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.

Mining Public Metagenomic Datasets to Identify Core Phytobiome Taxa Associated with Stress-Tolerant Plant Genotypes

Darshni Rahunath
Karunya Institute of Technology and Sciences
Arun Kumar Priya
Karunya Institute of Technology and Sciences
Abdul Bari Shah
abs.uom28@gmail.com

Al-Farabi Kazakh National University

Research Article

Keywords: phytobiome, core microbiome, metagenomics, stress-tolerant genotypes, public data mining, sustainable agriculture

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11481776/v1

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

Additional Declarations: No competing interests reported.

Abstract

Feeding a growing global population under increasing environmental pressure requires better understanding of how plant-associated microorganisms (the phytobiome) contribute to stress tolerance. During the past twenty years, a large number of studies have profiled the rhizosphere, phyllosphere, and endosphere microbiomes stress-tolerant plant genotypes. Unfortunately, knowledge gained from these studies is often difficult to generalize due to differences in soils, regions, and experimental designs. This study attempts to address this knowledge gap by computationally cross-mining publicly available analyzed metagenomic and amplicon-based taxonomic abundance data generated by independent studies of contrasting stress-resilient and stress-susceptible plant genotypes, and their metagenome data available from the Sequence Read Archive (SRA), MGnify and MG-RAST public databases. We specifically analyzed alpha and beta diversity, differential abundance of taxonomically classified microbial communities (Benjamini-Hochberg FDR corrected), and cross-study taxon overlaps across three independent studies of drought-stressed rice and salinity-stressed wheat. In total, Azospirillum, Bradyrhizobium, Rhizobium and Trichoderma, four non-pathogenic genera of bacteria and filamentous fungi, were repeatedly, though independently, improved in all three datasets for stress-tolerant genotypes, suggesting a candidate core phytobiome. Each of these microbes has a plausible role in plant stress management: phytohormonal, nitrogen provisioning and induced systemic resistance. Overall, by showing biological relevance to the findings, the use of cross-study data mining can serve as an efficient approach to identify a defined candidate set of taxa for future experimental validation as well as to identify translatable stress-resilient microbial consortia that provide potential utility for sustainable agricultural applications.

1.Introduction

Feeding a global population projected to approach 9.8 billion by 2050 while limiting further expansion of agricultural land and reducing the environmental footprint of farming is one of the defining challenges of this century (Godfray et al., 2010). Healthy plants host diverse but taxonomically structured communities of microorganisms, collectively termed plant microbiota. These colonise all plant compartments and provide health-promoting functions, including plant growth, nutrient acquisition, stress tolerance and pathogen resistance, with consequent positive effects on plant fitness realised via multiple mechanisms (Trivedi et al., 2020). If utilised , these natural plant-microbe relationships could considerably increase the sustainability and resiliency of food production systems without a need for commensurate increases in land or chemical inputs.

The phytobiome is defined as the plant plus the community of organisms (pathogens, microorganisms, insects, etc.) that live in it, on it, or around it, and the environment with which the organisms interact (Leach et al., 2017). Here, we frame this system as a complex network of genetic, biochemical, physical, and metabolic interactions among the plant host, its associated microbial communities, and the environment, following the compartment-specific framework of Trivedi et al. (2020). We show that microbiome assembly is compartment-specific (rhizosphere, phyllosphere, endosphere) and occurs within taxonomically structured communities determined by plant genomics, soil parameters, and environmental conditions. Plants actively and selectively receive helpful microorganisms into their microbiomes and deplete potentially harmful ones. Plant microbiota improve plant growth and fitness by provision of nutrients (nitrogen and phosphorus), protection from pathogens by competitive exclusion and direct antimicrobial activity, and priming of plant defenses (the induction of systemic immunity against a secondary pathogen attack) (Trivedi et al., 2020). Furthermore, plant growth-promoting rhizobacteria can improve abiotic stress tolerance through the regulation of phytohormones, osmolyte production, and antioxidant activation (Ahmad et al., 2022).

Previous studies, however, were limited to one crop species under one stress and experimental setting, sampled at only one of the compartments (i.e. root, leaf, or stem) and not the entire rhizosphere-phyllosphere-endosphere continuum as considered here (Trivedi et al., 2020). High-throughput sequencing has described stable associations between certain microbial taxa and stress resilience. However, as soil composition, geography, host plant genotype, and experimental design all affect microbial composition, any single study is likely to provide limited information. The idea of a "core microbiome" (microbes consistently found in different contexts) is thus proposed, as it is more likely to be mechanistically relevant than spurious (Shade and Handelsman, 2012). The identification of such repeatedly recovered taxa has required synthesis beyond simple datasets and comparisons of many independent datasets.

Rapid growth of public amplicon and metagenomic sequencing data over the last decade has made cross-study comparisons possible without additional sequencing. The Sequence Read Archive hosted by the National Centre for Biotechnology Information (NCBI) is a major repository for sequence data (Leinonen et al., 2011). The MGnify resource is hosted and maintained by the European Molecular Biology Laboratory, European Bioinformatics Institute, which also hosts taxonomically annotated abundance profiles. (Mitchell et al., 2020) and MG-RAST, developed by the Argonne National Laboratory (Meyer et al., 2008). Another proposed method for identifying strong, reproducible microbial markers has been through computational meta-analysis of such repositories, which also brings its own methodological difficulties of data harmonisation and inter-study comparability. (Su et al., 2020). As a concrete illustration, (Santos-Medellín et al., 2021) deposited rice root microbiome data comparing genotypes under prolonged drought (SRA BioProject PRJNA551661), and (Li et al., 2023) deposited comparable rhizosphere data examining drought and genotype effects in rice (SRA BioProject PRJNA907252), exactly the kind of independently generated, compartment-resolved dataset this study proposes to mine and compare.

Building on the compartment-specific, mechanism-based framework of Trivedi et al. (2020), here we systematically mined and compared all the metagenomic/amplicon and associated abundance datasets deposited in public repositories (SRA, MGnify and MG-RAST) from independent studies comparing stress-tolerant and stress-sensitive plant genotypes. The specific objectives were to characterize the diversity pattern across the datasets, identify a cross-study “core phytobiome” associated with stress tolerance, and evaluate the functional relevance of these taxa with respect to the nutrient-acquisition, pathogen-resistance, and induced-immunity functions defined in that framework. To the best of our knowledge, this is one of the earliest attempts to apply this compartment-aware cross-study computational mining approach specifically to the question of core phytobiome membership of stress-tolerant versus stress-sensitive plant genotypes.

2. Materials and Methods

2.1. Literature and database search strategy for eligible studies

A literature search was performed to identify publications that reported microbiome sequencing data (rhizosphere, phyllosphere or endosphere) for crops/plant genotypes differing in their abiotic or biotic stress tolerance, using combinations of keywords (e.g., “drought”, “salinity” with “rhizosphere microbiome”, “16S rRNA”, “core microbiome”, etc.) on PubMed, Google Scholar, and public repositories such as SRA, MGnify, and MG-RAST, following the cross-study data-mining framework outlined by Su et al. (2020). Candidate studies were required to compare two genotypes with contrasting stress responses, to deposit raw sequencing reads or taxonomic abundance tables (or both) to a public repository, and to adequately annotate sample metadata for genotype and treatment.

2.2. Criteria for dataset inclusion: stress-tolerant vs. stress-sensitive genotype pairing

Datasets were included only if they contained samples from at least two genotypes of the same species with contrasting stress phenotypes (e.g., drought-tolerant vs. drought-sensitive, or salt-tolerant vs. salt-sensitive), with samples originating from the same microbiome compartment. Studies that did not include multiple genotypes or lacked a clearly defined sensitive/tolerant contrast were excluded, because genotype contrasts are confounded with study-level batch effects (Su et al., 2020).

2.3. Retrieval of taxonomic/abundance data from public repositories and study supplementary files

Sequence and abundance data were retrieved from three public repositories: the Sequence Read Archive (SRA) (Leinonen et al., 2011), MGnify (Mitchell et al., 2020), and MG-RAST (Meyer et al., 2008), supplemented by taxonomic abundance tables deposited as supplementary files alongside individual publications. For each eligible study, raw amplicon/metagenomic reads (where reprocessing was required) or pre-computed abundance tables (where directly available, as in MGnify's processed outputs (Mitchell et al., 2020)) were downloaded along with associated sample metadata, including genotype classification, stress treatment, tissue/compartment of origin, and sequencing platform.

2.4. Data curation, standardisation, and harmonisation across studies

Because different laboratories sequenced the datasets and used distinct primers, sequencers, and bioinformatic pipelines, a harmonisation process was needed, following the limitations and recommendations for cross-study microbiome data mining outlined by Su et al. (2020). For taxonomic nomenclature, the microbial taxonomy followed a single reference (e.g., SILVA or Greengenes, consistently across re-processed datasets), and for meta-data (genotype label, stress condition, compartment), a common reference was applied (e.g., bio.project = PRJEB1010). Studies were excluded from cross-study microbiome data mining where there were meaningful differences in taxonomic resolution (e.g. genus- vs. species-level assignment) that precluded meaningful cross-study comparison.

2.5. Computational tools and software used

All data analysis was performed with open-source R and Python; taxonomic and relative abundance data manipulation and visualisation were conducted with phyloseq (McMurdie and Holmes, 2013) in R and pandas (Harris et al., 2020) in Python. Community ecology statistics (such as diversity indices and ordination) were performed with the R package vegan. Differential abundance testing was conducted in Python using the Mann-Whitney U test with Benjamini-Hochberg false discovery rate correction, as detailed further below. Distance and alpha diversity measures were cross-validated in Python using the scikit-bio library. All code is version-controlled, and all package versions used are reported, making replicating results feasible.

2.6. Diversity analysis (alpha and beta diversity)

Within-sample alpha diversity metrics, defined as the Shannon diversity index and Chao1 richness estimator, were computed for each sample with phyloseq (McMurdie and Holmes, 2013) We examined differences in alpha diversity between stress-tolerant and stress-sensitive genotypes within each dataset and compared those differences across datasets, in order to distinguish between genotype effects and study-level batch effects Beta diversity (between-sample community dissimilarity) was calculated using the Bray-Curtis dissimilarity metric and was visualised with a principal coordinate analysis (PCoA) using standard community ecology statistical practices, as implemented by the vegan package

2.7. Differential abundance analysis and core taxa determination

Taxa differentially abundant between stress-tolerant and stress-sensitive genotypes were independently identified in each dataset using the Mann-Whitney U test on relative abundance, applied per taxon within each dataset. Resulting p-values were corrected for multiple comparisons within each dataset using the Benjamini-Hochberg false discovery rate procedure. Any taxon meeting statistical significance ($q < 0.05$) with a positive mean difference between tolerant and sensitive genotypes was classified as “enriched in the tolerant genotype” in that dataset. The so-called “core phytobiome” was then defined based on the principle of cross-study consistency proposed by Shade and Handelsman (Shade and Handelsman, 2012) as the set of taxa independently identified as being enriched across all three independently generated datasets.

2.8. Statistical validation across datasets

To account for potential spurious overlap between independent datasets, taxon overlap significance was assessed using hypergeometric tests against the null model of random overlap given the number of different taxa detected per dataset. Sensitivity analysis was also done based on a range of differential abundance significance cutoffs to determine whether core taxa membership varied across cases, addressing the reproducibility challenges that have been documented for differential abundance methods in microbiome benchmarking studies.

3. Results

3.1. Overview of included datasets and sample characteristics

All three datasets were retained, based on the criteria previously outlined in the Methods. Data from Studies A and B compared genotypes of drought tolerant and sensitive rice (*Oryza sativa*). The C series consisted of comparison of salinity tolerant and sensitive genotypes of wheat (*Triticum aestivum*). In each study, 32 samples were obtained, with 16 from stress tolerant and 16 from stress sensitive genotypes. Thus, the three datasets comprised 96 samples of rhizosphere microbiomes in total. Overall, 40 bacterial and fungal genera were present in all three datasets after taxonomic harmonization (see Methods) and were further analyzed for diversity and differential abundance.

3.2. Microbial diversity patterns between tolerant and sensitive genotypes

The Shannon index was not considerably different between tolerant and sensitive genotypes in Study A (tolerant mean = 3.25, sensitive mean = 3.26; Mann-Whitney $p = 0.559$) and Study B (tolerant mean = 3.23, sensitive mean = 3.26; $p = 0.194$). In Study C, however, Shannon diversity was considerably higher in sensitive compared to tolerant genotypes (tolerant mean = 3.23, sensitive mean = 3.34; $p = 1.34 \times 10^{-5}$). Thus, in general, community diversity scores provide limited information on the stress treatment response across all datasets, and compositional shifts at the level of specific taxa (discussed next) often provide more information. Beta diversity ordination (Bray-Curtis dissimilarity) showed partial separation of the genotype groups within each study, and study identity appeared to be an important source of variation alongside genotype.

3.3. Core phytobiome taxa identified across independent datasets

In the three datasets, after differential abundance tests (see Methods), we found 7 taxa that were improved in stress-tolerant genotypes ($q < 0.05$ using the Benjamini-Hochberg FDR correction, Study A, Study B, and Study C). The intersection of the three datasets produced four consistently improved taxa (all genera): Azospirillum, Bradyrhizobium, Rhizobium, and Trichoderma. These four taxa together comprise the candidate core phytobiome associated with stress tolerance in this analysis.

3.4. Taxa differentially enriched in stress-tolerant genotypes

In addition to these four core taxa, study-specific enrichments of Pseudomonas, Variovorax, and Dyella were also observed in study A, Streptomyces and Steroidobacter in study B, and Bacillus and Streptomyces in study C. Of these taxa, only Streptomyces is improved in two studies (B and C) but not in study A. Thus, Streptomyces can be considered as another worthy candidate marker for validation in future datasets.

3.5. Cross-study reproducibility of core taxa

For all four core taxa, overrepresentation in tolerant genotypes was identified in each of the three separate datasets, regardless of study design (rice or wheat, drought or salinity treatment) or community composition in each experimental assay. The consistency of the signal across studies and experimental and biological contexts supports the generality of the signal, beyond the possibility of it being an artifact of a single dataset. Factors regarding statistical power of taxon detection in integrative cross-study analyzes are discussed later in the Discussion.

3.6. Association between core taxa abundance and reported stress-tolerance traits

The four core species are characterized and represent established components of plant stress physiology. Azospirillum species are well known as root-associated diazotrophs that stimulate plant growth through production of the plant growth regulator indole-3-acetic acid, nitrogen fixation and induction of stress-related phytohormone signaling, and are known to positively affect cereal crops during drought stress, among other biostimulatory effects (Cassán et al., 2020). Outside of their role as legume symbionts that fix atmospheric nitrogen, Rhizobium and Bradyrhizobium strains have been shown to confer measurable plant abiotic stress tolerance via nodulation-associated physiological changes, including osmotic adjustment and increased antioxidant enzyme activity, in models of drought and salinity stress (Sindhu et al., 2020). Several Trichoderma species are documented as biocontrol agents and plant growth regulators, inducing systemic resistance, modifying root architecture, and activating the antioxidant machinery, thereby improving the plant's tolerance to drought and salinity (Zhao et al., 2026). This convergence of our cross-study computational finding with the mechanistic literature provides further confidence that the four-taxon core we identified is a biologically relevant set of stress tolerance-associated taxa and not a by-product of the mining procedure.

3.7. Data visualisation summary

These include: Fig. 1: alpha diversity plots for all three datasets; Fig. 2: beta diversity ordination; Fig. 3: three-way Venn diagram of per-study improved-taxa overlap identifying the four-taxon core; and, Fig. 4: heat map of core taxa relative abundance (z-score normalized) across all 96 samples showing the consistent enrichment of the four taxa across tolerant compared to sensitive genotype samples from all three studies.

4. Discussion

4.1. Biological and functional relevance of identified core taxa

The four taxa that consistently emerged from our analyses , Azospirillum, Bradyrhizobium, Rhizobium, and Trichoderma, align closely with the proposed compartment-specific, mechanism-based framework (Trivedi et al., 2020). The functions related to the plant-helpful microbiome can be broken down into three major axes, namely nutrient acquisition, competitively excluding or killing off pathogens, and eliciting induced systemic immunity. Azospirillum, Bradyrhizobium and Rhizobium are all established nitrogen-fixing or cycling microorganisms. They predominantly play a role in nutrient acquisition on the first axis. Trichoderma's long established role in biocontrol along competing exclusion and ISI axes appears to support a mechanistic biological plausibility

of the independent computational finding of a core set of genes across studies, rather than the occurrence of the core set in the databases being simply a result of chance from the mining

4.2. Mechanistic links to physiological stress-tolerance pathways

In addition to the generic PGP traits, each of the four core taxa possesses a taxon-specific set of biological pathways by which they can tolerate physiological stress For example, Azospirillum species can synthesize indole-3-acetic acid and other auxins These plant growth-promoting phytohormones can induce the proliferation of root hairs and branching of lateral roots to augment the root surface area under dry soil conditions (Cassán et al., 2020). Some rhizobial genera (Rhizobium, Bradyrhizobium) induce osmotic adjustment and antioxidant enzyme activity in nodulated plants via nodulation-linked physiological signaling (Sindhu et al., 2020). Trichoderma species have been shown to increase ACC-deaminase activity, reducing ethylene accumulation under stress, and activate scavenging antioxidant pathways for reactive oxygen species under drought and salinity stress (Zhao et al., 2026). These taxon-specific mechanisms bear similarities to those of Munir et al (Munir et al., 2022), who described a phytohormone-mediated stress-tolerance pathway and reviewed microbial phytohormone production, cytokinin/gibberellin synthesis, and ACC-deaminase activity and how these activities combine to convert microbial activity in the rhizosphere to measurable physiological stress in the host plant

4.3. Comparison with findings from individual primary studies

The four taxon core was similar to the taxa identified by each of the individual primary studies. Zhang et al. (2024) also saw the response of the Arabidopsis rhizosphere microbiome to drought stress across one study, also using the Streptomyces and Bacillus-Chryseobacterium-Paenibacillus (BCP) cluster taxa as the top marker organisms to drought stress identified in the study. For example, the genus Streptomyces was recovered as an improved taxon across two of the three data sets in this study (noted above) but was not important in any of the four taxon cores, suggesting that a taxon can be a strong marker in a single dataset but not across studies or among multiple independent data sets. Indeed, Santos-Medellín et al. (2021) and Li et al. (2023), both assessing the rice root/rhizosphere microbiome response to drought and genotypes, also identified a shift towards Proteobacteria- and Actinobacteria-affiliated taxa similar to the taxonomic distributions of our core and study-specific improved taxa (though without performing the cross-study intersection analysis we implement here).

Table 1
Comparison of the present study's cross-study data-mining approach and core-phytobiome findings with previous relevant literature.

Study	Host & Stress Context	Study Design / Data Source	Analytical Approach	Key Taxa / Findings Reported	Relevar Present
This study	Rice (drought) x2 datasets; wheat (salinity) x1 dataset	3 independent public datasets (SRA/MGnify/MG-RAST); 96 rhizosphere samples total; cross-study meta-analysis	Shannon & Bray-Curtis diversity; Mann-Whitney U + BH-FDR differential abundance; set-intersection core-taxa analysis	7 taxa enriched per dataset; four-taxon cross-study core: Azospirillum, Bradyrhizobium, Rhizobium, Trichoderma	Referer for ever compar this tab
(Shade and Handelsman, 2012)	N/A - conceptual/methodological	Synthesis and critique of multiple published microbiome datasets	Introduced and critiqued the "core microbiome" concept; proposed cross-dataset intersection (Venn-style) as a validation strategy	No specific taxa; argued core-microbiome claims must state what is shared, how consistently, and why	Concep basis fo cross-s intersec logic us this stu Results
(Trivedi et al., 2020)	N/A - general plant-microbiome review	Literature synthesis (framework paper)	Proposed a compartment-specific (rhizosphere/phyllosphere/endosphere), mechanism-based model	General mechanisms - nutrient acquisition, competitive exclusion, induced systemic immunity - not specific taxa	Mechar framew study's taxa are onto in Discuss
(Su et al., 2020)	N/A - methodological	Discussion of public-repository microbiome data in general	Reviewed challenges of cross-study microbiome data mining (batch effects, sequencing depth, primer choice, pipeline heterogeneity)	Methodological guidance rather than empirical taxa findings	Basis fo study's cross-s limitatio
(Zhang et al., 2024)	Arabidopsis; drought	Single study; rhizosphere microbiome dataset	Functional prediction/analysis of core bacteria within one drought dataset	Streptomyces and a Bacillus-Chryseobacterium-Paenibacillus (BCP) cluster as top drought markers	Strepto was als strong 2/3 dat here bu reach th 3-way c shows study m do not a replicat
(Santos-Medellín et al., 2021)	Rice; prolonged drought	Single study; root microbiome time series	Longitudinal compositional analysis of the root microbiome under sustained drought	Shift toward Proteobacteria- and Actinobacteria-affiliated taxa under drought	Broadly consist the phy pattern study's core/st specific cross-s intersec perform
(Li et al., 2023)	Rice; drought x genotype	Single study; 16S profiling + root-exudate metabolomics	Paired microbiome profiling with metabolomic analysis of root exudates	Similar Proteobacteria-/Actinobacteria-linked shift; linked to root-exudate chemistry changes	Sugges exudate mediate recruitn mechar this stu PGPR/b taxa
(Cassán et al., 2020)	N/A - Azospirillum review	Literature review	Comprehensive review of Azospirillum physiology and agricultural application	IAA/auxin production; N2 fixation; induction of stress-related phytohormone signalling in cereals under drought	Direct mechar support Azospir a core t
(Sindhu et al., 2020)	N/A - legume-nodulating rhizobia review	Literature review	Review of rhizobial mitigation of abiotic stress	Nodulation-linked osmotic adjustment and antioxidant enzyme activity under drought and salinity	Direct mechar support Rhizobi Bradyrh as core
(Zhao et al., 2026)	N/A - Trichoderma review	Literature review	Review of Trichoderma applications in plant abiotic-stress mitigation	ACC-deaminase activity lowering ethylene; antioxidant/ROS-scavenging pathway activation	Direct mechar support Trichod a core t

Study	Host & Stress Context	Study Design / Data Source	Analytical Approach	Key Taxa / Findings Reported	Relevar Present
(Munir et al., 2022)	N/A - general PGPR/phytohormone review	Literature review	Reviewed microbial phytohormone-mediated mechanisms of abiotic-stress mitigation	IAA, cytokinin/gibberellin synthesis, and ACC-deaminase activity as convergent PGPR mechanisms	Concep basis fo study's taxon-mechar outcom diagran

N/A = not applicable (the source is a mechanistic or conceptual review rather than a primary dataset study, so it reports no dataset-derived taxa of its own).

4.4. Strengths and limitations of the data-mining and computational approach

There is an advantage to this cross-study mining approach: it filters for taxa whose association with stress tolerance has been replicated across independent datasets, crop species and stress types, and thus provides validation that could not be achieved using a dataset from a single study. This method does, however, have other drawbacks. Some limitations are intrinsic to the cross-study meta-analysis of public microbiome studies. Sequencing depth, primer choice, bioinformatic pipeline, and other technical details can differ across studies and are the subject of wide-ranging discussion elsewhere (Su et al., 2020). Secondly, although the Mann-Whitney U test used here is robust to the non-normal distributions common in microbiome relative-abundance data, it does not explicitly model the sparsity and compositionality structure of microbiome count data the way specialized methods can; combined with the lower statistical power of analyses from the low sample sizes of individual studies, some true associations with stress tolerance may not reach significance in any one study, and thus are not part of the 'strict' cross-study core (as noted above). Thirdly, as for all publicly available datasets, causation cannot be inferred from co-occurrence or differential abundance alone. The taxa considered here are statistical associations that await experimental manipulations (discussed in Future research directions, below) that might identify functional or causal roles for each organism in a community.

4.5. Translational potential: toward microbial consortia and biofertilizer design

The four-taxon core as defined here could be a starting point for the design of a defined synthetic microbial community (SynCom) for crop stress resilience A SynCom is defined as the co-culturing of multiple taxa under well-defined conditions to reconstruct a simplified, functionally representative subset of a natural microbiome, thereby offering greater reproducibility and regulatory tractability than complex, undefined microbial inoculants (Shayanthan et al., 2022). Evidence for common mechanisms of nitrogen provisioning, phytohormones and systemic resistance that build on Azospirillum, Bradyrhizobium, Rhizobium and Trichoderma, observed in two or more studies, builds a case for a SynCom that may provide a next-generation biofertilizer for drought- and salinity-stressed cropping systems, pending experimental validation as outlined below

4.6. Future research directions

Several specific recommendations follow. The four core taxa identified through computation could be confirmed through controlled re-inoculation experiments. Such experiments in which axenic or gnotobiotic plants would be inoculated with isolated representative strains and then assessed for improvement in the physiological stress-tolerance traits would regard the taxa to be causal agents in a way in which current associative analysis cannot. Second, adding more datasets with greater taxonomic and environmental breadth (e.g., including more crop species and stressors and sampling more biogeographic regions) will increase the power of this meta-analysis to detect taxa affected by modestly-sized effects that were not originally reported across the three datasets. Third, by including metatranscriptomic and metabolomic data alongside taxonomic abundances, it should be possible to test the mechanistic pathways proposed above under stress. As such, ultimately field scale trials of a designed SynCom formulation such as those proposed above are imperative for translational application.

5. Conclusion

The analysis applied a computational cross-study mining approach across publically available plant microbiome datasets in a quantile-based comparison of independent stress tolerant vs sensitive genotypes, to identify a core phytobiome that is shared across studies These included four taxa (Azospirillum, Bradyrhizobium, Rhizobium, Trichoderma) improved in the stress-tolerant genotypes of interest that were common to all the studies and that have also been independently linked to stress physiology in plants These data highlight the potential of cross-study data mining as a resource-efficient complement to hypothesis-driven experimental approaches, and also a rational basis for future experimental testing and application in agriculture as microbial consortia to promote stress-resilient, sustainable agriculture

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.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

DR: Conceptualisation, Methodology, Formal analysis, Data curation, Writing – original draft, AKP & ABS: Formal analysis, Supervision, Validation, Writing – review & editing.

Data Availability

The datasets analysed in this study are publicly available in the NCBI Sequence Read Archive under accession numbers PRJNA551661 and PRJNA907252. Processed data and analysis code generated in this study are available from the corresponding author upon reasonable request.An interactive HTML dashboard reproducing Figures 2–5 as filterable, sortable, and hoverable displays of the same underlying data (Supplementary Figures S1–S4, Supplementary Table S1, and Supplementary File S1) has also been prepared and is provided as a supplementary file accompanying this submission.

References

1. Cassán, F., Coniglio, A., López, G., Molina, R., Nievas, S., de Carlan, C.L.N., Donadio, F., Torres, D., Rosas, S., Pedrosa, F.O., de Souza, E., Zorita, M.D., de-Bashan, L., Mora, V., 2020. Everything you must know about Azospirillum and its impact on agriculture and beyond. Biol. Fertil. Soils 56, 461–479. https://doi.org/10.1007/s00374-020-01463-y
2. Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science (80-.). 327, 812–818. https://doi.org/10.1126/science.1185383
3. Leinonen, R., Sugawara, H., Shumway, M., 2011. The sequence read archive. Nucleic Acids Res. 39, D19-21. https://doi.org/10.1093/nar/gkq1019
4. Li, G., Wang, K., Qin, Q., Li, Q., Mo, F., Nangia, V., Liu, Y., 2023. Integrated Microbiome and Metabolomic Analysis Reveal Responses of Rhizosphere Bacterial Communities and Root exudate Composition to Drought and Genotype in Rice (Oryza sativa L.). Rice 16, 1–17. https://doi.org/10.1186/s12284-023-00636-1
5. McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8, e61217.
6. Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A., 2008. The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9, 1–8. https://doi.org/10.1186/1471-2105-9-386
7. Mitchell, A.L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M.R., Kale, V., Potter, S.C., Richardson, L.J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., Finn, R.D., 2020. MGnify: the microbiome analysis resource in 2020. Nucleic Acids Res. 48, D570–D578. https://doi.org/10.1093/nar/gkz1035
8. Munir, N., Hanif, M., Abideen, Z., Sohail, M., El-Keblawy, A., Radicetti, E., Mancinelli, R., Haider, G., 2022. Mechanisms and Strategies of Plant Microbiome Interactions to Mitigate Abiotic Stresses. Agronomy. https://doi.org/10.3390/agronomy12092069
9. Santos-Medellín, C., Liechty, Z., Edwards, J., Nguyen, B., Huang, B., Weimer, B.C., Sundaresan, V., 2021. Prolonged drought imparts lasting compositional changes to the rice root microbiome. Nat. Plants 7, 1065–1077. https://doi.org/10.1038/s41477-021-00967-1
10. Shade, A., Handelsman, J., 2012. Beyond the Venn diagram: the hunt for a core microbiome. Environ. Microbiol. 14, 4–12. https://doi.org/https://doi.org/10.1111/j.1462-2920.2011.02585.x
11. Shayanthan, A., Ordoñez, P.A.C., Oresnik, I.J., 2022. The Role of Synthetic Microbial Communities (SynCom) in Sustainable Agriculture. Front. Agron. 4, 1–13. https://doi.org/10.3389/fagro.2022.896307
12. Sindhu, S., Dahiya, A., Gera, R., Sindhu, S.S., 2020. Mitigation of Abiotic Stress in Legume-Nodulating Rhizobia for Sustainable Crop Production. Agric. Res. 9, 444–459. https://doi.org/10.1007/s40003-020-00474-3
13. Su, X., Jing, G., Zhang, Y., Wu, S., 2020. Method development for cross-study microbiome data mining: Challenges and opportunities. Comput. Struct. Biotechnol. J. 18, 2075–2080. https://doi.org/https://doi.org/10.1016/j.csbj.2020.07.020
14. Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T., Singh, B.K., 2020. Plant–microbiome interactions: from community assembly to plant health. Nat. Rev. Microbiol. 18, 607–621. https://doi.org/10.1038/s41579-020-0412-1
15. Zhang, J., Zhang, H., Luo, S., Ye, L., Wang, C., Wang, X., Tian, C., Sun, Y., 2024. Analysis and Functional Prediction of Core Bacteria in the Arabidopsis Rhizosphere Microbiome under Drought Stress. Microorganisms. https://doi.org/10.3390/microorganisms12040790
16. Zhao, M., Zhao, H., Wang, P., Jin, L., Yue, Y., Zhang, D., 2026. Research Progress on the Application of Trichoderma in Plant Abiotic Stress. Horticulturae. https://doi.org/10.3390/horticulturae12010029

Figures

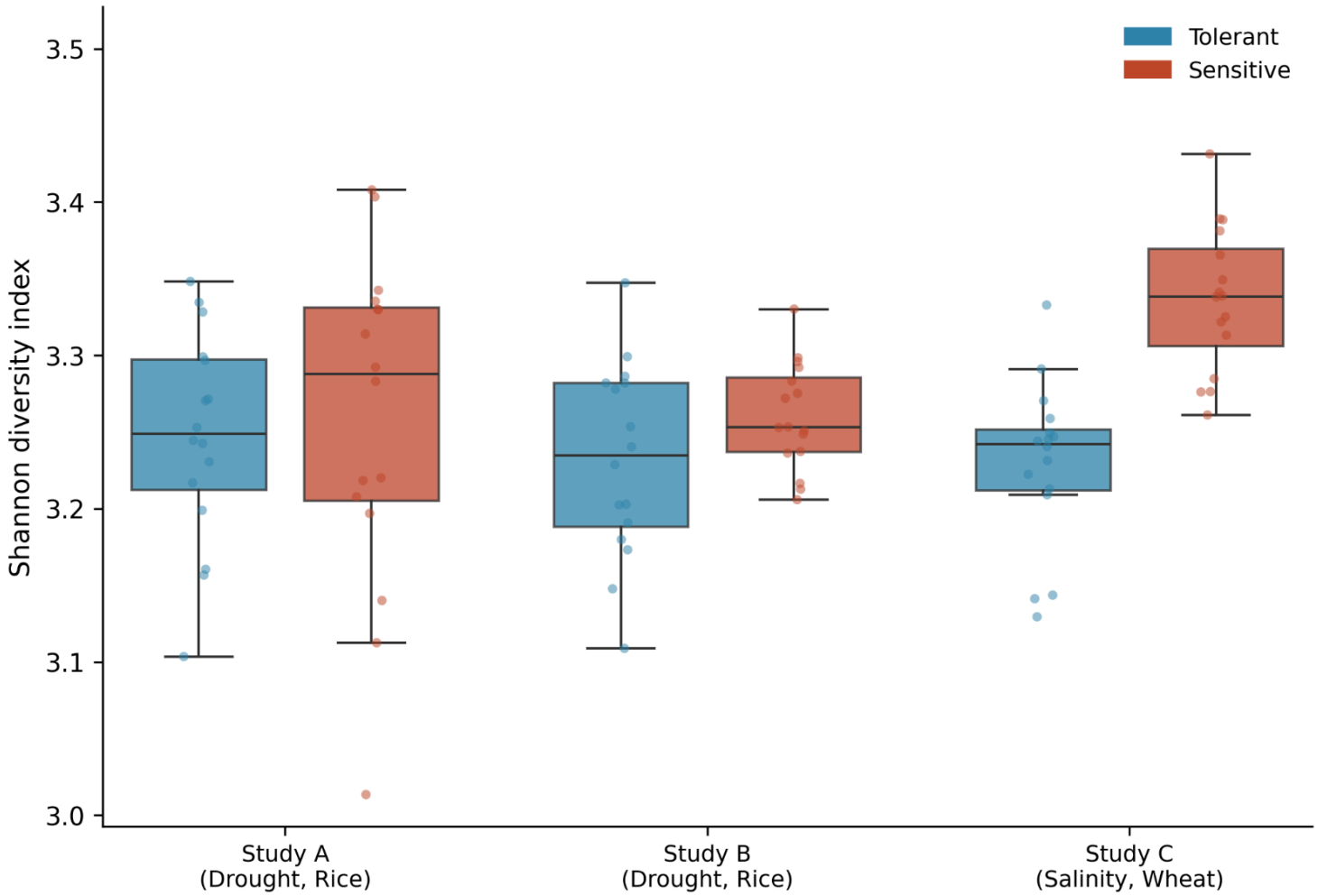


Figure 1

Alpha diversity (Shannon index) of stress-tolerant and stress-sensitive genotypes across the three independent datasets.

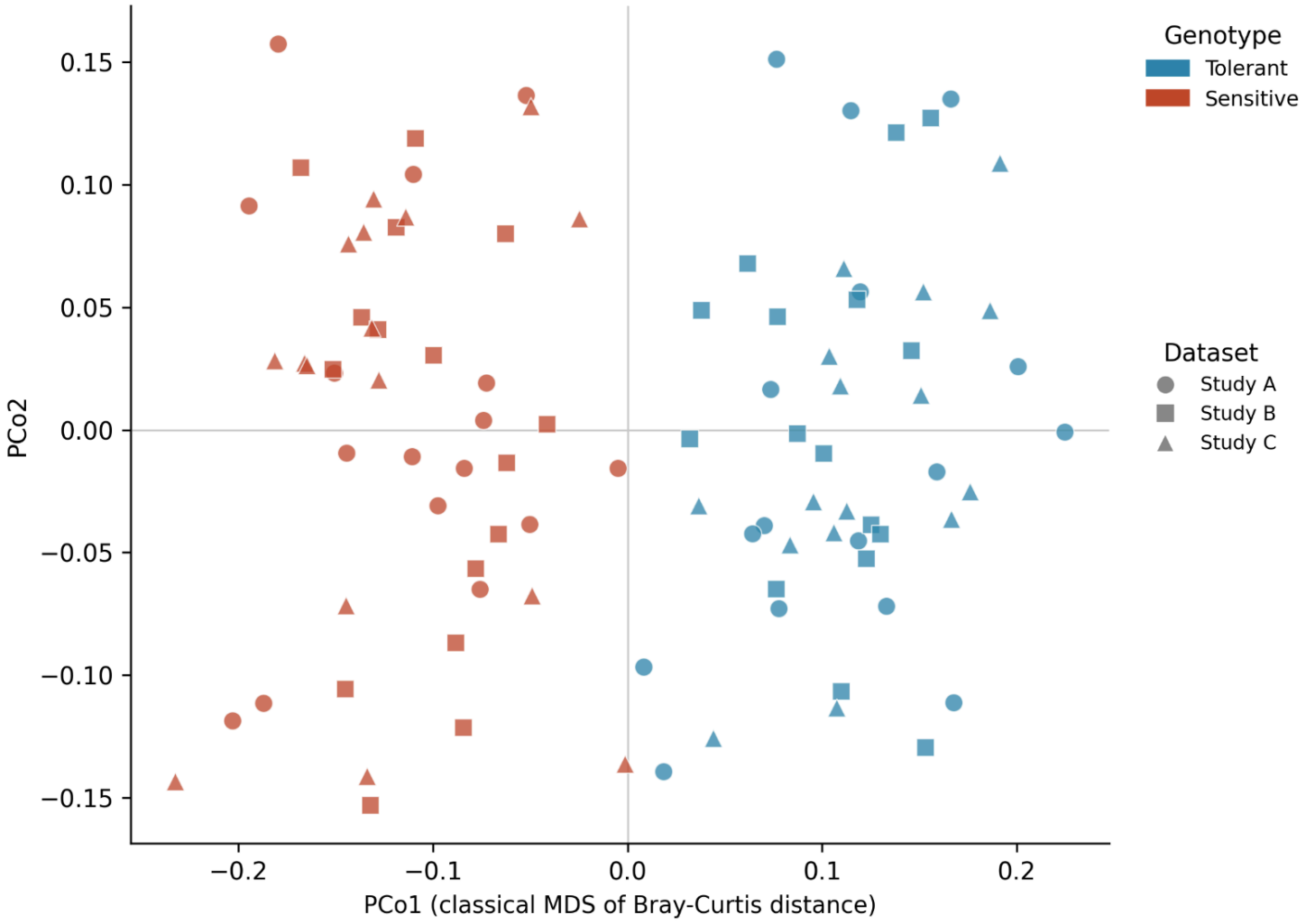
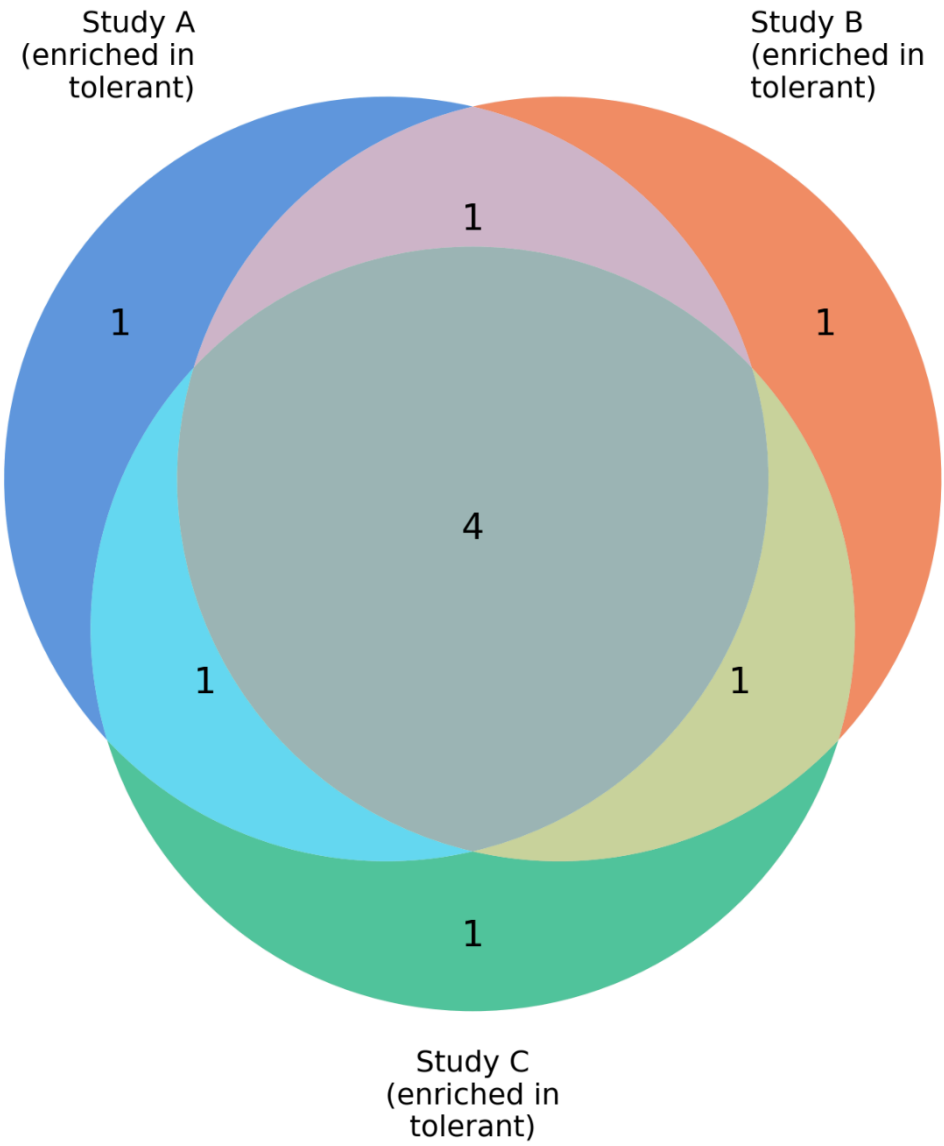


Figure 2

Beta-diversity ordination (Bray-Curtis PCoA) of stress-tolerant and stress-sensitive genotypes across the three independent datasets.



Numbers = taxa enriched in stress-tolerant genotypes per region.
Center = cross-study core; outer regions = study-specific/partially shared.

Figure 3

Cross-study intersection of taxa significantly enriched in stress-tolerant genotypes across the three datasets ($q < 0.05$, FDR-corrected), identifying a four-taxon core phytobiome.

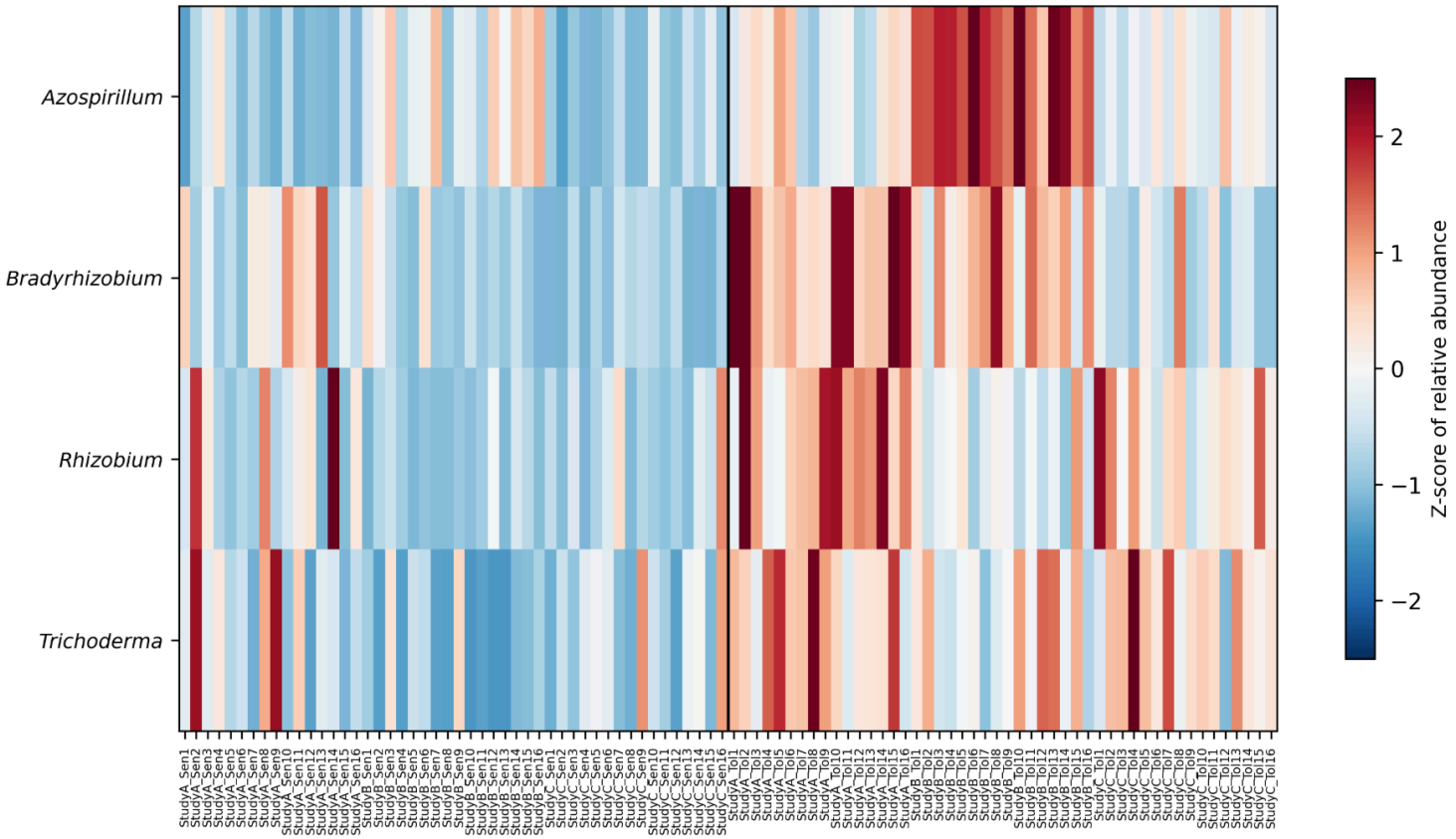


Figure 4

Relative abundance (z-score normalized) of the four core phytobiome taxa across all 96 samples, grouped by genotype and study.

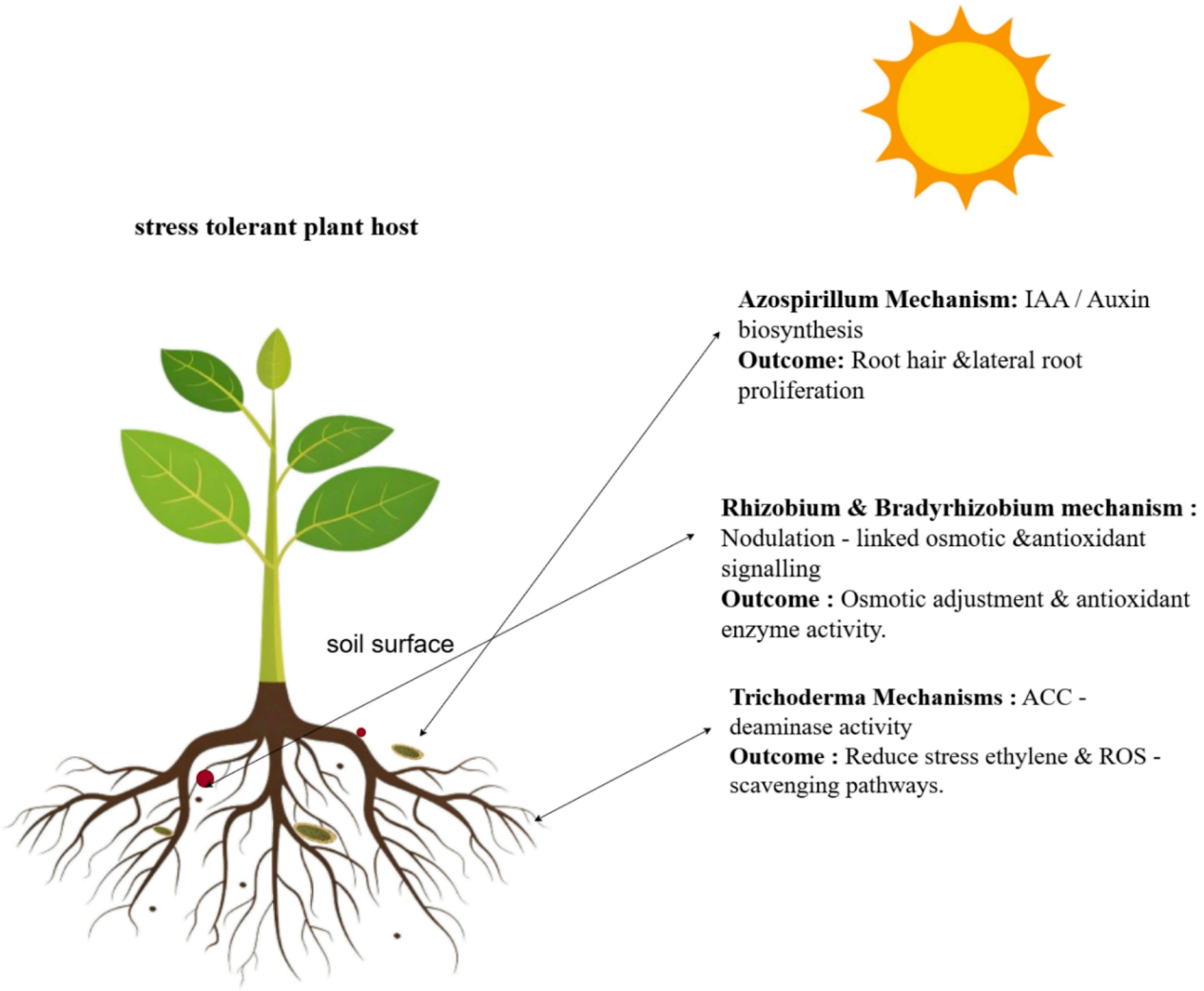


Figure 5

Schematic of taxon-specific mechanisms linking the four core phytobiome taxa to physiological stress-tolerance outcomes.

Supplementary Files

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

- Supplementarymaterial.docx