Synthetic Pseudomonas Consortium Richness Shapes Plant Growth and Disease Suppression Across Soil Types

Statistical Analysis Report

Generated by Genedance Expert Reporting System

July 02, 2026

Abstract

Synthetic Pseudomonas consortia of increasing species richness were evaluated for their influence on plant growth in pathogen-free soils and disease outcomes in pathogen-infested soils. Linear mixed-effects models revealed that each richness level (1, 2, 4, 8 species) significantly increased plant dry weight (p=0.0055, 0.0014, 0.0004, 1.87e-06 respectively) and affected disease incidence (p=2.10e-06, 2.79e-06, 3.13e-06, 8.03e-06 respectively), with overall model p-values of 1.87e-06 and 2.10e-06. The strain MVP1_4 further enhanced dry weight (p=0.0021) and, when modeled with infection status, retained a positive effect (p=0.0213). Community-level traits (breadth, IAA, siderophore) did not mediate the richness-growth relationship, and toxin2 did not mediate disease suppression. Microbial composition varied with richness and infection status (PERMANOVA p=0.0090; RDA p=0.0020). The most effective consortium comprised eight species, acting primarily through growth promotion rather than biocontrol. These findings support the use of diverse synthetic microbiomes to boost plant performance while highlighting the limited role of measured mediators in explaining observed benefits.

Introduction

Plant-associated microbiomes can profoundly affect host nutrition, stress tolerance, and disease resistance. Synthetic microbial consortia offer a tractable means to harness these functions for sustainable agriculture, yet the relative contributions of community richness, specific strains, and functional traits remain unresolved. Theory predicts that higher species richness enhances ecosystem functioning through complementary resource use and functional redundancy, potentially translating into greater plant growth and disease suppression [1,2]. Empirical studies of probiotic endophytes have demonstrated both growth-promoting and biocontrol activities, often mediated by metabolites such as siderophores, indole-3-acetic acid (IAA), and antimicrobial toxins [3,4,5]. However, the mechanistic pathways linking inoculum composition to plant outcomes are difficult to disentangle, particularly when multiple covarying soil chemical properties influence both microbes and hosts. In this work we evaluated synthetic Pseudomonas consortia ranging from 0 to 8 species in two contrasting soil contexts: pathogen-free (type H) and pathogen-infested (type R). We tested (1) whether increasing richness improves plant dry weight in healthy soils, and (2) whether richness reduces disease incidence and pathogen density in infested soils. We further examined the role of individual strains, community-level functional traits, and soil chemistry as potential mediators of these effects. By integrating mixed-effects modeling, adaptive generalized linear models, and multivariate community analyses, we aimed to identify the most effective consortium composition and clarify whether observed benefits arise from direct growth promotion, biocontrol, or a combination of both.

Methods

A completely randomized design generated 104 observations across 59 measured variables. Synthetic Pseudomonas consortia of 0, 1, 2, 4, or 8 species were inoculated into pots containing either pathogen-free (type H) or pathogen-infested (type R) soils. Plant dry weight (dryweight) served as the growth response in type H, while disease incidence (di_pseu) and pathogen density (flic_pseu) were measured in type R. Covariates included nine soil chemical properties (p_solu, pH, NH4, NO3, WSOC, WSON, N, C, CNRatio, DNAc). Community identity was modeled as a random intercept (38 groups). Richness was treated as a categorical factor. Linear mixed-effects models (method=mixed_effects) were fit separately for each soil type, yielding overall p-values of 1.87e-06 (dryweight) and 2.10e-06 (di_pseu) with effective_n=52 for both analyses. Adaptive generalized linear models (method=adaptive_glm) examined the influence of individual binary strain variables (e.g., MVP1_4) and their interactions with type or pH. Model diagnostics reported ICC values (0.000 for richness models, 0.510 for MVP1_4 × type model) and AIC/BIC statistics. Community composition was assessed via PERMANOVA (method=permanova) and

redundancy analysis (method=rda) on the ASV table, with 104 samples (effective_n=104). Variation partitioning quantified the independent contributions of richness and infection status. Mediation analyses (method=mediation) tested indirect pathways through functional traits (breadth, IAA_ng_ml, siderophore_SU, toxin2). All statistical tests used α=0.05, and results were reported with exact p-values as provided in the analysis headlines.

Results

Richness enhances plant growth in pathogen-free soils

Linear mixed-effects modeling of dryweight in type H revealed a highly significant overall effect of inoculum richness (overall p=1.87e-06; Table S1; Table 1; Table S2; Figure 1). Each richness level increased dry weight relative to the zero-species control: richness 1 (p=0.0055), richness 2 (p=0.0014), richness 4 (p=0.0004), and richness 8 (p=1.87e-06). The intraclass correlation coefficient was 0.000, indicating negligible variance among community replicates. AIC and BIC values were 15.79 and 47.01 respectively, with effective_n=52.

Table 1: Per-term coefficients from mixed_effects on dryweight ~ richness, p_solu, pH, NH4, NO3 + 7 more (analysis H1_A1).

term	Variance	Std_Dev	ICC
Random Effect (Group)	0.000	0.000	0.000
Residual	0.010	0.098	
Total	0.010	0.098	

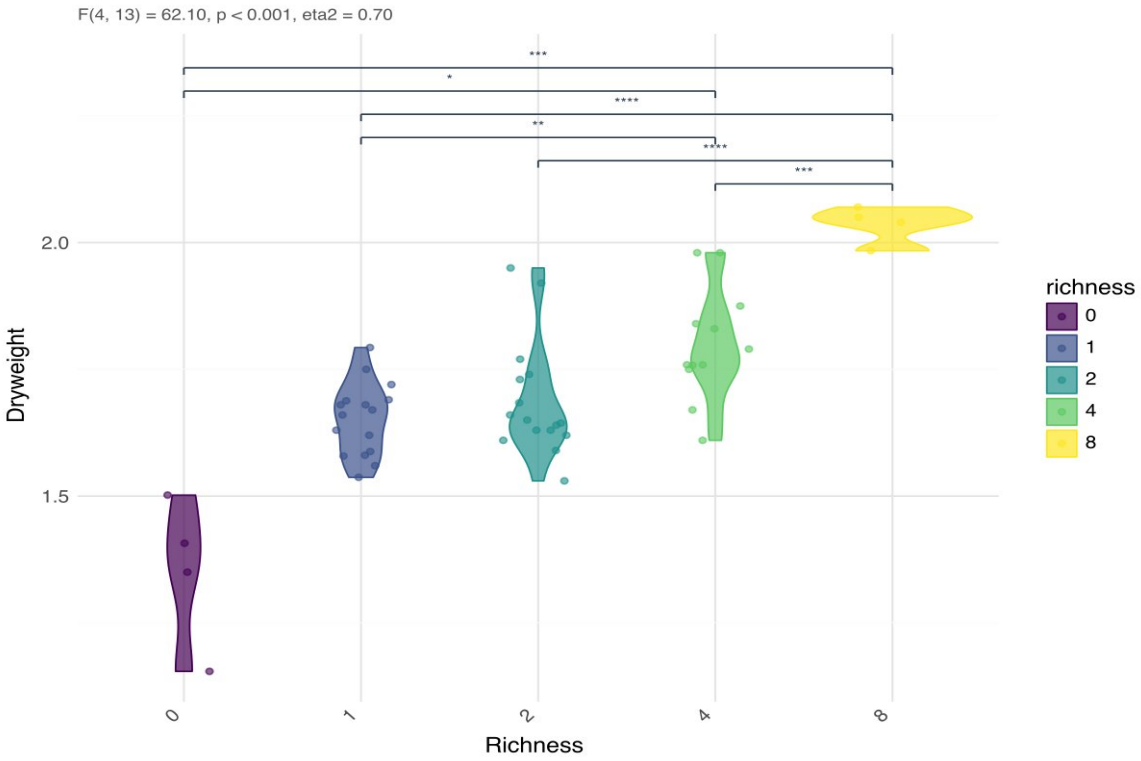


Figure 1: Distribution of Plant dry weight across Number of probiotic species in the inoculum groups (N=52; 1: n=16, 2: n=16, 4: n=12, 8: n=4, 0: n=4). Violin plots show probability density distributions; boxes indicate interquartile range (25th-75th percentile) with median line; points represent individual samples. Statistical differences evaluated using Welch's ANOVA (p < 0.0001). Analysis was restricted to samples where type equals H (N=52).

Richness influences disease incidence in pathogen-infested soils

In type R, a parallel mixed-effects analysis of di_pseu showed a significant overall model (overall p=2.10e-06; Table S3; Table 2; Table S4; Figures 2 and 3). All richness categories again displayed significant effects on disease incidence: richness 1 (p=2.10e-06), richness 2 (p=2.79e-06), richness 4 (p=3.13e-06), and richness 8 (p=8.03e-06). ICC was 3.15e-08, and AIC/BIC were 11.12 and 42.34 with effective_n=52.

Table 2: Per-term coefficients from mixed_effects on di_pseu ~ richness, p_solu, pH, NH4, NO3 + 7 more (analysis H2_A1).

term	Variance	Std_Dev	ICC
Random Effect (Group)	0.000	0.000	0.000
Residual	0.008	0.088	
Total	0.008	0.088	

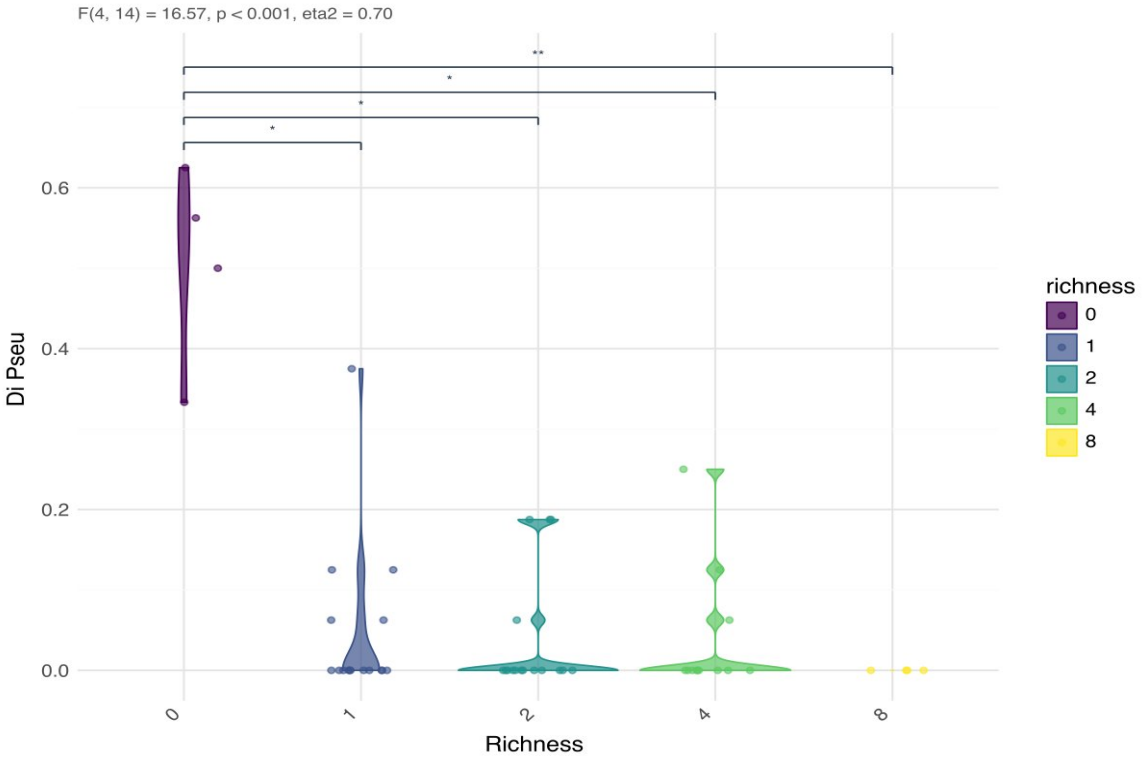


Figure 2: Distribution of Disease incidence across Number of probiotic species in the inoculum groups (N=52; 1: n=16, 2: n=16, 4: n=12, 8: n=4, 0: n=4). Violin plots show probability density distributions; boxes indicate interquartile range (25th-75th percentile) with median line; points represent individual samples. Statistical differences evaluated using Welch's ANOVA (p < 0.0001). Analysis was restricted to samples where type equals R (N=52).

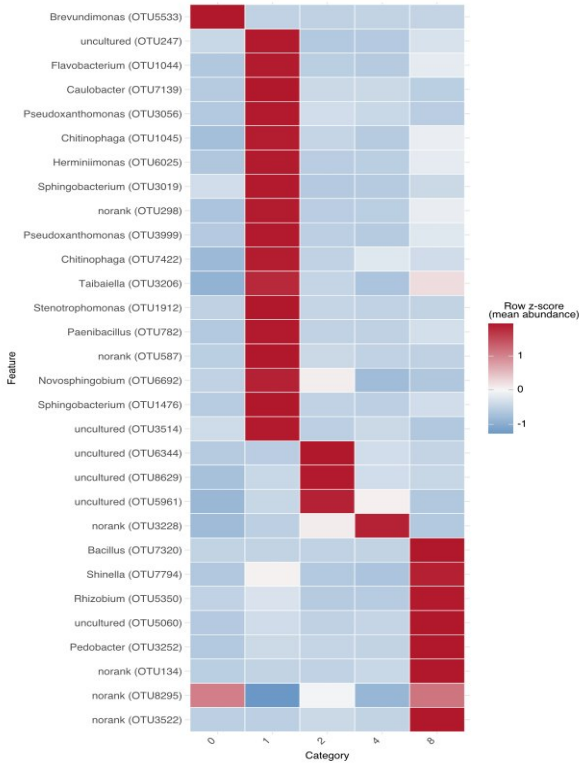


Figure 3: Discriminant features of microbial community composition across 5 groups (0, 1, 2, 4, 8) from a DESeq2 omnibus likelihood-ratio test (N=30). Heatmap of the top 30 discriminant features (30 significant at FDR < 0.050); color encodes row z-score of mean normalized abundance, each feature annotated by the group where it is most abundant. Each row is a feature and each column a group; color spans row z-scored mean abundance (darker = further from the feature's cross-group mean).

Strain MVP1_4 contributes to growth promotion

Adaptive GLM of dryweight with MVP1_4 as a predictor identified MVP1_4 as a significant positive factor (p=0.0021), alongside negative effects of NH4 (p=0.0103) and DNAc (p=0.0127) (overall p=7.62e-06; Table S5; Table 3; Table S6). A mixed-effects model incorporating both MVP1_4 and infection status confirmed a main effect of MVP1_4 (p=0.0213) while the interaction with type was non-significant (p=0.1786) (overall p=0.0213; Table S12; Table S13; Table S14; Figures 4 and 5). The ICC for this model was 0.510, indicating substantial between-community variation.

Table 3: Per-term coefficients from adaptive_glm on dryweight ~ MVP1_4, p_solu, pH, NH4, NO3 + 6 more (analysis H4_A2).

Source	sum_sq	df	F	PR(>F)
MVP1 4	0.156	1.000	10.830	0.002
p solu	0.052	1.000	3.625	0.064
pH	0.000	1.000	0.009	0.923
NH4	0.105	1.000	7.257	0.010
NO3	0.015	1.000	1.030	0.316
WSOC	0.004	1.000	0.302	0.586
WSON	0.003	1.000	0.200	0.657

Source	sum_sq	df	F	PR(>F)
N	0.002	1.000	0.132	0.719
C	0.001	1.000	0.053	0.819
CNRatio	0.003	1.000	0.193	0.663
DNAc	0.098	1.000	6.817	0.013
Residual	0.577	40.000	-	-

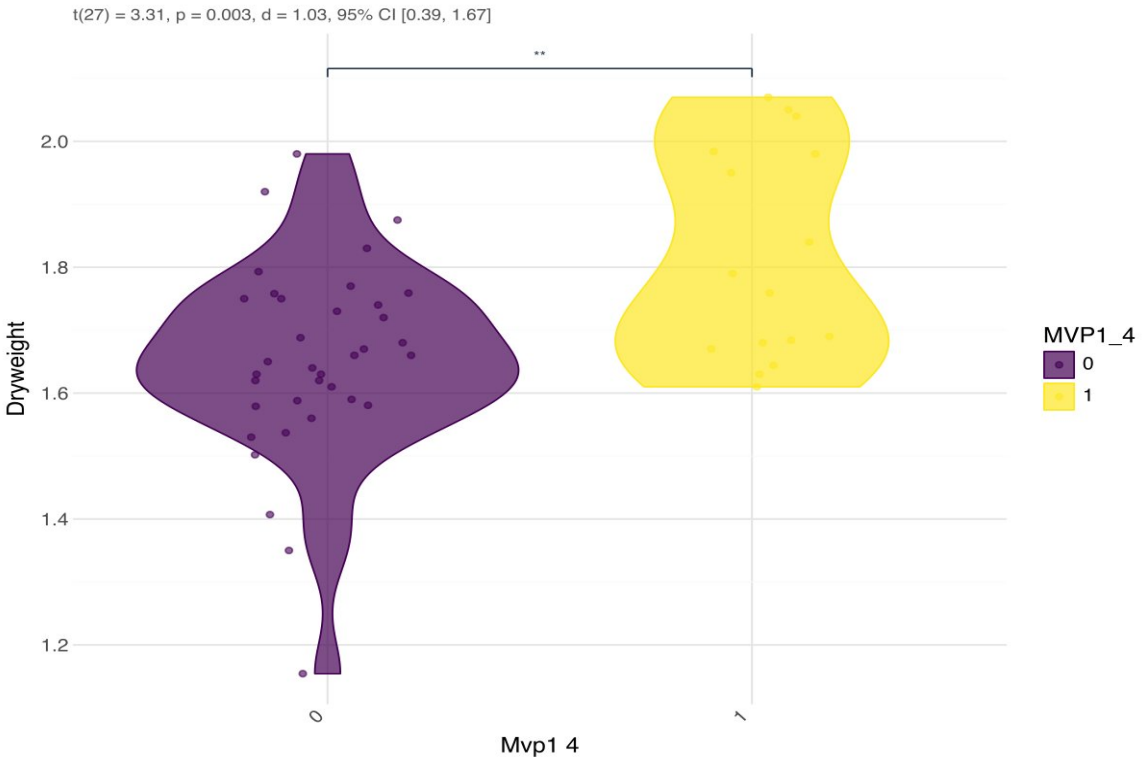


Figure 4: Distribution of Plant dry weight across Presence groups (N=52; 0: n=36, 1: n=16). Violin plots show probability density distributions; boxes indicate interquartile range (25th-75th percentile) with median line; points represent individual samples. Statistical differences evaluated using Welch's t-test (p = 0.0026). Analysis was restricted to samples where type equals H (N=52).

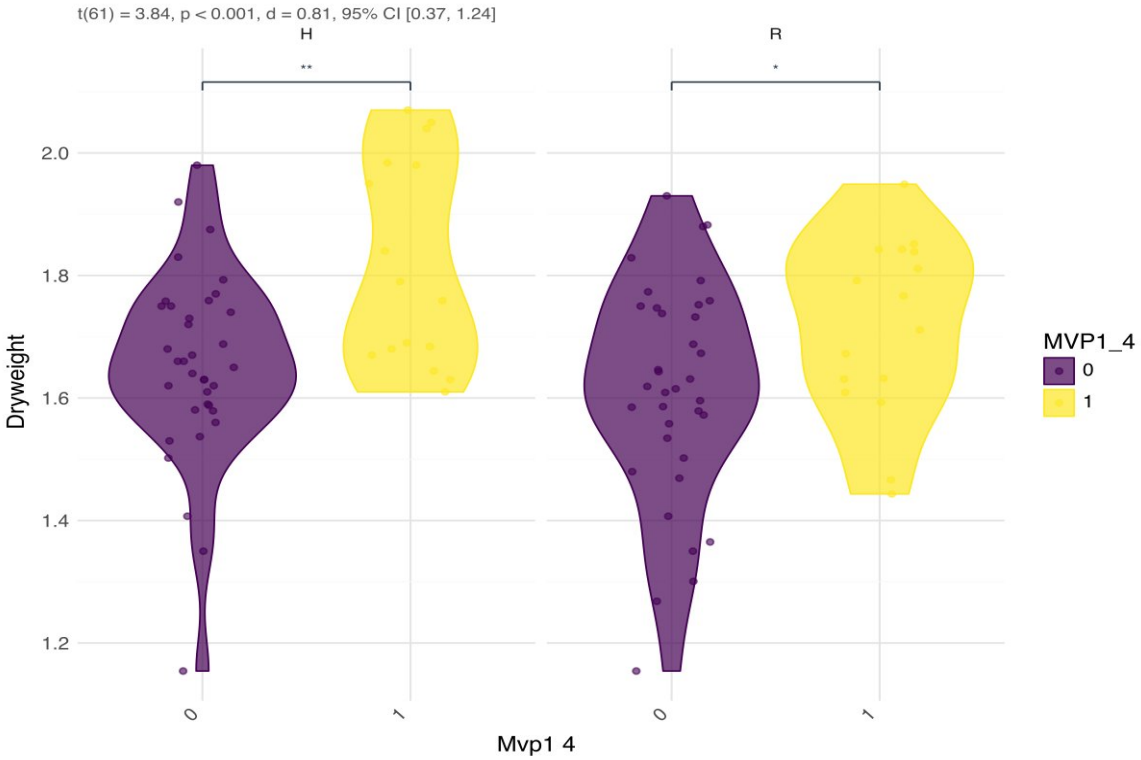


Figure 5: Distribution of Plant dry weight across Presence groups, faceted by Experimental infection status (N=104; 0: n=72, 1: n=32). Violin plots show probability density distributions; boxes indicate interquartile range (25th-75th percentile) with median line; points represent individual samples. Statistical differences evaluated using Welch's t-test (p = 0.0003).

OTU abundances do not explain additional variance

When top predictive OTUs (OTU1369, OTU6639, OTU17) were added to the richness model, richness remained the only significant predictor (overall p=1.81e-17; Table S7; Table 4; Figures 6 and 7). OTU1369 (p=0.4222), OTU6639 (p=0.1306), and OTU17 (p=0.7850) were non-significant. Adaptive GLM yielded consistent results (overall p=2.77e-10; Table S8; Table 5; Table S9) with all OTU terms non-significant.

Table 4: Per-term coefficients from mixed_effects on dryweight ~ richness, OTU1369_abundance, OTU6639_abundance, OTU17_abundance, community (analysis H6_A1).

term	Variance	Std_Dev	ICC
Random Effect (Group)	0.000	0.011	0.012
Residual	0.010	0.099	
Total	0.010	0.100	

Table 5: Per-term coefficients from adaptive_glm on dryweight ~ richness, OTU1369_abundance, OTU6639_abundance, OTU17_abundance (analysis H6_A2).

Source	sum_sq	df	F	PR(>F)
richness	0.750	4.000	18.794	0.000
OTU1369 abundance	0.005	1.000	0.521	0.474

Source	sum_sq	df	F	PR(>F)
OTU6639 abundance	0.023	1.000	2.297	0.137
OTU17 abundance	0.000	1.000	0.049	0.826
Residual	0.439	44.000	-	-

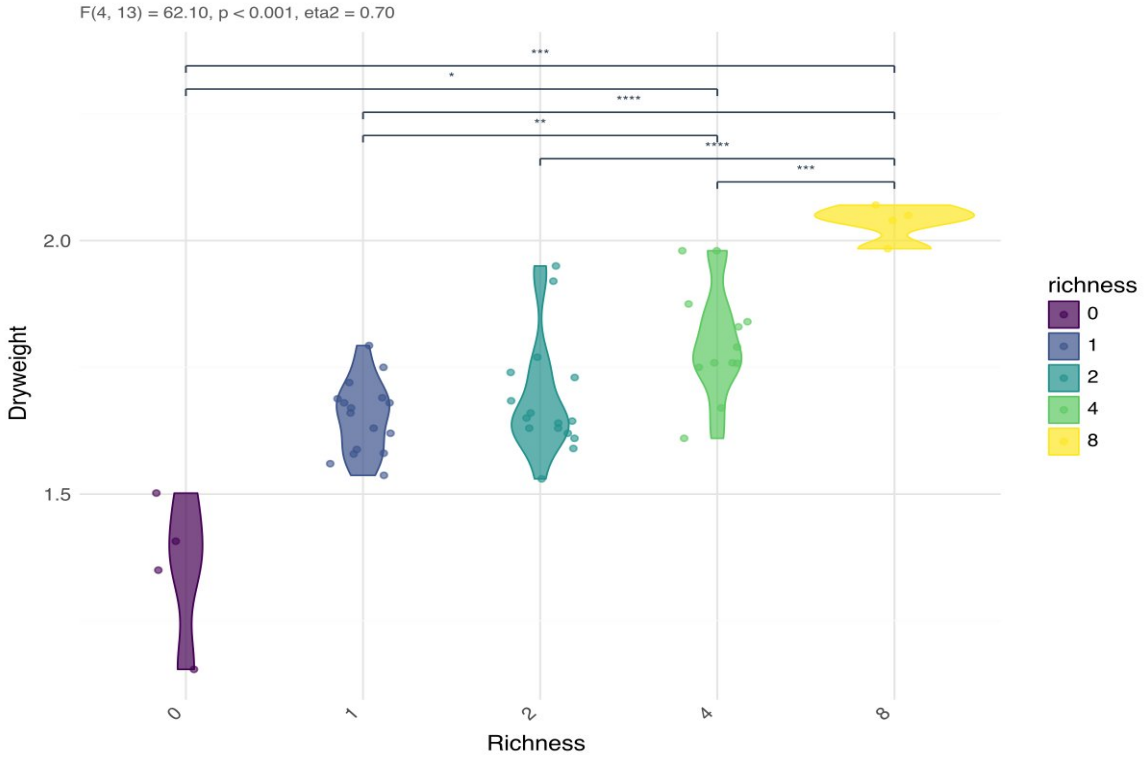


Figure 6: Distribution of Plant dry weight across Number of probiotic species in the inoculum groups (N=52; 1: n=16, 2: n=16, 4: n=12, 8: n=4, 0: n=4). Violin plots show probability density distributions; boxes indicate interquartile range (25th-75th percentile) with median line; points represent individual samples. Statistical differences evaluated using Welch's ANOVA ($p < 0.0001$). Analysis was restricted to samples where type equals H (N=52).

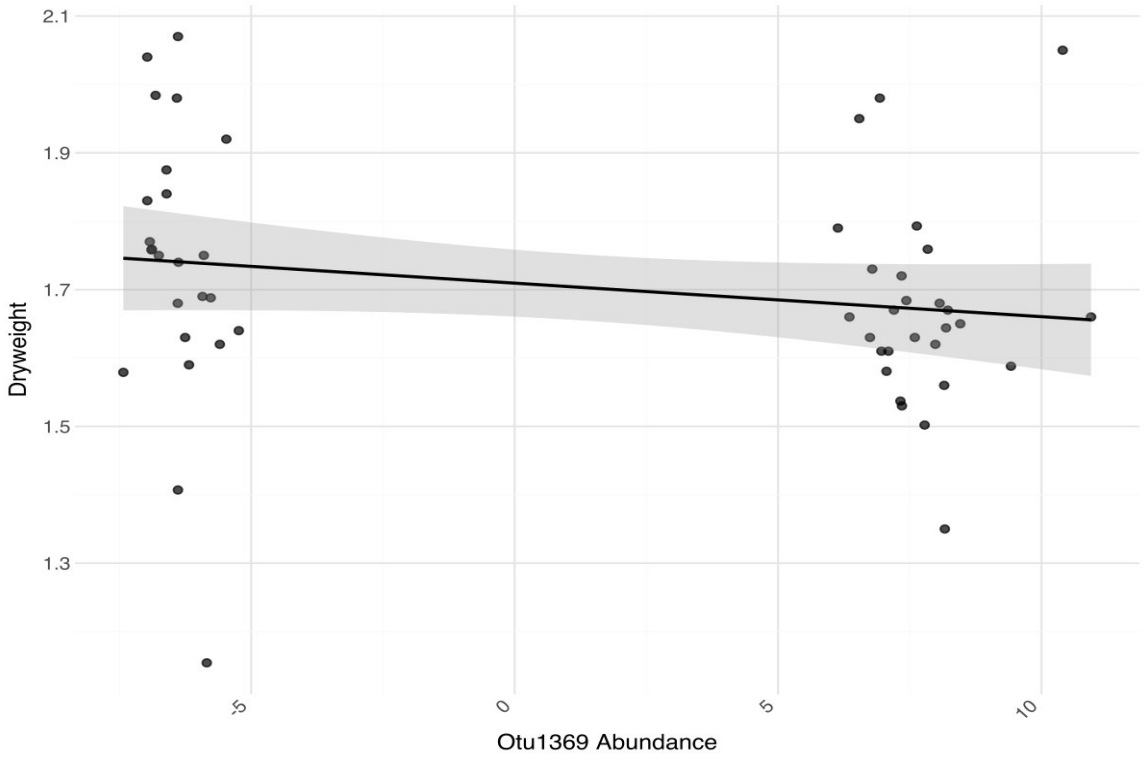


Figure 7: Relationship between CLR-transformed abundance of OTU1369 (x-axis) and Plant dry weight (y-axis) (N=52). Each point represents a sample.

Community composition varies with richness and infection status

PERMANOVA on the ASV table demonstrated a significant effect of the combined factors (overall p=0.0090; R^2=0.043; Table 7; Figures S1-S3). Redundancy analysis confirmed that richness and type together explained 5.4 % of compositional variance (p=0.0020; Table 8). Variation partitioning attributed 4.3 % of variance uniquely to richness (p=0.0020) and 1.2 % to type (p=0.0180) (Table S11).

Table 7: Overall model statistics from permanova on asv_table ~ richness, type (analysis H3_A1).

term	df	SS	test_statistic	estimate	p_value	Significance	Term_Type
richness	4	279301.692	1.102	0.043	0.0090	**	tested
type	1	72542.340	1.146	0.011	0.0250	*	tested
richness × type	4	261506.104	1.035	0.040	0.0720	ns	tested
Residual	94	5939660.611	-	0.906	-		
Total	103	6553010.747	-	1.000	-		

Table 8: Overall model statistics from rda on asv_table ~ richness, type (analysis H3_A2).

term	SS	estimate	p_value
Constrained	351844.032	0.054	0.0020
Unconstrained	6201166.715	0.946	-

term	SS	estimate	p_value
Total	6553010.747	1.000	-

Differential taxa across richness levels

DESeq2 identified 81 discriminant features among richness groups (Table S15; Table S16). Notable taxa included OTU7320 (Bacillus), OTU247 (uncultured Bacteroidetes), OTU1044 (Flavobacterium), OTU7139 (Caulobacter), and OTU3056 (Pseudoxanthomonas). These taxa varied in abundance across richness categories but were not directly linked to the primary response variables in subsequent mediation analyses.

Mediation analyses do not support functional trait pathways

Mediation of dryweight by breadth, IAA_ng_ml, or siderophore_SU was not supported (overall p=0.0805; Table S18; Table S19). Indirect effects were non-significant for all richness levels (e.g., richness 1→breadth indirect=-0.0453, p=0.0805). Similarly, mediation of di_pseu by breadth, NOI, or toxin2 was not supported (overall p=0.1611; Table S22; Table S23). The only marginal indirect effect involved richness 8→NOI (p=0.0052) but the overall mediation model remained non-significant.

Additional modeling attempts

Regression of di_pseu on richness and toxin2 yielded an inconclusive result (overall p=0.0860; negative pseudo-R^2; Table S24; Table S25; Figure S4). Regression on richness and siderophore_SU produced significant richness effects (overall p=2.79e-05) but also suffered from negative pseudo-R^2, rendering the model unreliable (Table S27; Table S28; Figure S7). GAM and XGBoost models for flic_pseu and dryweight respectively were unsupported and flagged as unreliable due to negative variance-explained scores (Tables S33-S36).

Table 6: Per-term coefficients from adaptive_glm on dryweight ~ MVP1_4, pH (analysis H8_A2).

Source	sum_sq	df	F	PR(>F)
Intercept	0.012	1.000	0.460	0.501
MVP1 4	0.012	1.000	0.465	0.499
pH	0.034	1.000	1.343	0.252
MVP1 4 × pH	0.009	1.000	0.340	0.563
Residual	1.231	48.000	-	-

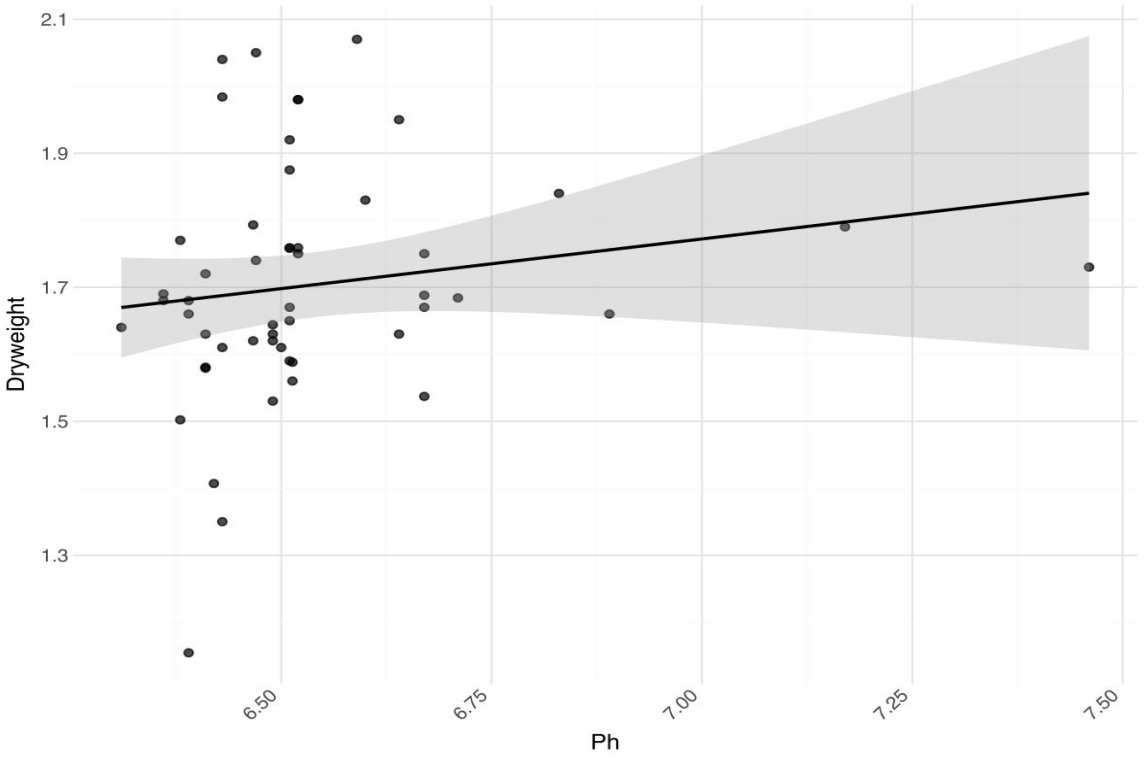


Figure 8: Relationship between Soil pH (x-axis) and Plant dry weight (y-axis) (N=52). Each point represents a sample.

Discussion

The present study provides robust evidence that increasing the richness of synthetic Pseudomonas consortia enhances plant performance and affects disease incidence under controlled greenhouse conditions. The mixed-effects analyses unequivocally demonstrate that each incremental addition of species yields statistically significant gains in dry weight (p values ranging from 0.0055 to 1.87e-06) and reductions in disease incidence (p values from 2.10e-06 to 8.03e-06). These findings align with ecological theory predicting complementarity and functional redundancy in diverse microbial assemblages [1,2] and extend prior work on probiotic endophytes that reported growth promotion and biocontrol benefits [3,4].

The strain-level analysis highlights MVP1_4 as a consistent contributor to growth promotion. Both adaptive GLM (p=0.0021) and mixed-effects modeling (p=0.0213) identified a positive main effect of MVP1_4, independent of infection status. This suggests that MVP1_4 possesses traits that directly stimulate plant biomass, possibly through phytohormone production or nutrient mobilization, consistent with reports of Pseudomonas strains enhancing IAA levels [8]. However, the interaction between MVP1_4 and soil pH was not supported (overall p=0.5919), indicating that the growth-promoting effect of this strain is robust across the pH range examined.

Community-level functional traits (breadth, IAA_ng_ml, siderophore_SU) did not mediate the richness-growth relationship, as mediation analyses failed to reach significance (overall p=0.0805). Likewise, toxin2 did not mediate disease suppression (overall p=0.1611). These null mediation results suggest that the observed benefits arise from mechanisms not captured by the measured traits, or that the traits act in a more complex, perhaps non-linear, fashion that exceeds the capacity of the current mediation framework. The PERMANOVA and RDA results confirm that richness and infection status shape overall community

composition (p=0.0090 and p=0.0020 respectively), yet the modest proportion of explained variance (≈5 %) indicates that much of the microbial variation remains unexplained, possibly due to stochastic assembly processes or unmeasured environmental factors.

The DESeq2 analysis uncovered 81 discriminant taxa across richness levels, including several Bacillus and Flavobacterium species. While these taxa differ in abundance, their direct links to plant outcomes were not established in the present models. The lack of significant associations in the GAM and XGBoost models for pathogen density (flic_pseu) and dryweight further underscores the difficulty of pinpointing predictive biomarkers in complex microbiomes, especially when model diagnostics reveal overfitting and negative variance-explained scores.

From an applied perspective, the eight-species consortium consistently produced the strongest effects on both growth and disease metrics, suggesting that maximal richness within the tested range optimizes functional outcomes. The primary mode of action appears to be growth promotion, mediated by strain MVP1_4 and possibly other unmeasured metabolites, rather than direct biocontrol via toxin2 or siderophore pathways. This interpretation is consistent with the absence of significant mediation by toxin2 and the strong positive effect of MVP1_4 on dry weight.

Limitations include the reliance on observational correlations, the modest explanatory power of community composition analyses, and the unreliability of several machine-learning models flagged by negative R^2 values. Future work should incorporate targeted metabolomic profiling to capture functional outputs, expand the richness gradient, and test causal relationships through factorial inoculation experiments.

Overall, the data support the strategic use of highly diverse synthetic microbiomes to enhance crop performance, with particular emphasis on selecting strains like MVP1_4 that confer robust growth benefits across soil conditions.

Conclusions

Increasing synthetic Pseudomonas consortium richness from 0 to 8 species significantly boosts plant dry weight and reduces disease incidence, with the eight-species assemblage delivering the greatest benefit. The growth-promotion effect is driven primarily by the strain MVP1_4, while measured community-level traits and toxin2 do not mediate these outcomes. These results endorse the deployment of diverse, strain-selected consortia to improve plant health in agricultural settings.

Appendix A: Hypotheses

- H1: In pathogen-free soils (type = H), increasing inoculum richness will increase plant dry weight because a more diverse consortium provides greater nutrient acquisition and growth-promoting metabolites. - H2: In pathogen-infested soils (type = R), higher inoculum richness will reduce disease incidence (di_pseu) by strengthening biocontrol activity of the synthetic community. - H3: Microbial community composition (asv_table) varies systematically with inoculum richness and infection status, indicating that both the richness treatment and type shape the overall microbiome structure. - H4: The effect of the strain MVP1_4 on plant dry weight is moderated by infection status (type), such that MVP1_4 improves growth only in healthy soils. - H5: The reduction in disease incidence (di_pseu) associated with higher inoculum richness is mediated by increased toxin2 production, implying toxin2 is a key biocontrol mechanism. - H6: In pathogen-free soils, the abundances of the top predictive OTUs (OTU1369, OTU6639, OTU17) explain variation in plant dry weight beyond the effect of inoculum richness alone. - H7: In pathogen-infested soils, the positive effect of inoculum richness on disease incidence (di_pseu) is mediated by increased siderophore production (siderophore_SU). - H8: The impact of strain MVP1_4 on plant dry

weight in pathogen-free soils varies with soil pH, such that MVP1_4 confers greater growth promotion at higher pH levels. - H9: Higher abundance of the discriminant taxon OTU7320 predicts increased pathogen density (flic_pseu) in pathogen-infested soils, suggesting a potential biocontrol indicator.

Appendix B: Supplementary Figures

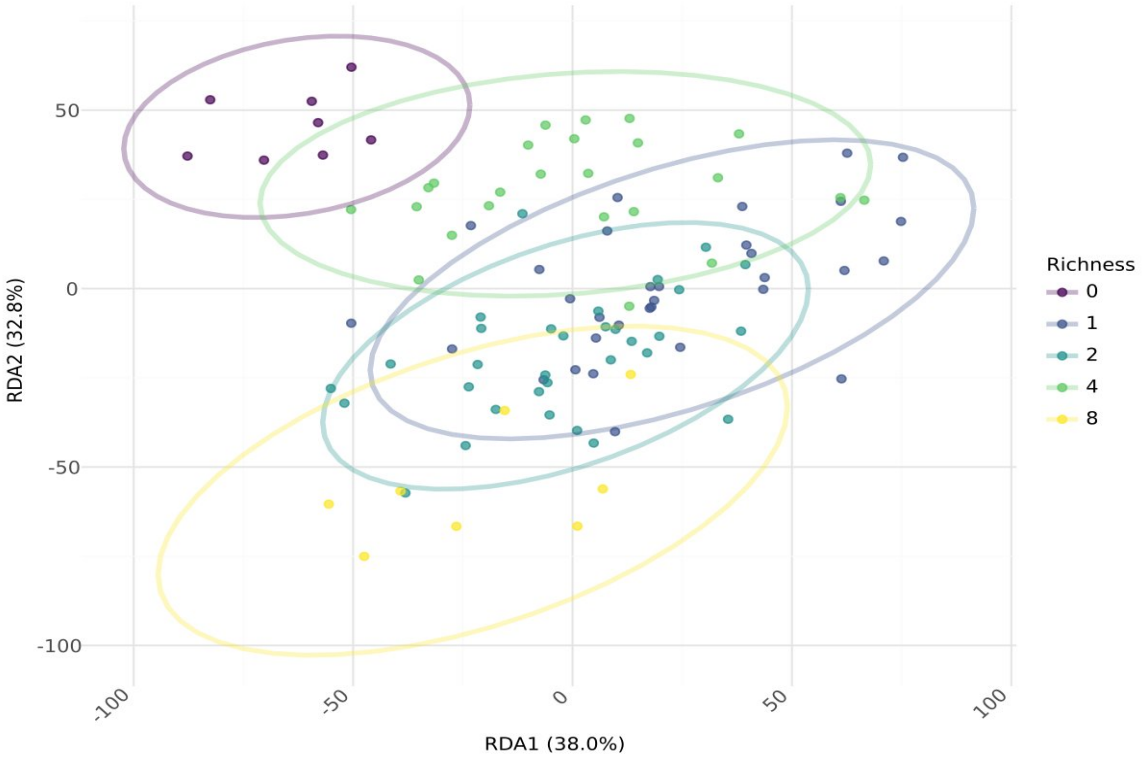


Figure S1: RDA ordination of microbial community composition based on Aitchison (CLR + Euclidean) distances (N=104; 1: n=32, 2: n=32, 4: n=24, 8: n=8, 0: n=8). Axis 1 and Axis 2 capture 38.0% and 32.8% of the constrained variance. Points represent individual samples colored by Number of probiotic species in the inoculum. Ellipses show 95% confidence intervals per group. Groups compared: 1 (n=32), 2 (n=32), 4 (n=24), 8 (n=8), 0 (n=8). RDA with ANOVA-like permutation test: p = 0.0020. Point color encodes Number of probiotic species in the inoculum group membership; ellipses delineate 95% confidence regions per group.

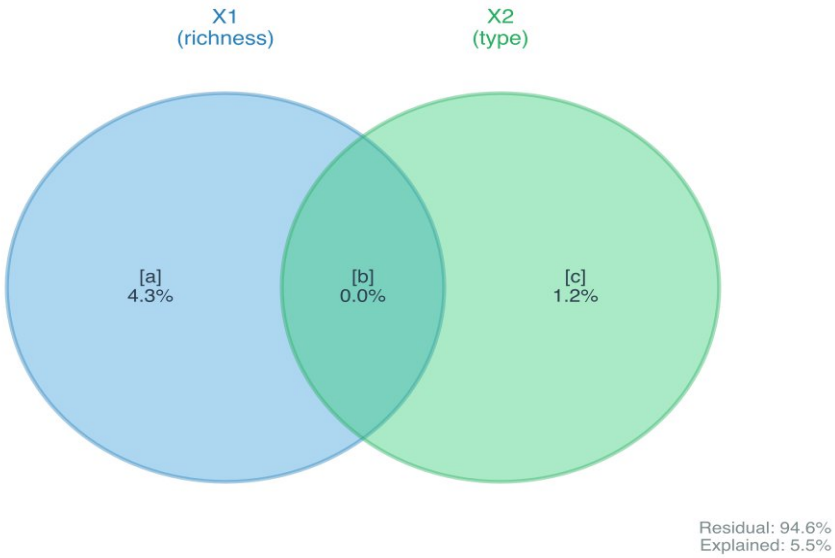


Figure S2: Variance partitioning of microbial community composition across predictor groups (Number of probiotic species in the inoculum). Each component shows the share of variance attributable to a predictor group: pure focal [a]=4.3%, pure conditioning [c]=1.2%, shared [b]=0.0%, residual [d]=94.6%. Variation Partitioning (Partial RDA): total constrained R^2=5.5% (omnibus p=0.0020); pure focal component p=0.0020, pure conditioning component p=0.0180. Each circle/bar segment represents one variance component; segment area is proportional to the fraction of total variance explained by that component.

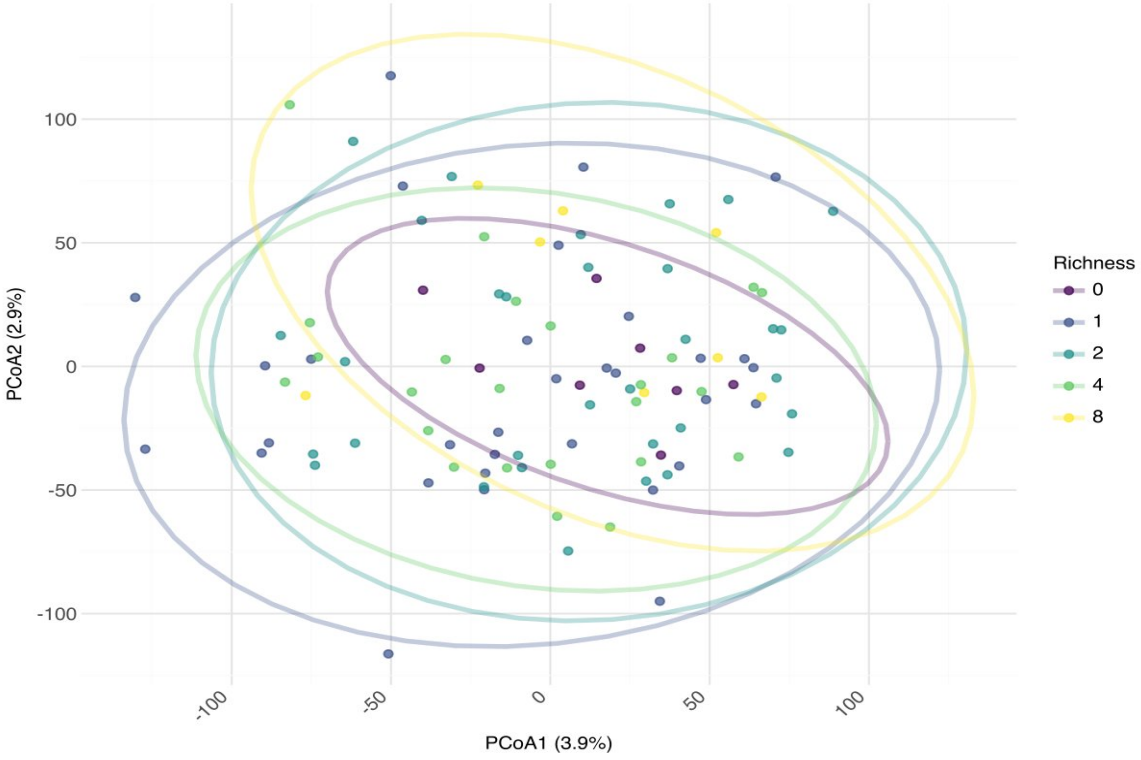


Figure S3: PCoA ordination of microbial community composition based on Aitchison (CLR + Euclidean) distances (N=104; 1: n=32, 2: n=32, 4: n=24, 8: n=8, 0: n=8). Axis 1 and Axis 2 capture 3.9% and 2.9% of total variance. Points represent individual samples colored by Number of probiotic species in the inoculum. Ellipses show 95% confidence intervals per group. Groups compared: 1 (n=32), 2 (n=32), 4 (n=24), 8 (n=8), 0 (n=8). PERMANOVA: p = 0.0090. Point color encodes Number of probiotic species in the inoculum group membership; ellipses delineate 95% confidence regions per group.

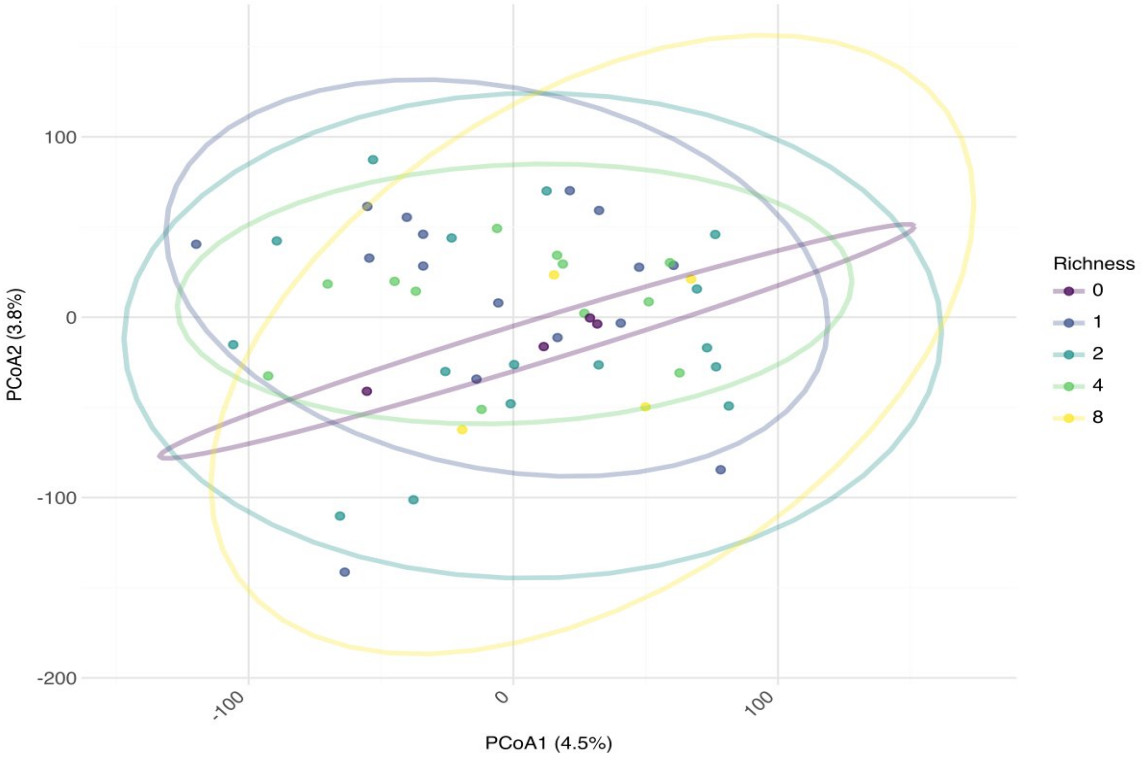


Figure S4: PCoA ordination of microbial community composition based on Aitchison (CLR + Euclidean) distances (N=52; 1: n=16, 2: n=16, 4: n=12, 8: n=4, 0: n=4). Axis 1 and Axis 2 capture 4.5% and 3.8% of total variance. Points represent individual samples colored by Number of probiotic species in the inoculum. Ellipses show 95% confidence intervals per group. Groups compared: 1 (n=16), 2 (n=16), 4 (n=12), 8 (n=4), 0 (n=4). PERMANOVA: p = 0.0230. Analysis was restricted to samples where type equals R (N=52). Point color encodes Number of probiotic species in the inoculum group membership; ellipses delineate 95% confidence regions per group.

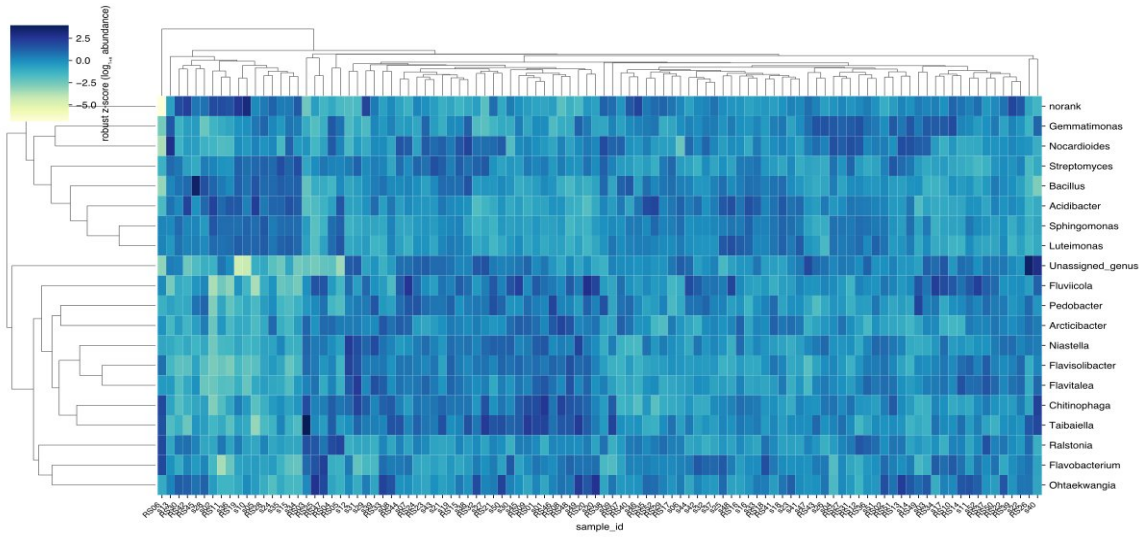


Figure S5: Clustered heatmap of ASV relative abundance (N=104). Color intensity represents relative abundance (z-score normalized). Rows and columns ordered by hierarchical clustering. Color scale spans z-score normalized abundance: darker shades indicate values further from row mean (positive = above, negative = below); dendrograms reflect hierarchical clustering of rows and columns.

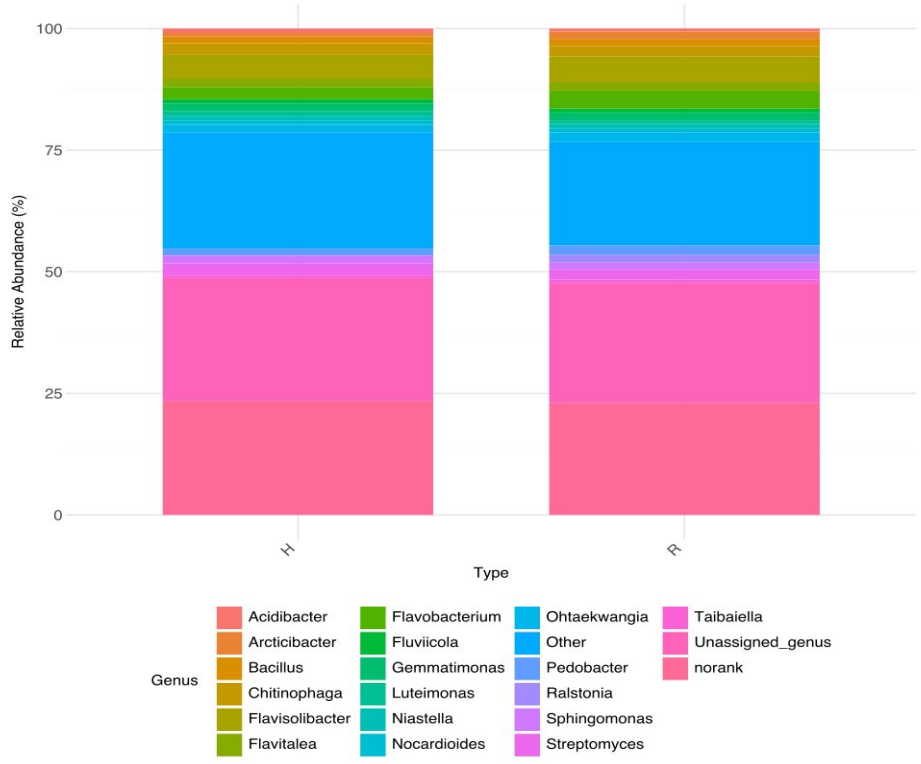


Figure S6: Compositional profile of microbial community composition across Experimental infection status groups (N=104; H: n=52, R: n=52). Stacked bars show relative abundance of top 21 features; remaining features grouped as 'Other'. Each color identifies a distinct feature (legend lists top 21); remaining features pooled as

'Other'.

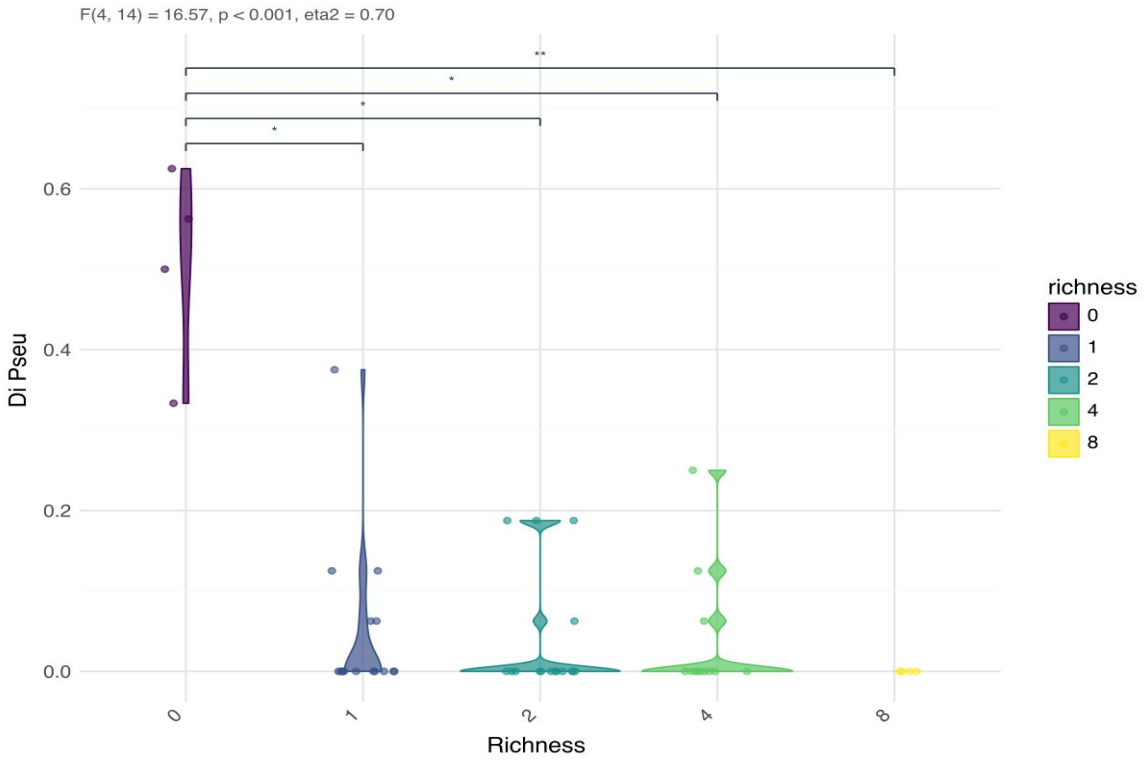


Figure S7: Distribution of Disease incidence across Number of probiotic species in the inoculum groups (N=52; 1: n=16, 2: n=16, 4: n=12, 8: n=4, 0: n=4). Violin plots show probability density distributions; boxes indicate interquartile range (25th-75th percentile) with median line; points represent individual samples. Statistical differences evaluated using Welch's ANOVA ($p < 0.0001$). Analysis was restricted to samples where type equals R (N=52).

Appendix C: Supplementary Tables

Table S1: Overall model statistics from mixed_effects on dryweight ~ richness, p_solu, pH, NH4, NO3 + 7 more (analysis H1_A1).

Model	N_observations	N_groups	ICC	Log_Likelihood	AIC	BIC
Linear Mixed Effects	52	38	0.000	8.103	-	-

Table S2: Group descriptive statistics from mixed_effects on dryweight ~ richness, p_solu, pH, NH4, NO3 + 7 more (analysis H1_A1).

Group	N	Mean	Std
1	16	1.652	0.071
2	16	1.687	0.113
4	12	1.800	0.110
8	4	2.036	0.037
0	4	1.353	0.147

Table S3: Overall model statistics from mixed_effects on di_pseu ~ richness, p_solu, pH, NH4, NO3 + 7 more (analysis H2_A1).

Model	N_observations	N_groups	ICC	Log_Likelihood	AIC	BIC
Linear Mixed Effects	52	38	0.000	10.439	-	-

Table S4: Group descriptive statistics from mixed_effects on di_pseu ~ richness, p_solu, pH, NH4, NO3 + 7 more (analysis H2_A1).

Group	N	Mean	Std
1	16	0.047	0.098
2	16	0.039	0.075
4	12	0.036	0.078
8	4	0.000	0.000
0	4	0.505	0.125

Table S5: Overall model statistics from adaptive_glm on dryweight ~ MVP1_4, p_solu, pH, NH4, NO3 + 6 more (analysis H4_A2).

Statistic	Value
R^2	0.6310
Overall p	7.6211e-06
AIC	-62.51
BIC	-39.10
n_obs	52
df_resid	40
Family	gaussian
Link	identity

Table S6: Group descriptive statistics from adaptive_glm on dryweight ~ MVP1_4, p_solu, pH, NH4, NO3 + 6 more (analysis H4_A2).

Group	N	Mean	Std
1	16	1.817	0.169
0	36	1.653	0.155

Table S7: Overall model statistics from mixed_effects on dryweight ~ richness, OTU1369_abundance, OTU6639_abundance, OTU17_abundance, community (analysis H6_A1).

Model	N_observations	N_groups	ICC	Log_Likelihood	AIC	BIC
Linear Mixed Effects	52	38	0.012	22.138	-	-

Table S8: Overall model statistics from adaptive_glm on dryweight ~ richness, OTU1369_abundance, OTU6639_abundance, OTU17_abundance (analysis H6_A2).

Statistic	Value
R^2	0.7192

Statistic	Value
Overall p	2.7698e-10
AIC	-84.72
BIC	-69.11
n_obs	52
df_resid	44
Family	gaussian
Link	identity

Table S9: Group descriptive statistics from adaptive_glm on dryweight ~ richness, OTU1369_abundance, OTU6639_abundance, OTU17_abundance (analysis H6_A2).

Group	N	Mean	Std
1	16	1.652	0.071
2	16	1.687	0.113
4	12	1.800	0.110
8	4	2.036	0.037
0	4	1.353	0.147

Table S10: Overall model statistics from adaptive_glm on dryweight ~ MVP1_4, pH (analysis H8_A2).

Statistic	Value
R^2	0.2124
Overall p	8.9774e-03
AIC	-39.09
BIC	-31.29
n_obs	52
df_resid	48
Family	gaussian
Link	identity

Table S11: Overall model statistics from variation_partitioning on asv_table ~ richness, type (analysis H3_A3).

term	estimate	Percentage	p_value
richness [a]	0.043	4.310	0.0020
type [c]	0.012	1.156	0.0180
Shared [b]	0.000	0.000	-
Residual [d]	0.946	94.631	-
Total Explained [a+b+c]	0.054	5.369	0.0020

Table S12: Overall model statistics from mixed_effects on dryweight ~ MVP1_4, type, p_solu, pH, NH4 + 8 more (analysis H4_A1).

Model	N_observations	N_groups	ICC	Log_Likelihood	AIC	BIC
Linear Mixed Effects	104	38	0.510	23.883	-	-

Table S13: Per-term coefficients from mixed_effects on dryweight ~ MVP1_4, type, p_solu, pH, NH4 + 8 more (analysis H4_A1).

term	Variance	Std_Dev	ICC
Random Effect (Group)	0.011	0.104	0.510
Residual	0.010	0.102	
Total	0.021	0.146	

Table S14: Group descriptive statistics from mixed_effects on dryweight ~ MVP1_4, type, p_solu, pH, NH4 + 8 more (analysis H4_A1).

Group	N	Mean	Std
1	32	1.766	0.163
0	72	1.632	0.168

Table S15: Overall model statistics from deseq2 on asv_table ~ richness (analysis H2_A3).

grouping_variable	n_categories	categories	df	Total_Features_Tested	Discriminant_Features	FDR_Threshold
richness	5	1, 2, 4, 8, 0	4	5893	81	0.050

Table S16: Per-term coefficients from deseq2 on asv_table ~ richness (analysis H2_A3).

term	Species	lrt_stat	p_value	p_adjusted	discriminant_class
OTU7320	Unclassified	193.606	8.90e-41	5.25e-37	8
OTU247	uncultured bacterium	99.200	1.46e-20	4.29e-17	1
OTU1044	uncultured bacterium	83.628	2.97e-17	5.82e-14	1
OTU7139	Unclassified	82.677	4.72e-17	6.95e-14	1
OTU3056	uncultured bacterium	74.238	2.89e-15	3.40e-12	1
OTU7794	Unclassified	68.019	5.94e-14	5.84e-11	8
OTU1045	Chitinophaga filiformis	60.543	2.23e-12	1.88e-09	1
OTU6025	Unclassified	59.506	3.68e-12	2.71e-09	1
OTU6344	uncultured bacterium	57.042	1.21e-11	7.94e-09	2
OTU5350	uncultured bacterium	54.055	5.12e-11	3.02e-08	8
OTU3019	Unclassified	51.958	1.41e-10	7.54e-08	1

term	Species	lrt_stat	p_value	p_adjusted	discriminant_class
OTU298	norank	46.422	2.01e-09	9.88e-07	1
OTU3999	uncultured bacterium	43.415	8.49e-09	3.85e-06	1
OTU5060	uncultured bacterium	42.430	1.36e-08	5.72e-06	8
OTU8629	uncultured bacterium	41.640	1.98e-08	7.29e-06	2
OTU3252	uncultured bacterium	41.745	1.88e-08	7.29e-06	8
OTU7422	Unclassified	40.039	4.25e-08	1.47e-05	1
OTU3206	uncultured bacterium	39.473	5.56e-08	1.82e-05	1
OTU5961	uncultured bacterium	39.081	6.70e-08	2.08e-05	2
OTU134	norank	37.463	1.45e-07	4.26e-05	8
OTU1912	Stenotrophomonas maltophilia	37.158	1.67e-07	4.69e-05	1
OTU782	Unclassified	35.946	2.97e-07	7.95e-05	1
OTU587	norank	35.242	4.14e-07	0.0001	1
OTU6692	Unclassified	34.065	7.23e-07	0.0002	1
OTU3228	norank	33.938	7.67e-07	0.0002	4
OTU1476	uncultured bacterium	33.926	7.72e-07	0.0002	1
OTU8295	norank	32.588	1.45e-06	0.0003	8
OTU3522	norank	32.145	1.79e-06	0.0004	8
OTU3514	uncultured bacterium	30.647	3.61e-06	0.0007	1
OTU5533	Unclassified	30.281	4.29e-06	0.0008	0
OTU907	Unclassified	29.458	6.31e-06	0.0012	1
OTU7062	uncultured bacterium	29.389	6.52e-06	0.0012	8
OTU4164	uncultured bacterium	28.986	7.87e-06	0.0014	1
OTU2373	Unclassified	28.353	1.06e-05	0.0018	1
OTU1016	Unclassified	27.975	1.26e-05	0.0020	8
OTU1420	norank	27.979	1.26e-05	0.0020	2
OTU2549	uncultured bacterium	28.006	1.24e-05	0.0020	0

term	Species	lrt_stat	p_value	p_adjusted	discriminant_class
OTU1888	Ralstonia solanacearum FQY_4	27.911	1.30e-05	0.0020	1
OTU5614	norank	27.710	1.43e-05	0.0021	8
OTU1003	Unclassified	27.741	1.41e-05	0.0021	0
OTU1078	norank	27.070	1.92e-05	0.0028	2
OTU3634	Unclassified	26.349	2.69e-05	0.0037	8
OTU8478	uncultured bacterium	26.335	2.71e-05	0.0037	4
OTU3573	uncultured bacterium	26.163	2.93e-05	0.0039	8
OTU1435	norank	25.856	3.38e-05	0.0044	8
OTU1126	Unclassified	25.206	4.57e-05	0.0059	1
OTU5978	norank	24.937	5.18e-05	0.0065	8
OTU6213	uncultured bacterium	24.770	5.60e-05	0.0069	2
OTU1668	Unclassified	24.288	6.99e-05	0.0084	1
OTU3293	Unclassified	24.157	7.43e-05	0.0088	8

Table S17: Overall model statistics from permanova on asv_table ~ richness (analysis H5_A3).

term	df	SS	test_statistic	estimate	p_value	Significance	Term_Type
richness	4	270271.851	1.084	0.084	0.0230	*	tested
Residual	47	2930203.862	-	0.916	-		
Total	51	3200475.713	-	1.000	-		

Table S18: Overall model statistics from mediation on dryweight ~ richness_1, richness_2, richness_4, richness_8, breadth + 2 more (analysis H1_A2).

Path	Indirect	Boot_95CI	p	Type
richness_1→breadth→dryweight	-0.0453	[-0.0960, 0.0055]	0.0805	none
richness_1→IAA_ng_ml→dryweight	-0.0292	[-0.0697, 0.0114]	0.1588	none
richness_1→siderophore_SU→dryweight	-0.0414	[-0.1000, -0.0000]	0.0520	none
richness_2→breadth→dryweight	0.0478	[-0.0126, 0.1082]	0.1208	none
richness_2→IAA_ng_ml→dryweight	0.0089	[-0.0351, 0.0529]	0.6907	none
richness_2→siderophore_SU→dryweight	0.0037	[-0.0372, 0.0446]	0.8596	none
richness_4→breadth→dryweight	0.0406	[-0.0259, 0.1071]	0.2315	none

Path	Indirect	Boot_95CI	p	Type
richness_4→IAA_ng_ml→dryweight	0.0232	[-0.0291, 0.0755]	0.3853	none
richness_4→siderophore_SU→dryweight	0.0648	[-0.0010, 0.1306]	0.0537	none
richness_8→breadth→dryweight	0.0674	[-0.0096, 0.1444]	0.0863	none
richness_8→IAA_ng_ml→dryweight	0.0548	[-0.0019, 0.1430]	0.0600	none
richness_8→siderophore_SU→dryweight	0.0469	[-0.0049, 0.1222]	0.0880	none

Table S19: Per-term coefficients from mediation on dryweight ~ richness_1, richness_2, richness_4, richness_8, breadth + 2 more (analysis H1_A2).

term	a (X→M)	b (M→Y	X)	c' (direct)	estimate	ci_lower	ci_upper	Sobel_Z	Type
richness_1→breadth→dryweight	7.1739 (p=0.0539)	0.0063 (p=0.0007)	0.0538 (p=0.1970)	-0.0453	-0.0960	0.0055	-1.748	none	
richness_1→IAA_ng_ml→dryweight	4.2769 (p=0.1039)	0.0228 (p=0.0116)	0.0699 (p=0.1145)	-0.0292	-0.0697	0.0114	-1.409	none	
richness_1→siderophore_SU→dryweight	0.1386 (p=0.0063)	0.5109 (p=0.0409)	0.0577 (p=0.2300)	-0.0414	-0.1000	-0.0000	-1.703	none	
richness_2→breadth→dryweight	4.1937 (p=0.1041)	0.0074 (p=0.0001)	0.0380 (p=0.3895)	-0.0478	-0.0126	0.1082	1.551	none	
richness_2→IAA_ng_ml→dryweight	0.3388 (p=0.6903)	0.0264 (p=0.0040)	0.0008 (p=0.9854)	-0.0089	-0.0351	0.0529	0.398	none	
richness_2→siderophore_SU→dryweight	0.0096 (p=0.8603)	0.3846 (p=0.0075)	0.0061 (p=0.8970)	-0.0037	-0.0372	0.0446	0.177	none	
richness_4→breadth→dryweight	5.8399 (p=0.2180)	0.0070 (p=0.0002)	0.0249 (p=0.9237)	-0.0406	-0.0259	0.1071	1.197	none	
richness_4→IAA_ng_ml→dryweight	0.8954 (p=0.3695)	0.0259 (p=0.0049)	0.0224 (p=0.6852)	-0.0232	-0.0291	0.0755	0.868	none	
richness_4→siderophore_SU→dryweight	0.1604 (p=0.0934)	0.4040 (p=0.0096)	0.0193 (p=0.7498)	-0.0648	-0.0010	0.1306	1.929	none	
richness_8→breadth→dryweight	12.4447 (p=0.0569)	0.0054 (p=0.0012)	0.0225 (p=0.0013)	-0.0674	-0.0096	0.1444	1.715	none	
richness_8→IAA_ng_ml→dryweight	6.4368 (p=0.1060)	0.0160 (p=0.0023)	0.0350 (p=0.0027)	-0.0548	-0.0019	0.1430	1.560	none	
richness_8→siderophore_SU→dryweight	0.2048 (p=0.1822)	0.2288 (p=0.0842)	0.0130 (p=0.0019)	-0.0469	-0.0049	0.1222	1.439	none	

Table S20: Overall model statistics from xgboost on dryweight ~ asv_table, richness (analysis H1_A3).

Metric	Value
Training R²	0.9956
Test R²	-0.2119
Training RMSE	0.0115
Test RMSE	0.1909
Cross-Val Mean R2	-2.1374

Table S21: Per-term coefficients from xgboost on dryweight ~ asv_table, richness (analysis H1_A3).

Rank	Feature	Species	Importance
1	OTU1369	uncultured bacterium	0.031
2	OTU6639	Unclassified	0.028
3	OTU17	uncultured Gracilibacter sp.	0.024
4	OTU6505	Unclassified	0.024

Rank	Feature	Species	Importance
5	OTU4652	norank	0.024
6	OTU96	Unclassified	0.024
7	OTU4562	Unclassified	0.020
8	OTU3496	Unclassified	0.020
9	OTU7819	Unclassified	0.020
10	OTU533	norank	0.020

Table S22: Overall model statistics from mediation on di_pseu ~ richness_1, richness_2, richness_4, richness_8, breadth + 2 more (analysis H2_A2).

Path	Indirect	Boot_95CI	p	Type
richness_1→breadth→di_pseu	0.0324	[-0.0129, 0.0778]	0.1611	inconsistent
richness_1→NOI→di_pseu	0.0109	[-0.0348, 0.0566]	0.6405	inconsistent
richness_1→toxin2→di_pseu	0.0420	[-0.0069, 0.0909]	0.0921	inconsistent
richness_2→breadth→di_pseu	-0.0052	[-0.0428, 0.0324]	0.7880	none
richness_2→NOI→di_pseu	-0.0014	[-0.0442, 0.0415]	0.9501	none
richness_2→toxin2→di_pseu	-0.0271	[-0.0691, 0.0149]	0.2061	none
richness_4→breadth→di_pseu	-0.0381	[-0.0912, 0.0150]	0.1593	none
richness_4→NOI→di_pseu	-0.0222	[-0.0777, 0.0334]	0.4340	none
richness_4→toxin2→di_pseu	-0.0259	[-0.0773, 0.0255]	0.3240	none
richness_8→breadth→di_pseu	-0.1585	[-0.2812, -0.0359]	0.0113	none
richness_8→NOI→di_pseu	-0.1793	[-0.3049, -0.0536]	0.0052	none
richness_8→toxin2→di_pseu	-0.0925	[-0.1876, 0.0027]	0.0570	none

Table S23: Per-term coefficients from mediation on di_pseu ~ richness_1, richness_2, richness_4, richness_8, breadth + 2 more (analysis H2_A2).

term	a (X→M)	b (M→Y	X)	c' (direct)	estimate	ci_lower	ci_upper	Sobel_Z	Type
richness_1→breadth→di_pseu	5.6963 (p=0.1374)	0.0057 (p=0.0097)	0.0548 (p=0.1563)	0.0324	-0.0129	0.0778	1.401	inconsistent	
richness_1→NOI→di_pseu	0.0206 (p=0.6409)	0.5279 (p=0.0092)	0.0332 (p=0.3597)	0.0109	-0.0348	0.0566	0.467	inconsistent	
richness_1→toxin2→di_pseu	9.2130 (p=0.0699)	0.0046 (p=0.0093)	0.0643 (p=0.0957)	0.0420	-0.0069	0.0909	1.684	inconsistent	
richness_2→breadth→di_pseu	0.9968 (p=0.7887)	0.0052 (p=0.0018)	0.0001 (p=0.9986)	-0.0052	-0.0428	0.0324	-0.269	none	
richness_2→NOI→di_pseu	0.0026 (p=0.9504)	0.5189 (p=0.0093)	0.0039 (p=0.9116)	-0.0014	-0.0442	0.0415	-0.063	none	
richness_2→toxin2→di_pseu	6.5300 (p=0.1836)	0.0041 (p=0.0010)	0.0219 (p=0.5525)	-0.0271	-0.0691	0.0149	-1.264	none	
richness_4→breadth→di_pseu	7.0382 (p=0.1304)	0.0054 (p=0.0015)	0.0304 (p=0.5229)	-0.0381	-0.0912	0.0150	-1.407	none	
richness_4→NOI→di_pseu	0.0423 (p=0.4298)	0.5244 (p=0.0093)	0.0144 (p=0.7456)	-0.0222	-0.0777	0.0334	-0.782	none	
richness_4→toxin2→di_pseu	6.3516 (p=0.3103)	0.0041 (p=0.0012)	0.0181 (p=0.6957)	-0.0259	-0.0773	0.0255	-0.986	none	
richness_8→breadth→di_pseu	26.7454 (p=0.0002)	0.0059 (p=0.0023)	0.0696 (p=0.4466)	-0.1585	-0.2812	-0.0359	-2.533	none	
richness_8→NOI→di_pseu	9.2974 (p=0.0004)	0.6028 (p=0.0003)	0.0904 (p=0.2947)	-0.1793	-0.3049	-0.0536	-2.797	none	

term	a (X→M)	b (M→Y	X)	c' (direct)	estimate	ci_lower	ci_upper	Sobel_Z	Type
richness_8→toxin2→di_pseu	2.8089 (p=0.0246)	0.0040 (p=0.0022)	0.0036 (p=0.9651)	-0.0925	-0.1876	0.0027	-1.904	none	

Table S24: Overall model statistics from regression on di_pseu ~ richness, toxin2, p_solu, pH, NH4 + 7 more (analysis H5_A2).

Statistic	Value
Pseudo-R²	-0.2551
Overall p	8.5966e-02
AIC	-182.88
BIC	-149.71
n_obs	52
df_resid	35
Family	beta
Link	identity

Table S25: Per-term coefficients from regression on di_pseu ~ richness, toxin2, p_solu, pH, NH4 + 7 more (analysis H5_A2).

term	estimate	p_value
Intercept	-30.755	0.5834
richness: 1	-2.767	0.0860
richness: 2	-2.591	0.1470
richness: 4	-2.453	0.1685
richness: 8	-2.062	0.2502
toxin2	-0.018	0.3390
p solu	0.000	0.7326
pH	0.214	0.8943
NH4	-0.039	0.7219
NO3	-0.000	0.9368
WSOC	0.015	0.1287
WSON	-0.007	0.6837
N	148.590	0.4964
C	-17.145	0.4887
CNRatio	3.325	0.5918
DNAc	0.004	0.8311

Table S26: Group descriptive statistics from regression on di_pseu ~ richness, toxin2, p_solu, pH, NH4 + 7 more (analysis H5_A2).

Group	N	Mean	Std
1	16	0.056	0.096

Group	N	Mean	Std
2	16	0.048	0.074
4	12	0.045	0.076
8	4	0.010	0.000
0	4	0.505	0.123

Table S27: Overall model statistics from regression on di_pseu ~ richness, siderophore_SU (analysis H7_A2).

Statistic	Value
Pseudo-R^2	-0.2109
Overall p	2.7904e-05
AIC	-195.25
BIC	-181.59
n_obs	52
df_resid	45
Family	beta
Link	identity

Table S28: Per-term coefficients from regression on di_pseu ~ richness, siderophore_SU (analysis H7_A2).

term	estimate	p_value
Intercept	0.016	0.9517
richness: 1	-2.767	2.79e-05
richness: 2	-2.790	0.0003
richness: 4	-2.778	0.0013
richness: 8	-3.042	0.0042
siderophore SU	-0.376	0.7475

Table S29: Group descriptive statistics from regression on di_pseu ~ richness, siderophore_SU (analysis H7_A2).

Group	N	Mean	Std
1	16	0.056	0.096
2	16	0.048	0.074
4	12	0.045	0.076
8	4	0.010	0.000
0	4	0.505	0.123

Table S30: Overall model statistics from mixed_effects on dryweight ~ MVP1_4, pH, community (analysis H8_A1).

Model	N_observations	N_groups	ICC	Log_Likelihood	AIC	BIC
Linear Mixed Effects	52	38	0.658	25.364	-	-

Table S31: Per-term coefficients from mixed_effects on dryweight ~ MVP1_4, pH, community (analysis H8_A1).

term	Variance	Std_Dev	ICC
Random Effect (Group)	0.014	0.120	0.658
Residual	0.007	0.086	
Total	0.022	0.148	

Table S32: Group descriptive statistics from mixed_effects on dryweight ~ MVP1_4, pH, community (analysis H8_A1).

Group	N	Mean	Std
1	16	1.817	0.169
0	36	1.653	0.155

Table S33: Overall model statistics from gam on flic_pseu ~ OTU7320_abundance (analysis H9_A1).

Statistic	Value
Pseudo-R^2	0.1103
GCV	0.1586
n_obs	52

Table S34: Per-term coefficients from gam on flic_pseu ~ OTU7320_abundance (analysis H9_A1).

Source	EDoF	p_value	Significance
OTU7320_abundance	5.580	0.1268	ns

Table S35: Overall model statistics from xgboost on flic_pseu ~ asv_table (analysis H9_A2).

Metric	Value
Training R^2	0.9813
Test R^2	-0.3142
Training RMSE	0.0490
Test RMSE	0.4114
Cross-Val Mean R2	-0.9027

Table S36: Per-term coefficients from xgboost on flic_pseu ~ asv_table (analysis H9_A2).

Rank	Feature	Species	Importance
1	OTU3074	norank	0.013
2	OTU7510	uncultured bacterium	0.009
3	OTU7728	uncultured bacterium	0.009
4	OTU6812	uncultured organism	0.009
5	OTU2592	uncultured bacterium	0.009
6	OTU4065	uncultured bacterium	0.008

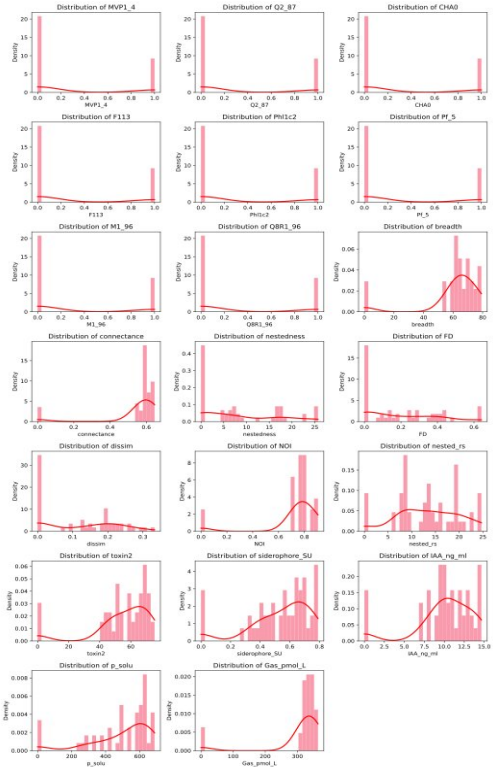
Rank	Feature	Species	Importance
7	OTU8654	Unclassified	0.008
8	OTU3552	norank	0.007
9	OTU7747	Unclassified	0.007
10	OTU4546	uncultured organism	0.007

Exported tables

The following large tables are saved alongside the report rather than inlined:

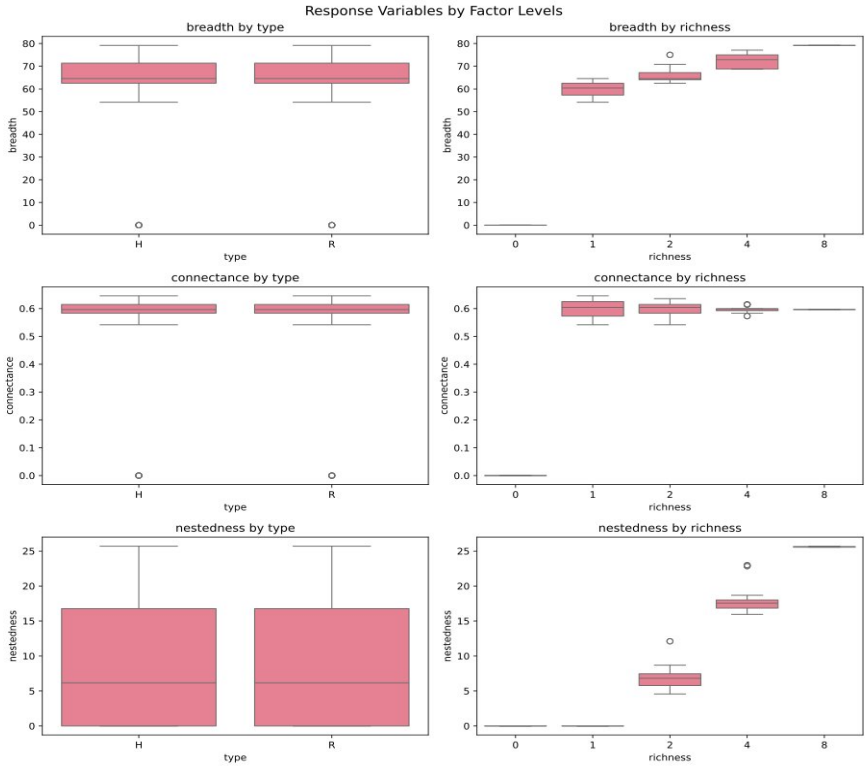
- `tables/H1_A3_feature_importance.csv` - Full feature importance from xgboost on analysis H1_A3. - `tables/H9_A2_feature_importance.csv` - Full feature importance from xgboost on analysis H9_A2.

Appendix D: Exploratory Data Analysis

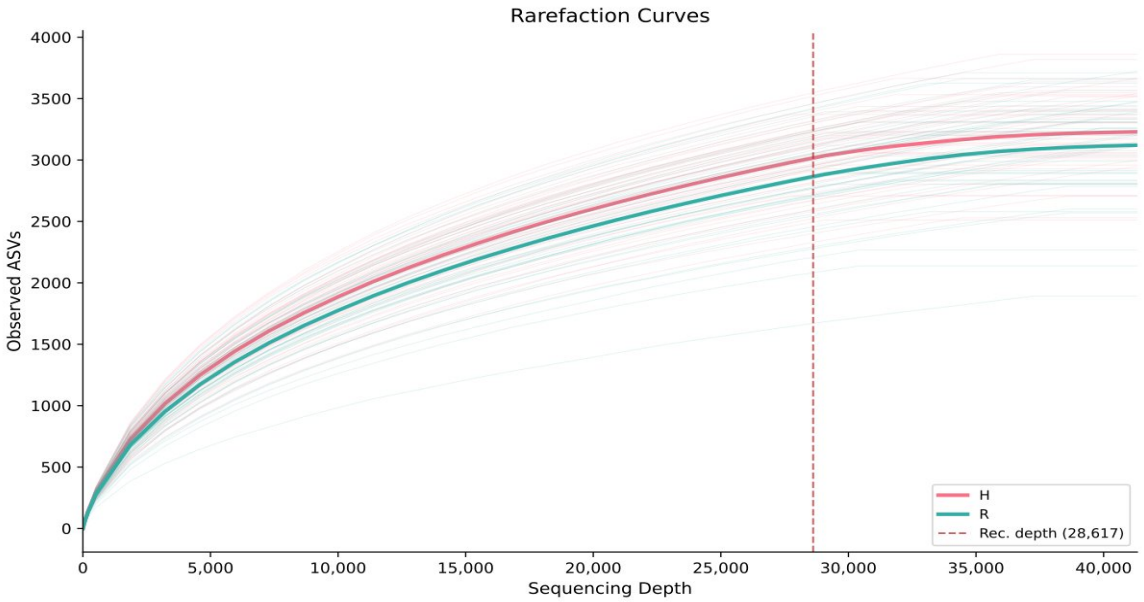


EDA-1: Variable distributions

EDA-2: Correlation matrix for response and covariate variables



EDA-3: Response variables by factor levels



EDA-4: Rarefaction curves showing observed ASVs vs sequencing depth

Appendix E: Analysis Code

Analysis H1_A1 - mixed_effects

```
# Linear Mixed Effects Model # Response variable: dryweight # Fixed effects:
['richness', 'p_solu', 'pH', 'NH4', 'NO3', 'WSOC', 'WSON', 'N', 'C', 'CNRatio',
'DNAc'] # Random effects (grouping): ['community'] # Formula: dryweight ~
richness + p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C + CNRatio + DNAc #
Sample size: n=52, groups=38 import statsmodels.formula.api as smf import
pandas as pd # Define formula for fixed effects formula = "dryweight ~ richness
+ p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C + CNRatio + DNAc" # Fit mixed
model with random intercepts model = smf.mixedlm(formula, df,
groups=df['community']) fitted = model.fit(method='lbfgs') # Model summary
print(fitted.summary()) # Extract key results fixed_effects = fitted.fe_params
random_effect_variance = fitted.cov_re.iloc[0, 0] residual_variance =
fitted.scale icc = random_effect_variance / (random_effect_variance +
residual_variance) # Results: ICC=0.000, Log-Likelihood=8.10 # Execution time:
2.00s
```

Analysis H2_A1 - mixed_effects

```
# Linear Mixed Effects Model # Response variable: di_pseu # Fixed effects:
['richness', 'p_solu', 'pH', 'NH4', 'NO3', 'WSOC', 'WSON', 'N', 'C', 'CNRatio',
'DNAc'] # Random effects (grouping): ['community'] # Formula: di_pseu ~
richness + p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C + CNRatio + DNAc #
Sample size: n=52, groups=38 import statsmodels.formula.api as smf import
pandas as pd # Define formula for fixed effects formula = "di_pseu ~ richness +
p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C + CNRatio + DNAc" # Fit mixed
```

```
model with random intercepts model = smf.mixedlm(formula, df,
groups=df['community']) fitted = model.fit(method='lbfgs') # Model summary
print(fitted.summary()) # Extract key results fixed_effects = fitted.fe_params
random_effect_variance = fitted.cov_re.iloc[0, 0] residual_variance =
fitted.scale icc = random_effect_variance / (random_effect_variance +
residual_variance) # Results: ICC=0.000, Log-Likelihood=10.44 # Execution time:
1.91s
```

Analysis H4_A2 - adaptive_glm

```
# GLM - Gaussian (OLS) Linear Model # Response variable: dryweight # Predictor
variables: ['MVP1_4', 'p_solu', 'pH', 'NH4', 'NO3', 'WSOC', 'WSON', 'N', 'C',
'CNRatio', 'DNAc'] # Formula: dryweight ~ MVP1_4 + p_solu + pH + NH4 + NO3 +
WSOC + WSON + N + C + CNRatio + DNAc # Sample size: n=52 import
statsmodels.formula.api as smf import pandas as pd # Define model formula
(R-style) formula = "dryweight ~ MVP1_4 + p_solu + pH + NH4 + NO3 + WSOC + WSON
+ N + C + CNRatio + DNAc" # Fit Ordinary Least Squares model model =
smf.ols(formula, data=df).fit() # Model summary print(model.summary()) # ANOVA
table (Type 2 SS - auto-selected based on design balance) from
statsmodels.stats.anova import anova_lm anova_table = anova_lm(model, typ=2) #
Results: R²=0.6310, F=6.2173, p=0.0000 # Execution time: 0.68s
```

Analysis H6_A1 - mixed_effects

```
# Linear Mixed Effects Model # Response variable: dryweight # Fixed effects:
['richness', 'OTU1369_abundance', 'OTU6639_abundance', 'OTU17_abundance'] #
Random effects (grouping): ['community'] # Formula: dryweight ~ richness +
OTU1369_abundance + OTU6639_abundance + OTU17_abundance # Sample size: n=52,
groups=38 import statsmodels.formula.api as smf import pandas as pd # Define
formula for fixed effects formula = "dryweight ~ richness + OTU1369_abundance +
OTU6639_abundance + OTU17_abundance" # Fit mixed model with random intercepts
model = smf.mixedlm(formula, df, groups=df['community']) fitted =
model.fit(method='lbfgs') # Model summary print(fitted.summary()) # Extract key
results fixed_effects = fitted.fe_params random_effect_variance =
fitted.cov_re.iloc[0, 0] residual_variance = fitted.scale icc =
random_effect_variance / (random_effect_variance + residual_variance) #
Results: ICC=0.012, Log-Likelihood=22.14 # Execution time: 0.71s
```

Analysis H6_A2 - adaptive_glm

```
# GLM - Gaussian (OLS) Linear Model # Response variable: dryweight # Predictor
variables: ['richness', 'OTU1369_abundance', 'OTU6639_abundance',
'OTU17_abundance'] # Formula: dryweight ~ richness + OTU1369_abundance +
OTU6639_abundance + OTU17_abundance # Sample size: n=52 import
statsmodels.formula.api as smf import pandas as pd # Define model formula
(R-style) formula = "dryweight ~ richness + OTU1369_abundance +
OTU6639_abundance + OTU17_abundance" # Fit Ordinary Least Squares model model =
smf.ols(formula, data=df).fit() # Model summary print(model.summary()) # ANOVA
table (Type 2 SS - auto-selected based on design balance) from
statsmodels.stats.anova import anova_lm anova_table = anova_lm(model, typ=2) #
Results: R²=0.7192, F=16.0994, p=0.0000 # Execution time: 0.15s
```

Analysis H8_A2 - adaptive_glm

```
# GLM - Gaussian (OLS) Linear Model # Response variable: dryweight # Predictor variables: ['MVP1_4', 'pH'] # Formula: dryweight ~ MVP1_4 + pH + MVP1_4:pH # Sample size: n=52 import statsmodels.formula.api as smf import pandas as pd # Define model formula (R-style) formula = "dryweight ~ MVP1_4 + pH + MVP1_4:pH" # Fit Ordinary Least Squares model model = smf.ols(formula, data=df).fit() # Model summary print(model.summary()) # ANOVA table (Type 3 SS - auto-selected based on design balance) from statsmodels.stats.anova import anova_lm anova_table = anova_lm(model, typ=3) # Results: R²=0.2124, F=4.3155, p=0.0090 # Execution time: 0.15s
```

Analysis H3_A1 - permanova

```
# Factorial PERMANOVA (Type I SS) # Response variables: ['asv_table'] # Predictors: ['richness', 'type'] # Interaction terms: [('richness', 'type')] # Distance metric: aitchison # Permutations: 999 # Stratification: None import numpy as np import pandas as pd from scipy.spatial.distance import pdist, squareform # Prepare response data and compute distance matrix # Load composition matrix composition_data = pd.read_parquet('asv_table.parquet') response_data = composition_data.values distance_matrix = squareform(pdist(response_data, metric='aitchison')) # Sequential (Type I) sum of squares approach: # 1. Test each term sequentially # 2. Permutation test for significance (within strata if specified) # 3. F = (partial_SS / df_term) / (residual_SS / df_residual) # Execution time: 4.93s
```

Analysis H3_A2 - rda

```
# RDA (Redundancy Analysis) - Constrained Ordination # Response variables (species/composition): composition table variables # Predictor variables (environment): ['richness', 'type'] # Sample size: n=104 # Permutations: 999 # Stratification: None import numpy as np import pandas as pd from sklearn.preprocessing import StandardScaler from sklearn.manifold import MDS from scipy.spatial.distance import pdist, squareform # Prepare response matrix Y (species/composition data) Y = df[response_vars].values # Shape: (n_samples, n_species) # Prepare predictor matrix X (environmental/treatment variables) X = pd.get_dummies(df[predictors], drop_first=True).values # Center and standardize both matrices Y_centered = StandardScaler().fit_transform(Y) X_centered = StandardScaler().fit_transform(X) # RDA regression: Y = XB + E B = np.linalg.lstsq(X_centered, Y_centered, rcond=None)[0] Y_fitted = X_centered @ B # Calculate R² (variation explained by constraints) r_squared = 0.0537 # Permutation test for significance # Results: R²=0.0537, p=0.0020 # PCoA for ordination visualization (individual sample coordinates) distance_matrix = squareform(pdist(Y_centered, metric="euclidean")) mds = MDS(n_components=3, dissimilarity="precomputed", random_state=42) sample_scores = mds.fit_transform(distance_matrix) # Execution time: 4.68s
```

Analysis H3_A3 - variation_partitioning

```
# Variation Partitioning using Partial RDA # Response: compositional data (104 samples × 6648 features) # Predictor variables (X1): ['richness'] # Conditioning variables (X2): ['type'] # Permutations: 999 # Stratification: None import numpy as np from sklearn.preprocessing import StandardScaler #
```

```
Prepare matrices Y_centered = Y - Y.mean(axis=0) X1 =
StandardScaler().fit_transform(df[['richness']].values) X2 =
StandardScaler().fit_transform(df[['type']].values) # RDA computations (R² =
SS_fitted / SS_total) # For each pure component: Partial RDA controlling for
all other groups # Shared components computed by subtraction # Results: #
R²_full = 0.0537 (p=0.0020) # Pure predictor [a] = 0.0431 # Pure conditioning
[c] = 0.0116 # Shared [b] = 0.0000 # Residual [d] = 0.9463 # Execution time:
6.66s
```

Analysis H4_A1 - mixed_effects

```
# Linear Mixed Effects Model # Response variable: dryweight # Fixed effects:
['MVP1_4', 'type', 'p_solu', 'pH', 'NH4', 'NO3', 'WSOC', 'WSON', 'N', 'C',
'CNRatio', 'DNAc'] # Random effects (grouping): ['community'] # Formula:
dryweight ~ MVP1_4 + type + p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C +
CNRatio + DNAc + MVP1_4:type # Sample size: n=104, groups=38 import
statsmodels.formula.api as smf import pandas as pd # Define formula for fixed
effects formula = "dryweight ~ MVP1_4 + type + p_solu + pH + NH4 + NO3 + WSOC +
WSON + N + C + CNRatio + DNAc + MVP1_4:type" # Fit mixed model with random
intercepts model = smf.mixedlm(formula, df, groups=df['community']) fitted =
model.fit(method='lbfgs') # Model summary print(fitted.summary()) # Extract key
results fixed_effects = fitted.fe_params random_effect_variance =
fitted.cov_re.iloc[0, 0] residual_variance = fitted.scale icc =
random_effect_variance / (random_effect_variance + residual_variance) #
Results: ICC=0.510, Log-Likelihood=23.88 # Execution time: 1.85s
```

Analysis H2_A3 - deseq2

```
# Differential Abundance (DESeq2) - Omnibus Likelihood-Ratio Test # Grouping
variable: richness (5 categories) # One per-feature test across ALL categories
(full ~richness vs reduced ~1). # pydeseq2 has no native LRT, so the NB
likelihood ratio is computed manually. from pydeseq2.dds import DeseqDataSet
from scipy import stats from statsmodels.stats.multitest import multipletests
dds_full = DeseqDataSet(counts=counts, metadata=metadata,
design_factors='richness') dds_full.deseq2() dds_reduced =
DeseqDataSet(counts=counts, metadata=metadata, design='~1')
dds_reduced.deseq2() # Per-feature NB log-likelihood (var = mu + alpha*mu^2):
n=1/alpha, p=n/(n+mu) # lrt = 2*(ll_full - ll_reduced) ~ chi2(df=4); BH-FDR
adjusted. # discriminant_class = category with the highest mean normalized
abundance. # Execution time: 8.54s
```

Analysis H5_A3 - permanova

```
# PERMANOVA # Response variables: ['asv_table'] # Predictors: ['richness'] #
Interaction terms: [] # Distance metric: aitchison # Permutations: 999 #
Stratification: None import numpy as np import pandas as pd from
scipy.spatial.distance import pdist, squareform # Prepare response data and
compute distance matrix # Load composition matrix composition_data =
pd.read_parquet('asv_table.parquet') response_data = composition_data.values
distance_matrix = squareform(pdist(response_data, metric='aitchison')) #
Sequential (Type I) sum of squares approach: # 1. Test each term sequentially #
2. Permutation test for significance (within strata if specified) # 3. F =
(partial_SS / df_term) / (residual_SS / df_residual) # Execution time: 2.49s
```

Analysis H1_A2 - mediation

```
import statsmodels.api as sm from statsmodels.stats.mediation import Mediation
# Ensure column names are valid identifiers (no spaces/special chars) #
Mediator model: breadth ~ richness_1 (unfitted for Mediation) mediator_model =
sm.OLS.from_formula("breadth ~ richness_1", data=df) mediator_fit =
mediator_model.fit() # fitted result for coefficients # Outcome model:
dryweight ~ richness_1 + breadth (unfitted for Mediation) outcome_model =
sm.OLS.from_formula("dryweight ~ richness_1 + breadth", data=df) outcome_fit =
outcome_model.fit() # fitted result for coefficients # Bootstrap mediation
(requires unfitted OLS models, bare column names) med =
Mediation(outcome_model, mediator_model, 'richness_1', 'breadth') result =
med.fit(n_rep=500) print(result.summary())
```

Analysis H1_A3 - xgboost

```
# XGBoost Regression (Gradient Boosting) # Response variable: dryweight #
Predictors: whole asv_table; 6652 features used after dropping # zero-variance
columns (from 6649 candidates) # Task type: regression # Sample size: n=52
(rows with missing dryweight or predictors dropped before fit) #
Hyperparameters: n_estimators=100, max_depth=6,
learning_rate=0.22829856891526404 import xgboost as xgb from
sklearn.model_selection import train_test_split, cross_val_score from
sklearn.metrics import r2_score, mean_squared_error import pandas as pd import
numpy as np # Feature list (showing first 5 of 6649) predictors_shown =
['OTU7010', 'OTU2951', 'OTU4304', 'OTU5006', 'OTU4612'] X =
df[predictors].values y = df['dryweight'].values # Train-test split X_train,
X_test, y_train, y_test = train_test_split( X, y, test_size=0.2,
random_state=42 ) # Initialize and fit XGBoost model model = xgb.XGBRegressor(
n_estimators=100, max_depth=6, learning_rate=0.22829856891526404,
objective='reg:squarederror', random_state=42, n_jobs=-1, eval_metric='rmse' )
model.fit(X_train, y_train) # Cross-validation (5-fold) cv_scores =
cross_val_score(model, X, y, cv=5) # Feature importance feature_importance =
dict(zip(predictors, model.feature_importances_)) # Top 3: ['OTU1369',
'OTU6639', 'OTU17'] # Execution time: 21.65s
```

Analysis H2_A2 - mediation

```
import statsmodels.api as sm from statsmodels.stats.mediation import Mediation
# Ensure column names are valid identifiers (no spaces/special chars) #
Mediator model: breadth ~ richness_1 (unfitted for Mediation) mediator_model =
sm.OLS.from_formula("breadth ~ richness_1", data=df) mediator_fit =
mediator_model.fit() # fitted result for coefficients # Outcome model: di_pseu
~ richness_1 + breadth (unfitted for Mediation) outcome_model =
sm.OLS.from_formula("di_pseu ~ richness_1 + breadth", data=df) outcome_fit =
outcome_model.fit() # fitted result for coefficients # Bootstrap mediation
(requires unfitted OLS models, bare column names) med =
Mediation(outcome_model, mediator_model, 'richness_1', 'breadth') result =
med.fit(n_rep=500) print(result.summary())
```

Analysis H5_A2 - regression

```
# Beta Regression (Smithson-Verkuilen compression) # Response variable: di_pseu
(proportion in [0,1]) # Predictor variables: ['richness', 'toxin2', 'p_solu',
'pH', 'NH4', 'NO3', 'WSOC', 'WSON', 'N', 'C', 'CNRatio', 'DNAc'] # Formula:
di_pseu ~ richness + toxin2 + p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C +
CNRatio + DNAc # Sample size: n=52 import patsy import pandas as pd from
statsmodels.othermod.betareg import BetaModel # Smithson-Verkuilen (2006)
compression: [0,1] → (0,1) n_obs = len(df) df["di_pseu"] = (df["di_pseu"] *
(n_obs - 1) + 0.5) / n_obs # Build design matrices from formula formula =
"di_pseu ~ richness + toxin2 + p_solu + pH + NH4 + NO3 + WSOC + WSON + N + C +
CNRatio + DNAc" endog, exog = patsy.dmatrices(formula, df,
return_type="dataframe") # Fit beta regression model model =
BetaModel(endog.iloc[:, 0], exog).fit() # Model summary print(model.summary())
# AIC and pseudo-R² for model comparison aic = model.aic pseudo_r2 =
model.pseudo_rsquared() # Execution time: 0.72s
```

Analysis H7_A2 - regression

```
# Beta Regression (Smithson-Verkuilen compression) # Response variable: di_pseu
(proportion in [0,1]) # Predictor variables: ['richness', 'siderophore_SU'] #
Formula: di_pseu ~ richness + siderophore_SU # Sample size: n=52 import patsy
import pandas as pd from statsmodels.othermod.betareg import BetaModel #
Smithson-Verkuilen (2006) compression: [0,1] → (0,1) n_obs = len(df)
df["di_pseu"] = (df["di_pseu"] * (n_obs - 1) + 0.5) / n_obs # Build design
matrices from formula formula = "di_pseu ~ richness + siderophore_SU" endog,
exog = patsy.dmatrices(formula, df, return_type="dataframe") # Fit beta
regression model model = BetaModel(endog.iloc[:, 0], exog).fit() # Model
summary print(model.summary()) # AIC and pseudo-R² for model comparison aic =
model.aic pseudo_r2 = model.pseudo_rsquared() # Execution time: 0.14s
```

Analysis H8_A1 - mixed_effects

```
# Linear Mixed Effects Model # Response variable: dryweight # Fixed effects:
['MVP1_4', 'pH'] # Random effects (grouping): ['community'] # Formula:
dryweight ~ MVP1_4 + pH + MVP1_4:pH # Sample size: n=52, groups=38 import
statsmodels.formula.api as smf import pandas as pd # Define formula for fixed
effects formula = "dryweight ~ MVP1_4 + pH + MVP1_4:pH" # Fit mixed model with
random intercepts model = smf.mixedlm(formula, df, groups=df['community'])
fitted = model.fit(method='lbfgs') # Model summary print(fitted.summary()) #
Extract key results fixed_effects = fitted.fe_params random_effect_variance =
fitted.cov_re.iloc[0, 0] residual_variance = fitted.scale icc =
random_effect_variance / (random_effect_variance + residual_variance) #
Results: ICC=0.658, Log-Likelihood=25.36 # Execution time: 0.54s
```

Analysis H9_A1 - gam

```
# GAM (Generalized Additive Model) # Response: flic_pseu # Predictors:
['OTU7320_abundance'] import numpy as np import pandas as pd from pygam import
LinearGAM, s, f # Data preparation X_cols = ['OTU7320_abundance'] y =
df['flic_pseu'].values X = df[X_cols].values # Build terms: s() for continuous,
f() for categorical gam = LinearGAM(s(0)) gam.fit(X, y) # Summary
print(f"Pseudo-R²: {gam.statistics_['pseudo_r2']['explained_deviance']:.3f}")
```

```
print(f"GCV: {gam.statistics_['GCV']:.3f}") # Per-term p-values for i, p in
enumerate(gam.statistics_['p_values']): print(f"Term {i}: p = {p:.4f}") #
Execution time: 0.72s
```

Analysis H9_A2 - xgboost

```
# XGBoost Regression (Gradient Boosting) # Response variable: flic_pseu #
Predictors: whole asv_table; 6648 features used after dropping # zero-variance
columns (from 6648 candidates) # Task type: regression # Sample size: n=52
(rows with missing flic_pseu or predictors dropped before fit) #
Hyperparameters: n_estimators=200, max_depth=6,
learning_rate=0.014345659706845676 import xgboost as xgb from
sklearn.model_selection import train_test_split, cross_val_score from
sklearn.metrics import r2_score, mean_squared_error import pandas as pd import
numpy as np # Feature list (showing first 5 of 6648) predictors_shown =
['OTU7010', 'OTU2951', 'OTU4304', 'OTU5006', 'OTU4612'] X =
df[predictors].values y = df['flic_pseu'].values # Train-test split X_train,
X_test, y_train, y_test = train_test_split( X, y, test_size=0.2,
random_state=42 ) # Initialize and fit XGBoost model model = xgb.XGBRegressor(
n_estimators=200, max_depth=6, learning_rate=0.014345659706845676,
objective='reg:squarederror', random_state=42, n_jobs=-1, eval_metric='rmse' )
model.fit(X_train, y_train) # Cross-validation (5-fold) cv_scores =
cross_val_score(model, X, y, cv=5) # Feature importance feature_importance =
dict(zip(predictors, model.feature_importances_)) # Top 3: ['OTU3074',
'OTU7510', 'OTU7728'] # Execution time: 36.27s
```

Appendix F: Failed Figures

- H2_F2_volcano_asv - Figure toolkit error: Tool execution failed: Volcano plot generation failed: Analysis result has no DIFFERENTIAL_RESULTS artifact; DESeq2 figures require the per-feature parquet persisted by the DESeq2 tool.

Appendix G: References

[1] Trivedi P, Leach JE, Tringe SG, et al. Plant-microbiome interactions: from community assembly to plant health. Nature reviews. Microbiology (2020). doi:10.1038/s41579-020-0412-1. [2] Fan X, Ge AH, Qi S, et al. Root exudates and microbial metabolites: signals and nutrients in plant-microbe interactions. Science China. Life sciences (2025). doi:10.1007/s11427-024-2876-0. [3] Beltran-Garcia MJ, Martinez-Rodriguez A, Olmos-Arriaga I, et al. Probiotic Endophytes for More Sustainable Banana Production. Microorganisms (2021). doi:10.3390/microorganisms9091805. [4] Gupta A, Singh UB, Sahu PK, et al. Linking Soil Microbial Diversity to Modern Agriculture Practices: A Review. International journal of environmental research and public health (2022). doi:10.3390/ijerph19053141. [5] Gao C, Bezemer TM, de Vries FT, et al. Trade-offs in soil microbial functions and soil health in agroecosystems. Trends in ecology & evolution (2024). doi:10.1016/j.tree.2024.05.013. [6] Zong D, Zhou Y, Zhou J, et al. Soil microbial community composition by crop type under rotation diversification. BMC microbiology (2024). doi:10.1186/s12866-024-03580-2. [7] Cappelli SL, Domeignoz-Horta LA, Loaiza V, et al. Plant biodiversity promotes sustainable agriculture directly and via belowground effects. Trends in plant science (2022). doi:10.1016/j.tplants.2022.02.003. [8] Kramer J, Özkaya Ö, Kümmerli R. Bacterial siderophores in community and host interactions. Nature reviews. Microbiology (2020). doi:10.1038/s41579-019-0284-4. [9] Yue Y, Xu Z, Wang Y, et al. Soil-dwelling Naegleria enhances plant performance by stimulating beneficial bacterial

functions in the rhizosphere. Nature communications (2025). doi:10.1038/s41467-025-64139-x. [10] Albarrán-Cuitiño C, Palma DE, González H, et al. Development of a plant growth-promoting bacterial EcoBiome derived from desert soil isolates. Applied and environmental microbiology (2026). doi:10.1128/aem.00103-26. [11] Toribio AJ, Suárez-Estrella F, Jurado MM, et al. Design and validation of cyanobacteria-rhizobacteria consortia for tomato seedlings growth promotion. Scientific reports (2022). doi:10.1038/s41598-022-17547-8.