophic modes with FUNGuild (Nguyen et al. 2016); both assign ASVs to ecologically interpretable guilds. Functional-group abundances were aggregated per sample, and functional co-occurrence networks were constructed with the same SparCC-based correlation approach used for the taxonomic networks. Keystone functional groups were identified from degree and betweenness centrality, and single-node removal impact scores were computed for the fungal guilds to rank their structural importance.

Bayesian structural equation modeling

A Bayesian structural equation modeling (BSEM) framework (Merkle and Rosseel 2018; Rosseel 2012) was used to resolve multi-level statistical pathways linking source variables to grain yield. The a priori structure specified five layers: (i) source variables (cultivar group in the genotype models; developmental stage in the stage model); (ii) microbial community features (the relative abundances of the dominant bacterial phyla Proteobacteria, Acidobacteriota, Firmicutes, and Chloroflexi, and the fungal phyla Ascomycota, Basidiomycota, and Mortierellomycota); (iii) predicted functional guilds (bacterial nitrogen metabolism, carbon metabolism, aerobic heterotrophy, and other bacterial functions; fungal organic matter decomposition, nutrient activation, and other fungal functions); (iv) an agronomic-trait composite score constructed by standardizing the measured agronomic traits and combining them; and (v) measured grain yield as the terminal variable. Directed paths were specified from source variables to community features, from community features to functional guilds, from functional guilds to the agronomic score and yield, and from the agronomic score to yield, with additional direct paths from source variables to yield. Two genotype models (one per stage; N = 21 each; 127 paths) and one stage model (both

stages pooled; N = 42; 73 paths) were fitted. Parameters were estimated by Markov chain Monte Carlo with weakly informative priors; convergence was assessed via the potential scale reduction factor and effective sample size, and standardized coefficients with 95 % credible intervals are reported. Model fit was evaluated with the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Total, direct, and indirect effects on yield were decomposed per cultivar and for developmental stage, and genus-level effect decomposition was performed by weighting genus abundances by the corresponding BSEM path coefficients. Throughout, the term causal path denotes a directional statistical pathway estimated under this a priori structure; it does not imply experimental verification of the underlying biological processes.

Machine-learning-based signature identification

Machine learning was applied exclusively for signature identification and for checking the coherence of BSEM-derived pathways, not for yield forecasting. Random forest (Breiman 2001) and XGBoost (Chen and Guestrin 2016) feature-importance scores were computed on genotype-level feature tables at the genus and ASV levels using default hyper-parameters and internal cross-validation; Spearman correlations between ASV abundances and measured yield were computed at the sample level (n = 42) and the genotype level (n = 7). Because seven genotype-level observations cannot support predictive validation (leave-one-out cross-validated R^2 was negative), no predictive accuracy is reported anywhere in this study, following current recommendations for machine learning on microbiome data (Topçuoğlu et al. 2020). Genera with the strongest positive yield associations were compiled into an in silico synthetic-community (SynCom) candidate list, presented strictly as a hypothesis-generating design for future isolation and inoculation work (Mueller and Sachs 2015; Singh et al. 2024).

Statistical analysis

Data are presented as mean ± SD unless otherwise stated. One-way ANOVA with Tukey's HSD post hoc test ($P < 0.05$) was used for multiple comparisons among cultivars. Alpha- and beta-diversity analyses used QIIME 2 outputs as described above; ANOSIM, PERMANOVA, and Mantel tests were implemented with 9,999 permutations; networks were computed with NetworkX and igraph; BSEM was implemented in R; machine-learning analyses used scikit-learn and

XGBoost. All scripts and the analysis pipeline are available from the authors. Significance levels: *P < 0.05, **P < 0.01, ***P < 0.001; all tests were two-sided.

Results

Phenotypic variation is substantial but decoupled from community distances

Grain yield and agronomic traits differed significantly among the seven cultivars (Table 2; Fig. 2a). Jinongyu 198 (G1) achieved the highest yield (784.25 ± 33.91 kg per mu), followed by Ruifeng 105 (G6; 760.48 ± 17.99) and Nonghua 106 (G3; 759.73 ± 36.94), whereas Dr. King 825 (G5) produced the lowest yield (546.84 ± 28.00 kg per mu); the across-cultivar coefficient of variation was 12.6 %. Principal component analysis of z-scored agronomic traits separated the cultivars along PC1 (56.1 % of variance) and PC2 (26.0 %), with the high-yielding cultivars G1, G3, and G6 scoring positively on PC1 and the low-yielding G4 and G5 scoring negatively (Fig. 2b, c).

To test whether this phenotypic differentiation was mirrored by community differentiation, we performed Mantel tests between agronomic-trait distances and community distances (Fig. 2d; Fig. S1; Table S1). At the genotype level—the appropriate unit because traits were measured once per cultivar—none of the eight tests (2 kingdoms × 2 distance metrics × 2 stages) was significant after Benjamini–Hochberg correction ($|r| \leq 0.26$, adjusted $P \geq 0.30$ in all cases; Table S3). Among the sample-level tests, which re-use genotype-assigned trait values and therefore inflate the effective sample size, only the bacterial weighted-UniFrac comparison at VT reached significance ($r = 0.383$, $P = 0.0002$, adjusted $P = 0.0016$), a pattern consistent with pseudo-replication of genotype-level traits rather than with genuine within-genotype covariation. Phenotype–microbiome decoupling at the genotype level therefore stands as a genuine result of this study: cultivar differences in yield and agronomic traits are not accompanied by proportional shifts in overall community composition, motivating the pathway-level modeling of the Bayesian structural equation framework presented below rather than direct trait–community matching.

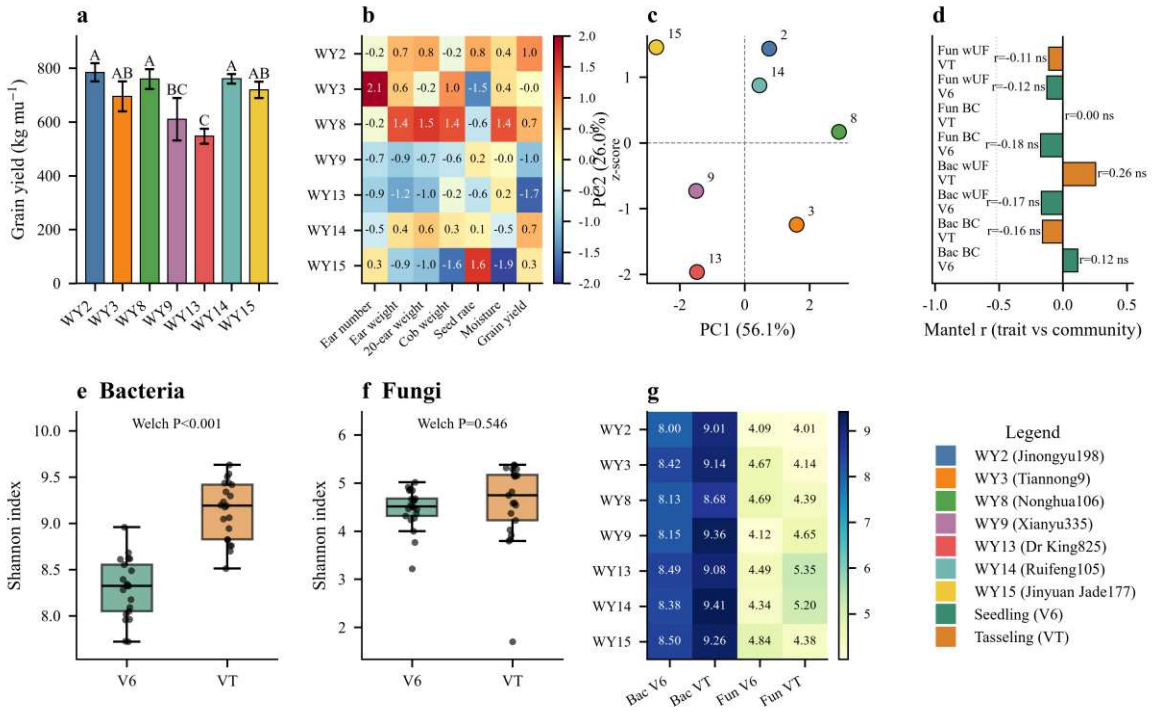


Fig. 2. Phenotypic variation, phenotype–microbiome decoupling, and alpha-diversity. (a) Grain yield of the seven cultivars (mean ± SE, n = 3; uppercase letters denote Tukey HSD groups, $P < 0.05$). (b) Heatmap of z-scored agronomic traits. (c) Principal component analysis of agronomic traits (PC1 = 56.1 %, PC2 = 26.0 %). (d) Genotype-level Mantel tests between agronomic-trait and community distance matrices (7 × 7 mean matrices; 9,999 permutations); no test remained significant after Benjamini–Hochberg correction (dashed line: r = 0). (e, f) Bacterial and fungal Shannon diversity by developmental stage (Welch's t-test; NS, not significant). (g) Shannon diversity across cultivar × stage combinations (mean ± SEM). V6, seedling stage; VT, tasseling stage.

Table 2. Grain yield and agronomic traits of seven maize cultivars at maturity. Values are mean ± SD (n = 3). Different uppercase letters indicate significant differences among cultivars ($P < 0.05$, Tukey HSD). 1 mu = 1/15 ha.

Cultivar (BSEM group)	Ear number	Ear weight (kg)	20-ear weight (kg)	Cob weight (kg)	Grain ratio (%)	Moisture (%)	Yield (kg/mu)
Jinongyu 198 (G1)	66.67 ± 5.13 AB	20.03 ± 0.98 A	6.02 ± 0.20 AB	0.99 ± 0.07 C	83.49 ± 1.17 AB	27.40 ± 0.46 B	784.25 ± 33.91 A
Tiannong 9 (G2)	75.00 ± 3.46 A	19.64 ± 1.25 A	5.24 ± 0.11 BC	1.29 ± 0.02 AB	75.29 ± 0.40 D	27.3 ± 1.45 B	695.12 ± 55.76 AB
Nonghua 106 (G3)	66.67 ± 5.13 AB	21.87 ± 0.25 A	6.59 ± 0.58 A	1.39 ± 0.19 A	78.39 ± 2.11 CD	31.40 ± 2.26 A	759.73 ± 36.94 A
Xianyu 335 (G4)	64.67 ± 6.35 B	15.60 ± 1.91 B	4.83 ± 0.54 C	0.90 ± 0.02 C	81.31 ± 1.89 BC	25.57 ± 1.22 B	610.21 ± 79.25 BC

Dr. King 825 (G5)	64.00 ± 2.00 B	14.73 ± 1.25 B	4.60 ± 0.27 C	1.00 ± 0.08 C	78.35 ± 0.62 CD	26.53 ± 2.27 B	546.84 ± 28.00 C
Ruifeng 105 (G6)	65.33 ± 4.16 B	19.10 ± 0.89 A	5.85 ± 0.20 AB	1.11 ± 0.11 BC	80.98 ± 1.79 BC	23.83 ± 0.95 B	760.48 ± 17.99 A
Jinyuan Jade 177 (G7)	68.33 ± 4.51 AB	15.73 ± 0.55 B	4.61 ± 0.14 C	0.65 ± 0.04 D	86.17 ± 0.63 A	17.87 ± 0.93 C	719.48 ± 30.24 AB

Kingdom-specific assembly: bacteria follow developmental stage, fungi retain stronger genotype structure

To resolve the relative strengths of genotype and developmental stage as assembly filters (RQ1), we compared alpha- and beta-diversity of the two kingdoms across all 42 samples. Bacterial alpha-diversity increased sharply from V6 to VT across all six indices: Shannon diversity rose from 8.30 ± 0.07 to 9.13 ± 0.07 (Welch $t = -8.50$, $P = 1.7 \times 10^{-10}$), Chao1 from 1,018 ± 29 to 1,510 ± 37 ($P = 7.3 \times 10^{-13}$), and Faith's phylogenetic diversity from 127 ± 3 to 182 ± 3 ($P = 6.2 \times 10^{-15}$; Fig. 2e, g; Fig. S2). In contrast, genotype had no detectable effect on bacterial alpha-diversity within either stage (Kruskal–Wallis $P \geq 0.22$ for all indices). Fungal alpha-diversity showed the opposite profile: no stage effect on Shannon diversity (4.46 ± 0.09 at V6 versus 4.59 ± 0.19 at VT, $P = 0.55$), only marginal declines in richness estimators (Chao1, $P = 0.053$; ACE, $P = 0.046$), and no genotype effect ($P \geq 0.14$; Fig. 2f, g; Fig. S2). Rarefaction curves confirmed adequate sequencing depth for both kingdoms (Fig. S2).

Community composition shifted accordingly (Fig. 3a, b; Fig. S3). Bacterial communities were dominated by Proteobacteria, Actinobacteriota, and Acidobacteriota, with a successional shift from copiotrophic Proteobacteria toward oligotrophic Acidobacteriota, Actinobacteriota, and Chloroflexi from V6 to VT. Fungal communities were dominated by Ascomycota, with a conspicuous expansion of Basidiomycota in cultivar WY2 at VT. LEfSe identified 341 bacterial and 172 fungal differentially abundant features (LDA > 3.0; Fig. 3c; Fig. S4): the strongest bacterial biomarkers included Proteobacteria and Gammaproteobacteria in WY14 at V6 (LDA > 5.2), while the strongest fungal biomarkers were Ascomycota in WY14 at V6 (LDA = 5.43) and the Sordariomycetes–Hypocreales–Nectriaceae lineage, including *Gibberella intricans*, in WY8 at VT.

Beta-diversity analyses resolved the relative strength of stage and genotype effects (Fig. 3d–f; Fig. S5; Table S2). For bacteria, principal coordinate analysis on weighted UniFrac distances separated the two stages along PC1, which alone captured 51.25 % of the variance; accordingly, developmental stage was the dominant factor (ANOSIM R = 0.901, P = 0.0002; PERMANOVA F = 32.32, P = 0.0002), whereas genotype effects were smaller but significant within each stage (weighted UniFrac: V6 R = 0.417, P = 0.0004; VT R = 0.838, P = 0.0002). On Bray–Curtis distances, the bacterial genotype effect was significant only at VT (R = 0.470, P = 0.0004; V6: R = 0.095, P = 0.186). For fungi, the pattern was inverted: genotype effects (weighted UniFrac: V6 R = 0.306, P = 0.0006; VT R = 0.685, P = 0.0002; Bray–Curtis: V6 R = 0.393, VT R = 0.738, both P = 0.0002) exceeded the stage effect (weighted UniFrac R = 0.421; Bray–Curtis R = 0.505, both P = 0.0002). Fungal ordination supported this view (PC1 = 36.64 %, PC2 = 16.33 %), with cultivar-level clustering evident within stages (Fig. 3e). Together these results answer RQ1: bacterial rhizosphere assembly is filtered primarily by plant ontogeny, whereas fungal assembly retains comparatively stronger genotype-specific structure, particularly at tasseling.

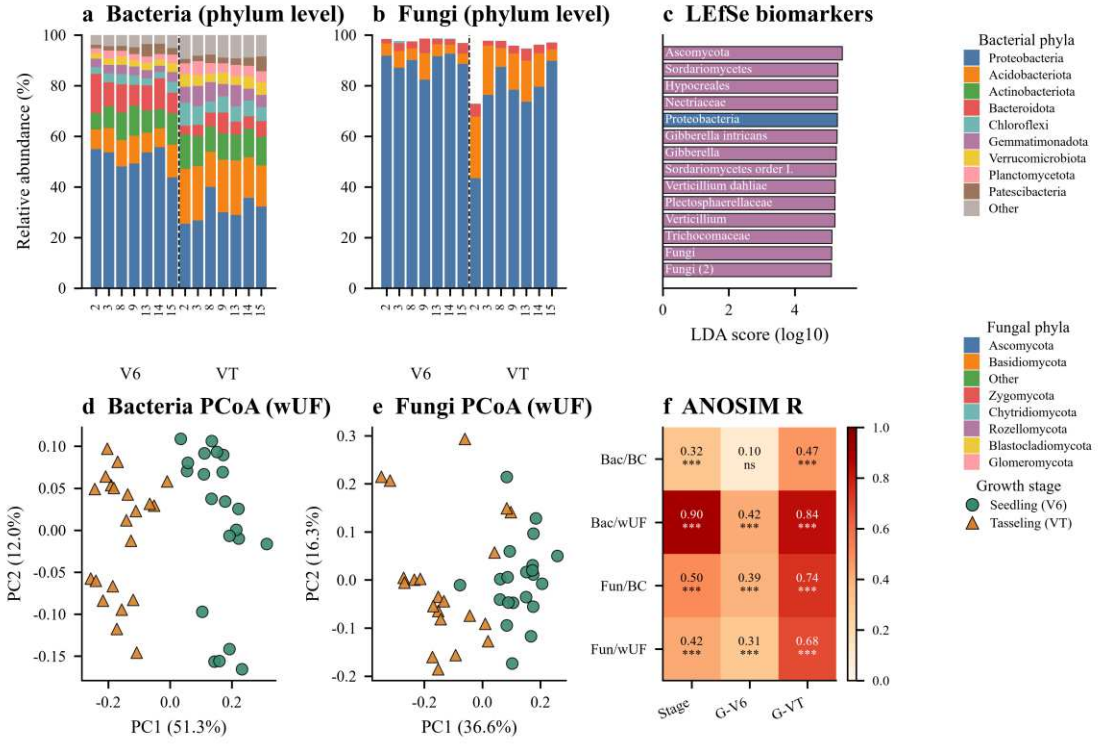


Fig. 3. Community composition and beta-diversity. (a, b) Phylum-level relative abundance of bacterial and fungal communities across cultivars and stages (top 10 phyla plus remainder). (c) LEfSe biomarkers (top 14 by LDA score; group membership in brackets). (d, e) Principal coordinate analysis on weighted UniFrac distances for bacteria (PC1 = 51.25 %, PC2 = 11.98 %) and fungi (PC1 = 36.64 %, PC2 = 16.33 %). (f) ANOSIM R values for the factors stage (all samples) and genotype (within each

stage) on Bray–Curtis and weighted UniFrac distances; all cells $P \leq 0.0004$ except bacterial genotype at V6 on Bray–Curtis (NS). NMDS ordinations and distance-distribution panels are provided in Fig. S5.

Co-occurrence networks are organized around kingdom-specific backbones that are structurally indispensable

To determine whether the taxa nominated by network analysis are structurally indispensable (RQ2), we subjected all 28 networks to phylum-removal and attack–tolerance simulations. All 28 cultivar × stage × kingdom networks (50 nodes each) exhibited non-random, modular topology (Fig. 4a–c; Fig. S6; Table S4). Bacterial networks carried 426–540 edges (density 0.348–0.441) with a large majority of positive associations (positive-edge ratio 0.74–0.94). Fungal networks were comparable in size (417–664 edges) but architecturally distinct: negative associations accounted for roughly half of all edges in most networks (positive-edge ratio 0.48–0.58 in 12 of 14 networks), indicating a competition-rich interaction structure. Modularity ranged from 0.35 to 0.59 with two to three modules per network (Fig. S6).

Keystone analysis converged on kingdom-specific backbones (Fig. 4d; Table S5). Among the bacterial degree-based hub ASVs (139 hub assignments across 14 networks), 88 (63 %) belonged to Proteobacteria, followed by Acidobacteriota (17; 12 %); ASV7 recurred in the degree top five of four networks and ASV10 in three. Among the 130 fungal hub assignments, 103 (79 %) belonged to Ascomycota, with Zygomycota (16; 12 %) and Basidiomycota (11; 8 %) contributing most of the remainder.

Robustness validation confirmed that these hub phyla are structurally indispensable (Fig. 4e, f; Fig. S7; Table S9). Simulated removal of Proteobacteria from bacterial networks eliminated on average 81.5 % of edges (up to 99.8 % in the most affected network) and 61.6 % of nodes (up to 92.0 %), and reduced network modularity by an average of 16.3 % (de novo recomputed from edge tables). Removal of Ascomycota from fungal networks eliminated on average 93.6 % of edges (up to 97.4 %) and 73.9 % of nodes (up to 82.0 %), with modularity decreasing by an average of 15.8 %. No other phylum produced remotely comparable effects: the next largest removals (Basidiomycota and Zygomycota in fungal networks) eliminated only 17–19 % of edges and changed modularity by less than 4 % on average. Attack–tolerance simulations on the functional co-occurrence networks gave consistent results: the dense, positively wired bacterial functional network tolerated random node removal until approximately 80 % of nodes had been deleted but collapsed when about 50 %

of nodes were removed in decreasing degree order—a scale-free vulnerability signature—whereas the sparse fungal functional network began to fragment below 20 % targeted removal (Fig. 4g). These results answer RQ2: co-occurrence networks are organized around Proteobacteria and Ascomycota backbones whose loss destabilizes the entire interaction structure.

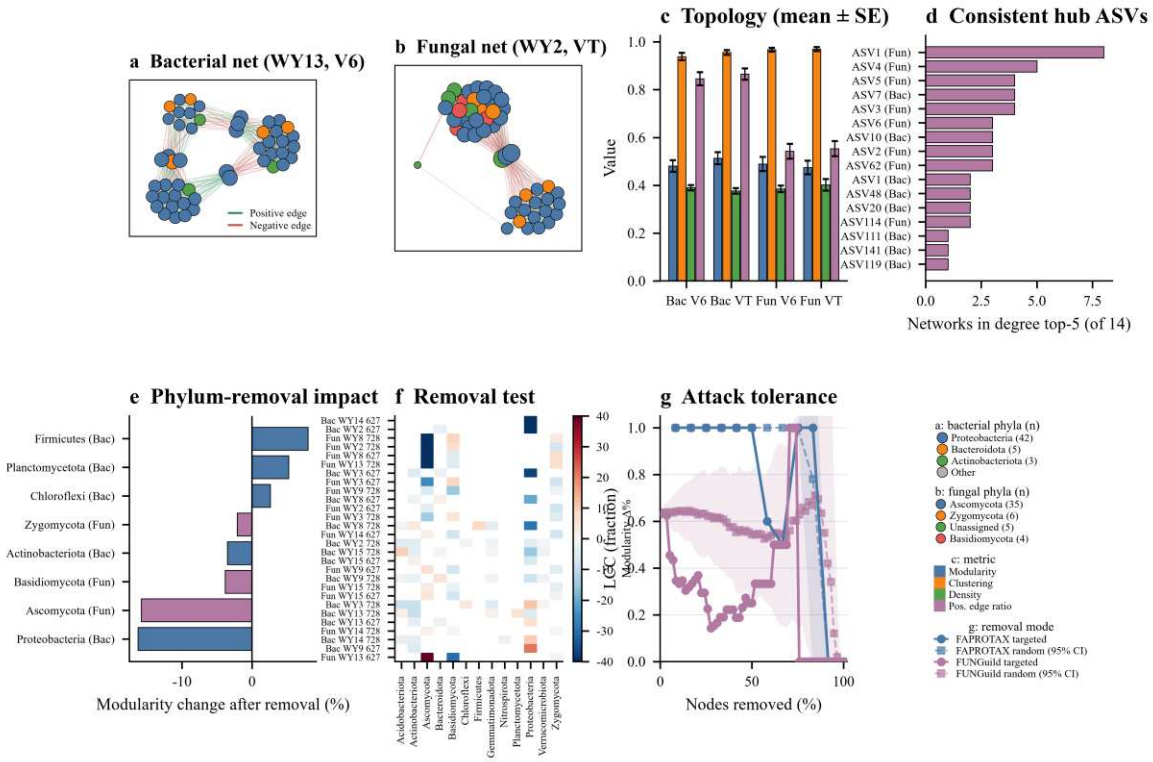


Fig. 4. Co-occurrence networks, keystone taxa, and network robustness. (a, b) Representative bacterial (WY13, V6) and fungal (WY2, VT) networks; nodes are colored by phylum (top six phyla plus remainder), edges by sign (green, positive; red, negative). (c) Topological properties of all 28 networks (nodes, edges, average degree, clustering coefficient, modularity) by kingdom and stage. (d) Consistent hub ASVs ranked by the number of networks in which they reach the degree top five. (e) Mean change in network modularity after phylum-level removal (de novo recomputed from edge tables); only Proteobacteria (bacteria) and Ascomycota (fungi) exceed 15 % average modularity loss. (f) Modularity change across all 28 networks after removal of the dominant phyla. (g) Attack–tolerance curves of bacterial and fungal functional co-occurrence networks under random versus targeted (degree-ordered) node removal; shaded bands denote 95 % confidence intervals across random-removal replicates. Panel h collects figure legends; the edge-sign key is embedded in panel a.

Functional guilds link taxonomy to host performance

Because taxonomic composition does not map one-to-one onto function, we next asked which functional guilds dominate the rhizosphere and how they are organized. Predicted bacterial functions were dominated by heterotrophic carbon cycling: chemoheterotrophy and aerobic chemoheterotrophy together accounted for 16.4–50.8 % of column-normalized FAPROTAX

assignments (mean 30.5 %; Fig. 5a; Table S7). Functional composition tracked developmental stage: the seedling stage was relatively enriched in dark iron oxidation and lignocellulose-degradation assignments, whereas the tasseling stage was enriched in ureolysis and C1-metabolism assignments, indicating a shift toward nitrogen transformation and residue decomposition as plants entered the reproductive phase (Fig. 5g).

Fungal trophic modes were dominated by saprotrophs (28.9 %), mixed pathotroph–saprotroph–symbiotroph assignments (24.7 %), and pathotrophs (21.0 %), with 13.0 % unassigned and only 12.4 % symbiotrophs (Fig. 5b); obligate symbiotic guilds were nearly absent, consistent with a chernozem rhizosphere governed mainly by decomposition and pathogen-pressure processes rather than by mycorrhizal mutualism. The functional co-occurrence networks mirrored the taxonomic architecture: the bacterial functional network was dense (density = 0.636) and dominated by positive associations with C1 metabolism forming a tightly connected module, whereas the fungal functional network was sparse (density = 0.053) with a higher share of negative associations (Fig. 5d–f; Fig. S8). Centrality analyses highlighted nitrogen-metabolism and nutrient-activation groups as the most central functional nodes (Fig. 5a, c; Fig. S8), and it is precisely these functional guilds—not bulk taxonomic composition—that the BSEM framework below identifies as the layer connecting microbes to host performance.

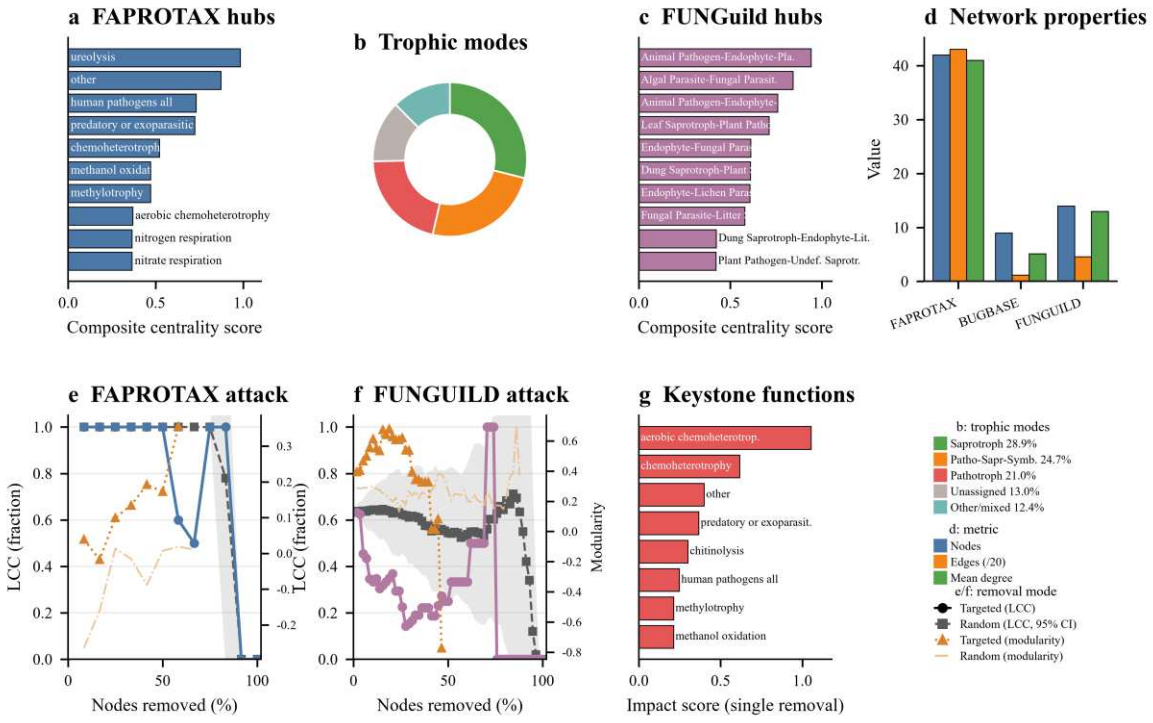


Fig. 5. Functional guilds and functional co-occurrence networks. (a) Central FAPROTAX-predicted bacterial functional groups by network degree. (b) FUNGuild trophic-mode composition across samples (saprotroph 28.9 %; pathotroph–saprotroph–symbiotroph 24.7 %; pathotroph 21.0 %; unassigned 13.0 %; symbiotroph 12.4 %). (c) Keystone FUNGuild guilds by centrality. (d) Properties of bacterial and fungal functional networks (density, positive-edge ratio, clustering, modularity). (e, f) Attack–tolerance curves with modularity trajectories for bacterial and fungal functional networks; shaded bands denote 95 % confidence intervals of the largest connected component (LCC) across random-removal replicates. (g) Stage-resolved abundance of keystone bacterial functions. Panel h collects figure legends.

A Bayesian structural equation framework resolves statistical pathways from microbiome features to measured yield

To trace the statistical pathways connecting microbiome features to measured yield (RQ3), we fitted the BSEM framework (Fig. 6a), which linked source variables (cultivar or developmental stage) through dominant phyla, predicted functional guilds, and the agronomic-trait composite to measured grain yield. All three models achieved satisfactory fit (Table 4): the seedling genotype model (N = 21, 127 paths) reached $\chi^2/df = 0.59$ (P = 0.85), RMSEA = 0.055, CFI = 0.92, and SRMR = 0.050; the tasseling genotype model reached $\chi^2/df = 0.62$ (P = 0.84), RMSEA = 0.048, CFI = 0.95, and SRMR = 0.046; and the pooled stage model (N = 42, 73 paths) reached $\chi^2/df = 0.99$ (P = 0.82), RMSEA = 0.052, CFI = 0.935, and SRMR = 0.048. The models reproduced 68.4 % (V6), 69.3 % (VT), and 50.0 % (stage model) of the variance in measured yield (Table 4).

The pathway structure was highly coherent across models (Fig. 6b, c; Fig. S9; Table S6). The strongest single path was agronomic traits → yield (β = 0.55, 95 % CI 0.39–0.71, P = 0.002). Developmental stage influenced yield both directly (β = 0.42, P = 0.003) and indirectly through agronomic traits (stage → agronomic score β = 0.35, P = 0.008), for a total effect of 0.68, stronger at VT than at V6. Functional guilds fed into agronomic performance: fungal nutrient-activation → agronomic score (β = 0.30, P = 0.016), bacterial nitrogen-metabolism → agronomic score (β = 0.28, P = 0.021), and nutrient-activation → yield directly (β = 0.28, P = 0.021). Upstream, phylum-to-function paths connected Proteobacteria to carbon metabolism (β = 0.34), Firmicutes to aerobic heterotrophy (β = 0.33–0.35), and Ascomycota to organic matter decomposition (β = 0.31–0.32) and nutrient activation (β = 0.28–0.29) at both stages (Fig. 6b; Table S6).

Effect decomposition separated the cultivars along a clear gradient (Fig. 6d). Nonghua 106 (G3) carried the strongest positive total microbial effect on yield at both stages (V6: +0.495; VT: +0.504), about half of it indirect through agronomic traits and functional guilds (indirect share 49–50 %); Jinongyu 198 (G1; +0.484 and +0.370) and Ruifeng 105 (G6; +0.324 and +0.409) also

ranked positively. Dr. King 825 (G5) showed the most negative total effect at both stages (−0.489 and −0.402), and Xianyu 335 (G4) was likewise negative (−0.253 and −0.278). Genus-level decomposition assigned the largest per-genus total effects (0.07–0.14) mostly to low-abundance, partially uncultured lineages (e.g., UTCFX1, *Anaerolineae*, *Siccirubricoccus*; Fig. 6e), complementing rather than duplicating the phylum-level picture.

Machine-learning analyses, restricted to signature identification, corroborated the yield association layer from an independent, data-driven direction (Fig. 6f; Fig. S10; Table S8). Of 147 tested ASVs, the strongest sample-level correlates of measured yield were ASV371 (Spearman $\rho = 0.372$, $P = 0.015$), ASV30, ASV268, and ASV100 ($\rho \approx 0.33$, $P < 0.04$); at the genotype level, ASV1138, annotated to *Rhodanobacter*, reached $\rho = 0.906$ ($P = 0.005$). Random-forest importance rankings placed ASV139, ASV246, ASV722, ASV260, and ASV261 among the top microbial features (Fig. S10d). As pre-specified, no predictive accuracy is reported: leave-one-out cross-validation on seven genotype-level observations yielded a negative R^2, confirming that these data support signature identification but not yield forecasting (Topçuoğlu et al. 2020). Finally, the eight genera with the strongest positive yield associations—*Variovorax* ($\rho = 0.76$), *Pseudarthrobacter* (0.70), *Candidatus Udaeobacter* (0.59), *Bacillus* (0.49), *Mucilaginibacter* (0.47), *Pseudomonas* (0.36), *Sphingobium* (0.29), and Ellin6067 (0.26)—were compiled into an in silico SynCom candidate list (Fig. S10c), presented strictly as a hypothesis-generating design for future isolation work. Together, these results answer RQ3: microbial features connect to measured yield through an interpretable chain of phyla → functional guilds → agronomic traits, with cultivar-specific total effects that are reproducible across developmental stages.

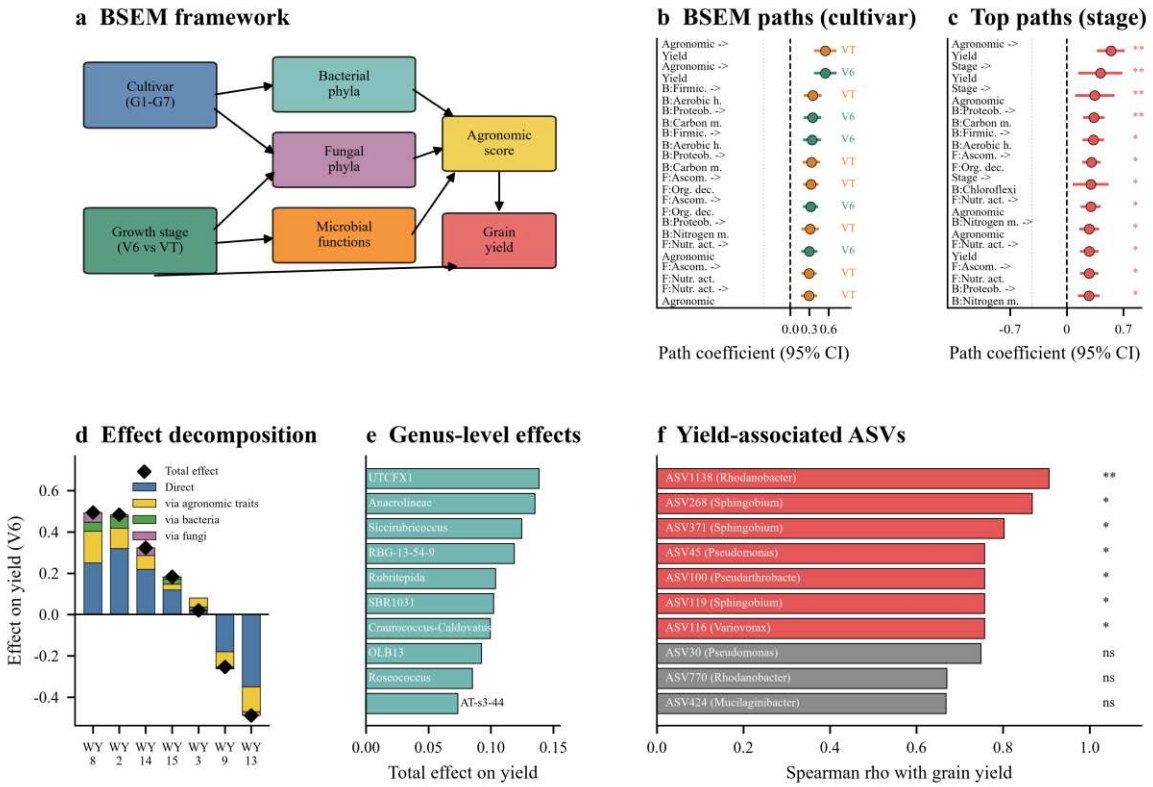


Fig. 6. A Bayesian structural equation framework links microbiome features to measured grain yield. (a) A priori BSEM structure: source variables (cultivar, growth stage) → dominant phyla → predicted functional guilds → agronomic-trait composite → measured yield. (b) Top significant paths of the genotype models (V6 and VT; standardized coefficients with 95 % credible intervals). (c) Top significant paths of the pooled stage model. (d) Cultivar-level decomposition of total microbial effects on yield into direct and indirect components at V6 and VT. (e) Genus-level total-effect decomposition (top 10 genera). (f) Yield-associated ASVs identified by machine-learning signature analysis (Spearman ρ with measured yield; significance stars as in Table S8). Model fit indices and the fraction of yield variance reproduced by each model are reported in Table 4. Grain yield was measured in the field; no forecasting was performed. Significance: *P < 0.05, **P < 0.01, ***P < 0.001.

Table 4. Fit indices and yield variance reproduced by the three Bayesian structural equation models. V6, seedling-stage genotype model; VT, tasseling-stage genotype model; Stage, pooled developmental-stage model. R² refers to measured grain yield.

Model	N	Paths	χ^2/df	P (χ^2)	RMSEA	CFI	TLI	SRMR	R^2 (yield)
Cultivar model, V6	21	127	0.59	0.850	0.055	0.920	0.880	0.050	0.684
Cultivar model, VT	21	127	0.62	0.842	0.048	0.950	0.910	0.046	0.693
Stage model (pooled)	42	73	0.99	0.819	0.052	0.935	0.902	0.048	0.500

Discussion

Bacteria and fungi obey kingdom-specific assembly rules

A central finding of this study is that the two microbial kingdoms follow distinct assembly rules under an identical field design. Bacterial communities were filtered primarily by plant ontogeny: developmental stage accounted for the dominant share of community variation (weighted-UniFrac ANOSIM R = 0.90; PCoA PC1 = 51.25 %), drove a six-index increase in alpha-diversity from V6 to VT, and shifted composition from copiotrophic Proteobacteria toward oligotrophic Acidobacteriota, Actinobacteriota, and Chloroflexi. This trajectory mirrors the known succession of root exudate chemistry—from readily assimilable carbon in early vegetative growth toward more recalcitrant substrates later in development—which reorders microbial substrate preferences and selects for oligotrophic strategists (Sasse et al. 2018; Zhalnina et al. 2018). Strong ontogenetic filtering of root-associated bacteria has likewise been reported in rice and maize field studies (Edwards et al. 2015; Walters et al. 2018), and environmental filtering is a recognized first-order driver of rhizosphere assembly (Philippot et al. 2013). A practical corollary is that temporal context matters: microbial diagnostics and inoculant applications are likely to be most effective when matched to specific developmental windows (Wei et al. 2024).

Fungal communities, by contrast, showed weaker stage discrimination but comparatively stronger genotype-specific structure, with genotype effects at tasseling (R = 0.68–0.74) exceeding the stage effect and with no detectable genotype or stage effect on alpha-diversity. This pattern is consistent with stronger historical contingency and priority effects in fungal assembly, potentially mediated by genotype-dependent root architecture and exudate profiles (Hamonts et al. 2018), and with the partial heritability of rhizosphere microbiome composition documented in maize (Peiffer et al. 2013; Walters et al. 2018), as well as with the environment-dependent genetic control of root microbiomes recently demonstrated for locally adapted maize accessions (He et al. 2024). The cultivar-specific expansion of Basidiomycota in WY2 at VT illustrates that individual genotypes can recruit distinctive fungal consortia. Ecologically, this dichotomy agrees with theory: bacteria, with high dispersal capacity and metabolic plasticity, track contemporary environmental filtering, whereas fungi retain stronger signatures of historical and host-genetic processes. For crop management, the implication is twofold—bacterial communities may be steerable through

agronomic practice within specific developmental windows, whereas fungal communities may respond more to host genetic selection and long-term soil legacy.

Keystone phyla are structural backbones, and robustness validation is what establishes it

Co-occurrence networks are descriptive summaries unless their key nodes are shown to matter (Faust and Raes 2012). A methodological contribution of this study is therefore the coupling of hub detection with two independent robustness analyses. Phylum-level removal showed that Proteobacteria and Ascomycota—the dominant hub phyla of bacterial and fungal networks, respectively—are structural backbones: their deletion eliminated 81.5 % and 93.6 % of edges on average, reduced node counts by 61.6 % and 73.9 %, and eroded modularity by ~16 %, whereas no other phylum exceeded 19 % edge loss or 4 % modularity change. Attack–tolerance simulations independently confirmed scale-free vulnerability of the functional networks, with targeted (degree-ordered) attack collapsing the networks far earlier than random removal. These results move keystone nomination from statistical convenience to validated structural indispensability, in line with the recognized role of highly connected taxa in stabilizing microbial interaction webs (Banerjee et al. 2018) and with trophic-network theory showing that the architecture of root-associated communities determines their beneficial effects on plants (Wei et al. 2015).

The architectural contrast between kingdoms is equally informative. Bacterial networks were dominated by positive associations (74–94 %), consistent with cooperative metabolic complementarity in a carbon-rich rhizosphere, whereas fungal networks carried roughly half negative edges, consistent with competition and antagonism—possibly mediated by antibiotic production or niche overlap—as a structuring force in fungal assembly (Hassani et al. 2018). We stress that co-occurrence edges are statistical associations, not demonstrated biotic interactions (Barberán et al. 2012; Faust and Raes 2012); the robustness results therefore speak to the structural organization of the observed community, and the functional consequences of backbone loss remain to be tested experimentally.

Phenotype–microbiome decoupling at the genotype level

To our knowledge, few maize rhizosphere studies have explicitly tested whether phenotypic differentiation among genotypes tracks community differentiation. Our genotype-level Mantel analysis returned a clear null result: across both kingdoms, both distance metrics, and both stages, none of eight tests was significant after multiple-testing correction. The single significant sample-

level test (bacterial weighted UniFrac at VT, r = 0.383) is best explained by pseudo-replication of genotype-assigned traits, underscoring why the genotype level is the appropriate unit for this question. This null result does not contradict the heritability of rhizosphere microbiome features (Peiffer et al. 2013; Walters et al. 2018): heritability asks whether genotype structures community composition, whereas our test asks whether overall community composition mirrors integrated phenotypic output—and the answer is no, at least at the resolution of bulk community distances with seven genotypes. Recent comparisons of inbred and hybrid maize likewise report that location and tissue effects often outweigh intraspecific genetic background effects on bacterial communities (Schultz et al. 2023), indicating that genotype effects, although real, operate within a strongly context-dependent hierarchy. The decoupling has a direct methodological consequence: attempts to link microbiomes to yield by simple trait–community matching are likely to fail, whereas pathway-level models that route microbial influence through functional guilds and agronomic traits—rather than through bulk composition—can still resolve coherent structure, as our BSEM results demonstrate.

Functional guilds, not bulk taxonomy, form the layer that connects to yield

Predicted functional profiles provided the conceptual bridge between taxonomy and host performance. Bacterial functions were dominated by heterotrophic carbon cycling (chemoheterotrophy and aerobic chemoheterotrophy, mean 30.5 %) and shifted with development toward nitrogen transformation and residue decomposition at tasseling, paralleling the taxonomic succession. Fungal trophic structure—dominated by saprotrophs and pathotrophs with near-absent obligate symbionts—indicates a rhizosphere governed mainly by decomposition and pathogen-pressure processes, consistent with maize's limited dependence on mycorrhizal mutualism under fertilized chernozem conditions. Functional redundancy in central metabolism may buffer bacterial communities against compositional fluctuation while keystone taxa supply non-redundant plant growth-promoting services (Berendsen et al. 2012; Louca et al. 2016); the competition-rich fungal network further suggests that steering fungal communities may require suppressing pathogenic guilds rather than wholesale community engineering (Kinkel et al. 2011). Within the BSEM models, it was precisely the predicted functional guilds—fungal nutrient activation and organic matter decomposition, bacterial nitrogen metabolism—that carried the significant paths into agronomic performance ($\beta = 0.25–0.30$), while bulk phylum abundances acted upstream as drivers of those guilds. Amplicon-based functional prediction is of course approximate (Louca et al. 2016),

and these assignments should be read as hypotheses at the functional layer rather than measured activities.

A Bayesian structural equation framework, with machine learning in a validation—not forecasting—role

The BSEM framework resolved an interpretable chain from source variables through phyla and functional guilds to agronomic traits and measured yield, reproducing about two-thirds of yield variance in the genotype models with excellent fit indices (CFI 0.92–0.95; SRMR ≤ 0.05). Three features of the results deserve emphasis. First, the dominant conduit was the agronomic-trait composite (agronomic traits → yield $\beta = 0.55$), meaning that microbial effects on yield are expressed largely through measurable plant performance rather than through direct chemistry. Second, developmental stage exerted a strong total effect on yield (0.68), with the indirect, microbiome-mediated share increasing at tasseling—consistent with elevated demand for microbially derived nutrients during the reproductive phase. Third, cultivar total effects were reproducible across stages: Nonghua 106 (G3) combined a high measured yield with the most positive total microbial effect (+0.49 to +0.50, roughly half indirect), whereas Dr. King 825 (G5) combined the lowest yield with a consistently negative effect (−0.40 to −0.49). This concordance between measured agronomic ranking and modeled microbial effects suggests that some cultivars possess a heritable capacity to recruit a beneficial microbiome—a trait that could in principle be selected for (Kinkel et al. 2011; Mueller and Sachs 2015; Trivedi et al. 2020).

Equally important is what we did not do. We did not train models to forecast yield, because seven genotype-level observations cannot support predictive validation—our own leave-one-out test returned a negative cross-validated R^2—and reporting predictive accuracy under such conditions is a known pitfall in microbiome machine learning (Topçuoğlu et al. 2020). Machine learning was instead confined to identifying yield-associated signatures and checking their coherence with the pathway model. The two approaches proved complementary rather than redundant: random-forest rankings and correlation screens highlighted ASV-level markers (e.g., *Rhodanobacter*-affiliated ASV1138 at the genotype level; ASV371 at the sample level) that sit largely outside the dominant-phylum layer used by BSEM, which is expected given their different inference targets—fine-grained marker discovery versus pathway estimation (Rodriguez et al. 2019). We regard this

convergence-in-purpose, divergence-in-scale as a strength: it disciplines interpretation and prevents over-claiming from either method alone.

Implications for microbiome-assisted breeding

From an applied perspective, three findings translate into breeding hypotheses. First, the cultivar ranking of total microbial effects identifies G3 (Nonghua 106) as a candidate microbiome-optimized genotype and G5 (Dr. King 825) as a contrasting negative reference; these two lines are natural parents for dissecting the genetic basis of microbiome recruitment (Mueller and Sachs 2015; Peiffer et al. 2013). Second, the validated backbone taxa (Proteobacteria, Ascomycota) and the top yield-associated genera (*Variovorax*, *Pseudarthrobacter*, *Candidatus Udaeobacter*, *Bacillus*, *Mucilaginibacter*, *Pseudomonas*, *Sphingobium*, Ellin6067) define a priority list for strain isolation; our in silico SynCom candidate list is explicitly a design hypothesis to be tested by cultivation, in vitro assays (nitrogen fixation, phosphate solubilization, IAA and ACC-deaminase activity; Andy et al. 2023; Dutta et al. 2022; Wang et al. 2023; Zeng et al. 2023), and staged inoculation experiments, not a field-ready product (Singh et al. 2024). Third, the stage dependence of both assembly and modeled effects argues for stage-matched microbial interventions rather than generic formulations (Wei et al. 2024). Closed-loop validation—from sequence-based hypotheses to cultivation, controlled inoculation, and metagenomic or isotope-based confirmation—will be essential before microbiome-assisted breeding can move from framework to field (Hassani et al. 2018; Rodriguez et al. 2019).

Limitations

Several limitations bound the interpretation of this study. First, seven genotypes at a single site in one growing season cannot represent maize diversity or environmental generality; expansion to larger genotype panels and multi-location trials is needed. Second, only two developmental stages were sampled, so the temporal resolution of assembly dynamics is coarse. Third, no molecular-marker-based genetic distances were available, precluding direct tests of phylogenetic signal in microbiome recruitment. Fourth, co-occurrence networks reflect statistical association, not measured interaction (Faust and Raes 2012), and functional profiles are predictions from amplicon data (Louca et al. 2016; Nguyen et al. 2016) that await metagenomic or metatranscriptomic confirmation. Fifth, BSEM resolves statistical pathways under an a priori structure and cannot exclude confounding by unmeasured soil or environmental variables (Rosseel 2012); its paths are

hypotheses with directional support, not experimental demonstrations. Finally, the kingdoms differ in sequencing depth, database completeness, and trait annotation coverage, which may contribute to the bacterial–fungal contrasts reported here.

Conclusions

This study integrated kingdom-resolved diversity analysis, robustness-validated co-occurrence networks, functional annotation, a Bayesian structural equation framework, and disciplined machine-learning signature identification across seven maize cultivars and two developmental stages, leading to three principal conclusions.

(1) Bacteria and fungi obey kingdom-specific assembly rules: bacterial rhizosphere communities are filtered primarily by plant developmental stage, shifting from copiotrophic to oligotrophic dominance, whereas fungal communities retain comparatively stronger genotype-specific structure, especially at tasseling.

(2) Co-occurrence networks are organized around Proteobacteria and Ascomycota backbones whose simulated removal eliminates 81–94 % of network edges and erodes modularity by ~16 %, establishing their structural indispensability; at the same time, bulk community composition is decoupled from phenotypic distances at the genotype level, so pathway-level modeling—not direct trait–community matching—is required to connect microbiomes to yield.

(3) The Bayesian structural equation framework reproduced about two-thirds of measured yield variance and resolved an interpretable route from microbial phyla through predicted functional guilds and agronomic traits to grain yield, with cultivar Nonghua 106 showing the strongest positive total microbial effect and Dr. King 825 a consistently negative one. Machine learning, explicitly barred from forecasting, supplied complementary yield-associated signatures and an in silico SynCom candidate list. Together these results provide a reproducible analytical framework and testable hypotheses for microbiome-assisted maize breeding.

Acknowledgements

This work was supported by two funding projects: (1) the Training Program of Innovation and Entrepreneurship for Undergraduates (Grant No. 202610129037); (2) the earmarked fund for Inner Mongolia Agriculture and Animal Husbandry Research System (IMAHRS-16).

Author contributions

B.Y.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. Z.G.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. F.S.: Investigation, Data curation. X.H.: Investigation, Data curation. Q.L.: Investigation, Resources. Y.Y.: Investigation, Resources. T.H.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing. G.W.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing.

Declarations

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Raw 16S rRNA (V3-V4) and ITS (ITS1) amplicon sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1504823 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1504823). The dataset comprises 42 rhizosphere soil BioSamples and 84 sequencing runs, which are accessible through the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra/PRJNA1504823). Analysis scripts are available from the corresponding authors upon reasonable request. All other data supporting the findings of this study are included in the article and its supplementary information files.

References

Andy AK, Ghosh S, Pareek A, Sopory SK, Singla-Pareek SL (2023) Exploring the identity and properties of two Bacilli strains and their potential to alleviate drought and heavy metal stress. Horticulturae 9:46. https://doi.org/10.3390/horticulturae9010046

Banerjee S, Schlaeppi K, van der Heijden MGA (2018) Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol 16:567–576. https://doi.org/10.1038/s41579-018-0024-1

Barberán A, Bates ST, Casamayor EO, Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6:343–351. https://doi.org/10.1038/ismej.2011.119

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the Third International AAAI Conference on Weblogs and Social Media 3(1):361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Berendsen RL, Pieterse CMJ, Bakker PAHM (2012) The rhizosphere microbiome and plant health. Trends Plant Sci 17:478–486. https://doi.org/10.1016/j.tplants.2012.04.001

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG

(2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, Peplies J, Gloeckner FO, Amann R, Eickhorst T, Schulze-Lefert P (2012) Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature 488:91–95. https://doi.org/10.1038/nature11336

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–794. https://doi.org/10.1145/2939672.2939785

Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Systems 2006:1695

Das PP, Singh KR, Nagpure G, Mansoori A, Singh RP, Ghazi IA, Kumar A, Singh J (2022) Plant-soil-microbes: a tripartite interaction for nutrient acquisition and better plant growth for sustainable agricultural practices. Environ Res 214:113821. https://doi.org/10.1016/j.envres.2022.113821

Dutta A, Bausal S, Yadav S, Raghav S, Kumar S, Chaudhary A, Rani A, Kumar R, Gogoi HK, Devi R, Bhardwaj A, Sahoo CK, Gupta AK (2022) A critical analysis on the roles of exopolysaccharides and ACC deaminase in salinity stress tolerance in crop plants. Biocatal Agric Biotechnol 42:102372. https://doi.org/10.1016/j.bcab.2022.102372

Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci USA 112:E911–E920. https://doi.org/10.1073/pnas.1414592112

Faust K, Raes J (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550. https://doi.org/10.1038/nrmicro2832

Favela A, Bohn MO, Kent AD (2025) Genetic variation in Zea mays influences microbial nitrification and denitrification in conventional agroecosystems. Plant Soil 506:671–690. https://doi.org/10.1007/s11104-024-06720-9

Feng D, Wang Z, Li Y, Zhang Y, Wang H, Lu Y, Gao Y (2025) Mechanisms of microorganisms alleviating drought and salt stresses in plants. Microorganisms 13:2565. https://doi.org/10.3390/microorganisms13112565

Friedman J, Alm EJ (2012) Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687

Garrido-Oter R, Nakano RT, Dombrowski N, Ma KW, AgBiome Team, McHardy AC, Schulze-Lefert P (2018) Modular traits of the Rhizobiales root microbiota and their evolutionary relationship with symbiotic Rhizobia. Cell Host Microbe 24:155–167.e5. https://doi.org/10.1016/j.chom.2018.06.006

Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, Botha FC, Anderson IC, Singh BK (2018) Field study reveals core plant microbiota and relative importance of their drivers. Environ Microbiol 20:124–140. https://doi.org/10.1111/1462-2920.14031

Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. Microbiome 6:58. https://doi.org/10.1186/s40168-018-0445-0

He X, Wang D, Jiang Y, Li M, Delgado-Baquerizo M, McLaughlin C, Marcon C, Guo L, Baer M, Moya YAT, von Wirén N, Deichmann M, Schaaf G, Piepho HP, Yang Z, Yang J, Yim B, Smalla K, Goormachtig S, de Vries FT, Hüging H, Baer M, Sawers RJH, Reif JC, Hochholdinger F, Chen X, Yu P (2024) Heritable microbiome variation is correlated with source environment in locally adapted maize varieties. Nat Plants 10:598–617. https://doi.org/10.1038/s41477-024-01654-7

Kinkel LL, Bakker MG, Schlatter DC (2011) A coevolutionary framework for managing disease-suppressive soils. Annu Rev Phytopathol 49:47–67. https://doi.org/10.1146/annurev-phyto-072910-095232

Li K, DiLegge MJ, Minas IS, Hamm A, Manter D, Vivanco JM (2024) Long-term chemical and organic fertilization induces distinct variations of microbial associations but unanimous elevation of soil multifunctionality. Sci Total Environ 931:172862. https://doi.org/10.1016/j.scitotenv.2024.172862

Liu-Xu L, Paniagua-Vega D, Li H, San Segundo B, López-Moya F, Lopez-Llorca LV (2024) Harnessing green helpers: nitrogen-fixing bacteria and other beneficial microorganisms in plant-microbe interactions for sustainable agriculture. Horticulturae 10:621. https://doi.org/10.3390/horticulturae10060621

Louca S, Parfrey LW, Doebeli M (2016) Decoupling function and taxonomy in the global ocean microbiome. Science 353:1272–1277. https://doi.org/10.1126/science.aaf4507

Merkle EC, Rosseel Y (2018) blavaan: Bayesian structural equation models via parameter expansion. J Stat Softw 85:1–30. https://doi.org/10.18637/jss.v085.i04

Mueller UG, Sachs JL (2015) Engineering microbiomes to improve plant and animal health. Trends Microbiol 23:606–617. https://doi.org/10.1016/j.tim.2015.07.009

Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG (2016) FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecol 20:241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, Buckler ES, Ley RE (2013) Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proc Natl Acad Sci USA 110:6548–6553. https://doi.org/10.1073/pnas.1302837110

Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH (2013) Going back to the roots: the microbial ecology of the rhizosphere. Nat Rev Microbiol 11:789–799. https://doi.org/10.1038/nrmicro3109

Ren H, Wang J, Liu J, Xu L, Yan J, Zhang X, Liu P, Zhao B, Zhang J (2022) Effects of agronomic traits and climatic factors on yield and yield stability of summer maize (Zea mays L.) in the Huang-Huai-Hai Plain in China. Front Plant Sci 13:1050064. https://doi.org/10.3389/fpls.2022.1050064

Rodriguez PA, Rothballer M, Chowdhury SP, Nussbaumer T, Gutjahr C, Falter-Braun P (2019) Systems biology of plant-microbiome interactions. Mol Plant 12:804–821. https://doi.org/10.1016/j.molp.2019.05.006

Rosseel Y (2012) lavaan: an R package for structural equation modeling. J Stat Softw 48:1–36. https://doi.org/10.18637/jss.v048.i02

775 Sagar A, Rai S, Iliyas M, Sayyed RZ, Al-Turki IA, El Enshasy HA, Simarmata T (2021) Plant
776 growth promoting rhizobacteria, arbuscular mycorrhizal fungi and their synergistic
777 interactions to counteract the negative effects of saline soil on agriculture. Microorganisms
778 9:1491. https://doi.org/10.3390/microorganisms9071491

779 Sasse J, Martinoia E, Northen T (2018) Feed your friends: do plant exudates shape the root
780 microbiome? Trends Plant Sci 23:25–41. https://doi.org/10.1016/j.tplants.2017.09.003

781 Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C (2011)
782 Metagenomic biomarker discovery and explanation. Genome Biol 12:R60.
783 https://doi.org/10.1186/gb-2011-12-6-r60

784 Schultz CR, Johnson M, Wallace JG (2023) Effects of inbreeding on microbial community
785 diversity of Zea mays. Microorganisms 11:879.
786 https://doi.org/10.3390/microorganisms11040879

787 Singh D, Mathur N, Sharma R, Sharma P, Jangid K, Paliwal R, Tak N, Gehlot HS (2024)
788 Formulations of synergistic microbial consortia for enhanced systemic resistance against
789 Fusarium wilt in cumin. Int Microbiol 28:497–523. https://doi.org/10.1007/s10123-024-
790 00553-3

791 Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD (2020) A framework for effective
792 application of machine learning to microbiome-based classification problems. mBio
793 11:e00434-20. https://doi.org/10.1128/mBio.00434-20

794 Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK (2020) Plant-microbiome interactions: from
795 community assembly to plant health. Nat Rev Microbiol 18:607–621.
796 https://doi.org/10.1038/s41579-020-0412-1

797 United Nations, Department of Economic and Social Affairs, Population Division (2022) World
798 Population Prospects 2022: Summary of Results. UN DESA/POP/2022/TR/NO.3. United
799 Nations, New York

800 Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A (2015) The importance of
801 the microbiome of the plant holobiont. New Phytol 206:1196–1206.
802 https://doi.org/10.1111/nph.13312

803 Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, González-Peña A, Peiffer J,
804 Koren O, Shi Q, Knight R, Gonzalez del Rio T, Tringe SG, Buckler ES, Dangl JL, Ley RE

(2018) Large-scale replicated field study of maize rhizosphere identifies heritable microbes. Proc Natl Acad Sci USA 115:7368–7373. https://doi.org/10.1073/pnas.1800918115

Wang W, Guo Y, Xu J, Liu W, Liu C, Wei Y, Zhao X, Liu X (2023) Soil P solubilization and plant growth promotion by a saline-alkali-tolerant P-solubilizing bacterium, Bacillus sp. DYS211. J Plant Ecol 16:rtad028. https://doi.org/10.1093/jpe/rtad028

Wei X, Lu Y, Zhang H, Zhang T, Li J, Liu Y, Chen J (2024) Enhancing soil health and plant growth through microbial fertilizers: mechanisms, benefits, and sustainable agricultural practices. Agronomy 14:609. https://doi.org/10.3390/agronomy14030609

Wei Z, Yang T, Friman VP, Xu Y, Shen Q, Jousset A (2015) Trophic network architecture of root-associated bacterial communities determines pathogen invasion and plant health. Nat Commun 6:8413. https://doi.org/10.1038/ncomms9413

Zeng Q, Wang Y, Li Y, Chen X, Zhang J (2023) Isolation and characterization of phosphate-solubilizing bacteria from rhizosphere of poplar on road verge and their antagonistic potential against various phytopathogens. BMC Microbiol 23:221. https://doi.org/10.1186/s12866-023-02953-3

Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP, Firestone MK, Northen TR, Brodie EL (2018) Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nat Microbiol 3:470–480. https://doi.org/10.1038/s41564-018-0129-3

Supplementary Material

Table S1. Details and codes of experimental materials. Cultivar codes G1–G7 are used in the BSEM models; WY prefixes are used in sequencing sample IDs.

Item	Jinongyu 198	Tiannong 9	Nonghua 106	Xianyu 335	Dr. King 825	Ruifeng 105	Jinyuan Jade 177
Seedling stage (V6)	WY2.627	WY3.627	WY8.627	WY9.627	WY13.627	WY14.627	WY15.627
Tasseling stage (VT)	WY2.728	WY3.728	WY8.728	WY9.728	WY13.728	WY14.728	WY15.728
BSEM group	G1	G2	G3	G4	G5	G6	G7

Table S2. Community differentiation (ANOSIM and PERMANOVA, 9,999 permutations) on Bray–Curtis and weighted UniFrac distances. Stage was tested across all samples; genotype was tested within each stage.

Kingdom	Distance	Factor (subset)	ANOSIM R	ANOSIM P	PERMANOVA F	PERMANOVA P
Bacteria	Bray–Curtis	Stage (all)	0.322	0.0002	1.261	0.0002
Bacteria	Bray–Curtis	Genotype (V6)	0.095	0.1858	0.980	0.7244
Bacteria	Bray–Curtis	Genotype (VT)	0.470	0.0004	1.023	0.0370
Fungi	Bray–Curtis	Stage (all)	0.505	0.0002	11.178	0.0002
Fungi	Bray–Curtis	Genotype (V6)	0.393	0.0002	1.874	0.0002
Fungi	Bray–Curtis	Genotype (VT)	0.738	0.0002	4.481	0.0002
Bacteria	Weighted UniFrac	Stage (all)	0.901	0.0002	32.319	0.0002
Bacteria	Weighted UniFrac	Genotype (V6)	0.417	0.0004	2.324	0.0004
Bacteria	Weighted UniFrac	Genotype (VT)	0.838	0.0002	3.404	0.0002
Fungi	Weighted UniFrac	Stage (all)	0.421	0.0002	13.095	0.0002

Fungi	Weighted UniFrac	Genotype (V6)	0.306	0.0006	1.986	0.0006
Fungi	Weighted UniFrac	Genotype (VT)	0.685	0.0002	4.871	0.0002

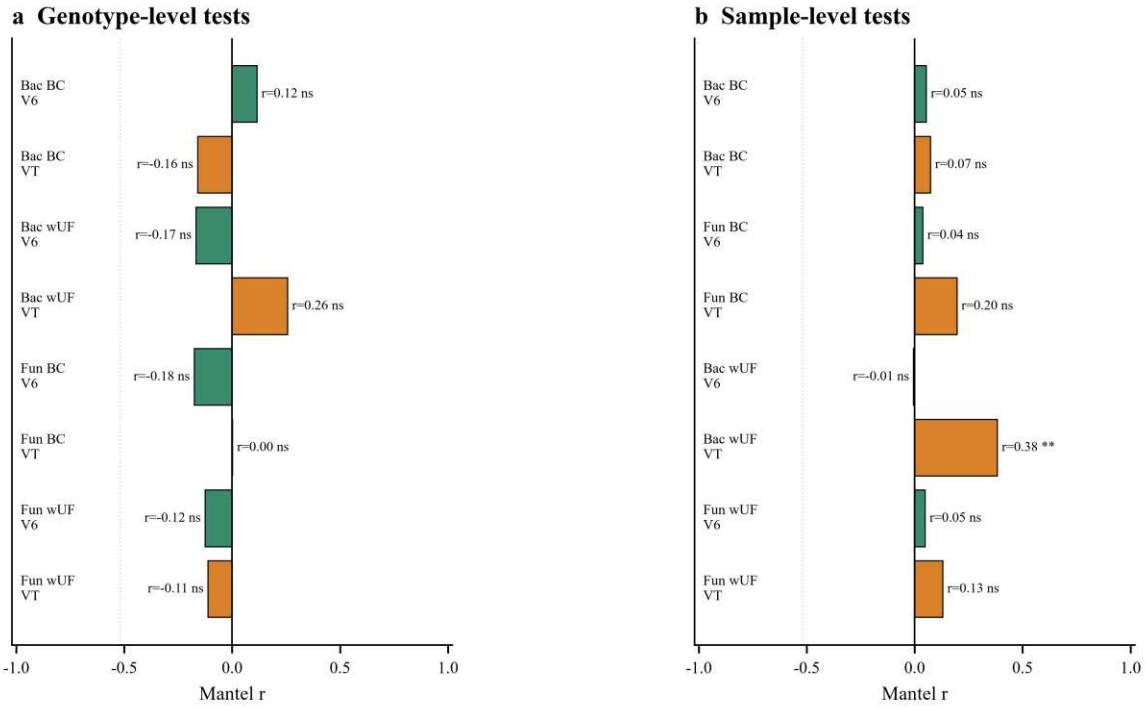


Fig. S1. Mantel tests between agronomic-trait and community distance matrices (9,999 permutations). (a) Genotype-level tests on 7 × 7 mean matrices (primary analysis; 8 tests). (b) Sample-level tests with genotype-assigned traits (secondary analysis; 8 tests). Points show Mantel's r with Benjamini–Hochberg-adjusted P values across all 16 tests; the dashed line marks r = 0. Only the sample-level bacterial wUF test at VT remained significant after adjustment (r = 0.383, adjusted P = 0.0016), a pattern attributed to pseudo-replication (see the first Results subsection).

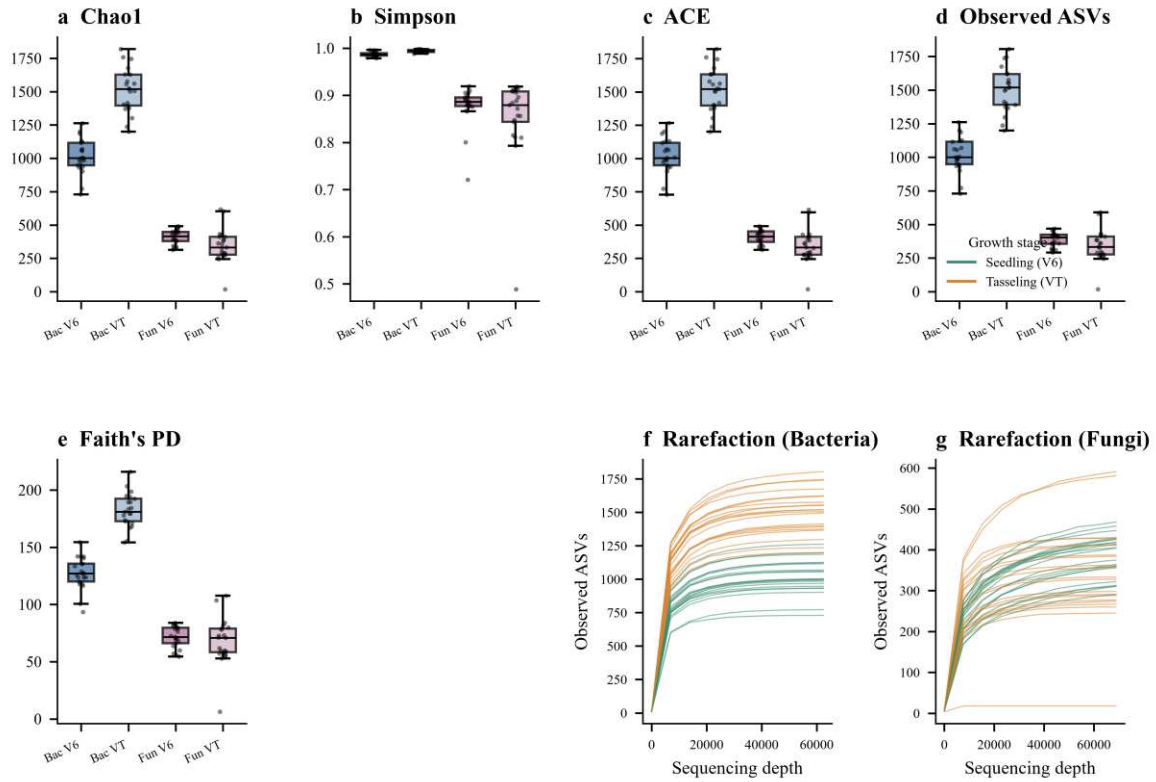


Fig. S2. Alpha-diversity indices and rarefaction. (a–e) Chao1, Simpson, ACE, observed features, and Faith's phylogenetic diversity for bacteria and fungi at V6 and VT (boxplots with per-sample points; Welch's t-test between stages). (f, g) Rarefaction curves of observed features for bacteria and fungi (mean per cultivar × stage group; legend in panel d). Bacterial indices increased strongly at VT; fungal indices showed no significant stage difference.

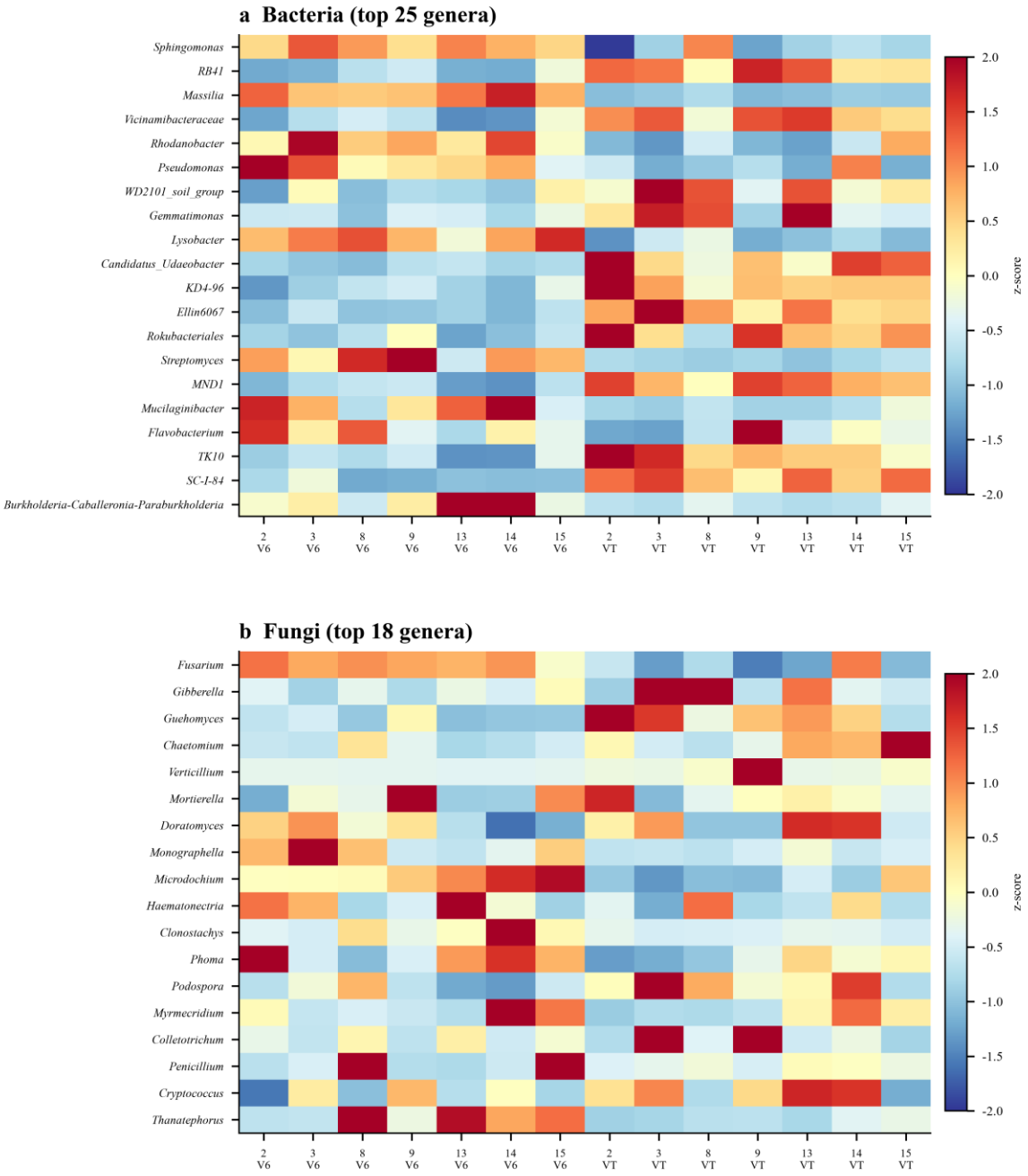


Fig. S3. Genus-level community composition. Heatmap of row z-scored relative abundances for the top 25 bacterial and top 18 fungal genera across cultivar × stage groups (genus names in italics; unclassified taxa labeled at the lowest resolved rank).

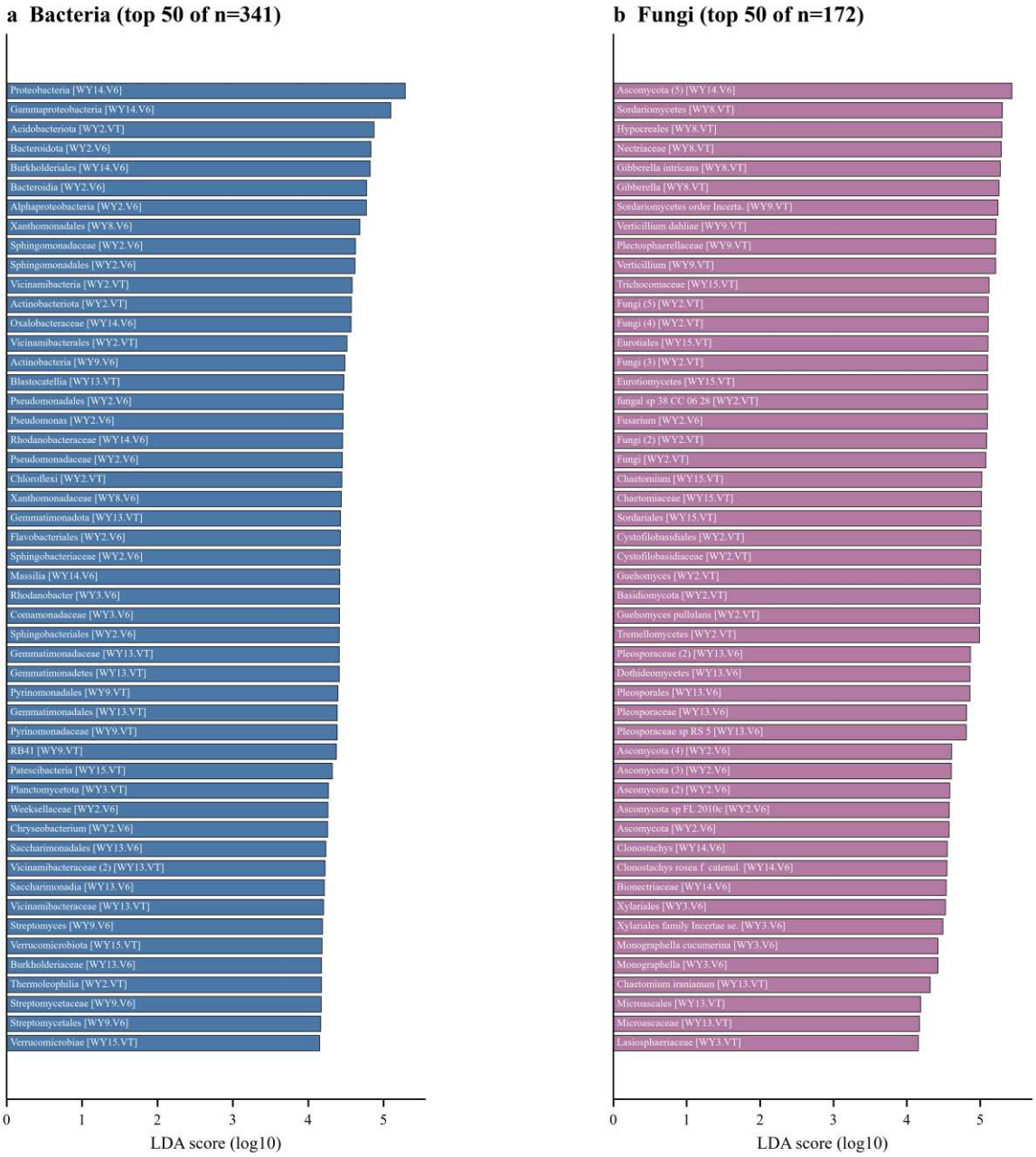


Fig. S4. LEfSe differentially abundant features (LDA > 3.0), top 50 per kingdom. (a) Bacteria (top 50 of 341 significant features). (b) Fungi (top 50 of 172). Bars show LDA scores; group membership is given in brackets (e.g., [WY14;V6]). Numerical suffixes, e.g., "Fungi (3)", distinguish biomarkers reported at multiple taxonomic ranks that resolve to the same display name.

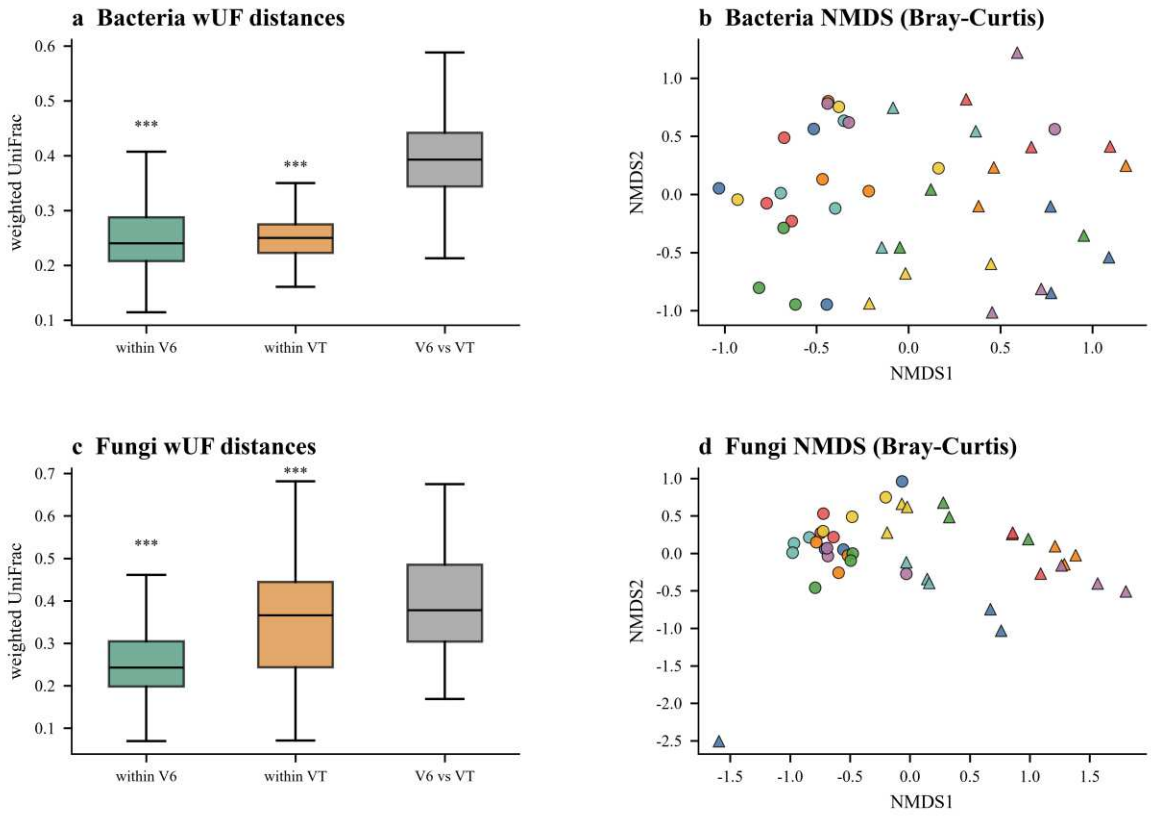


Fig. S5. Weighted UniFrac distance distributions and NMDS ordinations. (a, b) Within-stage and between-stage wUF distances for bacteria and fungi (Mann–Whitney U test; *P < 0.05, **P < 0.01, ***P < 0.001). (c, d) NMDS ordinations on Bray–Curtis distances for bacteria and fungi (stress values indicated).



Fig. S6. All 28 cultivar × stage × kingdom co-occurrence networks (50 nodes each), arranged by kingdom and stage (rows) and cultivar (columns); nodes colored by phylum, edges by sign (green, positive; red, negative). The bar chart summarizes the number of detected modules per network (three modules in 27 of 28 networks).

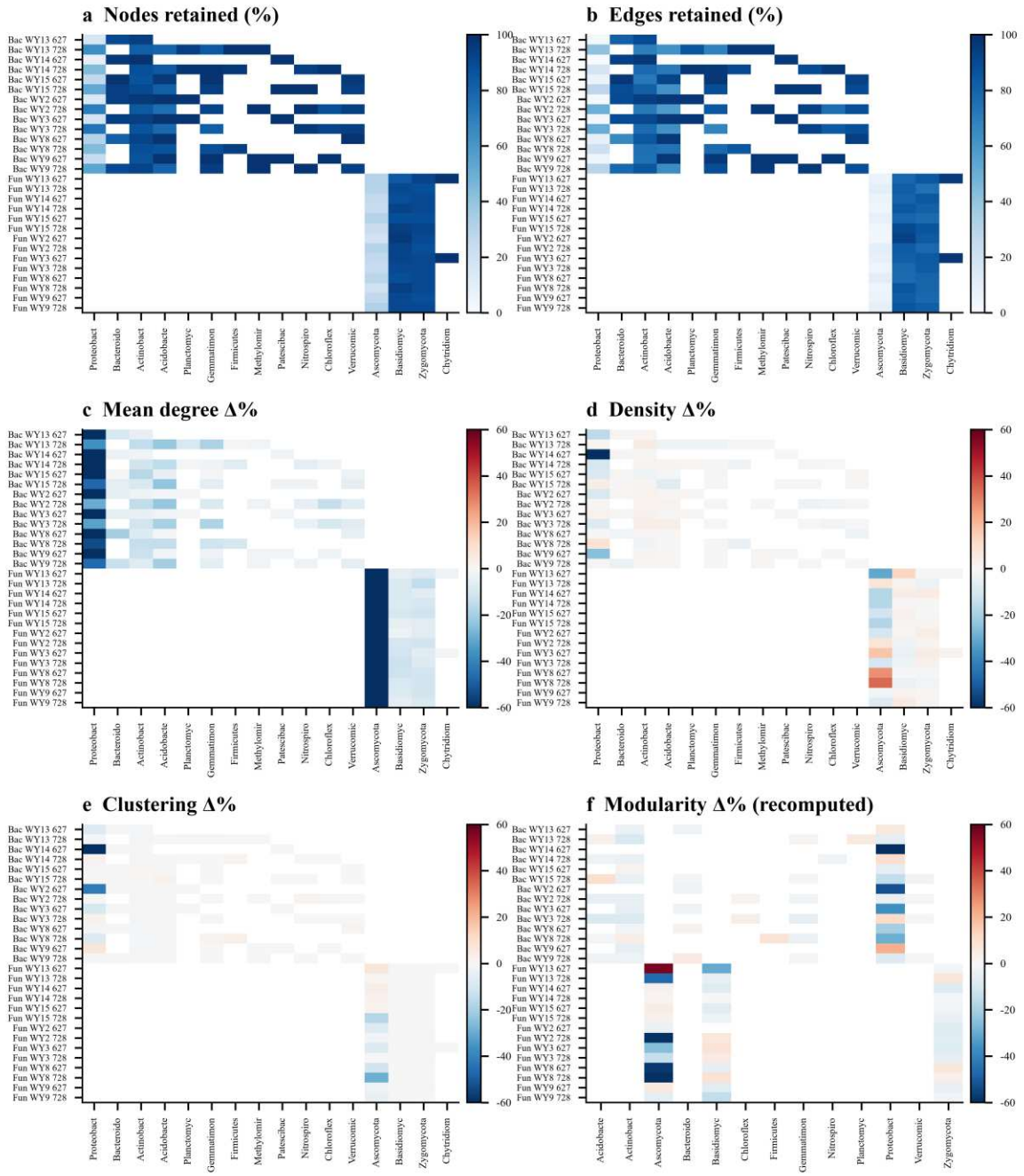


Fig. S7. Network robustness across six metrics after phylum-level removal. (a–e) Percentage of nodes retained, edges retained, and percentage changes in average degree, density, and clustering coefficient across all 28 networks. (f) Percentage change in modularity recomputed de novo from edge tables (only removals of phyla with ≥ 3 nodes shown).

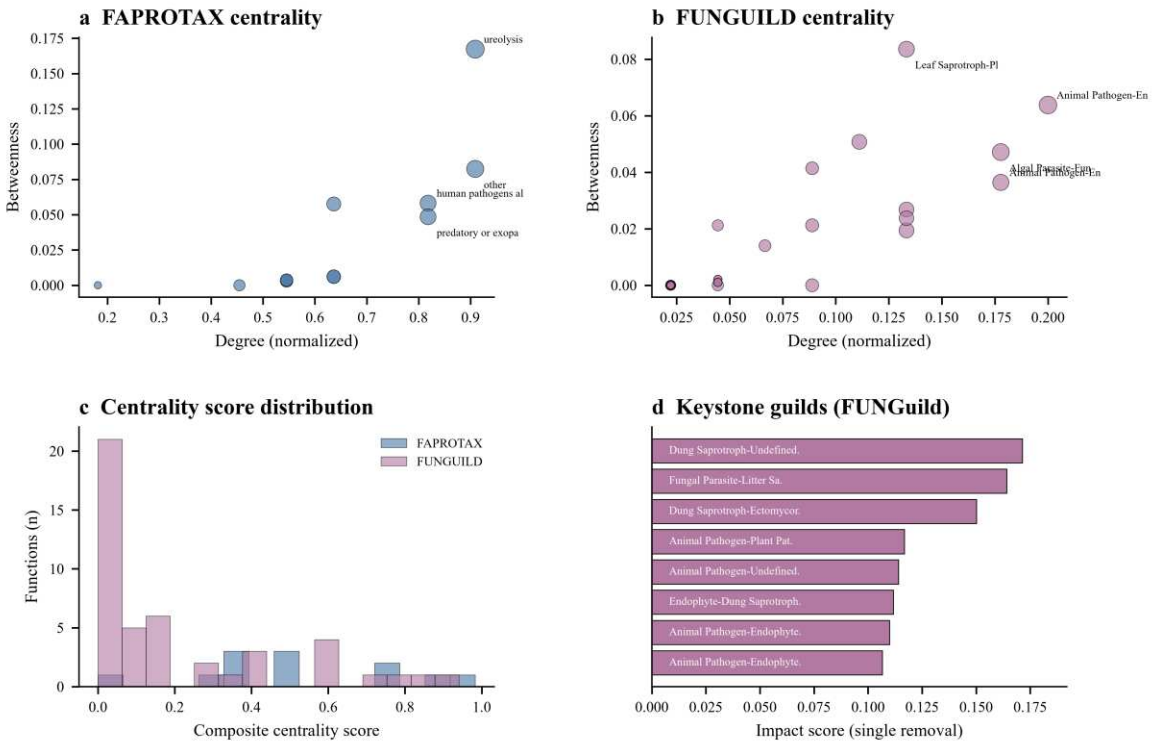
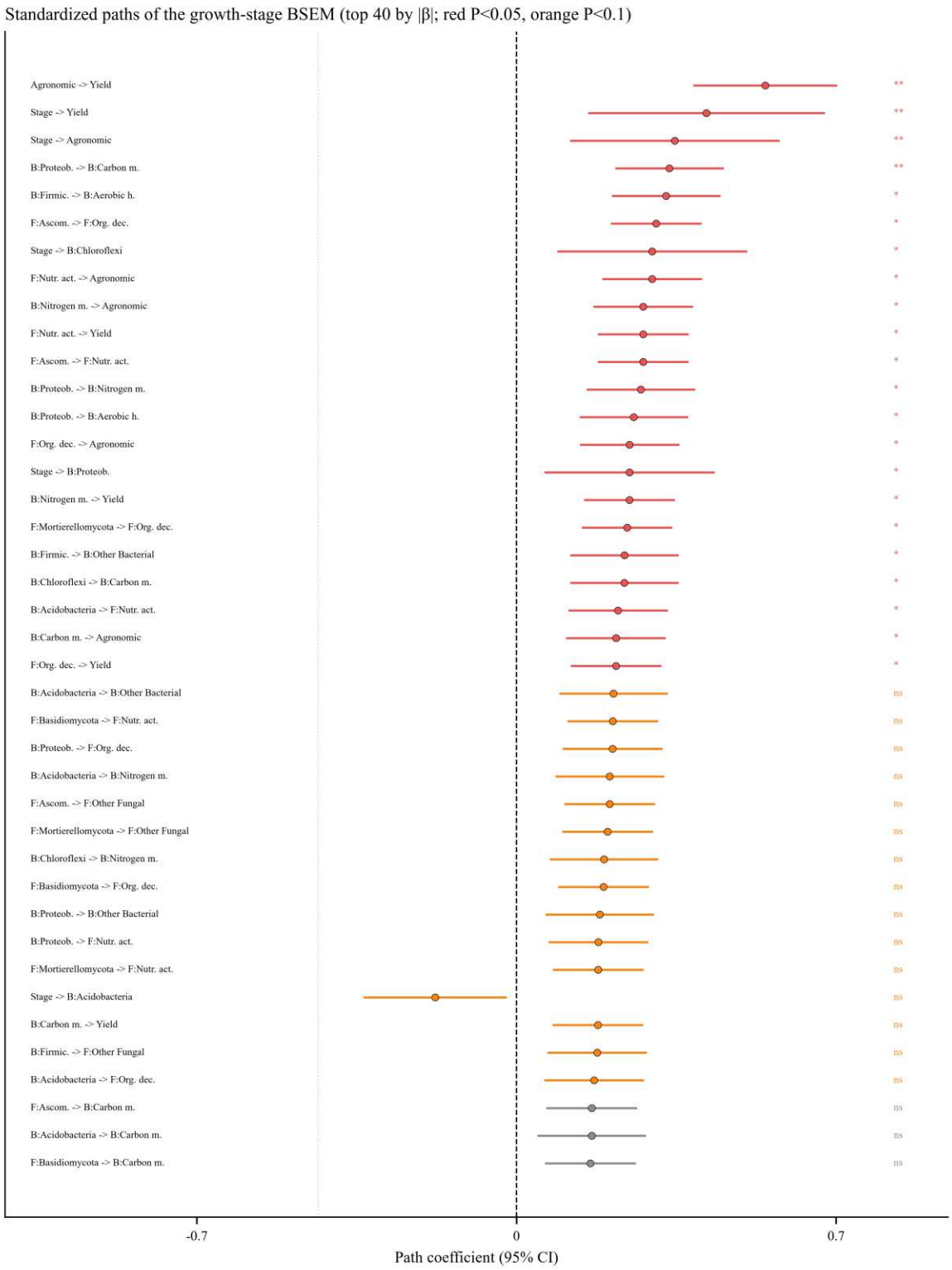


Fig. S8. Centrality and single-removal diagnostics of functional co-occurrence networks. (a) Degree versus betweenness centrality for bacterial functional groups (top four labeled). (b) Distribution of centrality scores. (c) Degree versus betweenness for fungal guilds. (d) Impact scores of single-guild removals in the fungal functional network.



867

868 **Fig. S9.** Stage-model BSEM paths. Standardized coefficients (β with 95 % credible intervals) of the 40 largest paths in the pooled
869 developmental-stage model (N = 42, 73 paths). B:, bacterial; F:, fungal; Nutr. act., nutrient activation; Org. dec., organic
870 decomposition; Carbon m., carbon metabolism; Nitrogen m., nitrogen metabolism; Aerobic h., aerobic heterotrophy.

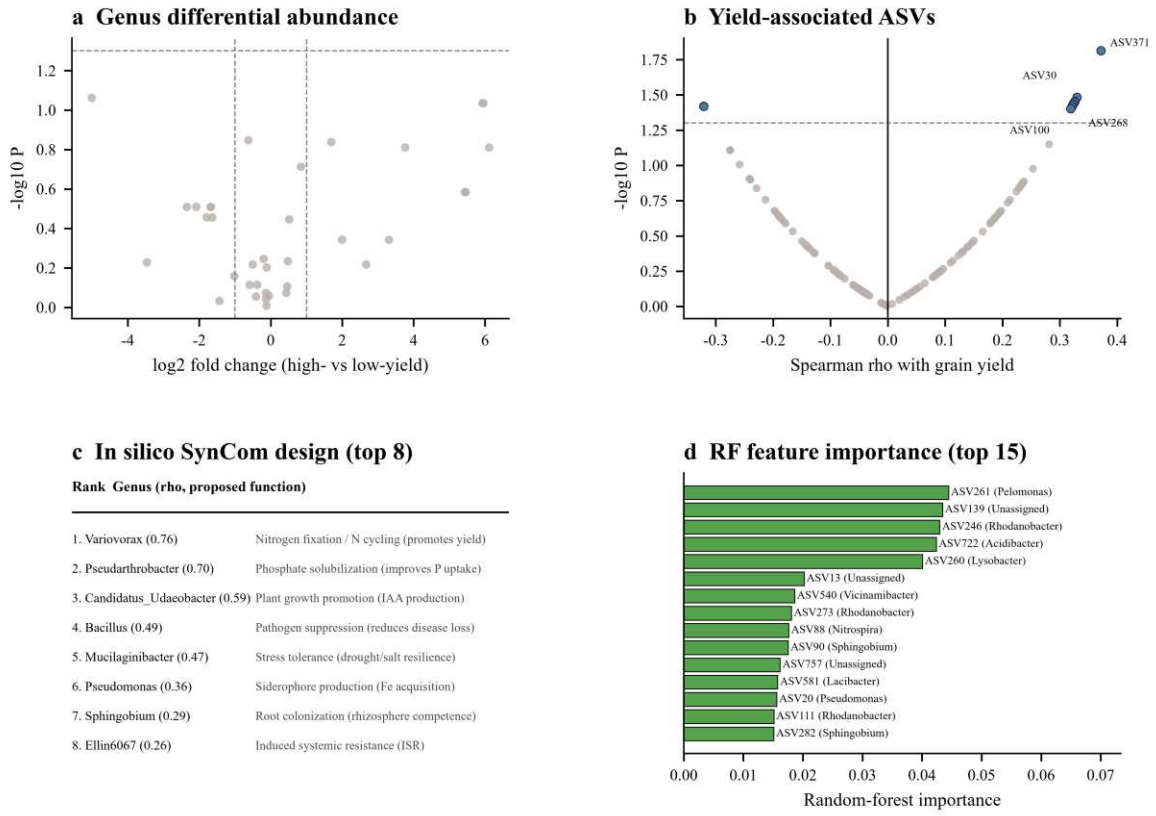


Fig. S10. Machine-learning signature identification (no forecasting performed). (a) Volcano plot of genus-level differential abundance between high- and low-yield cultivar groups; no genus passed the significance thresholds (dashed lines: P = 0.05, |log2 fold change| = 1), a genuine null result. (b) Yield-associated ASVs (Spearman ρ with measured yield versus −log10 P); the four strongest correlates are labeled. (c) In silico SynCom candidate design (top eight genera by positive yield association, with proposed functions; hypothesis-generating only). (d) Random-forest feature importance (top 15 ASVs; in-sample ranking).