



# Integration of Phenotypes in Microbiome Networks for Designing Synthetic Communities: a Study of Mycobiomes in the Grafted Tomato System

 Ravin Poudel,<sup>a,b,c,d</sup> Ari Jumpponen,<sup>e</sup> Megan M. Kennelly,<sup>d</sup> Cary Rivard,<sup>f</sup> Lorena Gomez-Montano,<sup>a,b,c,d</sup>  Karen A. Garrett<sup>a,b,c,d</sup>

<sup>a</sup>Plant Pathology Department, University of Florida, Gainesville, Florida, USA

<sup>b</sup>Global Food Systems Institute, University of Florida, Gainesville, Florida, USA

<sup>c</sup>Emerging Pathogens Institute, University of Florida, Gainesville, Florida, USA

<sup>d</sup>Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA

<sup>e</sup>Division of Biology and Ecological Genomics Institute, Kansas State University, Manhattan, Kansas, USA

<sup>f</sup>Department of Horticulture and Natural Resources, Kansas State University, Olathe, Kansas, USA

**ABSTRACT** Understanding factors influencing microbial interactions, and designing methods to identify key taxa that are candidates for synthetic communities, or SynComs, are complex challenges for achieving microbiome-based agriculture. Here, we study how grafting and the choice of rootstock influences root-associated fungal communities in a grafted tomato system. We studied three tomato rootstocks (BHN589, RST-04-106, and Maxifort) grafted to a BHN589 scion and profiled the fungal communities in the endosphere and rhizosphere by sequencing the internal transcribed spacer (ITS2). The data provided evidence for a rootstock effect (explaining ~2% of the total captured variation,  $P < 0.01$ ) on the fungal community. Moreover, the most productive rootstock, Maxifort, supported greater fungal species richness than the other rootstocks or controls. We then constructed a phenotype-operational taxonomic unit (OTU) network analysis (PhONA) using an integrated machine learning and network analysis approach based on fungal OTUs and associated tomato yield as the phenotype. PhONA provides a graphical framework to select a testable and manageable number of OTUs to support microbiome-enhanced agriculture. We identified differentially abundant OTUs specific to each rootstock in both endosphere and rhizosphere compartments. Subsequent analyses using PhONA identified OTUs that were directly associated with tomato fruit yield and others that were indirectly linked to yield through their links to these OTUs. Fungal OTUs that are directly or indirectly linked with tomato yield may represent candidates for synthetic communities to be explored in agricultural systems.

**IMPORTANCE** The realized benefits of microbiome analyses for plant health and disease management are often limited by the lack of methods to select manageable and testable synthetic microbiomes. We evaluated the composition and diversity of root-associated fungal communities from grafted tomatoes. We then constructed a phenotype-OTU network analysis (PhONA) using these linear and network models. By incorporating yield data in the network, PhONA identified OTUs that were directly predictive of tomato yield and others that were indirectly linked to yield through their links to these OTUs. Follow-up functional studies of taxa associated with effective rootstocks, identified using approaches such as PhONA, could support the design of synthetic fungal communities for microbiome-based crop production and disease management. The PhONA framework is flexible for incorporation of other phenotypic data, and the underlying models can readily be generalized to accommodate other microbiome or 'omics data.

**KEYWORDS** fungi, phenotypes, microbiome networks, model integration, grafting, tomato, variable phenotypes

**Editor** Irina S. Druzhinina, Royal Botanic Gardens

**Copyright** © 2023 Poudel et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ravin Poudel, rp3448@ufl.edu, or Karen A. Garrett, karengarrett@ufl.edu.

The authors declare no conflict of interest.

**Received** 1 November 2022

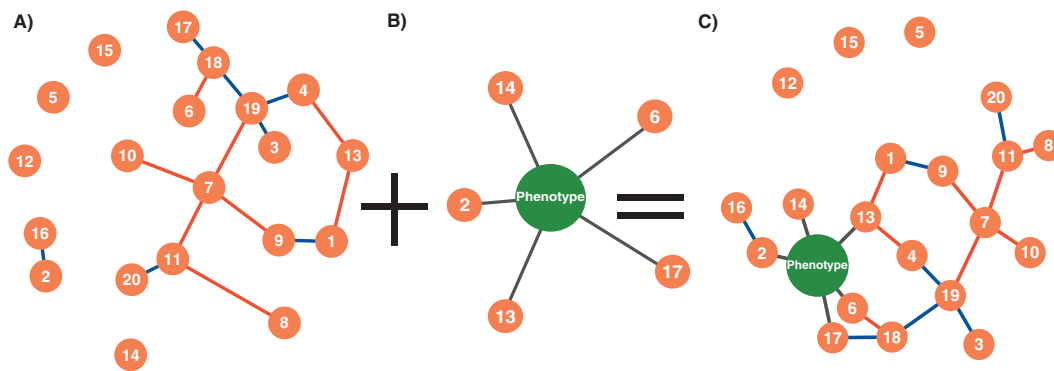
**Accepted** 24 April 2023

**Published** 24 May 2023

Interactions are key to defining system behaviors, structures, and outcomes. In microbial systems, interactions among organisms define their distribution, assemblies, and ecosystem functions. In addition to microbe-microbe interactions, microbes interact with their hosts and are essential to host health and performance (1–8). In agriculture, plant-microbe interactions improve plant productivity by providing access to nutrients (9–11), reducing infection by plant pathogens (5, 12), triggering plant growth-promoting factors (13, 14), and enhancing plant resistance (15, 16) and tolerance to abiotic stresses (17–20). Although the importance of microbes and host-microbe interactions to host health and ecological processes is well known, interaction-based approaches to manage crop production remain a scientific frontier. Past attempts to translate information about microbial interactions to design biocontrol agents or biofertilizers have often had limited efficacy and durability (21, 22). Most microbial inocula have been applied as single species, often selected based on pairwise relations of microbes with a pathogen or the host. Interactions among microbes as well as with the host are important, and the net outcome of these complex interactions defines host health and ecosystem functions (23). Thus, it is critical to understand the ecology of microbes selected for biological applications, and systems approaches centered on host-microbe interaction can help guide the selection of microbes for synthetic communities (24, 25).

Among the tools to better understand microbial interactions, network models of microbial communities, and studies of network structures and key groups, have proven popular for generating hypotheses about how to engineer microbial consortia. In such network models of microbiomes, a node represents an operational taxonomic unit (OTU), and a link exists between two OTUs if their sequence proportions are significantly associated across samples. When evaluated with other conventional measures of microbial community structures, such as diversity indices, network models can be used to identify hub taxa that may be key to maintaining microbial assemblages and diversity (26) or to evaluate changes in community complexity and interactions in response to experimental treatments (27). Microbiome networks are useful for describing general community structures and their key properties and are often the most practical option when additional information about species interactions is missing or the goal is to compare across studies with different types of data (28, 29). The utility of network analysis for identifying candidate assemblages for biocontrol can be enhanced by incorporating nodes that represent other additional types of features (30, 31). For instance, a novel association of host metadata with the microbiome was revealed in an integrated microbiome-metadata network (32), where a feature strongly associated with hub microbes can serve as a marker to measure host performance. In agriculture, plant yield or other phenotypic traits can be integrated in microbiome networks. Because such models include host phenotypes, they facilitate finding candidate sets of OTUs that may directly or indirectly affect host phenotypic traits. Visualization of networks is often valuable for this purpose, but the real value of phenotype-based network models is their ability to infer potential candidate taxa or consortia. The hypothesized beneficial sets of OTUs may represent targets for pure culturing efforts, or if cultures exist, the sets can be further evaluated in laboratory or field studies.

In our current study, building on previously described agricultural experiments in grafted tomato systems (33, 34), we characterized the root-associated fungal (RAF) communities and implemented an interaction-based approach to select potential candidate fungi that are predictive of tomato yield and/or that are in significant association with other fungal taxa. The new phenotype-OTU network analysis (PhONA) is a method for network-visualization and a framework to support the selection of candidate taxa and to integrate system traits (such as host yield) in microbe-microbe association networks. PhONA first identifies OTUs predictive of phenotype using lasso regression and then uses the predictive OTUs from lasso regression to build a reduced generalized linear model (GLM). PhONA then combines the GLM results indicating positive or negative associations of the predicted OTUs with the host phenotype as well as with other OTUs in a network model (Fig. 1). Due to the large number of OTUs compared to the sample size, lasso regression was used because it is suited for minimizing overfitting when



**FIG 1** (A to C) The phenotype-OTU network analysis (PhONA) combines (A) an OTU-OTU association network with (B) the nodes selected based on a predictive model for their association with a host phenotype variable such as yield to create (C) a PhONA.

applied with a relatively small sample size (35) and has been implemented in microbiome studies (36, 37).

While phenotype-based network models have the potential to identify key taxa, application of such models should be integrated with findings from other community analyses so that the inference about key taxa is biologically and ecologically meaningful. For example, plant microbiome studies indicate that a small but consistent proportion of variation in microbial communities is often explained by the host genotype (38–44), indicating the potential for genotype-based modulation of microbial communities in crop production on a broader scale. These results support the idea of host-specific microbial community selection (45). Many such microbes may be taxa that are evolutionarily essential for the survival and function of plants (46, 47). In addition, the extent of host genotype filtering of microbes differs across the rhizosphere, rhizoplane, and endosphere and varies from one host species to another (48–50). Results that indicate microbial filtering by different crop hosts, plant compartments, geographic locations, and environmental factors (51, 52) are promising for designing experiments to minimize the search space, or necessary sample numbers, to identify candidate taxa for synthetic communities. For instance, factors that explain great variation in microbial community composition but that are outside the control of management can be treated as blocks in experimental designs, so that host- or compartment-specific effects on the microbial community can be searched to identify the most desirable candidate taxa. Synthetic communities can potentially be matched to the environments in which they are most useful (53).

Phenotype-based selection of microbial consortia is promising as an effective approach to select representative microbial taxa and could support the design of microbiome-based products. Changes in abundance (54), successive selection over multiple generations (3), or analyses of binary host-microbe relationships (55) are some of the recent phenotype-based applications to select candidate taxa for biological applications. Despite the importance of biological test-based approaches, difficulty in culturing all the microbes makes computational approaches instrumental to define microbe-microbe and host-microbiome associations and to identify the biological and ecological key taxa. Tools to describe the community structures based on the cooccurrence matrix or covariance structures (56, 57) are more common, whereas tools to integrate host phenotype or environmental factors are at an earlier phase of development. Relatively small sample sizes combined with a great number of features may limit applications of the recent graph-based approaches. Such methods allow measurement of direct associations via conditional dependence structures and offer options to include environmental and phenotypic information in the model (58). CoNet (59) and FlashWeave (58) allow representation of the phenotype or an environmental variable as an extra node or a column in the adjacency matrix, and the same statistical method can be applied to define the associations among microbes and between microbes and phenotypes (taxa and metadata). PhONA is generic, as it allows the user to select data structure-specific models for microbe-microbe and microbe-phenotype associations.

In the current case study, we used lasso regression to identify the subset of OTUs and then fitted them using GLMs to predict OTU-phenotype associations, whereas the OTU-OTU associations were defined using sparse correlations for compositional data (SparCC). Additionally, we contrasted the RAF community's diversity and interactions among the rootstocks and the controls for endosphere and rhizosphere compartments. Based on our yield data, rootstock vigor, and previous studies of microbial interactions (27), we expected a greater number of fungal OTUs and of microbial associations for more productive rootstocks. Moreover, in our previous studies of bacterial communities in the tomato rhizobiome, we observed compartment-specific (endosphere versus rhizosphere) effects of grafting and rootstocks on bacterial community composition and diversity (34) and expected similar effects on RAF diversity and composition. All the code and vignettes for PhONA are available at <https://ravinpoudel.github.io/PhONA/index.html> and archived at Zenodo (doi: 10.5281/zenodo.6600986).

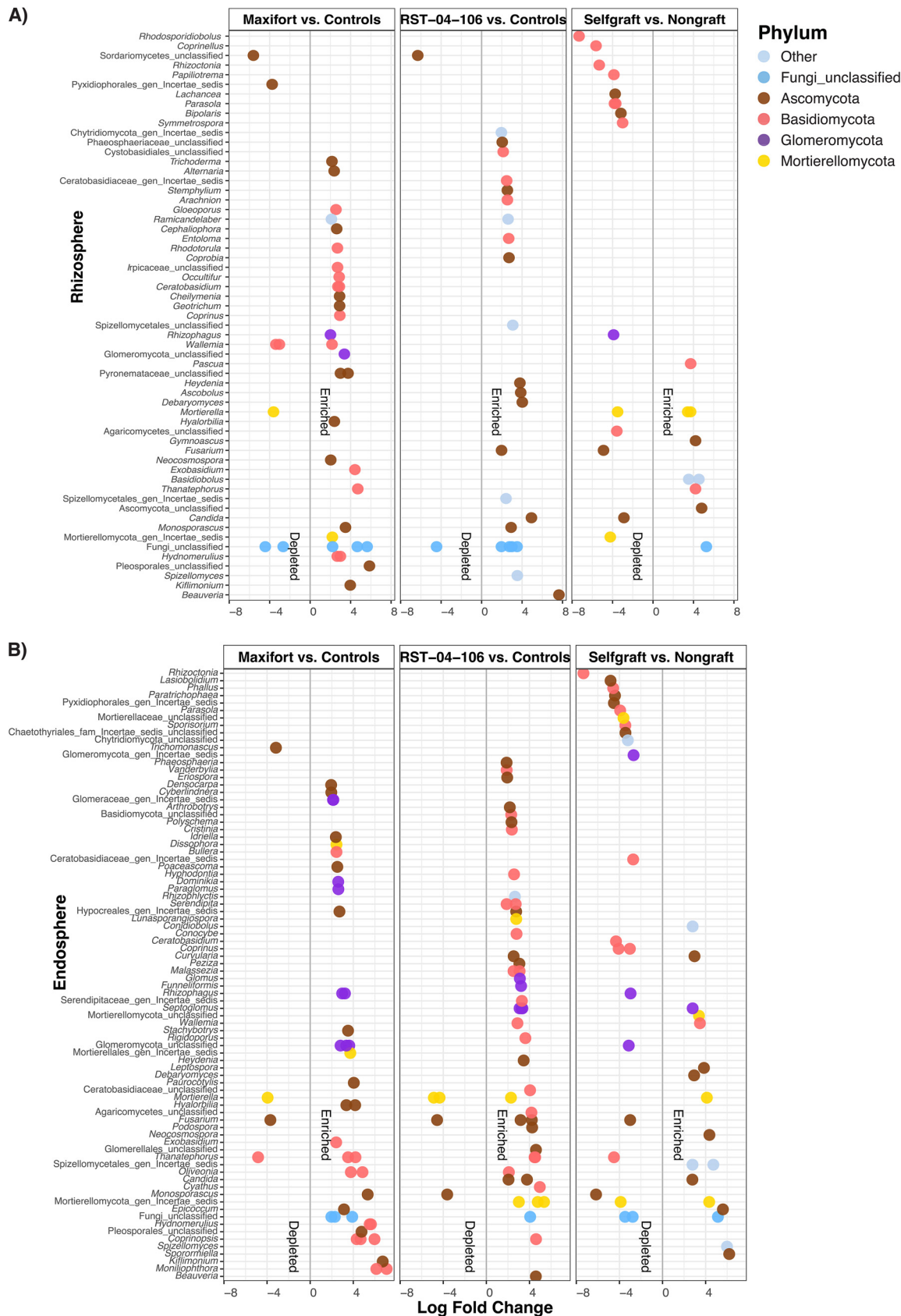
## RESULTS

**RAF in the grafted tomato system.** Once rare OTUs (<10 sequence counts, which accounted for more than 90% of the observed OTUs) were removed, the community consisted of 1,318 OTUs and 1,060,225 sequences. Of these sequences, 1.9% remained unclassified at the phylum level (see Fig. S1 in the supplemental material). The classified sequences represented Ascomycota (54.5%), Basidiomycota (25.3%), Glomeromycota (1.8%), Mortierellomycota (11.1%), Olpidiomyces (3.3%), and others (2.2%; phyla with relative abundance less than 1%) (Fig. S1). At the class level, Pezizomycetes, Agaricomycetes, and Dothideomycetes were the most abundant across all the rootstocks. At the order level, the communities were dominated by Pezizales, Pleosporales, Mortierellales, Cantharellales, and Hypocreales. Analyses at the family level revealed that Pyrenomataceae, Mortierellaceae, Ceratobasidiaceae, and Pleosporaceae were the most common taxa overall. At the genus level, *Kotlabaea*, *Thanatephorus*, and unclassified *Pyrenomataceae* were the most abundant (Tables S1 to S5).

**Effects of grafting and rootstock on  $\alpha$ -diversity.** There was strong evidence for a rootstock effect on OTU richness ( $F_{1,3} = 8.05$ ,  $P = 0.008$ ) and Shannon entropy ( $F_{1,3} = 3.39$ ,  $P = 0.02$ ) of tomato RAF communities. Mean species richness was higher in both the endosphere ( $P = 0.02$ ) and rhizosphere ( $P = 0.002$ ) of one of the hybrid rootstocks, Maxifort, compared to the nongrafted control (Fig. S2B). Shannon entropy followed trends similar to richness trends with a higher estimate for Maxifort; however, there was evidence for higher Shannon entropy in Maxifort only for the rhizosphere ( $P = 0.004$ ), but not for the endosphere ( $P = 0.5$ ) (Fig. S2A). Both species richness and Shannon entropy were higher ( $P < 0.001$ ) in the rhizosphere than in the endosphere across all the rootstock genotypes (Fig. S2).

**Effects of grafting and rootstock on RAF composition.** Based on previous studies of the plant genotype effect on the rhizobiome (34), we expected a significant rootstock effect on community composition. Rootstock explained 2% of the variation in the RAF community composition (permutational multivariate analysis of variance [PERMANOVA];  $P < 0.05$ ), whereas compartment, study site, and year explained a greater proportion of the variation than rootstock (Fig. S3 and Table S2). Endosphere-rhizosphere compartments accounted for 9.17% of the variation. Study site and interannual variation explained 8.44% and 5.33% of the total variation, respectively (Table S2). There was strong evidence ( $P < 0.001$ ) for the interaction of rootstocks with study sites but little evidence ( $P = 0.1$ ) for the interaction with year, suggesting more consistent shaping of fungal communities by rootstocks in a particular location across the study years.

**Comparison of differentially abundant OTUs (DAOTUs).** The analysis of differential abundance found effects of rootstock genotype at the individual OTU level. While analyses of alpha diversity indicated higher diversity in the rhizosphere than in the endosphere, we observed the opposite in the analysis of DAOTUs, with nearly twice as many DAOTUs in the endosphere ( $n = 134$ , i.e., 10.2% of the total OTUs) as in the rhizosphere ( $n = 87$  i.e., 6.6% of total OTUs) (Fig. 2 and Fig. S4). Comparison across rootstocks indicated a greater number of DAOTUs in Maxifort ( $n = 82$ ) than in RST-04-106



**FIG 2** (A and B) Enriched and depleted OTUs across tomato rootstock genotype combinations (nongraft BHN589, self-graft BHN589, and BHN589 grafted on two hybrid rootstocks [RST-04-106 and Maxifort]) evaluated for the rhizosphere (A) and the endosphere (B), (Continued on next page)



( $n = 75$ ) and the self-graft control ( $n = 64$ ). Compared to the hybrid rootstocks, the number of depleted taxa was greater in the self-graft control ( $n = 38$ ). Among the enriched OTUs in Maxifort, 25 belonged to Basidiomycota, 25 to Ascomycota, and 11 to Glomeromycota, whereas 3 basidiomycete, 4 ascomycete, and 2 Mortierellomycota OTUs were depleted in Maxifort. In RST-04-106, enriched taxa included 23 OTUs in Basidiomycota, 25 OTUs in Ascomycota, and five OTUs in Mortierellomycota, whereas the depleted OTUs included 2 in Mortierellomycota and 3 in Ascomycota. Comparing the self- and nongraft controls, 9 OTUs in Ascomycota, 3 OTUs in Basidiomycota, and 5 OTUs in Mortierellomycota were enriched in the self-graft treatment, whereas 17 OTUs in Basidiomycota, 10 OTUs in Ascomycota, 4 OTUs in Mortierellomycota, and 4 OTUs in Glomeromycota were depleted in the self-graft treatment.

**Network analysis/general network structures.** Fungal community complexity, defined in terms of mean node degree and community structures/motifs, varied among the rootstocks in both the endosphere and the rhizosphere, with a greater mean node degree in one of the hybrid rootstocks, Maxifort, compared to both controls and RST-04-106 (Fig. 3, Fig. S5 and 6, and Table 1). Complexity was higher in the rhizosphere than in the endosphere compartment (Fig. 3, Fig. S5 and 6, and Table 1). In addition to the total number of links, the link type (either positive or negative) differed among the rootstocks in both compartments (Table 1), with a higher ratio of negative to positive links in Maxifort in both the endosphere and the rhizosphere compartments. Rhizosphere fungal communities had a higher ratio of negative to positive links than those in the endosphere, for all rootstocks. Although we observed rootstock-specific or compartment-specific effects on the node degree and ratio of negative to positive links, the numbers of modules defined using a simulated annealing (SA) algorithm were similar in both the endosphere and rhizosphere compartments and across the rootstocks (Table 1). Our analyses of node types divided the observed nodes in the association network into four categories: peripherals, module hubs, network hubs, and connectors. More taxa in the rhizosphere were identified as key nodes than in the endosphere (Fig. 4 and 5).

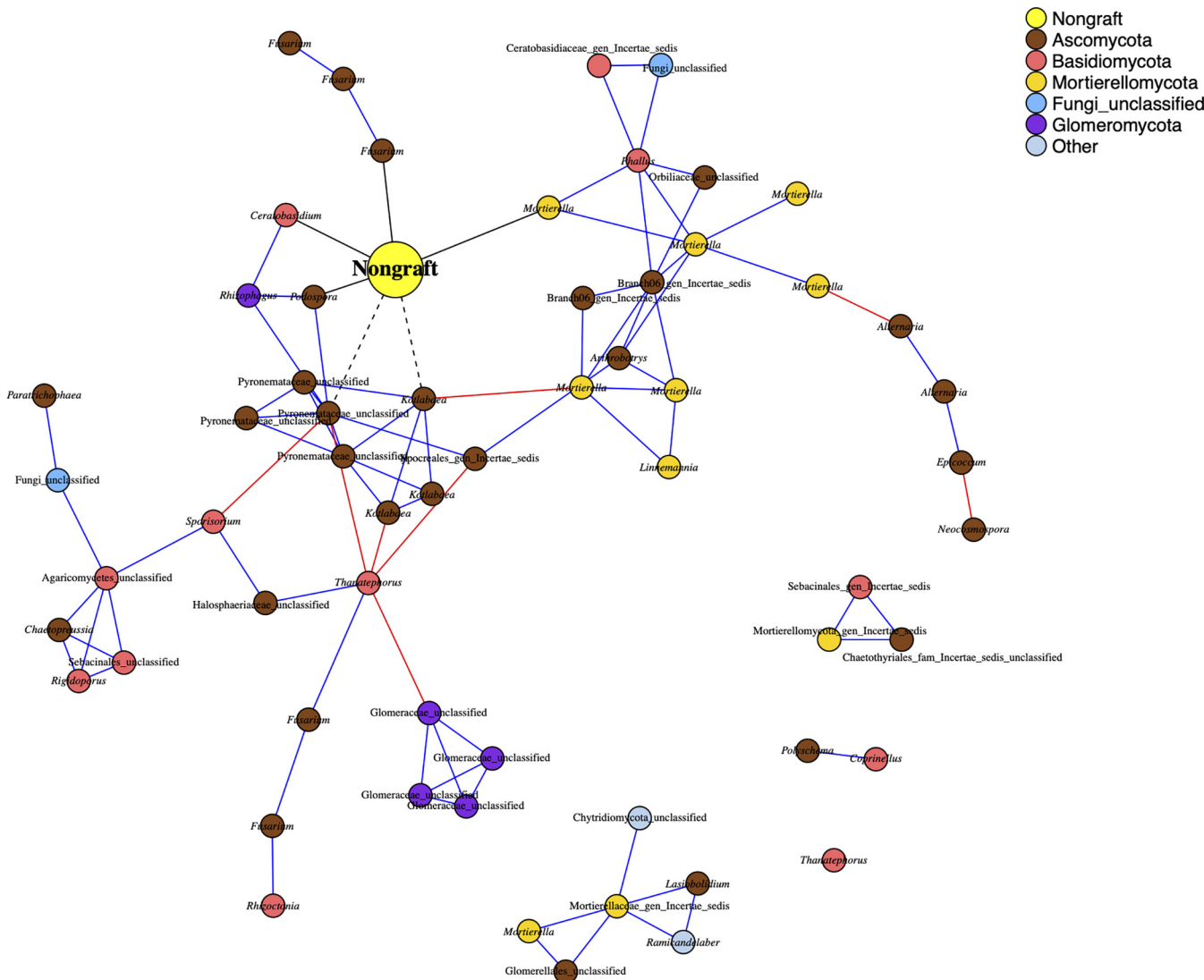
**Lasso regression, GLM, and PhONA.** Using lasso regression and GLM models, we identified the OTUs predictive of tomato yield in each compartment in each rootstock. The number of predictive OTUs identified using the varImp function was about the same across the rootstock treatments in both the compartments. However, not all the predicted OTUs were associated with other OTUs in the network models. The Maxifort rhizosphere had the highest number of OTUs (10) associated with other OTUs in the network models. Only a subset of the entire microbiome was predictive of the yield, among which only a few microbes were associated with other microbes in the network models.

## DISCUSSION

This study demonstrated the effect of rootstocks on RAF community composition and structure. General diversity-based analyses indicated a rootstock effect. The integrated host phenotype and OTU network in the PhONA identified potential candidate taxa for each rootstock and community structures in the endosphere and rhizosphere compartments. In contrast to network models that portray only microbe-microbe interactions, PhONA integrates the results of GLM models of microbe association with phenotypic traits to support inferences about candidate taxa and predictive microbiome analyses. Thus, candidate taxa can be selected not only because they have a direct association with the host response variable(s), but also because they are indirectly associated with the host response variable through their associations with community members that have direct associations with host traits. For instance, a node that has a

### FIG 2 Legend (Continued)

using OTU counts from self-grafts and nongrafts as controls. All the tests were adjusted to control the false-discovery rate (FDR,  $P < 0.01$ ) using the Benjamini-Hochberg method. Each point represents an OTU labeled at the genus level and colored based on phylum. Phyla with less than 1% relative abundance are labeled as "other." The position along the x axis represents the abundance fold change contrast with controls (except for the self-graft versus nongraft comparison, where the nongraft treatment was used as a control for the contrast).



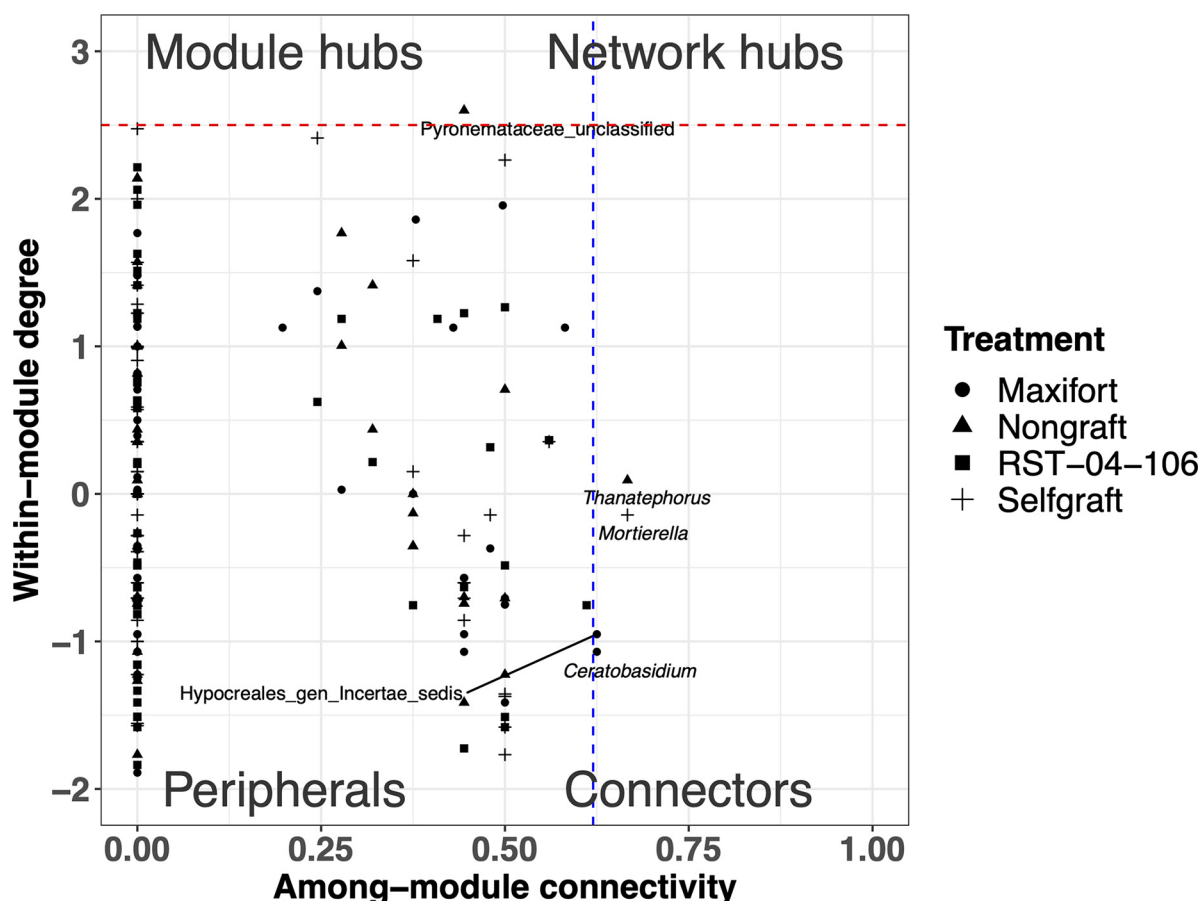
**FIG 3** Phenotype-OTU network analysis (PhONA) of endosphere fungal taxa, where the phenotype considered here is yield of the nongrafted tomato BHN589 (large yellow node). Other node colors indicate the OTU phylum, where phyla with less than 1% relative abundance are categorized as “Other.” Nodes connected to the yield node with black links are taxa that were predictive of yield, where dotted and solid lines indicate negative and positive associations, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the finest-resolution taxonomic categorization available.

positive association with a system phenotype node (in our case, yield) might have negative or positive associations with other OTUs. Such OTUs with indirect positive associations with the desired phenotype might also be included in biofertilizer consortia. PhONA provides data-driven selection of a consortium based on association with the

**TABLE 1** Network attributes and links observed in the fungal association networks for four tomato rootstock treatments<sup>a</sup>

Compartment	Rootstocks	No.	Nodes	Edges	Node degree	Density	Modules	Negative edges	Positive edges	Negative/positive edges
Endosphere	Nongraft	20	60	88	1.5	0.05	11	8	75	0.11
	Self-graft	20	67	101	1.5	0.05	10	9	86	0.10
	RST-04-106	20	75	123	1.6	0.04	11	8	110	0.07
	Maxifort	20	77	136	1.8	0.05	12	24	108	0.22
Rhizosphere	Nongraft	20	113	264	2.3	0.04	11	62	195	0.32
	Self-graft	20	102	252	2.5	0.05	7	76	175	0.43
	RST-04-106	20	136	341	2.5	0.04	12	62	271	0.23
	Maxifort	20	158	721	4.6	0.06	9	288	425	0.68

<sup>a</sup>The four rootstocks are nongraft and self-graft BHN589 and BHN589 grafted on two hybrid rootstocks (Maxifort and RST-04-106).

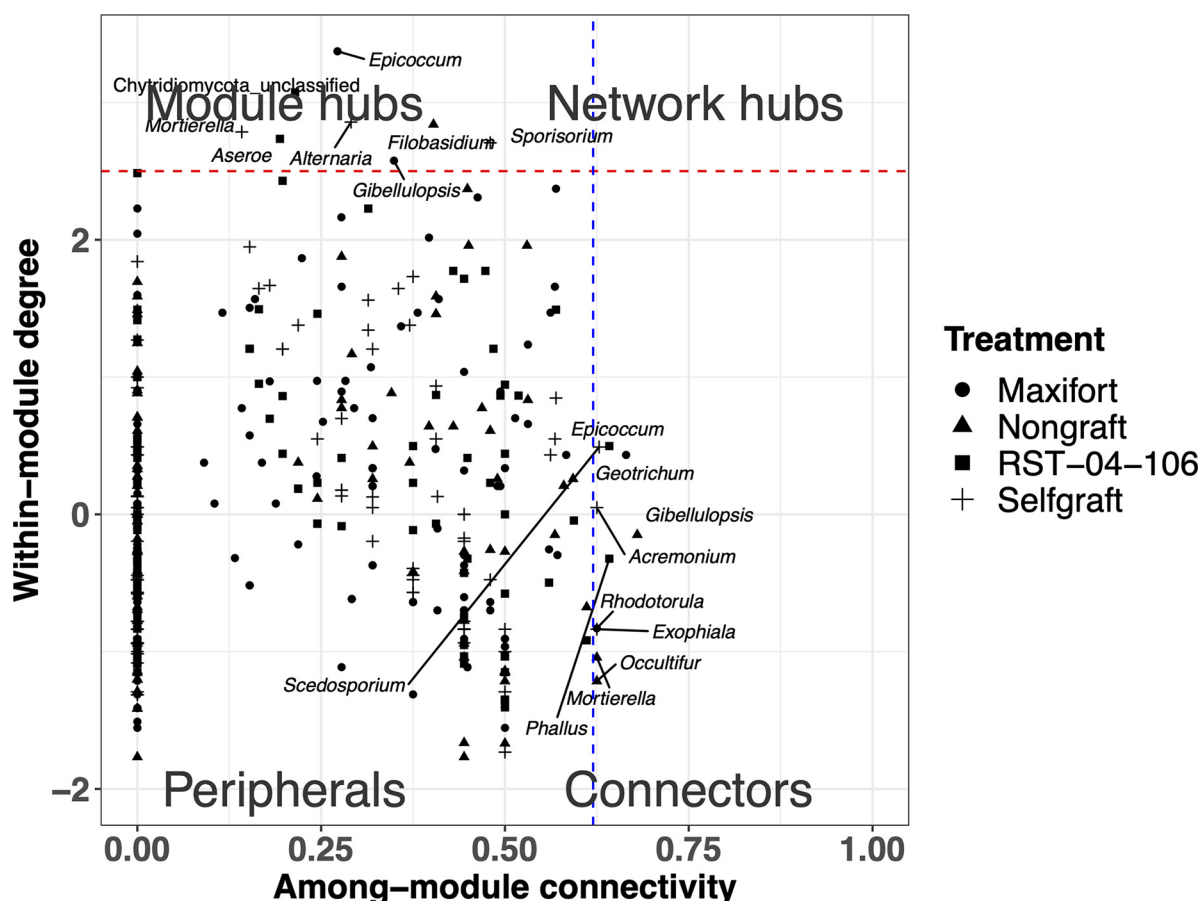


**FIG 4** Partitioning of endosphere fungal OTUs according to their network roles. Nodes were divided into four categories based on within-module degree and among-module connectivity. The blue dashed line represents a threshold value (0.62) for among-module connectivity, and the red dashed line represents a threshold value (2.5) for within-module degree. Nodes were categorized as peripherals, connectors, module hubs, and network hubs. Node shape indicates rootstock treatment (nongraft BHN589, self-graft BHN589, and BHN589 grafted on two hybrid rootstocks [RST-04-106 and Maxifort]).

phenotype of interest. Applying PhONA for disease phenotypes or pathogen resistance phenotypes could be useful for designing rational biocontrol consortia. We also observed some OTUs with direct negative associations with the yield node. Efforts to control taxa negatively associated with desired phenotypes, as well as the taxa that have positive associations with these taxa, might contribute to maximizing yield. Although we did not observe any disease symptoms in our experiments, OTUs negatively associated with the yield node might represent a case of asymptomatic negative microbial effects on yield. Efforts to explore negatively associated OTUs might provide opportunities to minimize asymptomatic yield loss in crops.

The main goal of PhONA is to provide a systems framework to generate hypotheses about the role of microbiome components in host function and performance and to support the potential for mechanistic/predictive models to better understand host-microbiome interactions. *In planta* experiments with fungal cultures are essential to test the hypotheses generated by these models, to help to differentiate between associations that are based on consistent biological interactions and not simply based on shared (or opposing) environmental niche preferences. It is important to be cautious in attributing biological interactions to the key structures in network attributes because the links in the PhONA may or may not depict biological interactions. That is, many links may represent only correlative relationships and not causal ones (30). For instance, the hub node in the network is often regarded as a key node, but the high number of links with the hub node in the association network could be due to shared niches, biological interactions, or a mixture thereof. If the associations are mostly due to shared niches, removing such a





**FIG 5** Partitioning of rhizosphere fungal OTUs according to their network roles. Nodes were divided into four categories based on within-module degree and among-module connectivity. The blue dashed line represents a threshold value (0.62) for among-module connectivity, and the red dashed line represents a threshold value (2.5) for within-module degree. Nodes were categorized as peripherals, connectors, module hubs, and network hubs. Node shape indicates rootstock treatment (nongraft BHN589, self-graft BHN589, and BHN589 grafted on two hybrid rootstocks [RST-04-106 and Maxifort]).

hub node will have a more limited effect, whereas removal of a hub node involved in many biological interactions could lead to significant effects on the microbial community.

The information gained through network analyses can be used in conjunction with prior knowledge of the observed organisms to select the taxa for consortia. For example, in the case of the nongraft treatment (Fig. 3), we observed many OTUs assigned to the potentially plant growth-promoting genus *Mortierella*. One of the *Mortierella* nodes was positively associated with yield, and another *Mortierella* module was positively associated with a node directly linked to yield. Taken together, these observations suggest that many *Mortierella* species may contribute to plant yield through direct or indirect mechanisms. In the same network, the plant pathogen *Thanatephorus* was a connector and was negatively associated with two modules, whereas many OTUs within these two modules were positively associated with each other. If the goal is to suppress this pathogen, including organisms from these two modules that are negatively associated with *Thanatephorus* may aid in the design of biocontrols. Network structure or prior biological knowledge about key nodes can guide consortium design but should be considered as a hypothesis-generating tool that requires further experimentation.

RAF community composition, diversity, and interactions differed between the endosphere and rhizosphere compartments. Although the endosphere and the rhizosphere are physically adjacent, they are distinct in community composition and diversity. Compartment specificity in community composition and diversity has been reported for other plant species, in both natural and agricultural settings (38, 60, 61). Usually, bulk soil is considered a source of plant-associated microbial communities, a subset of

which is selected for in the rhizosphere (38, 62), mainly as a function of root exudates and rhizodeposits (50, 63–66). Selection of the rhizosphere microbiome could be specific (e.g., antagonistic to plant pathogens) (67, 68) or more general with less influence of host genotype. In comparison, the endosphere of host plants often supports lower microbial diversity than the rhizosphere does (61, 62), as host tissues and defense systems act as biotic filters (2). As a result, the microbiome is more specialized in the endosphere than in the rhizosphere. RAF compartment specificity may also be an important consideration for microbe-based disease management strategies—especially for the management of pathogens or pests that are compartment specific, such as endoparasites and ectoparasites.

RAF community composition and diversity also differed among the rootstock genotypes, and the observed effects were due to the plant genotypes rather than grafting. Plant genotypes can structure root-associated fungal communities (38, 61, 69). The commercial rootstocks in our study have been bred to provide resistance against specific soilborne pathogens and pests. Small host genotypic differences could alter the physiological and immunological responses in the root systems, thereby selecting genotype-specific RAF communities (70). For example, some root exudates and metabolites could be specific to a plant genotype (71–73) and provide specific control of microbial communities (74–76). In some cases, the host genotype effect can be directly attributed to root anatomy (68, 76, 77). Efficient root types and architectures are desired agronomic traits to cope with biotic and abiotic stresses (78), and root systems vary among and within plant species (76). Moreover, the effect of plant genotypes on microbial communities in the root system may be linked to the flow of nutrients between the above-ground scion and below-ground rootstocks, where vigorous rootstock genotypes could drive greater resources to the microbial communities by supporting larger scion biomass. In such a positive nutrient feedback between the scion and rootstock, rootstock genotype appears to be a more critical driver than scion genotype (79). Rootstock genotypes supporting higher yield and biomass may support higher microbial diversity by excreting a greater volume of photosynthates as root exudates and metabolites. Although we did not evaluate root exudates, and used the same scion across the study, our study is consistent with higher yield and biomass (as for the Maxifort rootstock) being associated with higher fungal diversity. Additionally, we observed an effect of rootstock on the RAF community composition. Collectively, the results support our expectations of rootstock-specific control of the RAF community.

Our definition of complexity is based on interactions in networks, using a definition similar to that used in other microbiome network analyses (27, 41, 80). However, a greater number of interactions and complex network structures/motifs would tend to be observed whenever more nodes exist in these association networks, an inherent relationship not always considered in studies of complexity in microbiomes. The higher number of OTUs associated with Maxifort would tend to result in higher complexity than to rootstocks with fewer OTUs. Another potential measure of complexity is network density, the proportion of links observed in a network relative to the total number of possible links. For all the rootstocks we studied, network density was similar (0.04) in both compartments, indicating similar community complexity. Statistical methods comparable to rarefaction, designed to equalize the number of nodes across networks or methods to balance OTU richness for sampling efforts (81), will be a valuable future effort for understanding how network complexity responds to treatments and for making comparisons across studies. In addition, methods to optimize and automate the selection of association thresholds to define the pairwise relationships in a microbiome network are a gap and an opportunity for improving microbiome network analyses. For graphics in the figures in this analysis, we selected a level of association such that an interpretable number of OTUs were depicted for visual consideration. Studies directly applied to identify potential microbial assemblages for agricultural applications could benefit from exploring results of a range of thresholds.

Our study indicates a rootstock genotype-specific effect on RAF diversity, composition, and interactions and also demonstrates integration of system phenotypes such as

plant yield in a network-based model to support selection of candidate taxa for biological use. However, in sequence-based studies such as ours, the biological and functional significance of the candidate OTUs remains unknown. Follow-up experiments with fungal cultures will be necessary to determine the biological roles of the candidate OTUs and to differentiate causal associations from correlations based on niche preference. Similarly, further development of PhONA to incorporate temporal microbiome data and Bayesian learning and inference methodologies (82, 83) has the potential to support causal inference, including an understanding of directionality in microbiome networks. PhONA utilizes a lasso regression and GLM to link OTUs with a system phenotype, although many other models, such as random forest and other machine learning approaches (84), could also be used. Given the nature of microbiome data, having a high number of features ( $p$ ) and relatively small number of samples ( $n$ ), other models to address the  $n \times p$  problem can improve PhONA predictions. Rather than pure prediction, our methods aim to find the key predictors and use them in the GLM model for evaluating associations with a phenotypic response such as plant yield. PhONA focused on finding the attributable predictors/OTUs that are key to biological interventions, which are missed in approaches that are focused purely on prediction (85). Small sample size was a limitation in our current study, reflecting the challenge of processing a large number of plant replicates, and we did not validate the results from our model by splitting data into training and test sets. A rigorous model validation step would improve the accuracy of PhONA. As lab-based technologies and computational resources become less expensive, studies with large sample sizes are becoming more practical and, when combined with an analytical framework such as PhONA, microbial community analyses can go beyond simple analyses of diversity to help make microbiome-based agriculture a reality.

## MATERIALS AND METHODS

**Experimental plots, rootstocks, and study sites.** We studied grafted tomato plants in high tunnels in an experimental design similar to that described in detail by Poudel et al. (34). Tomato plants were grafted following a tube-grafting protocol described in Meyer et al. (33). Our study included three rootstocks (BHN589, RST-04-106, and Maxifort) in four graft treatments: (i) nongrafted BHN589 plants, (ii) self-grafted BHN589 plants (plants grafted to their own rootstock), (iii) BHN589 grafted to RST-04-106, and (iv) BHN589 grafted to Maxifort. We chose BHN589 as the scion primarily based on its popularity due to high yield and high-quality fruit with a long shelf life. For rootstocks, we selected Maxifort because it is a productive and popular rootstock and RST-04-106 as a new rootstock variety based on tomato breeders' recommendations.

Our study included two sites: Olathe Horticulture Research and Extension Center (OHREC) and Common Harvest Farm, a farm managed by a collaborating farmer. For more information about the sites, see Table S1. At each study site, the four graft treatments were assigned to four plots per block in a randomized complete block design. Each plot consisted of 5 to 8 plants, and one middle plant per plot was sampled during the peak growth stage. There were six blocks at OHREC and four blocks at Common Harvest Farm, such that for each year, each graft treatment was replicated 10 times. The experiment was repeated for 2 years (2014 and 2015) with a similar design, with the blocks and rootstocks randomly and independently assigned each year.

**Sample preparation, DNA extraction, and amplicon generation.** To compare the fungal communities, we selected a center plant from each plot and carefully dug the whole plant out such that the majority of the roots remained intact. Endosphere and rhizosphere samples were prepared as previously described (34), and the total genomic DNA was extracted using a DNA extraction kit (UltraClean soil DNA isolation kit; MoBio, Carlsbad, CA, USA) as per the manufacturer's protocol, with slight modification during the homogenization step (34). To recover high genetic diversity, we opted for the two-step PCR approach. The primary PCR amplicons were generated in 50- $\mu$ L reactions under the following conditions: 1  $\mu$ M forward and reverse primers, 10 ng template DNA, 200  $\mu$ M each dioxynucleotide, 1.5 mM  $MgCl_2$ , 10  $\mu$ L 5 $\times$  Phusion Green HF buffer (Finnzymes, Vantaa, Finland), 24.5  $\mu$ L molecular biology-grade water, and 1 unit (0.5  $\mu$ L) Phusion Hot Start II DNA polymerase (Finnzymes). PCR cycle parameters consisted of a 98°C initial denaturing step for 30 s, followed by 30 cycles at 98°C for 10 s, 57°C annealing temperature for 30 s, and a 72°C extension step for 30 s, followed by a final extension step at 72°C for 10 min. All samples were PCR-amplified in triplicate to minimize stochasticity, pooled, and cleaned using Diffinity RapidTip (Diffinity Genomics, West Chester, PA, USA). In this PCR, we amplified the entire internal transcribed spacer (ITS) region of fungal rRNA genes using primers ITS1F-CTTGGTCATTTAGAGGAAGTAA and ITS4-TCCTCCGCTTATTGATATGC (see, e.g., 86). The average amplicon length of the ITS region in fungi is about 600 bp and could not reliably be fully covered with the Illumina MiSeq platform (v.3 chemistry) in a single read. Thus, in the following nested PCR, only ITS2 of the ITS region was amplified using ITS2-ITS4 primers (87) incorporating unique molecular identifier tags (MIDs) at the 5' end of the reverse primer (ITS4). For the nested PCR, we used similar reagents and PCR conditions as in the primary PCR, with some modifications: the number of PCR cycles was reduced to 10, total reaction volume was reduced to 25  $\mu$ L, and 5  $\mu$ L of cleaned PCR product from the first PCR

amplification was used as the DNA template. The nested PCR was also run in triplicate, pooled by experimental unit, and cleaned with an Agencourt AmPure cleanup kit using a SPRIplate 96-ring magnet (Beckman Coulter, Beverly, MA, USA) as per the manufacturer's protocol. Then, 200 ng of cleaned, barcoded amplicons was combined per experimental unit, and the final pool was cleaned again using an Agencourt AmPure cleanup kit as described above. MiSeq adaptors were ligated to the library and paired-end sequenced on a MiSeq personal sequencing system (Illumina, San Diego, CA, USA) using a MiSeq reagent kit V3 with 600 cycles. The endosphere and the rhizosphere amplicon libraries were sequenced separately in two runs. Adaptor ligation and sequencing were performed at the Integrated Genomics Facility at Kansas State University. All sequence data generated in this study were deposited in the NCBI Sequence Read Archive depository (BioProject no. [PRJNA496268](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA496268)).

**Bioinformatics and OTU designation.** The sequence library of fastq files was curated using the mothur pipeline (version 1.33.3; 88) following steps modified from the MiSeq standard operating protocol (SOP; [www.mothur.org/wiki/MiSeq\\_SOP](http://www.mothur.org/wiki/MiSeq_SOP)). Briefly, the forward and reverse reads were assembled into contigs using the default alignment algorithm. Any sequences shorter than 250 bp or containing an ambiguous base call or more than 8 homopolymers or missing MID were removed from the library. Barcoded sequences were assigned to experimental units, and the data for endosphere and rhizosphere libraries were merged and processed together for the remaining steps in the mothur pipeline. The pairwise distance matrix based on the filtered sequences was created, and sequence data were clustered into OTUs at 97% sequence similarity using the nearest neighbor joining algorithm. The clustered OTUs were assigned to a putative taxonomic identity using a Bayesian classifier (89) referencing the UNITE plus International Nucleotide Sequence Database (INSD) nonredundant ITS database (version 29.11.2022; 90). To minimize the bias resulting from unequal sequence counts per sample, samples were rarefied to the lowest sequence count among the samples (6,698). The final curated OTU database included 1,071,678 total sequences representing 13,367 fungal OTUs, including singletons (4,522).

**Statistical analyses.** We evaluated the network of associations among fungal OTUs with network models to better understand the community composition and the interactions therein. The observed OTU database was divided into eight subsets, each combination of the four rootstocks and two compartments (endosphere and rhizosphere), such that we constructed eight networks in total. In our network models, a node represents an OTU and a link exists between a pair of OTUs if there is evidence ( $P < 0.05$ ) that their frequencies are correlated (positively or negatively) across samples. Reducing false associations due to compositional bias in network modeling of microbiome data is important for clearer interpretation (57). Thus, we used a sparse correlations for compositional data (SparCC) method to evaluate the pairwise associations (57), designed to minimize the compositional bias effect due to normalization. In our analyses, associations were defined in 20 iterations, and the significance of a pairwise association was determined from 500 bootstrapped data sets. Once the matrix defining all the pairwise associations was derived, we selected only those OTUs for which the absolute value of at least one association was greater than 0.5 (and  $P < 0.05$ ) in the network analyses for each of the rootstock genotypes.

To identify the OTUs associated with tomato yield in each rootstock, a regression-based model was fitted to the observed data. Marketable tomato yield data (Tables S6 and S7) were the response variable, and fungal OTUs were potential predictors. We used the caret package (91) to evaluate the lasso regression and selected OTUs using varImp functions. Lasso regression used an L1 regularization approach to shrink the less important variables' coefficients to zero and to reduce the number of variables in the model. In lasso regression, lambda determines the penalty of regularization, and its value can range from zero to infinity; when it is zero, the results are similar to the least square lines. A grid-based approach was used to tune the lambda parameter using repeated (500 iterations) 5-fold cross-validation, and the value of lambda with lowest variance was selected. Only the OTUs with nonzero coefficients were selected, based on the lasso-regression model, to build the reduced GLM model, and the association type of each OTU with phenotype was estimated. Given the small sample size, we did not evaluate the model performance by splitting the data into training and test cases, although this would be a valuable step in future studies with larger sample sizes. PHONA then integrates the results from the GLM model for yield with the OTU-OTU association network. We plotted the resulting network using the igraph package (92) in R. To evaluate the role of nodes in the network, we placed each node in one of four categories—peripherals, module hubs, network hubs, and connectors—based on the within-module degree and among-module connectivity (30, 93). The role analyses consider the presence or absence of links in the network and do not account for the link types (i.e., positive or negative associations).

To evaluate the effects of rootstocks on fungal diversity, Shannon entropy and species richness were evaluated using the vegan package (94) wrapped by the phyloseq package (95) in R (96). Differences in diversity across the rootstocks were compared using a mixed model analysis of variance (ANOVA) in the lme4 package in R (97). Rootstock treatment and compartment were treated as fixed factors, while study site and sampling year were treated as random factors, with study site used as a blocking factor. Changes in fungal community composition across the samples were estimated based on a Bray-Curtis dissimilarity distance matrix and visualized in nonmetric multidimensional scaling (NMDS) plots. The contribution of factors to the observed variation in fungal composition was estimated in a permutational multivariate analysis of variance (PERMANOVA, using 1,000 permutations) using the adonis function in the vegan package (94). To identify OTUs that were sensitive to the rootstock treatments, the observed frequency (proportion) of each OTU was evaluated by fitting a generalized linear model (GLM) with negative binomial distribution, to identify depleted or enriched OTUs (differentially abundant OTUs [DAOTUs]). Likelihood ratio tests and contrast analyses (between the hybrid rootstock and controls) were performed for the fitted GLM to identify the DAOTUs. We used OTU frequencies from self-grafts and nongrafts as controls, in comparisons with other rootstocks, using contrasts. All tests were adjusted to control the false-discovery rate (FDR,  $P < 0.01$ ) using

the Benjamini-Hochberg method (98). A differential abundance test was performed within the controls (self-graft versus nongraft) to identify the OTUs responsive to the grafting procedure itself.

**Data availability.** Sequencing data for this project are available at NCBI under BioProject no. PRJNA496268. All the code and vignettes for PhONA are available at <https://ravinpoudel.github.io/PhONA/index.html> and archived at Zenodo (doi: 10.5281/zenodo.6600986).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 3.5 MB.

**SUPPLEMENTAL FILE 2**, XLSX file, 0.2 MB.

## ACKNOWLEDGMENTS

We appreciate support from the Ceres Trust, USDA NCR SARE Research and Education Grant LNC13-355, Foundation for Food and Agriculture Research Grant FF-NIA19-0000000050, USDA NIFA grant 2022-51181-38242, the Kansas Agricultural Experiment Station, the National Institute for Mathematical and Biological Synthesis (NIMBioS), and the University of Florida. We thank members of the Garrett Lab for their comments on an early version of the manuscript.

R.P., K.A.G., A.J., M.M.K., and C.R. designed the study. R.P. and L.G.-M. collected and processed the samples. R.P. analyzed and interpreted the data. R.P., K.A.G., and A.J. wrote the manuscript. R.P., K.A.G., A.J., M.M.K., C.R., and L.G.-M. revised the manuscript. All authors read and approved the final manuscript.

## REFERENCES

- Berendsen RL, Pieterse CM, Bakker PA. 2012. The rhizosphere microbiome and plant health. *Trends Plant Sci* 17:478–486. <https://doi.org/10.1016/j.tplants.2012.04.001>.
- Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, del Rio TG, Jones CD, Tringe SG, Dangl JL. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 349:860–864. <https://doi.org/10.1126/science.aaa8764>.
- Panke-Buisse K, Poole AC, Goodrich JK, Ley RE, Kao-Kniffin J. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J* 9:980–989. <https://doi.org/10.1038/ismej.2014.196>.
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JL. 2009. A core gut microbiome in obese and lean twins. *Nature* 457:480–484. <https://doi.org/10.1038/nature07540>.
- Xue C, Penton CR, Shen Z, Zhang R, Huang Q, Li R, Ruan Y, Shen Q. 2015. Manipulating the banana rhizosphere microbiome for biological control of Panama disease. *Sci Rep* 5:11124. <https://doi.org/10.1038/srep11124>.
- Gould AL, Zhang VV, Lamberti L, Jones EW, Obadia B, Korasidis N, Gavryushkin A, Carlson JM, Beerewinkel N, Ludington WB. 2018. Microbiome interactions shape host fitness. *Proc Natl Acad Sci U S A* 115: E11951–E11960. <https://doi.org/10.1073/pnas.1809349115>.
- Zhou X, Wang JT, Wang WH, Tsui CK, Cai L. 2021. Changes in bacterial and fungal microbiomes associated with tomatoes of healthy and infected by *Fusarium oxysporum* f. sp. *lycopersici*. *Microb Ecol* 81:1004–1017. <https://doi.org/10.1007/s00248-020-01535-4>.
- Kaur S, Egidi E, Qiu Z, Macdonald CA, Verma JP, Trivedi P, Wang J, Liu H, Singh BK. 2022. Synthetic community improves crop performance and alters rhizosphere microbial communities. *J Sustain Agric Environ* 1:118–131. <https://doi.org/10.1002/sae2.12017>.
- Jin CW, Ye YQ, Zheng SJ. 2014. An underground tale: contribution of microbial activity to plant iron acquisition via ecological processes. *Ann Bot* 113:7–18. <https://doi.org/10.1093/aob/mct249>.
- Koide RT. 1991. Nutrient supply, nutrient demand and plant response to mycorrhizal infection. *New Phytol* 117:365–386. <https://doi.org/10.1111/j.1469-8137.1991.tb00001.x>.
- Mishra PK, Bisht SC, Mishra S, Selvakumar G, Bisht JK, Gupta HS. 2012. Co-inoculation of *Rhizobium leguminosarum*-pr1 with a cold tolerant *Pseudomonas* sp. improves iron acquisition, nutrient uptake and growth of field pea (*Pisum sativum* L.). *J Plant Nutr* 35:243–256. <https://doi.org/10.1080/01904167.2012.636127>.
- Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, Piceno YM, DeSantis TZ, Andersen GL, Bakker PAHM, Raaijmakers JM. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 332:1097–1100. <https://doi.org/10.1126/science.1203980>.
- Gaiero JR, McCall CA, Thompson KA, Day NJ, Best AS, Dunfield KE. 2013. Inside the root microbiome: Bacterial root endophytes and plant growth promotion. *Am J Bot* 100:1738–1750. <https://doi.org/10.3732/ajb.1200572>.
- Glick BR. 1995. The enhancement of plant growth by free-living bacteria. *Can J Microbiol* 41:109–117. <https://doi.org/10.1139/m95-015>.
- Conrath U. 2009. Priming of induced plant defense responses. *Adv Botanical Res* 51:361–395. [https://doi.org/10.1016/S0065-2296\(09\)51009-9](https://doi.org/10.1016/S0065-2296(09)51009-9).
- Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Van Wees SCM, Bakker PAHM. 2014. Induced systemic resistance by beneficial microbes. *Annu Rev Phytopathol* 52:347–375. <https://doi.org/10.1146/annurev-phyto-082712-102340>.
- Marasco R, Rolli E, Ettoumi B, Viganì G, Mapelli F, Borin S, Abou-Hadid AF, El-Behairy UA, Sorlini C, Cherif A, Zocchi G, Daffonchio D. 2012. A drought resistance-promoting microbiome is selected by root system under desert farming. *PLoS One* 7:e48479. <https://doi.org/10.1371/journal.pone.0048479>.
- Yang J, Kloepper JW, Ryu CM. 2009. Rhizosphere bacteria help plants tolerate abiotic stress. *Trends Plant Sci* 14:1–4. <https://doi.org/10.1016/j.tplants.2008.10.004>.
- Yuan ZL, Druzhinina IS, Labbe J, Redman R, Qin Y, Rodriguez R, Zhang CL, Tuskan GA, Lin FC. 2016. Specialized microbiome of a halophyte and its role in helping non-host plants to withstand salinity. *Sci Rep* 6:32467. <https://doi.org/10.1038/srep32467>.
- Poudel M, Mendes R, Costa LAS, Bueno CG, Meng Y, Folimonova SY, Garrett KA, Martins SJ. 2021. The role of plant-associated bacteria, fungi, and viruses in drought stress mitigation. *Front Microbiol* 12:743512. <https://doi.org/10.3389/fmicb.2021.743512>.
- Mazzola M, Freilich S. 2017. Prospects for biological soilborne disease control: application of indigenous versus synthetic microbiomes. *Phytopathology* 107:256–263. <https://doi.org/10.1094/PHYTO-09-16-0330-RVV>.
- Xu XM, Jeffries P, Pautasso M, Jeger MJ. 2011. Combined use of biocontrol agents to manage plant diseases in theory and practice. *Phytopathology* 101:1024–1031. <https://doi.org/10.1094/PHYTO-08-10-0216>.
- Shayanthan A, Ordoñez PAC, Oresnik IJ. 2022. The role of synthetic microbial communities (SynCom) in sustainable agriculture. *Front Agron* 4:896307. <https://doi.org/10.3389/fagro.2022.896307>.
- de Souza RSC, Armanhi JSL, Arruda P. 2020. From microbiome to traits: Designing synthetic microbial communities for improved crop resiliency. *Front Plant Sci* 11:1179. <https://doi.org/10.3389/fpls.2020.01179>.
- Martins SJ, Pasche JM, Silva HA, Seltén GG, Savastano N, Abreu L, Bais H, Garrett KA, Kraisitdomsook N, Pieterse CM, Cernava T. 2023. The use of



- synthetic microbial communities (SynComs) to improve plant health. Phytopathology. <https://doi.org/10.1094/PHYTO-01-23-0016-IA>.
26. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, Kemen EM. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol* 14:e1002352. <https://doi.org/10.1371/journal.pbio.1002352>.
  27. Shi S, Nuccio EE, Shi ZJ, He Z, Zhou J, Firestone MK. 2016. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol Lett* 19:926–936. <https://doi.org/10.1111/ele.12630>.
  28. Ascunce MS, Shin K, Hugueta-Tapia JC, Poudel R, Garrett KA, van Bruggen AHC, Goss EM. 2019. Penicillin trunk injection affects bacterial community structure in citrus trees. *Microb Ecol* 78:457–469. <https://doi.org/10.1007/s00248-018-1302-1>.
  29. Garrett KA, Alcalá-Briseño RI, Andersen KF, Buddenhagen CE, Choudhury RA, Fulton JC, Hernandez Nopsa JF, Poudel R, Xing Y. 2018. Network analysis: a systems framework to address grand challenges in plant pathology. *Annu Rev Phytopathol* 56:559–580. <https://doi.org/10.1146/annurev-phyto-080516-035326>.
  30. Poudel R, Jumpponen A, Schlatter DC, Paulitz TC, Gardener BBM, Kinkel LL, Garrett KA. 2016. Microbiome networks: A systems framework for identifying candidate microbial assemblages for disease management. *Phytopathology* 106:1083–1096. <https://doi.org/10.1094/PHYTO-02-16-0058-FI>.
  31. Röttgers L, Faust K. 2018. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol Rev* 42:761–780. <https://doi.org/10.1093/femsre/fuy030>.
  32. Wang Z, Bafadhel M, Halder K, Spivak A, Mayhew D, Miller BE, Tal-Singer R, Johnston SL, Ramsheh MY, Barer MR, Brightling CE, Brown JR. 2016. Lung microbiome dynamics in COPD exacerbations. *Eur Respir J* 47:1082–1092. <https://doi.org/10.1183/13993003.01406-2015>.
  33. Meyer L. 2016. Grafting to increase high tunnel tomato productivity in the Central United States. Master's thesis. Kansas University, Manhattan, KS. <http://hdl.handle.net/2097/32736>.
  34. Poudel R, Jumpponen A, Kennelly MM, Rivard CL, Gomez-Montano L, Garrett KA. 2019. Rootstocks shape the rhizosphere: Rhizosphere and endosphere bacterial communities in the grafted tomato system. *Appl Environ Microbiol* 85:e01765-18. <https://doi.org/10.1128/AEM.01765-18>.
  35. Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 58:267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
  36. Roguet A, Eren AM, Newton RJ, McLellan SL. 2018. Fecal source identification using random forest. *Microbiome* 6:185. <https://doi.org/10.1186/s40168-018-0568-3>.
  37. Belk A, Xu ZZ, Carter DO, Lynne A, Bucheli S, Knight R, Metcalf JL. 2018. Microbiome data accurately predicts the postmortem interval using random forest regression models. *Genes* 9:104. <https://doi.org/10.3390/genes9020104>.
  38. Greger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, Schadt CW. 2018. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome* 6:31. <https://doi.org/10.1186/s40168-018-0413-8>.
  39. Edwards JA, Santos-Medellin CM, Liechty ZS, Nguyen B, Lurie E, Eason S, Phillips G, Sundaresan V. 2018. Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol* 16:e2003862. <https://doi.org/10.1371/journal.pbio.2003862>.
  40. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. 2018. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc Natl Acad Sci U S A* 115:E1157–E1165. <https://doi.org/10.1073/pnas.1717617115>.
  41. Marasco R, Rolli E, Fusi M, Michoud G, Daffonchio D. 2018. Grapevine rootstocks shape underground bacterial microbiome and networking but not potential functionality. *Microbiome* 6:3. <https://doi.org/10.1186/s40168-017-0391-2>.
  42. Wagner MR, Lundberg DS, del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T. 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* 7:12151. <https://doi.org/10.1038/ncomms12151>.
  43. Herrera J, Poudel R, Bokati D. 2013. Assessment of root-associated fungal communities colonizing two species of tropical grasses reveals incongruence to fungal communities of North American native grasses. *Fungal Ecol* 6:65–69. <https://doi.org/10.1016/j.funeco.2012.08.002>.
  44. French E, Tran T, Iyer-Pascuzzi AS. 2020. Tomato genotype modulates selection and responses to root microbiota. *Phytobiomes J* 4:314–326. <https://doi.org/10.1094/PBIOMES-02-20-0020-R>.
  45. Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. 2013. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol* 64:807–838. <https://doi.org/10.1146/annurev-arplant-050312-120106>.
  46. Hol WHG, Bezemer TM, Biere A. 2013. Getting the ecology into interactions between plants and the plant growth-promoting bacterium *Pseudomonas fluorescens*. *Front Plant Sci* 4:81. <https://doi.org/10.3389/fpls.2013.00081>.
  47. Ortiz-Castro R, Contreras-Cornejo HA, Macías-Rodríguez L, López-Bucio J. 2009. The role of microbial signals in plant growth and development. *Plant Signal Behav* 4:701–712. <https://doi.org/10.4161/psb.4.8.9047>.
  48. Edwards J, Johnson C, Santos-Medellin C, Lurie E, Podshetty NK, Bhatnagar S, Eisen JA, Sundaresan V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A* 112:E911–E920. <https://doi.org/10.1073/pnas.1414592112>.
  49. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, del Rio TG, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tringe SG, Dangl JL. 2012. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488:86–90. <https://doi.org/10.1038/nature11237>.
  50. Wang P, Marsh EL, Ainsworth EA, Leakey ADB, Sheflin AM, Schachtman DP. 2017. Shifts in microbial communities in soil, rhizosphere and roots of two major crop systems under elevated CO<sub>2</sub> and O<sub>3</sub>. *Sci Rep* 7:15019. <https://doi.org/10.1038/s41598-017-14936-2>.
  51. Herrera J, Poudel R, Nebel KA, Collins SL. 2011. Precipitation increases the abundance of some groups of root-associated fungal endophytes in a semi-arid grassland. *Ecosphere* 2:1–14. <https://doi.org/10.1890/ES11-00001.1>.
  52. Bokati D, Herrera J, Poudel R. 2016. Soil influences colonization of root-associated fungal endophyte communities of maize, wheat, and their progenitors. *J Mycol* 2016:8062073. <https://doi.org/10.1155/2016/8062073>.
  53. Buddenhagen CE, Xing Y, Andrade-Piedra JL, Forbes GA, Kromann P, Navarrete I, Thomas-Sharma S, Choudhury RA, Andersen Onofre KF, Schulte-Geldermann E, Etherton BA, Plex Sulá AI, Garrett KA. 2022. Where to invest project efforts for greater benefit: a framework for management performance mapping with examples for potato seed health. *Phytopathology* 112: 1431–1443. <https://doi.org/10.1094/PHYTO-05-20-0202-R>.
  54. Niu B, Paulson JN, Zheng XQ, Kolter R. 2017. Simplified and representative bacterial community of maize roots. *Proc Natl Acad Sci U S A* 114: E2450–E2459. <https://doi.org/10.1073/pnas.1616148114>.
  55. Paredes SH, Gao TX, Law TF, Finkel OM, Mucyn T, Teixeira PJPL, Gonzalez IS, Feltcher ME, Powers MJ, Shank EA, Jones CD, Jovic V, Dangl JL, Castrillo G. 2018. Design of synthetic bacterial communities for predictable plant phenotypes. *PLoS Biol* 16:e2003962. <https://doi.org/10.1371/journal.pbio.2003962>.
  56. Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comp Biol* 11:1004226. <https://doi.org/10.1371/journal.pcbi.1004226>.
  57. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. *PLoS Comput Biol* 8:e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>.
  58. Tackmann J, Matias Rodrigues JF, von Mering C. 2019. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Syst* 9:286–296.e8. <https://doi.org/10.1016/j.cels.2019.08.002>.
  59. Faust K, Raes J. 2016. CoNet app: Inference of biological association networks using Cytoscape. *F1000Res* 5:1519. <https://doi.org/10.12688/f1000research.9050.2>.
  60. Coleman-Derr D, Desgarennes D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T, North G, Visel A, Partida-Martinez LP, Tringe SG. 2016. Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol* 209:798–811. <https://doi.org/10.1111/nph.13697>.
  61. Poli A, Lazzari A, Prigione V, Voyron S, Spadaro D, Varese GC. 2016. Influence of plant genotype on the cultivable fungi associated to tomato rhizosphere and roots in different soils. *Fungal Biol* 120:862–872. <https://doi.org/10.1016/j.funbio.2016.03.008>.
  62. Gottel NR, Castro HF, Kerley M, Yang ZM, Pelletier DA, Podar M, Karpinet T, Uberbacher E, Tuskan GA, Vilgalys R, Doktycz MJ, Schadt CW. 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. *Appl Environ Microbiol* 77:5934–5944. <https://doi.org/10.1128/AEM.05255-11>.
  63. Badri DV, Vivanco JM. 2009. Regulation and function of root exudates. *Plant Cell Environ* 32:666–681. <https://doi.org/10.1111/j.1365-3040.2009.01926.x>.

64. Dennis PG, Miller AJ, Hirsch PR. 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiol Ecol* 72:313–327. <https://doi.org/10.1111/j.1574-6941.2010.00860.x>.
65. van der Heijden MG, Schlaeppi K. 2015. Root surface as a frontier for plant microbiome research. *Proc Natl Acad Sci U S A* 112:2299–2300. <https://doi.org/10.1073/pnas.1500709112>.
66. Seitz VA, McGivern BB, Daly RA, Chaparro JM, Borton MA, Sheflin AM, Kresovich S, Shields L, Schipanski ME, Wrighton KC, Prenni JE. 2022. Variation in root exudate composition influences soil microbiome membership and function. *Appl Environ Microbiol* 88:e0022622. <https://doi.org/10.1128/aem.00226-22>.
67. Berg G, Opelt K, Zachow C, Lottmann J, Gotz M, Costa R, Smalla K. 2006. The rhizosphere effect on bacteria antagonistic towards the pathogenic fungus *Verticillium* differs depending on plant species and site. *Environ Microbiol Rep* 56:250–261. <https://doi.org/10.1111/j.1574-6941.2005.00025.x>.
68. Tian T, Reverdy A, She Q, Sun B, Chai Y. 2020. The role of rhizodeposits in shaping rhizomicrobiome. *FEMS Microbiol Ecol* 12:160–172. <https://doi.org/10.1111/1758-2229.12816>.
69. Reazin C, Baird R, Clark S, Jumpponen A. 2019. Chestnuts bred for blight resistance depart nursery with distinct fungal rhizobiomes. *Mycorrhiza* 29:313–324. <https://doi.org/10.1007/s00572-019-00897-z>.
70. Cardarelli M, Roupheal Y, Kyriacou MC, Colla G, Pane C. 2020. Augmenting the sustainability of vegetable cropping systems by configuring rootstock-dependent rhizomicrobiomes that support plant protection. *Agronomy* 10:1185. <https://doi.org/10.3390/agronomy10081185>.
71. Nasholm T, Huss-Danell K, Hogberg P. 2000. Uptake of organic nitrogen in the field by four agriculturally important plant species. *Ecology* 81:1155–1161. <https://doi.org/10.2307/177188>.
72. Reeve JR, Smith JL, Carpenter-Boggs L, Reganold JP. 2008. Soil-based cycling and differential uptake of amino acids by three species of strawberry (*Fragaria* spp.) plants. *Soil Biol Biochem* 40:2547–2552. <https://doi.org/10.1016/j.soilbio.2008.06.015>.
73. U'Ren N. 2007. Types, amounts, and possible functions of compounds released into the rhizosphere by soil-grown plants, p 1–21. *In* The rhizosphere. CRC Press, Boca Raton, FL.
74. Beattie GA. 2018. Metabolic coupling on roots. *Nat Microbiol* 3:396–397. <https://doi.org/10.1038/s41564-018-0139-1>.
75. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi SJ, Cho HJ, Karaoz U, Loque D, Bowen BP, Firestone MK, Northen TR, Brodie EL. 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat Microbiol* 3:470–480. <https://doi.org/10.1038/s41564-018-0129-3>.
76. Iannucci A, Canfora L, Nigro F, De Vita P, Beleggia R. 2021. Relationships between root morphology, root exudate compounds and rhizosphere microbial community in durum wheat. *Appl Soil Ecol* 158:103781. <https://doi.org/10.1016/j.apsoil.2020.103781>.
77. Warschefsky EJ, Klein LL, Frank MH, Chitwood DH, Londo JP, von Wettberg EJB, Miller AJ. 2016. Rootstocks: diversity, domestication, and impacts on shoot phenotypes. *Trends Plant Sci* 21:418–437. <https://doi.org/10.1016/j.tplants.2015.11.008>.
78. Malamy JE. 2005. Intrinsic and environmental response pathways that regulate root system architecture. *Plant Cell Environ* 28:67–77. <https://doi.org/10.1111/j.1365-3040.2005.01306.x>.
79. Song F, Bai F, Wang J, Wu L, Jiang Y, Pan Z. 2020. Influence of citrus scion/rootstock genotypes on arbuscular mycorrhizal community composition under controlled environment condition. *Plants* 9:901–916. <https://doi.org/10.3390/plants9070901>.
80. Cao X, Zhao D, Xu H, Huang R, Zeng J, Yu Z. 2018. Heterogeneity of interactions of microbial communities in regions of Taihu Lake with different nutrient loadings: a network analysis. *Sci Rep* 8:8890. <https://doi.org/10.1038/s41598-018-27172-z>.
81. Carlson CJ, Zipfel CM, Garnier R, Bansal S. 2019. Global estimates of mammalian viral diversity accounting for host sharing. *Nat Ecol Evol* 3:1070–1075. <https://doi.org/10.1038/s41559-019-0910-6>.
82. McGeachie MJ, Sordillo JE, Gibson T, Weinstock GM, Liu YY, Gold DR, Weiss ST, Litonjua A. 2016. Longitudinal prediction of the infant gut microbiome with dynamic Bayesian networks. *Sci Rep* 6:20359. <https://doi.org/10.1038/srep20359>.
83. Peñalver Bernabé B, Cralle L, Gilbert JA. 2018. Systems biology of the human microbiome. *Curr Opin Biotechnol* 51:146–153. <https://doi.org/10.1016/j.copbio.2018.01.018>.
84. Topçuoğlu BD, Lesniak NA, Ruffin MT, Wiens J, Schloss PD. 2020. A framework for effective application of machine learning to microbiome-based classification problems. *mBio* 11:e00434-20. <https://doi.org/10.1128/mBio.00434-20>.
85. Efron B. 2020. Prediction, estimation, and attribution. *J the American Statistical Association* 115:636–655. <https://doi.org/10.1080/01621459.2020.1762613>.
86. Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes: application to the identification of mycorrhizae and rusts. *Mol Ecol* 2:113–118. <https://doi.org/10.1111/j.1365-294x.1993.tb00005.x>.
87. Ihrmark K, Bodeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandstrom-Durling M, Clemmensen KE, Lindahl BD. 2012. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol Ecol* 82:666–677. <https://doi.org/10.1111/j.1574-6941.2012.01437.x>.
88. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
89. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.
90. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schöbeler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson K-H. 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* 22:5271–5277. <https://doi.org/10.1111/mec.12481>.
91. Kuhn M. 2008. Building predictive models in R using the caret package. *J Stat Soft* 28:1–26. <https://doi.org/10.18637/jss.v028.i05>.
92. Csárdi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJ. Complex Syst* 1695:1–9. <https://cran.r-project.org/web/packages/igraph/index.html>.
93. Guimerà R, Nunes Amaral LA. 2005. Functional cartography of complex metabolic networks. *Nature* 433:895–900. <https://doi.org/10.1038/nature03288>.
94. Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGinn D, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M, Szoecs E, Wagner H. 2018. *Vegan: Community ecology package*. R package version 2.4–6. <https://CRAN.R-project.org/package=vegan>.
95. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
96. R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
97. Bates D, Machler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. *J Stat Soft* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
98. Ferreira JA. 2007. The Benjamini-Hochberg method in the case of discrete test statistics. *Int J Biostat* 3:Article 11.